

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Sommersemester 2020

30.03.2020 - 19.06.2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 1

Stand: März 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim

# Einführung

## 1 Einleitung

Der Begriff *Statistik* wird häufig verwendet, oft ist aber nicht klar, was damit gemeint ist. Das liegt auch daran, dass der Begriff nicht eindeutig ist. Einerseits verstehen wir unter Statistik die Gesamtdarstellung einer konkret vorliegenden Datenmenge, etwa die Statistik der Verkehrstoten in einem Jahr oder die Statistik der gemeldeten Einwohner einer Gemeinde, in der eine gesamt vorliegende Datenmenge in Zahlen, Tabellen und Graphiken zusammengefasst wird. Andererseits ist *Statistik* aber auch die Gesamtheit aller Methode, die für die Gewinnung und Verarbeitung empirischer Informationen relevant sind.

Typische Beispiele für Anwendungsgebiete dieser statistischen Methodenlehre sind etwa politische Umfragen, bei der aus der Befragung weniger ausgewählter Personen auf das politische Stimmungsbild der Gesamtbevölkerung geschlossen werden soll. Ähnlich verhält es sich mit der Erforschung des Konsumentenverhaltens, etwa durch die GfK, oder die Bestimmung der Marktanteile der diversen Fernseh- oder Rundfunksender bzw. -programme, die ebenfalls auf Stichproben beruhen.

Die statistische Methodenlehre beschäftigt sich dementsprechend mit den folgenden drei Grundfragen:

- Datenbeschreibung (*Deskription*).
- Datenerhebung (*Exploration*).
- Schlußfolgerung (*Induktion*).

Die vorliegende Vorlesung basiert in Teilen auf

FAHRMEIR L., C. HEUMANN, R. KÜNSTLER, I. PIGEOT, G. TUTZ: *Statistik: Der Weg zur Datenanalyse*. 8. Auflage, Springer Verlag, Heidelberg, Dordrecht, London, New York, 2016.

Bei einigen Themen weicht die Vorgehensweise allerdings von diesem Buch ab (speziell im Rahmen der Wahrscheinlichkeitsrechnung). Diese Abschnitte werden jedoch in diesem Skriptum besonders ausführlich behandelt.

## 2 Deskriptive Statistik

### 2.1 Grundbegriffe der Datenerhebung

Bei der Datenerhebung stellt sich zunächst die Grundfrage, was von wem erfasst werden soll. Hierfür benötigen wir einige Begrifflichkeiten:

**Definition 2.1.** (*Grundbegriffe*)

- Ein *Merkmalsträger* oder eine *statistische Einheit* ist ein Objekt, an dem interessierende Größen erfasst werden.
- Die *Grundgesamtheit* oder *statistische Masse* ist die Menge aller für die Fragestellung relevanten Merkmalsträger.
- Eine *Stichprobe* ist die tatsächlich untersuchte Teilmenge der Grundgesamtheit.
- Ein *Merkmal* oder eine *Variable* ist die interessierende Größe.
- Eine *Merkmalsausprägung* ist der konkrete Wert eines Merkmals bei einem konkreten Merkmalsträger.

Die Grundgesamtheit wird üblicherweise mit  $\Omega$  bezeichnet. Dabei unterscheiden wir zwischen endlichen Grundgesamtheiten,  $\Omega = \{\omega_1, \dots, \omega_n\}$  und unendlichen Grundgesamtheiten, also etwa  $\Omega = \mathbb{N}$  (abzählbar unendlich) oder  $\Omega = \mathbb{R}$  (überabzählbar unendlich).

**Beispiel 2.1.**

- a) Untersuchen wir die Entwicklung der Verkaufszahlen für Klopapier (in Rollen), so ist eine hierfür naheliegende Grundgesamtheit die Menge aller Tage (ab einem Startzeitpunkt), also eine unendliche, aber abzählbare Grundgesamtheit.
- b) Untersuchen wir die Temperaturkurve an einem Ort, so ist eine naheliegende Grundgesamtheit die Zeitskala (ab einem bestimmten Startzeitpunkt), also eine unendliche und überabzählbare Grundgesamtheit.

Ferner unterscheiden wir Grundgesamtheiten auch nach zwei Grundtypen:

- Bestandsmassen: Grundgesamtheit zu einem fixierten Zeitpunkt.
- Bewegungsmassen: Grundgesamtheit über eine Zeitperiode.

Für die Datenanalyse ist es wichtig, die Merkmale hinsichtlich gewisser Eigenschaften ihrer Ausprägungen zu klassifizieren.

**Definition 2.2.** (*Skalentypen*)

- Ein Merkmal heißt *nominalskaliert*, wenn die Merkmalsausprägungen Namen oder Kategorien sind, die den Merkmalen zugeordnet werden können, und diese Ausprägungen zwar unterschieden werden können aber keine natürliche Rangfolge aufweisen.
- Ein Merkmal heißt *ordinalskaliert*, wenn die Merkmalsausprägungen geordnet und in eine natürliche Reihenfolge gebracht werden können, ohne dass aber die Abstände zwischen den Merkmalsausprägungen interpretiert werden können.
- Ein Merkmal heißt *kardinalskaliert* oder *metrisch skaliert*, wenn die Merkmalsausprägungen in eine Rangordnung gebracht werden können, und zusätzlich noch bestimmt werden kann, in welchem Ausmaß sich je zwei verschiedene Merkmalsausprägungen unterscheiden.

**Bemerkung 2.1.** Bei Kardinalskalen unterscheiden wir noch nach *Intervallskalen*, bei denen die Abstände zwischen zwei Ausprägungen verglichen werden können und *Verhältnisskalen* bei denen zusätzlich noch ein natürlicher Nullpunkt dazukommt und damit auch die Quotienten von Merkmalsausprägungen sinnvoll interpretiert werden können.

**Beispiel 2.2.** In einem Fragebogen werden folgende Daten erhoben:

- Geschlecht (1 = männlich, 2 = weiblich).
- Geburtsjahr
- Körpergröße
- Nettohaushaltseinkommen  $NHE$  pro Monat (1 :  $NHE < 1000 N$ , 2:  $1000 N \leq NHE \leq 3000 N$ , 3 :  $NHE > 3000 N$ ).

Das Merkmal *Geschlecht* ist nominalskaliert.

Das Merkmal *Geburtszeitpunkt* ist kardinalskaliert. Allerdings handelt es sich nicht um eine Verhältnisskala, da auf der Zeitachse der natürliche Nullpunkt fehlt.

Das Merkmal *Körpergröße* ist kardinalskaliert. Bei der Körpergröße handelt es sich auch um eine Verhältnisskala, da ein natürlicher Nullpunkt  $0\text{ cm}$  vorhanden ist, und die beiden

Körpergrößen  $k_1 = 100\text{ cm}$  und  $k_2 = 200\text{ cm}$  auch die Interpretation  $k_2$  ist doppelt so groß wie  $k_1$  zulassen.

Das Nettohaushaltseinkommen (wie hier erhoben) ist ordinalskaliert. Würde das Nettohaushaltseinkommen dagegen als tatsächlicher Eurowert erhoben, so wäre es verhältnisskaliert.

**Bemerkung 2.2.** Nominalskalen sind invariant gegenüber eineindeutigen Abbildungen. Ordinalskalen sind invariant gegenüber streng monotonen Abbildungen. Intervallskalen sind invariant gegenüber linearen Abbildungen und Verhältnisskalen sind invariant gegenüber linearen Abbildungen, die den Nullpunkt erhalten.

**Bemerkung 2.3.** Eine andere Typisierung ist die Einteilung von Merkmalen in diskrete Merkmale, bei denen nur endlich viele oder abzählbar unendlich viele Merkmalsausprägungen möglich sind, und stetige Merkmale, bei denen die Merkmalsausprägung eine stetige Funktion auf der Grundgesamtheit definiert. Hier sind jedoch auch Zwischenstufen möglich.

**Definition 2.3.** Ein Merkmal heißt *kategorisiert* oder *klassiert* oder *gruppiert*, wenn die eigentlichen Merkmalsausprägungen zu Klassen oder Gruppen zusammengefasst werden.

**Beispiel 2.3.** Das Nettohaushaltseinkommen aus Beispiel 2.2 ist ein kategorisiertes Merkmal.

**Bemerkung 2.4.** Bei Merkmalen unterscheiden wir auch nach *qualitativ* (Merkmalen die nur endlich viele Ausprägungen haben und höchstens ordinalskaliert sind) und *quantitativ* (Merkmale, deren Ausprägungen eine Intensität bzw. ein Ausmaß wiedergeben).

Eine wesentliche Grundfrage der Statistik ist nun natürlich die der Datengewinnung. Statistische Verfahren werden ja üblicherweise dann eingesetzt, wenn Forschungsfragen empirisch überprüft werden sollen und die Einflüsse bestimmter Größen auf interessierende Merkmale untersucht werden sollen, etwa die Wirksamkeit von Medikamenten oder der Einfluß der Ausbildung auf das Einkommen.

Um eine solche Beziehungsstruktur empirisch zu überprüfen, ist es notwendig, die vermuteten Beziehungen zwischen Einflussgrößen und Merkmalen zu präzisieren und vor allem die Einfluss- und die Zielvariablen exakt festzulegen und zu messen.

Als erste Stufe ist also ein genauer Versuchsplan zu erstellen, in dem festgelegt wird, welches Ziel mit der Studien erreicht werden soll, wie dieses Ziel erreicht werden kann und welche statistischen Methoden zum Einsatz kommen. Dazu ist eine Vielzahl von einzelnen Aspekten zu berücksichtigen, angefangen von der Festlegung der relevanten

Grundgesamtheit bis hin zur Erfassung und Kontrolle von Störvariablen (etwa durch Homogenisierung, Randomisierung, durch Parallelisierung oder durch die Auswahl eines geeigneten statistischen Modells).

Ist nun der Versuchplan festgelegt, so sind als nächstes die Daten für die Studie durch eine Erhebung zu gewinnen. Im Idealfall ist eine Vollerhebung möglich, in der von allen Merkmalsträger die Werte der Einfluss- und der Zielgrößen erfasst werden. Ist jedoch eine Vollerhebung nicht möglich, so greift man in der Regel auf das Ziehen einer *Stichprobe* zurück, bei der nur ein Teil der Grundgesamtheit auf die Merkmalsausprägung untersucht werden soll.

Bei einer *einfachen Stichprobe* werden Teilmengen der Grundgesamtheit so erhoben, dass jede dieser Teilmengen dieselbe Wahrscheinlichkeit besitzt, gezogen zu werden. Notwendig ist dazu, dass die Grundgesamtheit numerierbar ist und (zumindest theoretisch) als Liste vorliegt. Zur Realisierung wird dann jede Nummer auf ein Los geschrieben und in eine Urne gesteckt. Dann wird die Stichprobe blind aus dieser Urne gezogen.

In der Praxis ist das natürlich schwer umsetzbar. Deshalb wird hier häufig systematisch gezogen, indem etwa für eine Haushaltsuntersuchung jeder elfte Haushalt in einem Straßenzug ausgewählt wird oder für eine Telefonbefragung jede siebte Telefonnummer aus einem ausgewählten Abschnitt der Telefonbuchs.

Eine andere Variante ist, die Grundgesamtheit zunächst in disjunkte Teilmengen zu zerlegen und dann aus jeder Teilmenge eine Zufallsstichprobe zu ziehen (*geschichtete Zufallsstichprobe* oder auch *Klumpenstichprobe*, falls die Grundgesamtheit in natürlicher Weise in disjunkte Teile zerfällt).

Häufig sind die Auswahlverfahren für Stichproben auch mehrstufig (erst werden zufällig einige Gemeinden ausgewählt, dann in jeder Gemeinde einige Straßen, dann in jeder Straße einige Probanden). Oft sind dabei auch Quoten zu berücksichtigen.

## 2.2 Eindimensionale deskriptive Statistik

In diesem Abschnitt befassen wir uns mit den grundlegenden statistischen Methoden zur Verarbeitung *univariater Daten* oder *eindimensionaler Daten*, die aus der Beobachtung eines einzigen Merkmals hervorgehen, dessen Ausprägungen wir in geeigneter Weise (in der Regel durch reelle Zahlen) darstellen.

Wir untersuchen ein Merkmal  $X$ . Dazu machen wir eine Erhebung von  $n$  Merkmalsträgern dieses Merkmals und beobachten an diesen die Merkmalsausprägungen  $x_1, \dots, x_n \in \mathbb{R}$  (*Roh- oder Primärdaten, Urliste*).

**Beispiel 2.4.** In einer Vorlesung wird das Geschlecht der Vorlesungsteilnehmer erhoben

(1 = weiblich, 2 = männlich und 3 = divers). Die Erhebung ergibt die Urliste

$$(2, 2, 2, 1, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2)$$

Die Merkmalswerte der Urliste seien  $a_1, \dots, a_k$ , wobei wir annehmen, dass  $a_1 < a_2 < \dots < a_k$ .

**Definition 2.4.** (*Häufigkeiten*)

- Die *absolute Häufigkeit*  $h_l = h(a_l)$  der Ausprägung  $a_l$  ist die Anzahl der  $i$  mit  $x_i = a_l$ ,

$$h(a_l) = |\{i \in \{1, \dots, n\} \mid x_i = a_l\}|$$

- Die *relative Häufigkeit*  $f(a_l)$  von  $a_l$  ist  $f(a_l) = \frac{h(a_l)}{n}$ .
- $(h(a_1), \dots, h(a_k))$  heißt *absolute Häufigkeitsverteilung* der Stichprobe.
- $(f(a_1), \dots, f(a_k))$  heißt *relative Häufigkeitsverteilung* der Stichprobe.

**Beispiel 2.5.** Bei der Notenerhebung aus Beispiel 2.4 treten die Merkmalsausprägungen  $a_1 = 1, a_2 = 2, a_3 = 3, a_4 = 4$  und  $a_5 = 5$  auf, und zwar mit den folgenden Häufigkeiten

$$h(1) = 3, h(2) = 6, h(3) = 1, h(4) = 3, h(5) = 2$$

Die Häufigkeitsverteilung ist

$$H = (3, 6, 1, 3, 2)$$

Die zugehörige relative Häufigkeitsverteilung ist

$$F = \left( \frac{3}{15}, \frac{6}{15}, \frac{1}{15}, \frac{3}{15}, \frac{2}{15} \right)$$

**Bemerkung 2.5.** Häufigkeiten werden üblicherweise in Tabellen, in Stabdiagrammen oder in Kreissektordiagrammen dargestellt.

**Bemerkung 2.6.** Es ist  $\sum_{l=1}^k h(a_l) = n$ ,  $\sum_{l=1}^k f(a_l) = 1$ .

**Bemerkung 2.7.** Bei stetigen Merkmalen ist es häufig sinnvoll, die Merkmalsausprägungen in Klassen zusammenzufassen (Histogramme), wie wir das etwa beim Haushaltseinkommen in Beispiel 2.2 gemacht haben. Die absoluten Häufigkeiten sind in diesem Fall also Klassenhäufigkeiten.



Häufig sind nicht die einzelnen Ausprägungen für sich und ihre Häufigkeiten gefragt (vor allem, wenn das viele sind), sondern wir interessieren uns dafür, wie oft die Merkmalsausprägung unter einem kritischen Wert  $x$  bleibt.

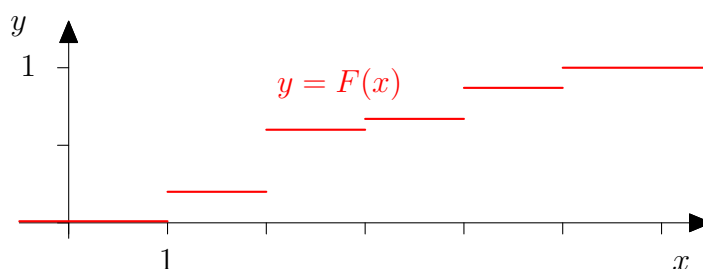
**Definition 2.5.** (*kumulierte Häufigkeiten*)

- $H(x) = \sum_{a_l \leq x} h(a_l)$  heißt *absolute kumulierte Häufigkeit* der Stichprobe zum Wert  $x$ .
- $F(x) = \sum_{a_l \leq x} f(a_l)$  heißt *relative kumulierte Häufigkeit* der Stichprobe zum Wert  $x$ .

**Beispiel 2.6.** Bei der Notenerhebung aus Beispiel 2.4 treten die folgenden kumulierten Häufigkeiten auf

	$x < 1$	$1 \leq x < 2$	$2 \leq x < 3$	$3 \leq x < 4$	$4 \leq x < 5$	$x \geq 5$
$H(x)$	0	3	9	10	13	15
$F(x)$	0	$\frac{3}{15}$	$\frac{9}{15}$	$\frac{10}{15}$	$\frac{13}{15}$	1

Eine Skizze der kumulierten relativen Häufigkeit hat etwa folgende Gestalt.



**Bemerkung 2.8.** Es gilt  $F(x) = \frac{H(x)}{n}$ .

**Bemerkung 2.9.** Bei  $F(x)$  spricht man auch von der empirischen Verteilungsfunktion.

Zur Beschreibung von Verteilungen benutzen wir die sogenannten *Lagemaße*, die in einem geeigneten Sinn das Zentrum der Verteilung beschreiben.

Wir betrachten zunächst ein (mindestens) nominal skaliertes Merkmal  $X$ .

**Definition 2.6.** Der *Modus*  $x_{mod}$  der Stichprobe ist der Merkmalswert mit der größten Häufigkeit,

$$x_{mod} = a_l \iff h(a_l) \geq h(a_j) \quad \forall j \neq l$$

**Bemerkung 2.10.** Der Modus muss nicht eindeutig sein, es können mehrere Werte mit der gleichen (größten) Häufigkeit auftreten.

**Beispiel 2.7.** Bei der Notenerhebung aus Beispiel 2.4 ist  $x_{\text{mod}} = 2$ .

Wir betrachten nun ein mindestens ordinal skaliertes Merkmal  $X$ , und wir nehmen an, dass schon die  $x_i$  angeordnet sind, also  $x_1 \leq x_2 \leq \dots \leq x_n$ .

**Definition 2.7.** Der *Median*  $x_{\text{med}}$  ist definiert wie folgt

- Falls  $n$  ungerade ist, so ist

$$x_{\text{med}} = x_{\frac{n+1}{2}}$$

- Falls  $n$  gerade ist, so ist

$$x_{\text{med}} = \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$

**Bemerkung 2.11.** Es gilt

- 50% der Daten der Stichprobe sind kleiner oder gleich  $x_{\text{med}}$ .
- 50% der Daten der Stichprobe sind größer oder gleich  $x_{\text{med}}$ .

**Beispiel 2.8.** Bei der Notenerhebung aus Beispiel 2.4 ist  $x_{\text{med}} = 2$ .

**Definition 2.8.** Für einen Wert  $p \in ]0, 1[$  ist ein  $p$ -Quantil  $x_p$  der Stichprobe eine Merkmalsausprägung  $x_p$ , für die gilt

$$\frac{|\{i \in \{1, \dots, n\} \mid x_i \leq x_p\}|}{n} \geq p \quad \text{und} \quad \frac{|\{i \in \{1, \dots, n\} \mid x_i \geq x_p\}|}{n} \geq 1 - p$$

**Bemerkung 2.12.** Das  $p$ -Quantil ist so gewählt, dass mindestens  $p \cdot n$  der Merkmalswerte kleiner oder gleich  $x_p$  sind und mindestens  $(1 - p) \cdot n$  der Merkmalswerte größer oder gleich  $x_p$  sind.

Ein  $p$ -Quantil existiert immer. Wenn  $n \cdot p \notin \mathbb{Z}$ , so ist es eindeutig bestimmt. Falls  $n \cdot p \in \mathbb{Z}$ , so ist jede Zahl  $x_p \in [x_{np}, x_{np+1}]$  ein  $p$ -Quantil. Es ist üblich, in diesem Fall

$$x_p = \frac{1}{2} \cdot (x_{np} + x_{np+1})$$

zu wählen.

**Bemerkung 2.13.** Spezielle  $p$ -Quantile sind die Quartile,

$$x_{0,25} = 25\% \text{-Quantil} = \text{unteres Quartil.}$$

$x_{0,75} = 75\%$ -Quantil = oberes Quartil.

Das "mittlere Quartil" einer Stichprobe ist der Median der Stichprobe.

**Bemerkung 2.14.** Die *Fünf-Punkte-Zusammenfassung* einer Verteilung besteht aus den Werten

$$x_{\min}, x_{0,25}, x_{\text{med}}, x_{0,75}, x_{\max}$$

Wir betrachten nun ein metrisch skaliertes Merkmal  $X$ .

**Definition 2.9.** Das **arithmetische Mittel**  $\bar{x}$  der Stichprobe ist definiert als

$$\bar{x} = \frac{1}{n} \cdot (x_1 + \dots + x_n) = \frac{1}{n} \sum_{l=1}^n x_l$$

**Bemerkung 2.15.** Das arithmetische Mittel berechnet sich aus den relativen Häufigkeiten als

$$\bar{x} = a_1 f(a_1) + \dots + a_k f(a_k) = \sum_{l=1}^n a_l f(a_l)$$

**Beispiel 2.9.** Wir betrachten das Merkmal *Körpergröße* von fünf Personen

$k$	1	2	3	4	5
$x_k$	176	182	190	182	174

Dann ist

$$\bar{x} = \frac{1}{5} \cdot (176 + 182 + 190 + 182 + 174) = 180.8$$

Beachten Sie, dass es für das arithmetische Mittel nicht nötig ist, die Merkmalsausprägungen der Größe nach aufsteigend zu sortieren. Der Median dagegen kann nicht direkt aus der gegebenen Form abgelesen werden, denn

$$x_{\text{med}} = 182 \neq x_3 = 190$$

**Beispiel 2.10.** Wir betrachten das Merkmal *Körpergewicht* von sechs Personen

$k$	1	2	3	4	5	6
$x_k$	84	72	69	80	94	81

Dann ist

$$\bar{x} = \frac{1}{6} \cdot (84 + 72 + 69 + 80 + 94 + 81) = 80.0$$

**Bemerkung 2.16.** Das arithmetische Mittel reagiert relativ empfindlich auf Ausreißer: Betritt Bill Gates ein Restaurant, so ist jeder in dem Restaurant, einschließlich Köche und Bedienungen im arithmetischen Mittel ein Multimillionär.

**Bemerkung 2.17.** Das arithmetische Mittel ist der Wert, von dem alle Merkmalswerte die geringste mittlere quadratische Abweichung haben, also dasjenige  $\mu$ , für das gilt

$$\sum_{i=1}^n (x_i - \mu)^2 \quad \text{ist minimal}$$

**Bemerkung 2.18.** Neben arithmetischem Mittel und Median gibt es auch noch weitere Mittelwerte, die bei der Betrachtung von Daten eine Rolle spielen.

Für positive Zahlen  $a_1, \dots, a_n$  heißt

$$\bar{a}_{\text{geom}} := \sqrt[n]{a_1 \cdots a_n}$$

das **geometrische Mittel** von  $a_1, \dots, a_n$  und

$$\bar{a}_{\text{harm}} := \frac{1}{\frac{1}{n} \sum_{k=1}^n \frac{1}{a_k}}$$

da **harmonische Mittel** von  $a_1, \dots, a_n$ .

Neben den Lageparametern interessieren uns aber auch noch die Streuungsparameter der Stichprobe, also Parameter, die angeben, wie weit die Stichprobenwerte von Mittelwerten abweichen. Wir beschränken uns hierzu auf metrische Merkmale und Abweichungen vom arithmetischen Mittel.

**Definition 2.10.** (*Streuungsparameter*)

Die *Varianz*  $\tilde{s}^2$  der Stichprobenwerte ist erklärt als

$$\tilde{s}^2 = \frac{1}{n} \cdot ((x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2) = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

Die *Standardabweichung*  $\tilde{s}$  der Stichprobe ist erklärt als

$$\tilde{s} = \sqrt{\tilde{s}^2} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

**Bemerkung 2.19.** Die Varianz ist das Mittel der quadratischen Abweichungen der Merkmalswerte vom arithmetischen Mittel.

**Definition 2.11.** Die *Stichprobenvarianz*  $s^2$  der Stichprobe ist definiert als

$$s^2 = \frac{1}{n-1} \cdot ((x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2) = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

**Bemerkung 2.20.** Die Stichprobenvarianz unterscheidet sich von der Varianz im Normierungsfaktor. Eine Begründung für diesen anderen (zunächst seltsam erscheinenden) Normierungsfaktor werden wir in der induktiven Statistik geben.

**Beispiel 2.11.** Wir betrachten wieder das Merkmal *Körpergröße* von fünf Personen

$k$	1	2	3	4	5
$x_k$	176	182	190	182	174

von dem wir schon wissen, das  $\bar{x} = 180.8$ .

Dann ist

$$\begin{aligned}
 \tilde{s}^2 &= \frac{1}{5} \cdot (((176 - 180.8)^2 + (182 - 180.8)^2 + (190 - 180.8)^2 \\
 &\quad + (182 - 180.8)^2 + (174 - 180.8)^2) \\
 &= \frac{1}{5} \cdot 156.8 \\
 &= 31.36
 \end{aligned}$$

und

$$s^2 = \frac{1}{4} \cdot 156.8 = 39.2$$

Ferner ist

$$\tilde{s} = 5.6, \quad s = 6.26$$

**Beispiel 2.12.** Wir betrachten das Merkmal *Körpergewicht* von sechs Personen

$k$	1	2	3	4	5	6
$x_k$	84	72	69	80	94	81

Dann ist

$$\begin{aligned}
 \tilde{s}^2 &= \frac{1}{6} \cdot (((84 - 80)^2 + (72 - 80)^2 + (69 - 80)^2 \\
 &\quad + (80 - 80)^2 + (94 - 80)^2 + (81 - 80)^2) \\
 &= \frac{1}{6} \cdot 398
 \end{aligned}$$

und

$$s^2 = \frac{1}{5} \cdot 398 = 79.6$$

Ferner ist

$$\tilde{s} = 8.14, \quad s = 8.92$$

**Satz 2.1.** (*Verschiebungssatz*)

Für jedes  $c \in \mathbb{R}$  gilt

$$\frac{1}{n} \cdot \sum_{i=1}^n (x_i - c)^2 = \tilde{s}^2 + (\bar{x} - c)^2$$

Speziell gilt also

$$\tilde{s}^2 = \frac{1}{n} \cdot \sum_{i=1}^n x_i^2 - \bar{x}^2$$

**Beweis:** Wir haben

$$\begin{aligned} \sum_{i=1}^n (x_i - c)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - c)^2 \\ &= \sum_{i=1}^n ((x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - c) + (\bar{x} - c)^2) \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + 2(\bar{x} - c) \cdot \sum_{i=1}^n (x_i - \bar{x}) + \sum_{i=1}^n (\bar{x} - c)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (\bar{x} - c)^2 \end{aligned}$$

wobei wir ausgenutzt haben, dass

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n \cdot \bar{x} - n \cdot \bar{x} = 0$$

Wie wir gesehen haben, sind metrisch skalierte Merkmale invariant unter linearen Abbildungen. Auch ihre Parameter können unter diesen Transformationen kontrolliert werden:

**Satz 2.2.** (*Transformationssatz*)

Ist  $Y = aX + b$  (also ist  $y_i = ax_i + b$  für alle  $i$ ), so gilt

$$\bar{y} = a \cdot \bar{x} + b, \quad \tilde{s}_y = |a| \cdot \tilde{s}_x$$

Der Nachweis ist eine einfache Rechnung, die wir dem Leser überlassen.

**Bemerkung 2.21.** Die Lage- und Verteilungsparameter geben noch kein vollständiges Bild der Verteilung der Ausprägungen eines Merkmals ab. Ein wichtiger Faktor ist auch, in welchen Teilen sich die Werte konzentrieren.

Eine Häufigkeitsverteilung heißt **symmetrisch**, wenn die Häufigkeiten symmetrisch um den Mittelwert angeordnet sind. Andernfalls heißt sie **schief**. Dabei heißt die Verteilung **linkssteil**, wenn der überwiegende Anteil der Daten auf der linken Seite konzentriert ist, und **rechtssteil**, wenn der überwiegende Anteil der Daten auf der rechten Seite konzentriert ist.

Die Verteilung

$n$	1	2	3	4	5	6	7
$a_n$	1	2	3	4	3	2	1

ist symmetrisch, die Verteilung

$n$	1	2	3	4	5	6	7
$a_n$	5	6	5	4	3	2	1

ist linkssteil und die Verteilung

$n$	1	2	3	4	5	6	7
$a_n$	1	2	3	4	5	6	5

ist rechtssteil.

Bei empirischen Verteilungen ist eine exakte Symmetrie praktisch nicht erreichbar. Deshalb spricht man auch von symmetrischen Häufigkeitsverteilungen, wenn die Werte annähernd symmetrisch um eine Symmetrieachse verteilt sind.

Eine speziell für Fragen der Wirtschaftswissenschaften wichtige Verteilung ist die Gleichverteilung. Wenn etwa mehrere Unternehmen ein Produkt anbieten, so ist es für die Wettbewerbssituation von entscheidender Bedeutung, wie gleichmäßig die Marktanteile auf die Unternehmen verteilt sind. Eine andere Situation, in der diese Frage eine Rolle spielt, ist die der Einkommensverteilung (Gerechtigkeitsdebatte). Wie wir schon gesehen haben, wird der Tellerwäscher eines Restaurants dadurch zum Millionär (zumindest im arithmetischen Mittel), wenn ein Milliardär das Restaurant betritt. Konkret hilft das jedoch im Tellerwäscher nicht weiter. Daher interessiert man sich bei diesbezüglichen Untersuchungen auch stark für die Frage, wie weit eine gegebene Häufigkeitsverteilung von der Gleichverteilung abweicht.

Wir betrachten dazu ein Merkmal  $X$  mit Merkmalsausprägungen  $x_1, \dots, x_n$ , wobei wir bereits annehmen, dass die Merkmalsausprägungen angeordnet sind, dass also gilt

$$x_1 \leq x_2 \leq \dots \leq x_{n-1} \leq x_n$$

und dass die Merkmalsausprägungen nicht-negativ sind ( $x_i \geq 0$ ). Für  $k = 1, \dots, n$  setzen wir:

$$u_k = \frac{k}{n}$$

$$v_k = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i}$$

und für  $k = 0$ :

$$u_0 = 0, \quad v_0 = 0$$

**Definition 2.12.** Der Polygonzug, der die Punkte

$$(0, 0) = (u_0, v_0), (u_1, v_1), \dots, (u_{n-1}, v_{n-1}), (u_n, v_n) = (1, 1)$$

miteinander verbindet, heißt **Lorenzkurve** zu den Merkmalsausprägungen  $x_1, \dots, x_n$ .

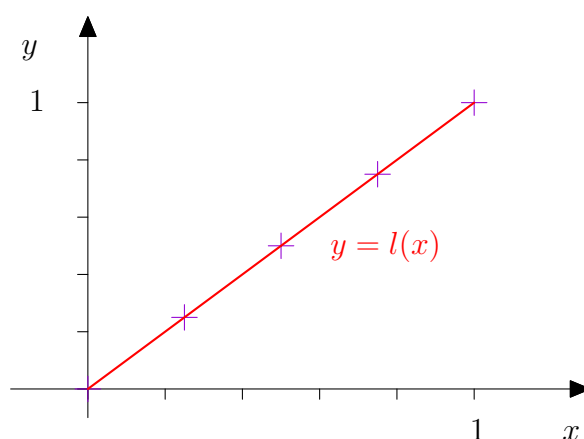
**Beispiel 2.13.** Wir betrachten eine Marktsituation in der sich die Umsätze (in Millionen €) auf vier Unternehmen wie folgt aufteilen:

Unternehmen	1	2	3	4
Umsatz	25	25	25	25

Wir haben also eine vollkommene Gleichverteilung. Es gilt

$k$	1	2	3	4
$u_k$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$v_k$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1

und wir erhalten als Lorenzkurve



also die Gerade durch die Punkte  $(0, 0)$  und  $(1, 1)$ .

Betrachten wir dagegen eine Aufteilung der Umsätze wie folgt

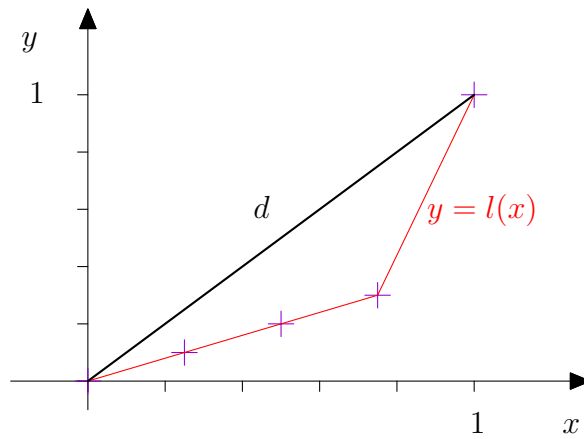
Unternehmen	1	2	3	4
Umsatz	10	10	10	70

also eine starke Konzentration auf Unternehmen 4, so gilt

$k$	1	2	3	4
$u_k$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$v_k$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	1

und wir erhalten als Lorenzkurve





In diesem Fall liegt die Lorenzkurve also deutlich unterhalb der Geraden, die  $(0, 0)$  und  $(1, 1)$  verbindet (die wir bei Gleichverteilung erhalten haben).

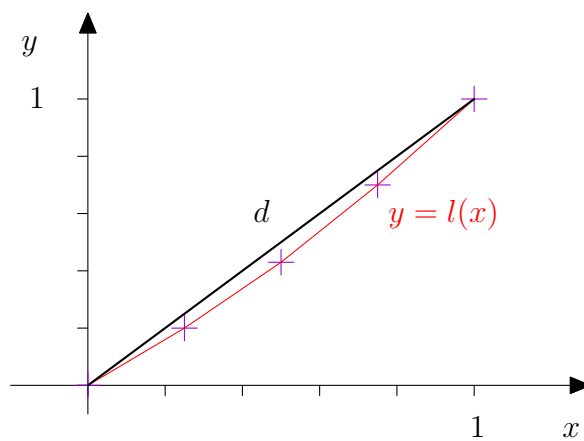
Betrachten wir schließlich noch eine Aufteilung der Umsätze wie folgt

Unternehmen	1	2	3	4
Umsatz	20	23	27	30

also eine relativ gleichmäßige Verteilung mit einer leichten Konzentration in Richtung von Unternehmen 3 und 4, so gilt

$k$	1	2	3	4
$u_k$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$v_k$	$\frac{20}{100}$	$\frac{43}{100}$	$\frac{70}{100}$	1

und wir erhalten als Lorenzkurve



In diesem Fall liegt die Lorenzkurve also ebenfalls unterhalb der Geraden, die  $(0, 0)$  und  $(1, 1)$  verbindet (die wir bei Gleichverteilung erhalten haben), jedoch deutlich weniger ausgeprägt als in der zweiten Situation.

Beispiel 2.13 deutet darauf hin, dass die Abweichung der Lorenzkurve von der Geraden, die  $(0, 0)$  und  $(1, 1)$  verbindet (also von der Diagonalen), ein Maß für die Abweichung einer Häufigkeitsverteilung von der Gleichverteilung ist.

**Definition 2.13.** Die Größe

$$G = \frac{\text{Fläche zwischen Diagonale und Lorenzkurve}}{\text{Fläche zwischen Diagonale und } u\text{-Achse}}$$

heißt **Gini-Koeffizient** zu den Merkmalsausprägungen  $x_1, \dots, x_n$ .

Für  $n \geq 2$  ist der **normierte Gini-Koeffizient**  $G^*$  definiert als

$$G^* = \frac{n}{n-1} \cdot G$$

**Bemerkung 2.22.** Es gilt

$$G = 2 \cdot \text{Fläche zwischen Diagonale und Lorenzkurve}$$

Der Ginikoeffizient kann unmittelbar aus den Daten  $(x_1, \dots, x_n)$  berechnet werden:

**Satz 2.3.** *Es gilt*

$$G = \frac{2 \cdot \sum_{k=1}^n k \cdot x_k}{n \cdot \sum_{k=1}^n x_k} - \frac{n+1}{n}$$

**Beweis:**

Der Gini-Koeffizient ist definiert als

$$G = \frac{\text{Fläche zwischen Diagonale und Lorenzkurve}}{\text{Fläche zwischen Diagonale und } u\text{-Achse}}$$

Dabei gilt

$$\text{Fläche zwischen Diagonale und } u\text{-Achse} = \frac{1}{2}$$

Die Lorenzkurve ist der Polygonzug durch die Punkte  $(u_k, v_k)$ , wobei

$$\begin{aligned} u_k &= \frac{k}{n} \\ v_k &= \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i} \end{aligned}$$

und für  $k = 0$ :

$$u_0 = 0, \quad v_0 = 0$$

Die Fläche  $F_k$  unterhalb der Lorenzkurve zwischen  $u_k$  und  $u_{k+1}$  ermittelt sich als

$$\begin{aligned} F_k &= (u_{k+1} - u_k) \cdot \frac{v_{k+1} + v_k}{2} \\ &= \frac{1}{n \cdot \sum_{i=1}^n x_i} \cdot \left( \sum_{i=1}^k x_i + \frac{x_{k+1}}{2} \right) \end{aligned}$$

und daher ermittelt sich die Gesamtfläche  $F$  unterhalb der Lorenzkurve als

$$\begin{aligned} F &= \sum_{k=0}^{n-1} F_k \\ &= \sum_{k=0}^{n-1} \frac{1}{n \cdot \sum_{i=1}^n x_i} \cdot \left( \sum_{i=1}^k x_i + \frac{x_{k+1}}{2} \right) \\ &= \frac{1}{n \cdot \sum_{i=1}^n x_i} \sum_{k=0}^{n-1} \left( \sum_{i=1}^k x_i + \frac{x_{k+1}}{2} \right) \\ &= \frac{1}{n \cdot \sum_{i=1}^n x_i} \sum_{k=0}^{n-1} k \cdot x_k + \frac{1}{n \cdot \sum_{i=1}^n x_i} \sum_{k=0}^{n-1} \frac{x_{k+1}}{2} \\ &= \frac{1}{n \cdot \sum_{i=1}^n x_i} \sum_{k=0}^{n-1} (n - k) \cdot x_k + \frac{1}{2n} \\ &= 1 + \frac{1}{2n} - \frac{1}{n \cdot \sum_{i=1}^n x_i} \sum_{k=0}^{n-1} k \cdot x_k \end{aligned}$$

Damit gilt für die Fläche  $A$  zwischen Diagonale und Lorenzkurve

$$\begin{aligned} A &= \frac{1}{2} - 1 - \frac{1}{2n} + \frac{1}{n \cdot \sum_{i=1}^n x_i} \sum_{k=0}^{n-1} k \cdot x_k \\ &= \frac{1}{n \cdot \sum_{i=1}^n x_i} \sum_{k=0}^{n-1} k \cdot x_k - n + 12n \end{aligned}$$

und für den Gini-Koeffizienten

$$\begin{aligned} G &= \frac{A}{\frac{1}{2}} \\ &= 2 \cdot A \\ &= \frac{2 \cdot \sum_{k=0}^{n-1} k \cdot x_k}{n \cdot \sum_{i=1}^n x_i} - n + 12n \end{aligned}$$

wie gewünscht.

**Beispiel 2.14.** In den Situation von Beispiel 2.13 erhalten wir:

- Für die Verteilung

Unternehmen	1	2	3	4
Umsatz	25	25	25	25

die Gini-Koeffizienten

$$G = 0, \quad G^* = 0$$

- Für die Verteilung

Unternehmen	1	2	3	4
Umsatz	10	10	10	70

den Gini-Koeffizienten

$$G = 0.45 \quad G^* = 0.60$$

- Für die Verteilung

Unternehmen	1	2	3	4
Umsatz	20	23	27	30

den Gini-Koeffizienten

$$G = 0.085 \quad G^* = 0.1133$$

**Satz 2.4.** Genau dann gilt  $G^* = 0$ , wenn die Merkmalsausprägungen gleichverteilt sind. Genau dann gilt  $G^* = 1$ , wenn  $x_1 = x_2 = \dots = x_{n-1} = 0, x_n \neq 0$  (also wenn maximale Konzentration vorliegt).

**Beweis:**

Da  $0 \leq x_1 \leq x_2 \leq \dots \leq x_n$  und  $x_n > 0$  gilt sicherlich

$$v_k = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i} \leq \frac{k}{n} = u_k$$

Damit verschwindet der Gini-Koeffizient genau dann, wenn die Lorenzkurve mit der Diagonale übereinstimmt.

Falls nun  $x_1 = x_2 = \dots = x_n = c$ , so gilt für alle  $k$ :

$$v_k = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i} = \frac{k \cdot c}{n \cdot c} = \frac{k}{n} = u_k$$

und damit stimmt in diesem Fall die Lorenzkurve mit der Diagonale überein, und daher ist  $G = 0$ .

Falls umgekehrt  $G = 0$ , so muss für alle  $k$  gelten

$$\frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i} = \frac{k}{n}$$

also

$$\frac{1}{k} \cdot \sum_{i=1}^k x_i = \frac{1}{n} \sum_{i=1}^n x_i$$

Falls aber nicht alle  $x_i$  gleich wären, so müsste, da  $x_1 \leq x_2 \leq \dots \leq x_n$  und  $x_1 < x_n$  (da ja nicht alle  $x_i$  gleich sind) gelten

$$x_1 = \frac{1}{n} \cdot n \cdot x_1 < \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

ein Widerspruch. Damit ist auch diese Richtung gezeigt.

Falls  $x_1 = x_2 = \dots = x_{n-1} = 0$ , so gilt

$$v_1 = v_2 = \dots = v_{n-1} = 0, \quad v_n = 1$$

die Lorenzkurve stimmt also bis  $x = \frac{n-1}{n}$  mit der  $x$ -Achse überein und ist dann das Geradenstück, dass  $(\frac{n-1}{n}, 0)$  mit  $(1, 1)$  verbindet. Daher hat die Fläche unterhalb der Lorenzkurve den Inhalt  $\frac{1}{2n}$ , und es ist

$$G = \frac{\frac{1}{2} - \frac{1}{2n}}{\frac{1}{2}} = \frac{n-1}{n}$$

Wir betrachten nun den Fall, dass  $G = \frac{n-1}{n}$  und nehmen an, dass auch nicht gilt

$$x_1 = x_2 = \dots = x_{n-1} = 0$$

Notwendig muss dann  $x_{n-1} > 0$  gelten, und deshalb hat bereits die Fläche unter der Lorenzkurve zwischen den Punkten  $v_{n-1} = \frac{n-1}{n}$  und  $v_n = 1$  den Inhalt  $F > \frac{1}{2n}$ , so dass also

$$G \leq \frac{\frac{1}{2} - F}{\frac{1}{2}} > \frac{\frac{1}{2} - \frac{1}{2n}}{\frac{1}{2}} = \frac{n-1}{n}$$

ein Widerspruch. Also muss notwendigerweise gelten

$$x_1 = x_2 = \dots = x_{n-1} = 0$$

Der Ginikoeffizient wird häufig herangezogen, um Vermögensverteilungen zu untersuchen.

**Beispiel 2.15.** Die Gini-Koeffizienten in einigen Ländern für das Vermögen  $V$  und das Einkommen  $E$

Land	Gini( $V/2000$ )	Gini( $V/2016$ )	Gini( $E$ )
Russland	0.699	0.923	0.377
Indien	0.669	0.876	0.368
Italien	0.609	0.687	0.354
USA	0.801	0.862	0.415
Deutschland	0.667	0.789	0.291
UK	0.697	0.732	0.332
Bangladesh	0.660	0.786	0.324
Japan	0.547	0.631	0.336
Frankreich	0.730	0.720	0.327

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 2

Stand: März 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

e-mail: [reinhold.huebl@dhbw-mannheim.de](mailto:reinhold.huebl@dhbw-mannheim.de)

©Fakultät für Technik DHBW Mannheim

## 2.2 Mehrdimensionale deskriptive Statistik

In vielen Anwendungen interessiert man sich nicht nur für ein einziges Merkmal sondern gleichzeitig für mehrere. Bei Konsumforschungen etwa interessieren natürlich die Gesamtausgaben der Haushalte für Konsumgüter (univariate Statistik), aber für Marktanalysen ist es wichtiger, die Ausgaben der Haushalte für einzelne Produkte oder Produktgruppen (Lebensmittel, Bildung und Kultur, Kosmetika, Unterhaltung etc.) zu wissen. Darüberhinaus ist es für Marketingüberlegungen auch relevant, zu wissen, welche Konsumentengruppe welche Ausgaben tätigt, also neben den Ausgaben auch noch Merkmale wie Geschlecht oder Bildungsniveau zu erheben.

Bei der Untersuchen von zwei Merkmalen  $X$  und  $Y$  interessieren dabei grundsätzlich zwei Fragen:

- Beeinflussen sich die Merkmalsausprägungen von  $X$  und  $Y$  gegenseitig (Korrelationsanalyse)?
- Hängen die Ausprägungen eines Merkmals (etwa  $Y$ ) funktional von den Ausprägungen des anderen Merkmals (also  $X$ ) ab (Regressionsanalyse).

Wir wollen uns hier zunächst auf zwei Merkmale (zweidimensionale Daten) beschränken. Wir betrachten dazu zwei diskrete (mindestens nominalskalierte) Merkmale, die relativ wenige Ausprägungen zulassen, z.B. Geschlecht und Noten in Mathematik. Dieser Zustand kann auch herbeigeführt werden, indem wir ein (oder beide) Merkmale klassieren, etwa Alter (eingeteilt in drei Altersgruppen) und Parteienpräferenz oder Alter (eingeteilt in drei Altersgruppen) und Konsumausgaben für Unterhaltung (gruppiert in vier Klassen).

**Beispiel 2.16.** Bei der sogenannten Sonntagsfrage wird regelmäßig erhoben, welche Partei von ausgewählten Befragten gewählt würde, wenn am nächsten Sonntag Bundestagswahl wäre. Am 3.2.2016 ergab sich dabei (bei Infratest dimap) das folgende Gesamtbild (Angaben in %):

	CDU/CSU	SPD	GRÜ	Linke	FPD	AfD	Sonst
Ges	35	24	10	9	5	12	5

Aufgegliedert nach Ostdeutschland (OD) und Westdeutschland (WD) ergeben sich jedoch deutliche Unterschiede:

	CDU/CSU	SPD	GRÜ	Linke	FPD	AfD	Sonst
WD	35	26	11	7	5	11	5
OD	32	19	6	18	3	16	6



Bei 1000 Befragten, davon 200 aus Ostdeutschland (das ist die übliche Samplegröße) entspricht das also in absoluten Zahlen etwa:

	CDU/CSU	SPD	GRÜ	Linke	FPD	AfD	Sonst
WD	283	205	90	54	43	87	38
OD	64	38	12	36	6	32	12

In der zweiten (und dritten) Tabelle werden also zwei Merkmale, *Parteienpräferenz* und *Region*, untersucht.

**Beispiel 2.17.** Untersucht werden die beiden Merkmale  $X = \text{Einkommen}$  gestaffelt nach

- *niedrig* ( $n$ ): höchstens € 2000.
- *mittel* ( $m$ ): mehr als € 2000 aber höchstens € 4000.
- *hoch* ( $h$ ): mehr als € 4000.

und  $Y = \text{Alter}$ , gestaffelt nach

- *jung* ( $jg$ ): höchstens 30 Jahre alt.
- *mittleres Alter* ( $ma$ ): älter als 30 Jahre aber höchstens 50 Jahre.
- *älter* ( $al$ ): älter als 50 Jahre.

Eine Untersuchung von 100 Personen führte zu folgenden Ergebnissen

	jg	ma	al
$n$	20	15	7
$m$	8	17	14
$h$	2	8	9

**Beispiel 2.18.** Die Aufgliederung des Beschäftigungsverhältnisses nach Geschlecht ergibt folgende Daten (in Tausend)

	arbeitslos	Teilzeit	vollbeschäftigt
weiblich	500	4200	7300
männlich	500	800	12700

Wir betrachten nun allgemeiner zwei Merkmale  $X$  und  $Y$  mit den möglichen Ausprägungen

$$\begin{array}{ll} a_1, \dots, a_k & \text{für } X \\ b_1, \dots, b_l & \text{für } Y \end{array}$$

In den Rohdaten liegen für jeden Merkmalsträger die gemeinsamen Messwerte für die Merkmale  $X$  und  $Y$  vor, d.h. wir haben Paare  $(x_1, y_1), \dots, (x_n, y_n)$ . In Analogie zum univariaten Fall bilden wir nun die (*absoluten*) *Häufigkeiten*

$$h_{i,j} = h(a_i, b_j) = |\{(x_r, y_r) \mid x_r = a_i, y_r = b_j\}|$$

**Definition 2.14.** Die Werte  $h_{i,j}, i = 1, \dots, k, j = 1, \dots, l$  heißen *gemeinsame Verteilung der Merkmale  $X$  und  $Y$  in absoluten Häufigkeiten* und die Tabelle

	$b_1$	$\dots$	$b_l$
$a_1$	$h_{1,1}$	$\dots$	$h_{1,l}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$a_k$	$h_{k,1}$	$\dots$	$h_{k,l}$

heißt *Kontingenztabelle* der Merkmale  $X$  und  $Y$ .

Aus der Kontingenztabelle für zwei Merkmale kommen wir zurück auf die Häufigkeiten eines Merkmals

**Definition 2.15** (Randhäufigkeiten). Die *Randhäufigkeiten des Merkmals  $Y$*  sind die Spaltensummen der Kontingenztabelle

$$h_{\bullet,j} = h_{1,j} + \dots + h_{k,j} \quad (j = 1, \dots, l)$$

und die *Randhäufigkeiten des Merkmals  $X$*  sind die Zeilensummen der Kontingenztabelle

$$h_{i,\bullet} = h_{i,1} + \dots + h_{i,l} \quad (i = 1, \dots, k)$$

**Bemerkung 2.23.** Die Kontingenztabelle wird oft mit den Randhäufigkeiten ergänzt und hat dann die folgende Gestalt

	$b_1$	$\dots$	$b_l$	
$a_1$	$h_{1,1}$	$\dots$	$h_{1,l}$	$h_{1,\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$a_k$	$h_{k,1}$	$\dots$	$h_{k,l}$	$h_{k,\bullet}$
	$h_{\bullet,1}$	$\dots$	$h_{\bullet,l}$	

In unserem Beispiel mit der Sonntagsfrage etwa

	CDU/CSU	SPD	GRÜ	Linke	FPD	AfD	Sonst	
WD	283	205	90	54	43	87	38	800
OD	64	38	12	36	6	32	12	200
	347	243	102	90	49	119	50	

Die rechte Spalte gibt also an, wie viele Ost- bzw. Westdeutsche befragt wurden und die unterste Spalte enthält die Gesamtstimmen, die dabei für die jeweilige Partei abgegeben wurden.

**Beispiel 2.19.** Bei den Zahlen aus Beispiel 2.17 hat die um die Randhäufigkeiten erweiterte Tabelle die folgende Gestalt

	jg	ma	al	
$n$	20	15	7	42
$m$	8	17	14	39
$h$	2	8	9	19
	30	40	30	100

**Definition 2.16.** (*relative Häufigkeiten*)

Bei einer Stichprobengröße  $n$  heißt

$$f_{i,j} = f(a_i, b_j) = \frac{1}{n} \cdot h(a_i, b_j)$$

die *relative (gemeinsame) Häufigkeit* der Merkmalsausprägungen  $(a_i, b_j)$ , und

$$f(a_i) = f_X(a_i) = \frac{1}{n} \cdot h_{i,\bullet}, \quad f(b_j) = f_Y(b_j) = \frac{1}{n} \cdot h_{\bullet,j}$$

heißen relative Randhäufigkeiten der Merkmalsausprägungen  $a_i$  von  $X$  und  $b_j$  von  $Y$ .

Aus den relativen Häufigkeiten erstellen wir die Kontingenztafel der relativen Häufigkeiten, im Beispiel der Sonntagsumfrage etwa

	CDU/CSU	SPD	GRÜ	Linke	FPD	AfD	Sonst	
WD	0.283	0.205	0.09	0.054	0.143	0.087	0.038	0.80
OD	0.064	0.038	0.012	0.036	0.006	0.032	0.012	0.20
	0.347	0.243	0.102	0.09	0.049	0.119	0.05	

Die Zahlen stimmen hier nicht genau mit den weiter oben angegebenen Prozentwerten überein, da diese Werte auf ganze Prozent gerundet wurden.

**Beispiel 2.20.** Bei den Zahlen aus Beispiel 2.17 hat die Tabelle der relativen Häufigkeiten die folgende Gestalt

	jg	ma	al
$n$	0.20	0.15	0.07
$m$	0.08	0.17	0.14
$h$	0.02	0.08	0.09

Zur Untersuchung der Zusammenhänge zwischen  $X$  und  $Y$  ist es wichtig, zu untersuchen, ob die Ausprägungen von  $X$  einen Einfluß auf die Ausprägungen von  $Y$  haben (und umgekehrt). Dazu setzen wir

**Definition 2.17.** (*Bedingte Häufigkeiten*)

Die *bedingten Häufigkeiten von  $Y$  unter der Bedingung  $X = a_i$* , kurz auch  $Y|X = a_i$  sind gegeben durch

$$f_Y(b_1|a_i) = \frac{h_{i,1}}{h_{i,\bullet}}, \quad \dots, \quad f_Y(b_l|a_i) = \frac{h_{i,l}}{h_{i,\bullet}}$$

Die *bedingten Häufigkeiten von  $X$  unter der Bedingung  $Y = b_j$* , kurz auch  $X|Y = b_j$  sind gegeben durch

$$f_X(a_1|b_j) = \frac{h_{1,j}}{h_{\bullet,j}}, \quad \dots, \quad f_X(a_k|b_j) = \frac{h_{k,j}}{h_{\bullet,j}}$$

Die Merkmale  $X$  und  $Y$  heiße *empirisch unabhängig*, wenn gilt

$$\begin{aligned} f_Y(b_j|a_i) &= \frac{1}{n} \cdot h_{\bullet,j} && \text{für alle } j, i \\ f_X(a_i|b_j) &= \frac{1}{n} \cdot h_{i,\bullet} && \text{für alle } j, i \end{aligned}$$

**Beispiel 2.21.** Bezeichnen wir mit  $a_1$  die Merkmalsausprägung *OD* und mit  $a_2$  die Merkmalsausprägung *WD* sowie mit  $b_1$  die Merkmalsausprägung *CDU/CSU*, so gilt

$$f_Y(b_1|a_1) = 0.32, \quad f_Y(b_1|a_2) = 0.35375$$

Die Merkmale *Parteienpräferenz* und *Region* sind also nicht empirisch unabhängig. Noch besser erkennbar ist dies an den Werten für die Linke: ist  $b_4$  die Merkmalsausprägung *Linke*, so gilt

$$f_Y(b_4|a_1) = 0.18, \quad f_Y(b_4|a_2) = 0.0675$$

**Bemerkung 2.24.** Sind  $X$  und  $Y$  empirisch unabhängig, so gilt

$$h_{i,j} = \frac{h_{i,\bullet} \cdot h_{\bullet,j}}{n}$$

Deshalb nennen wir auch

$$\widetilde{h}_{i,j} = \frac{h_{i,\bullet} \cdot h_{\bullet,j}}{n}$$

die Häufigkeiten, die zu erwarten sind, wenn kein Zusammenhang vorliegt.

**Satz 2.5.** Zwei Merkmale  $X$  und  $Y$  mit Merkmalsausprägungen  $a_1, \dots, a_k$  und  $b_1, \dots, b_l$  sind genau dann empirisch unabhängig, wenn

$$h_{i,j} = \widetilde{h}_{i,j} \quad \text{für alle } i \in \{1, \dots, k\}, j \in \{1, \dots, l\}$$

**Beispiel 2.22.** Bei der Sonntagsumfrage ergibt sich als Tabelle der erwarteten Häufigkeiten bei empirische Unabhängigkeit

	CDU/CSU	SPD	Linke	Grüne	FDP	AfD	Sonst	
WD	333	180	81	67.5	85.5	108	45	900
OD	111	60	27	22.5	28.5	36	15	300
	444	240	108	90	114	144	60	

Die Merkmale sind also offensichtlich nicht empirisch unabhängig, starke Abweichungen sind allerdings auf wenige Zellen konzentriert.

**Beispiel 2.23.** Bei den Zahlen aus Beispiel 2.17 hat die Tabelle der erwarteten Häufigkeiten bei empirischer Unabhängigkeit die folgende Gestalt

	jg	ma	al	
$n$	12.6	16.8	12.6	42
$m$	11.7	15.6	11.7	39
$h$	5.7	7.6	5.7	19
	30	40	30	100

Diese Werte weichen also (an mehreren Stellen) teilweise deutlich von den erhobenen Werten ab. Damit sind diese Merkmale nicht empirisch unabhängig.

**Definition 2.18** (Kontingenzkoeffizienten). Die Größe

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(h_{i,j} - \widetilde{h}_{i,j})^2}{\widetilde{h}_{i,j}}$$

heißt  $\chi^2$ -Koeffizient von  $X$  und  $Y$ .

Die Größe

$$K = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

heißt *Kontingenzkoeffizient von  $X$  und  $Y$* .

Die Größe

$$K^* = \frac{K}{\sqrt{\frac{M-1}{M}}} \quad \text{wobei } M = \min\{k, l\}$$

heißt *korrigierter oder normierter Kontingenzkoeffizient von  $X$  und  $Y$* .

**Bemerkung 2.25.** Es gilt

1. Die Kontingenzkoeffizienten bilden ein Maß für den Zusammenhang zwischen  $X$  und  $Y$ , also ob die Merkmalsausprägungen für  $X$  mit denen für  $Y$  zusammenhängen.
2. Die Kontingenzkoeffizienten sind geeignet für die Untersuchung mindestens nominalskaliertter Merkmale.
3.  $\chi^2 \in [0, \infty[$ ,  $K \in [0, \sqrt{\frac{M-1}{M}}]$  und  $K^* \in [0, 1]$ .
4.  $\chi^2$  hängt ab von der Skalierung der Merkmale; werden alle Häufigkeiten verdoppelt, so verdoppelt sich auch  $\chi^2$ . Die Kontingenzkoeffizienten  $K$  und  $K^*$  sind unabhängig von solchen linearen Transformationen der Häufigkeiten. Darüberhinaus ist  $K^*$  normiert.
5. Genau dann sind die Kontingenzkoeffizienten 0, wenn  $X$  und  $Y$  empirisch unabhängig sind.
6. Je größer  $K$  und  $K^*$  (und unter Berücksichtigung der Normierungsfrage auch  $\chi^2$ ) sind, desto stärker ist der Zusammenhang zwischen  $X$  und  $Y$ .

**Beispiel 2.24.** Bei der Sonntagsumfrage gilt

$$\begin{aligned}\chi^2 &= 36.112 \\ K &= 0.187 \\ K^* &= 0.264\end{aligned}$$

Diese Zahlen zeigen, dass empirische Unabhängigkeit nicht gegeben ist, dass allerdings auch nicht von einer starken empirischen Abhängigkeit gesprochen werden kann.

**Beispiel 2.25.** Bei den Zahlen aus Beispiel 2.17 erhalten wir die Koeffizienten

$$\begin{aligned}\chi^2 &= 13.1 \\ K &= 0.34 \\ K^* &= 0.42\end{aligned}$$

Diese Zahlen zeigen, dass empirische Unabhängigkeit nicht gegeben ist, und dass diese schon deutlich stärker verletzt ist als im Beispiel mit der Sonntagsumfrage.

**Beispiel 2.26.** Bei den Zahlen aus Beispiel 2.18 erhalten wir die Koeffizienten

$$\begin{aligned}\chi^2 &= 3637.7 \\ K &= 0.350 \\ K^* &= 0.496\end{aligned}$$

Diese Zahlen zeigen, dass empirische Unabhängigkeit nicht gegeben ist, und dass diese schon deutlich stärker verletzt ist als im Beispiel mit der Sonntagsumfrage.

Der Koeffizient  $\chi^2$  ist geeignet für alle mindestens nominal skalierten Merkmale. Er gibt allerdings keine Auskunft über die Richtung des Zusammenhangs. Bei nur nominalskalierten Merkmalen ist das auch nicht möglich, da die Merkmalsausprägungen nicht quantitativ interpretiert werden können. Deshalb wollen wir jetzt metrisch skalierte Merkmale betrachten. Dazu betrachten wir die arithmetischen Mittelwerte

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i$$

von  $X$  und  $Y$ .

**Definition 2.19.** (*Korrelationskoeffizienten*)

Die Größe

$$r = r_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

heißt *Bravais–Pearson–Korrelationskoeffizient* oder *empirischer Korrelationskoeffizient* von  $X$  und  $Y$ .

**Bemerkung 2.26.** Es gilt

1.  $r_{X,Y} \in [-1, 1]$ .
2.  $r_{X,Y}$  untersucht einen linearen Zusammenhang zwischen  $X$  und  $Y$ : Genau dann gilt  $r_{X,Y} = \pm 1$ , wenn  $y_i = ax_i + b$  für alle  $i \in \{1, \dots, n\}$ , wobei  $a, b \in \mathbb{R}, a \neq 0$  fest sind. Dabei ist  $r_{X,Y} = 1$  wenn  $a > 0$  und  $r_{X,Y} = -1$  wenn  $a < 0$ .
3. Ist  $r_{X,Y} > 0$  so heiße  $X$  und  $Y$  positiv oder gleichsinnig linear korreliert, ist  $r_{X,Y} < 0$ , so sind sie negativ oder gegensinnig linear korreliert. Ist  $r_{X,Y} = 0$ , so sind sie unkorreliert.

**Bemerkung 2.27.** Die Größe

$$s_{X,Y}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

wird auch als *korrigierte empirische Kovarianz* bezeichnet. Damit schreibt sich der Bravais–Pearson–Korrelationskoeffizient auch als

$$r_{X,Y} = \frac{s_{X,Y}^2}{\sqrt{s_X^2 \cdot s_Y^2}} = \frac{s_{X,Y}^2}{s_X \cdot s_Y}$$

**Beispiel 2.27.** Wir betrachten die Merkmale  $X = \text{Körpergröße}$  und  $Y = \text{Körpergewicht}$  von vier Personen und erhalten folgende Daten:

Proband	1	2	3	4
Größe	175	184	179	182
Gewicht	82	81	96	81

Dann gilt  $\bar{x} = 180$  und  $\bar{y} = 85$ . Ferner ist

$$\begin{aligned} \sum_{i=1}^4 (x_i - \bar{x})^2 &= (-5)^2 + 4^2 + (-1)^2 + 2^2 &= 46 \\ \sum_{i=1}^4 (y_i - \bar{y})^2 &= (-3)^2 + (-4)^2 + 11^2 + (-4)^2 &= 162 \\ \sum_{i=1}^4 (x_i - \bar{x}) \cdot (y_i - \bar{y}) &= (-5) \cdot (-3) + 4 \cdot (-4) + (-1) \cdot 11 + 2 \cdot (-4) &= -20 \end{aligned}$$

und damit

$$r_{x,y} = \frac{-20}{\sqrt{46 \cdot 162}} = -\frac{20}{\sqrt{7452}} = -0.232$$

also sind  $X$  und  $Y$  in diesem Beispiel schwach negativ korreliert.

**Bemerkung 2.28.** Der Bravais–Pearson–Korrelationkoeffizient ist nur für metrische Variablen geeignet. Eine Ausnahme bilden sogenannte *dichotome* oder *binäre* Merkmale, die also nur zwei Werte annehmen können. Diese werden üblicherweise auf 0 und 1 gesetzt. In diesem Fall schreibt man auch  $\Phi = \Phi_{X,Y}$  für  $r_{X,Y}$  und nennt diesen Wert den  $\Phi$ –Koeffizienten von  $X$  und  $Y$ . Es gilt

$$\Phi = \frac{\chi^2}{n}$$

Ein alternativer Korrelationskoeffizient für mindestens ordinal skalierte Merkmale ist möglich, wenn wir von den ursprünglichen Werten von  $X$  und  $Y$  zu ihren *Rängen* übergehen, also etwa

$x_i$	87	32	55	91	64
$rg(x_i)$	2	5	4	1	3



Eine entsprechende Vergabe von Rangplätzen wird für  $y$  durchgeführt. Falls unter den  $x_i$  oder  $y_i$  identische Werte auftreten, ist eine eindeutige Rangzuordnung nicht möglich. In diesem Fall wird jedem der in Frage kommenden  $x_i$  (bzw.  $y_i$ ) der *Druchschnittsrank*, also das arithmetische Mittel der in Frage kommenden Ränge zugewiesen, also etwa

$x_i$	87	32	55	91	55
$rg(x_i)$	2	5	3.5	1	3.5

Ferner bezeichnen wir mit

$$\begin{aligned}\overline{rg}_X &= \frac{1}{n} \cdot \sum_{i=1}^n rg(x_i) = \frac{n+1}{2} \\ \overline{rg}_Y &= \frac{1}{n} \cdot \sum_{i=1}^n rg(y_i) = \frac{n+1}{2}\end{aligned}$$

den Mittelwert der Ränge.

**Definition 2.20.** Die Größe

$$r_{SP} = \frac{\sum_{i=1}^n (rg(x_i) - \overline{rg}_X) \cdot (rg(y_i) - \overline{rg}_Y)}{\sqrt{\sum_{i=1}^n (rg(x_i) - \overline{rg}_X)^2 \cdot \sum_{i=1}^n (rg(y_i) - \overline{rg}_Y)^2}}$$

heißt *Spearman-Korrelationskoeffizient*.

**Bemerkung 2.29.** Es gilt

1.  $r_{SP} \in [-1, 1]$ .
2.  $r_{SP}$  ist der Bravais-Pearson-Korrelationskoeffizient der Rangaussprägungen  $(rg(x_i), rg(y_i))$  der Merkmale  $X$  und  $Y$ .
3. Genau dann ist  $r_{SP} > 0$ , wenn es einen gleichsinnigen monotonen Zusammenhang zwischen  $X$  und  $Y$  gibt.
4. Genau dann ist  $r_{SP} < 0$ , wenn es einen gegensinnigen monotonen Zusammenhang zwischen  $X$  und  $Y$  gibt.
5.  $r_{SP} \approx 0$  wenn es keinen monotonen Zusammenhang zwischen  $X$  und  $Y$  gibt.

Wir wollen nun wieder zwei metrisch skalierte Merkmale  $X$  und  $Y$  betrachten. Falls der Bravais-Pearson-Korrelationskoeffizient darauf hindeutet, dass es einen (linearen)

Zusammenhang zwischen  $X$  und  $Y$  gibt, soll dieser Zusammenhang auch quantifiziert werden, es soll also eine Gerade

$$y = f(x) = ax + b$$

gefunden werden, so dass die  $f(x_i)$  die  $y_i$  "möglichst gut" approximieren.

Wir betrachten dazu jetzt Beobachtungen  $(x_1, y_1), \dots, (x_n, y_n)$  der Merkmale  $X$  und  $Y$  und wollen eine Gerade  $y = f(x) = ax + b$  durch die Zeichenebene legen, so dass die quadratische Abweichung der Punkte  $(y_i, x_i)$  möglichst gering ist, dass also  $\sum_{i=1}^n (y_i - f(x_i))^2$  minimal wird (**Methode der kleinsten Quadrate**). Eine Gerade, die diese Quadratsumme minimiert nennen wir **Ausgleichsgerade**.

Dazu setzen wir

$$\hat{a} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{b} = \bar{y} - \hat{a} \cdot \bar{x}$$

**Definition 2.21.** (*Lineare Regression*)

Die Gerade

$$f(x) = \hat{a} \cdot x + \hat{b}$$

heißt *lineare Einfachregression* der Merkmale  $X$  und  $Y$ .

Die Werte

$$\varepsilon_i = y_i - f(x_i)$$

heißen Residuen der Einfachregression.

**Definition 2.22.** Die Größe

$$R^2 = \frac{\sum_{i=1}^n (f(x_i) - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

heißt Bestimmtheitsmaß der linearen Regression.

**Bemerkung 2.30.** Es gilt

$$R^2 = r_{X,Y}^2$$

**Bemerkung 2.31.** Es gilt

$$\sum_{i=1}^n \left( y_i - \hat{b} - \hat{a} \cdot x_i \right)^2 = \min \left\{ \sum_{i=1}^n (y_i - b - a \cdot x_i)^2 \mid a, b \in \mathbb{R} \right\}$$

**Beispiel 2.28.** Wir betrachten zwei Merkmale  $X$  und  $Y$  mit den folgenden Ausprägungen

$k$	1	2	3	4	5	6	7
$x_k$	3	5	6	7	8	10	12
$y_k$	24	30	32	30	28	44	42

Hierfür gilt

$$r_{X,Y} = 0.8701$$

es besteht also offensichtlich ein starker linearer Zusammenhang zwischen den Merkmalsausprägungen, und die Suche nach einer Ausgleichsgeraden ist sinnvoll.

Zunächst gilt

$$\bar{x} = 7.2875, \quad \bar{y} = 32.7142$$

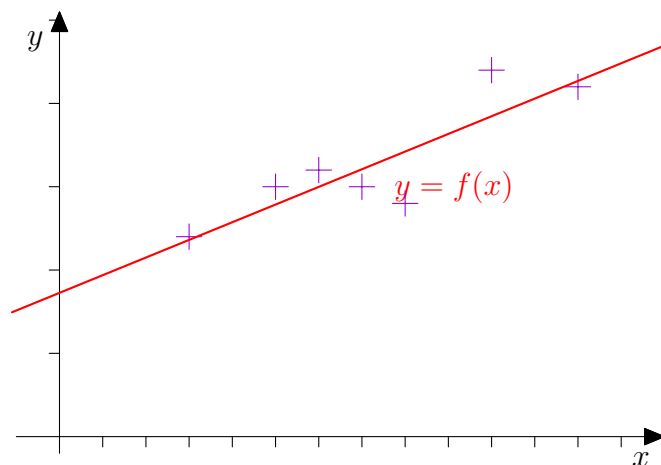
und damit

$$\hat{a} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 2.1211 \quad \text{und} \quad \hat{b} = \bar{y} - \hat{a} \cdot \bar{x} = 17.2603$$

Die lineare Regressionsgerade ist also gegeben durch

$$f(x) = 17.2603 + 2.1211 \cdot x$$

Das ergibt folgendes Bild:



Sie beschreibt die Werte sehr gut, was auch durch das hohe Bestimmtheitsmaß von

$$R^2 = 0.757$$

bestätigt wird.

**Beispiel 2.29.** Wir betrachten die beiden Merkmale  $X = \text{Größe}$  und  $Y = \text{Gewicht}$  und erhalten aus einer Erhebung die folgenden Daten

$k$	1	2	3	4	5	6	7
$x_k$	160	165	170	175	180	185	190
$y_k$	90	85	58	95	65	68	78

Hierfür gilt

$$r_{X,Y} = -0.352$$

es besteht also offensichtlich ein negativer linearer Zusammenhang zwischen den Merkmalsausprägungen, der allerdings eher schwach ausgeprägt ist.

Zunächst gilt

$$\bar{x} = 175, \quad \bar{y} = 77$$

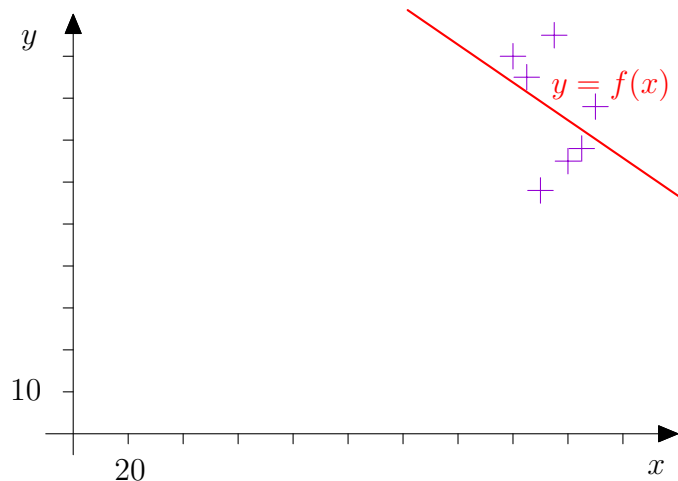
und damit

$$\hat{a} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = -0.45 \quad \text{und} \quad \hat{b} = \bar{y} - \hat{a} \cdot \bar{x} = 155.75$$

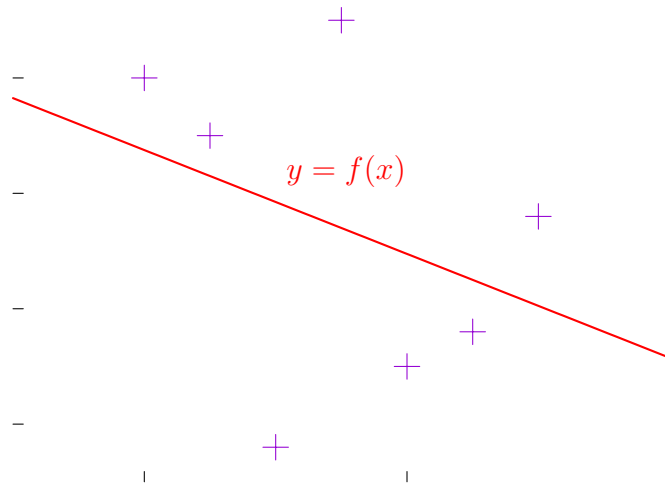
Die lineare Regressionsgerade ist also gegeben durch

$$f(x) = 155.75 - 0.45 \cdot x$$

Das ergibt folgendes Bild:



In diesem Fall ist die Skalierung noch etwas zu grob. Ein Blick nur auf den Datenbereich liefert



Die Ausgleichsgerade führt also in diesem Fall zu einer eher schlechten Näherung für die Punkte. Das wird auch bestätigt durch ds Bestimmtheitsmaß, dass hier nur bei

$$R^2 = 0.124$$

liegt.

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 3

Stand: April 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim

## 3 Wahrscheinlichkeiten

### 3.1 Mengenalgebren und $\sigma$ -Algebren

Die Grundlage für die Beschäftigung mit Wahrscheinlichkeiten und Statistiken ist ein profundes Verständnis von Mengen und Mengensystemen. Dazu wollen wir uns zunächst ein wenig mit Mengen beschäftigen.

Mengen spielen eine zentrale Rolle in der Mathematik (und nicht nur dort). Wir folgen hier dem Zugang Georg Cantors (1845 - 1918) und verzichten auf eine axiomatische Einführung.

**Definition 3.1.** Eine **Menge** ist eine Zusammenfassung  $M$  von bestimmten wohlunterschiedenen Objekten  $m$ , genannt die Elemente von  $M$ , unseres Anschauungsraums oder unseres Denkens zu einem Ganzen.

Ist  $m$  ein Element von  $M$ , so schreiben wir  $m \in M$ , andernfalls schreiben wir  $m \notin M$ . Für jedes Objekt  $m$  unserer Anschauung und jede Menge  $M$  gilt also genau entweder  $m \in M$  oder  $m \notin M$ , nicht aber beides.

**Beispiel 3.1.** • Die Menge, die keine Elemente enthält, bezeichnet man als **leere Menge**. Hierfür schreiben wir  $\emptyset$  oder  $\{\}$ .

- Eine Menge  $M$ , die nur endlich viele Elemente hat, lässt sich durch vollständige Auflistung aller ihrer Elemente zwischen Mengenklammern beschreiben, z.B.  $M = \{1, 3, 5, 7, 9\}$  ist die Menge, die als Elemente die Zahlen 1, 3, 5, 7, 9 enthält, und  $M = \{\text{rot}, \text{blau}, \text{grün}\}$  ist die Menge, die als Elemente die Farben *rot*, *blau* und *grün* enthält.
- Auch unendliche Mengen lassen sich oft in aufzählender Form darstellen. So wird die Menge  $\mathbb{N}$  der natürlichen Zahlen oft in der Form

$$\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}$$

dargestellt, und die Menge  $\mathbb{Z}$  der ganzen Zahlen als

$$\mathbb{Z} = \{\dots - 4, -3, -2, -1, 0, 1, 2, 3, 4, \dots\}$$

- Manche Mengen lassen sich nicht oder nur schwer aufzählend darstellen. Häufig ist hierfür jedoch eine beschreibende Darstellung möglich. Die Menge

$$\mathbb{P} = \{x \in \mathbb{N} \mid x \text{ ist eine Primzahl} \}$$

etwa enthält als Elemente alle natürlichen Zahlen  $x$ , die Primzahlen sind, d.h. alle natürlichen Zahlen  $x \neq 1$ , die nur durch 1 und sich selbst teilbar sind.

- Eine wichtige Rolle in diesem Skript spielt die Menge  $\mathbb{Q}$  der rationalen Zahlen,

$$\mathbb{Q} = \left\{ \frac{m}{n} \mid m, n \in \mathbb{Z} \text{ und } n \neq 0 \right\}.$$

**Definition 3.2.** Eine Menge  $N$  heißt **Teilmenge** einer Menge  $M$ , geschrieben  $N \subseteq M$ , wenn jedes Element von  $N$  auch Element von  $M$  ist, d.h. wenn die Implikation  $x \in N \implies x \in M$  gilt. In diesem Fall heißt  $M$  auch **Obermenge** von  $N$ .

**Definition 3.3.** Zwei Mengen  $N$  und  $M$  heißen gleich, geschrieben  $M = N$ , wenn ein Objekt  $x$  genau dann Element von  $N$  ist, wenn es Element von  $M$  ist, d.h. wenn die Äquivalenz  $x \in N \iff x \in M$  gilt.

**Bemerkung 3.1.** Eine Menge  $N$  heißt echte Teilmenge von  $M$ , geschrieben  $N \subset M$ , wenn  $N$  Teilmenge von  $M$  mit  $N \neq M$  ist.

Die Gleichheit von Mengen lässt sich durch die Teilmengenbeziehung beschreiben.

**Satz 3.1.** Für zwei Mengen  $M$  und  $N$  gilt genau dann  $M = N$ , wenn  $M \subseteq N$  und  $N \subseteq M$ .

Für Teilmengen  $A$  und  $B$  einer gegebenen Menge  $M$  lassen sich verschiedene Beziehungen definieren:

**Definition 3.4.** Die **Schnittmenge** von  $A$  und  $B$ , geschrieben  $A \cap B$  besteht aus den Elementen von  $M$ , die sowohl in  $A$  als auch in  $B$  sind:

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

Falls  $A \cap B = \emptyset$ , so nennen wir  $A$  und  $B$  **disjunkt**.

**Definition 3.5.** Die **Vereinigungsmenge** von  $A$  und  $B$ , geschrieben  $A \cup B$  besteht aus den Elementen von  $M$ , die entweder in  $A$  oder in  $B$  sind:

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

Unmittelbar aus der Definition sehen wir

**Bemerkung 3.2.** Ist  $A \subseteq B$  so gilt

$$\begin{aligned} A \cup B &= B \\ A \cap B &= A \end{aligned}$$

Für die Bildung von Vereinigungs- und Durchschnittsmengen gelten folgende wichtige Regeln



**Satz 3.2.** Für Teilmengen  $A, B, C \subseteq M$  gilt

Kommutativgesetz  $A \cup B = B \cup A$

$$A \cap B = B \cap A$$

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

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

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

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

Verschmelzungsgesetz  $A \cap (A \cup B) = A$

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

**Beweis:** Kommutativ- und Assoziativgesetze sind klar.

Wir zeigen zunächst das erste Distributivgesetz (unter Ausnutzung des ersten Distributivgesetzes der Logik):

$$\begin{aligned} x \in (A \cup B) \cap C &\iff (x \in (A \cup B)) \wedge (x \in C) \\ &\iff ((x \in A) \vee (x \in B)) \wedge (x \in C) \\ &\iff ((x \in A) \wedge (x \in C)) \vee ((x \in B) \wedge (x \in C)) \\ &\iff (x \in A \cap C) \vee (x \in B \cap C) \\ &\iff x \in (A \cap C) \cup (B \cap C) \end{aligned}$$

Wir haben also gezeigt:

$$x \in (A \cup B) \cap C \iff x \in (A \cap C) \cup (B \cap C)$$

und damit

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

wie gewünscht.

Das zweite Distributivgesetz zeigt man analog mit dem entsprechenden Distributivgesetz der Logik.

Den Beweis der Verschmelzungsgesetze überlassen wir dem Leser als Übungsaufgabe.

**Aufgabe 1.** Welche der folgenden Aussagen sind richtig für Mengen  $A$ ,  $B$  und  $C$ , welche sind falsch?

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

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

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

**Lösung 1.** 1.  $(A \cup B) \cap C = A \cup (B \cap C)$  ist falsch:

Dazu betrachten wir die folgenden Teilmengen  $A = \{1, 2, 3, 4\}$ ,  $B = \{3, 4, 5, 6\}$  und  $C = \{5, 6, 7, 8\}$  von  $\mathbb{N}$ . Dann gilt

$$(A \cup B) \cap C = \{5, 6\}$$

aber

$$A \cup (B \cap C) = \{1, 2, 3, 4, 5, 6\}$$

2.  $(A \cap B) \cap (C \cup B) = A \cap B$  ist richtig:

Es gilt sicherlich

$$A \cap B \subseteq B \subseteq C \cup B$$

und damit gilt nach Bemerkung 3.2

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

3.  $(A \cup B) \cup (B \cap C) = A \cup B$  ist richtig:

Es gilt sicherlich

$$B \cap C \subseteq B \subseteq A \cup B$$

und damit gilt nach Bemerkung 3.2

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

**Definition 3.6.** Die **Differenzmenge** von  $A$  und  $B$ , geschrieben  $A \setminus B$  besteht aus den Elementen von  $M$ , die in  $A$  aber nicht in  $B$  sind:

$$A \setminus B = \{x \in M \mid x \in A \text{ und } x \notin B\}$$

Falls  $B \subseteq A$  nennen wir  $A \setminus B$  auch das **Komplement** von  $B$  in  $A$  und schreiben hierfür  $\overline{B}^A$ . Falls  $A = M$ , schreiben wir hierfür auch kurz  $\overline{B}$  und nennen es das Komplement von  $B$ .

Auch für die Komplementbildung gelten einige einfache Regeln:

**Regel 3.3.** Für eine Teilmenge  $A \subseteq M$  gilt:

1.  $\overline{\overline{A}} = A.$

2.  $\overline{A} \cup A = M.$

3.  $\overline{A} \cap A = \emptyset$ .

4.  $\overline{\emptyset} = M$

5.  $\overline{M} = \emptyset$ .

**Aufgabe 2.** Welche der folgenden Aussagen ist korrekt?

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

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

**Lösung 2.** 1.  $A \setminus (B \cap C) = A \setminus B \cap A \setminus C$  ist falsch:

Dazu betrachten wir die folgenden Teilmengen  $A = \{1, 2, 3, 4\}$ ,  $B = \{3, 4, 5, 6\}$  und  $C = \{5, 6, 7, 8\}$  von  $\mathbb{N}$ . Dann gilt

$$A \setminus (B \cap C) = \{1, 2, 3, 4\}$$

aber

$$A \setminus B \cap A \setminus C = \{1, 2\}$$

2.  $A \setminus (B \cup C) = A \setminus B \cap A \setminus C$  ist richtig:

Es gilt

$$\begin{aligned} x \in A \setminus (B \cup C) &\iff x \in A \wedge x \notin (B \cap C) \\ &\iff x \in A \wedge \neg(x \in B \cap C) \\ &\iff x \in A \wedge \neg(x \in B \vee x \in C) \end{aligned}$$

Nun fixieren wir eine Grundmenge  $\Omega$  und bezeichnen mit  $\mathfrak{P}(\Omega)$  ihre Potenzmenge, also die Menge aller Teilmengen von  $\Omega$ :

$$\mathfrak{P}(\Omega) = \{A \mid A \subseteq \Omega\}$$

**Definition 3.7.** Eine Teilmenge  $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$  heißt **Mengenalgebra** (auf  $\Omega$ ), wenn gilt:

1.  $\mathfrak{A} \neq \emptyset$ .

2.  $A \in \mathfrak{A} \implies \overline{A} \in \mathfrak{A}$ .

3.  $A, B \in \mathfrak{A} \implies A \cup B \in \mathfrak{A}$ .

**Beispiel 3.2.**  $\mathfrak{A} = \{\emptyset, \Omega\}$  ist eine Mengenalgebra (triviale Mengenalgebra).

**Beispiel 3.3.**  $\mathfrak{P}(\Omega)$  ist eine Mengenalgebra

**Beispiel 3.4.** Ist  $\Omega = \{1, 2, 3, 4\}$ , so ist

$$\mathfrak{A} = \{\emptyset, \{1\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\}$$

eine Mengenalgebra.

**Beispiel 3.5.** Ist  $\Omega = \mathbb{R}$  so ist

$$\mathfrak{A} = \{M \subseteq \mathbb{R} \mid M \text{ endlich oder } \overline{M} \text{ endlich}\}$$

eine Mengenalgebra.

Die einzige Bedingung, die hier eventuell Schwierigkeiten verursachen könnte, ist die dritte. Hier unterscheiden wir drei Fälle:

1. Fall:  $A$  und  $B$  sind endlich: In diesem Fall ist auch  $A \cup B$  endlich, und damit ist  $A \cup B \in \mathfrak{A}$ .
2. Fall:  $A$  und  $\overline{B}$  sind endlich (oder  $\overline{A}$  und  $B$  sind endlich): In diesem Fall gilt (nach den de Morganschen Gesetzen)

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \subseteq \overline{B}$$

und damit ist  $\overline{A \cup B}$  endlich, also  $A \cup B \in \mathfrak{A}$ .

3. Fall:  $\overline{A}$  und  $\overline{B}$  sind endlich (oder  $\overline{A}$  und  $B$  sind endlich): In diesem Fall gilt (nach den de Morganschen Gesetzen)

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \subseteq \overline{B}$$

und damit ist  $\overline{A \cup B}$  endlich, also  $A \cup B \in \mathfrak{A}$ .

Damit sind die drei Bedingungen erfüllt und  $\mathfrak{A}$  ist eine Mengenalgebra.

Einige einfache Regeln ergeben sich unmittelbar aus den Axiomen

**Regel 3.4.** Für jede Mengenalgebra  $\mathfrak{A}$  auf  $\Omega$  gilt:

- i)  $\emptyset \in \mathfrak{A}$  und  $\Omega \in \mathfrak{A}$ .
- ii) Sind  $A, B$  in  $\mathfrak{A}$ , so auch  $A \cap B$ .
- iii) Sind  $A, B$  in  $\mathfrak{A}$ , so auch  $A \setminus B$ .
- iv) Sind  $A_1, \dots, A_n$  in  $\mathfrak{A}$  so auch  $A_1 \cup \dots \cup A_n$  und  $A_1 \cap \dots \cap A_n$ .

**Beweis:** Nach Definition (1) gibt es sicher eine Menge  $A \in \mathfrak{A}$ . Damit gilt aber nach (2) auch  $\overline{A} \in \mathfrak{A}$ , und daher nach (3):

$$\Omega = A \cup \overline{A} \in \mathfrak{A}$$

Damit gilt, wiederum nach (2) aber auch

$$\emptyset = \overline{\Omega} \in \mathfrak{A}$$

und damit haben wir i) gezeigt.

Für  $A, B \in \mathfrak{A}$  gilt nach (2) zunächst

$$\overline{A} \in \mathfrak{A}, \quad \overline{B} \in \mathfrak{A}$$

Damit gilt nach (3)

$$\overline{A} \cup \overline{B} \in \mathfrak{A}$$

und daher (wieder nach (2))

$$\overline{\overline{A} \cup \overline{B}} \in \mathfrak{A}$$

Nach dem de Morganschen Gesetz für Mengen gilt aber

$$\overline{\overline{A} \cup \overline{B}} \in \mathfrak{A} = \overline{\overline{A}} \cap \overline{\overline{B}} = A \cap B$$

und es folgt ii).

Nach Definition ist mit  $A$  und  $B$  auch  $\overline{\overline{A} \cup \overline{B}}$  in  $\mathfrak{A}$ . Da

$$A \setminus B = \overline{\overline{A} \cup \overline{B}}$$

folgt hieraus iii)

iv) schließlich ergibt sich induktiv aus der Definition bzw. aus ii).

**Definition 3.8.** Eine Teilmenge  $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$  heißt  $\sigma$ -**Algebra** (auf  $\Omega$ ), wenn gilt:

1.  $\mathfrak{A} \neq \emptyset$ .
2.  $A \in \mathfrak{A} \implies \overline{A} \in \mathfrak{A}$ .
3.  $A_n \in \mathfrak{A}$  für alle  $n \in \mathbb{N} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{A}$ .

Ein  $A \in \mathfrak{A}$  heißt **Ereignis**.

**Beispiel 3.6.**  $\mathfrak{P}(\Omega)$  ist eine  $\sigma$ -Algebra. Ebenso ist die triviale Mengenalgebra  $\mathfrak{A} = \{\emptyset, \Omega\}$  eine  $\sigma$ -Algebra.

**Beispiel 3.7.** Das Mengensystem

$$\mathfrak{A} = \{\emptyset, \{1\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\}$$

auf  $\Omega = \{1, 2, 3, 4\}$  (aus Beispiel 3.4) ist eine  $\sigma$ -Algebra.

**Beispiel 3.8.** Das Mengensystem

$$\mathfrak{A} = \{M \subseteq \mathbb{R} \mid M \text{ ist endlich oder } \overline{M} \text{ ist endlich}\}$$

aus Beispiel 3.5 ist keine  $\sigma$ -Algebra. So sind etwa die Mengen  $A_n = \{n\}$  in  $\mathfrak{A}$ , aber nicht

$$\bigcup_{n \in \mathbb{N}} A_n = \mathbb{N}.$$

Dagegen ist das Mengensystem

$$\mathfrak{A} = \{M \subseteq \mathbb{R} \mid M \text{ ist abzählbar oder } \overline{M} \text{ ist abzählbar}\}$$

eine  $\sigma$ -Algebra.

Auch für  $\sigma$ -Algebren gelten einige leicht einzusehende Regeln:

**Regel 3.5.** Für eine  $\sigma$ -Algebra  $\mathfrak{A}$  gilt

- i)  $\mathfrak{A}$  ist eine Mengenalgebra.
- ii)  $\emptyset \in \mathfrak{A}$  und  $\Omega \in \mathfrak{A}$ .
- iii) Sind  $A, B$  in  $\mathfrak{A}$ , so auch  $A \setminus B$ .
- iv)  $A_n \in \mathfrak{A}$  für alle  $n \in \mathbb{N} \implies \bigcap_{n \in \mathbb{N}} A_n \in \mathfrak{A}$ .

**Beweis:** Aussage i) ist offensichtlich, und ii) und iii) folgen hieraus. iv) folgt wiederum aus dem Satz von de Morgan, der auch für beliebige Vereinigungen und Durchschnitte gilt.

## 3.2 Wahrscheinlichkeiten

Um mit Zufallsvorgängen arbeiten zu können benötigen wir noch den Begriff der Wahrscheinlichkeit. Dieser hängt eng zusammen mit Mengensystemen und auch mit der Integration.

**Definition 3.9.** Eine Paar  $(\Omega, \mathfrak{A})$ , bestehend aus einer Menge  $\Omega$  und einer  $\sigma$ -Algebra  $\mathfrak{A}$  auf  $\Omega$  heißt Messraum.

Wir fixieren jetzt einen Messraum  $(\Omega, \mathfrak{A})$ .

**Definition 3.10.** Eine Abbildung  $p : \mathfrak{A} \rightarrow [0, 1]$  heißt **Wahrscheinlichkeitsmaß** oder eine **Wahrscheinlichkeit** auf  $(\Omega, \mathfrak{A})$ , wenn sie die **Axiome von Kolmogoroff** erfüllt:

(K1)  $p(A) \geq 0$  für alle  $A \in \mathfrak{A}$ .

(K2)  $p(\Omega) = 1$ .

(K3) Sind  $A_n \in \mathfrak{A}$  paarweise disjunkt (also  $A_n \cap A_m = \emptyset$  für  $m \neq n$ ), so gilt

$$p\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} p(A_n)$$

**Bemerkung 3.3.** Ist  $p$  eine Wahrscheinlichkeit auf  $(\Omega, \mathfrak{A})$  und sind  $A, B$  zwei disjunkte Teilmengen, so gilt

$$p(A \cup B) = p(A) + p(B)$$

**Beispiel 3.9.** Wir betrachten  $\Omega = \{1, 2, 3, 4, 5, 6\}$  mit der  $\sigma$ -Algebra  $\mathfrak{A} = \mathfrak{P}(\Omega)$  und definieren für ein Teilmenge  $A \subseteq \Omega$ :

$$p(A) = \frac{|A|}{6}$$

Dann ist  $p$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathfrak{A})$ . Dieses Wahrscheinlichkeitsmaß beschreibt die Wahrscheinlichkeit mit einem (ungezinkten) Würfel eine bestimmte Zahl zu würfeln. So ist etwa die Wahrscheinlichkeit, eine gerade Zahl zu würfeln, genau

$$p(A) = \frac{3}{6} = \frac{1}{2}$$

wobei  $A = \{2, 4, 6\}$ .

**Beispiel 3.10.** Wir betrachten  $\Omega = \{1, 2\}$  mit der  $\sigma$ -Algebra  $\mathfrak{A} = \mathfrak{P}(\Omega)$  und definieren

$$p(\emptyset) = 0, \quad p(1) = 0.6, \quad p(2) = 0.5, \quad p(\{1, 2\}) = 1$$

Dann ist  $p$  kein Wahrscheinlichkeitsmaß auf  $\Omega$ , denn sonst müsste einerseits wegen (K3) gelten

$$p(\{1, 2\}) = p(\{1\} \cup \{2\}) = p(1) + p(2) = 1.1$$

andererseits aber wegen (K2)

$$p(\{1, 2\}) = p(\Omega) = 1$$

Setzen wir dagegen

$$p(\emptyset) = 0, \quad p(1) = 0.6, \quad p(2) = 0.4, \quad p(\{1, 2\}) = 1$$

so erhalten wir dadurch eine Wahrscheinlichkeit.

**Beispiel 3.11.** Ist  $\Omega$  eine beliebige endliche Menge der Mächtigkeit  $n$ , und ist  $\mathfrak{A} = \mathfrak{P}(\Omega)$ , so wird durch

$$p(A) = \frac{|A|}{n} \quad \text{für } A \subseteq \Omega$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathfrak{A})$  definiert. Bei zufälliger Auswahl eines Elements aus  $\Omega$  beschreibt dieses Wahrscheinlichkeitsmaß genau die Wahrscheinlichkeit, ein Element aus einer gegebenen Menge zu ziehen.

**Definition 3.11.** Ein **Wahrscheinlichkeitsraum** ist ein Tripel  $(\Omega, \mathfrak{A}, p)$ , bestehend aus einer Grundmenge  $\Omega$ , einer  $\sigma$ -Algebra  $\mathfrak{A}$  auf  $\Omega$  und einer Wahrscheinlichkeit  $p$  auf  $(\Omega, \mathfrak{A})$ .

**Regel 3.6.** Für eine Wahrscheinlichkeit  $p$  auf  $(\Omega, \mathfrak{A})$  gilt:

1.  $0 \leq p(A) \leq 1$  für alle  $A \in \mathfrak{A}$ .
2.  $p(\emptyset) = 0$ .
3.  $p(A) \leq p(B)$  für  $A \subseteq B$ .
4.  $p(\overline{A}) = 1 - p(A)$  für alle  $A \in \mathfrak{A}$
5.  $p(A \cup B) = p(A) + p(B) - p(A \cap B)$  und  $p(p(A \cap B) = p(A) + p(B) - p(A \cup B))$ .

**Beweis:**

1. Aus  $\Omega = A \cup \overline{A}$  und (K2) und (K3) folgt

$$1 = p(A) + p(\overline{A})$$

also nach (K1) notwendig  $0 \leq p(A) \leq 1$ .

2. Wir haben  $\emptyset = \emptyset \cup \emptyset$ , so dass also nach (K3) gelten muss

$$p(\emptyset) = 2p(\emptyset)$$

also  $p(\emptyset) = 0$ .

3. Wir schreiben  $B = A \cup C$  wobei  $C = B \setminus A$ . Damit gilt nach (K3) und (K1):

$$p(B) = p(A) + p(C) \geq p(A)$$



4. Das haben wir im Prinzip beim Nachweis von (1) schon gezeigt: Aus  $\Omega = A \cup \bar{A}$  und (K2) und (K3) folgt

$$1 = p(A) + p(\bar{A})$$

also  $p(\bar{A}) = 1 - p(A)$ .

5. Wir schreiben  $B = (A \cap B) \cup (B \setminus A)$  und  $A \cup B = A \cup (B \setminus A)$  (jeweils als disjunkte Vereinigung). Dann gilt

$$\begin{aligned} p(B) &= p(A \cap B) + p(B \setminus A) \\ p(A \cup B) &= p(A) + p(B \setminus A) \end{aligned}$$

woraus beide Formeln folgen.

**Bemerkung 3.4.** Ist  $\Omega$  endlich, so wollen wir im Zweifel immer die Potenzmenge als  $\sigma$ -Algebra auf  $\Omega$  betrachten.

Ist  $\Omega = \{\omega_1, \dots, \omega_n\}$ , so heißen die  $\{\omega_i\}$  auch die Elementarereignisse von  $\Omega$ , und wir schreiben kurz  $p_\omega$  für  $p(\{\omega\})$ .

**Bemerkung 3.5.** Ist  $\Omega$  endlich, so ist

$$p(A) = \frac{|A|}{|\Omega|} \quad \text{für } A \subseteq \Omega$$

eine Wahrscheinlichkeit auf  $\Omega$ . Dieses Wahrscheinlichkeitsmaß heißt **Laplace-Maß** oder **Laplace-Wahrscheinlichkeit** auf  $\Omega$ .

**Beispiel 3.12.** Die Wahrscheinlichkeit, mit einem (nicht-manipulierten) Würfel eine bestimmte Zahl zwischen 1 und 6 zu werfen, beträgt genau  $\frac{1}{6}$ . Es ist also eine Laplace-Wahrscheinlichkeit.

**Beispiel 3.13.** Bei einer (nicht-manipulierten) Münze mit Wappen  $W$  und Zahl  $Z$  ist die Wahrscheinlichkeit *Wappen* oder *Zahl* zu werfen jeweils gegeben durch

$$p(W) = p(Z) = \frac{1}{2}$$

Es ist also eine Laplace-Wahrscheinlichkeit.

**Beispiel 3.14.** Eine manipulierte Münze mit Wappen  $W$  und Zahl  $Z$  ist dadurch gekennzeichnet, dass die Wahrscheinlichkeit *Wappen* oder *Zahl* zu werfen unterschiedlich sind, etwa

$$p(W) = 0.60, \quad p(Z) = 0.40$$

Es ist also keine Laplace-Wahrscheinlichkeit.

**Beispiel 3.15.** Wählen wir aus dem Intervall  $[0, 5]$  zufällig eine beliebige reelle Zahl aus, so hat jede Zahl die gleich Wahrscheinlichkeit, gezogen zu werden. Da es aber in diesem Fall unendlich viele Zahlen gibt, bedeutet das auch, dass für jede einzelne Zahl nur die „Wahrscheinlichkeit 0“ übrigbleibt.

Tatsächlich werden aber natürlich Zahlen gewählt, und bei einer reinen Zufallsauswahl ist die Wahrscheinlichkeit (zumindest die, die wir erwarten würden), dass die gewählte Zahl höchstens 2 ist gerade 0.40, und die Wahrscheinlichkeit, dass eine Zahl zwischen 3 und 4 gezogen ist, gleich 0.20. Positive Wahrscheinlichkeiten können also erst größeren Bereichen (etwa Intervallen) zugeordnet werden. Für  $0 \leq a \leq b \leq 5$  gilt dann

$$p([a, b]) = \frac{b - a}{5}$$

Dadurch lässt sich tatsächlich eine Wahrscheinlichkeit definieren. Die passende  $\sigma$ -Algebra ist die kleinste  $\sigma$ -Algebra, die alle Teilintervalle  $I \subseteq [0, 5]$  enthält (die sogenannte  $\sigma$ -Algebra der Borelmengen). Die korrekte Definition der Wahrscheinlichkeit auf einem Ereignis  $A$  ist dann

$$p(A) = \int_A \frac{1}{5} dx$$

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 4

Stand: April 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim

### 3.4 Kombinatorik und Zufallsstichproben

Bei endlichen Wahrscheinlichkeitsräumen und Laplace-Experimenten lassen sich die Wahrscheinlichkeiten sehr gut durch Abzählen bestimmen. Empirisch arbeiten wir hier mit sogenannten *Zufallsstichproben*, die üblicherweise mit dem Urnenmodell dargestellt wird:

Jeder Einheit  $\omega \in \Omega$  aus der Grundgesamtheit wird eine numerierte Kugel  $E$  zugeordnet und in die Urne gelegt. Eine Folge gezogener Kugeln, also ein geordnetes  $n$ -Tupel  $(E_1, \dots, E_n)$  nennen wir eine Stichprobe vom Umfang  $n$ .

**Definition 3.12.** Eine *einfache Zufallsstichprobe* liegt vor, wenn jedes Stichprobe vom Umfang  $n$  aus einer Grundgesamtheit vom Umfang  $N$  mit derselben Wahrscheinlichkeit gezogen wird.

Die Berechnungen, welche Wahrscheinlichkeiten sich daraus ergeben, führen sehr schnell zu sehr komplizierten Formeln. Dazu benötigen wir einige kombinatorische Grundlagen.

**Satz 3.7.** (*Fundamentalprinzip der Kombinatorik*)

*Sind  $M_1, \dots, M_n$  endliche Mengen mit jeweils  $a_1, \dots, a_n$  Elementen, so gibt es genau  $a_1 \cdots a_n$  viele Möglichkeiten aus jeder Menge jeweils genau ein Element auszuwählen.*

**Beweis:** Die Auswahl je eines Elements aus  $M_1, \dots, M_n$  ist gegeben durch ein  $n$ -Tupel  $(m_1, \dots, m_n)$ , wobei  $m_i \in M_i$ , und umgekehrt legt auch jedes solche  $n$ -Tupel  $(m_1, \dots, m_n)$  genau eine Auswahl von je einem Element aus  $M_1, \dots, M_n$  fest. Damit entspricht also die Anzahl, die wir suchen, der Mächtigkeit des kartesischen Produktes  $M_1 \times \cdots \times M_n$ . Diese Zahl kann nun leicht induktiv ermittelt werden:

Für  $n = 1$  gilt nach Voraussetzung  $|M_1| = a_1$ .

Für  $n = 2$  erhalten wir das Ergebnis unmittelbar durch Nachzählen:

$$|M_1 \times M_2| = |M_1| \cdot |M_2| = a_1 \cdot a_2$$

Hieraus erhalten wir induktiv die allgemeine Aussage, denn

$$M_1 \times \cdots \times M_{n-1} \times M_n = (M_1 \times \cdots \times M_{n-1}) \times M_n$$

**Beispiel 3.16.** Sie haben fünf Paar Schuhe und neun Pullover, und jedes Paar Schuhe kann mit jedem Pullover kombiniert werden. Dann können Sie 45 Tage lang jeden Morgen mit einer neuen Schuhe/Pullover-Kombination aus dem Haus gehen.

**Beispiel 3.17.** Ein Passwort muss aus drei Großbuchstaben gefolgt von zwei bis vier Ziffern bestehen. Dann gibt es hierfür insgesamt

$$n = 26^3 \cdot (10^2 + 10^3 + 10^4) = 195\,093\,600$$

viele Möglichkeiten (also  $26^3 \cdot 10^2$  für ein Passwort mit drei Buchstaben, zwei Ziffern,  $26^3 \cdot 10^3$  für ein Passwort mit drei Buchstaben, drei Ziffern und  $26^3 \cdot 10^4$  für ein Passwort mit drei Buchstaben, vier Ziffern).

In unserem Satz 3.7 müssen die Mengen  $M_1, \dots, M_n$  nicht notwendig verschieden sein. Der Fall, dass alle Mengen gleich sind,  $M_i = M$  für alle  $i$ , ist sogar von besonderen Interesse, denn er liefert die Antwort auf die folgende Fragestellung:

$n$  Personen ziehen der Reihe nach eine Zahl aus einem Hut mit  $a$  Zahlen. Jeder Teilnehmer merkt sich seine Nummer und legt den Zettel dann zurück in den Hut (*Ziehen mit Zurücklegen*). Wieviele mögliche Zahlenfolgen können dadurch entstehen? Hier werden die möglichen Kombinationen offensichtlich gegeben durch die Elemente des kartesischen Produktes

$$\underbrace{M \times \dots \times M}_{n\text{-mal}} = M^n$$

wobei  $M$  die Menge der  $a$  Zahlen ist, und daher gilt für das Ziehen eine Zufallsstichprobe mit Zurücklegen:

**Satz 3.8.** *Ist  $M$  eine endliche Menge mit  $a$  Elementen aus der  $n$ -mal ein Element mit Zurücklegen gezogen wird, so gibt es insgesamt  $Z(a, n) = a^n$  verschiedene mögliche Elementekombinationen.*

Ein ebenfalls sehr häufiger Fall ist der des Ziehens ohne Zurücklegen. Jeder Teilnehmer zieht also eine Zahl aus einem Hut und behält den Zettel mit dieser Zahl. Der nächste Teilnehmer kann nur noch einen Zettel mit einer noch nicht gezogenen Zahl bekommen.

**Definition 3.13.** Ist  $a \in \mathbb{N}$  so definieren wir die Zahl  $a! \in \mathbb{N}$ , genannt " $a$ -Fakultät", induktiv wie folgt:

- $0! = 1$
- Ist  $(a - 1)!$  schon definiert, so setzen wir  $a! = a \cdot (a - 1)!$ .

**Beispiel 3.18.** Es gilt

1.  $1! = 1$ .

2.  $2! = 2$ .

3.  $3! = 6$ .

4.  $4! = 24$ .

**Satz 3.9.** *Ist  $M$  eine endliche Menge mit  $a$  Elementen aus der  $n$ -mal ein Element ohne Zurücklegen gezogen wird, so gibt es insgesamt  $V(a, n) := \frac{a!}{(a-n)!}$  verschiedene mögliche Elementekombinationen.*

**Beweis:** Auch das lässt sich am besten induktiv beweisen:

Für  $n = 1$  wird ein Element aus einer Menge von  $a$  Elementen gezogen, und dafür gibt es natürlich genau  $a = \frac{a!}{(a-1)!}$  verschiedene Möglichkeiten.

Sei jetzt  $n \geq 2$  und die Behauptung für  $(n-1)$ -maliges Ziehen schon gezeigt. Unabhängig von der Kombination, die in den  $n-1$  ersten Zügen gezogen wurde, sind beim  $n$ -ten Zug noch  $a-n+1$  Elemente in der zur Verfügung stehenden Menge, wir haben also für jede der schon gewählten  $(n-1)$ -elementigen Kombinationen genau  $a-n+1$  Möglichkeiten, diese zu einer  $n$ -elementigen zu ergänzen. Damit ergibt sich für die Gesamtzahl gemäß Induktionsvoraussetzung

$$(a-n+1) \cdot \frac{a!}{(a-n+1)!} = (a-n+1) \cdot \frac{a!}{(a-n+1) \cdot (a-n)!} = \frac{a!}{(a-n)!}$$

**Beispiel 3.19.** Anton wählt eine dreiziffrige PIN aus, indem er in einen Topf 10 Zettel mit den Ziffern  $0, 1, \dots, 9$  legt, dreimal hintereinander daraus einen Zettel zieht und die Zahlen auf diesen Zetteln in der gezogenen Reihenfolge aufschreibt. Legt er die gezogene Zahl nach dem Aufschreiben und vor dem nächsten Ziehen gleich wieder in den Topf zurück, so ergeben sich für die PIN insgesamt  $10^3 = 1000$  Möglichkeiten, nämlich alle Zahlen zwischen 000 und 999. Legt er die gezogenen Zahlen dagegen nicht zurück (und lässt damit nur PINs zu, bei denen keine Ziffer doppelt vorkommt), so gibt es noch  $10 \cdot 9 \cdot 8 = 720$  mögliche Geheimnummern.

**Beispiel 3.20.** Die Anzahl der Kombinationen für ein Passwort, das aus drei voneinander verschiedenen Großbuchstaben, gefolgt von zwei bis vier jeweils paarweise verschiedenen Ziffern bestehen muss, ist

$$26 \cdot 25 \cdot 24 \cdot (10 \cdot 9 + 10 \cdot 9 \cdot 8 + 10 \cdot 9 \cdot 8 \cdot 7) = 91\,260\,000$$

Eine besonders interessante Situation tritt auf, wenn  $a = n$ . In diesem Fall werden also alle Elemente aus der Menge  $M$  gezogen.

**Satz 3.10.** *Es gibt genau  $a!$  verschiedene Möglichkeiten, die Elemente einer Menge  $M$  der Mächtigkeit  $a$  anzuordnen.*

Das klärt aber noch nicht alle für uns relevanten kombinatorischen Fragen. Für viele interessant ist etwa die Frage nach den Möglichkeiten beim Lotto 6 aus 49. Dort wird zwar auch sechsmal eine Kugel aus einem Topf mit 49 Kugeln gezogen und die gezogenen Kugeln werden nicht mehr zurückgelegt. Für das Endergebnis spielt aber auch die Reihenfolge der gezogenen Elemente keine Rolle mehr, die Zugfolge 43, 2, 7, 28, 19, 3 ist also etwa für den Lottospieler gleichbedeutend mit der Zugfolge 2, 3, 7, 43, 19, 28 oder jeder anderen Reihenfolge dieser Zahlen. Wir interessieren uns in diesem Fall also nicht mehr für die Anzahl der (geordneten) 6-Tupel, die wir durch Ziehen ohne Zurücklegen aus der Menge  $\{1, \dots, 49\}$  erhalten können, sondern nur noch für die Anzahl der (ungeordneten) 6-elementigen Teilmengen aus der Menge  $\{1, \dots, 49\}$ . Auch hierfür gibt es eine Formel

**Satz 3.11.** *Aus einer Menge  $M$  von  $a$  Elementen lassen sich*

$$B(a, n) := \frac{a!}{(a-n)! \cdot n!}$$

*verschiedene  $n$ -elementige Teilmengen auswählen.*

**Beweis:** Wie wir schon in Satz 3.9 gesehen haben, lassen sich aus  $M$  insgesamt  $V(a, n) = \frac{a!}{(a-n)!}$  verschiedene  $n$ -Tupel  $(m_1, \dots, m_n)$  mit  $m_i \neq m_j$  für  $i \neq j$  auswählen. Dabei gibt es für eine gegebene  $n$ -elementige Teilmenge  $\{m_1, \dots, m_n\}$  nach Satz 3.10 genau  $n!$  verschiedenen Möglichkeiten, alle ihre Elemente als geordnete  $n$ -Tupel anzuordnen, und damit gilt

$$B(a, n) = \frac{V(a, n)}{n!} = \frac{a!}{(a-n)! \cdot n!}$$

**Definition 3.14.**  $\binom{a}{n} := \frac{a!}{(a-n)!n!}$  heißt Binomialkoeffizient von  $a$  und  $n$ , kurz auch "a über n" oder "n aus a".

**Beispiel 3.21.** Damit können wir endlich unsere Lottoaufgabe lösen: Es gibt insgesamt  $\binom{49}{6} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 13983186$  verschiedene Kombinationen, die als Ergebnis einer Lottoziehung in Frage kommen.

**Beispiel 3.22.** Die Anzahl der Kombinationen für ein Passwort, das aus drei voneinander verschiedenen Großbuchstaben und zwei bis vier jeweils paarweise verschiedenen

Ziffern bestehen muss (wobei die Buchstaben jetzt an beliebigen Stellen stehen können), ist

$$\binom{5}{3} \cdot 26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 + \binom{6}{3} \cdot 26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 + \binom{7}{3} \cdot 26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 2\,990\,520\,000$$

**Beispiel 3.23.** Ein Verein mit einer Vorstandschaft von 9 Mitgliedern schickt eine Vorstandsdelegation zum jährlichen Frühjahrsumzug der Stadt. Da dort immer in Zweierreihen marschiert wird hat der Verein

$$\binom{9}{2} + \binom{9}{4} + \binom{9}{6} + \binom{9}{8} = 255$$

viele Möglichkeiten eine (nichtleere) Delegation zusammenzustellen.

Binomialkoeffizienten spielen eine so wichtige Rolle in der Mathematik, dass wir sie hier noch ein wenig genauer studieren wollen:

**Bemerkung 3.6.** Es gilt

1.  $\binom{a}{a} = 1$ .
2.  $\binom{a}{a-1} = a$ .
3.  $\binom{a}{n} = 0$  für  $n > a$ .
4.  $\binom{a}{n} = \binom{a}{a-n}$ .

**Satz 3.12.** Für  $1 \leq n \leq a$  gilt:

$$\binom{a}{n} = \binom{a-1}{n-1} + \binom{a-1}{n}$$

**Beweis:** Für  $n = a$  stimmt die Formel offensichtlich (nach obiger Bemerkung), und für  $1 \leq n \leq a-1$  gilt

$$\begin{aligned} \binom{a-1}{n-1} + \binom{a-1}{n} &= \frac{(a-1)!}{(n-1)!(a-n)!} + \frac{(a-1)!}{n!(a-n-1)!} \\ &= \frac{n(a-1)! + (a-n)(a-1)!}{n!(a-n)!} \\ &= \frac{a!}{n!(a-n)!} \\ &= \binom{a}{n} \end{aligned}$$



Hieraus ergibt sich das bekannte Dreiecksschema, das zur Berechnung der binomischen Zahlen herangezogen werden kann

$$\begin{array}{ccccccc}
 & & & & 1 & & \\
 & & & & 1 & & 1 \\
 & & & 1 & & 2 & & 1 \\
 & & 1 & & 3 & & 3 & & 1 \\
 & 1 & & 4 & & 6 & & 4 & & 1 \\
 & \vdots & & & & \vdots & & & & \vdots
 \end{array}$$

bei der die Binomialzahl  $\binom{a}{n}$  in der  $a + 1$ -ten Zeile an der  $n + 1$ -ten Stelle steht und sich durch Addition der beiden schräg links und schräg rechts darüberstehenden Binomialkoeffizienten  $\binom{a-1}{n-1}$  und  $\binom{a-1}{n}$  ergibt.

Aus der Schule kennen Sie alle die binomischen Formeln, etwa

$$(x + y)^2 = x^2 + 2xy + y^2$$

Mit den Binomialkoeffizienten bekommen wir nun eine gewaltige Verallgemeinerung hiervon, nämlich

**Satz 3.13.** *Für beliebige Zahlen  $x, y$  gilt*

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} \cdot x^i \cdot y^{n-i}$$

**Beweis:** Wir beweisen mit vollständiger Induktion nach  $n$ :

Für  $n = 1$  ist die Aussage  $((x + y)^1 = x + y = \binom{1}{0} x^0 y^1 + \binom{1}{1} x^1 y^0)$  trivial.

Sei also jetzt  $n \geq 2$  und die Aussage für  $n - 1$  schon bewiesen. Wir rechnen

$$\begin{aligned}
 (x + y)^n &= (x + y)^{n-1} \cdot (x + y) \\
 &= \left( \sum_{i=0}^{n-1} \binom{n-1}{i} x^i y^{n-1-i} \right) \cdot (x + y) && \text{(Induktion)} \\
 &= \sum_{i=0}^{n-1} \binom{n-1}{i} x^{i+1} y^{n-1-i} + \sum_{i=0}^{n-1} \binom{n-1}{i} x^i y^{n-i} \\
 &= x^n + \sum_{i=1}^{n-1} \left[ \binom{n-1}{i} + \binom{n-1}{i-1} \right] x^i y^{n-i} + y^n \\
 &= x^n + \sum_{i=1}^{n-1} \binom{n}{i} x^i y^{n-i} + y^n && \text{(Satz 3.12)} \\
 &= \sum_{i=0}^n \binom{n}{i} x^i y^{n-i}
 \end{aligned}$$

**Folgerung 3.14.**  $\sum_{i=0}^n \binom{n}{i} = 2^n$ .

**Beweis:** Das erhält man aus Satz 3.13, wenn man  $x = 1$  und  $y = 1$  setzt.

**Folgerung 3.15.** Ist  $A$  eine Menge mit  $a$  Elementen, so hat  $A$  genau  $2^a$  viele verschiedene Teilmengen, also

$$|\mathfrak{P}(A)| = 2^{|A|}$$

wobei  $\mathfrak{P}(A)$  die Potenzmenge von  $A$  bezeichnet.

**Beweis:** Die Menge  $A$  hat  $\binom{a}{n}$  viele  $n$ -elementige Teilmengen. Lässt man nun  $n$  von 0 bis  $a$  laufen, so erhält man insgesamt

$$\sum_{n=0}^a \binom{a}{n} = 2^a$$

viele Möglichkeiten.

**Folgerung 3.16** (Vandermonde-Identität). Für ganze Zahlen  $N, M$  und  $n \geq 0$  gilt

$$\sum_{k=0}^n \binom{N}{k} \cdot \binom{M}{n-k} = \binom{N+M}{n}$$

**Beweis:** Wir wenden den binomischen Satz 3.13 auf das Polynom  $(1+x)^{N+M}$  an. Hierfür gilt einerseits

$$(1+x)^{N+M} = \sum_{n=0}^{N+M} \binom{N+M}{n} x^n$$

und andererseits

$$\begin{aligned} (1+x)^{N+M} &= (1+x)^N \cdot (1+x)^M \\ &= \left( \sum_{n=0}^N \binom{N}{n} \cdot x^n \right) \cdot \left( \sum_{m=0}^M \binom{M}{m} \cdot x^m \right) \\ &= \sum_{n=0}^N \sum_{m=0}^M \binom{N}{n} \binom{M}{m} \cdot x^m \cdot x^n \\ &= \sum_{n=0}^{N+M} \sum_{k=0}^n \binom{N}{k} \binom{M}{n-k} \cdot x^n \end{aligned}$$

Durch Koeffizientenvergleich erhalten wir die Behauptung.

Eine kombinatorische Verallgemeinerung der Binomialkoeffizienten sind die Multinomialkoeffizienten:

**Bemerkung 3.7.** Die Anzahl der Möglichkeiten, die  $a$  Elemente einer Menge  $M$  in Gruppen  $M_1, \dots, M_r$  mit festen Größen  $n_1, \dots, n_r$  aufzuteilen (wobei  $n_1 + \dots + n_r = a$  gelten muss), ist gegeben durch den Multinomialkoeffizienten

$$\binom{a}{n_1, n_2, \dots, n_r} = \frac{a!}{n_1! \cdot n_2! \cdot \dots \cdot n_r!}$$

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 5

Stand: April 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim

### 3.5 Bedingte Wahrscheinlichkeiten

Wir können nun verschiedene Ereignisse betrachten und uns damit beschäftigen, ob diese Ereignisse zusammenhängen, und ob das Auftreten eines Ereignisses die Wahrscheinlichkeit, dass ein anderes eintritt, beeinflusst. Bei der Urlaubsplanung etwa ist es nicht so sehr von Interesse, zu wissen, wie hoch die Regenwahrscheinlichkeit an einem durchschnittlichen Tag im Jahr ist, sondern wie hoch sie ist, wenn schon bekannt ist, dass dieser Tag im August liegt. Mit ähnlichen Fragestellungen haben wir uns ja schon in der deskriptiven Statistik beschäftigt.

**Definition 3.15.** Die *bedingte Wahrscheinlichkeit*  $p(A|B)$  für das Ereignis  $A$  unter der Bedingung  $B$  ist die Wahrscheinlichkeit dafür, dass das Ereignis  $A$  eintritt, wenn das Ereignis  $B$  vorliegt.

**Bemerkung 3.8.** Die Definition gibt uns keinerlei Handhabe, die bedingte Wahrscheinlichkeit zu berechnen. In der allgemeinen Situation ist das auch sehr kompliziert. Falls aber  $p(B) > 0$  ist, so gilt

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

**Beispiel 3.24.** Wir betrachten einen Laplace-Würfel und die beiden Ereignisse

$$\begin{aligned} A &= \text{eine 3 wird gewürfelt} \\ B &= \text{eine ungerade Zahl wird gewürfelt} \end{aligned}$$

Dann gilt

$$p(A|B) = \frac{1}{3}$$

denn  $A \cap B = A$  und  $p(A) = \frac{1}{6}$ ,  $p(B) = \frac{1}{2}$ .

**Beispiel 3.25.** Wir betrachten einen Laplace-Würfel und die beiden Ereignisse

$$\begin{aligned} A &= \text{eine der Zahlen 1, 2, 3, oder 4 wird gewürfelt} \\ B &= \text{eine gerade Zahl wird gewürfelt} \end{aligned}$$

Dann gilt

$$p(A|B) = \frac{p(\{2, 4\})}{p(\{2, 4, 6\})} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$$

und

$$p(B|A) = \frac{p(\{2, 4\})}{p(\{1, 2, 3, 4\})} = \frac{\frac{2}{6}}{\frac{4}{6}} = \frac{1}{2}$$

**Bemerkung 3.9.** Ist  $(\Omega, \mathfrak{A}, p)$  ein Wahrscheinlichkeitsraum und ist  $B \in \mathfrak{A}$ , so ist

$$\mathfrak{A}_B := \{A \cap B \mid A \in \mathfrak{A}\}$$

eine  $\sigma$ -Algebra auf  $B$  und

$$p(\cdot \mid B) : \mathfrak{A}_B \longrightarrow \mathbb{R}$$

mit  $p(\cdot \mid B) = p(A|B)$  ist eine Wahrscheinlichkeit auf  $B$ .

**Definition 3.16.** Zwei Ereignisse  $A$  und  $B$  heißen (*stochastisch*) *unabhängig*, wenn gilt

$$p(A|B) = p(A), \quad p(B|A) = p(B)$$

**Bemerkung 3.10.** Stochastische Unabhängigkeit bedeutet, dass die Wahrscheinlichkeit dafür, dass ein Ereignis  $A$  eintritt unabhängig davon ist, ob man weiss, dass ein anderes Ereignis  $B$  vorliegt oder nicht.

**Beispiel 3.26.** Die beiden Ereignisse  $A$  und  $B$  aus Beispiel 3.24 sind nicht stochastisch unabhängig. In diesem Fall gilt nämlich

$$p(A) = \frac{1}{6}$$

allerdings ist

$$p(A|B) = \frac{1}{3}$$

**Beispiel 3.27.** Die beiden Ereignisse  $A$  und  $B$  aus Beispiel 3.24 sind stochastisch unabhängig. In diesem Fall gilt nämlich

$$p(A|B) = \frac{2}{3} = p(A), \quad p(B|A) = \frac{1}{2} = p(B)$$

Alternativ können wir auch so argumentieren:

$$p(B \cap A) = p(\{2, 4\}) = \frac{1}{3} = \frac{2}{3} \cdot \frac{1}{2} = p(A) \cdot p(B)$$

**Beispiel 3.28.** Wir würfeln zweimal (unabhängig voneinander) mit einem Laplace-Würfel und schreiben die Ergebnisse als geordnetes Zahlenpaar  $(a, b)$  auf. Dann gilt für jedes Elementarereignis  $E = \{(a, b)\}$ :

$$P(E) = \frac{1}{36}$$

da insgesamt 36 (gleichwahrscheinliche) Zahlenpaare auftreten können. Betrachten wir die Ereignisse

$$\begin{aligned} A &= \text{im ersten Wurf wird eine 6 gewürfelt} \\ B &= \text{im zweiten Wurf wird eine 6 gewürfelt} \end{aligned}$$

so gilt  $|A| = 6 = |B|$  also

$$p(A) = p(B) = \frac{1}{6}$$

Ferner ist  $A \cap B = \{(6, 6)\}$  und damit gilt

$$p(A \cap B) = \frac{1}{36}$$

Also gilt

$$p(A|B) = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} = p(A), \quad p(B|A) = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} = p(B)$$

und damit sind die beiden Ereignisse unabhängig.

**Beispiel 3.29.** Wir würfeln zweimal hintereinander mit einem Laplace-Würfel, gehen dabei aber vor wie folgt: Fällt im ersten Wurf eine gerade Zahl, so wird ganz normal mit dem zweiten Laplacewürfel geworfen und das Ergebnis der beiden Würfe wird als geordnetes Zahlenpaar  $(a, b)$  notiert. Fällt dagegen im ersten Wurf eine ungerade Zahl, so wird auf dem zweiten Laplacewürfel die 1 mit einer 2, die 3 mit einer 4 und die 5 mit einer 6 überklebt und erst dann geworfen (wobei aber immer noch jede Seite gleichwahrscheinlich sein soll) und das Ergebnis der beiden Würfe wird als geordnetes Zahlenpaar  $(a, b)$  notiert. Dann gilt hierfür z.B.

$$p((2, 4)) = \frac{1}{36}$$

aber

$$p((1, 4)) = \frac{2}{36}$$

(denn genau betrachtet ist ja  $(1, 4) = \{(1, 4_1), (1, 4_2)\}$ , und sowohl  $(1, 4_1)$  als auch  $(1, 4_2)$  haben die Wahrscheinlichkeit  $\frac{1}{36}$ . (Betrachten wir die Ereignisse

$$\begin{aligned} A &= \text{im ersten Wurf wird eine 6 gewürfelt} \\ B &= \text{im zweiten Wurf wird eine 6 gewürfelt} \end{aligned}$$

so sind diese jetzt nicht mehr stochastisch unabhängig, denn es gilt zwar  $|A| = 6 = |B|$  aber

$$p(A) = \frac{6}{36}, \quad p(B) = \frac{9}{36}$$

Ferner ist  $A \cap B = \{(6, 6)\}$  und hierfür gilt

$$p(A \cap B) = \frac{1}{36}$$

Also gilt

$$p(A|B) = \frac{\frac{1}{36}}{\frac{9}{36}} = \frac{1}{9} \neq p(A), \quad p(B|A) = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} \neq p(B)$$

und damit sind die beiden Ereignisse nicht stochastisch unabhängig.

**Beispiel 3.30.** Beim Ziehen aus einer Urne mit Zurücklegen sind die Ergebnisse der einzelnen Ziehungen unabhängig, beim Ziehen ohne Zurücklegen sind sie abhängig.

**Satz 3.17.** *Genau dann sind zwei Ereignisse  $A$  und  $B$  mit  $p(A) > 0$  und  $p(B) > 0$  stochastisch unabhängig, wenn gilt*

$$p(A \cap B) = p(A) \cdot p(B)$$

**Beweis:** Nach Bemerkung 3.8 gilt

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \quad \text{und} \quad p(B|A) = \frac{p(A \cap B)}{p(A)}$$

Da  $A$  und  $B$  genau dann unabhängig sind, wenn  $p(A) = p(A|B)$  und  $p(B) = p(B|A)$ , folgt hieraus die Behauptung.

**Satz 3.18** (Satz von der totalen Wahrscheinlichkeit). *Ist  $A_1, \dots, A_k$  eine disjunkte Zerlegung von  $\Omega$ , gilt also*

$$\Omega = A_1 \cup A_2 \cup \dots \cup A_k$$

*wobei  $A_i \cap A_j = \emptyset$  für  $i \neq j$ , und ist  $p(A_i) > 0$  für alle  $i$ , so gilt für jedes Ereignis  $B$ :*

$$p(B) = \sum_{i=1}^k p(B|A_i) \cdot p(A_i)$$

**Beweis:** Es ist

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_k)$$

eine disjunkte Zerlegung von  $B$ , und damit gilt

$$\begin{aligned} p(B) &= \sum_{i=1}^k p(B \cap A_i) \\ &= \sum_{i=1}^k \frac{p(B \cap A_i)}{p(A_i)} \cdot p(A_i) \\ &= \sum_{i=1}^k p(B|A_i) \cdot p(A_i) \end{aligned}$$

Besonders wichtig ist der Fall  $k = 2$ :

**Folgerung 3.19.** *Ist  $A$  ein Ereignis mit  $0 < p(A) < 1$ , so gilt für jedes Ereignis  $B$ :*

$$p(B) = p(B|A) \cdot p(A) + p(B|\bar{A}) \cdot p(\bar{A})$$



**Beispiel 3.31.** Ein Test auf eine Infektion ist zu 99.9 % positiv, wenn der Proband infiziert ist. Der Test ist aber auch mit einer Wahrscheinlichkeit von 5.0 % positiv, wenn der Proband nicht infiziert ist.

Insgesamt sind 3.0 % der getesteten Personen tatsächlich infiziert.

Dann kann daraus die Wahrscheinlichkeit für einen positiven Test nach dem Satz von der totalen Wahrscheinlichkeit bestimmt werden. Dazu bezeichne

$$\begin{aligned} P : & \quad \text{Test ist positiv} \\ K : & \quad \text{Patient ist infiziert} \end{aligned}$$

Dann wissen wir

$$p(P|K) = 0.999, \quad p(P|\bar{K}) = 0.05, \quad p(K) = 0.03$$

Damit wissen wir natürlich auch

$$p(\bar{K}) = 1 - p(K) = 0.97$$

und erhalten damit

$$\begin{aligned} p(P) &= p(P|K) \cdot p(K) + p(P|\bar{K}) \cdot p(\bar{K}) \\ &= 0.999 \cdot 0.03 + 0.05 \cdot 0.97 \\ &= 0.07847 \end{aligned}$$

Der Test fällt als mit einer Wahrscheinlichkeit von 7.847 % positiv aus.

In gewissen Situationen lässt sich aus  $p(B|A)$  auch  $p(A|B)$  berechnen:

**Satz 3.20** (Satz von Bayes). *Ist  $A_1, \dots, A_k$  eine disjunkte Zerlegung von  $\Omega$ , und ist  $p(A_i) > 0$  für alle  $i$ , so gilt für jedes Ereignis  $B$  mit  $p(B) > 0$  und jedes  $j \in \{1, \dots, k\}$ :*

$$p(A_j|B) = \frac{p(B|A_j) \cdot p(A_j)}{\sum_{i=1}^k p(B|A_i) \cdot p(A_i)} = \frac{p(B|A_j) \cdot p(A_j)}{p(B)}$$

**Beweis:** Die zweite Umformung ergibt sich sofort aus dem Satz von der totalen Wahrscheinlichkeit, und daher ist hier nur

$$p(A_j|B) = \frac{p(B|A_j) \cdot p(A_j)}{p(B)}$$

zu zeigen. Das ergibt sich aber nun leicht aus der Definition:

$$\frac{p(B|A_j) \cdot p(A_j)}{p(B)} = \frac{\frac{p(B \cap A_j)}{p(A_j)} \cdot p(A_j)}{p(B)} = \frac{p(B \cap A_j)}{p(B)} = p(A_j|B)$$

und die Formel ist gezeigt.

Auch hier ist der Fall  $k = 2$  von speziellem Interesse:

**Folgerung 3.21.** Ist  $A$  ein Ereignis mit  $0 < p(A) < 1$ , so gilt für jedes Ereignis  $B$ :

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B|A) \cdot p(A) + p(B|\bar{A}) \cdot p(\bar{A})} = \frac{p(B|A) \cdot p(A)}{p(B)}$$

**Beispiel 3.32.** Wir betrachten einen medizinischen Test auf eine Virusinfektion, und wir betrachten die Ereignisse

$A$  = der Patient hat die Virusinfektion  
 $B$  = das Testergebnis ist positiv

Bekannt ist

$$\begin{aligned} p(B|A) &= 0.99 \\ p(B|\bar{A}) &= 0.02 \\ p(A) &= 0.002 \end{aligned}$$

Wie hoch ist die Wahrscheinlichkeit, dass der Patient in der Tat krank ist, wenn das Testergebnis positiv ist, also wie hoch ist  $p(A|B)$ ? Nach dem Satz von Bayes gilt

$$\begin{aligned} p(A|B) &= \frac{p(B|A) \cdot p(A)}{p(B|A) \cdot p(A) + p(B|\bar{A}) \cdot p(\bar{A})} \\ &= \frac{0.99 \cdot 0.002}{0.99 \cdot 0.002 + 0.02 \cdot 0.998} \\ &\approx 0.09 \end{aligned}$$

Die Wahrscheinlichkeit, an dem Virus erkrankt zu sein, wenn das Testergebnis positiv ist, liegt also nur bei ca. 9 %.

**Beispiel 3.33.** Wir greifen Beispiel 3.31 noch einmal auf und wollen auch in diesem Fall wissen, wie hoch die Wahrscheinlichkeit ist, dass ein positiv getesteter Proband tatsächlich infiziert ist. Dazu erhalten wir mit den Daten von Beispiel 3.31 aus dem Satz von Bayes

$$p(K|P) = \frac{p(P|K) \cdot p(K)}{p(P)} = \frac{0.999 \cdot 0.03}{0.07847} = 0.382$$

In diesem Beispiel liegt also die Wahrscheinlichkeit, infiziert zu sein, wenn man positiv getestet worden ist, bei etwa 38.2 %.

**Beispiel 3.34.** In einem Obstverarbeitungsbetrieb werden Äpfel je nach Qualität unterschiedlich verarbeitet. Minderwertige Früchte werden versaftet, die verbleibenden Äpfel werden in die Handelsklassen I und II aufgeteilt.

Bekannt ist, dass 25.00 % der Ernte minderwertig sind. Vom Rest erfüllt 40.00 % die Anforderungen der Handelsklasse I. Zur Aussortierung der minderwertigen Äpfel wird

eine Sortiermaschine verwendet. Diese erkennt 96.00 % der minderwertigen Äpfel und sortiert sie aus. Von den Äpfeln der Handelsklasse II werden jedoch auch 8.00 % als minderwertig aussortiert und von denen der Handelsklasse I immer noch 2.00 %.

- a) Wie groß ist die Wahrscheinlichkeit, dass ein Apfel als minderwertig aussortiert wird.
- b) Wie groß ist die Wahrscheinlichkeit, dass ein Apfel minderwertig ist, wenn er aussortiert wird.

Wir bezeichnen dazu mit  $A$  das Ereignis

$A$  : der Apfel wird aussortiert

und betrachten zudem die Ereignisse

$M$  : der Apfel ist minderwertig

$II$  : Der Apfel hat die Qualität der Handelsklasse II

$I$  : Der Apfel hat die Qualität der Handelsklasse I

Wir wissen

$$p(M) = 0.25, \text{ also } p(\overline{M}) = 0.75,$$

$$p(I) = 0.75 \cdot 0.40 = 0.30,$$

$$p(II) = 0.75 - 0.30 = 0.45$$

und

$$p(A|M) = 0.96, \quad p(A|II) = 0.08, \quad p(A|I) = 0.02$$

- a) Gesucht ist  $p(A)$ . Nach dem **Satz von der totalen Wahrscheinlichkeit** gilt

$$\begin{aligned} p(A) &= p(A|M) \cdot p(M) + p(A|II) \cdot p(II) + p(A|I) \cdot p(I) \\ &= 0.96 \cdot 0.25 + 0.08 \cdot 0.45 + 0.02 \cdot 0.30 \\ &= 0.2820 \end{aligned}$$

Ein Apfel wird also mit einer Wahrscheinlichkeit von 28.20 % aussortiert.

- b) Gesucht ist die Wahrscheinlichkeit  $p(M|A)$ . Dazu verwenden wir den **Satz von Bayes**:

$$\begin{aligned} p(M|A) &= \frac{p(A|M) \cdot p(M)}{p(A)} \\ &= \frac{0.96 \cdot 0.25}{0.2820} \\ &\approx 0.8511 \end{aligned}$$

Die Wahrscheinlichkeit dafür, dass ein Apfel minderwertig ist, wenn er aussortiert wird, beträgt also 85.11 %.

**Beispiel 3.35.** In dem Beispiell 3.34 wird eine neue Sortiermaschine getestet, von der bekannt ist, dass Sie 98.00 % der minderwertigen Äpfel erkennt und von denen der Handelsklasse I nur noch 1.00 % aussortiert. Welcher Prozentsatz der Äpfel der Handelsklasse II wird von dieser Maschine als minderwertig klassifiziert, wenn sie insgesamt 29.30 % der Äpfel aussortiert?

Wir bezeichnen in diesem Fall mit  $B$  das Ereignis *Der Apfel wird vom neuen Prüfgerät aussortiert*. Nach den Voraussetzungen gilt nach wie vor

$$\begin{aligned}p(M) &= 0.25, \text{ also } p(\overline{M}) = 0.75, \\p(I) &= 0.75 \cdot 0.40 = 0.30, \\p(II) &= 0.75 - 0.30 = 0.45\end{aligned}$$

und ferner ist bekannt, dass

$$p(B|M) = 0.98, \quad p(B|I) = 0.01$$

und gesucht ist  $p(B|II)$ .

Wir setzen  $x = p(B|II)$  wenden den **Satz von der totalen Wahrscheinlichkeit** an:

$$\begin{aligned}p(B) &= p(B|M) \cdot p(M) + p(B|I) \cdot p(I) + p(B|II) \cdot p(II) \\&= 0.98 \cdot 0.25 + 0.01 \cdot 0.30 + x \cdot 0.45 \\&= 0.2480 + 0.45 \cdot x\end{aligned}$$

Da wir wissen, dass  $p(B) = 0.2930$ , erhalten wir

$$0.45 \cdot x = 0.2930 - 0.2480 = 0.045$$

also

$$x = \frac{0.045}{0.45} = 0.10$$

Der Anteil der Äpfel der Handelsklasse I, den dieses Prüfgerät aussortiert, ist also 10.0 %. Die höhere Rate an minderwertigen bzw. hochwertigen Äpfeln, die die Maschine erkennt, muss also durch eine höhere Aussortiertrate bei Äpfeln der Handelsklasse II erkaufte werden.

**Beispiel 3.36.** Eine Produktionsanlage stellt Mikrochips mit eine Ausschussanteil von 8.0 % her. Zur Aussortierung der defekten Chips wird ein Prüfgerät verwendet, das 99.0 % aller defekten Chips aussortiert, allerdings auch 3.0 % der nicht defekten Chips.

a) Wie groß ist die Wahrscheinlichkeit dafür, dass ein Teil aussortiert wird?

Hierfür führen wir die folgenden Bezeichnungen für Ereignisse ein:

$A$  :     Das Teil wird aussortiert  
 $D$  :     Das Teil ist defekt

Wir wissen

$$\begin{aligned}p(D) &= 0.080, \text{ also } p(\overline{D}) = 0.92, \\p(A|D) &= 0.990, \\p(A|\overline{D}) &= 0.03\end{aligned}$$

Gesucht ist jetzt  $p(A)$ . Dafür können wir den Satz von der totalen Wahrscheinlichkeit anwenden und erhalten

$$\begin{aligned}p(A) &= p(A|D) \cdot p(D) + p(A|\overline{D}) \cdot p(\overline{D}) \\&= 0.99 \cdot 0.08 + 0.03 \cdot 0.92 \\&= 0.1068\end{aligned}$$

Ein Chip wird also mit einer Wahrscheinlichkeit von 10.68 % aussortiert.

- b) Wie groß ist die Wahrscheinlichkeit dafür, dass ein aussortiertes Teil auch tatsächlich defekt ist?

Hierfür können wir den Satz von Bayes anwenden. Mit den Bezeichnungen aus Teil a) suchen wir die Wahrscheinlichkeit  $p(D|A)$ . Diese ist (nach dem Satz von Bayes) gegeben durch

$$p(D|A) = \frac{p(A|D) \cdot p(D)}{p(A)} = \frac{0.99 \cdot 0.08}{0.1068} = 0.7416$$

Damit ist ein aussortiertes Teil also mit einer Wahrscheinlichkeit von 74.16 % auch defekt.

- c) Ein neues Prüfgerät wird getestet, dass ebenfalls 99.0 % der defekten Chips erkennt und das insgesamt 9.8 % Prozent der Chips aussortiert. Wie hoch ist der Anteil an brauchbaren Chips, der durch dieses Gerät als defekt klassifiziert wird?

Wir bezeichnen hierfür mit  $A^*$  das Ereignis, dass das neue Prüfgerät ein Teil aussortiert, und wir wissen

$$p(A^*) = 0.098, \quad p(A^*|D) = 0.990$$

Gesucht ist in diesem Aufgabenteil  $p(A^*|\overline{D})$ . Das können wir nach dem Satz von der totalen Wahrscheinlichkeit berechnen. Demnach gilt nämlich

$$p(A^*) = p(A^*|D) \cdot p(D) + p(A^*|\overline{D}) \cdot p(\overline{D})$$

Stellen wir das um, so erhalten wir

$$p(A^*) - p(A^*|D) \cdot p(D) = p(A^*|\overline{D}) \cdot p(\overline{D})$$

also

$$p(A^*|\overline{D}) = \frac{p(A^*) - p(A^*|D) \cdot p(D)}{p(\overline{D})}$$

Es folgt

$$p(A^*|\overline{D}) = \frac{0.098 - 0.990 \cdot 0.080}{0.920} = \frac{0.0188}{0.92} = 0.0204$$

Damit sortiert das neue Prüfgerät ziemlich genau 2% der brauchbaren Teile aus.

**Beispiel 3.37.** Bei einer Quizzsendung wird der Gewinner der Fragerunde in einen Raum mit 3 Türen geführt. Hinter zwei Türen liegt jeweils eine Sandale eines Beduinenstammes, hinter der dritten steht ein fabrikneuer SUV, der den Namen dieses Beduinenstammes trägt. Diese werden dabei zu jeder Show neu zufällig hinter den drei Türen verteilt. Der Gewinner der Show nimmt als Preis mit nach Hause, was hinter der Tür liegt, die er auswählt. Dabei wird allerdings, nachdem er eine Auswahl getroffen hat, noch nicht sofort die gewählte Tür geöffnet. Vielmehr öffnet der Showmaster eine der Türen, die der Kandidat **nicht** gewählt hat, und hinter denen sich eine Sandale befindet (falls der Kandidat eine Tür mit Sandale gewählt hat, gibt es genau eine weitere Tür, hinter der auch eine Sandale liegt, und der Showmaster öffnet diese, falls der Kandidat die Tür mit dem SUV gewählt hat, gibt es für den Showmaster zwei Möglichkeiten, aus denen er jeweils zufällig eine auswählt). Dann hat der Kandidat die Möglichkeit, seine Wahl zu überdenken und auf die eine noch verschlossene Tür zu wechseln. Soll der Kandidat diese Möglichkeit nutzen und die Wahl ändern oder erlangt er dadurch keinen Vorteil?

Die Antwort ist, dass der Kandidat wechseln sollte, denn dadurch erhöht sich seine Gewinnwahrscheinlichkeit, und zwar von  $\frac{1}{3}$  auf  $\frac{2}{3}$ .

Dazu bezeichnen wir mit  $G_i$  das Ereignis

$$G_i : \text{Der Gewinn (SUV) liegt hinter Tür } i$$

Nach Voraussetzung gilt

$$p(G_i) = \frac{1}{3}$$

Die Wahrscheinlichkeit, dass der SUV hinter der vom Kandidaten gewählten Tür ist, beträgt also  $\frac{1}{3}$ . Ferner bezeichnen wir die mit  $T_l$  das Ereignis

$$T_l : \text{der Showmaster öffnet die Tür } l$$

Wir wollen nun die folgende Situation betrachten

- Der Kandidat wählt die Tür 1.
- Der Showmaster öffnet die Tür 2.

(nach Umnummerierung der Türen, können wir immer annehmen, dass wir in dieser Situation sind).

Zu untersuchen ist nun, wie hoch in dieser Situation die Wahrscheinlichkeit dafür ist, dass der SUV hinter Tür 3 ist, also wie hoch  $p(G_3|T_2)$  ist.

Nach den Angaben gilt

1.  $p(T_2|G_1) = \frac{1}{2}$  (ist der SUV hinter Tür 1, so hätte der Showmaster mit gleicher Wahrscheinlichkeit die Türen 2 oder 3 wählen können).
2.  $p(T_2|G_2) = 0$ .
3.  $p(T_2|G_3) = 1$  (in diesem Fall muss der Showmaster zwangsläufig Tür 2 öffnen).

Damit erhalten wir nach dem Satz von Bayes:

$$\begin{aligned}
 p(G_3|T_2) &= \frac{p(T_2|G_3) \cdot p(G_3)}{p(T_2|G_1) \cdot p(G_1) + p(T_2|G_2) \cdot p(G_2) + p(T_2|G_3) \cdot p(G_3)} \\
 &= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 + 1 \cdot \frac{1}{3}} \\
 &= \frac{\frac{1}{3}}{\frac{1}{2}} \\
 &= \frac{2}{3}
 \end{aligned}$$

Die Chancen dafür, dass sich das Auto hinter Tür 3 befindet, liegen also bei  $\frac{2}{3}$  und sind damit doppelt so hoch wie die Chancen, dass das Auto hinter Tür 1, der ursprünglich gewählten Tür, liegen.

Beachten Sie dabei, dass sich die Wahrscheinlichkeit für Tür 1 durch die Zusatzinformation nicht erhöht, denn der Satz von Bayes ergibt auch, dass  $p(G_1|T_2) = \frac{1}{3}$  (wie vorher).

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 6

Stand: April 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim



## 4 Zufallsvariablen

Die Ereignisse, die im Rahmen der Wahrscheinlichkeitstheorie betrachtet werden, können sehr unterschiedlicher Natur sein. Insbesondere sind sie in vielen Fällen keine Zahlen sondern abstrakte Ereignisse. Damit wir damit rechnen können, ist es in der Regel erforderlich, dass wir die Ereignisse in Zahlen (oder Zahlentupel) übertragen.

### 4.1 Mess- oder Zufallsvariablen

Wir betrachten einen Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, p)$ .

**Definition 4.1.** Eine **Zufallsvariable**  $X$  ist eine Abbildung

$$X : \Omega \longrightarrow \mathbb{R}$$

so dass für jedes offene Intervall  $I \subseteq \mathbb{R}$  gilt:  $X^{-1}(I) \in \mathfrak{A}$ .

**Bemerkung 4.1.** Eine Abbildung

$$X : \Omega \longrightarrow \mathbb{R}$$

ist genau dann eine Zufallsvariable, wenn  $\{\omega \in \Omega \mid X(\omega) \leq a\} \in \mathfrak{A}$  für jedes  $a \in \mathbb{R}$ .

**Beispiel 4.1.** Wird eine Laplacemünze geworfen und interessieren wir uns für die geworfene Augenzahl, so ist  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , und die relevante Zufallsvariable

$$X : \Omega \longrightarrow \mathbb{R}$$

ist gegeben durch  $X(\omega) = \omega$ .

Interessieren wir uns dagegen nur dafür, ob die Zahl 6 geworfen wird oder nicht, würden wir eine Zufallsvariable

$$X : \Omega \longrightarrow \mathbb{R}$$

der Form

$$X(\omega) = \begin{cases} 1 & \text{falls } \omega = 6 \\ 0 & \text{sonst} \end{cases}$$

heranziehen.

**Beispiel 4.2.** Wir betrachten den viermaligen Wurf einer Münze (mit Wappen und Zahl) und interessieren uns dafür, wie oft Wappen oben liegt. Damit gilt

$$\Omega = \{\omega_W, \omega_Z\}^4 = \{(\omega_1, \omega_2, \omega_3, \omega_4) \mid \omega_i \in \{\omega_W, \omega_Z\}\}$$

wobei  $\omega_W$  für "Wappen wird geworfen" und  $\omega_Z$  für "Zahl wird geworfen" steht.  
Die Zufallsvariable

$$X : \Omega \longrightarrow \mathbb{R}$$

die unsere Untersuchung beschreibt, ist in diesem Fall gegeben durch

$$X((\omega_1, \omega_2, \omega_3, \omega_4)) = |\{i \mid \omega_i = \omega_W\}|$$

Damit werden alle für uns interessanten Ereignisse durch Zahlen beschrieben. Die Ereignisse lassen sich nun wie etwa folgt interpretieren

$$\begin{aligned} \{X = 1\} &= \text{es tritt genau einmal Wappen auf} \\ \{X \leq 1\} &= \text{es tritt höchstens einmal Wappen auf} \end{aligned}$$

also

$$\begin{aligned} \{X = 1\} &= \{(\omega_W, \omega_Z, \omega_Z, \omega_Z), (\omega_Z, \omega_W, \omega_Z, \omega_Z), (\omega_Z, \omega_Z, \omega_W, \omega_Z), \\ &\quad (\omega_Z, \omega_Z, \omega_Z, \omega_W)\} \\ \{X \leq 1\} &= \{(\omega_W, \omega_Z, \omega_Z, \omega_Z), (\omega_Z, \omega_W, \omega_Z, \omega_Z), (\omega_Z, \omega_Z, \omega_W, \omega_Z), \\ &\quad (\omega_Z, \omega_Z, \omega_Z, \omega_W), (\omega_Z, \omega_Z, \omega_Z, \omega_Z)\} \end{aligned}$$

Ist die Münze eine Laplacemünze, so erhalten wir also durch Abzählen (da  $|\Omega| = 16$ )

$$p(X \leq 1) = \frac{5}{16}$$

**Bemerkung 4.2.** Für ein  $x \in \mathbb{R}$  und eine Zufallsvariable  $X$  setzen wir

$$\{X \leq x\} = \{\omega \in \Omega \mid X(\omega) \leq x\}$$

und allgemeiner für eine beliebige Teilmenge  $B \subseteq \mathbb{R}$ :

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

Die Definition einer Zufallsvariable stellt sicher, dass für jedes  $B \subseteq \mathbb{R}$  wobei  $B$  ein (offenes, halboffenes oder abgeschlossenes) Intervall ist oder durch Bildung abzählbarer Durchschnitt bzw. abzählbarer Vereinigung und Komplemente hiervon entsteht, gilt

$$\{X \in B\} \in \mathfrak{A}$$

Speziell also

$$\{X \leq x\} \in \mathfrak{A}$$

Speziell ist also die Wahrscheinlichkeit für diese Mengen wohldefiniert, d.h.

$$p(\{X \leq x\}) \quad \text{existiert}$$

Wir schreiben kurz  $p(X \leq x)$  für  $p(\{X \leq x\})$

**Definition 4.2.** Eine Menge  $B \subseteq \mathbb{R}$  heißt *messbar*, wenn  $B$  ein (offenes, halboffenes oder abgeschlossenes) Intervall ist oder durch Bildung abzählbarer Durchschnitt bzw. abzählbarer Vereinigung und Komplemente hiervon entsteht

Die Menge aller Wahrscheinlichkeiten  $p(X \in B)$  wobei  $B$  messbar ist, nennen wir die **Wahrscheinlichkeitsverteilung** von  $X$ .

**Bemerkung 4.3.** Die Wahrscheinlichkeitsverteilung einer Zufallsvariable ist schon bestimmt durch die Wahrscheinlichkeiten

$$p(X \leq a), \quad a \in \mathbb{R}$$

Jede messbare Menge in  $\mathbb{R}$  entsteht nämlich aus den Mengen

$$\{x \in \mathbb{R} \mid x \leq a\}, a \in \mathbb{R}$$

durch (wiederholte) Bildung von Komplementen und abzählbaren Durchschnitten und Vereinigungen.

**Definition 4.3.**

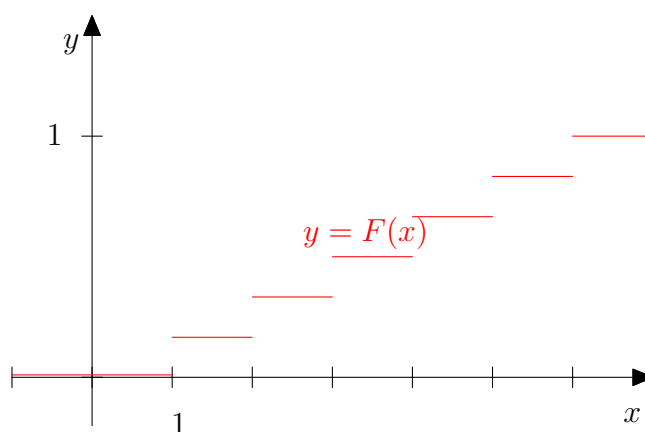
$$F : \mathbb{R} \longrightarrow \mathbb{R}$$

mit

$$F(x) = p(X \leq x)$$

heißt die **Verteilungsfunktion** der Zufallsvariable  $X$ .

**Beispiel 4.3.** Wir betrachten einen Laplacewürfel und die Zufallsvariable  $X$ , die die geworfene Augenzahl notiert. Dann hat die Verteilungsfunktion von  $X$  die folgende Gestalt



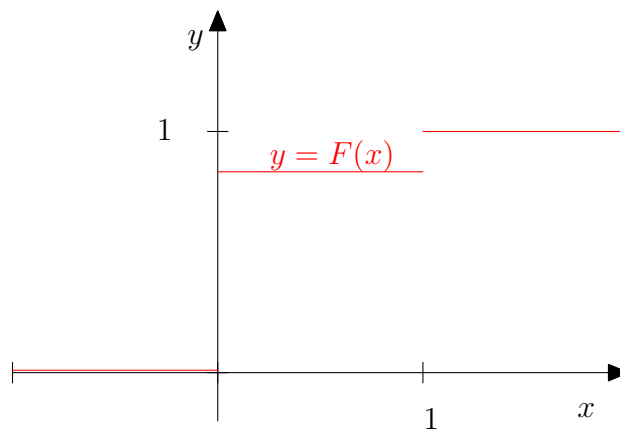
**Beispiel 4.4.** Wir betrachten einen Laplacewürfel und die Zufallsvariable  $X$ , die gegeben ist durch

$$X(\omega) = \begin{cases} 1 & \text{falls } \omega = 6 \\ 0 & \text{sonst} \end{cases}$$

Dann ist die Verteilungsfunktion von  $X$  gegeben durch

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{5}{6} & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1 \end{cases}$$

und hat die folgende Gestalt



**Bemerkung 4.4.** Für die Wahrscheinlichkeitsverteilung sind nur  $B$  relevant, die Werte enthalten, die auch tatsächlich angenommen werden. In Beispiel 4.2 etwa ist die Wahrscheinlichkeitsverteilung schon bestimmt durch

$$p(X \in B) \quad \text{wobei } B \subseteq \{0, 1, 2, 3, 4\}$$

**Definition 4.4.** Zwei Zufallsvariablen  $X$  und  $Y$  heißen **stochastisch unabhängig**, wenn für je zwei messbare Mengen  $A, B \subseteq \mathbb{R}$  gilt

$$p((X \in A) \cap (Y \in B)) = p(X \in A) \cdot p(Y \in B)$$

Zufallsvariablen  $X_1, \dots, X_n$  heißen stochastisch unabhängig, wenn für alle messbaren Mengen  $A_1, \dots, A_n$  gilt

$$p((X_1 \in A_1) \cap \dots \cap (X_n \in A_n)) = p(X_1 \in A_1) \cdot \dots \cdot p(X_n \in A_n)$$

**Bemerkung 4.5.** Die stochastische Unabhängigkeit von  $X$  und  $Y$  bedeutet, dass sich die beiden Zufallsvariablen nicht gegenseitig beeinflussen.

**Bemerkung 4.6.** Stochastische Unabhängigkeit können wir auch für unendlich viele Zufallsvariablen  $X_1, \dots, X_n, \dots$  betrachten.

**Beispiel 4.5.** Wir würfeln zweimal mit einem Laplacewürfel. Mit  $X$  bezeichnen wir das Ergebnis des ersten Wurfes, und mit  $Y$  das des zweiten. Dann sind  $X$  und  $Y$  stochastisch unabhängig.

**Beispiel 4.6.** Wir würfeln zweimal hintereinander mit einem Laplacewürfel und gehen dabei vor wie folgt: Fällt im ersten Wurf eine gerade Zahl, so wird ganz normal mit dem zweiten Laplacewürfel geworfen. Fällt dagegen im ersten Wurf eine ungerade Zahl, so wird auf dem zweiten Laplacewürfel die 1 mit einer 2, die 3 mit einer 4 und die 5 mit einer 6 überklebt und erst dann geworfen (wobei aber immer noch jede Seite gleichwahrscheinlich sein soll). Mit  $X$  bezeichnen wir das Ergebnis des ersten Wurfes, und mit  $Y$  das des zweiten. Dann sind  $X$  und  $Y$  nicht stochastisch unabhängig.

Es gilt nämlich

$$p(X = 6) = \frac{6}{36}, \quad p(Y = 6) = \frac{9}{36}$$

(wie wir schon in Beispiel 3.29 festgestellt haben), und (wie wir ebenfalls dort schon berechnet haben)

$$p(X = 6, Y = 6) = \frac{1}{36} \neq p(X = 6) \cdot p(Y = 6)$$

## 4.2 Diskrete Zufallsvariablen

**Definition 4.5.** Eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  heißt *diskrete Zufallsvariable*, falls sie nur endlich viele Werte  $x_1, \dots, x_k$  oder abzählbar unendlich viele Werte  $x_1, x_2, \dots, x_k, \dots$  annehmen kann.

**Bemerkung 4.7.** Die Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariable mit Werten  $x_1, x_2, \dots, x_k \dots$  ist schon bestimmt durch die *Elementarwahrscheinlichkeiten*

$$p_i = p(X = x_i)$$

In Beispiel 4.2 etwa ist die Wahrscheinlichkeitsverteilung schon bestimmt durch

$$\begin{aligned} p_0 = p(X = 0) &= \frac{1}{16}, & p_1 = p(X = 1) &= \frac{4}{16}, & p_2 = p(X = 2) &= \frac{6}{16}, \\ p_3 = p(X = 3) &= \frac{4}{16}, & p_4 = p(X = 4) &= \frac{1}{16} \end{aligned}$$

**Beispiel 4.7.** Wir würfeln zweimal unabhängig mit einem Laplacwürfel und bezeichnen mit  $X$  die Zufallsvariable *Summe der beiden Augenzahlen*. Dann sind die Elementarwahrscheinlichkeiten für  $X$  gegeben durch

$$\begin{array}{llll} p(X=2) & = & \frac{1}{36} & p(X=3) & = & \frac{1}{18} & p(X=4) & = & \frac{1}{12} \\ p(X=5) & = & \frac{1}{9} & p(X=6) & = & \frac{5}{36} & p(X=7) & = & \frac{1}{6} \\ p(X=8) & = & \frac{5}{36} & p(X=9) & = & \frac{1}{9} & p(X=10) & = & \frac{1}{12} \\ p(X=11) & = & \frac{1}{18} & p(X=12) & = & \frac{1}{36} \end{array}$$

**Bemerkung 4.8.** Zwei diskrete Zufallsvariablen  $X$  und  $Y$  sind schon dann stochastisch unabhängig, wenn

$$p(X=x, Y=y) = p(X=x) \cdot p(Y=y) \quad \text{für alle } x, y \in \mathbb{R}$$

**Definition 4.6.** Die **Wahrscheinlichkeitsfunktion**  $f(x)$  einer diskreten Zufallsvariable  $X$  mit Werten  $x_1, x_2, \dots, x_k, \dots$  ist definiert durch

$$f(x) = \begin{cases} p_i = p(X=x_i) & \text{für } x = x_i \in \{x_1, x_2, \dots, x_n, \dots\} \\ 0 & \text{sonst} \end{cases}$$

**Bemerkung 4.9.** Die **Verteilungsfunktion**  $F(x)$  einer diskreten Zufallsvariable  $X$  mit Werten  $x_1, x_2, \dots, x_k, \dots$  ist gegeben durch

$$F(x) = p(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

**Beispiel 4.8.** Die Verteilungsfunktion zur Zufallsvariable  $X$  aus Beispiel 4.2 ist gegeben durch

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{16} & \text{für } 0 \leq x < 1 \\ \frac{5}{16} & \text{für } 1 \leq x < 2 \\ \frac{11}{16} & \text{für } 2 \leq x < 3 \\ \frac{15}{16} & \text{für } 3 \leq x < 4 \\ 1 & \text{für } x \geq 4 \end{cases}$$

**Definition 4.7.** Ist  $X$  eine diskrete Zufallsvariable mit endlich vielen Werten  $x_1, \dots, x_k$ , so heißt  $X$  **gleichverteilt**, wenn

$$p(X=x_i) = \frac{1}{k} \quad \text{für alle } i \in \{1, \dots, k\}$$

**Beispiel 4.9.** Beim Werfen eines Laplacewürfels mit Zufallsvariable  $X$  mit

$$X(\omega) = \text{Augenzahl des Wurfes } \omega$$

ist eine gleichverteilte Zufallsvariable mit

$$p(X = i) = \frac{1}{6} \quad \text{für alle } i \in \{1, \dots, 6\}$$

**Definition 4.8.** Eine diskrete Zufallsvariable  $X$  heißt **Bernoulli-Variable**, wenn  $X$  nur die beiden Werte 0 und 1 annimmt. Die zugehörige Bernoulli-Verteilung ist schon festgelegt durch

$$p := p(X = 1)$$

denn notwendig gilt dann  $p(X = 0) = 1 - p$ .

**Beispiel 4.10.** Eine Produktionsanlage produziert Teile, die mit einer Wahrscheinlichkeit von 20 % von minderer Qualität sind. Dann ist die Zufallsvariable  $X$  mit

$$X(\omega) = \begin{cases} 0 & \text{falls } \omega \text{ von minderer Qualität} \\ 1 & \text{sonst} \end{cases}$$

eine Bernoulli-Variable mit Parameter  $p = 0.20$

**Definition 4.9.** Der **Erwartungswert** einer diskreten Zufallsvariable  $X$  mit Werten  $x_1, \dots, x_k, \dots$ , und Elementarwahrscheinlichkeiten  $p_k = p(X = x_k)$  ist definiert als

$$\mu = E(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_k \cdot p_k + \dots = \sum_{k \geq 1} x_k \cdot p_k$$

falls diese Summe existiert.

Die **Varianz** einer diskreten Zufallsvariable  $X$  mit Erwartungswert  $\mu$  und mit Werten  $x_1, \dots, x_k, \dots$ , und Elementarwahrscheinlichkeiten  $p_k = p(X = x_k)$  ist definiert als

$$\begin{aligned} \sigma^2 = \text{Var}(X) &= (x_1 - \mu)^2 \cdot p_1 + (x_2 - \mu)^2 \cdot p_2 + \dots + (x_k - \mu)^2 \cdot p_k + \dots \\ &= \sum_{k \geq 1} (x_k - \mu)^2 \cdot p_k \end{aligned}$$

falls diese Summe existiert. Die *Standardabweichung* von  $X$  ist in diesem Fall erklärt als

$$\sigma = +\sqrt{\text{Var}(X)}$$

**Regel 4.1.** Für das Rechnen mit Erwartungswerten gilt

1. Ist  $Y = aX + b$ , so gilt

$$E(Y) = aE(X) + b$$

2. Sind  $X$  und  $Y$  zwei diskrete Zufallsvariablen, so gilt

$$E(X + Y) = E(X) + E(Y)$$

3. Sind  $X$  und  $Y$  zwei stochastisch unabhängige Zufallsvariablen, so gilt

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

**Regel 4.2.** Für das Rechnen mit Varianzen gilt:

1. Ist  $\mu = E(X)$ , so ist

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

2. Ist  $Y = aX + b$ , so ist

$$\text{Var}(Y) = a^2 \cdot \text{Var}(X)$$

3. Sind  $X$  und  $Y$  zwei stochastisch unabhängige Zufallsvariablen, so gilt

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

**Beispiel 4.11.** Für eine Bernoulli-Variable  $X$  mit  $p(X = 1) = p$  gilt

$$E(X) = p$$

$$\text{Var}(X) = p(1 - p)$$

Wir betrachten ein Experiment, das durch eine Zufallsvariable  $X$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, p)$  beschrieben wird (also etwa das Experiment "Werfen einer Münze" mit  $X(\omega_W) = 0$ , und  $X(\omega_Z) = 1$ ). Wird dieses Experiment  $n$  mal stochastisch unabhängig voneinander wiederholt, so bezeichnen wir mit  $X_i$  die Zufallsvariable, die die  $i$ -te Wiederholung des Experiments beschreibt. Wir sagen hierfür kurz

$X_1, \dots, X_n$  sind unabhängige Wiederholungen von  $X$

Dabei ist auch  $n = \infty$  zugelassen (das Experiment wird also beliebig oft wiederholt).

**Definition 4.10.** Wir betrachten unabhängige Wiederholungen  $X_1, X_2, \dots, X_n, \dots$  einer Bernoulli-Variable  $X$  mit  $p(X = 1) = p (> 0)$ . Dann heißt die Zufallsvariable  $Y$  mit

$$Y(\omega) = \min\{n \in \mathbb{N} \mid X_n(\omega) = 1\}$$

geometrisch verteilte Zufallsvariable mit Parameter  $p$ .



**Bemerkung 4.10.** Für eine geometrisch verteilte Zufallsvariable  $Y$  mit Parameter  $p$  gilt

$$p(Y = n) = (1 - p)^{n-1} \cdot p$$

und damit

$$\begin{aligned} E(Y) &= \frac{1}{p} \\ \text{Var}(Y) &= \frac{1-p}{p^2} \end{aligned}$$

**Bemerkung 4.11.** Für eine geometrisch verteilte Zufallsvariable  $X$  mit Parameter  $p$  schreiben wir kurz  $X \sim G(p)$ .

**Beispiel 4.12.** Wir betrachten einen Laplacewürfel und die Zufallsvariable  $X$  Augenzahl, sowie unabhängige Wiederholungen  $X_1, X_2, \dots, X_n, \dots$  von  $X$  und die Zufallsvariable  $Y$  mit

$$Y(\omega) = \min\{n \in \mathbb{N} \mid X_n(\omega) = 6\}$$

Dann ist  $Y \sim G\left(\frac{1}{6}\right)$  mit

$$E(Y) = 6, \quad \text{Var}(Y) = \frac{\frac{5}{6}}{\frac{1}{36}} = 30$$

**Definition 4.11.** Wir betrachten unabhängige Wiederholungen  $X_1, X_2, \dots, X_n, \dots$  einer Bernoulli-Variable  $X$  mit  $p(X = 1) = p (> 0)$ . Dann heißt die Zufallsvariable  $Y$  mit

$$Y(\omega) = |\{i \in \{1, \dots, n\} \mid X_i(\omega) = 1\}|$$

**binomial verteilte Zufallsvariable** mit Parametern  $n$  und  $p$ . Hierfür gilt

$$p(Y = k) = \binom{n}{k} \cdot (1 - p)^{n-k} \cdot p^k$$

und damit

$$\begin{aligned} E(Y) &= n \cdot p \\ \text{Var}(Y) &= n \cdot p \cdot (1 - p) \end{aligned}$$

**Bemerkung 4.12.** Ist  $X$  binomialverteilt mit Parametern  $n$  und  $p$ , so schreiben wir hierfür kurz  $X \sim B(n, p)$ .

**Beispiel 4.13.** Wir betrachten das viermalige unabhängige Würfeln eines Laplacewürfels, die Zufallsvariable  $X_k$ , die die Augenzahl im  $k$ -ten Wurf registriert und die Zufallsvariable

$$Y(\omega) = \text{Anzahl der } k \in \{1, \dots, 4 \text{ mit } X_k(\omega) = 6\}$$

Dann ist  $Y$  binomialverteilt mit Parametern  $n = 4$  und  $p = \frac{1}{6}$ .

Es ist

$$p(Y = 2) = \binom{4}{2} \cdot \left(\frac{5}{6}\right)^2 \cdot \left(\frac{1}{6}\right)^2 = \frac{150}{1296}$$

und

$$p(Y \leq 2) = p(Y = 0) + p(Y = 1) + p(Y = 2) = \frac{1275}{1296}$$

**Beispiel 4.14.** Wir betrachten die Produktionsanlage mit einem Anteil von 20 % fehlerhafter Teile.

Mit welcher Wahrscheinlichkeit sind von 12 zufällig herausgegriffenen Teilen mindestens 4 fehlerhaft?

Die Zufallsvariable  $Y = \text{Anzahl der fehlerhaften Teile}$  ist binomialverteilt mit den Parametern  $p = 0.2$  und  $n = 12$ . Damit berechnet sich diese Wahrscheinlichkeit zu

$$\begin{aligned} p(Y \geq 4) &= 1 - p(Y \leq 3) \\ &= 1 - \binom{12}{3} \cdot 0.2^3 \cdot 0.8^9 - \binom{12}{2} \cdot 0.2^2 \cdot 0.8^{10} \\ &\quad - \binom{12}{1} \cdot 0.2^1 \cdot 0.8^{11} - \binom{12}{0} \cdot 0.2^0 \cdot 0.8^{12} \\ &= 1 - 0.7946 \\ &= 0.2054 \end{aligned}$$

**Definition 4.12.** Eine diskrete Zufallsvariable  $X$  mit Wertebereich  $\mathbb{N}$  heißt *Poisson-verteilt* mit Parameter  $\lambda$ , wenn gilt

$$p(X = n) = \frac{\lambda^n}{n!} \cdot e^{-\lambda} \quad \text{für alle } n \in \mathbb{N}$$

**Beispiel 4.15.** Der radioaktive Zerfall wird durch die Poisson-Verteilung beschrieben: Misst man bei einer gegebenen Menge einer radioaktiven Substanz die Anzahl der Teilchen, die in einem bestimmten Zeitintervall  $\Delta_T$  zerfällt, das sehr kurz im Vergleich zur Halbwertszeit ist, so ergibt sich eine Poisson-Verteilung. Der Parameter  $\lambda$  ist die Anzahl der im Mittel in  $\Delta_T$  zerfallenden Teilchen.

**Bemerkung 4.13.** Ist  $X$  Poisson-verteilt mit Parameter  $\lambda$ , so schreiben wir kurz  $X \sim Po(\lambda)$ .

**Bemerkung 4.14.** Ist  $X$  Poisson-verteilt mit Parameter  $\lambda$ , so gilt

$$\begin{aligned} E(X) &= \lambda \\ \text{Var}(X) &= \lambda \end{aligned}$$

**Beispiel 4.16.** Von einer radioaktiven Probe mit sehr hoher Halbwertszeit zerfallen in einer Sekunde im Mittel 4 Teilchen. Dann ist die Zufallsvariable

$$X = \text{Anzahl der Teilchen, die in einer Sekunde zerfällt}$$

Poisson-verteilt mit Parameter  $\lambda = 4$ , und es gilt

$$p(X = 0) = \frac{4^0}{0!} \cdot e^{-4} = e^{-4} = 0.0183$$

**Beispiel 4.17.** Die Anzahl  $X$  der Blitze die pro Jahr in einem Hektar Land in Deutschland einschlägt ist (in sehr guter Näherung) Poisson-verteilt mit Parameter  $\lambda = 0.10$ . Damit gilt

$$p(X = 0) = \frac{0.10^0}{0!} \cdot e^{-0.10} = 0.9048 \quad p(X = 1) = \frac{0.10^1}{1!} \cdot e^{-0.10} = 0.09048$$

und

$$p(X \geq 4) = 1 - p(X \leq 3) = 0.000003847$$

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 7

Stand: Mai 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim

### 4.3 Stetige Zufallsvariablen

Neben den diskreten Zufallsvariablen ist noch eine weitere Klasse von besonderem Interesse. Eine typische Eigenschaft stetiger Zufallsvariablen ist, dass einzelne Werte immer mit Wahrscheinlichkeit 0 auftreten und erst bei Intervallen mit positiver Wahrscheinlichkeit zu rechnen ist.

**Beispiel 4.18.** Beim zufälligen Ziehen einer reellen Zahl aus dem Intervall  $[0, 10]$  (bei dem also jede Zahl gleichwahrscheinlich ist), wird etwa mit einer Wahrscheinlichkeit von  $p = 0.30$  eine Zahl aus dem Intervall  $[2, 5]$  gezogen und mit einer Wahrscheinlichkeit von  $p = 0.05$  eine Zahl zwischen 6.00 und 6.50. Eine einzelne Zahl dagegeben wird immer mit einer Wahrscheinlichkeit von  $p = 0$  gezogen.

**Definition 4.13.** Eine Zufallsvariable  $X$  heißt **stetig**, wenn es eine integrierbare Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$ , so dass für alle  $a \leq b$  gilt

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

Die Funktion  $f(x)$  heißt **Dichte** von  $X$ .

**Bemerkung 4.15.** Genau dann ist  $X$  eine stetige Zufallsvariable mit Dichte  $f(x)$ , wenn für die Verteilungsfunktion  $F(x)$  von  $X$  gilt:

$$F(x) = \int_{-\infty}^x f(t) dt$$

**Beispiel 4.19.** Die Zufallsvariable  $X$  aus Beispiel 4.18 ist stetig mit Dichte  $f : \mathbb{R} \rightarrow \mathbb{R}$  gegeben durch

$$f(x) = \begin{cases} \frac{1}{10} & \text{für } 0 \leq x \leq 10 \\ 0 & \text{sonst} \end{cases}$$

**Bemerkung 4.16.** Ist  $X$  eine stetige Zufallsvariable, so gilt für jedes  $a \in \mathbb{R}$ :

$$p(X = a) = 0$$

Es ist nämlich

$$p(X = a) = p(X \in [a, a]) = \int_a^a f(x) dx = 0$$

**Regel 4.3.** Für die Verteilungsfunktion  $F$  und die Dichtefunktion  $f$  einer stetigen Zufallsvariable  $X$  gilt:

1.  $\int_{-\infty}^{\infty} f(x) dx = 1$ . Insbesondere existiert dieses uneigentliche Integral.
2.  $F$  ist stetig und monoton wachsend mit Werten im Intervall  $[0, 1]$
3.  $\lim_{x \rightarrow -\infty} F(x) = 0$  und  $\lim_{x \rightarrow \infty} F(x) = 1$ .
4. Ist  $x$  eine Stetigkeitsstelle von  $f(x)$ , so ist  $F$  differenzierbar in  $x$  mit  $F'(x) = f(x)$ .
5. Für  $a, b \in \mathbb{R}$  mit  $a \leq b$  gilt

$$p(a \leq X \leq b) = F(b) - F(a), \quad p(X \geq a) = 1 - F(a)$$

### Beispiel 4.20.

- a) Die Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$f(x) = \begin{cases} x & \text{für } -1 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

kann nicht die Dichtefunktion einer Zufallsvariable  $X$  sein, denn  $f(x) < 0$  für  $-1 \leq x < 0$ .

- b) Die Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$f(x) = \begin{cases} x & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

kann nicht die Dichtefunktion einer Zufallsvariable  $X$  sein. Hierfür gilt zwar  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$ , und  $f(x)$  ist auch integrierbar, allerdings würde für die Verteilungsfunktion  $F(x)$  von  $X$  gelten

$$\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} \int_{-\infty}^x f(t) dt = \int_0^1 t dt = \frac{1}{2}$$

im Widerspruch zu Regel 4.3.

- c) Die Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$f(x) = \begin{cases} \frac{1}{2} \cdot x & \text{für } 0 \leq x \leq 2 \\ 0 & \text{sonst} \end{cases}$$

könnte die Dichtefunktion einer Zufallsvariable  $X$  sein, denn hierfür gilt

1.  $f(x)$  ist integrierbar.
2.  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$ .
3.  $\int_{-\infty}^{\infty} f(x) dx = \int_0^2 \frac{x}{2} dx = \left[ \frac{x^2}{4} \right]_0^2 = 1.$

**Bemerkung 4.17.** Ist  $f : \mathbb{R} \longrightarrow \mathbb{R}$  eine Funktion mit

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

so gibt es eine stetige Zufallsvariable  $X$ , die  $f(x)$  als Dichte hat.

**Beispiel 4.21.** Die Funktion  $f : \mathbb{R} \longrightarrow \mathbb{R}$  mit

$$f(x) = \begin{cases} \frac{3}{2} \cdot x - \frac{3}{4} \cdot x^2 & \text{für } 0 \leq x \leq 2 \\ 0 & \text{sonst} \end{cases}$$

ist die Dichtefunktion einer Zufallsvariable  $X$ , denn hierfür gilt

1.  $f(x)$  ist integrierbar.
2.  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$ .
3.  $\int_{-\infty}^{\infty} f(x) dx = \int_0^2 \frac{3x}{2} - \frac{3x^2}{4} dx = \left[ \frac{3x^2}{4} - \frac{x^3}{4} \right]_0^2 = 1.$

**Beispiel 4.22.** Die Funktion  $f : \mathbb{R} \longrightarrow \mathbb{R}$  mit  $f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$  ist die Dichtefunktion einer Zufallsvariable  $X$ , denn hierfür gilt

1.  $f(x)$  ist integrierbar.
2.  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$ .
3.  $\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\pi} \cdot \left( \lim_{R \rightarrow \infty} \arctan(R) - \lim_{S \rightarrow -\infty} \arctan(S) \right) = \frac{1}{\pi} \cdot \left( \frac{\pi}{2} - \left( -\frac{\pi}{2} \right) \right) = 1.$

**Definition 4.14.** Der **Erwartungswert** einer stetigen Zufallsvariable  $X$  ist definiert als

$$E(X) = \mu = \int_{-\infty}^{\infty} x f(x) dx$$

falls dieses uneigentliche Integral existiert.

Die **Varianz** einer stetigen Zufallsvariable  $X$  mit Erwartungswert  $\mu$  ist definiert als

$$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx$$

(falls dieses uneigentliche Integral existiert), und die **Standardabweichung** einer stetigen Zufallsvariabel  $X$  ist

$$\sigma = +\sqrt{\text{Var}(X)}$$

**Regel 4.4.** Für stetige Zufallsvariablen gilt:

- Ist  $Y = aX + b$ , so ist

$$E(Y) = aE(X) + b$$

- Ist  $Z = X + Y$ , so ist

$$E(Z) = E(X) + E(Y)$$

- Ist  $\mu = E(X)$ , so gilt

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

- Ist  $Y = aX + b$ , so gilt

$$\text{Var}(Y) = a^2 \cdot \text{Var}(X)$$

- Sind  $X$  und  $Y$  unabhängige Zufallsvariablen, so gilt

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

**Beispiel 4.23.** Für die Zufallsvariable aus Beispiel 4.19, die das zufällige Ziehen einer Zahl aus dem Intervall  $[0, 10]$  beschreibt, gilt

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot f(x) dx \\ &= \int_0^{10} \frac{x}{10} dx \\ &= \left[ \frac{x^2}{20} \right]_0^{10} \\ &= 5 \end{aligned}$$

und

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{\infty} (x - 5)^2 \cdot f(x) dx \\ &= \int_0^{10} \frac{(x-5)^2}{10} dx \\ &= \left[ \frac{(x-5)^3}{30} \right]_0^{10} \\ &= \frac{25}{3} \end{aligned}$$



**Definition 4.15.** Eine stetige Zufallsvariable  $X$  heißt **gleichverteilt** auf dem Intervall  $[a, b]$ , wenn sie eine Dichte

$$f : \mathbb{R} \longrightarrow \mathbb{R}$$

mit

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

besitzt. Eine auf dem Intervall  $[0, 1]$  gleichverteilte Zufallsvariable heißt auch **standardgleichverteilt**.

**Bemerkung 4.18.** Ist  $X$  gleichverteilt auf dem Intervall  $[a, b]$ , so gilt

$$\begin{aligned} E(X) &= \frac{a+b}{2} \\ \text{Var}(X) &= \frac{(b-a)^2}{12} \end{aligned}$$

**Definition 4.16.** Eine stetige Zufallsvariable  $X$  heißt **exponentialverteilt** mit Parameter  $\lambda > 0$ , wenn sie die Dichte

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$$

besitzt.

**Bemerkung 4.19.** Ist  $X$  exponentialverteilt mit Parameter  $\lambda$ , so gilt

$$\begin{aligned} E(X) &= \lambda \cdot \int_0^{\infty} x \cdot e^{-\lambda \cdot x} dx = \frac{1}{\lambda} \\ \text{Var}(X) &= \lambda \cdot \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \cdot e^{-\lambda \cdot x} dx = \frac{1}{\lambda^2} \end{aligned}$$

**Beispiel 4.24.** Die Exponentialfunktion beschreibt häufig Lebensdauerprozesse von Objekten, bei denen die Wahrscheinlichkeit für das Ende ihrer Lebensdauer in einer gegebenen Zeitperiode fester Länge immer gleich groß ist, vorausgesetzt, sie erleben den Beginn einer Periode.

Ein typisches Beispiel hierfür ist der radioaktive Zerfall. Die Wahrscheinlichkeit für den Zerfall eines Teilchens in der nächsten Sekunde ist immer gleich groß.

Caesium-137 ist ein radioaktives Material mit einer Halbwertszeit von 11 000 Tagen, dh. die Wahrscheinlichkeit dafür, dass ein Teilchen innerhalb der nächsten 11 000 Tage zerfällt ist exakt 0.50. Der Zerfallszeitpunkt eines zufällig ausgewählten Teilchens definiert also eine stetige, exponentialverteilte Zufallsvariable  $X$  mit Dichte

$$f(t) = \begin{cases} \lambda \cdot e^{-\lambda \cdot t} & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

wobei wir  $\lambda$  wie folgt ermitteln:

Nach den Angaben gilt

$$0.50 = p(X \leq 11\,000) = F(11\,000) = \int_0^{11\,000} \lambda \cdot e^{-\lambda \cdot t} dt = [-e^{-\lambda \cdot x}]_0^{11\,000} = 1 - e^{-11\,000\lambda}$$

Daraus erhalten wir

$$e^{-11\,000\lambda} = 0.50$$

also (durch Logarithmieren)

$$-11\,000 \cdot \lambda = \ln(0.50) = -\ln(2)$$

und damit

$$\lambda = \frac{\ln(2)}{11\,000}$$

(wobei  $\lambda$  in  $\frac{1}{\text{Tag}}$  gemessen wird).

Damit ist die Wahrscheinlichkeit dafür, dass ein gegebenes Teilchen innerhalb der nächsten 100 Tage zerfällt, genau

$$p = \int_0^{100} \frac{\ln(2)}{11\,000} \cdot e^{-\frac{\ln(2)}{11\,000} \cdot t} dt = 1 - e^{-\frac{\ln(2)}{11\,000} \cdot 100} = 0.0063$$

Da die Anzahl der radioaktiven Teilchen in einer Probe selbst in einer sehr kleinen) sehr hoch ist, kann man sehr verlässlich davon ausgehen, dass von einer Caesium-137-Probe tatsächlich innerhalb der nächsten 100 Tage exakt 0.63 % zerfallen.

**Beispiel 4.25.** Die Ausfallwahrscheinlichkeit von Maschinen und Bauteilen (sofern sie immer gut gewartet und gepflegt werden) kann in guter Näherung durch eine Exponentialverteilung beschrieben werden.

Ist von einem Bauteil bekannt, dass es im Mittel nach 180 Tagen ausfällt, so wird die Wahrscheinlichkeit dafür, dass ein beliebig gewähltes Bauteil innerhalb der nächsten  $t$  Tage ausfällt, durch eine stetige, exponentialverteilte Zufallsvariable  $X$  mit Parameter  $\lambda$  beschrieben, wobei (nach Angaben)

$$\frac{1}{\lambda} = E(X) = 180$$

Also  $\lambda = \frac{1}{180}$  (in  $\frac{1}{\text{Tag}}$ ). Damit ist die Wahrscheinlichkeit dafür, dass ein zufällig ausgewähltes Teilchen in 200 Tagen noch funktioniert gegeben durch

$$p = 1 - p(X \leq 200) = 1 - \int_0^{200} \frac{1}{180} \cdot e^{-\frac{1}{180} \cdot t} dt = e^{-\frac{200}{180}} = 0.3292$$

Die Wahrscheinlichkeit dafür, dass ein Bauteil nach 180 Tagen noch funktioniert, ist dann übrigens

$$p = e^{-1} = 0.3679$$

und nicht 0.50, wie man vielleicht erwarten würde.

**Definition 4.17** (*Die Normalverteilung*). Eine stetige Zufallsvariable  $X$  heißt **normalverteilt** mit den Parametern  $\mu$  und  $\sigma^2$ , wenn sie die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

besitzt. Sie heißt **standardnormalverteilt**, wenn  $\mu = 0$  und  $\sigma^2 = 1$ .

**Bemerkung 4.20.** Ist  $X$  normalverteilt mit Parametern  $\mu$  und  $\sigma^2$ , so schreiben wir hierfür

$$X \sim N(\mu, \sigma^2)$$

**Bemerkung 4.21.** Die Dichte der Standardnormalverteilung ist gegeben durch

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

**Bemerkung 4.22.** Ist  $X \sim N(\mu, \sigma^2)$ , so gilt

$$\begin{aligned} E(X) &= \mu \\ \text{Var}(X) &= \sigma^2 \end{aligned}$$

Die Verteilungsfunktion von  $X$  ist gegeben durch

$$F(X) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^x \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt$$

Die Verteilungsfunktion einer standardnormalverteilten Zufallsvariable ist gegeben durch

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

Diese Integrale lassen sich nicht mehr in einer geschlossenen Form schreiben.

Die Werte von  $\Phi(x)$  finden sich jedoch in Tabellen (in jedem Buch zur Statistik).

**Regel 4.5.** Ist  $X \sim N(\mu, \sigma^2)$ , so ist die transformierte Zufallsvariable

$$Y = \frac{X - \mu}{\sigma}$$

standardnormalverteilt.

Ist  $X \sim N(\mu, \sigma^2)$ , so ist die transformierte Zufallsvariable

$$Y = aX + b$$

normalverteilt mit Parametern  $a\mu + b$  und  $a^2\sigma^2$ .

Sind  $X_1 \sim N(\mu_1, \sigma_1^2)$  und  $X_2 \sim N(\mu_2, \sigma_2^2)$  zwei unabhängige, normalverteilte Zufallsvariablen, so ist auch  $X_1 + X_2$  normalverteilt und es gilt

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

**Bemerkung 4.23.** Ist  $X \sim N(\mu, \sigma^2)$ , so gilt

- $p(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6827$ .
- $p(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9545$ .
- $p(\mu - 0.675 \cdot \sigma \leq X \leq \mu + 0.675 \cdot \sigma) = 0.50$ .
- $p(\mu - 1.645 \cdot \sigma \leq X \leq \mu + 1.645 \cdot \sigma) = 0.90$ .

**Bemerkung 4.24.** Ist  $X \sim N(\mu, \sigma^2)$ , so hat die Dichtefunktion  $f(x)$  von  $X$  Wendepunkte in  $\mu - \sigma$  und  $\mu + \sigma$  und das sind die beiden einzigen Wendepunkte von  $f(x)$ .

**Beispiel 4.26.** Normalverteilungen treten sehr häufig auf. So beschreiben sie etwa (in sehr guter Näherung) die Größen- und Gewichtsverteilungen in Deutschland. Sie treten sehr häufig auch bei Maschinen bei Abweichungen von Voreinstellungen auf.

**Beispiel 4.27.** Die Größe deutscher Männer ist (in guter Näherung) normalverteilt mit einem Mittelwert von 180.0 cm und einer Standardabweichung von 5.0 cm.

- a) Wie groß ist die Wahrscheinlichkeit dafür, dass ein zufällig ausgewählter deutscher Mann zwischen 1.74 m und 1.84 m groß ist?

Dazu bezeichnen wir mit  $X$  die Zufallsvariable *Körpergröße*. Dann ist nach den Angaben  $X \sim N(180, 25)$ , also ist

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 180}{5} \sim N(1, 0)$$

Damit gilt

$$\begin{aligned} 174 \leq X \leq 184 & \iff \frac{174-180}{5} \leq Z \leq \frac{184-180}{5} \\ & \iff -1.20 \leq Z \leq 0.80 \end{aligned}$$

und damit ist

$$\begin{aligned}
 p(174 \leq X \leq 184) &= p(-1.20 \leq Z \leq 0.80) \\
 &= \Phi(0.80) - \Phi(-1.20) \\
 &= \Phi(0.80) - (1 - \Phi(1.20)) \\
 &= \Phi(0.80) + \Phi(1.20) - 1 \\
 &= 0.8849 + 0.7881 - 1 \\
 &= 0.6730
 \end{aligned}$$

- b) In welchem symmetrischen Intervall um den Mittelwert 180 cm liegt die Größe von 80 % der deutschen Männer?

Gesucht ist hier ein  $a > 0$  so, dass  $p(180 - a \leq X \leq 180 + a) = 0.80$ . Wir führen das wieder auf die Standardnormalverteilung zurück indem wir wie in a) ausnutzen, dass

$$180 - a \leq X \leq 180 + a \iff \frac{180 - a - 180}{5} \leq Z \leq \frac{180 + a - 180}{5} \iff -\frac{a}{5} \leq Z \leq \frac{a}{5}$$

Damit erhalten wir die Bedingung

$$\begin{aligned}
 0.80 &= p(180 - a \leq X \leq 180 + a) \\
 &= p\left(-\frac{a}{5} \leq Z \leq \frac{a}{5}\right) \\
 &= \Phi\left(\frac{a}{5}\right) - \Phi\left(-\frac{a}{5}\right) \\
 &= \Phi\left(\frac{a}{5}\right) - (1 - \Phi\left(\frac{a}{5}\right)) \\
 &= 2 \cdot \Phi\left(\frac{a}{5}\right) - 1
 \end{aligned}$$

Daraus erhalten wir

$$\Phi\left(\frac{a}{5}\right) = 0.90$$

und aus der Tabelle der Standardnormalverteilung lesen wir ab, dass

$$\frac{a}{5} = 1.28$$

also

$$a = 6.40$$

Das gesuchte Intervall ist also  $[173.60, 186.40]$ .

**Beispiel 4.28.** Vom Durchmesser einer Frucht ist bekannt, dass er normalverteilt mit Mittelwert 14.00 cm ist, und dass 85.00 % aller Früchte einen Durchmesser von höchstens 15.50 cm haben.

Wie groß ist die Standardabweichung für den Durchmesser dieser Frucht?

Wir bezeichnen mit  $X$  die Zufallsvariable *Fruchtdurchmesser*. Dann ist  $X \sim N(14, \sigma^2)$  (mit unbekanntem  $\sigma^2$ ), und daher ist

$$Z = \frac{X - 14}{\sigma} \sim N(0, 1)$$

Damit gilt

$$X \leq a \iff \frac{X - 14}{\sigma} \leq \frac{a - 14}{\sigma} \iff Z \leq \frac{a - 14}{\sigma}$$

und wir erhalten

$$\begin{aligned} 0.85 &= p(X \leq 15.50) \\ &= p\left(Z \leq \frac{15.50 - 14.00}{\sigma}\right) \\ &= \Phi\left(\frac{1.50}{\sigma}\right) \end{aligned}$$

Aus der Tabelle der Standardnormalverteilung erhalten wir

$$\frac{1.50}{\sigma} = 1.035$$

und damit

$$\sigma = \frac{1.50}{1.035} = 1.45$$

Einige weitere Verteilungen spielen eine große Rolle bei statistischen Tests.

**Definition 4.18.** Eine Zufallsvariable  $Z$  heißt **Chi-Quadrat-verteilt** mit  $n$  Freiheitsgraden, wenn es stochastisch unabhängige und standardnormalverteilte Zufallsvariablen  $X_1, \dots, X_n$  gibt mit

$$Z = X_1^2 + \dots + X_n^2$$

**Bemerkung 4.25.** Ist  $Z$  Chi-Quadrat-verteilt mit  $n$  Freiheitsgraden, so schreiben wir

$$Z \sim \chi^2(n)$$

Es gilt

$$\begin{aligned} E(Z) &= n \\ \text{Var}(Z) &= 2n \end{aligned}$$

**Definition 4.19.** Eine Zufallsvariable  $T$  heißt  **$t$ -verteilt** mit  $n$  Freiheitsgraden, wenn es eine standardnormalverteilte Zufallsvariable  $X$  und eine  $\chi^2$ -verteilte Zufallsvariable  $Z$  mit  $n$  Freiheitsgraden gibt, die stochastisch unabhängig sind, mit

$$T = \frac{X}{\sqrt{\frac{1}{n} Z}}$$

**Bemerkung 4.26.** Ist  $T$  eine  $t$ -verteilte Zufallsvariable mit  $n$  Freiheitsgraden, so schreiben wir

$$T \sim t(n)$$

und sprechen auch von einer **Student-Verteilung** mit  $n$ -Freiheitsgraden.

**Bemerkung 4.27.** Ist  $X \sim t(n)$ , so gilt

$$E(X) = 0$$

Außerdem gilt für  $n > 2$ :

$$\text{Var}(X) = \frac{n}{n-2}$$

Für  $n = 1$  und  $n = 2$  existiert die Varianz von  $t(n)$  nicht.

**Bemerkung 4.28.** Für  $n > 30$  ist die  $t(n)$ -Verteilung sehr ähnlich zur Standardnormalverteilung und kann durch die Standardnormalverteilung ersetzt werden.

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 8

Stand: Mai 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

e-mail: [reinhold.huebl@dhbw-mannheim.de](mailto:reinhold.huebl@dhbw-mannheim.de)

©Fakultät für Technik DHBW Mannheim



## 4.4 Kovarianz und gemeinsame Verteilungen

Wir betrachten nun zwei Zufallsvariablen  $X$  und  $Y$ .

**Definition 4.20.** Sind  $X$  und  $Y$  diskret verteilt, mit Wertebereichen  $W_X$  und  $W_Y$ , so heißt

$$f(x, y) = \begin{cases} p(X = x, Y = y) & \text{falls } x \in W_X, y \in W_Y \\ 0 & \text{sonst} \end{cases}$$

die gemeinsame (diskrete) Dichtefunktion von  $X$  und  $Y$ , und

$$F(x, y) = \sum_{\xi \leq x, \eta \leq y} f(\xi, \eta)$$

heißt die gemeinsame Verteilungsfunktion von  $X$  und  $Y$ .

**Bemerkung 4.29.** Die diskreten Dichten  $f_X$  von  $X$  und  $f_Y$  von  $Y$  ergeben sich aus der gemeinsamen Dichtefunktion  $f$  als Randdichten, also

$$\begin{aligned} f_X(x) &= p(X = x) = \sum_{y \in W_Y} f(x, y) \\ f_Y(y) &= p(Y = y) = \sum_{x \in W_X} f(x, y) \end{aligned}$$

**Definition 4.21.** Sind  $X$  und  $Y$  diskret verteilt, so ist die bedingte Wahrscheinlichkeitsfunktion von  $X$  gegeben  $Y = y$  definiert als

$$f_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)} \quad \text{falls } f_Y(y) > 0$$

**Beispiel 4.29.** Wir betrachten das zweimalige unabhängige Würfeln eines Laplacewürfels und die beiden Zufallsvariablen  $X$  : *Ergebnis des ersten Wurfes* und  $Y$  : *Ergebnis des zweiten Wurfes*.

Dann ist die gemeinsame diskrete Dichtefunktion von  $X$  und  $Y$  gegeben durch

$$f(x, y) = \begin{cases} \frac{1}{36} & \text{falls } x \in \{1, \dots, 6\}, y \in \{1, \dots, 6\} \\ 0 & \text{sonst} \end{cases}$$

**Beispiel 4.30.** Wir betrachten ein Zufallsexperiment, dass aus dem Würfeln eines Laplacewürfels und einem Laplacemünzwurf besteht und wie folgt aufgebaut ist:

Der Laplacewürfel wird geworfen und die Zufallsvariable  $X$  notiert sein Ergebnis. Ist das Ergebnis gerade, so wird die Laplacemünze geworfen und  $Y$  notiert das Ergebnis dieses Wurfes (0 für Zahl und 1 für Kopf). Ist das Würfelergebnis ungerade, wird  $Y = 1$  notiert.

Dann ist die gemeinsame Dichte von  $X$  und  $Y$  gegeben durch

$$f(x, y) = \begin{cases} \frac{1}{12} & \text{für } x \in \{2, 4, 6\}, y \in \{0, 1\} \\ \frac{1}{6} & \text{für } x \in \{1, 3, 5\}, y = 1 \\ 0 & \text{sonst} \end{cases}$$

**Definition 4.22.** Zwei Zufallsvariablen  $X$  und  $Y$  heißen gemeinsam stetig verteilt, wenn es eine zweidimensionale (integrierbare) Dichtefunktion  $f(x, y)$  mit  $f(x, y) \geq 0$  für alle  $x, y \in \mathbb{R}^2$  gibt, so dass

$$p(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dy dx$$

für alle  $a \leq b$  und  $c \leq d$ .

**Bemerkung 4.30.** Sind  $X$  und  $Y$  gemeinsam stetig verteilt mit gemeinsamer Dichte  $f$ , so sind  $X$  und  $Y$  jeweils für sich betrachtet stetig verteilt mit den Randdichten

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

**Beispiel 4.31.** Das Gewicht  $X$  des Inhalts einer Pralinschachtel ist normalverteilt mit Mittelwert  $\mu_1 = 250.00$  g und Streuung  $\sigma_1 = 1.20$  g und das Gewicht  $Y$  der Packung ist normalverteilt mit Mittelwert  $\mu_2 = 40.00$  g und Streuung  $\sigma_2 = 0.40$  g, wobei Gewicht von Inhalt und Verpackung unabhängig voneinander sind. Dann besitzen  $X$  und  $Y$  eine gemeinsame stetige Dichte  $f(x, y)$ , gegeben durch

$$\begin{aligned} f(x, y) &= \frac{1}{\sqrt{2\pi} \cdot 1.20} \cdot \exp\left(-\frac{(x-250)^2}{2.88}\right) \cdot \frac{1}{\sqrt{2\pi} \cdot 0.40} \cdot \exp\left(-\frac{(y-40)^2}{0.32}\right) \\ &= \frac{1}{2\pi \cdot 0.48} \cdot \exp\left(-\frac{(x-250)^2}{2.88} - \frac{(y-40)^2}{0.32}\right) \end{aligned}$$

**Beispiel 4.32.** Das Größe  $X$  von Männern in der Altersgruppe von 18 bis 24 Jahren in Deutschland ist (in guter Näherung) normalverteilt mit Mittelwert  $\mu_1 = 181$  cm und Streuung  $\sigma_1 = 5.0$  cm und das Gewicht  $Y$  in dieser Altersgruppe ist (in guter Näherung) normalverteilt mit Mittelwert  $\mu_2 = 78.8$  kg und Streuung  $\sigma_2 = 4.5$  kg. Die Größen sind aber nicht stochastisch unabhängig sondern positiv korreliert mit Korrelationskoeffizienten  $r_{X,Y} = 0.47$ .

Die Größen  $X$  und  $Y$  besitzen dann eine gemeinsame stetige Verteilung, die (in guter Näherung) beschrieben wird durch

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi \cdot \sigma_1 \cdot \sigma_2 \cdot \sqrt{1 - r_{X,Y}^2}} \cdot \\ &= \exp\left(-\frac{1}{2 \cdot (1 - r_{X,Y}^2)} \cdot \left\{ \frac{(x - \mu_1)^2}{\sigma_1^2} - 2 \cdot r_{X,Y} \cdot \frac{(x - \mu_1)}{\sigma_1} \cdot \frac{(y - \mu_2)}{\sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right\}\right) \end{aligned}$$

**Beispiel 4.33.** Von den beiden stetigen Zufallsvariablen  $X$  und  $Y$  ist bekannt, dass sie die gemeinsame stetige Dichte

$$f(x, y) = \begin{cases} x + y & \text{für } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

besitzen. Dann gilt

$$\begin{aligned} p(X \leq 0.5, Y \leq 0.6) &= \int_0^{0.5} \int_0^{0.6} x + y \, dy \, dx \\ &= \int_0^{0.5} \left[ xy + \frac{y^2}{2} \right]_0^{0.6} dx \\ &= \int_0^{0.5} 0.6 \cdot x + 0.18 \, dx \\ &= [0.3 \cdot x^2 + 0.18 \cdot x]_0^{0.5} \\ &= 0.075 + 0.09 \\ &= 0.165 \end{aligned}$$

Es gilt

$$\int_0^1 x + y \, dy = \left[ xy + \frac{y^2}{2} \right]_0^1 = x + \frac{1}{2}$$

und daher erhalten wir als Randdichte von  $X$ :

$$f_X(x) = \begin{cases} x + 0.5 & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

und als Randdichte von  $Y$  (aus Symmetriegründen):

$$f_Y(y) = \begin{cases} y + 0.5 & \text{für } 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

**Definition 4.23.** Sind  $X$  und  $Y$  gemeinsam stetig verteilt, so ist die bedingte Wahrscheinlichkeitsfunktion von  $X$  gegeben  $Y = y$  definiert durch

$$f_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)} \quad \text{falls } f_Y(y) > 0$$

**Bemerkung 4.31.** Die gemeinsame Verteilungsfunktion  $F$  von gemeinsam stetig verteilten Variablen  $X$  und  $Y$  berechnet sich aus der gemeinsamen Dichte  $f$  nach der Formel

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(\xi, \eta) \, d\eta \, d\xi$$

**Bemerkung 4.32.** Sind zwei Zufallsvariablen unabhängig, so besitzen  $X$  und  $Y$  eine gemeinsame Dichte  $f(x, y)$ , und diese ist gegeben durch

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{für alle } x, y \in \mathbb{R}$$

Ist umgekehrt

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{für alle } x, y \in \mathbb{R}$$

eine gemeinsame Dichte von  $X$  und  $Y$ , so sind  $X$  und  $Y$  stochastisch unabhängig. Dann gilt nämlich im stetigen Fall

$$\begin{aligned} p(a \leq X \leq b, c \leq Y \leq d) &= \int_a^b \int_c^d f(x, y) \, dy \, dx \\ &= \int_a^b \int_c^d f_X(x) \cdot f_Y(y) \, dy \, dx \\ &= \int_a^b f_X(x) \, dx \cdot \int_c^d f_Y(y) \, dy \\ &= p(a \leq X \leq b) \cdot p(c \leq Y \leq d) \end{aligned}$$

und damit sind  $X$  und  $Y$  stochastisch unabhängig. Den diskreten Fall sieht man ähnlich.

**Beispiel 4.34.** Die beiden stetigen Zufallsvariablen  $X$  und  $Y$  mit gemeinsamer Dichte

$$f(x, y) = \begin{cases} x + y & \text{für } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

aus Beispiel 4.33 sind nicht stochastisch unabhängig, denn  $f_X(x) \cdot f_Y(y)$  ist keine gemeinsame Dichte von  $X$  und  $Y$ .

**Definition 4.24.** (*Kovarianz und Korrelation*)

Für zwei stetige oder diskrete Zufallsvariablen  $X$  und  $Y$  heißt

$$\text{Cov}(X, Y) = E((X - E(X)) \cdot (Y - E(Y)))$$

die **Kovarianz** von  $X$  und  $Y$ , falls dieser Ausdruck existiert, und

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

heißt die **Korrelation** von  $X$  und  $Y$ . Dabei heißen  $X$  und  $Y$  *unkorreliert*, wenn

$$\rho(X, Y) = 0$$

und andernfalls heißen sie *korreliert*.

**Bemerkung 4.33.** Sind  $X$  und  $Y$  diskret mit gemeinsamer Dichte  $f$ , so gilt

$$\text{Cov}(X, Y) = \sum_{x, y} (x - E(X)) \cdot (y - E(Y)) \cdot f(x, y)$$

Sind  $X$  und  $Y$  gemeinsam stetig verteilt mit gemeinsamer Dichte  $f$ , so gilt

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X)) \cdot (y - E(Y)) \cdot f(x, y) \, dy \, dx$$

**Regel 4.6.** Für zwei Zufallsvariablen  $X$  und  $Y$  gilt

1. *Verschiebungssatz:*  $\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$ .
2. *Symmetrie*  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ .
3. *Transformation:* Für  $X_1 = a \cdot X + b$ ,  $Y_1 = c \cdot Y + d$  gilt

$$\text{Cov}(X_1, Y_1) = a \cdot c \cdot \text{Cov}(X, Y)$$

4. *Summenregel:*  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ .

**Beispiel 4.35.** Wir betrachten das zweimalige unabhängige Werfen einer Laplacemünze und die Zufallsvariablen  $X$ : *Ergebnis des ersten Wurfes* und  $Y$ : *Ergebnis des zweiten Wurfes*, wobei wir 0 für *Zahl* und 1 für *Kopf* notieren. Dann gilt

$$E(X) = \frac{1}{2}, \quad E(Y) = \frac{1}{2}$$

und damit

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{1}{4} \cdot \left(0 - \frac{1}{2}\right) \cdot \left(0 - \frac{1}{2}\right) + \frac{1}{4} \cdot \left(0 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{2}\right) \\ &\quad + \frac{1}{4} \cdot \left(1 - \frac{1}{2}\right) \cdot \left(0 - \frac{1}{2}\right) + \frac{1}{4} \cdot \left(1 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{2}\right) \\ &= \frac{1}{16} - \frac{1}{16} - \frac{1}{16} + \frac{1}{16} \\ &= 0 \end{aligned}$$

also auch

$$\rho(X, Y) = 0$$

**Beispiel 4.36.** Wir betrachten das zweimalige unabhängige Werfen eines Laplacewürfels und die Zufallsvariablen  $X$ : *Ergebnis des ersten Wurfes* und  $Y$ : *Ergebnis des zweiten Wurfes*. Dann gilt

$$E(X) = \frac{7}{2}, \quad E(Y) = \frac{7}{2}$$

und damit

$$\begin{aligned}\text{Cov}(X, Y) &= \sum_{i,j=1}^6 \frac{1}{36} \cdot \left(i - \frac{7}{2}\right) \cdot \left(j - \frac{7}{2}\right) \\ &= \frac{1}{36} \cdot \left(\sum_{i=1}^6 \left(i - \frac{7}{2}\right)\right) \cdot \left(\sum_{j=1}^6 \left(j - \frac{7}{2}\right)\right) \\ &= 0\end{aligned}$$

also auch

$$\rho(X, Y) = 0$$

**Beispiel 4.37.** Wir betrachten das Werfen einer Laplacemünze und die Zufallsvariablen  $X$ : *Ergebnis des Wurfes*, wobei wir 0 für *Zahl* und 1 für *Kopf*. Gilt  $X = 0$ , so wird mit der Laplacemünze noch einmal geworfen und  $Y$  bezeichnet das Ergebnis des zweiten Wurfes. Gilt dagegen  $X = 1$ , so wird für  $Y$  automatisch 0 notiert. Die drei möglichen Ergebnisse des Experiments sind  $(0, 0)$ ,  $(0, 1)$  und  $(1, 0)$ , wobei  $(1, 0)$  mit Wahrscheinlichkeit  $\frac{1}{2}$  auftritt, die beiden anderen Ereignisse jeweils mit Wahrscheinlichkeit  $\frac{1}{4}$ . Dann ist

$$E(X) = \frac{1}{2}, \quad E(Y) = \frac{1}{4}$$

und damit

$$\begin{aligned}\text{Cov}(X, Y) &= \frac{1}{4} \cdot \left(0 - \frac{1}{2}\right) \cdot \left(0 - \frac{1}{4}\right) + \frac{1}{4} \cdot \left(0 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{4}\right) \\ &\quad + \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) \cdot \left(0 - \frac{1}{4}\right) \\ &= \frac{1}{32} - \frac{3}{32} - \frac{1}{16} \\ &= -\frac{1}{8}\end{aligned}$$

und damit

$$\rho(X, Y) = -\frac{\frac{1}{8}}{\frac{1}{2} \cdot \frac{\sqrt{3}}{\sqrt{8}}} = -\frac{\sqrt{2}}{2 \cdot \sqrt{3}}$$

**Bemerkung 4.34.** Sind  $X$  und  $Y$  stochastisch unabhängig, so gilt

$$\begin{aligned}\text{Cov}(X, Y) &= 0 \\ \rho(X, Y) &= 0\end{aligned}$$

Die Umkehrung dieser Aussage gilt nicht.

**Bemerkung 4.35.** Es ist  $-1 \leq \rho(X, Y) \leq 1$ .

**Regel 4.7.** *Es gilt:*

1.  $\text{Cov}(X, X) = \text{Var}(X)$ .
2. Ist  $Y = a \cdot X + b$ , so gilt

- $\rho(X, Y) = 1$ , falls  $a > 0$ .
- $\rho(X, Y) = -1$ , falls  $a < 0$ .

**Beweis:**

Die erste Aussage ist klar:

$$\begin{aligned}\text{Cov}(X, X) &= E((X - E(X)) \cdot (X - E(X))) \\ &= E((X - E(X))^2) \\ &= \text{Var}(X)\end{aligned}$$

Die zweite Aussage folgt nun leicht daraus, denn

$$\begin{aligned}\text{Cov}(X, Y) &= a \cdot \text{Cov}(X, X) = a \cdot \text{Var}(X) \\ \text{Var}(Y) &= a^2 \cdot \text{Var}(X)\end{aligned}$$

nach den Verschiebungssätzen für Varianz und Kovarianz. Setzen wir diese Ergebnisse in die Definition der Korrelation ein, so erhalten wir die Behauptung.

**Bemerkung 4.36.** Schwieriger zu zeigen ist die Umkehrung des zweiten Teils der obigen Regel: Wenn  $\rho(X, Y) = \pm 1$ , dann sind  $X$  und  $Y$  (fast überall) linear korreliert.

## 5 Grenzwertsätze

In diesem Abschnitt betrachten wir nur stetige oder diskrete Zufallsvariablen  $X$  mit existierendem Erwartungswert  $\mu$  und existierender Varianz  $\sigma^2$ , obwohl die Aussagen (in geeigneter Formulierung) auch für beliebige Zufallsvariablen gelten. Zum Testen von Hypothesen über  $X$  ist es notwendig, das zugrundeliegende Experiment mehrfach zu wiederholen, und zwar unabhängig voneinander. Wir haben dann also mehrere "Kopien"  $X_1, \dots, X_n$  dieser Zufallsvariable  $X$ .

**Definition 5.1.** Ein *stochastischer Prozess* ist eine Folge  $\mathbb{X} = (X_n)_{n \geq 1}$  von Zufallsvariablen  $X_1, X_2, \dots, X_n, \dots$ .

**Beispiel 5.1.** Die unabhängige Wiederholung eines Würfelexperiments mit

$$X_n = \text{Ergebnis des } n\text{-ten Wurfs}$$

definiert einen stochastischen Prozess  $\mathbb{X}$ .

Wir betrachten nun einen diskreten stochastischen Prozess  $\mathbb{X}$ , also einen stochastischen Prozess  $\mathbb{X} = (X_n)_{n \geq 1}$  mit diskreten Zufallsvariablen  $X_n$ .

**Definition 5.2.** Der Prozess  $\mathbb{X}$  heißt *Markovkette* (der Ordnung 1), wenn für jedes  $n \in \mathbb{N}$ , jedes  $t < n$  und alle Werte  $x_t \in W_{X_t}, \dots, x_n \in \mathbb{W}_{X_n}$  gilt

$$p(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_t = x_t) = p(X_n = x_n | X_{n-1} = x_{n-1})$$

**Bemerkung 5.1.** Ein Prozess  $\mathbb{X}$  ist eine Markovkette, wenn die Wahrscheinlichkeit für ein bestimmtes Ereignis nur von der unmittelbaren Vergangenheit, nicht aber von länger zurückliegenden Ereignissen abhängt.

**Definition 5.3.** Die Wahrscheinlichkeit  $p(X_{n+1} = x_{n+1} | X_n = x_n)$  heißt **Übergangswahrscheinlichkeit** von  $x_n$  zu  $x_{n+1}$  zum Zeitpunkt  $n$ .

Eine Markovkette heißt **homogen**, wenn alle  $X_n$  die gleichen Zustände  $x_1, x_2, \dots$  annehmen können, und wenn die Übergangswahrscheinlichkeiten nicht vom Zeitpunkt abhängen, dh. wenn

$$p(X_{n+1} = x_j | X_n = x_i) = p(X_1 = x_j | X_0 = x_i)$$

Ist  $\mathbb{X}$  eine homogene Markovkette, so setze  $p_{i,j} = p(X_1 = x_j | X_0 = x_i)$ .

**Definition 5.4.** Die Matrix  $P = (p_{i,j})$  heißt Matrix der Übergangswahrscheinlichkeiten von  $\mathbb{X}$ .

**Beispiel 5.2.** Wir betrachten ein Teilchen, dass auf drei verschiedenen Energie-Niveaus I, II und III liegen kann. Dabei geht ein Teilchen auf Niveau I in einer Zeiteinheit mit einer Wahrscheinlichkeit von 0.4 in Energieniveau II über, mit einer Wahrscheinlichkeit von 0.6 bleibt es auf Niveau I. Ein Teilchen auf Niveau II geht in einer Zeiteinheit mit einer Wahrscheinlichkeit von 0.3 in Energieniveau III über, mit einer Wahrscheinlichkeit von 0.2 in Energieniveau I und mit einer Wahrscheinlichkeit von 0.5 bleibt es auf Niveau II. Ein Teilchen auf Niveau III geht in einer Zeiteinheit mit einer Wahrscheinlichkeit von 0.3 in Energieniveau II über und mit einer Wahrscheinlichkeit von 0.7 bleibt es auf Niveau III. Dann ist  $\mathbb{X} = (X_n)_{n \in \mathbb{N}}$  mit

$$X_n(\omega) = \text{Niveau des Teilchens } \omega \text{ nach } n \text{ Zeiteinheiten}$$

eine homogene Markovkette mit Matrix

$$P = \begin{pmatrix} 0.6 & 0.2 & 0.0 \\ 0.4 & 0.5 & 0.3 \\ 0.0 & 0.3 & 0.7 \end{pmatrix}$$

der Übergangswahrscheinlichkeiten.



Eine homogene Markovkette heißt **zusammenhängend** wenn es möglich ist, von einem beliebigen Zustand  $x_i$  in endlich vielen Schritten in jeden beliebigen anderen Zustand  $x_j$  zu kommen.

**Bemerkung 5.2.** Jede zusammenhängende homogene Markovkette mit  $T$  vielen Zuständen  $x_1, \dots, x_T$  hat einen Gleichgewichtszustand, dh. es gibt einen Vektor

$$\vec{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_T \end{pmatrix}$$

mit  $z_i \geq 0$ ,  $z_1 + \dots + z_T = 1$  und

$$P \cdot \vec{z} = \vec{z}$$

Dabei ist  $\vec{z}$  eine Eigenvektor von  $P$  zu Eigenwert 1.

**Beispiel 5.3.** In Beispiel 5.2 ist

$$\vec{z} = \begin{pmatrix} 0.2 \\ 0.4 \\ 0.4 \end{pmatrix}$$

ein Gleichgewichtszustand, dh. die Teilchen werden sich langfristig im Verhältnis 0.2 : 0.4 : 0.4 auf die drei Energiezustände aufteilen.

**Beispiel 5.4.** Ein Fahrradverleih hat zwei Verleihstationen an einem Ort, eine am Hauptbahnhof und eine in der Innenstadt. Ein Fahrrad, dass am Hauptbahnhof ausgeliehen wird, wird mit einer Wahrscheinlichkeit von 0.6 am Ende des Tages wieder dort abgegeben, mit einer von 0.4 in der Innenstadt, ein Fahrrad, dass in der Innenstadt ausgeliehen wird, wird mit einer Wahrscheinlichkeit von 0.7 am Ende des Tages wieder dort abgegeben, mit einer von 0.3 am Hauptbahnhof. Dann ist  $\mathbb{X} = (X_n)_{n \in \mathbb{N}}$  mit

$$X_n(\omega) = \text{Standort des Fahrrades } \omega \text{ nach } n \text{ Tagen}$$

eine homogene Markovkette mit Matrix

$$P = \begin{pmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{pmatrix}$$

der Übergangswahrscheinlichkeiten.

Ein stationärer Zustand ist der Vektor

$$\vec{z} = \begin{pmatrix} \frac{3}{7} \\ \frac{4}{7} \end{pmatrix}$$

**Definition 5.5.** Der Prozess  $\mathbb{X}$  heißt *Prozess mit unabhängigen Zuwächsen* oder *random walk*, wenn für jedes  $n \in \mathbb{N}$  und jedes Tupel  $t_1 < t_2 < \dots < t_n$  in  $\mathbb{N}$  die Zufallsvariablen  $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$  stochastisch unabhängig sind.

**Bemerkung 5.3.** Ein random walk ist auch eine Markovkette. Die Umkehrung stimmt nicht.

Für statistische Analysen sind insbesondere Prozesse wichtige, die die unabhängige Wiederholung eines Experiments beschreiben.

**Definition 5.6.** Ein stochastischer Prozess  $\mathbb{X} = (X_n)_{n \geq 1}$  heißt *i.i.d* (*independent and identically distributed*), wenn  $X_1, \dots, X_n, \dots$ , unabhängig sind und alle die gleiche Verteilungsfunktion haben.

**Bemerkung 5.4.** Die unabhängige Wiederholung eines Zufallsexperiments  $X$  definiert einen i.i.d. Prozess.

Wir betrachten nun einen i.i.d-Prozess  $\mathbb{X}$  und definieren für  $n \in \mathbb{N}$  neue Zufallsvariablen, das arithmetische Mittel

$$\tilde{X} = \tilde{X}_n := \frac{1}{n} \cdot (X_1 + \dots + X_n)$$

**Bemerkung 5.5.** Es gilt

$$\begin{aligned} E(\tilde{X}_n) &= \mu \\ \text{Var}(\tilde{X}_n) &= \frac{\sigma^2}{n} \end{aligned}$$

**Satz 5.1.** (*Gesetz der großen Zahl*)

Für jedes  $c > 0$  gilt

$$\lim_{n \rightarrow \infty} p(|\tilde{X}_n - \mu| \leq c) = 1$$

d.h.  $\tilde{X}_n$  konvergiert nach Wahrscheinlichkeit gegen  $\mu = E(X)$ .

Das Gesetz der großen Zahl ist so zu interpretieren:

Wird ein Zufallsexperiment, dass durch die Zufallsvariable  $X$  beschrieben wird, beliebig oft unabhängig wiederholt, und ist  $x_n$  der beobachtete Wert der Zufallsvariable  $X$  im  $n$ -ten Versuch, so konvergiert das arithmetische Mittel

$$\tilde{x} = \widetilde{x}_n = \frac{1}{n} (x_1 + \dots + x_n)$$

mit Wahrscheinlichkeit 1 gegen den Erwartungswert  $\mu$  von  $X$ .

Das Gesetz der großen Zahl gibt aber keinen Anhaltspunkt über die Konvergenzgeschwindigkeit. Als Faustregel hat es sich aber eingebürgert, dass mindestens 30 Versuche durchgeführt werden müssen.

Als Folgerung erhalten wir eine Aussage über die Approximation von Verteilungsfunktionen. Dazu setzen wir

$$\tilde{F}_n(x) = \frac{1}{n} \cdot |\{i \in \{1, \dots, n\} | x_i \leq x\}|$$

(empirische Verteilungsfunktion bei  $n$  Versuchen). Ist dann  $F(x)$  die gemeinsame Verteilungsfunktion der  $X_n$ , so gilt

**Folgerung 5.2.** (*Satz von Glivenko–Cantelli*)

Für jedes  $x > 0$  gilt:

$$\lim_{n \rightarrow \infty} p(\sup |\tilde{F}_n(x) - F(x)| \leq c) = 1$$

d.h. die empirische Verteilungsfunktion konvergiert mit Wahrscheinlichkeit 1 gegen  $F(x)$ .

Die Standardnormalverteilung spielt eine wesentliche Rolle in der Statistik, was auch durch folgendes Resultat zum Ausdruck gebracht wird:

**Satz 5.3.** (*zentraler Grenzwertsatz*)

Ist  $\mathbb{X} = (X_n)_{n \geq 1}$  ein i.i.d.-Prozess mit  $E(X_n) = \mu$  und  $\text{Var}(X_n) = \sigma^2$ , und setzen wir

$$Z_n = \frac{1}{\sqrt{n}\sigma} \cdot (X_1 + \dots + X_n - n\mu) = \frac{1}{\sqrt{n}} \cdot \sum_{k=1}^n \frac{X_k - \mu}{\sigma}$$

so gilt für die Verteilungsfunktion  $F_n(x) = p(Z_n \leq x)$  von  $Z_n$ :

$$\lim_{n \rightarrow \infty} F_n(x) = \Phi(x) \quad \text{für alle } x \in \mathbb{R}$$

wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung ist.

**Bemerkung 5.6.** In der Situaition des zentralen Grenzwertsatzes schreiben wir auch

$$Z_n \stackrel{a}{\sim} N(0, 1)$$

und sagen, dass  $(Z_n)_{n \in \mathbb{N}}$  approximativ standardnormalverteilt ist.

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 9

Stand: Mai 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim

## 6 Parameterschätzung

Statistische Untersuchungen, bei denen Stichproben gezogen und betrachtet werden, dienen dem Zweck, ein möglichst gutes Bild der Gesamtsituation zu gewinnen und Daten zu ermitteln, die für die Parameter der Grundgesamtheit aussagekräftig sind. Schätzverfahren dienen also dazu, aus einer Zufallsstichprobe auf ein Merkmal der Grundgesamtheit zurückzuschliessen, dass durch eine Zufallsvariable beschrieben wird. Dabei konzentrieren wir uns auf bestimmte Aspekte (Parameter) dieses Merkmals, speziell

- Erwartungswert und Varianz einer Zufallsvariable.
- Median oder  $p$ -Quantile einer Zufallsvariable.
- Korrelation zweier Zufallsvariablen.

Ausgangspunkt sind  $n$  Stichprobenziehungen, die durch  $n$  unabhängige und identisch verteilte Zufallsvariablen  $X_1, \dots, X_n$  repräsentiert werden und aus der wir auf einen Parameter  $\theta$  der Zufallsvariable  $X$  schließen wollen. Die Realisationen dieser Zufallsvariablen, also die gemessenen Werte, bezeichnen wir mit  $x_1, \dots, x_n$ .

**Definition 6.1.** Eine **Schätzfunktion** oder **Schätzstatistik** für den Parameter  $\theta$  der Grundgesamtheit ist eine Funktion

$$T = g(X_1, \dots, X_n)$$

der Stichprobenvariablen  $X_1, \dots, X_n$ . Der aus den Realisationen  $x_1, \dots, x_n$  gewonnene Wert

$$t = g(x_1, \dots, x_n)$$

heißt der dazugehörige **Schätzwert**.

**Beispiel 6.1.** Oft betrachtete Schätzfunktionen sind

- $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$  für den Erwartungswert  $\mu = E(X)$ .
- $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$  für die Varianz  $\sigma^2 = \text{Var}(X)$ .
- $\tilde{S}^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2$  für die Varianz  $\sigma^2 = \text{Var}(X)$ .

**Definition 6.2.** Eine Schätzstatistik  $T$  heißt **erwartungstreu** oder unverzerrt für den Parameter  $\theta$ , wenn

$$E(T) = \theta$$

Der **Bias** der Schätzstatistik  $T$  ist

$$\text{Bias}(T) = E(T) - \theta$$

Der **Standardfehler** der Schätzstatistik  $T$  ist die Standardabweichung der Schätzstatistik

$$\sigma_g = \sqrt{\text{Var}(g(X_1, \dots, X_n))}$$

**Bemerkung 6.1.** Ein Schätzer ist also erwartungstreu, wenn der Erwartungswert der Zufallsvariable  $T$  der zu schätzende Parameter  $\theta$  ist.

**Beispiel 6.2.** Der Schätzer

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$$

für den Erwartungswert  $\mu = E(X)$  ist erwartungstreu, denn

$$E(\bar{X}) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{1}{n} \sum_{k=1}^n \mu = \mu$$

**Beispiel 6.3.** Der Schätzer

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

für die Varianz  $\sigma^2 = \text{Var}(X)$  ist erwartungstreu, der Schätzer

$$\tilde{S}^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2$$

für die Varianz  $\sigma^2 = \text{Var}(X)$  ist dagegen nicht erwartungstreu.

Dazu berechnen wir zunächst

$$E \left( \sum_{k=1}^n (X_k - \bar{X})^2 \right)$$

Es gilt

$$\begin{aligned}
E\left(\sum_{k=1}^n (X_k - \bar{X})^2\right) &= \sum_{k=1}^n E(X_k - \bar{X})^2 \\
&= \sum_{k=1}^n E((X_k - \mu) + (\mu - \bar{X}))^2 \\
&= \sum_{k=1}^n E((X_k - \mu)^2 + 2 \cdot (X_k - \mu) \cdot (\mu - \bar{X}) + (\mu - \bar{X})^2) \\
&= \sum_{k=1}^n (E((X_k - \mu)^2) + 2 \cdot E((X_k - \mu) \cdot (\mu - \bar{X})) + E((\mu - \bar{X})^2)) \\
&= \sum_{k=1}^n (E((X - \mu)^2) + 2 \cdot E((X_k - \mu) \cdot (\mu - \bar{X})) + E((\bar{X})^2 - \mu)) \\
&= \sum_{k=1}^n (\text{Var}(X) - 2 \cdot E((X_k - \mu) \cdot (\bar{X} - \mu)) + \frac{1}{n} \text{Var}(X)) \\
&= (n+1) \cdot \text{Var}(X) - 2 \sum_{k=1}^n E((X_k - \mu) \cdot (\bar{X} - \mu)) \\
&= (n+1) \cdot \text{Var}(X) - 2E\left(\sum_{k=1}^n (X_k - \mu) \cdot (\bar{X} - \mu)\right) \\
&= (n+1) \cdot \text{Var}(X) - 2E\left((\bar{X} - \mu) \cdot \sum_{k=1}^n (X_k - \mu)\right) \\
&= (n+1) \cdot \text{Var}(X) - 2E((\bar{X} - \mu) \cdot n \cdot (\bar{X} - \mu)) \\
&= (n+1) \cdot \text{Var}(X) - 2n \cdot E(\bar{X} - \mu)^2 \\
&= (n+1) \cdot \text{Var}(X) - 2n \cdot \text{Var}(\bar{X}) \\
&= (n+1) \cdot \text{Var}(X) - 2 \cdot \text{Var}(X) \\
&= (n-1) \cdot \text{Var}(X)
\end{aligned}$$

Beachten Sie dabei, dass wir ausgenutzt haben, dass

$$\sum_{k=1}^n (X_k - \mu) = \sum_{k=1}^n X_k - \sum_{k=1}^n \mu = n \cdot \frac{1}{n} \cdot \sum_{k=1}^n X_k - n \cdot \mu = n \cdot \bar{X} - n \cdot \mu = n \cdot (\bar{X} - \mu)$$

Damit gilt also

$$E(S^2) = \frac{1}{n-1} \cdot E\left(\sum_{k=1}^n (X_k - \bar{X})^2\right) = \sigma$$

und

$$E(\tilde{S}^2) = \frac{1}{n} \cdot E\left(\sum_{k=1}^n (X_k - \bar{X})^2\right) = \frac{n-1}{n} \cdot \sigma$$

also ist  $S^2$  erwartungstreu und  $\tilde{S}^2$  nicht.

Wir lassen nun  $n$  variieren und betrachten Schätzstatistiken

$$T_n = g_n(X_1, \dots, X_n)$$

**Beispiel 6.4.** Die in Beispiel 6.1 betrachteten Schätzer lassen sich alle für variables  $n$  erklären.

**Definition 6.3.** Eine Schätzstatistik  $(T_n)_{n \in \mathbb{N}}$  heißt **asymptotisch erwartungstreu**, wenn

$$\lim_{n \rightarrow \infty} E(T_n) = \theta$$

**Beispiel 6.5.** Alle drei Schätzstatistiken aus Beispiel 6.1 sind asymptotisch erwartungstreu.

Die Erwartungstreue ist eine wichtige Eigenschaft von Schätzern. Darüberhinaus ist aber auch noch von Interesse, wie stark die Schätzwerte um den zu schätzenden Parameter streuen können. Ideal wäre dabei natürlich ein Schätzer für den Parameter  $\theta$ , der für jede Realisation  $x_1, \dots, x_n$  den Wert

$$g(x_1, \dots, x_n) = \theta$$

liefert. Ein Maß für die Güte bezüglich der Abweichungen ist die mittlere quadratische Abweichung MSE (für **mean squared error**)

**Definition 6.4.** Der Wert

$$\text{MSE}(T) = E((T - \theta)^2)$$

heißt **erwartete mittlere Abweichung** der Schätzstatistik  $T$ .

**Bemerkung 6.2.** Es gilt

$$\text{MSE}(T) = \text{Var}(T) + \text{Bias}(T)^2$$

**Definition 6.5.** Eine Schätzstatistik  $(T_n)_{n \in \mathbb{N}}$  heißt **konsistent im quadratische Mittel**, wenn

$$\lim_{n \rightarrow \infty} \text{MSE}(T_n) = 0$$

Eine Schätzstatistik  $(T_n)_{n \in \mathbb{N}}$  heißt **schwach konsistent**, wenn für jedes  $\varepsilon > 0$  gilt:

$$\lim_{n \rightarrow \infty} p(|T_n - \theta| < \varepsilon) = 1$$

**Bemerkung 6.3.** Eine Schätzstatistik  $(T_n)_{n \in \mathbb{N}}$  ist genau dann schwach konsistent, wenn für jedes  $\varepsilon > 0$  gilt

$$\lim_{n \rightarrow \infty} p(|T_n - \theta| > \varepsilon) = 0$$

**Bemerkung 6.4.** Ist eine Schätzstatistik konsistent im quadratischen Mittel, so ist sie auch schwach konsistent.



**Beispiel 6.6.** Das arithmetische Mittel

$$\bar{X} = \frac{1}{n} \cdot \sum_{k=1}^n X_k$$

ist ein im quadratischen Mittel konsistenter Schätzer für den Erwartungswert.

Wir betrachten nun zwei Schätzstatistiken  $T_1$  und  $T_2$  für den Parameter  $\theta$ .

**Definition 6.6.** Schätzstatistik  $T_1$  heißt **MSE–wirksamer** als  $T_2$ , wenn

$$\text{MSE}(T_1) \leq \text{MSE}(T_2)$$

Sind  $T_1$  und  $T_2$  erwartungstreu, so heißt  $T_1$  **wirksamer** als  $T_2$ , wenn

$$\text{Var}(T_1) \leq \text{Var}(T_2)$$

Eine Schätzstatistik  $T$  heißt **MSE–wirksamst**, wenn sie **MSE–wirksamer** ist als alle andere zugelassenen Schätzer, und eine erwartungstreue Schätzstatistik heißt **wirksamst** oder **effizient**, wenn sie **wirksamer** ist als alle andern zugelassenen erwartungstreuen Schätzer.

**Beispiel 6.7.** Bei  $n = 2$  sind die Schätzstatistiken

$$T_1 = \frac{1}{10} \cdot X_1 + \frac{9}{10} \cdot X_2$$

und

$$T_2 = \frac{1}{4} \cdot X_1 + \frac{3}{4} \cdot X_2$$

erwartungstreu für  $\mu = E(X)$ , aber  $T_2$  ist wirksamer als  $T_1$ .

Es ist

$$E(T_1) = \frac{1}{10} \cdot E(X_1) + \frac{9}{10} \cdot E(X_2) = \frac{1}{10} \cdot E(X) + \frac{9}{10} \cdot E(X) = E(X)$$

und

$$E(T_2) = \frac{1}{4} \cdot E(X_1) + \frac{3}{4} \cdot E(X_2) = \frac{1}{4} \cdot E(X) + \frac{3}{4} \cdot E(X) = E(X)$$

also sind beide erwartungstreu.

Ferner gilt (wegen der stochastischen Unabhängigkeit)

$$\begin{aligned} \text{Var}(T_1) &= \frac{1}{100} \cdot \text{Var}(X_1) + \frac{81}{100} \cdot \text{Var}(X_2) \\ &= \frac{1}{100} \cdot \text{Var}(X) + \frac{81}{100} \cdot \text{Var}(X) \\ &= \frac{82}{100} \cdot \text{Var}(X) \end{aligned}$$

und

$$\begin{aligned}\text{Var}(T_2) &= \frac{1}{16} \cdot \text{Var}(X_1) + \frac{9}{16} \cdot \text{Var}(X_2) \\ &= \frac{1}{16} \cdot \text{Var}(X) + \frac{9}{16} \cdot \text{Var}(X) \\ &= \frac{5}{8} \cdot \text{Var}(X)\end{aligned}$$

also ist  $T_2$  effizienter als  $T_1$ .

**Beispiel 6.8.** Wir betrachten Zufallsvariable  $X$  mit  $\mu$  und Varianz  $\sigma^2$ . Die 5 Stichprobenvariablen  $X_1, \dots, X_5$  seien unabhängige Wiederholungen von  $X$ . Für  $\mu$  betrachten wir die Schätzfunktionen

$$\begin{aligned}T_1 &= \bar{X} = \frac{1}{5} (X_1 + X_2 + X_3 + X_4 + X_5) \\ T_2 &= \frac{1}{2} (X_1 + X_2) \\ T_3 &= \frac{1}{6} (X_1 + X_2 + X_3) + \frac{1}{4} (X_4 + X_5) \\ T_4 &= \frac{1}{4} (X_1 + X_2 + X_3) + \frac{1}{6} (X_4 + X_5) \\ T_5 &= X_1 + X_2 + X_4\end{aligned}$$

a) Welche Schätzer sind erwartungstreu?

Die unabhängigen Ziehungen  $X_1, \dots, X_5$  sind unabhängige Wiederholungen von  $X$ , insbesondere gilt also  $X_k \sim N(\mu, \sigma^2)$  ( $k = 1, \dots, 5$ ), und damit vor allem

$$E(X_k) = \mu, \quad \text{Var}(X_k) = \sigma^2 \quad (k = 1, \dots, 5)$$

Es gilt nach den Regeln zum Rechnen mit Erwartungswerten

$$\begin{aligned}E(T_1) &= \frac{1}{5} \cdot (E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5)) &= \mu \\ E(T_2) &= \frac{1}{2} \cdot (E(X_1) + E(X_2)) &= \mu \\ E(T_3) &= \frac{1}{6} \cdot (E(X_1) + E(X_2) + E(X_3)) + \frac{1}{4} \cdot (E(X_4) + E(X_5)) &= \mu \\ E(T_4) &= \frac{1}{4} \cdot (E(X_1) + E(X_2) + E(X_3)) + \frac{1}{6} \cdot (E(X_4) + E(X_5)) &= \frac{13}{12} \cdot \mu \\ E(T_5) &= E(X_1) + E(X_2) + E(X_4) &= 3 \cdot \mu\end{aligned}$$

Damit sind als  $T_1$ ,  $T_2$ , und  $T_3$  erwartungstreue Schätzer,  $T_4$  und  $T_5$  nicht.

b) Welcher der erwartungstreuen Schätzer ist der wirksamste unter diesen?

Es ist jeweils die Varianz der drei erwartungstreuen Schätzer  $T_1$ ,  $T_2$  und  $T_3$  zu bestimmen (da für erwartungstreue Schätzer  $\text{MSE}(T) = \text{Var}(T)$ ).

Nach den Regeln zum Rechnen mit Varianzen (für unabhängige Zufallsvariablen) gilt

$$\begin{aligned}
\text{Var}(T_1) &= \frac{1}{5^2} \cdot (\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) + \text{Var}(X_4) + \text{Var}(X_5)) \\
&= \frac{1}{5} \cdot \sigma^2 \\
\text{Var}(T_2) &= \frac{1}{2^2} \cdot (\text{Var}(X_1) + \text{Var}(X_2)) \\
&= \frac{1}{2} \cdot \sigma^2 \\
\text{Var}(T_3) &= \frac{1}{6^2} \cdot (\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3)) + \frac{1}{4^2} \cdot (\text{Var}(X_4) + \text{Var}(X_5)) \\
&= \frac{5}{24} \cdot \sigma^2
\end{aligned}$$

Damit ist  $T_1$  der wirksamste Schätzer.

Wir betrachten nun wieder unabhängige und identisch verteilte Wiederholungen eines (diskreten oder stetigen) Zufallsexperiments  $X$ , also einen i.i.d.-Prozess  $\mathbb{X} = (x_n)_{n \geq 1}$  mit  $X_n \sim X$ , für den wir den Parameter  $\theta$  schätzen wollen. Ist  $\theta_0$  der tatsächliche Wert des Parameters, so bezeichnen wir mit  $f(x|\theta_0)$  die (diskrete oder stetige) Dichte in diesem Fall. Dann gilt für die Dichtefunktion der  $n$ -fachen Wiederholung

$$f(x_1, \dots, x_n | \theta_0) = \prod_{i=1}^n f(x_i | \theta_0) = f(x_1 | \theta_0) \cdots f(x_n | \theta_0)$$

(da  $X_1, \dots, X_n$  ja stochastisch unabhängig sind und sich die gemeinsame Dichte daher als Produkt der einzelnen Dichten ergibt), wobei wir hier  $\theta_0$  als festen Parameter und die  $x_1, \dots, x_n$  als variable Realisationen betrachten.

Aus Sicht der Schätzung wollen wir aber eine andere Perspektive wählen: Wir gehen davon aus, dass wir eine feste Realisation  $x_1, \dots, x_n$  der Stichprobe haben und den Parameter  $\theta$  suchen, wir betrachten also die Funktion

$$L(\theta) = f(x_1, \dots, x_n | \theta)$$

**Definition 6.7.**  $L(\theta)$  heißt **Likelihoodfunktion**.

**Definition 6.8.** Eine Schätzfunktion  $\hat{\theta}$  für den Parameter  $\theta$  ist nach dem Maximum-Likelihood-Prinzip konstruiert oder kurz **Maximum-Likelihood-Schätzer**, wenn

$$L(\hat{\theta}) \geq L(\theta) \quad \text{für alle zugelassenen } \theta$$

**Bemerkung 6.5.** Oft wird die **log-Likelihoodfunktion**  $l(\theta) = \ln(L(\theta))$  untersucht, da diese häufig einfacher zu handhaben ist. Da  $\ln$  streng monoton wachsend ist, stimmen Maxima von  $L(\theta)$  und  $l(\theta)$  überein.

**Bemerkung 6.6.** Zur Konstruktion eines Maximum-Likelihood-Schätzers für den Parameter  $\theta$  einer Zufallsvariable  $X$  mit (diskreter oder stetiger) Dichte  $f(x|\theta)$  gehen wir bei  $n$  stochastisch unabhängigen Wiederholungen  $X_1, \dots, X_n$  von  $X$  vor wie folgt:

1. Bestimme die Likelihood-Funktion  $L(\theta) = f(x_1|\theta) \cdot f(x_n|\theta)$ .
2. Bestimme die Log-Likelihood-Funktion  $l(\theta) = \ln(L(\theta))$ .
3. Bestimme die Ableitung  $l'(\theta)$  (bzw.  $L'(\theta)$ , je nachdem, was einfacher wird).
4. Bestimme  $\hat{\theta}$  mit  $l'(\hat{\theta}) = 0$  (bzw.  $L'(\hat{\theta}) = 0$ ).
5. Überprüfe, ob  $\hat{\theta}$  ein Maximum von  $l(\theta)$  (bzw.  $L(\theta)$  ist) (z.B. über die zweite Ableitung).

**Beispiel 6.9.** Wir betrachten die  $n$ -fache Wiederholung einer Poisson-verteilten Zufallsvariable  $X$  mit Parameter  $\lambda$  (den wir schätzen wollen), also  $X \sim Po(\lambda)$  und  $n$  Realisationen  $x_1, \dots, x_n$ . Dann gilt

$$\begin{aligned} L(\lambda) &= e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} \cdots e^{-\lambda} \frac{\lambda^{x_n}}{x_n!} \\ &= e^{-n \cdot \lambda} \frac{\lambda^{x_1 + x_2 + \cdots + x_n}}{x_1! \cdot x_2! \cdots x_n!} \end{aligned}$$

Logarithmieren der Likelihoodfunktion liefert

$$l(\lambda) = \ln(L(\lambda)) = -n \cdot \lambda + (x_1 + \cdots + x_n) \cdot \ln(\lambda) - \ln(x_1! \cdot x_2! \cdots x_n!)$$

und ableiten ergibt

$$l'(\lambda) = -n + \left( \sum_{k=1}^n x_k \right) \cdot \frac{1}{\lambda}$$

Einen Kandidaten  $\hat{\lambda}$  für ein Minimum erhalten wir durch Nullsetzen der Ableitung, also

$$0 = -n + \left( \sum_{k=1}^n x_k \right) \cdot \frac{1}{\hat{\lambda}}$$

woraus folgt

$$\hat{\lambda} = \frac{1}{n} \cdot \left( \sum_{k=1}^n x_k \right) = \bar{x}$$

Dabei handelt es sich in der Tat um ein Maximum, denn

$$f''(\lambda) = -\frac{\bar{x}}{\lambda^2}$$

und da  $\bar{x}$  bei einem Poisson-Prozess positiv ist, ist  $f''(\hat{\lambda}) < 0$ , und damit ist das arithmetische Mittel ein Maximum-Likelihoodschätzer für  $\lambda$ .

**Beispiel 6.10.** Wir betrachten eine geometrisch verteilte Zufallsvariable  $X$ , mit unbekanntem Parameter  $\theta = p$ , d.h.

$$p(X = k) = \theta \cdot (1 - \theta)^{k-1} \quad (k = 1, 2, 3, \dots)$$

mit einem unbekannten Parameter  $0 < \theta < 1$ . Ferner betrachten wir  $n$  unabhängige Wiederholungen  $X_1, \dots, X_n$  des Zufallsexperiments mit den Ergebnissen  $x_1, \dots, x_n$ . Dann ermittelt sich ein Maximum-Likelihood-Schätzer für den Parameter  $\theta$  wie folgt: Es gilt für die diskrete Dichte von  $X$ , (falls  $\theta$  der Parameter von  $X$  ist):

$$f(x|\theta) = \theta \cdot (1 - \theta)^{x-1} \quad (x = 1, 2, 3, \dots)$$

Damit hat die Likelihoodfunktion die Gestalt

$$L(\theta) = f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) = \prod_{i=1}^n \theta \cdot (1 - \theta)^{x_i-1} = \theta^n \cdot (1 - \theta)^{\sum_{i=1}^n x_i - n}$$

Logarithmieren ergibt

$$l(\theta) = \ln L(\theta) = n \cdot \ln(\theta) + \left( \sum_{i=1}^n x_i - n \right) \cdot \ln(1 - \theta)$$

und es reicht, diese logarithmierte Likelihoodfunktion zu maximieren.

$$l'(\theta) = \frac{n}{\theta} - \frac{\sum_{i=1}^n x_i - n}{1 - \theta}$$

Notwendige Bedingung für ein Maximum ist

$$0 = \frac{n}{\theta} - \frac{\sum_{i=1}^n x_i - n}{1 - \theta}$$

oder äquivalent dazu

$$\frac{n}{\theta} = \frac{\sum_{i=1}^n x_i - n}{1 - \theta}$$

bzw.

$$n \cdot (1 - \theta) = \theta \cdot \left( \sum_{i=1}^n x_i - n \right)$$

bzw.

$$n = \theta \cdot \left( \sum_{i=1}^n x_i \right)$$

mit der einzigen Lösung

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i}$$

(also das Inverse des arithmetischen Mittels). Beachten Sie dabei, dass nach Voraussetzung  $0 < \theta < 1$  gelten muss; der Extremfall  $\hat{\theta} = 1$  (also alle  $x_i = 1$ ) muss also hier ausgeschlossen werden (in diesem Fall ist auch  $f'(\hat{\theta})$  nicht definiert).

Ferner ist

$$l''(\theta) = -\frac{n}{\theta^2} - \frac{\sum_{i=1}^n x_i - n}{(1 - \theta)^2}$$

so dass also  $l''(\hat{\theta}) < 0$ , und damit ist  $\hat{\theta}$  in der Tat ein Maximum-Likelihood Schätzer.

**Beispiel 6.11.** Von der stetigen Zufallsvariable  $X$  ist bekannt, dass ihre Dichtefunktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  die Gestalt

$$f(x|\theta) = \begin{cases} (\theta + 1) \cdot (1 - x)^\theta & \text{für } 0 < x < 1 \\ 0 & \text{sonst} \end{cases}$$

mit einem unbekannten Parameter  $\theta > 0$  hat. Bei  $n$  unabhängige Wiederholungen  $X_1, \dots, X_n$  des Zufallsexperiments  $X$  werden die Ergebnissen  $x_1, \dots, x_n$  beobachtet.

Dann ermittelt sich ein Likelihood-Schätzer für  $\theta$  wie folgt:

Die Likelihoodfunktion hat die Gestalt

$$\begin{aligned} L(\theta) &= f(x_1, \dots, x_n | \theta) \\ &= \prod_{i=1}^n f(x_i | \theta) \\ &= (\theta + 1) \cdot (1 - x_1)^\theta \cdots (\theta + 1) \cdot (1 - x_n)^\theta \\ &= (\theta + 1)^n \cdot (1 - x_1)^\theta \cdots (1 - x_n)^\theta \end{aligned}$$

Logarithmieren ergibt

$$\begin{aligned} l(\theta) &= \ln L(\theta) = \ln((\theta + 1)^n \cdot (1 - x_1)^\theta \cdots (1 - x_n)^\theta) \\ &= n \cdot \ln(\theta + 1) + \theta \cdot \ln(1 - x_1) + \cdots + \theta \cdot \ln(1 - x_n) \\ &= n \cdot \ln(\theta + 1) + \theta \cdot \sum_{i=1}^n \ln(1 - x_i) \end{aligned}$$

und es reicht, diese logarithmierte Likelihoodfunktion zu maximieren.

$$l'(\theta) = \frac{n}{\theta + 1} + \sum_{i=1}^n \ln(1 - x_i)$$

Notwendige Bedingung für ein Maximum ist daher

$$0 = \frac{n}{\theta + 1} + \sum_{i=1}^n \ln(1 - x_i)$$

und Auflösen nach  $\theta$  zeigt, dass

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^n \ln(1 - x_i)} - 1$$

die einzige Lösung dieser Gleichung ist. Dabei handelt es sich in der Tat um einen Maximum-Likelihood Schätzer, da

$$l''(\hat{\theta}) = -\frac{n}{(\hat{\theta} + 1)^2} < 0$$

**Beispiel 6.12.** Wir betrachten die  $n$ -fache Wiederholung einer normalverteilten Zufallsvariable  $X$  mit Parametern  $\mu$  und  $\sigma^2$ , die wir schätzen wollen (also  $\theta(\mu, \sigma^2)$  und  $n$  Realisationen  $x_1, \dots, x_n$ ). Dann gilt

$$\begin{aligned} L(\mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1 - \mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_n - \mu)^2}{2\sigma^2}} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\sum_{k=1}^n \frac{(x_k - \mu)^2}{2\sigma^2}} \end{aligned}$$

Logarithmieren der Likelihoodfunktion liefert

$$l(\mu, \sigma) = \ln(L(\mu, \sigma)) = -n \cdot \ln(\sqrt{2\pi} \cdot \sigma) + \left( -\sum_{k=1}^n \frac{(x_k - \mu)^2}{2\sigma^2} \right)$$

Die partiellen Ableitungen der logarithmierten Likelihoodfunktion sind

$$\begin{aligned} \frac{\partial l(\mu, \sigma)}{\partial \mu} &= \sum_{k=1}^n \frac{x_k - \mu}{\sigma^2} \\ \frac{\partial l(\mu, \sigma)}{\partial \sigma} &= \sum_{k=1}^n \left( -\frac{1}{\sigma} + \frac{2(x_k - \mu)^2}{2\sigma^3} \right) \end{aligned}$$

Eine notwendige Bedingung für die Existenz eines Extremums ist das Verschwinden beider partieller Ableitungen, was uns folgendes Gleichungssystem gibt:

$$\begin{aligned} 0 &= \sum_{k=1}^n \frac{x_k - \hat{\mu}}{\hat{\sigma}^2} \\ 0 &= \sum_{k=1}^n \left( -\frac{1}{\hat{\sigma}} + \frac{2(x_k - \hat{\mu})^2}{2\hat{\sigma}^3} \right) \end{aligned}$$

Aus der ersten Gleichung folgt (für  $\sigma \neq 0$ )

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{k=1}^n x_k$$

und die zweite Gleichung liefert (nach Multiplikation mit  $\sigma$ )

$$n = \sum_{k=1}^n \frac{2(x_k - \hat{\mu})^2}{2\hat{\sigma}^2}$$

woraus folgt

$$\hat{\sigma} = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^n (x_k - \hat{\mu})^2}$$

Hierbei handelt es sich in der Tat um ein Maximum, wie die Betrachtung der Hessematrix zeigt.

**Beispiel 6.13.** Von einer stetigen Zufallsvariable  $X$  ist bekannt, dass sie die Wahrscheinlichkeitsdichte

$$f(x|\alpha) = \begin{cases} \alpha \cdot e^{-\alpha \cdot x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

mit einem unbekannten Parameter  $\alpha > 0$  hat. Bei  $n$  unabhängige Wiederholungen  $X_1, \dots, X_n$  des Zufallsexperiments  $X$  werden die Ergebnissen  $x_1, \dots, x_n$  beobachtet.

Einen Maximum-Likelihood-Schätzer für den Parameter  $\alpha$  können wir wie folgt ermitteln:

Die Likelihoodfunktion hat die Gestalt

$$\begin{aligned} L(\alpha) &= f(x_1, \dots, x_n | \alpha) \\ &= \prod_{i=1}^n f(x_i | \alpha) \\ &= \alpha \cdot e^{-\alpha \cdot x_1} \dots \alpha \cdot e^{-\alpha \cdot x_n} \\ &= \alpha^n \cdot e^{-\sum_{i=1}^n \alpha \cdot x_i} \end{aligned}$$

Logarithmieren ergibt

$$l(\alpha) = \ln L(\alpha) = \ln \left( \alpha^n \cdot e^{-\sum_{i=1}^n \alpha \cdot x_i} \right) = n \cdot \ln(\alpha) + \sum_{i=1}^n -\alpha \cdot x_i = n \cdot \ln(\alpha) - \alpha \cdot (x_1 + \dots + x_n)$$

und es reicht, diese logarithmierte Likelihoodfunktion zu maximieren.

$$l'(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^n x_i$$



Notwendige Bedingung für ein Maximum ist daher

$$0 = \frac{n}{\alpha} - \sum_{i=1}^n x_i$$

und Multiplikation mit  $\alpha$  und Auflösen nach  $\alpha$  zeigt, dass

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n x_i}$$

(also das inverse des arithmetischen Mittels) die einzige Lösung dieser Gleichung ist. Dabei handelt es sich in der Tat um einen Maximum-Likelihood Schätzer, denn

$$l''(\alpha) = -\frac{n}{\alpha^2}$$

und damit gilt, dass  $l''(\hat{\alpha}) < 0$ .

**Bemerkung 6.7.** Weitere Methoden, gute Schätzstatistiken zu erhalten, ergeben sich aus der Kleinst-Quadrate Schätzung (auf die wir im Rahmen der Regressionsanalyse nochmal kommen) und der Bayes-Schätzung, die eine enge Verbindung zur Maximum-Likelihood-Methode hat.

Wir betrachten nun Stichprobenvariablen  $X_1, \dots, X_n, \dots$ , ein  $0 < \alpha < 1$  und Schätzstatistiken

$$G_u = G_u(X_1, \dots, X_n), \quad G_o = G_o(X_1, \dots, X_n)$$

**Definition 6.9.** Die Schätzstatistiken  $G_u$  und  $G_o$  liefern ein  $1 - \alpha$ -Konfidenzintervall für  $\theta$ , wenn

$$\begin{aligned} p(G_u \leq G_o) &= 1 \\ p(G_u \leq \theta \leq G_o) &\geq 1 - \alpha \end{aligned}$$

Dann heißt  $\alpha$  die *Irrtumswahrscheinlichkeit* und  $1 - \alpha$  die *Konfidenzwahrscheinlichkeit* der Schätzer  $[G_u, G_o]$ .

**Bemerkung 6.8.** Aus den Realisationen  $x_1, \dots, x_n$  der Stichprobe ergeben sich realisierte Konfidenzintervalle der Form

$$I = [g_u, g_o]$$

wobei  $g_u = G_u(x_1, \dots, x_n)$  und  $g_o = G_o(x_1, \dots, x_n)$ .

**Bemerkung 6.9.** Setzt man  $G_u = -\infty$ , so erhält man ein einseitiges Konfidenzintervall  $] - \infty, G_o]$  mit

$$p(\theta \leq G_o) \geq 1 - \alpha$$

(und entsprechend für  $G_o = \infty$ .)

Wir betrachten nun einen i.i.d. Prozess  $\mathbb{X} = (X_n)_{n \geq 1}$  mit normalverteilten Zufallsvariablen  $X_n \sim N(\mu, \sigma)$  und suchen eine Konfidenzintervallschätzung für  $\mu$ .

### 1. Fall: $\sigma$ ist bekannt

In dieser Situation gilt für das arithmetische Mittel

$$\bar{X} = \frac{1}{n} \cdot \sum_{k=1}^n X_k$$

nach unseren Rechenregeln für Erwartungswert und Varianz

$$\begin{aligned} E(\bar{X}) &= \mu \\ \text{Var}(\bar{X}) &= \frac{\sigma^2}{n} \end{aligned}$$

und ausserdem ist  $\bar{X}$  normalverteilt (als normierte Summe unabhängiger Wiederholungen identisch normalverteilter Variablen), d.h.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Damit erhalten wir

$$\tilde{X} := \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Bezeichnen wir nun mit  $z = z_{1-\frac{\alpha}{2}}$  das  $1 - \frac{\alpha}{2}$ -Quantil der Standardnormalverteilung (also den Wert  $z$  mit  $\Phi(z) = 1 - \frac{\alpha}{2}$ ), so gilt (aufgrund der Symmetrie von  $N(0, 1)$ ):

$$p\left(-z \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right) = 1 - \alpha$$

also, durch einfaches Umrechnen

$$\begin{aligned} 1 - \alpha &= p\left(-z \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right) \\ &= p\left(\bar{X} - z \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z \cdot \frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

und wir erhalten als  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$  das Intervall

$$I = \left[\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Die Breite des Konfidenzintervalls nimmt also mit zunehmendem Stichprobenumfang ab.

### 2. Fall: $\sigma$ ist unbekannt

In diesem Fall ersetzen wir die Schätzvariable  $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$  durch den Schätzer

$$\tilde{X} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X} - \mu}{S} \cdot \sqrt{n}$$

wobei

$$S = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2}$$

also ein Schätzer für die Streuung. Diese Variable ist  $t$ -verteilt mit  $n-1$  Freiheitsgraden,

$$\frac{\bar{X} - \mu}{S} \cdot \sqrt{n} \sim t(n-1)$$

Ist daher  $t = t_{1-\frac{\alpha}{2}, n-1}$  das  $1 - \frac{\alpha}{2}$ -Quantil der  $t$ -Verteilung mit  $n-1$  Freiheitsgraden, so erhalten wir

$$p\left(-t \leq \frac{\bar{X} - \mu}{S} \cdot \sqrt{n} \leq t\right) = 1 - \alpha$$

und damit in Analogie zum ersten Fall das  $(1 - \alpha)$ -Konfidenzintervall

$$J = \left[ \bar{X} - t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}} \right]$$

Ist nun  $\mathbb{X} = (X_n)_{n \geq 1}$  ein i.i.d.-Prozess mit einer beliebigen Verteilung (mit Erwartungswert  $\mu$  und Varianz  $\sigma^2$ ), so können wir den zentralen Grenzwertsatz anwenden und erhalten immer noch dass der Schätzer

$$\tilde{X} := \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

approximativ standardnormalverteilt ist. Damit sind

$$I = \left[ \bar{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

bei bekanntem  $\sigma$  und

$$J = \left[ \bar{X} - t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}} \right]$$

bei unbekanntem  $\sigma$  für hinreichend hohe  $n$  immer noch approximative  $(1 - \alpha)$ -Konfidenzintervalle für  $\mu$ .

**Bemerkung 6.10.** Als Faustregel für die Verwendung der approximativen Konfidenzintervalle gilt, dass der Stichprobenumfang  $n > 30$  sein muss.

**Bemerkung 6.11.** Zur Schätzung von  $\sigma^2$  bei einem i.i.d.-Prozess  $\mathbb{X} = (X_n)_{n \geq 1}$  mit normalverteilten Zufallsvariablen  $X_n \sim N(\mu, \sigma)$  benutzen wir die Schätzstatistik

$$S^2 = \frac{1}{n-1} \cdot \sum_{k=1}^n (X_k - \bar{X})^2$$

Dann gilt

$$\frac{n-1}{\sigma^2} \cdot S^2 \sim \chi^2(n-1)$$

ist  $\chi^2$ -verteilt mit  $n-1$  Freiheitsgraden. Diese Verteilung ist nicht symmetrisch, daher betrachten wir das  $\frac{\alpha}{2}$ -Quantil  $q_1 = q_{\frac{\alpha}{2}, n-1}$  und das  $1 - \frac{\alpha}{2}$ -Quantil  $q_2 = q_{\frac{1-\alpha}{2}, n-1}$  der  $\chi^2(n-1)$ -Verteilung und erhalten mit einer Rechnung wie oben, dass

$$K = \left[ \frac{(n-1)S^2}{q_2}, \frac{(n-1)S^2}{q_1} \right]$$

ein  $(1 - \alpha)$ -Konfidenzintervall für  $\sigma^2$  ist.

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 10

Stand: Mai 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

`e-mail:reinhold.huebl@dhbw-mannheim.de`

©Fakultät für Technik DHBW Mannheim

## 7 Hypothesentests

Der Hypothesentest gibt uns bei Entscheidungen eine Richtlinie für die Wahl einer Alternativenentscheidung. Wir treffen unsere Entscheidung aufgrund dessen, was wir für richtig halten. Das nennen wir die *Nullhypothese*. Die Alternativenentscheidung bezeichnen wir als *alternative Hypothese*.

**Beispiel 7.1.** Die Produktion einer Firma soll auf ihre Qualität hin geprüft werden. Dabei wird nur unterschieden, ob ein Produkt den Qualitätsstandards genügt (*gut*) oder nicht (*fehlerhaft, Ausschuß*). Um Vertragsstrafen zu vermeiden, darf der *Ausschußanteil*  $a$  einen maximalen Wert  $\pi$  nicht überschritten werden. In der Regel ist es aber nicht möglich, alle Produkte einzeln auf ihre Qualität zu überprüfen, sei es aus Kosten- oder Zeitgründen, sei es weil die Prüfung das Produkt zerstört (wie etwa bei Lebensmitteln). Wir betrachten also

**Nullhypothese**  $H_0: a \leq \pi$ .

**Alternativhypothese**  $H_1: a > \pi$ .

Aufgrund statistischer Stichproben soll nun eine Entscheidung getroffen werden. Diese Entscheidung ist immer von der Form  $H_0$  **wird abgelehnt** oder  $H_0$  **kann nicht abgelehnt werden**.

**Beispiel 7.2.** Sie haben die Vermutung, dass eine Münze manipuliert wurde und „Wappen“ jetzt häufiger auftritt als „Zahl“. Wenn wir mit  $p_0$  die Wahrscheinlichkeit für „Wappen“ bezeichnen, so betrachten in diesem Fall

**Nullhypothese**  $H_0: p_0 = \frac{1}{2}$ .

**Alternativhypothese**  $H_1: p_0 > \frac{1}{2}$ .

Aufgrund statistischer Stichproben soll auch hier eine Entscheidung getroffen werden.

Zur Formalisierung dieser Situation betrachten wir zu unserem Zufallsexperiment eine Zufallsstichprobe  $Y_1, Y_2, \dots, Y_n, \dots$  und ein Ereignis  $A$ , das wir betrachten wollen. Damit definieren wir einen neuen i.i.d.-Prozess  $\mathbb{X} = (X_n)_{n \geq 1}$  mit dichotomen Zufallsvariablen mit

$$X_n(\omega) = \begin{cases} 1 & \text{falls } A \text{ in der } n\text{-ten Stichprobe eintritt} \\ 0 & \text{falls } A \text{ in der } n\text{-ten Stichprobe nicht eintritt} \end{cases}$$

und als statistische Prüfgröße nehmen wir bei einem Stichprobenumfang von  $n$  die Zufallsvariable

$$\Sigma = \Sigma_n = \sum_{k=1}^n X_k$$

Dann ist  $\Sigma$  binomialverteilt mit Parametern  $n$  und  $p_0$ , wobei  $p_0$  die Wahrscheinlichkeit für das Auftreten von  $A$  ist, also

$$\Sigma \sim B(n, p_0)$$

so dass sich die Wahrscheinlichkeitswerte nach der Formel

$$p(\Sigma = k) = \binom{n}{k} p_0^k (1 - p_0)^{n-k}$$

berechnen.

Betrachten wir etwa das Beispiel mit dem Münzwurf, und interessieren wir uns für die Frage, ob „Wappen“ tatsächlich mit Wahrscheinlichkeit  $\frac{1}{2}$  fällt, so betrachten wir also die Zufallsvariable  $X$  wobei

$$X(\omega) = \begin{cases} 1 & \text{falls } \omega = \text{„Wappen“ wird geworfen} \\ 0 & \text{sonst} \end{cases}$$

Dann ist  $X$  eine Bernoullivariablen mit Parameter  $p$ , wobei  $p$  die wahre Wahrscheinlichkeit für „Wappen“ ist. Machen wir eine Stichprobe von 12 Würfeln, also 12 stochastisch unabhängige Wiederholungen von  $X$ , so könnte eine statistische Herangehensweise etwa wie folgt strukturiert sein:

1. Wiederhole den Münzwurf 12 mal, dh. führe 12 stochastisch unabhängige Wiederholungen von  $X$  durch.
2. Zähle, wie oft „Wappen“ bei der Durchführung auftritt, dh. betrachte die Zufallsvariable  $A = X_1 + \dots + X_{12}$ .
3. Ist die Nullhypothese korrekt, so würden wir etwa sechs mal „Wappen“ erwarten. Eine sehr hohe Anzahl von „Wappen“ ist dann ziemlich unwahrscheinlich.
4. Wir betrachten also die Realisation  $a$  von  $A$  und vergleichen  $a$  mit dem Wert 6, den wir bei Gültigkeit der Nullhypothese erwarten würden.
5. Ist  $a$  sehr viel größer als 6, so spricht das aus der Sicht der Alternativhypothese gegen die Nullhypothese, und wir würden  $H_0$  ablehnen.

Zur Formalisierung dieser Herangehensweise setzen wir

$$\Sigma = X_1 + X_2 + \cdots + X_{12}$$

Ist  $\Sigma \sim B(12, 0.5)$ , so erhalten wir folgende Tabelle

$k$	0	1	2	3	4	5	6
$p(\Sigma = k)$	0.00025	0.0029	0.0161	0.0537	0.1208	0.1934	0.2256

Wegen  $p(\Sigma = 12 - k) = p(\Sigma = k)$  ist hierdurch schon die gesamte Wahrscheinlichkeitsverteilung festgelegt.

Da unter der Annahme der Nullhypothese gleich oft „Wappen“ und „Zahl“ zu erwarten ist, sprechen mehr als 6 Wappen für die Alternativhypothese  $H_1 : p_0 > \frac{1}{2}$ , und die Frage, die sich jetzt stellt, lautet:

Wie groß muss die Anzahl der „Wappen“ sein, damit die Nullhypothese sehr oder gar extrem unwahrscheinlich wird?

Aus obiger Binomialverteilung können wir nun direkt ablesen, wie klein die Wahrscheinlichkeit für hohe Anzahlen von „Wappen“ wird. Übliche Wahrscheinlichkeitswerte, die hier betrachtet werden, sind 0.01, 0.05 oder 0.1. Diese Wahrscheinlichkeiten  $\alpha$  heißen **maximal erlaubte Fehlerwahrscheinlichkeiten**. Den Komplementärwert  $1 - \alpha$  bezeichnen wir als **Signifikanzniveau**, und damit (und mit den Werten der Binomialverteilung) konstruieren wir **Ablehnungsbereich**, also Wertebereiche für  $\Sigma$ , so daß  $\Sigma$  mit einer Wahrscheinlichkeit  $\leq \alpha$  in diesem Bereich liegt, wenn  $H_0$  zutrifft. Wir schreiben  $p_{H_0}(\Sigma \in B)$  für die Wahrscheinlichkeit, dass  $\Sigma \in B$ , wenn  $H_0$  zutrifft.

Bei dem Würfelbeispiel etwa ergibt sich

$$\begin{aligned} p_{H_0}(\Sigma \geq 11) &= 0.0032 \\ p_{H_0}(\Sigma \geq 10) &= 0.0193 \\ p_{H_0}(\Sigma \geq 9) &= 0.0730 \\ p_{H_0}(\Sigma \geq 8) &= 0.1938 \end{aligned}$$

Damit ist bei einem Signifikanzniveau von 0.99 die Nullhypothese nur dann abzulehnen, wenn mindestens elf mal „Wappen“ fällt, bei einem Signifikanzniveau von 0.95 reichen 10 solche Ereignisse, und bei einem Signifikanzniveau von 0.90 reichen 9 für eine Ablehnung aus.

Ändern wir dieses Beispiel dahingehend ab, dass wir vermuten, dass die Münze keine Laplacemünze ist (ohne zu vermuten, welche Seite bevorzugt wird), so sind die Hypothesen wie folgt zu formulieren:

**Nullhypothese**  $H_0: p_0 = \frac{1}{2}$ .



**Alternativhypothese**  $H_1: p_0 \neq \frac{1}{2}$ .

In diesem Fall sprechen nicht nur sehr hohe Anzahlen von „Wappen“ gegen die Nullhypothese sondern auch sehr niedrige. Damit ist bei einem Signifikanzniveau von 0.99 die Nullhypothese dann abzulehnen, wenn mindestens elf mal oder höchstens einmal „Wappen“ fällt, bei einem Signifikanzniveau von 0.95 ist der Ablehnungsbereich  $\{0, 1, 2, 10, 11, 12\}$ , und bei einem Signifikanzniveau von 0.9 erhalten wir in diesem Fall den gleichen Ablehnungsbereich.

**Regel 7.1** (Exakter Binomialtest). *Wir betrachten die folgenden Testprobleme über den Parameter  $p_0$  einer Bernoulli-verteilten Zufallsvariable  $X$ :*

1. Nullhypothese  $H_0 : p = p_0$  gegen Alternativhypothese  $H_1 : p \neq p_0$ .
2. Nullhypothese  $H_0 : p = p_0$  gegen Alternativhypothese  $H_1 : p < p_0$ .
3. Nullhypothese  $H_0 : p = p_0$  gegen Alternativhypothese  $H_1 : p > p_0$ .

Basierend auf  $n$  unabhängigen Wiederholungen  $X_1, \dots, X_n$  von  $X$  und der Prüfgröße

$$Z = \sum_{i=1}^n X_i$$

mit gegebenen Realisation  $z$  und einem vorgegebenen Signifikanzniveau  $1 - \alpha$  fällt die Entscheidung gegen  $H_0$  im Testproblem

1. falls  $z \leq k$  oder  $z \geq l$ , wobei  $k$  maximal mit  $p(B(n, p_0) \leq k) \leq \frac{\alpha}{2}$  und  $l$  minimal mit  $p(B(n, p_0) \geq l) \leq \frac{\alpha}{2}$ .
2. falls  $z \leq k$ , wobei  $k$  maximal mit  $p(B(n, p_0) \leq k) \leq \alpha$ .
3. falls  $z \geq l$ , wobei  $l$  minimal mit  $p(B(n, p_0) \geq l) \leq \alpha$ .

**Beispiel 7.3.** Die Besitzerin einer Losbude auf einem Volksfest beschäftigt einen Angestellten, der Besucher anspricht und sie zum Kauf eines Loses animieren soll. Sie ist mit der Erfolgsquote des Angestellten unzufrieden und möchte ihm das Gehalt kürzen, wenn weniger als 15 % der angesprochenen Personen Lose kaufen. Zur Entscheidung über die Gehaltskürzung will sie einen Test mit 60 (zufällig ausgewählten) angesprochenen Besuchern des Volksfests durchführen, wobei vermieden werden soll, dem Angestellten den Lohn zu unrecht zu kürzen. Ferner soll bei der Entscheidung höchstens ein Fehler von 10 % auftreten.

In diesem Fall ist ein Binomialtest durchzuführen. Dazu bezeichnen wir mit  $X$  die Zufallsvariable mit

$$X(\omega) = \begin{cases} 1 & \text{falls der angesprochene Besucher } \omega \text{ ein Los kauft} \\ 0 & \text{sonst} \end{cases}$$

Dann ist  $X$  eine Bernoulli-Variable mit Parameter  $p$ , wobei  $p$  die tatsächliche Wahrscheinlichkeit dafür ist, dass der Angestellte erfolgreich ist und den Besucher zum Loskauf überredet.

**Nullhypothese**  $H_0: p = 0.15$ .

**Alternativhypothese**  $H_1: p < 0.15$ .

Wir betrachten 60 unabhängige Wiederholungen  $X_1, \dots, X_{60}$  von  $X$  (also eine Zufallsstichprobe von 60 angesprochenen Besuchern). Die Entscheidungsgröße ist

$$A = X_1 + \dots + X_{60}$$

Dann ist  $A \sim B(60, p)$ . Ist also die Nullhypothese richtig, so ist  $A \sim B(60, 0.15)$ , und wir werden die Nullhypothese ablehnen, wenn für die Realisation  $a$  von  $A$  gilt:  $a \leq k$  wobei  $k$  maximal mit  $p(B(60, 0.15) \leq k) \leq 0.10$ .

Die relevanten Werte für eine binomialverteilte Zufallsvariable  $B \sim B(60, 0.15)$  ermitteln sich aus der Formel

$$p(B = k) = \binom{60}{k} \cdot 0.15^k \cdot 0.85^{60-k}$$

Für kleine Werte von  $k$  gilt also

$k$	0	1	2	3	4	5	6	7
$p(B = k)$	0.00006	0.00062	0.0032	0.0110	0.0275	0.0544	0.0880	0.1199
$p(B \leq k)$	0.00006	0.00068	0.0039	0.0148	0.0424	0.0968	0.1848	0.3047

Speziell wird bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha = 0.10$  die Nullhypothese also genau dann abgelehnt, wenn  $z \leq 5$ .

Die Auswertung der Untersuchung liefert, dass genau 6 der angesprochenen 60 Besucher ein Los gekauft haben.

Es ist also  $a = 6$ . Das  $6 \not\leq 5$ , kann die Nullhypothese in diesem Fall nicht abgelehnt werden.

**Beispiel 7.4.** Sie untersuchen einen Würfel, da sie vermuten, dass bei diesem Würfel die Wahrscheinlichkeit dafür, die Zahl „6“ zu würfeln weniger als  $\frac{1}{6}$  beträgt. Sie wollen mit einer maximalen Fehlerwahrscheinlichkeit von  $\alpha = 0.05$  zu einer Entscheidung kommen. Dazu betrachten wir die Bernoullivariablen  $X$  mit

$$X(\omega) = \begin{cases} 1 & \text{die Zahl '6' wird gewürfelt} \\ 0 & \text{eine andere Zahl fällt} \end{cases}$$

(wobei  $p$  die tatsächliche Wahrscheinlichkeit für die Zahl „6“ ist) und untersuchen die Hypothesen

**Nullhypothese**  $H_0: p = \frac{1}{6}$ .

**Alternativhypothese**  $H_1: p < \frac{1}{6}$ .

Wir betrachten 18 unabhängige Wiederholungen  $X_1, \dots, X_{18}$  von  $X$  (also eine Zufallsstichprobe von 18 Würfeln). Die Entscheidungsgröße ist

$$A = X_1 + \dots + X_{18}$$

Dann ist  $A \sim B(18, p)$ . Ist also die Nullhypothese richtig, so ist  $A \sim B(18, \frac{1}{6})$ , und wir werden die Nullhypothese ablehnen, wenn für die Realisation  $a$  von  $A$  gilt:  $a \leq k$  wobei  $k$  maximal mit  $p(B(18, \frac{1}{6}) \leq k) \leq 0.05$ .

Angenommen, eine Zufallsbeobachtung von 18 unabhängigen Würfeln dieses Würfels liefert die folgenden Ergebnisse

Wurf	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Wert	1	1	3	5	1	3	5	6	4	4	2	1	4	2

Wurf	15	16	17	18
Wert	5	4	1	3

Für  $B \sim B(18, \frac{1}{6})$  ist  $p(B = 0) = 0.0376$  und  $p(B \leq 1) = 0.1728$ , daher kann  $H_0$  bei  $\alpha = 0.05$  nur abgelehnt werden, wenn kein einziges Mal die „6“ fällt.

In diesem Beispiel kann also  $H_0$  nicht abgelehnt werden.

Die Binomialverteilung ist für große Stichproben numerisch nicht einfach zu ermitteln, allerdings ist sie für große  $n$  approximativ normalverteilt. Genauer gilt für  $\Sigma$  wie oben

$$\Sigma \stackrel{a}{\sim} N(n \cdot p_0, n \cdot p_0(1 - p_0))$$

also

$$Z := \frac{\Sigma - np_0}{\sqrt{np_0(1 - p_0)}} \stackrel{a}{\sim} N(0, 1)$$

unter der Annahme, dass die Nullhypothese  $H_0$  korrekt ist, dass also  $p_0$  die korrekte Wahrscheinlichkeit für das Auftreten des Ereignisses  $A$  ist.

Durch Ausklammern von  $n$  aus Zähler und Nenner können wir  $Z$  schreiben als

$$Z = \frac{\Sigma - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1 - p_0)}} = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0 \cdot (1 - p_0)}{n}}} = \frac{\bar{X} - p_0}{\sqrt{p_0 \cdot (1 - p_0)}} \cdot \sqrt{n}$$

mit dem arithmetischen Mittel  $\bar{X} = \frac{1}{n} \cdot \sum_1^k X_k$ , das wir dann durch das arithmetische Mittel  $\bar{x} = \frac{1}{n} \cdot \sum_1^k x_k$  der Beobachtungen ersetzen.

**Regel 7.2** (Approximativer Binomialtest). *Wir betrachten die folgenden Testprobleme über den Parameter  $p_0$  einer Bernoulli-verteilten Zufallsvariable  $X$ :*

1. Nullhypothese  $H_0 : p = p_0$  gegen Alternativhypothese  $H_1 : p \neq p_0$ .
2. Nullhypothese  $H_0 : p = p_0$  gegen Alternativhypothese  $H_1 : p < p_0$ .
3. Nullhypothese  $H_0 : p = p_0$  gegen Alternativhypothese  $H_1 : p > p_0$ .

Basierend auf  $n$  unabhängigen Wiederholungen  $X_1, \dots, X_n$  von  $X$  mit  $n \cdot p_0 \cdot (1 - p_0) > 9$  und der (unter der Annahme der Gültigkeit der Nullhypothese) approximativ standard-normalverteilten Prüfgröße

$$Z = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{a}{\sim} N(0, 1)$$

mit gegebenen Realisation  $z$  und einem vorgegebenen Signifikanzniveau  $1 - \alpha$  fällt die Entscheidung gegen  $H_0$  im Testproblem

1. , falls  $|z| > z_{1-\frac{\alpha}{2}}$ .
2. , falls  $z < -z_{1-\alpha}$ .
3. , falls  $z > z_{1-\alpha}$ .

Dabei bezeichnet zu vorgegebenem  $\beta \in ]0, 1[$  die Zahl  $z_\beta$  das  $\beta$ -Quantil der Standardnormalverteilung.

Beachten Sie, dass  $n \cdot p_0 \cdot (1 - p_0) > 9$  gelten muss, um den approximativen Binomialtest anwenden zu können.

**Beispiel 7.5.** Wir greifen das Beispiel 7.3 mit der Losbude nochmal auf. Die Besitzerin wiederholt ihren Test, diesmal mit 300 zufällig ausgewählten Besuchern (unter sonst gleichen Voraussetzungen und Bedingungen), die von dem Angestellten angesprochen werden.

Da  $300 \cdot 0.15 \cdot 0.85 = 38.25 > 9$ , kann sie jetzt auf einen approximativen Binomialtest zurückgreifen. Null- und Alternativhypothesen und die untersuchten Zufallsvariablen  $X$  und  $X_1, \dots, X_n$  bleiben gleich, ebenso  $A = X_1 + \dots + X_{300}$ .

Setzt Sie jetzt außerdem noch  $\bar{A} = \frac{1}{300} \cdot \sum_{k=1}^{300} X_k$  und  $Z = \frac{\bar{A} - 0.15}{\sqrt{0.15 \cdot 0.85}} \cdot \sqrt{300}$ , so kann sie  $H_0$  (bei gleichbleibendem  $\alpha = 0.10$ ) ablehnen, falls für die Realisation  $z$  von  $Z$  gilt:

$$z < -z_{1-0.10} = -1.28$$

Die Auswertung der Stichprobe liefert, dass 36 der angesprochenen 300 Personen tatsächlich ein Los gekauft haben. Damit gilt

$$\bar{a} = \frac{36}{300} = \frac{3}{25} = 0.120$$

also

$$z = \frac{0.120 - 0.150}{\sqrt{0.1275}} \cdot \sqrt{300} = -1.455$$

Da  $-1.455 < -1.28$ , kann die Nullhypothese jetzt bei einem  $\alpha = 0.10$  abgelehnt werden. Bei  $\alpha = 0.05$  wäre das nicht mehr möglich.

Viele typische Testprobleme sind jedoch keine Aussagen über dichotome Zufallsvariablen bzw. Binomialverteilungen sondern Aussagen über die Erwartungswerte beliebiger Zufallsvariablen.

**Beispiel 7.6.** Bei einer Maschine soll überprüft werden, ob für die gefertigten Teile eine bestimmte Sollgröße  $\mu_0$  erreicht wird. Bezeichnen wir also mit  $X$  die Zufallsvariable "Länge des gefertigten Teils", so geht es um die Frage, ob  $E(X) = \mu_0$  oder nicht.

**Beispiel 7.7.** Ein Lieferant gibt an, dass die von ihm gelieferten Werkstücke ein mittleres Gewicht von 100.00 g haben, dass das tatsächliche Gewicht produktionsbedingt aber gewissen statistischen Schwankungen unterworfen ist. Sie als Abnehmer sind bereit, diese Schwankungen in Kauf zu nehmen, es ist ihnen aber wichtig, dass das angegebenen mittlere Gewicht stimmt. In diesem Fall würden Sie mit der Zufallsvariable  $X$ : *Gewicht eines Werkstücks* einen Hypothesentest wie folgt aufsetzen:

- Nullhypothese  $H_0 : E(X) = 100.00 \text{ g}$

- Alternativhypothese  $H_1 : E(X) \neq 100.00 \text{ g}$ .

**Beispiel 7.8.** Ein Pralinenhersteller gibt an, dass in einer Schachtel seiner Marke *Exklusiv* im Mittel 200.00 g enthalten sind, dass das tatsächliche Gewicht produktionsbedingt aber gewissen statistischen Schwankungen unterworfen ist. In diesem Fall ist es für Sie als Abnehmer vermutlich eher wichtig, dass das angegebene Gewicht im Mittel nicht unterschritten wird. Daher wäre mit der Zufallsvariable  $X$  : *Gewicht des Inhalts einer Pralinschachtel* ein sinnvoller Hypothesentest in diesem Fall also

- Nullhypothese  $H_0 : E(X) = 200.00 \text{ g}$
- Alternativhypothese  $H_1 : E(X) < 200.00 \text{ g}$ .

Für das Testproblem nehmen wir jetzt stets an, dass die Zufallsvariable  $X$  normalverteilt und die Varianz  $\sigma^2$  von  $X$  bekannt ist (bei Untersuchungen zu Maschinen kann das als gegeben angenommen werden, da die Hersteller hier Fehlertoleranzen angeben), und wir betrachten die Testprobleme

**Regel 7.3** (Gaußtest). *Wir betrachten die folgenden Testprobleme für den Parameter  $E(X) = \mu$  (bei bekannter Varianz  $\sigma^2$ ):*

1. Nullhypothese  $H_0 : E(X) = \mu_0$  gegen Alternativhypothese  $H_1 : E(X) \neq \mu_0$ .
2. Nullhypothese  $H_0 : E(X) = \mu_0$  gegen Alternativhypothese  $H_1 : E(X) < \mu_0$ .
3. Nullhypothese  $H_0 : E(X) = \mu_0$  gegen Alternativhypothese  $H_1 : E(X) > \mu_0$ .

Dazu definieren wir den i.i.d.-Prozess  $\mathbb{X} = (X_n)_{n \geq 1}$ , wobei die  $X_n$  unabhängige Wiederholungen von  $X$  sind, und als Prüfgröße nehmen wir das arithmetische Mittel

$$\bar{X} = \bar{X}_n = \frac{1}{n} \cdot \sum_{k=1}^n X_k$$

bzw. die normierte Form

$$Z = Z_n = \frac{\bar{X} - \mu_0}{\sigma} \cdot \sqrt{n}$$

Ist  $X$  normalverteilt mit Erwartungswert  $\mu_0$  und Varianz  $\sigma^2$ ,  $X \sim N(\mu_0, \sigma^2)$ , so ist  $Z$  standardnormalverteilt, im allgemeinen Fall ist  $Z$  (unter Annahme der Gültigkeit der

Nullhypothese) nach dem Gesetz der große Zahl zumindest approximativ standardnormalverteilt,

$$Z \stackrel{a}{\sim} N(0, 1)$$

Zu gegebener Realisation  $z$  und einem vorgegebenen Signifikanzniveau  $\alpha$  fällt die Entscheidung gegen  $H_0$  im Testproblem

1. , falls  $|z| > z_{1-\frac{\alpha}{2}}$ .
2. , falls  $z < -z_{1-\alpha}$ .
3. , falls  $z > z_{1-\alpha}$ .

**Beispiel 7.9.** Wir greifen noch einmal Beispiel 7.7 und nehmen jetzt an, dass der Hersteller zusätzlich angibt, dass das Gewicht des Werkstückes normalverteilt mit einer Varianz von 0.1 g ist. (Die Normalverteilungsannahme ist bei technischen Produktionsprozessen in der Regel gegeben). In diesem Fall kann das Testszenario wie folgt beschrieben werden:

**Testszenario:**

Wir bezeichnen mit  $X$  die Zufallsvariable *Gewicht eines Werkstücks* und führen einen Hypothesentest

- Nullhypothese  $H_0 : E(X) = 100.00 \text{ g}$
- Alternativhypothese  $H_1 : E(X) \neq 100.00 \text{ g}$ .

durch. Bei einer zu untersuchenden Zufallsstichprobe vom Umfang  $n$  gehen wir vor wie folgt:  $X_1, \dots, X_n$  sind stochastisch unabhängige Wiederholungen von  $X$  (das entspricht der Zufallsstichprobe), und

$$\bar{X} = \frac{1}{n} \cdot \sum_{k=1}^n X_k$$

bezeichnet das arithmetische Mittel von  $X_1, \dots, X_n$ . Die Testgröße hat die Gestalt

$$Z = \frac{\bar{X} - 100}{\sqrt{0.1}} \cdot \sqrt{n}$$

Ist  $H_0$  korrekt, so ist  $Z \sim N(0, 1)$ .

**Entscheidungsgrundlage:**

Bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha$  kommen wir zu einer Ablehnung der Nullhypothese, wenn für die Realisation  $z$  von  $Z$  gilt:

$$z < -z_{1-\frac{\alpha}{2}} \quad \text{oder} \quad z > z_{1-\frac{\alpha}{2}}$$

wobei  $z_\beta$  das  $\beta$ -Quantil der Standardnormalverteilung bezeichnet.

Speziell gilt: Bei einer zugelassenen Fehlerwahrscheinlichkeit von 5 % kommen wir zu einer Ablehnung der Nullhypothese, wenn für die Realisation  $z$  von  $Z$  gilt:

$$z < -1.96 \quad \text{oder} \quad z > 1.96$$

und bei einer zugelassenen Fehlerwahrscheinlichkeit von 1 % kommen wir zu einer Ablehnung der Nullhypothese, wenn für die Realisation  $z$  von  $Z$  gilt:

$$z < -2.58 \quad \text{oder} \quad z > 2.58$$

### Testdurchführung:

Angenommen bei einer konkreten Zufallsstichprobe vom Umfang  $n = 16$  erhalten wir als Testreihe (alle Angaben in Gramm)

100.34	100.40	99.62	100.31	100.28	99.93	100.59	100.32
100.21	99.57	100.32	99.68	99.80	100.12	100.02	99.73

In diesem Fall erhalten wir

$$\bar{x} = 100.09(\text{g})$$

und damit als Realisation der Testgröße

$$z = \frac{100.09 - 100.00}{\sqrt{0.1}} \cdot \sqrt{16} = 1.1384$$

Daher kann die Nullhypothese bei einer zugelassenen Fehlerwahrscheinlichkeit von 5 % nicht abgelehnt werden (erst recht nicht bei einer zugelassenen Fehlerwahrscheinlichkeit von 1 %).

Angenommen bei einer Zufallsstichprobe vom Umfang  $n = 200$  erhalten wir ein Durchschnittsgewicht der getesteten Werkstücke von  $\bar{x} = 99.94$  g. Dann ist

$$z = \frac{99.94 - 100.00}{\sqrt{0.1}} \cdot \sqrt{200} = -2.6833$$

Da  $z < -1.96$ , kann  $H_0$  bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha = 0.05$  abgelehnt werden. Da sogar  $z < -2.58$  kann  $H_0$  sogar bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha = 0.01$  abgelehnt werden.

**Beispiel 7.10.** Wir greifen jetzt auch noch Beispiel 7.8 und nehmen zusätzlich an, dass der Pralinenfabrikant angibt, dass das Gewicht des Inhalts normalverteilt mit einer Varianz von  $1.0 \text{ g}^2$  ist. In diesem Fall kann das Testszenario wie folgt beschrieben werden:

### Testszenario:

Wir bezeichnen mit  $X$  die Zufallsvariable *Gewicht des Inhalts einer Pralinschachtel* und führen einen Hypothesentest



- Nullhypothese  $H_0 : E(X) = 200.00 \text{ g}$
- Alternativhypothese  $H_1 : E(X) < 200.00 \text{ g}$ .

durch. Bei einer zu untersuchenden Zufallsstichprobe vom Umfang  $n$  gehen wir vor wie folgt:  $X_1, \dots, X_n$  sind stochastisch unabhängige Wiederholungen von  $X$  (das entspricht der Zufallsstichprobe), und

$$\bar{X} = \frac{1}{n} \cdot \sum_{k=1}^n X_k$$

bezeichnet das arithmetische Mittel von  $X_1, \dots, X_n$ . Die Testgröße hat die Gestalt

$$Z = \frac{\bar{X} - 200}{\sqrt{1.0}} \cdot \sqrt{n} = (\bar{X} - 200) \cdot \sqrt{n}$$

Ist  $H_0$  korrekt, so ist  $Z \sim N(0, 1)$ .

### Entscheidungsgrundlage:

In diesem Fall haben wir einen einseitigen Test. Bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha$  kommen wir daher zu einer Ablehnung der Nullhypothese, wenn für die Realisation  $z$  von  $Z$  gilt:

$$z < -z_{1-\alpha}$$

wobei wieder  $z_\beta$  das  $\beta$ -Quantil der Standardnormalverteilung bezeichnet.

Speziell gilt: Bei einer zugelassenen Fehlerwahrscheinlichkeit von 5 % kommen wir zu einer Ablehnung der Nullhypothese, wenn für die Realisation  $z$  von  $Z$  gilt:

$$z < -1.65$$

und bei einer zugelassenen Fehlerwahrscheinlichkeit von 1 % kommen wir zu einer Ablehnung der Nullhypothese, wenn für die Realisation  $z$  von  $Z$  gilt: Betrachte

$$z < -2.33$$

### Testdurchführung:

Angenommen bei einer konkreten Zufallsstichprobe vom Umfang  $n = 16$  erhalten wir als Testreihe (alle Angaben in Gramm)

199.20	198.25	198.60	200.80	199.10	200.70	201.05	199.10
198.95	200.50	198.30	199.25	200.50	200.30	198.40	199.80

In diesem Fall erhalten wir

$$\bar{x} = 199.55(\text{g})$$

und damit als Realisation der Testgröße

$$z = (199.55 - 200.00) \cdot \sqrt{16} = -1.80$$

Daher kann die Nullhypothese bei einer zugelassenen Fehlerwahrscheinlichkeit von 5 % abgelehnt werden, da  $z = -1.80 < -1.65$ .

Bei einer zugelassenen Fehlerwahrscheinlichkeit von 1 % kann  $H_0$  jedoch nicht abgelehnt werden, da  $z = -1.80 \not< -2.33$ .

Angenommen bei einer Zufallsstichprobe vom Umfang  $n = 100$  erhalten wir ein Durchschnittsinhaltsgewicht der getesteten Pralinenschachteln von  $\bar{x} = 199.73$  g. Dann ist

$$z = (199.73 - 200.00) \cdot \sqrt{100} = -2.70$$

Da  $z < -1.96$ , kann  $H_0$  bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha = 0.05$  abgelehnt werden. Da sogar  $z < -2.58$  kann  $H_0$  auch bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha = 0.01$  abgelehnt werden.

**Bemerkung 7.1.** Ein **statistisches Testproblem** besteht aus einer Nullhypothese  $H_0$  und einer Alternative  $H_1$ , die sich gegenseitig ausschließen und Aussagen über die gesamte Verteilung oder über bestimmte Parameter eines Merkmals treffen.

Das Testproblem heißt **zweiseitig**, wenn

$$\begin{aligned} H_0 : \text{Merkmal} &= \mu_0 \\ H_1 : \text{Merkmal} &\neq \mu_0 \end{aligned}$$

und es heißt **einseitig**, wenn

$$\begin{aligned} H_0 : \text{Merkmal} &\leq \mu_0 \\ H_1 : \text{Merkmal} &> \mu_0 \end{aligned}$$

oder

$$\begin{aligned} H_0 : \text{Merkmal} &\geq \mu_0 \\ H_1 : \text{Merkmal} &< \mu_0 \end{aligned}$$

Ein statistischer Test heißt **Test zum Signifikanzniveau**  $1 - \alpha$  ( $0 < \alpha < 1$ ), wenn die Wahrscheinlichkeit dafür, dass  $H_0$  abgelehnt wird, obwohl  $H_0$  wahr ist, höchstens  $\alpha$  ist,

$$p(H_0 \text{ abgelehnt} \mid H_0 \text{ wahr}) \leq \alpha$$

Wichtig bei statistischen Entscheidungen ist eine Analyse der Fehlersituation, also des Falles, dass man sich falsch entscheidet. Beim statistischen Testproblem gibt es hier zwei mögliche Fehlentscheidungen:

**Definition 7.1.** Bei einem statistischen Testproblem  $H_0$  gegen  $H_1$  und einem dazugehörigen statistischen Test sprechen wir von einem

- **Fehler erster Art** oder  $\alpha$ -Fehler, wenn  $H_0$  verworfen wird, obwohl  $H_0$  wahr ist.
- **Fehler zweiter Art** oder  $\beta$ -Fehler, wenn  $H_0$  nicht verworfen wird, obwohl  $H_0$  falsch ist.

**Bemerkung 7.2.** Die maximale Fehlerwahrscheinlichkeit bei der Testkonstruktion eines Tests ist eine Schranke für die Wahrscheinlichkeit eines Fehler erster Art.

Schwieriger anzugehen ist der Fehler zweiter Art. Wir wollen das anhand eines Gaußtests

$$\begin{array}{ll} \text{Nullhypothese} & H_0 : E(X) \leq \mu_0 \\ \text{Alternativhypothese} & H_1 : E(X) > \mu_0 \end{array}$$

untersuchen und den Fehler zweiter Art wie folgt präzisieren:

**Definition 7.2.** Die **Gütefunktion**  $g$  des Gaußtests ist definiert als

$$g(\mu) = p(H_0 \text{ abgelehnt} \mid E(X) = \mu)$$

**Bemerkung 7.3.** Die Gütefunktion  $g(\mu)$  beschreibt also die Wahrscheinlichkeit dafür, dass  $H_0$  (richtigerweise) abgelehnt wird, wenn  $H_1$  wahr ist mit  $E(X) = \mu$ . Entsprechend beschreibt  $1 - g(\mu)$  die Wahrscheinlichkeit dafür, dass  $H_0$  in diesem Fall fälschlicherweise nicht abgelehnt wird.

Führen wir einen Gaußtest mit der Prüfgröße

$$Z = Z_n = \frac{\bar{X} - \mu_0}{\sigma} \cdot \sqrt{n}$$

durch, so ergibt sich bei unserem einseitigen Testproblem

Nullhypothese  $H_0 : E(X) \leq \mu_0$  gegen Alternativhypothese  $H_1 : E(X) > \mu_0$   
für die Gütefunktion

$$\begin{aligned} g(\mu) &= p\left(\frac{\bar{X} - \mu_0}{\sigma} \cdot \sqrt{n} > z_{1-\alpha} \mid E(X) = \mu\right) \\ &= p\left(\frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} + \frac{\mu - \mu_0}{\sigma} \cdot \sqrt{n} > z_{1-\alpha} \mid E(X) = \mu\right) \\ &= p\left(\frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} > z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \cdot \sqrt{n} \mid E(X) = \mu\right) \\ &= 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \cdot \sqrt{n}\right) \end{aligned}$$

Entsprechend erhalten wir für

$$\begin{array}{ll} \text{Nullhypothese} & H_0 : E(X) \geq \mu_0 \\ \text{Alternativhypothese} & H_1 : E(X) < \mu_0 \end{array}$$

als Gütefunktion

$$g(\mu) = \Phi \left( z_\alpha - \frac{\mu - \mu_0}{\sigma} \cdot \sqrt{n} \right)$$

und für das Testproblem

$$\text{Nullhypothese} \quad H_0 : E(X) = \mu_0$$

$$\text{Alternativhypothese} \quad H_1 : E(X) \neq \mu_0$$

als Gütefunktion

$$g(\mu) = \Phi \left( -z_{1-\frac{\alpha}{2}} + \frac{\mu - \mu_0}{\sigma} \cdot \sqrt{n} \right) + \Phi \left( -z_{1-\frac{\alpha}{2}} - \frac{\mu - \mu_0}{\sigma} \cdot \sqrt{n} \right)$$

**Bemerkung 7.4.** Entscheidungen aufgrund statistischer Tests sind immer mit gewissen Fehlerwahrscheinlichkeiten behaftet. Bei der Konstruktion versucht man üblicherweise, dafür zu sorgen, dass ein gewisses Signifikanzniveau erreicht wird, dass also der Fehler erster Art gering wird. Das führt zu einer Schranke, für die Wahrscheinlichkeit, die Nullhypothese abzulehnen, obwohl sie korrekt ist. Aber auch in diesem Fall ist keine absolute Sicherheit gegeben. Entgegen gängiger Interpretation bedeutet eine sehr niedrige Wahrscheinlichkeit nicht, dass ein Ereignis nicht eintritt und „es schon gut gehen wird“. Ganz im Gegenteil, ist die Wahrscheinlichkeit für ein (immer wieder mögliches) Ereignis positiv, so besagt die Wahrscheinlichkeitstheorie, dass dieses Ereignis irgendwann auch eintreten wird. In der Wahrscheinlichkeitstheorie gibt es hierzu sogar den *Satz von der Glückssträhne* der (etwas vereinfacht) besagt, dass ein Ereignis, das mit positiver Wahrscheinlichkeit eintreten kann (und sei sie auch noch so klein), irgendwann einmal mit Sicherheit (also mit Wahrscheinlichkeit 1) beliebig oft hintereinander eintreten wird. Beispielsweise sagt dieser Satz, dass die Zahlen 1, 2, 3, 4, 5, 6 nicht nur irgendwann als Lottozahlen gezogen werden, er besagt sogar, dass es irgendwann ein Jahr geben wird, in dem bei jeder Ziehung die Zahlen 1, 2, 3, 4, 5, 6 gezogen werden. Dieses Jahr kann in ferner oder fernster Zukunft liegen, es kann aber auch schon das Jahr 2021 sein, hierüber kann keine Aussage gemacht werden. Trotzdem sind diese Situationen mit der Wahrscheinlichkeitstheorie kompatibel, und es kann daraus nicht mit Sicherheit auf eine Manipulation des Ziehungsgeräts geschlossen werden.

DHBW Mannheim  
Studiengang Informatik

# Statistik

## Vorlesung 11

Stand: Mai 2020

Autor:

Reinhold Hübl, Fakultät für Technik, DHBW Mannheim

e-mail: [reinhold.huebl@dhbw-mannheim.de](mailto:reinhold.huebl@dhbw-mannheim.de)

©Fakultät für Technik DHBW Mannheim

## 8 Lage– und Verteilungsalternativen

In diesem Fall geht es um die Verteilung eines Untersuchungsmerkmals, z.B. das Haushaltseinkommen in einem Wohngebiet. Ausgehend von einer einfachen Stichprobe vom Umfang  $n$  sollen nun bestimmte Eigenschaften dieses Merkmals und Hypothesen über diese Eigenschaften statistisch untersucht werden. Dabei kann es sich um einzelne Parameter dieser Eigenschaft handeln, etwa

$$H_0 : \text{Das Durchschnittshaushaltseinkommen liegt bei } 2760 \text{ } N$$

oder um Eigenschaften der gesamten Verteilung des Merkmals, etwa

$$H_0 : \text{Das Haushaltseinkommen ist normalverteilt mit } \mu = 2760 \text{ } N$$

Im ersten Fall sprechen wir von einem Lageparameter-test. In dem angesprochenen Beispiel formuliert sich das Testproblem wie beim Gaußtest im vorangegangenen Abschnitt: Getestet werden sollen Hypothesen zum Erwartungswert  $\mu = E(X)$  einer Zufallsvariable  $X$ , also etwa

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Beim Gaußtest wurde angenommen, dass  $X$  normalverteilt mit bekannter Varianz  $\sigma^2$  ist. In diesem Fall ist bei einer Stichprobe vom Umfang  $n$  die standardnormalverteilte Schätzvariable

$$Z = \frac{\bar{X} - \mu_0}{\sigma} \cdot \sqrt{n}$$

die geeignete Größe. Die Normalverteilungsannahme kann in der Praxis in der Regel gemacht werden, jedoch ist die Varianz häufig nicht bekannt und muss ebenfalls geschätzt werden. Daher ist die korrekte Schätzvariable in diesem Fall der sogenannte  $t$ -Test

$$T = \frac{\bar{X} - \mu_0}{S} \cdot \sqrt{n}$$

wobei

$$S^2 = \frac{1}{n-1} \cdot \sum_{k=1}^n (X_k - \bar{X})^2, \quad S = +\sqrt{S^2}$$

Der Schätzer  $T$  unterscheidet sich also von  $Z$  nur im Nenner.

Die Anforderungen, die wir an einen Schätzer stellen, sind vor allem, dass

1. er sensibel für das Testproblem ist.
2. für ihn die Verteilung unter  $H_0$  bekannt ist.

Beides ist in diesem Fall gegeben, (1), da  $T$  eine normierte Version von  $\bar{X}$  ist, also auf Änderungen der Erwartungswertes reagiert, und (2), da  $T$  unter  $H_0$  eine  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden besitzt.

Neben dem oben beschriebenen zweiseitigen Testproblem lassen sich wie im vorangegangenen Abschnitt auch hier einseitige Testprobleme formulieren. Wir betrachten dazu unabhängige, identische verteilte Wiederholungen  $X_1, \dots, X_n, \dots$  einer Zufallsvariable  $X$ , wobei wir annehmen, dass entweder  $X \sim N(\mu, \sigma^2)$  (mit unbekanntem  $\sigma$ ) oder  $n > 30$  (mit beliebig verteiltem  $X$ ) und wir erhalten

**Regel 8.1** (Der  $t$ -Test im Ein-Stichprobenfall). *Für das Testproblem*

1. Nullhypothese  $H_0 : E(X) = \mu_0$  gegen Alternativhypothese  $H_1 : E(X) \neq \mu_0$ .
2. Nullhypothese  $H_0 : E(X) = \mu_0$  gegen Alternativhypothese  $H_1 : E(X) < \mu_0$ .
3. Nullhypothese  $H_0 : E(X) = \mu_0$  gegen Alternativhypothese  $H_1 : E(X) > \mu_0$ .

definieren wir die Schätzstatistik

$$T = T_n = \frac{\bar{X} - \mu_0}{S} \cdot \sqrt{n}$$

Ist  $X$  normalverteilt mit Erwartungswert  $\mu_0$  und Varianz  $\sigma^2$ ,  $X \sim N(\mu_0, \sigma^2)$ , so ist  $T$  verteilt wie  $t(n - 1)$ , im allgemeinen Fall ist  $T$  (unter Annahme der Gültigkeit der Nullhypothese) nach dem Gesetz der große Zahl zumindest approximativ standardnormalverteilt,

$$T \stackrel{a}{\sim} N(0, 1)$$

Zu gegebener Realisation  $t$  und einem vorgegebenen Signifikanzniveau  $1 - \alpha$  fällt die Entscheidung gegen  $H_0$  im Testproblem,

1. falls  $|t| > t_{1-\frac{\alpha}{2}}(n - 1)$ .
2. falls  $t < -t_{1-\alpha}(n - 1) = t_{\alpha}(n - 1)$ .
3. falls  $t > t_{1-\alpha}(n - 1)$ .

Für  $n > 30$  können die  $\alpha$ -Quantile der  $t(n - 1)$ -Verteilung durch die entsprechenden Quantile der Standardnormalverteilung ersetzt werden.

**Beispiel 8.1.** Wir greifen noch einmal Beispiel 7.7 bzw. Beispiel 7.9 auf und nehmen jetzt an, dass der Hersteller zwar angibt, dass das Gewicht des Werkstückes normalverteilt ist, aber keinen Wert für die Streuung bereitstellt. In diesem Fall kann das Testszenario wie folgt beschrieben werden:

**Testszenario:**

Wir bezeichnen mit  $X$  die Zufallsvariable *Gewicht eines Werkstücks* und führen einen Hypothesentest

- Nullhypothese  $H_0 : E(X) = 100.00 \text{ g}$
- Alternativhypothese  $H_1 : E(X) \neq 100.00 \text{ g}$ .

durch. Bei einer zu untersuchenden Zufallsstichprobe vom Umfang  $n$  gehen wir vor wie folgt:  $X_1, \dots, X_n$  sind stochastisch unabhängige Wiederholungen von  $X$  (das entspricht der Zufallsstichprobe), und

$$\bar{X} = \frac{1}{n} \cdot \sum_{k=1}^n X_k$$

bezeichnet das arithmetische Mittel von  $X_1, \dots, X_n$  und

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

den korrigierten Schätzer für die Varianz. Die Testgröße hat die Gestalt

$$T = \frac{\bar{X} - 100}{\sqrt{S^2}} \cdot \sqrt{n}$$

Ist  $H_0$  korrekt, so ist  $T \sim t(n-1)$ .

**Entscheidungsgrundlage:**

Bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha$  kommen wir zu einer Ablehnung der Nullhypothese, wenn für die Realisation  $t$  von  $T$  gilt:

$$t < -t(n-1)_{1-\frac{\alpha}{2}} \quad \text{oder} \quad t > t(n-1)_{1-\frac{\alpha}{2}}$$

wobei  $t(n-1)_\beta$  das  $\beta$ -Quantil der  $t$ -Verteilung mit  $n-1$  Freiheitsgraden bezeichnet.

**Testdurchführung:**

Angenommen bei einer konkreten Zufallsstichprobe vom Umfang  $n = 16$  erhalten wir als Testreihe (alle Angaben in Gramm) wieder wie oben die Werte

100.34	100.40	99.62	100.31	100.28	99.93	100.59	100.32
100.21	99.57	100.32	99.68	99.80	100.12	100.02	99.73



In diesem Fall erhalten wir

$$\bar{x} = 100.09 \text{ (g)}, \quad S^2 = 0.09425 \text{ (g}^2\text{)}$$

und damit als Realisation der Testgröße

$$t = \frac{100.09 - 100.00}{\sqrt{0.09425}} \cdot \sqrt{16} = 1.173$$

Da  $t(15)_{0.975} = 2.130$ , kann  $H_0$  in dieser Situation nicht abgelehnt werden.

Gauß- und  $t$ -Test arbeiten bei kleinem Stichprobenumfang nur dann gut, wenn das zugrundeliegende Merkmal  $X$  normalverteilt ist. Bei starken Abweichungen von dieser Annahme sind ihre Ergebnisse nicht sehr vertrauenswürdig. In diesem Fall empfehlen sich **verteilungsfreie Tests**. Hier stehen nicht Parameter der Verteilung (wie Erwartungswert oder Varianz oder  $\lambda$  in  $Po(\lambda)$ ) im Vordergrund sondern Charakteristika wie Median oder Quantile.

Das Mediantestproblem ist bestimmt durch das Hypothesenpaar

$$\text{Nullhypothese } H_0 : x_{\text{med}} = \delta_0 \text{ gegen Alternativhypothese } H_1 : x_{\text{med}} \neq \delta_0.$$

wobei  $x_{\text{med}}$  den Median des Merkmals  $X$  bezeichnet.

**Beispiel 8.2.** Eine Supermarktkette gibt in ihrer Werbung an, dass an ihren Kassen 50 % aller Kunden höchstens 2 Minuten anstehen. Falls das von einer Verbraucherorganisation kritisch überprüft werden soll, bietet sich hierfür der folgenden Hypothesentest an: Mit  $X$  bezeichnen wir die Zufallsvariable die die Wartezeit eines Kunden an der Kasse misst.

$$\text{Nullhypothese } H_0 : x_{\text{med}} = 2 \text{ min gegen}$$

$$\text{Alternativhypothese } H_1 : x_{\text{med}} > 2 \text{ min.}$$

(hier ist offensichtlich eine Unterschreitung von 2 Minuten nicht als Mangel anzusehen).

**Beispiel 8.3.** Bei einer Losbude ist jedes Los ein Treffer, wobei es Gewinne der Klassen 1 - 5 gibt, und wobei die Klasse 1 einen Hauptgewinn bezeichnet. Der Losverkäufer wirbt damit, dass die Hälfte der Lose Gewinnklasse 3 oder besser haben. Falls das kritisch überprüft werden soll, bietet sich hierfür der folgenden Hypothesentest an: Mit  $X$  bezeichnen wir die Zufallsvariable die die Gewinnklasse eines Loses misst.

$$\text{Nullhypothese } H_0 : x_{\text{med}} = 3 \text{ gegen}$$

Alternativhypothese  $H_1 : x_{\text{med}} > 3$ .

(hier ist offensichtlich eine höhere Anzahl an hochwertigen Gewinnen nicht als Mangel anzusehen).

Zum Test verwenden wir den sogenannten **Vorzeichentest**. Er beruht auf unabhängige Wiederholungen  $X_1, \dots, X_n, \dots$  von  $X$ . Daraus leiten wir neue Zufallsvariablen  $Y_k$  ab mit

$$Y_k(\omega) = \begin{cases} 1 & \text{falls } X_k(\omega) \leq \delta_0 \\ 0 & \text{falls } X_k(\omega) > \delta_0 \end{cases}$$

Die Prüfgröße, die wir betrachten, ist

$$A = A_n = Y_1 + \dots + Y_n$$

d.h. wir zählen die Wiederholungen, bei denen wir unter  $\delta_0$  bleiben. Ist  $X$  stetig verteilt und  $\delta_0$  in der Tat der Median, so ist  $A$  eine Bernoullivariablen mit

$$A \sim B(n, 0.5)$$

Die Nullhypothese wird abgelehnt, wenn  $A$  zu stark von dieser Verteilung abweicht. Der zugehörige Signifikanztest wird nun aus der Bernoulliverteilung abgeleitet. Dabei bezeichnet  $b_\gamma$  den größte Wert, für den die  $B(n, 0.5)$ -Verteilung den Wert  $\gamma$  nicht überschreitet.

### Regel 8.2. (*Vorzeichentest*)

*Für das Testproblem*

1. Nullhypothese  $H_0 : x_{\text{med}} = \delta_0$  gegen Alternativhypothese  $H_1 : x_{\text{med}} \neq \delta_0$ .
2. Nullhypothese  $H_0 : x_{\text{med}} \geq \delta_0$  gegen Alternativhypothese  $H_1 : x_{\text{med}} < \delta_0$ .
3. Nullhypothese  $H_0 : x_{\text{med}} \leq \delta_0$  gegen Alternativhypothese  $H_1 : x_{\text{med}} > \delta_0$ .

*definieren wir die Schätzstatistik*

$$A = A_n = Y_1 + \dots + Y_n$$

*Zu gegebener Realisation  $a$  und einem vorgegebenen Signifikanzniveau  $\alpha$  fällt die Entscheidung gegen  $H_0$  im Testproblem*

1. , falls  $a \leq b_{\frac{\alpha}{2}}$  oder  $n - a \leq b_{\frac{\alpha}{2}}$ .

2. , falls  $n - a \leq b_\alpha$ .

3. , falls  $a \leq b_\alpha$ .

**Beispiel 8.4.** Wir greifen noch einmal Beispiel 8.2 auf, bezeichnen wieder mit  $X$  die Zufallsvariable, die die Wartezeit eines Kunden an der Kasse misst und betrachten den Test

Nullhypothese  $H_0 : x_{\text{med}} = 2 \text{ min}$  gegen

Alternativhypothese  $H_1 : x_{\text{med}} > 2 \text{ min}$ .

Ferner betrachten wir 16 unabhängige Wiederholungen  $X_1, \dots, X_{16}$  von  $X$  (also eine Stichprobe vom Umfang  $n = 16$ ) und bezeichnen für  $k = 1, \dots, 16$  mit  $Y_k$  die Zufallsvariable

$$Y_k(\omega) = \begin{cases} 1 & \text{falls } X_k(\omega) \leq 2 \text{ min} \\ 0 & \text{falls } X_k(\omega) > 2 \text{ min} \end{cases}$$

Die Prüfgröße, die wir betrachten, ist

$$A = Y_1 + \dots + Y_{16}$$

Bei einer zugelassenen Fehlerwahrscheinlichkeit von  $\alpha = 0.05$  lehnen wir  $H_0$  ab, wenn für die Realisation  $a$  von  $A$  gilt:

$$a \leq 4$$

denn für eine binomialsverteilte Zufallsvariable  $B \sim B(16, 0.50)$  gilt

$$p(B \leq 4) = 0.038, \quad p(B \leq 5) = 0.105$$

**Beispiel 8.5.** Wir greifen auch noch einmal Beispiel 8.3 mit der Losbude auf, bezeichnen wieder mit  $X$  die Zufallsvariable, die die Gewinnklasse eines Loses notiert und betrachten den Test

Nullhypothese  $H_0 : x_{\text{med}} = 3$  gegen

Alternativhypothese  $H_1 : x_{\text{med}} > 3$ .

Ferner betrachten wir 100 unabhängige Wiederholungen  $X_1, \dots, X_{100}$  von  $X$  (also eine Stichprobe vom Umfang  $n = 100$ ) und bezeichnen für  $k = 1, \dots, 100$  mit  $Y_k$  die Zufallsvariable

$$Y_k(\omega) = \begin{cases} 1 & \text{falls } X_k(\omega) \leq 3 \\ 0 & \text{falls } X_k(\omega) > 3 \end{cases}$$

Da  $100 \cdot 0.5 \cdot 0.5 = 25 > 9$ , wenden wir hier einen approximativen Binomialtest an, dh. wir definieren zunächst die Größe

$$A = Y_1 + \dots + Y_{100}$$

setzen  $\bar{A} = \frac{1}{100} \cdot A$  und betrachten die Prüfgröße

$$Z = \frac{\bar{A} - 0.5}{\sqrt{0.5 \cdot 0.5}} \cdot \sqrt{100} = (\bar{A} - 0.5) \cdot 20$$

Bei einer zugelassenen Irrtumswahrscheinlichkeit von  $\alpha = 0.05$  lehnen wir die Nullhypothese ab, falls für die Realisation  $z$  von  $Z$  gilt

$$z < -1.65$$

Die Testdurchführung ergibt nun folgende Ergebnisse:

Gewinnklasse	1	2	3	4	5
Anzahl der Lose	3	8	31	26	32

Damit erhalten wir  $a = 42$ , also  $\bar{a} = 0.42$  und daher

$$z = (0.42 - 0.50) \cdot 20 = -1.60$$

Die Nullhypothese kann also nicht abgelehnt werden.

Oft ist es nicht der Median, der von Interesse ist sondern ein anderes Quantil einer Verteilung, wenn etwa untersucht werden soll, ob für einen bestimmten Prozentanteil der Gesamtheit ein vorgegebenes Qualitätsmerkmal erreicht oder überschritten wird. Hierzu kann der Vorzeichentest zu einem Test für  $p$ -Quantile  $x_p$  modifiziert werden:

1. Nullhypothese  $H_0 : x_p = \delta_0$  gegen Alternativhypothese  $H_1 : x_p \neq \delta_0$ .
2. Nullhypothese  $H_0 : x_p = \delta_0$  gegen Alternativhypothese  $H_1 : x_p < \delta_0$ .
3. Nullhypothese  $H_0 : x_p = \delta_0$  gegen Alternativhypothese  $H_1 : x_p > \delta_0$ .

Setze

$$Y_k(\omega) = \begin{cases} 1 & \text{falls } X_k(\omega) \leq \delta_0 \\ 0 & \text{falls } X_k(\omega) > \delta_0 \end{cases}$$

und betrachte die Testgröße

$$A = A_n = Y_1 + \dots + Y_n$$

**Regel 8.3** (modifizierter Vorzeichentest). Wir bezeichnen mit  $b_\gamma$  den größte Wert, für den die  $B(n, p)$ -Verteilung den Wert  $\gamma$  nicht überschreitet.

Zu gegebener Realisation  $a$  von  $A$  und einem vorgegebenen Signifikanzniveau  $1 - \alpha$  fällt die Entscheidung gegen  $H_0$  im Testproblem,

1. falls  $a \leq b_{\frac{\alpha}{2}}$  oder  $a > b_{1-\frac{\alpha}{2}}$ .
2. falls  $a > b_{1-\alpha}$ .
3. falls  $a \leq b_\alpha$ .

Auch hier wird mit einem approximativen Binomialtest gearbeitet, wenn  $n$  hinreichend groß ist.

**Beispiel 8.6.** Eine Arbeitsvermittlung wirbt damit, dass 75% der Kunden innerhalb von 60 Tagen vermittelt werden. Sie haben Ihre Zweifel und wollen die Behauptung überprüfen.

- a) Das Verfahren, das hier zur Anwendung kommt, ist der modifizierte Vorzeichentest, also ein Vorzeichentest für das 0.75-Quantil  $x_{0.75}$  der Zufallsvariable  $X$ , die die Zeitdauer misst, bis ein Kunde vermittelt wurde (dh. ein Binomialtest mit der Bernoullivariablen  $Y$ , die misst, ob ein Kunde in der angegebenen Zeit vermittelt werden konnte oder nicht). Da sie die Aussage der Arbeitsvermittlung anzweifeln, formulieren Sie folgenden Hypothesentest

**Nullhypothese**  $H_0: x_{0.75} = 60$  Tage.

**Alternativhypothese**  $H_1: x_{0.75} > 60$  Tage.

Dazu betrachten wir die Zufallsvariablen  $X_k$ , die die Vermittlungszeit des Kunden  $k$  (in Tagen) beschreibt, setzen

$$Y_k(\omega) = \begin{cases} 1 & \text{falls } X_k(\omega) \leq 60 \text{ min} \\ 0 & \text{falls } X_k(\omega) > 60 \text{ min} \end{cases}$$

und benutzen die Testgröße

$$A = Y_1 + \dots + Y_n$$

Dann ist ganz allgemein  $A \sim B(n, p)$ , wobei  $p$  die (tatsächliche) Wahrscheinlichkeit dafür ist, dass ein Kunde innerhalb von 60 Tagen vermittelt wird.

Unter der Nullhypothese gilt  $A \sim B(n, 0.75)$ , und, da unsere Befürchtung ist, dass weniger als 75 % aller Kunden innerhalb von 60 Tagen vermittelt werden (dass also  $p < 0.75$ ), führen wir einen Binomialtest durch für

**Nullhypothese**  $H_0: p = 0.75$ .

**Alternativhypothese**  $H_1: p < 0.75$ .

Bei gegebenen Irrtumswahrscheinlichkeit  $\alpha$  lehnen wir die Nullhypothese ab, wenn (bei gegebenen Realisation  $a$  von  $A$ )  $a \leq b = b_\alpha$ , wobei (wie üblich)  $b = b_\alpha$  das entsprechende  $\alpha$ -Quantil der  $B(n, 0.75)$ -Verteilung, also wenn  $b$  der größte Wert mit

$$p(B \leq b) \leq \alpha$$

ist, wobei  $B \sim B(n, 0.75)$ .

- b) Angenommen, eine Zufallsbeobachtung von 15 Kunden dieses Unternehmens liefert die folgenden Vermittlungszeiten (in Tagen):

Kunde	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Zeit	34	42	17	94	122	32	–	79	42	94	57	46	52	26	84

Dabei steht – für einen Kandidaten, der nicht (innerhalb von 60 Tagen) vermittelt werden konnte.

Der Entscheidung wird ein Stichprobenumfang von  $n = 15$  zugrundegelegt. Für eine  $B(15, 0.75)$ -verteilte Zufallsvariable  $B$  gilt

$$\begin{aligned} p(B \leq 6) &= 0.0042, & p(B \leq 7) &= 0.0173, \\ p(B \leq 8) &= 0.0566, & p(B \leq 9) &= 0.1484 \end{aligned}$$

und daher wird die Nullhypothese abgelehnt, wenn die Realisation  $a$  von  $A$  einen Wert kleiner oder gleich 7 liefert.

In der untersuchten Stichprobe Fall gilt  $a = 9$ , und daher kann die Nullhypothese nicht abgelehnt werden.

- c) Der Test wird wiederholt, allerdings diesmal mit einem Stichprobenumfang von  $n = 150$ . In diesem Fall gilt

$$150 \cdot 0.25 \cdot 0.75 = 28.125 > 9$$

und daher kann für die Entscheidung ein approximativer Binomialtest herangezogen werden. Die Hypothesen bleiben wie in Teil a), ebenso die Zufallsvariablen  $Y, Y_1, \dots, Y_n$  (hier mit  $n = 150$ ) und  $A = Y_1 + \dots + Y_{150}$ . Ferner setzen wir

$$\bar{A} = \frac{1}{500} \cdot A, \quad Z = \frac{\bar{A} - 0.75}{\sqrt{0.25 \cdot 0.75}} \cdot \sqrt{150} = 20 \cdot (\bar{A} - 0.75) \cdot \sqrt{2}$$

Ist  $H_0$  richtig, so ist  $Z \stackrel{a}{\sim} N(0, 1)$ , und wir lehnen  $H_0$  bei gegebener maximaler Irrtumswahrscheinlichkeit  $\alpha$  ab, wenn für die Realisation  $z$  von  $Z$  gilt

$$z < -z_{1-\alpha}$$

(wobei  $z_{1-\alpha}$  das  $(1 - \alpha)$ -Quantil der Standardnormalverteilung ist).

- d) Von den 150 zufällig ausgewählten befragten Kunden konnten exakt 90 innerhalb von 60 Tagen vermittelt werden. Dann gilt

$$\bar{a} = \frac{90}{150} = 0.60$$

und damit

$$z = 20 \cdot (0.60 - 0.75) \cdot \sqrt{2} = -4.24$$

Damit kann die Nullhypothese in diesem Fall bei  $\alpha = 0.05$  abgelehnt werden, und sogar bei  $\alpha = 0.01$  reicht die Datenlage noch für eine Ablehnung, denn

$$z < -1.65 = -z_{0.95} \quad \text{und} \quad z < -2.33 = -z_{0.99}$$

**Beispiel 8.7.** Eine Autoreparaturwerkstatt wirbt damit, dass mindestens 60 % aller Kunden die Qualität der Arbeiten mit der Bestnote 5 aus 5 Punkten bewerten. Im Auftrag einer Verbraucherschutzorganisation überprüfen Sie diese Aussage und führen dazu eine Test mit 12 zufällig ausgewählten Serviceaufträgen durch.

- a) Als Testansatz empfiehlt sich ein modifizierter Vorzeichentest. Allerdings müssen wir hierzu die Fragestellung erst umformulieren. Eine Möglichkeit besteht darin, die Aussage der Werkstatt dahingehend zu interpretieren, dass höchstens 40 % der Kunden die Arbeit der Werkstatt mit der Zufriedenheitsnote 1, ..., 4 bewerten und dann einen Test auf das 0.4-Quantil dieser Bewertungsskala durchzuführen, wobei wir den Test wie folgt aufsetzen müssen:

**Nullhypothese**  $H_0: x_{0.40} = 5$ .

**Alternativhypothese**  $H_1: x_{0.40} \leq 4$ .

Alternativ bietet sich eine „Umskalierung“ der Bewertungsskala wie folgt an: Bezeichnen wir mit  $\tilde{X}$  die Zufallsvariable *Zufriedenheitsbewertung durch den Kunden*, so können wir daraus die Zufallsvariable

$$X = 6 - \tilde{X}$$

ableiten, die auch Werte von 1 bis 5 annimmt, aber jetzt die Bestbewertung bei 1 hat. Die Behauptung der Reperaturwerkstatt übersetzt sich also in die Aussage, dass (bei dieser Bewertungsskala) mindestens 60 der Kunden höchstens den Wert 1 angegeben, dh. das Testszenario ist dann

**Nullhypothese**  $H_0: x_{0.60} = 1$ .

**Alternativhypothese**  $H_1: x_{0.60} > 1$ .

Da  $n = 12$  und  $12 \cdot 0.6 \cdot 0.4 = 2.88 \leq 9$ , müssen wir einen exakten Binomialtest durchführen.

Dazu bezeichnen wir mit  $X_1, \dots, X_{12}$  unabhängige Wiederholungen der Zufallsvariable  $X$  (*modifizierte Kundenzufriedenheit*) und leiten daraus neue Zufallsvariablen  $Y_k$  ab mit

$$Y_k(\omega) = \begin{cases} 1 & \text{falls } X_k(\omega) = 1 \\ 0 & \text{falls } X_k(\omega) = 5 \end{cases}$$

Die Prüfgröße, die wir betrachten, ist

$$A = A_{12} = Y_1 + \dots + Y_{12}$$

d.h. wir zählen die Kunden, die dem Service die Bestnote gegeben haben.

Unter der Nullhypothese gilt:  $A \sim B(12, 0.40)$ .

Bezeichnet wir bei zugelassener Fehlerwahrscheinlichkeit  $\alpha$  mit  $k$  den größten Wert, für den für eine  $B \sim B(12, 0.60)$ -verteilte Zufallsvariable gilt, dass  $p(B \leq k) \leq \alpha$  liegt, so lehnen wir die Nullhypothese ab, falls für die Realisation  $a$  von  $A$  gilt:

$$a \leq k$$

Aus den entsprechenden Tabellen der Binomialverteilung erhalten wir

$$p(B \leq 3) = 0.0153, \quad p(B \leq 8) = 0.0573$$

und damit  $k = 3$ , dh. wir lehnen  $H_0$  wenn höchstens 3 Kunden den Service mit der Bestnote bewerten.

Alternativ können wir auch wie folgt vorgehen: Wir bezeichnen mit  $X$  die Zufallsvariable *Kundenzufriedenheit* (mit den Werten 1 bis 5) und wir führen eine neue Zufallsvariable  $Y$  ein durch

$$Y(\omega) = \begin{cases} 1 & \text{falls } X(\omega) = 5 \\ 0 & \text{falls } X(\omega) < 5 \end{cases}$$



Dann ist  $Y$  eine Bernoulli-Variable mit Parameter  $p$ , wobei  $p$  die tatsächliche Wahrscheinlichkeit dafür ist, dass eine Kunde die Arbeit mit der Bestbewertung beurteilt.

Für diese Bernoulli-Variable führen wir einen Binomilatest durch, und zwar in der Form

**Nullhypothese**  $H_0: p = 0.60$ .

**Alternativhypothese**  $H_1: p < 0.60$ .

Da  $n = 12$  und  $12 \cdot 0.6 \cdot 0.4 = 2.88 \leq 9$ , müssen wir einen exakten Binomialtest durchführen.

Für die Testdurchführung betrachten wir  $n$  unabhängige Wiederholungen (hier mit  $n = 12$ ) von  $Y$  (d.h. wir nehmen eine Zufallsstichprobe von  $n$  Kunden und untersuchen bei jedem, ob er die Bestbewertung gegeben hat oder nicht).

Die Prüfgröße, die wir betrachten, ist

$$A = A_n = Y_1 + \cdots + Y_n$$

(hier mit  $n = 12$ ), dh. wir zählen die Kunden, die mit dem Service zufrieden waren.

Unter der Nullhypothese gilt:  $A \sim B(12, 0.60)$ .

Bezeichnen wir bei zugelassener Fehlerwahrscheinlichkeit  $\alpha$  mit  $b_\alpha$  den höchsten Wert, für den eine  $B \sim B(12, 0.60)$ -verteilte Zufallsvariable unterhalb  $\alpha$  liegt, so lehnen wir die Nullhypothese ab, falls für die Realisation  $a$  von  $A$  gilt:

$$a \leq b_\alpha$$

Im Fall  $n = 12$  und  $\alpha = 5\%$  ist  $b_\alpha = 3$ , ( $p(B \leq 3) = 1.25\%$  und  $p(B \leq 4) = 5.73\%$ ), wir können also  $H_0$  auch hier nur dann ablehnen, wenn höchstens 3 Kunden die Bestnote vergeben.

- b) Wir nehmen an, eine Untersuchung von 12 Serviceaufträgen liefert folgende Bewertungen:

Kunde	1	2	3	4	5	6	7	8	9	10	11	12
Bewertung	3	3	5	4	5	5	3	4	2	5	4	4

Wir haben also 4 Kunden, die die Höchstnote 5 vergeben haben und können daher die Nullhypothese nicht mehr ablehnen.

- c) Der Test soll nun mit der Untersuchung von  $n = 250$  zufällig ausgewählten Serviceaufträgen wiederholt werden.

In diesem Fall führen wir den Binomialtest nicht als exakten Binomialtest sondern in Form eines approximativen Binomialtests durch, da  $250 \cdot 0.60 \cdot 0.40 = 60 > 9$ . Nullhypothese und Alternativhypothese bleiben wie in Teil a), ebenso die Zufallsvariablen bzw.  $X_k, Y_k$  für  $k = 1, \dots, n$  (jetzt mit  $n = 250$ ) und  $A$ , (bzw.  $Y_k$  und  $A$ , wenn wir gleich einen Binomialtest durchführen). Darüberhinaus setzen wir

$$\bar{A} = \frac{A}{250}$$

d.h. wir betrachten das arithmetische Mittel zur Folge  $(Y_k)_{k=1, \dots, 250}$ .

Ferner betrachten wir die Teststatistik

$$Z = \frac{\bar{A} - 0.60}{\sqrt{\frac{6}{25}}} \cdot \sqrt{250} = \frac{5 \cdot \sqrt{250}}{\sqrt{6}} (\bar{A} - 0.60)$$

Unter Annahme der Gültigkeit der Nullhypothese ist  $Z$  approximativ standardnormalverteilt,

$$Z \stackrel{a}{\sim} N(0, 1)$$

Die Nullhypothese wird abgelehnt, wenn bei gegebenem Fehlerniveau  $\alpha$  für die Versuchsrealisation gilt:

$$z < -z_{1-\alpha}$$

Das Vorgehen ist dabei unabhängig davon, ob wir einen modifizierten Vorzeichen-test durchführen oder gleich zum Binomialtest übergegangen sind.

- d) Die Testdurchführung mit  $n = 250$  Serviceaufträgen ergibt, dass 136 Kunden die Leistung mit 5 Punkten bewertet haben. Dann gilt im Binomialtestansatz

$$\bar{a} = \frac{136}{250} = 0.544; \quad z = \frac{5 \cdot \sqrt{250}}{\sqrt{6}} (0.544 - 0.60) \approx -1.81$$

Wegen  $z_{0.95} = 1.65$  kann die Nullhypothese bei  $\alpha = 5\%$  abgelehnt werden, wegen  $z_{0.99} = 2.33$  kann sie bei  $\alpha = 1\%$  nicht mehr abgelehnt werden.

Bis jetzt haben wir uns nur mit der Lage der Verteilung, also mit entscheidenden Parametern wie Median oder Erwartungswert beschäftigt. Oft will man jedoch mehr wissen. Wenn etwa geprüft werden soll, ob ein Würfel fair ist (also jede Augenzahl gleichwahrscheinlich ist) oder allgemeiner eine diskrete Zufallsvariable  $X$  mit nur endlich vielen

möglichen Zuständen einem bestimmten Wahrscheinlichkeitsmuster genügt, reicht eine bloße Erwartungswertanalyse nicht aus. Gehen wir davon aus, dass die möglichen Werte von  $X$  durch die Zahlen  $1, \dots, l$  repräsentiert werden, so wollen wir also Nullhypothesen der Art

$$H_0 : p(X = k) = \pi_k, \quad (k = 1, \dots, l)$$

testen. Im Fall des Würfels also

$$H_0 : p(X = k) = \frac{1}{6}, \quad (k = 1, \dots, 6)$$

Die Alternativhypothese in diesem Fall ist die logische Verneinung von  $H_0$ , also hier

$$H_1 : p(X = k) \neq \frac{1}{6} \quad \text{für ein } k \in \{1, \dots, 6\}$$

Für eine einzelne Zelle  $X = k$  könnte die entsprechende Hypothese durch einen Binomialtest geprüft werden, für das simultane Testen aller Wahrscheinlichkeitswerte verwendet man jedoch als Schätzung den  $\chi^2$ -Test. Dazu betrachten wir wieder unabhängige Wiederholungen  $X_1, \dots, X_n, \dots$  der Zufallsvariable  $X$  und den daraus resultierenden i.i.d-Prozess  $\mathbb{X} = (X_n)_{n \in \mathbb{N}}$ . Daraus leiten wir neue Zufallsvariablen  $Y_{m,k}$  ab mit

$$Y_{m,k}(\omega) = \begin{cases} 1 & \text{falls } X_m(\omega) = k \\ 0 & \text{falls } X_m(\omega) \neq k \end{cases}$$

und wir setzen

$$H_k = H_{n,k} = Y_{1,k} + \dots + Y_{n,k}$$

d.h. wir zählen, wie oft der Wert  $k$  bei einer Stichprobe vom Umfang  $n$  auftritt.

**Regel 8.4** ( $\chi^2$ -Anpassungstest). *Für das Testproblem*

- Nullhypothese  $H_0 : p(X = k) = \pi_k, (k = 1, \dots, l)$  gegen
- Alternativhypothese  $H_1 : p(X = k) \neq \pi_k$  für ein  $k \in \{1, \dots, l\}$ .

definieren wir die Schätzstatistik

$$\chi^2 = \sum_{k=1}^l \frac{(H_{n,k} - n\pi_k)^2}{n\pi_k}$$

Unter der Annahme, dass  $H_0$  richtig ist, ist  $\chi^2$  approximativ  $\chi^2$ -verteilt mit  $l - 1$  Freiheitsgraden,

$$\chi^2 \stackrel{a}{\sim} \chi^2(l - 1)$$

Zu gegebener Realisation  $(h_{n,1}, \dots, h_{n,l})$  von  $H_{n,1}, \dots, H_{n,l}$  und einem vorgegebenen Signifikanzniveau  $\alpha$  fällt die Entscheidung gegen  $H_0$  im Testproblem wenn

$$\chi^2 > \chi_{1-\alpha}^2(l - 1)$$

**Bemerkung 8.1.** Die  $\chi^2$ -Verteilung kann für den Schätzer  $\chi^2$  zugrundegelegt werden, wenn

- $n \cdot \pi_k \geq 1$  für alle  $k \in \{1, \dots, l\}$

oder

- $n \cdot \pi_k \geq 5$  für mindestens 80% der Zellen.

**Beispiel 8.8.** Ein Würfel soll dahingehend untersucht werden, ob es sich dabei tatsächlich, wie behauptet wird, um einen Laplacewürfel handelt, ob also tatsächlich gilt, dass jede Zahl mit Wahrscheinlichkeit  $\frac{1}{6}$  fällt, und zwar mit einer maximal zugelassenen Fehlerwahrscheinlichkeit von  $\alpha = 0.10$  ( $\alpha = 0.05$  bzw.  $\alpha = 0.01$ ).

a) Für diese Untersuchung bietet sich der  $\chi^2$ -Anpassungstest an.

Dazu bezeichnen wir mit  $X$  die Zufallsvariable *Augenzahl beim Würfelwurf* und betrachten folgendes Testproblem

- Nullhypothese  $H_0 : p(X = k) = \frac{1}{6}, (k = 1, \dots, 6)$  gegen
- Alternativhypothese  $H_1 : p(X = k) \neq \frac{1}{6}$  für ein  $k \in \{1, \dots, 6\}$ .

Es ist also  $l = 6$ , und daher ist die  $\chi^2$ -Verteilung  $\chi^2(5)$  mit fünf Freiheitsgraden heranzuziehen. Die relevanten Quantile dieser Verteilung sind

$$\chi_{0.90}^2(5) = 9.24, \quad \chi_{0.95}^2(5) = 11.07, \quad \chi_{0.99}^2(5) = 15.09$$

Für die Durchführung betrachten wir  $n$  unabhängige Wiederholungen  $X_1, \dots, X_n$  von  $X$ . Die Testgrößen

$$Y_{m,k}(\omega) = \begin{cases} 1 & \text{falls } X_m(\omega) = k \\ 0 & \text{falls } X_m(\omega) \neq k \end{cases}$$

für  $k = 1, \dots, 6, m = 1, \dots, n$  und

$$H_k = H_{n,k} = Y_{1,k} + \dots + Y_{n,k}$$

für  $k = 1, \dots, 6$  und

$$\chi^2 = \sum_{k=1}^6 \frac{(H_{n,k} - n \cdot \frac{1}{6})^2}{n \cdot \frac{1}{6}}$$

setzen wir an wie oben.

Bei  $\alpha = 0.10$  wird  $H_0$  abgelehnt, wenn für die Realisation  $\chi^2$  der Testgröße gilt

$$\chi^2 > 9.24$$

- b) Eine Wurfserie von 30 unabhängigen Würfeln liefert folgende Ergebnisse

$k$	1	2	3	4	5	6
$h_k$	7	7	7	3	3	3

Dann erhalten wir als Realisation der Testgröße

$$\begin{aligned}
 \chi^2 &= \sum_{k=1}^6 \frac{(h_k - 5)^2}{5} \\
 &= \frac{2^2}{5} + \frac{2^2}{5} + \frac{2^2}{5} + \frac{(-2)^2}{5} + \frac{(-2)^2}{5} + \frac{(-2)^2}{5} \\
 &= 4.8
 \end{aligned}$$

Die Ergebnisse reichen also bei  $\alpha = 0.10$  nicht aus, um  $H_0$  abzulehnen (erst recht nicht bei  $\alpha = 0.05$  bzw.  $\alpha = 0.01$ ).

- c) Wir wiederholen den Test mit einer Wurfserie von 300 unabhängigen Würfeln, bei denen wir folgende Ergebnisse erhalten

$k$	1	2	3	4	5	6
$h_k$	60	60	60	40	40	40

Dann erhalten wir als Realisation der Testgröße

$$\begin{aligned}
 \chi^2 &= \sum_{k=1}^6 \frac{(h_k - 50)^2}{50} \\
 &= \frac{10^2}{50} + \frac{10^2}{50} + \frac{10^2}{50} + \frac{(-10)^2}{50} + \frac{(-10)^2}{50} + \frac{(-10)^2}{50} \\
 &= 12
 \end{aligned}$$

In dieser Situation können wir  $H_0$  bei  $\alpha = 0.10$  ablehnen, da  $\chi^2 > 9.24$  und auch noch bei  $\alpha = 0.05$ , da  $\chi^2 > 11.09$ , aber nicht mehr bei  $\alpha = 0.01$ , da  $\chi^2 \not> 15.09$ .

**Beispiel 8.9.** Es ist bekannt, dass die Verteilung der Blutgruppen in Deutschland wie folgt ist:

Blutgruppe	A	0	B	AB
Anteil	43%	41%	11%	5%

Sie vermuten, dass die Verteilung in den USA von der in Deutschland abweicht und wollen dazu einen Test mit einer maximalen Fehlerwahrscheinlichkeit von  $\alpha = 0.05$  durchführen.

- a) Diese Vermutung kann mit einem  $\chi^2$ -Anpassungstest überprüft werden:

Wir bezeichnen mit  $X$  die Zufallsvariable, die die Blutgruppenverteilung in den USA beschreibt, wobei wir die Werte numerisch beschreiben, d.h.  $1 = A$ ,  $2 = 0$ ,  $3 = B$  und  $4 = AB$ . Setzen wir dann noch

$$\pi_1 = 0.43, \quad \pi_2 = 0.41, \quad \pi_3 = 0.11, \quad \pi_4 = 0.05$$

so untersuchen wir folgenden Hypothesentest

**Nullhypothese**  $H_0: p(X = k) = \pi_k, (k = 1, 2, 3, 4).$

**Alternativhypothese**  $H_1: p(X = k) \neq \pi_k$  für ein  $k \in \{1, 2, 3, 4\}.$

Dazu benutzen wir einen  $\chi^2$ -Anpassungstest. Dazu betrachte  $n$  unabhängige Wiederholungen  $X_1, \dots, X_n$  von  $X$ , und die Testgrößen

$$Y_{m,k}(\omega) = \begin{cases} 1 & \text{falls } X_m(\omega) = k \\ 0 & \text{falls } X_m(\omega) \neq k \end{cases}$$

für  $k = 1, \dots, 4, m = 1, \dots, n$  und

$$H_k = H_{n,k} = Y_{1,k} + \dots + Y_{n,k}$$

für  $k = 1, \dots, 4$  und die Prüfgröße

$$\chi^2 = \sum_{k=1}^4 \frac{(H_{n,k} - n\pi_k)^2}{n\pi_k}$$

Die Prüfgröße  $\chi^2$  ist approximativ  $\chi^2$ -verteilt mit 3 Freiheitsgraden. Der  $\chi^2$ -Test kann angewendet werden, wenn der Stichprobenumfang so groß ist, dass  $n \cdot \pi_k > 1$  für all  $k \in \{1, 2, 3, 4\}.$

Da  $\chi_{0.95}^2(3) = 7.8147$ , wird  $H_0$  abgelehnt, wenn  $\chi^2 > 7.8147$ .

- b) Ankommen, eine Zufallsstichprobe von 1240 US-Amerikanern liefert die folgenden Daten

Blutgruppe	A	0	B	AB
Personen	520	546	125	49

Da  $n = 1240$  und  $n \cdot \pi_k \geq 1$  für alle  $k = 1, \dots, 4$ , kann der Test (mit der  $\chi^2$ -Verteilung als Bezugsgröße) zur Anwendung gebracht werden. Die Prüfgröße hat in diesem Fall den Wert

$$\begin{aligned} \chi^2 &= \sum_{k=1}^4 \frac{(h_k - 1240 \cdot \pi_k)^2}{1240 \cdot \pi_k} \\ &= \frac{(520 - 533.2)^2}{533.2} + \frac{(546 - 508.4)^2}{508.4} + \frac{(125 - 136.4)^2}{136.4} + \frac{(49 - 62)^2}{62} \\ &= 6.786 \end{aligned}$$

Da  $\chi_{0.95}^2(3) = 7.8147$ , kann die Nullhypothese nicht abgelehnt werden.

## 9 Regressionsanalyse

Bereits in der deskriptiven Statistik haben wir uns mit der Frage beschäftigt, wie der Einfluss einer statistischen Größe auf eine andere gemessen werden kann, und auch bei der Untersuchung der Zufallsvariablen hat sich diese Frage wieder gestellt. Bei der Untersuchung von Kovarianz und Korrelation haben wir schon festgestellt, dass die Korrelation eine gute Größe ist, um festzustellen, ob zwei Zufallsvariablen linear korreliert sind, d.h. ob es eine Beziehung der Form

$$Y = f(X) = \alpha + \beta X$$

gibt, oder ob dies zumindest näherungsweise der Fall ist (also  $Y = f(X) + \varepsilon$  mit einem geeigneten "kleinen"  $\varepsilon$ ). Allerdings gibt die Korrelation keine Aussage über die exakte Gestalt des Zusammenhangs, und selbst wenn die Beobachtungen auf einen linearen Zusammenhang hindeuten, wenn also  $|r_{X,Y}|$  nahe bei 1 liegt, wissen wir noch nichts über  $\alpha$  und  $\beta$ . Mit dieser Frage beschäftigt sich die *lineare Einfachregression*. Das Standardmodell der linearen Einfachregression geht davon aus, dass wir Beziehungen der Form

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, \dots, n)$$

vorliegen haben, wobei

$Y_1, \dots, Y_n$  unabhängige beobachtbare metrische Zufallsvariablen

$x_1, \dots, x_n$  gegebene Realisierungen einer metrischen Zufallsvariable  $X$  oder fest vorgegebene deterministische Werte der Variable  $X$ .

$\varepsilon_1, \dots, \varepsilon_n$  unbeobachtbare identische verteilte Störvariablen mit Erwartungswert  $E(\varepsilon_i) = 0$  und Varianz  $\text{Var}(\varepsilon_i) = \sigma^2$ .

Die Regressionskoeffizienten  $\alpha$  und  $\beta$  und die Varianz  $\sigma^2$  sind unbekannte Parameter, die aus den beobachteten Daten  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , zu schätzen sind.

Beispiele für solche Untersuchungen sind vermutete lineare Zusammenhänge zwischen Nettoeinkommen und Konsumausgaben oder zwischen der Entwicklung des Gesamtaktienmarktes und eines ausgewählten (repräsentativen) Aktie dieses Marktes.

Eine gängige Annahme, die wir auch machen werden, ist die Normalverteilungsannahme

$$\varepsilon \sim N(0, \sigma^2)$$

oder äquivalent dazu

$$Y_i = N(\alpha + \beta x_i, \sigma^2) \quad (i = 1, \dots, n)$$

Unter dieser Annahme lassen sich vernünftige Aussage über die Qualität von Schätzern erzielen.

Der Ansatz den wir hier verfolgen orientiert sich an der Kleinst-Quadrate-Methode aus der Regressionsanalyse der deskriptiven Statistik, wobei in dem Kleinst-Quadrate- oder KQ-Schätzer die Realisationen  $y_i$  durch die Zufallsvariablen  $Y_i$  ersetzt werden:

Bestimme Schätzwerte  $\hat{\alpha}$  und  $\hat{\beta}$  so, dass

$$\sum_{i=1}^n \left( Y_i - \hat{\alpha} - \hat{\beta} x_i \right)^2 = \min \left\{ \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \mid \alpha, \beta \in \mathbb{R} \right\}$$

(in einem geeigneten Sinne). Die Schätzstatistiken, die wir benutzen werden, orientieren sich an den Schätzwerten aus der deskriptiven Statistik. Dabei setzen wir

$$\begin{aligned} \bar{x} &= \frac{1}{n} \cdot (x_1 + \cdots + x_n) \\ \bar{Y} &= \frac{1}{n} \cdot (Y_1 + \cdots + Y_n) \end{aligned}$$

und erhalten

**Satz 9.1.** *Wir nehmen an, dass*

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2)$$

*Für die Schätzstatistiken*

$$\begin{aligned} \hat{\beta} &= \frac{\sum (x_i - \bar{x}) \cdot (Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\alpha} &= \bar{Y} - \hat{\beta} \cdot \bar{x} \\ \hat{\sigma}^2 &= \frac{1}{n-2} \cdot \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n-2} \cdot \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \end{aligned}$$

*gilt*

$$\begin{aligned} E(\hat{\alpha}) &= \alpha \\ E(\hat{\beta}) &= \beta \\ E(\hat{\sigma}^2) &= \sigma^2 \end{aligned}$$

*sowie*

$$\begin{aligned} \text{Var}(\hat{\alpha}) &= \sigma_{\hat{\alpha}}^2 = \frac{\sigma^2 \cdot \sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2 \cdot \sum_{i=1}^n x_i^2}{n \cdot \left( \sum_{i=1}^n x_i^2 \right) - n^2 \cdot \bar{x}^2} \\ \text{Var}(\hat{\beta}) &= \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{\left( \sum_{i=1}^n x_i^2 \right) - n \cdot \bar{x}^2} \end{aligned}$$



Insbesondere sind also  $\hat{\alpha}$ ,  $\hat{\beta}$  und  $\hat{\sigma}^2$  erwartungstreue Schätzer. Falls

$$\lim_{n \rightarrow \infty} \left( \sum_{i=1}^n (x_i - \bar{x})^2 \right) = \infty$$

so sind die Schätzer auch konsistent.

**Bemerkung 9.1.** Eine einfache Umrechnung zeigt

$$\begin{aligned}\hat{\beta} &= \sum_{i=1}^n b_i \cdot Y_i \\ \hat{\alpha} &= \sum_{i=1}^n a_i \cdot Y_i\end{aligned}$$

wobei

$$\begin{aligned}b_i &= \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ a_i &= \frac{1}{n} - \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \bar{x}\end{aligned}$$

so dass also  $\hat{\alpha}$  und  $\hat{\beta}$  lineare Funktionen in  $Y_1, \dots, Y_n$  sind.

Um das Schätzproblem noch besser in den Griff zu bekommen, ist es wichtig, auch die Verteilungsfunktion des Schätzers zu verstehen. Dabei benutzen wir wieder unsere Normalverteilungsannahme

$$\varepsilon_i \sim N(0, \sigma^2) \quad \text{bzw.} \quad Y_i \sim N(\alpha + \beta x_i, \sigma^2)$$

Da wir die  $Y_i$  als unabhängig annehmen, und da

$$\hat{\alpha} = \sum_{i=1}^n a_i Y_i, \quad \hat{\beta} = \sum_{i=1}^n b_i Y_i$$

sind auch  $\hat{\alpha}$  und  $\hat{\beta}$  normalverteilt mit

$$\hat{\alpha} \sim N(\alpha, \sigma_{\hat{\alpha}}^2), \quad \hat{\beta} \sim N(\beta, \sigma_{\hat{\beta}}^2)$$

wobei  $\sigma_{\hat{\alpha}}^2$  und  $\sigma_{\hat{\beta}}^2$  die Varianzen von  $\hat{\alpha}$  und  $\hat{\beta}$  aus Satz 9.1 sind. Ersetzt man dort die unbekannte Varianz  $\sigma^2$  durch den Schätzer  $\hat{\sigma}^2$ , so erhalten wir

**Satz 9.2.** Unter der Normalverteilungsannahme gilt

$$\frac{\hat{\alpha} - \alpha}{\hat{\sigma}_{\hat{\alpha}}} \sim t(n-2), \quad \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}} \sim t(n-2)$$

wobei

$$\hat{\sigma}_{\hat{\alpha}} = \hat{\sigma} \cdot \frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}}, \quad \hat{\sigma}_{\hat{\beta}} = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Damit gilt für die  $\gamma$ -Konfidenzintervalle für  $\alpha$ :

$$I = \left[ \hat{\alpha} - \hat{\sigma}_{\hat{\alpha}} \cdot t_{1-\frac{\gamma}{2}}(n-2), \hat{\alpha} + \hat{\sigma}_{\hat{\alpha}} \cdot t_{1-\frac{\gamma}{2}}(n-2) \right]$$

und für  $\beta$ :

$$J = \left[ \hat{\beta} - \hat{\sigma}_{\hat{\beta}} \cdot t_{1-\frac{\gamma}{2}}(n-2), \hat{\beta} + \hat{\sigma}_{\hat{\beta}} \cdot t_{1-\frac{\gamma}{2}}(n-2) \right]$$

**Bemerkung 9.2.** Ist  $n > 30$  können die  $\gamma$ -Quantile der  $t(n-2)$ -Verteilung näherungsweise durch die entsprechenden Quantile der Standardnormalverteilung ersetzt werden.

Die lineare Einfachregression untersucht den Fall, in dem eine abhängige Variable durch genau eine unabhängige Variable erklärt werden soll. Oft ist es aber so, dass mehrere unabhängige Variablen vorhanden sind, die alle die Zielgröße beeinflussen können, und das führt zur *multiplen linearen Regression*.

Das Standardmodell der multiplen linearen Regression geht davon aus, dass wir Beziehungen der Form

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p} + \varepsilon_i \quad (i = 1, \dots, n)$$

vorliegen haben, wobei

$Y_1, \dots, Y_n$  unabhängige beobachtbare metrische Zufallsvariablen

$x_{1,j}, \dots, x_{n,j}$  gegebene Realisierungen einer metrischen Zufallsvariable  $X_j$  oder fest vorgegebene deterministische Werte der Variable  $X_j$ .

$\varepsilon_1, \dots, \varepsilon_n$  unbeobachtbare identische verteilte Störvariablen mit Erwartungswert  $E(\varepsilon_i) = 0$  und Varianz  $\text{Var}(\varepsilon_i) = \sigma^2$ .

Die Regressionskoeffizienten  $\beta_0, \dots, \beta_p$  und die Varianz  $\sigma^2$  sind unbekannte Parameter, die aus den beobachteten Daten  $(x_{i,1}, \dots, x_{i,p}, y_i)$ ,  $(i = 1, \dots, n)$  zu schätzen sind.

Wie bei der linearen Einfachregression gehen wir auch hier von einer Normalverteilungsannahme

$$\varepsilon_i \sim N(0, \sigma^2)$$

aus, so dass auch

$$Y_i \sim N(\mu, \sigma^2), \quad \text{mit } \mu_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p}$$

Die Parameter  $\beta_0, \dots, \beta_p$  werden hier wiederum nach der Kleinst-Quadrate-Methode gesucht:

Bestimme  $\hat{\beta}_0, \dots, \hat{\beta}_p$  so, dass

$$\begin{aligned} & \sum_{i=1}^n \left( Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - \dots - \hat{\beta}_p x_{i,p} \right)^2 \\ &= \min \left\{ \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p})^2 \mid \beta_0, \dots, \beta_p \in \mathbb{R} \right\} \end{aligned}$$

Dieses Problem kann nun nicht mehr durch eine allgemeine Formel gelöst werden. Für gegebene Werte  $x_{1,j}, \dots, x_{n,j}$  handelt es sich hier um ein  $p+1$ -dimensionales Minimierungsproblem, das mit Mitteln der mehrdimensionalen Analysis behandelt werden kann. Das daraus resultierende lineare Gleichungssystem zur Bestimmung der Nullstellen der partiellen Ableitungen wird in der Praxis am besten numerisch gelöst. Als Schätzer für die Varianz fungiert dann

$$\hat{\sigma} = \frac{1}{n-p-1} \cdot \sum_{i=1}^n \hat{\varepsilon}_i^2$$

wobei

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

und

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \dots + \hat{\beta}_p x_{i,p}$$

Oft wird die multiple lineare Regression auch in Matrixschreibweise dargestellt, d.h. wir schreiben

$$\vec{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,p} \\ 1 & x_{2,1} & \dots & x_{2,p} \\ \vdots & & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,p} \end{pmatrix}$$

für die abhängigen und die unabhängige Variable und

$$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

so dass sich unser Grundmodell kompakt als

$$\vec{Y} = \mathbf{X} \cdot \vec{\beta} + \vec{\varepsilon}, \quad E(\vec{\varepsilon}) = \vec{0}$$

schreibt, und die Kleinst-Quadrate-Schätzung wird zu

$$\text{minimiere} \quad \left( \vec{Y} - \mathbf{X} \cdot \vec{\beta} \right)^T \cdot \left( \vec{Y} - \mathbf{X} \cdot \vec{\beta} \right)$$

(bezüglich  $\vec{\beta}$ ).