

Vorlesung Mathematik 3

Dr. Mandy Lange-Geisler

3. Juli 2023

Inhaltsverzeichnis

I	Grundlagen	7
1	Einführung in die Kombinatorik	9
1.1	Permutationen	9
1.1.1	Permutationen ohne Wiederholung	9
1.1.2	Permutationen mit Wiederholung	10
1.2	Auswahlen	11
1.2.1	Variationen ohne Wiederholung	11
1.2.2	Variationen mit Wiederholung	12
1.2.3	Kombinationen ohne Wiederholung	12
1.2.4	Kombinationen mit Wiederholung	14
II	Wahrscheinlichkeitstheorie	17
2	Grundlagen der Wahrscheinlichkeitsrechnung	19
2.1	Zufällige Ereignisse	19
2.2	Relationen zwischen zufälligen Ereignissen	21
2.3	Die σ -Algebra	22
3	Der Wahrscheinlichkeitsbegriff	25
3.1	Axiomatische Definition der Wahrscheinlichkeit	25
3.2	Klassische Definition der Wahrscheinlichkeit	26
3.3	Geometrische Definition der Wahrscheinlichkeit	27
3.4	Statistische Ermittlung der Wahrscheinlichkeit	29
3.5	Definition der bedingten Wahrscheinlichkeit	30

3.6	Unabhängigkeit zufälliger Ereignisse	32
3.7	Der Satz der totalen Wahrscheinlichkeit und der Satz von BAYES	34
4	Zufallsgrößen und Verteilungsfunktionen	37
4.1	Definition einer Zufallsgröße und deren Verteilungsfunktion	38
4.2	Grundlegende Begriffe diskreter Zufallsgrößen	40
4.3	Momente diskreter Zufallsgrößen	42
4.3.1	Das k -te Moment und der Erwartungswert	42
4.3.2	Das k -te zentrale Moment und die Varianz	43
4.4	Diskrete Verteilungsfunktionen	46
4.4.1	Die Null-Eins-Verteilung	46
4.4.2	Die Binomialverteilung	47
4.4.3	Die Poissonverteilung	51
4.4.4	Die Hypergeometrische Verteilung	54
4.5	Grundlegende Begriffe stetiger Zufallsgrößen	57
4.6	Momente stetiger Zufallsgrößen	59
4.6.1	Das k -te Moment und der Erwartungswert	59
4.6.2	Moment, zentrales Moment, Varianz	60
4.7	Stetige Verteilungsfunktionen	61
4.7.1	Die Exponentialverteilung	61
4.7.2	Die Normalverteilung	64
4.7.2.1	Die Standardisierte Normalverteilung $N(0, 1)$	66
4.7.2.2	Die 3σ -Grenze der Normalverteilung	67
4.8	Quantile einer Zufallsgröße	68
5	Grenzwertsätze	73
5.1	Die Ungleichung von Tschebyscheff	73
5.2	Konvergenzarten für Folgen von Zufallsgrößen	76
5.3	Gesetz der großen Zahlen	77
5.3.1	Das schwache Gesetz der großen Zahlen	78
5.3.2	Das starke Gesetz der großen Zahlen	80
5.4	Zentraler Grenzwertsatz	81
5.4.1	Globaler Grenzwertsatz nach LINDBERG/LEVY	81
5.4.2	Globaler Grenzwertsatz von MOIVRE/LAPLACE	83
5.4.3	Lokaler Grenzwertsatz von MOIVRE/LAPLACE	84

III	Statistik	87
6	Beschreibende Statistik	89
6.1	Grundlegende Begriffe der deskriptiven Statistik	90
6.1.1	Deskriptive Statistik für eindimensionale Daten	91
6.1.1.1	Häufigkeitstabellen und Häufigkeitsverteilungen	91
6.1.1.2	Grafische Darstellungsmöglichkeiten für Häufigkeiten	93
6.1.1.3	Mittelwerte (Lageparameter)	94
6.1.1.4	Streuungsmaße	95
6.1.2	Deskriptive Statistik für zweidimensionale Daten	96
6.2	Regressionsanalyse	99
6.2.1	Allgemeines Vorgehen bei einer Regressionsanalyse	99
6.2.2	Einfache lineare Regression	100
7	Beurteilende Statistik	105
7.1	Statistische Schätzmethoden	105
7.1.1	Punktschätzungen	105
7.1.1.1	Momentenmethode	106
7.1.1.2	Maximum-Likelihood-Methode	107
7.2	Konfidenzschätzungen	109
7.2.1	Konstruktion eines Konfidenzintervalls	111
7.2.1.1	Das zweiseitige Konfidenzintervall für den Erwartungs- wert	111
7.2.1.2	Einseitige Konfidenzintervalle für den Erwartungswert	115
7.2.1.3	Das Konfidenzintervall für die Varianz	115
7.2.1.4	Einseitige Konfidenzintervalle für die Varianz	117
7.2.1.5	Das Konfidenzintervall für einen Anteilswert	118
7.3	Statistische Tests	119
7.3.1	Parametertests für eine einfache Stichprobe	120
7.3.2	Test für den Erwartungswert	122
7.3.2.1	Der Einstichproben-Gauß-Test	123
7.3.2.2	Der Einstichproben- t -Test	125
7.3.2.3	Asymptotischer Einstichproben-Gauß-Test	126
7.3.3	Test für die Varianz - Der χ^2 -Streuungstest	126

7.3.4	Test für den Anteilswert - Asymptotischer Einstichproben-Gauß-	
	Test für den Anteilswert	127

Teil I

Grundlagen

Kapitel 1

Einführung in die Kombinatorik

Grob formuliert ist die Kombinatorik die Wissenschaft des Zählens. Zum Beispiel Fragestellungen der Form „Wie viele Möglichkeiten gibt es y, x, z in verschiedenen Reihenfolgen anzuordnen?“ oder „Wie viele Möglichkeiten gibt es y, x, z aus einer gegebenen endlichen Menge M auszuwählen?“ sind Gegenstand der Kombinatorik. Kombinatorik ist ein eigenständiger Bereich der Mathematik, gehört aber zu den Grundlagen der Wahrscheinlichkeitstheorie. Die klassische Definition der Wahrscheinlichkeit basiert auf Anzahlen, die mit Hilfe der Kombinatorik ermittelt werden können. Insbesondere die diskreten Verteilungen nutzen die Kombinatorik.

Ein Grundproblem der Kombinatorik ist das Abzählen der Elemente einer gegebenen endlichen Menge. Dabei unterscheidet man die

1. Anordnungen (Permutationen)
2. Auswahlen (Kombinationen, Variationen)
3. Verteilungen und Zerlegungen (Partitionen).

1.1 Permutationen

Permutationen werden unterschieden zwischen ohne oder mit Wiederholung der vorkommenden Objekte.

1.1.1 Permutationen ohne Wiederholung

Die Permutation beschäftigt sich mit den Problemen der Anordnung oder Reihenfolge von Dingen.

Frage: Auf wie viel verschiedene Arten kann man n Objekte anordnen?

Definition 1.1. Eine Anordnung von Objekten heißt Permutation. Die Anzahl der Permutationen von n Objekten wird mit P_n bezeichnet.

Um die Anzahl der Permutationen für n Objekte zu bestimmen sind nachfolgende Überlegungen hilfreich.

Herleitung: Für die Untersuchung der möglichen Anordnungen von n Objekten betrachtet man die Permutationen der Menge $\mathbb{N}_n = \{1, \dots, n\}$. Zuerst wird die Menge mit nur einem Element betrachtet. Für \mathbb{N}_1 gibt es nur eine Permutation. Nun soll die Anzahl der Permutationen P_{n-1} der Menge $\mathbb{N}_{n-1} = \{1, \dots, n-1\}$ bestimmt werden. Jeder Permutation der $(n-1)$ -Elemente von \mathbb{N}_{n-1} kann als geordnetes $(n-1)$ -Tupel der Form $(a_1, a_2, \dots, a_{n-1})$ geschrieben werden. Wie kann nun ein neues n -tes Element a_n in eine gegebene Permutation von \mathbb{N}_{n-1} eingefügt werden?

$$\begin{array}{ccccccccc} a_n & & a_n & & a_n & & a_n & & a_n & & a_n \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ & a_1 & , & a_2 & , & a_3 & , & \dots & , & a_{n-1} \end{array}$$

Aus der Grafik wird klar, dass es genau n Plätze (n Möglichkeiten) gibt. Da P_{n-1} die Anzahl der Permutationen des Tupels $(a_1, a_2, \dots, a_{n-1})$ folgt

$$P_n = n \cdot P_{n-1}.$$

Dass heißt, es kann P_n rekursiv ermittelt werden. Mit $P_1 = 1$ kann auch die explizite Formel angegeben werden:

$$P_n = n! = \prod_{k=1}^n k = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n, \quad n \in \mathbb{N}. \quad (1.1)$$

Die Gleichung (1.1) entspricht der Anzahl aller Anordnungen einer n -elementigen Menge.

Beispiel 1.2. Gegeben seien Buchstaben eines Alphabets (n -Objekte). Gesucht sind alle n -stelligen Wörter, so dass kein Buchstabe mehrfach auftritt.

1.1.2 Permutationen mit Wiederholung

Betrachtet wird zunächst folgendes Beispiel. Wie viel verschiedene 7-stellige Wörter lassen sich aus den Buchstaben $\{A, A, A, B, B, C, C\}$ bilden? $7!$ ist hier die falsche Antwort, da A, B und C mehrfach auftreten und die Vertauschung gleicher Buchstaben nicht zu einem neuen Wort führen. In diesem Beispiel ist eine sogenannte *Multimenge* gegeben.

Definition 1.3. Eine Zusammenfassung von Objekten, die einzelne Elemente auch mehrfach umfassen kann, nennt man eine *Multimenge* und wird geschrieben als

$$M = \left\{ \underbrace{a_1, \dots, a_1}_{k_1\text{-mal}}, \underbrace{a_2, \dots, a_2}_{k_2\text{-mal}}, \dots, \underbrace{a_r, \dots, a_r}_{k_r\text{-mal}} \right\}.$$

Die Mächtigkeit von M beträgt $n = k_1 + k_2 + \dots + k_r$.

Definition 1.4. Die Anzahl der Permutationen einer Multimenge M beträgt

$$\binom{n}{k_1, k_2, \dots, k_r} = \frac{n!}{k_1! k_2! \dots k_r!}. \quad (1.2)$$

Dabei wird der links stehende Ausdruck in Gleichung (1.2) als *Multinomialkoeffizient* bezeichnet, welcher eine Verallgemeinerung des *Binomialkoeffizienten* ist.

Beispiel 1.5. Fortsetzung des obigen Beispiels: Wie viel verschiedene Wörter lassen sich aus den Buchstaben $\{A, A, A, B, B, C, C\}$ bilden?

$$\binom{7}{3, 2, 2} = \frac{7!}{3!2!2!} = \frac{5040}{6 \cdot 2 \cdot 2} = 210$$

1.2 Auswahlen

Allgemein wird zwischen den *Variationen* und den *Kombinationen* unterschieden. Variationen sind Auswahlen bei denen die Reihenfolge der gewählten Elemente von Bedeutung ist und Kombinationen sind Auswahlen bei denen die Anordnung der Elemente nicht berücksichtigt wird.

1.2.1 Variationen ohne Wiederholung

Definition 1.6. Eine geordnete Auswahl von k aus n Elementen der Menge M ohne Wiederholung, welches jedes Element aus M höchstens einmal enthält wird Variation ohne Wiederholung genannt und mit V_n^k bezeichnet.

Herleitung: Um die Anzahl geordneten Auswahlen von k aus n Elementen zu bestimmen sind nachfolgende Überlegungen hilfreich.

- Für das erste Element gibt es n Wahlmöglichkeiten.
- Für das zweite Element gibt es $(n - 1)$ Wahlmöglichkeiten.

- ...
- Für das k -te Element gibt es $(n - k + 1)$ Wahlmöglichkeiten.

Die Anzahl der geordneten Auswahlen ohne Wiederholung ist somit

$$V_n^k = n(n-1) \cdots (n-k+1) = n^{\underline{k}} = \frac{n!}{(n-k)!}.$$

Bemerkung 1.7. Im Falle $k > n$ gibt es keine derartige Auswahl.

Beispiel 1.8. Wieviele vierstellige Zahlen kann man aus den Ziffern 4, 7, 3, 2, 9 und 5 bilden?

Mit $n = 6$ und $k = 4$ erhält man $V_6^4 = 6 \cdot (6-1) \cdot (6-2) \cdot (6-3) = 6^{\underline{4}} = \frac{6!}{(6-4)!} = \frac{6!}{2!} = 360$.

1.2.2 Variationen mit Wiederholung

Was passiert, wenn die Elemente wieder mehrfach vorkommen können?

Definition 1.9. Eine geordnete Auswahl von k aus n Elementen der Menge M mit Wiederholung, welches jedes Element aus M mehrfach enthalten kann wird Variation mit Wiederholung genannt und mit ${}^wV_n^k$ bezeichnet. Es gilt

$${}^wV_n^k = n^k.$$

Beispiel 1.10. Ein Autokennzeichen soll aus drei Buchstaben und vier Ziffern bestehen. Es besitzt dann die Struktur $BBBZZZZ$, wobei B für einen beliebigen Buchstaben und Z für eine beliebige Zahl steht. Wie viel Fahrzeuge können auf diese Weise gekennzeichnet werden, wenn das verwendete Alphabet aus 26 Buchstaben umfasst?

$$V = {}^wV_{26}^3 \cdot {}^wV_{10}^4 = 26^3 \cdot 10^4 = 1757600$$

1.2.3 Kombinationen ohne Wiederholung

Definition 1.11. Eine ungeordnete Auswahl von k aus n Elementen der Menge M ohne Wiederholung, welches jedes Element aus M höchstens einmal enthält wird Kombination ohne Wiederholung genannt und mit C_n^k bezeichnet.

Die Anzahl der Kombinationen ohne Wiederholung kann aus der Anzahl der geordneten Auswahlen abgeleitet werden. Da die Reihenfolge der Elemente bei den Kombinationen unwesentlich ist, braucht man nur die Anzahl der Anordnungen zu berücksichtigen (d.h. dividieren).

Definition 1.12. Die Anzahl der ungeordneten Auswahlen von k aus n Elementen ist definiert als

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!} = \frac{n!}{k!(n-k)!}.$$

Bemerkung 1.13. $\binom{n}{k}$ wird als Binomialkoeffizient bezeichnet.

Eigenschaften des Binomialkoeffizienten:

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

$$2. \quad \binom{n}{0} = 1$$

$$3. \quad \binom{n}{n} = 1$$

$$4. \text{ Symmetriebeziehung: } \binom{n}{k} = \binom{n}{n-k}, \text{ da } \binom{n}{n-k} = \frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Beispiel 1.14. Es werden die Gewinnchancen beim Lotto 6 aus 49 betrachtet. Genauer ist die Anzahl der möglichen Tipps gesucht.

Die Anzahl der geordneten Tipps beträgt $49^{\underline{6}} = 10068347520$. Allerdings tritt jeder Tipp $6! = 720$ mal auf. Demzufolge muss noch mit $6!$ dividiert werden.

$\curvearrowright \frac{49^{\underline{6}}}{6!}$ ist die gesuchte Anzahl.

Der Binomische Lehrsatz

Eine der bekanntesten Eigenschaften des Binomialkoeffizienten kommt im Binomialsatz oder binomischen Lehrsatz zum Ausdruck.

Satz 1.15. Es seien x und y zwei reelle (oder komplexe) Variablen und $n \in \mathbb{N}$. Dann gilt

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}. \quad (1.3)$$

Interpretation: $(x + y)^n$ kann als Produkt von n nebeneinander stehenden Klammern geschrieben werden, das heißt $\underbrace{(x + y)(x + y)(x + y) \dots (x + y)}_{n\text{-mal}}$. Wie erhält man die Po-

tenz $x^k y^{n-k}$? Indem aus k der n Klammern x ausgewählt wird und aus den restlichen Klammern y . Die Anzahl der möglichen Auswahlen ist aber gerade die kombinatorische Bedeutung des Binomialkoeffizienten.

Beispiel 1.16.

$$\begin{aligned} (x + y)^3 &= \sum_{k=0}^3 \binom{n}{k} x^k y^{3-k} \\ &= \binom{3}{0} x^3 + \binom{3}{1} x^2 y + \binom{3}{2} x y^2 + \binom{3}{3} y^3 \\ &= x^3 + 3x^2 y + 3x y^2 + y^3 \end{aligned}$$

Satz 1.17. Für $x = y = 1$ in (1.3) folgt

$$2^n = \sum_{k=0}^n \binom{n}{k}.$$

1.2.4 Kombinationen mit Wiederholung

Definition 1.18. Eine ungeordnete Auswahl von k aus n Elementen der Menge M mit Wiederholung, welches jedes Element aus M auch mehrfach enthalten kann wird Kombination mit Wiederholung genannt und mit ${}^w C_n^k$ bezeichnet. Die Anzahl der ungeordneten Auswahlen von k aus n Elementen mit Wiederholung ist definiert als

$$\binom{n + k - 1}{k}.$$

Herleitung am Beispiel:

Gegeben sei eine Kiste mit je 10 blauen, grünen, roten, schwarzen und weißen Kugeln. Wie viel verschiedene Verteilungen der Farben können auftreten, wenn 6 Kugeln aus der Kiste entnommen werden?

Bei der Entnahme dieser Kugeln ist die Reihenfolge unwesentlich. Es handelt sich somit um eine Auswahl ohne Berücksichtigung der Anordnung. Es ist allerdings eine wiederholte Wahl möglich. Eine spezielle Auswahl der 6 Kugeln wäre zum Beispiel *bbggs w* (die Farben werden mit den Anfangsbuchstaben abgekürzt, das heißt 2 blaue Kugeln, 2 grüne Kugeln, eine schwarze Kugel und eine weiße Kugel). Wenn man die Reihenfolge der Farben, wie

oben im Text angegeben, beibehält ist die Auswahl der 6 Kugeln auch eindeutig durch eine Zeichenkette gekennzeichnet:

- das Symbol $|$ entspricht dem Übergang zur nächsten Farbe
- das Symbol \bullet steht für eine gewählte Kugel

Damit kann die Auswahl *bbggs*w dargestellt werden mit der Zeichenkette $\bullet\bullet | \bullet\bullet || \bullet | \bullet$. Welche Zeichenketten entstehen für die Auswahlen *bbbbbb* und *bgrsw*w?

Folglich entspricht jeder Auswahl von 6 Kugeln einer Zeichenkette mit genau 10 Symbolen. Dieses Problem ist 6 aus 10 oder allgemein k aus $n + k - 1$. Demzufolge ist die Anzahl der verschiedenen Verteilungen der Farben

$$\binom{n+k-1}{k} = \binom{5+6-1}{6} = \binom{10}{6} = \frac{5040}{24} = 210.$$

Teil II

Wahrscheinlichkeitstheorie

Kapitel 2

Grundlagen der Wahrscheinlichkeitsrechnung

Die Wahrscheinlichkeitsrechnung befasst sich mit Zufallsexperimenten. Ein Zufallsexperiment ist ein Vorgang, bei dem der Ausgang nicht (exakt) vorhersagbar ist, weil selbst unter „gleichen Versuchsbedingungen“ jeweils verschiedene Ergebnisse auftreten. Einfache Beispiele sind Glücksspiele mit Würfeln oder Münzen. Diese eignen sie sich für Illustrationen der grundlegenden Begriffe und Sachverhalte.

2.1 Zufällige Ereignisse

Definition 2.1. Ein *Zufallsexperiment* ist ein beliebig oft wiederholbarer Vorgang mit einem unbestimmten Ausgang (z. B. Würfel, Werfen einer Münze).

Definition 2.2. Die möglichen, einander ausschließenden Ergebnisse eines Versuches nennt man *Elementarereignisse*.

Bezeichnungen:

- ω_i ist das i -te Elementarereignis
- Ω ist die Menge aller Elementarereignisse

Folglich ist $\omega_i \in \Omega$, $\forall i \in \mathbb{N}_0$.

Definition 2.3. Ein *zufälliges Ereignis* ist eine Teilmenge A der Menge Ω der Elementarereignisse, dass heißt $A \subseteq \Omega$.

Beispiel 2.4. Werfen einer Münze: $\omega_0 = \text{Zahl}$ und $\omega_1 = \text{Kopf}$, $\Omega = \{\omega_0, \omega_1\}$

Hinweis: Zufällige Ereignisse sind Mengen und die Elementarereignisse sind die Elemente der Menge Ω .

Satz 2.5. Ein zufälliges Ereignis A tritt ein genau dann, wenn eines der Elementarereignisse aus denen A besteht eintritt.

Definition 2.6. Ein Ereignis ist unmöglich, wenn es kein Elementarereignis der Menge Ω umfasst.

Folglich ist die leere Menge $\emptyset = A \subseteq \Omega$ das *unmögliche Ereignis*, d. h. das Ereignis, welches nie eintritt (z. B. Werfen der Augenzahl 8 mit einem Würfel nummeriert von 1 bis 6).

Definition 2.7. Ein Ereignis ist sicher, wenn es alle Elementarereignisse der Menge Ω umfasst.

Folglich ist die Menge $\Omega = A \subseteq \Omega$ das *sichere Ereignis*, d. h. das Ereignis, welches stets eintritt (z. B. Werfen irgend einer Augenzahl zwischen 1 und 6 mit einem Würfel).

Definition 2.8. Die Menge aller zufälligen Ereignisse (einschließlich \emptyset und Ω) heißt *Ereignisraum* \mathbb{E} .

\mathbb{E} ist eine Menge von Mengen von Ω . Häufig ist der Ereignisraum die Potenzmenge von Ω , d.h. $\mathbb{E} = P(\Omega)$.

Beispiel 2.9. Werfen einer Münze: Es treten zwei Elementarereignisse auf: ω_0 - Zahl und ω_1 - Wappen. Folglich ist die Menge der Elementarereignisse $\Omega = \{\omega_0, \omega_1\}$ eine endliche Menge. Das Ereignis $A_1 = \{\omega_0\}$ tritt nur ein, wenn die Zahl fällt. Das Ereignis $A_2 = \{\omega_0, \omega_1\}$ tritt ein, wenn Zahl oder Wappen fällt. Da $A_2 = \Omega$ liegt hier sicheres Ereignis vor. Weiter bezeichnet man A_2 auch als zusammengesetztes Ereignis, da es sich aus mehreren Elementarereignissen zusammensetzt. Der Ereignisraum ist $\mathbb{E} = \{\emptyset, \{\omega_0\}, \{\omega_1\}, \{\omega_0, \omega_1\}\}$ also die Potenzmenge von Ω .

Beispiel 2.10. Würfeln: Es treten sechs Elementarereignisse auf: ω_i -Zahl i , $i = 1, \dots, 6$. Damit ist die Menge der Elementarereignisse $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$.

- Ein zufälliges Ereignis A sei zum Beispiel „es fällt eine gerade Zahl“. Dieses Ereignis tritt ein, wenn ω_2, ω_4 oder ω_6 eintreten, dass heißt $A = \{\omega_2, \omega_4, \omega_6\}$.

- Das Ereignis B sei eine Zahl kleiner als 7 zu werfen. Damit ist $B = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ und entspricht damit Ω dem sicheren Ereignis.

- Das Ereignis C sei „es wird die Zahl 3 geworfen“. Damit ist $C = \{\omega_3\}$ ein Elementarereignis.

Ein Ereignisraum \mathbb{E} ist die Potenzmenge von Ω und besteht aus $2^6 = 64$ Ereignissen.

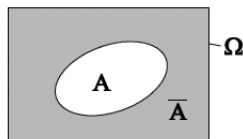
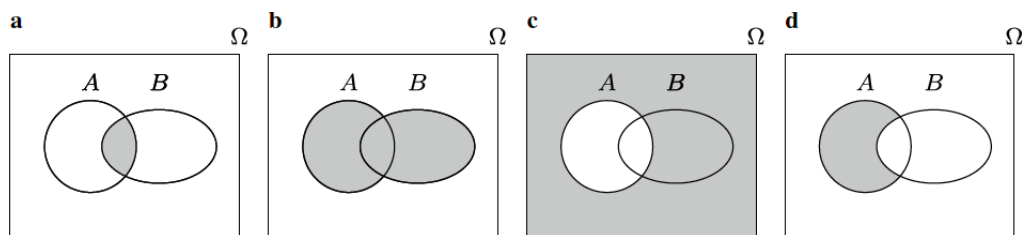


Abbildung 2.1: Darstellung des Komplementärereignisses.

Abbildung 2.2: Darstellung der Ereignisse $A \cap B$, $A \cup B$, $B - A$ und $A - B$.

2.2 Relationen zwischen zufälligen Ereignissen

Da zufällige Ereignisse als Teilmengen der Ereignismenge eingeführt wurden, können Mengenoperationen auf zufällige Ereignisse Anwendung finden. Im folgenden seien $A \subset \Omega$, $B \subset \Omega$ und $C \subset \Omega$ zufällige Ereignisse.

(1) Das *Komplementärereignis* $\bar{A} = \Omega - A$ zum Ereignis A : Das Ereignis \bar{A} tritt ein, wenn das Ereignis A nicht eintritt, siehe Abbildung 2.1.

(2) Die *Summe* $A \cup B$ der Ereignisse A und B : Das Ereignis $A \cup B$ tritt ein, wenn das Ereignis A oder das Ereignis B oder beide Ereignisse eintreten, siehe Abbildung 2.2 (b).

(3) Das Produkt $A \cap B$ der Ereignisse A und B : Das Ereignis $A \cap B$ tritt ein, wenn die Ereignisse A und B eintreten, siehe Abbildung 2.2 (a).

(4) Das Ereignis $A - B$ (bzw. $A \setminus B$) tritt ein, wenn A eintritt, aber B nicht eintritt, , siehe Abbildung 2.2 (d).

(5) Disjunkte (unvereinbare, konträre) Ereignisse: Die Ereignisse A und B heißen disjunkt, wenn A und B nicht gleichzeitig auftreten können. Sind die Ereignisse A und B disjunkt, so ist ihr Produkt $A \cap B = \emptyset$ das unmögliche Ereignis.

Bemerkung 2.11. Elementarereignisse sind disjunkt. Das Ereignis A und dessen komplementäres Ereignis \bar{A} sind disjunkt.

(6) Das Ereignis A zieht das Ereignis B nach sich, wenn $A \subset B$: Aus dem Eintreten des

Ereignisses A folgt zwangsläufig des Eintreten des Ereignisses B .

Beispiel 2.12. (Würfeln) Es sei $A = \{\omega_2\}$ das Ereignis für das Werfen einer 2 und $B = \{\omega_2, \omega_4, \omega_6\}$ das Ereignis für das Würfeln einer geraden Zahl. Da $A \subset B$ ist, zieht das Ereignis A das Ereignis B nach sich.

Allgemeine wichtige Eigenschaften:

- $\bar{\emptyset} = \Omega, \bar{\Omega} = \emptyset$
- $A \cup \bar{A} = \Omega, A \cap \bar{A} = \emptyset$
- $A - B = A \cap \bar{B}$

Distributivgesetze:

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

De MORGAN'sche Regeln:

1. $\overline{A \cup B} = \bar{A} \cap \bar{B}$ oder allgemein: $\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \bar{A}_i$
2. $\overline{A \cap B} = \bar{A} \cup \bar{B}$ oder allgemein: $\overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \bar{A}_i$

2.3 Die σ -Algebra

Bisher wurden Mengen von zufälligen Ereignissen betrachtet. Diesen Mengen soll nun ein *Maß* (später eine Wahrscheinlichkeit) zu geordnet werden. Dass heißt es wird eine Abbildung betrachtet, welche Mengen Zahlenwerte zuordnet, die gewisse Eigenschaften erfüllen. Erstens, ein Maß ist nicht negativ. Zweitens, wenn disjunkte Mengen vereinigt werden, so ist das Maß der Gesamtmenge die Summe der Maße der Einzelmengen. Dies nennt man auch *σ -additiv*. Der Begriff einer Wahrscheinlichkeit ist ein Spezialfall des eben erwähnten mathematischen Maßbegriffs. Außerdem stellt sich die Frage: Auf welchen Mengen ist ein Maß überhaupt definierbar? Dafür betrachtet man die sogenannte *messbare Menge* und den *sogenannten messbaren Raum*. Zuerst wird der Begriff *σ -Algebra* eingeführt.

Um also später den Wahrscheinlichkeitsbegriff für zufällige Ereignisse A sinnvoll definieren zu können, muss der Ereignisraum \mathbb{E} bestimmte Bedingungen erfüllen, die zur sogenannten σ -Algebra führen.

Definition 2.13. Es sei Ω eine nicht leere Menge und \mathbb{E} der zugehörige Ereignisraum mit folgenden Eigenschaften

1. Wenn $A \in \mathbb{E}$, dann ist auch $\bar{A} \in \mathbb{E}$ (stabil bezüglich der Komplementbildung)
2. Wenn $A_i \in \mathbb{E}, \forall i$, dann ist auch $\cup_{i=1}^{\infty} A_i \in \mathbb{E}$ (stabil bezüglich abzählbarer Vereinigungen)

so bildet \mathbb{E} eine σ -Algebra über Ω .

Folgerungen:

1. $\Omega \in \mathbb{E}$

Beweis: $\bar{A} \in \mathbb{E}, A \in \mathbb{E}, \bar{A} \cup A = \Omega \curvearrowright \Omega \in \mathbb{E}$

2. $\emptyset \in \mathbb{E}$

Beweis: $\Omega \in \mathbb{E}, \bar{\Omega} \in \mathbb{E}$ und $\bar{\Omega} = \emptyset \curvearrowright \emptyset \in \mathbb{E}$

3. Für $A_i \in \mathbb{E}, \forall i$ ist auch $\bigcap_{i=1}^{\infty} A_i \in \mathbb{E}$.

Beweis: $A_1, A_2, \dots \in \mathbb{E} \curvearrowright \bar{A}_1, \bar{A}_2, \dots \in \mathbb{E} \curvearrowright \bigcup_{i=1}^{\infty} \bar{A}_i \in \mathbb{E} \curvearrowright \overline{\bigcup_{i=1}^{\infty} \bar{A}_i} \in \mathbb{E} \curvearrowright \bigcap_{i=1}^{\infty} A_i \in \mathbb{E}$.

4. Für $A_1, A_2 \in \mathbb{E}$ folgt $A_1 \setminus A_2 \in \mathbb{E}$

Beweis: Es gilt $A_1 \setminus A_2 = A_1 \cap \bar{A}_2$. Weiter ist $A_1 \in \mathbb{E}$ und $\bar{A}_2 \in \mathbb{E}$ und damit auch $A_1 \cap \bar{A}_2 \in \mathbb{E}$. $\curvearrowright A_1 \setminus A_2 \in \mathbb{E}$

Beispiel 2.14. Es sei $\Omega = \{1, 2, 3\}$. Dann sind $\mathbb{E}_1 = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$ und $\mathbb{E}_2 = \{\emptyset, \{1, 3\}, \{2\}, \{1, 2, 3\}\}$ σ -Algebren über Ω . Dagegen ist $\mathbb{E}_3 = \{\emptyset, \{1\}, \{2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}$ keine σ -Algebra über Ω .

Definition 2.15. Für jede Menge Ω ist $\mathcal{A} = \{\emptyset, \Omega\}$ die kleinst mögliche σ -Algebra über Ω , die sogenannte *triviale σ -Algebra*.

Satz 2.16. Ist die Menge der Elementarereignisse endlich, so ist die Potenzmenge $\mathbb{E} = P(\Omega)$ stets eine σ -Algebra über Ω .

Die Potenzmenge ist die größtmögliche σ -Algebra über Ω .

Beispiel 2.17. 2.9 Werfen einer Münze mit $\mathbb{E} = \{\emptyset, \{\omega_0\}, \{\omega_1\}, \{\omega_0, \omega_1\}\}$, wobei ω_0 - Zahl und ω_1 - Wappen die zwei Elementarereignisse sind.

Beispiel 2.18. Betracht wird die Lebensdauer eines elektronischen Bauteils. Hier ist $\Omega = \mathbb{R}$. Die Menge aller offenen und abgeschlossenen Teilmengen (Intervalle) aus \mathbb{R} (die sogenannten BOREL-Mengen \mathbb{B}) bildet eine σ -Algebra über Ω .

Satz 2.19. Ω sei die Menge der Elementarereignisse und $A \subset \Omega$ ein beliebiges zufälliges Ereignis, dann ist $\mathbb{E} = \{\Omega, A, \bar{A}, \emptyset\}$ eine σ -Algebra. Es ist die kleinste σ -Algebra über Ω , die A enthält.

Beispiel 2.20. Es wird das Beispiel 2.10 (Würfeln) noch einmal betrachtet. Es sei $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. Das Ereignis A sei „es fällt eine gerade Zahl“, dass heißt $A = \{\omega_2, \omega_4, \omega_6\}$. Damit ist $\bar{A} = \{\omega_1, \omega_3, \omega_5\}$ das Ereignis für das Würfeln einer ungeraden Zahl. Dann ist $\mathbb{E} = \{\Omega, A, \bar{A}, \emptyset\}$ eine σ -Algebra.

Es soll nun die Frage geklärt werden auf welchen Mengen ein Maß überhaupt definierbar ist. Dafür definiert man die sogenannte *messbare Menge* und den sogenannten *messbaren Raum*.

Definition 2.21. Das Paar (Ω, \mathbb{E}) ist ein *messbarer Raum*, wobei \mathbb{E} eine σ -Algebra über Ω ist.

Beispiel 2.22. Im Beispiel 2.10 (Würfeln) ist (Ω, \mathbb{E}) ein messbarer Raum, wenn der Ereignisraum \mathbb{E} die Potenzmenge von Ω ist (bestehend aus 64 Ereignissen), da \mathbb{E} eine σ -Algebra über Ω bildet.

Beispiel 2.23. Im Beispiel 2.18 Lebensdauer eines elektronischen Bauteils ist (\mathbb{R}, \mathbb{B}) ein messbarer Raum, da die BOREL-Mengen \mathbb{B} eine σ -Algebra über $\Omega = \mathbb{R}$ bilden.

Definition 2.24. Eine *messbare Menge* ist jedes Element $A \in \mathbb{E}$, wobei \mathbb{E} eine σ -Algebra über Ω ist.

Satz 2.25. Zufällige Ereignisse $A \in \mathbb{E}$ sind messbare Mengen (unter der Bedingung, dass \mathbb{E} eine σ -Algebra über Ω ist).

Darauf basierend kann nun der Wahrscheinlichkeitsbegriff im folgenden Kapitel eingeführt werden.

Kapitel 3

Der Wahrscheinlichkeitsbegriff

Die allgemeinen Eigenschaften der Wahrscheinlichkeit werden durch die axiomatische Definition der Wahrscheinlichkeit beschrieben. Für die konkrete Berechnung der Wahrscheinlichkeit gibt es verschiedene Möglichkeiten, zum Beispiel die klassische, die geometrische und die statistische Definition der Wahrscheinlichkeit.

3.1 Axiomatische Definition der Wahrscheinlichkeit

In diesem Abschnitt soll ein Maß P (Wahrscheinlichkeit, P wie Probability) für die Sicherheit, mit der zufällige Ereignisse eintreten, eingeführt werden. Folgende Voraussetzungen müssen dafür erfüllt sein:

- Die Menge der Elementarereignisse sei nicht die leere Menge ($\Omega \neq \emptyset$).
- Der Ereignisraum \mathbb{E} ist eine σ -Algebra über Ω .

Definition 3.1. (KOLMOGOROW) Die *Wahrscheinlichkeit* ist eine reellwertige Funktion $P(A)$ über der σ -Algebra \mathbb{E} über Ω , die folgende Axiome erfüllt:

1. $P(A) \geq 0$ für beliebige $A \in \mathbb{E}$ (nicht negativ)
2. $P(\Omega) = 1$ (normiert)
3. $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$ (σ -additiv)

Durch diese Axiome werden die Eigenschaften des Wahrscheinlichkeitsbegriffes festgelegt.

Definition 3.2. Das Tripel (Ω, \mathbb{E}, P) heißt *Wahrscheinlichkeitsraum*, wobei (Ω, \mathbb{E}) ein messbarer Raum ist und P ein Wahrscheinlichkeitsmaß auf (Ω, \mathbb{E}) ist.

Ein Zufallsexperiment ist mit dem zugehörigen Wahrscheinlichkeitsraum (Ω, \mathbb{E}, P) vollständig charakterisiert. Jedem zufälligen Ereignis $A \in \mathbb{E}$ kann damit ein Wahrscheinlichkeitsmaß $P(A)$ zugeordnet werden.

Folgerungen:

1. $P(\emptyset) = 0$
2. $P(\bar{A}) = 1 - P(A)$
3. $A \subseteq B \Rightarrow P(A) \leq P(B)$

Satz 3.3. (*Additionssatz*)

Es sei $A \cap B = \emptyset$, dann gilt $P(A \cup B) = P(A) + P(B)$.

Es sei $A \cap B \neq \emptyset$, dann gilt $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Hinweis: Der Additionssatz kann auf endliche viele Ereignisse A_1, A_2, \dots, A_n erweitert werden, genannt Formel von SYLVESTER

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n P(A_i \cap A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n P(A_i \cap A_j \cap A_k) \\ \dots + (-1)^n P(A_i \cap \dots \cap A_n).$$

Satz 3.4. Es seien $A_1, A_2, \dots \in \mathbb{E}$ nicht disjunkte Ereignisse. Dann gilt $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$.

Bemerkung 3.5. Durch die Axiome werden die allgemeinen Eigenschaften der Wahrscheinlichkeit beschrieben. Die konkrete Bestimmung von $P(A)$ muss noch erfolgen. Dafür gibt es verschiedene Möglichkeiten. Wenn die Voraussetzung der Gleichwahrscheinlichkeit für alle Elementarereignisse gegeben ist, kann die Wahrscheinlichkeit eines Ereignisses A exakt angegeben werden mittels der klassischen Definition der Wahrscheinlichkeit (siehe Abschnitt 3.2) oder der geometrischen Definition der Wahrscheinlichkeit (siehe Abschnitt 3.3). Eine weitere Methode ist die Schätzung der Wahrscheinlichkeit eines zufälligen Ereignisses über dessen relative Häufigkeit (statistische Definition der Wahrscheinlichkeit).

3.2 Klassische Definition der Wahrscheinlichkeit

Für die Anwendung der klassischen Definition der Wahrscheinlichkeit müssen folgende Voraussetzungen erfüllt sein:

1. Ω enthält nur endlich viele Elementarereignisse.
2. Alle Elementarereignisse sind gleichwahrscheinlich.

Berechnung der Wahrscheinlichkeit eines Ereignisses A :

Definition 3.6. Es sei m die Anzahl der möglichen Elementarereignisse ω_i ($i = 1, \dots, m$) aus Ω , so ist die Wahrscheinlichkeit für das Eintreten jedes dieser Elementarereignisses

$$P(\omega_i) = \frac{1}{m}, \quad i = 1, \dots, m.$$

Definition 3.7. Das Ereignis $A \in \mathbb{E}$ enthalte n Elementarereignisse ($n \leq m$), dann gilt für die Wahrscheinlichkeit für das Eintreten von A

$$P(A) = \sum_{\omega_i \in A} P(\omega_i) = nP(\omega_i) = \frac{n}{m}.$$

Anders formuliert: $P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Elementarereignisse}}{\text{Anzahl der für } A \text{ möglichen Elementarereignisse}}.$

Der mit diesem Wahrscheinlichkeitsmaß definierte Wahrscheinlichkeitsraum (Ω, \mathbb{E}, P) heißt LAPLACEscher Wahrscheinlichkeitsraum.

Beispiel 3.8. (Fortsetzung Beispiel 2.10) Würfeln:

Das Ereignis A sei „es fällt eine gerade Zahl“. Dieses Ereignis enthält 3 für A günstige Elementarereignisse (ω_2, ω_4 und ω_6). Wegen $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ können jedoch 6 mögliche Elementarereignisse eintreten.

$$\cap P(A) = \frac{3}{6} = \frac{1}{2}.$$

Beispiel 3.9. (Werfen von zwei Würfeln):

Das Ereignis A sei „Werfen einer Augenzahl größer als 10“. Dieses Ereignis enthält 3 für A günstige Elementarereignisse $A = \{(5, 6), (6, 5), (6, 6)\}$. Wegen $\Omega = \{(i, j) \mid i, j = 1, \dots, 6\}$ existieren 36 mögliche Elementarereignisse. $\cap P(A) = \frac{3}{36} = \frac{1}{12}.$

3.3 Geometrische Definition der Wahrscheinlichkeit

Besitzt ein Zufallsexperiment unendlich viele, gleich mögliche Elementarereignisse, dann ist die Ermittlung der Wahrscheinlichkeit wegen des unendlich großen Nenners der obigen Formel nicht möglich. Einen Ausweg kann die geometrische Wahrscheinlichkeitsermittlung bieten.

Für die Anwendung der geometrischen Definition der Wahrscheinlichkeit müssen folgende Voraussetzungen erfüllt sein:

1. Es liegt ein Zufallsexperiment mit unendlich vielen Elementarereignissen vor, die einem geometrischen Inhalt entsprechen, dass heißt zum Beispiel als Punkte einer Strecke oder Fläche oder eines räumlichen Bereiches interpretiert werden können.
2. Alle Elementarereignisse sind gleichwahrscheinlich.

Berechnung der Wahrscheinlichkeit: Der Raum der Elementarereignisse entspricht also zum Beispiel der Länge einer Strecke (dem Flächen- oder Rauminhalt eines ebenen oder räumlichen Bereiches). Das zufällige Ereignis A ist eine Teilmenge von Ω und entspricht demzufolge einer Teilstrecke (Teilfläche oder Teilvolumen).

Definition 3.10. Es sei Ω die Menge der Elementarereignisse, A ein Ereignis und μ eine Mengenfunktion. Dann ist $\mu(A)$ der Inhalt von A und $\mu(\Omega)$ der Inhalt von Ω . Dann gilt

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}.$$

Anders formuliert: $P(A) = \frac{\text{Inhalt der für } A \text{ günstigen Teilmenge von } \Omega}{\text{Inhalt der Menge von } \Omega}.$

Beispiel 3.11. Zwei Personen wollen sich innerhalb einer Stunde treffen. Kommt nach einer Zeit von 20 Minuten kein Zusammentreffen zustande, so geht die entsprechende Person. Wie groß ist die Wahrscheinlichkeit für das Zusammentreffen der beiden Personen? Es sei x der Zeitpunkt für das Eintreffen der Person P_1 und y sei der Zeitpunkt für das Eintreffen der Person P_2 , siehe Abbildung 3.11.

Elementarereignisse: (x, y) mit $0 \leq x, y \leq 60$

Menge aller Elementarereignisse: $\Omega = \{(x, y) : x \in [0, 60], y \in [0, 60]\} \subset \mathbb{R}^2$

Die Menge der Elementarereignisse ist also überabzählbar und die Elementarereignisse sind gleichwahrscheinlich.

$$\mu(\Omega) = 60 \cdot 60 = 3600$$

Das Ereignis A ist das Zusammentreffen der beiden Studenten, dies entspricht $A = \{(x, y) \in \Omega : |x - y| \leq 20\}$, wobei gilt $|x - y| \leq 20 \leftrightarrow -20 \leq x - y \leq 20 \leftrightarrow y \geq -20 + x$ und $y \leq x + 20$.

$$\mu(A) = 60 \cdot 60 - 40 \cdot 40 = 2000$$

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{60 \cdot 60 - 40 \cdot 40}{60 \cdot 60} = \frac{20}{36} = \frac{5}{9} \approx 0.56$$

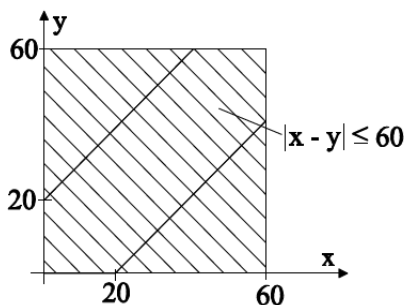


Abbildung 3.1: Grafische Veranschaulichung zum Beispiel 3.11.

3.4 Statistische Ermittlung der Wahrscheinlichkeit

Allgemein werde ein Zufallsexperiment n -mal unter den gleichen Bedingungen durchgeführt, wobei $n \in \mathbb{N}$. A sei ein zufälliges Ereignis. Bei jedem Versuchsdurchgang tritt A oder das Komplementärereignis \bar{A} ein. Die Anzahl der Versuche bei denen A eintritt, heißt *absolute Häufigkeit* des Ereignisses A und wird mit $h_n(A)$ bezeichnet.

Definition 3.12. Tritt bei n -maliger Durchführung eines Zufallsexperimentes $h_n(A)$ -mal (mit $0 \leq h_n(A) \leq n$) das Ereignis A ein, so heißt

$$H_n(A) = \frac{h_n(A)}{n}$$

relative Häufigkeit von A .

Bemerkung 3.13. Die relative Häufigkeit ist demnach vom Zufall abhängig. Verschiedene Versuchsdurchgänge gleichen Umfangs n werden im allgemeinen zu unterschiedlichen Ergebnissen $H_n(A)$ führen, aber diese werden in der unmittelbaren Nähe eines festen Wertes (der Grenzwert der Folge $H_n(A)$) liegen, falls n hinreichend groß ist. An dieser Stelle wird bereits das "Gesetz der großen Zahl" von Jakob Bernoulli angedeutet. Genauer, mit wachsender Zahl der Versuchsdurchgänge strebt die Wahrscheinlichkeit gegen Null, dass die absolute Differenz aus der relativen Häufigkeit und der Wahrscheinlichkeit größer als eine vorgegebene, beliebig kleine positive Zahl ε ist.

Beispiel 3.14. A sei das Ereignis, dass beim Werfen einer Münze das Wappen auftritt. Eine Münze werde 1000-mal geworfen. Berechnet man nach jedem Versuchsdurchgang die durch die bisherige Versuchsreihe bestimmte relative Häufigkeit, ergibt sich eine Zahlenfolge $H_n(A)$, $n \in \{1, \dots, 1000\}$. Für $n \geq 100$ sind die relativen Häufigkeiten in Abbildung 3.2 dargestellt.

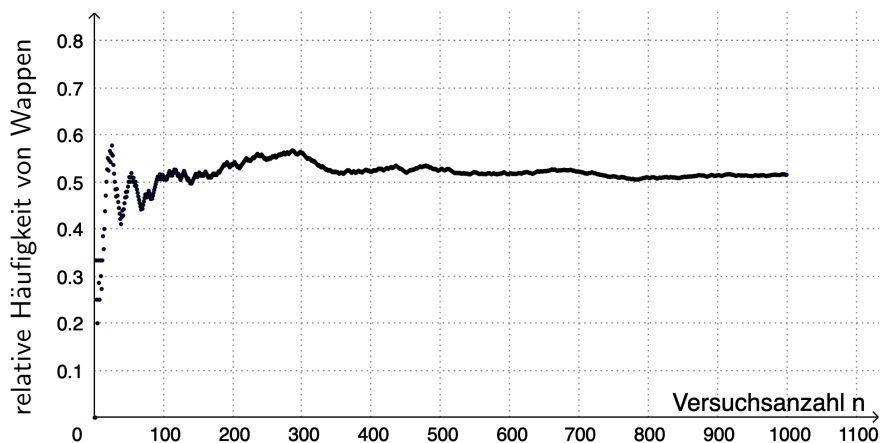


Abbildung 3.2: Relative Häufigkeit eines Münzwurfes mit $n = 1000$.

Eigenschaften:

- $0 \leq H_n(A) \leq 1$
- $H_n(\Omega) = 1$
- $H_n(A \cup B) = H_n(A) + H_n(B)$, falls $A \cap B = \emptyset$

3.5 Definition der bedingten Wahrscheinlichkeit

Gesucht ist die Wahrscheinlichkeit eines zufälligen Ereignisses B unter der Bedingung, dass ein Ereignis A bereits eingetreten ist. Bezeichnung: $P(B | A)$. Das Ereignis A kann als Ursache und das Ereignis B als Wirkung interpretiert werden. Im Folgenden wird vorausgesetzt, dass (Ω, \mathbb{E}, P) ein Wahrscheinlichkeitsraum ist und $A, B \in \mathbb{E}$ zufällige Ereignisse sind.

Definition 3.15. A und B seien Ereignisse aus \mathbb{E} mit $P(A) > 0$. Dann heißt der Ausdruck

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

bedingte Wahrscheinlichkeit des Ereignisses B unter Bedingung das Ereignis A schon eingetreten ist.

Die bedingte Wahrscheinlichkeit $P(B | A)$ genügt den Axiomen eines Wahrscheinlichkeitsmaßes, siehe Definition 3.1 auf Seite 25.

Satz 3.16. *Zwischen den bedingten Wahrscheinlichkeiten $P(B | A)$ und $P(A | B)$ besteht der Zusammenhang*

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

dass heißt aus $P(B | A)$ kann auch $P(A | B)$ bestimmt werden.

Begründung:

$P(B | A) = \frac{P(A \cap B)}{P(A)}$ und $P(A | B) = \frac{P(B \cap A)}{P(B)}$. Damit folgt $P(A \cap B) = P(B | A) P(A)$ und $P(B \cap A) = P(A | B) P(B)$.

$$\curvearrowright P(B | A) P(A) = P(A | B) P(B)$$

$$\curvearrowright P(A | B) = \frac{P(A)}{P(B)} P(B | A) \text{ bzw. } P(B | A) = \frac{P(B)}{P(A)} P(A | B)$$

Beispiel 3.17. In einer Kiste befinden sich 3 weiße und 2 schwarze Bausteine. Aus der Kiste wird jeweils ein Baustein ohne Zurücklegen entnommen. Gesucht ist die Wahrscheinlichkeit dafür, bei der zweiten Entnahme einen weißen Baustein zu ziehen, wenn bei der ersten Entnahme bereits ein weißer Baustein entnommen wurde.

Ereignis A : „Entnahme eines weißen Bausteins im ersten Versuch“

Ereignis B : „Entnahme eines weißen Bausteins im zweiten Versuch“

Ereignis $A \cap B$: „Ziehen von 2 weißen Bausteinen aus insgesamt 5 Bausteinen“

Für die Berechnung von $P(B | A)$ wird $P(A)$ und $P(A \cap B)$ benötigt.

$$P(A) = \frac{3}{5}$$

$$P(A \cap B) = \frac{\text{Anzahl der Möglichkeiten aus 3 weißen Bausteinen 2 davon zu ziehen}}{\text{Anzahl der Möglichkeiten aus 5 Bausteinen 2 Bausteine zu ziehen}} = \frac{\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10}$$

$$\curvearrowright P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{3}{10}}{\frac{3}{5}} = \frac{1}{2}$$

Bemerkung 3.18. Allgemein liegt hier ein sogenanntes *mehrstufiges Zufallsexperiment* vor. Ein mehrstufiges Zufallsexperiment ist eine Folge von einfachen Zufallsexperimenten, die miteinander verknüpft sind. Diese können mittels eines Baumdiagramms veranschaulicht werden. In Abbildung 3.3 wird das mehrstufige Zufallsexperiment aus dem Beispiel 3.17 dargestellt.

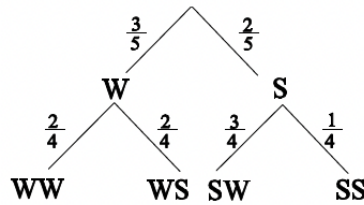


Abbildung 3.3: Baumdiagramm zum Beispiel 3.17.

3.6 Unabhängigkeit zufälliger Ereignisse

Es gibt Ereignisse die keinen kausalen Zusammenhang aufweisen. Genauer, zwischen zwei Ereignissen A und B liegt kein kausaler Zusammenhang vor, wenn das Ereignis A nicht B nach sich zieht und umgekehrt (B nicht A nach sich zieht). Dies wird nun genau definiert. Es sei wieder vorausgesetzt, dass (Ω, \mathbb{E}, P) ein Wahrscheinlichkeitsraum ist und $A, B \in \mathbb{E}$ zufällige Ereignisse sind.

Definition 3.19. (stochastische Unabhängigkeit von Ereignissen) Die Ereignisse A und B aus \mathbb{E} heißen *stochastisch unabhängig*, wenn gilt $P(A \cap B) = P(A)P(B)$.

Satz 3.20. Es sei $P(A) > 0$ und $P(B) > 0$. Die Ereignisse A und B sind genau dann stochastisch unabhängig, wenn $P(B | A) = P(B)$ und $P(A | B) = P(A)$ gilt.

Beispiel 3.21. Fortsetzung Beispiel 3.17. In einer Kiste befinden sich 3 weiße und 2 schwarze Bausteine. Die Bausteine werden nach der Entnahme (entgegen dem früheren Beispiel) jeweils wieder zurückgelegt.

Ereignis A : „Entnahme eines weißen Bausteins im ersten Versuch“ $\rightarrow P(A) = \frac{3}{5}$

Ereignis B : „Entnahme eines weißen Bausteins im zweiten Versuch“ $\rightarrow P(B) = \frac{3}{5}$

Die zweite Entnahme eines Bausteins ist unabhängig von der Entnahme des ersten Bausteins auf Grund des Zurücklegen, daher gilt $P(B | A) = P(B) = \frac{3}{5}$.

$$\cap P(A \cap B) = P(B | A)P(A) = P(B)P(A) = \frac{9}{25}.$$

Beispiel 3.22. Mit welcher Wahrscheinlichkeit erhält man beim Würfeln mit zwei Würfel die Augenzahl 12?

Ereignis A : „erster Würfel zeigt die Augenzahl 6“ $\rightarrow P(A) = \frac{1}{6}$

Ereignis B : zweiter Würfel zeigt die Augenzahl 6“ $\rightarrow P(B) = \frac{1}{6}$

Für $\Omega = \{(i, j) | i, j = 1, \dots, 6\}$ existieren 36 mögliche Elementarereignisse. Für die Augenzahl 12 müssen beide Würfel die Augenzahl 6 zeigen. Demzufolge gibt es nur eine

Möglichkeit die Augenzahl 12 zu erhalten. Mittels der klassischen Definition der Wahrscheinlichkeit ergibt $P(A \cap B) = \frac{1}{36}$.

$\curvearrowright P(A \cap B) = \frac{1}{36} = P(A)P(B)$, was die Unabhängigkeit von A und B zeigt.

Für die Definition der Unabhängigkeit von n Ereignissen gibt es zwei Möglichkeiten. Es kann die Unabhängigkeit paarweise betrachtet werden oder in der Gesamtheit.

Definition 3.23. (paarweise stochastische Unabhängigkeit) Die Ereignisse A_1, \dots, A_n heißen paarweise stochastisch unabhängig, wenn gilt

$$P(A_i \cap A_j) = P(A_i)P(A_j), \forall i \neq j.$$

Definition 3.24. (in der Gesamtheit stochastisch unabhängige Ereignisse) Die Ereignisse A_1, \dots, A_n heißen in ihrer Gesamtheit stochastisch unabhängig, wenn für jedes m -Tupel (i_1, \dots, i_m) mit $1 \leq i_1 \leq \dots \leq i_m \leq n$ und $m = 2, \dots, n$ gilt:

$$P(\cap_{k=1}^m A_{i_k}) = \prod_{k=1}^m P(A_{i_k}).$$

Satz 3.25. Sind die Ereignisse A_1, \dots, A_n in ihrer Gesamtheit unabhängig, so folgt daraus ihre paarweise Unabhängigkeit. Die Umkehrung dieser Aussage gilt jedoch im Allgemeinen nicht.

Beispiel 3.26. Eine Schachtel beinhalte 4 Zettel mit den Zahlenkombinationen 112, 121, 211, 222. Ein Zettel wird zufällig gezogen. Die Elementarereignisse sind $\omega_0 = 112$, $\omega_1 = 121$, $\omega_2 = 211$, und $\omega_3 = 222$.

Das Ereignis A_i ist definiert als "die Ziffer 1 steht an der i -ten Stelle", mit $i = 1, 2, 3$.

Damit ist $A_1 = \{112, 121\}$, $A_2 = \{112, 211\}$ und $A_3 = \{121, 211\}$.

Also gilt $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$.

Weiter ist $A_1 \cap A_2 = \{112\}$, $A_1 \cap A_3 = \{121\}$, $A_2 \cap A_3 = \{211\}$

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

$$P(A_1 \cap A_3) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_3)$$

$$P(A_2 \cap A_3) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_2)P(A_3)$$

Also liegt paarweise Unabhängigkeit vor. Allerdings sind A_1 , A_2 und A_3 nicht in der Gesamtheit unabhängig:

$$A_1 \cap A_2 \cap A_3 = \emptyset$$

$$\curvearrowright P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_2)P(A_3).$$

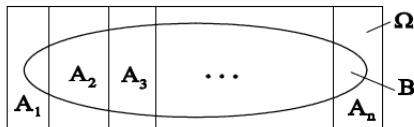


Abbildung 3.4: Ein beliebiges Ereignis B wird unterteilt mit den zufälligen Ereignissen A_1, \dots, A_n , wobei A_1, \dots, A_n ein vollständiges System von Ereignissen ist.

3.7 Der Satz der totalen Wahrscheinlichkeit und der Satz von BAYES

Ein wichtiges Hilfsmittel bei der Berechnung von Wahrscheinlichkeiten zufälliger Ereignisse ist der Satz der totalen Wahrscheinlichkeit und der Satz von BAYES, welche ein vollständiges Ereignissystem voraussetzen.

Definition 3.27. Die zufälligen Ereignisse A_1, \dots, A_n bilden ein *vollständiges System von Ereignissen*, falls gilt

1. $\cup_{i=1}^n A_i = \Omega$ und
2. $A_i \cap A_j = \emptyset, \forall i \neq j$.

Dies ist z. B. dann möglich, wenn endlich viele ω_i (Elementarereignisse) vorliegen. Es sei $A_i = \omega_i$ mit $i = 1, \dots, n$. Ein beliebiges Ereignis B ist dann darstellbar in der Form

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n),$$

siehe Abbildung 3.4. Die Ereignisse $B \cap A_i$ mit $i = 1, \dots, n$ sind dann paarweise disjunkt. Für die Berechnung von $P(B)$ gilt

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_n) \\ &= \sum_{i=1}^n P(B \cap A_i), \end{aligned}$$

wobei gilt

$$P(B \cap A_i) = P(B \mid A_i) P(A_i).$$

Satz 3.28. (Satz der totalen Wahrscheinlichkeit) Es sei (Ω, \mathbb{E}, P) ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n, B \in \mathbb{E}$. Falls A_1, \dots, A_n ein vollständiges System von Ereignissen bilden, dann gilt

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

Die genauere Betrachtung der Berechnung von $P(A_i | B)$ führt zu den Satz von BAYES. Aus der Definition der bedingten Wahrscheinlichkeit folgt

$$\begin{aligned} P(A_i \cap B) &= P(A_i | B) P(B) \\ &= P(B | A_i) P(A_i). \end{aligned}$$

Damit ergibt sich

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

Satz 3.29. (Satz von BAYES) Es sei (Ω, \mathbb{E}, P) ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n, B \in \mathbb{E}$. Falls A_1, \dots, A_n ein vollständiges System von Ereignissen bilden, dann gilt

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

Die Wahrscheinlichkeiten $P(A_i)$, $i = 1, \dots, n$ nennt man *apriori-Wahrscheinlichkeiten* und die Wahrscheinlichkeiten $P(A_i | B)$, $i = 1, \dots, n$ heißen *aposteriori-Wahrscheinlichkeiten*.

Beispiel 3.30. In einer Fabrik werden bestimmte Werkstücke an 3 Maschinen gefertigt, wobei folgende Anteile an der Gesamtproduktion sowie die zugehörigen Ausschussanteile auftreten:

Maschine 1: Anteil an der Gesamtproduktion (%) beträgt 50% mit 3% Ausschussanteil

Maschine 2: Anteil an der Gesamtproduktion (%) beträgt 30% mit 1% Ausschussanteil

Maschine 3 Anteil an der Gesamtproduktion (%) beträgt 20% mit 2% Ausschussanteil

(a) Wie groß ist die Wahrscheinlichkeit, dass ein zufällig aus der Gesamtproduktion entnommenes Werkstück Ausschuss ist?

A_i : „Werkstück wird auf Maschine i mit $i = 1, 2, 3$ gefertigt.“

B : „Werkstück ist Ausschuss“

Die erste Überlegung gilt der Überprüfung des Vorliegen eines vollständiges System von den Ereignissen A_1, \dots, A_n .

Es gilt hier $\Omega = A_1 \cup A_2 \cup A_3$ und $A_i \cap A_j = \emptyset$, $\forall i \neq j$. Daher bildet A_1, A_2, A_3 vollständiges System von Ereignissen.

Weiter können die Wahrscheinlichkeiten für A_1, A_2, A_3 aus den gegebenen Informationen abgelesen werden.

$$P(A_1) = 0.5, P(A_2) = 0.3, P(A_3) = 0.2$$

Außerdem können auch die bedingten Wahrscheinlichkeiten abgelesen werden.

$$P(B | A_1) = 0,03, P(B | A_2) = 0,01, P(B | A_3) = 0,02$$

$P(B)$ kann nun mittels des Satzes der totalen Wahrscheinlichkeit bestimmt werden:

$$\begin{aligned} P(B) &= \sum_{i=1}^3 P(B | A_i) P(A_i) \\ &= 0,03 \cdot 0,5 + 0,01 \cdot 0,3 + 0,02 \cdot 0,2 \\ &= 0,022. \end{aligned}$$

\curvearrowright Der Gesamtausschussanteil liegt bei 2,2%.

(b) Ein Werkstück ist Ausschuss. Mit welcher Wahrscheinlichkeit stammt es von Maschine 1?

Gesucht ist hier die Wahrscheinlichkeit $P(A_1 | B)$, welche mit dem Satz von BAYES bestimmt werden kann.

$$\begin{aligned} P(A_1 | B) &= \frac{P(B | A_1) P(A_1)}{P(B)} \\ &= \frac{0,03 \cdot 0,5}{0,022} \\ &= 0,6818. \end{aligned}$$

Analog kann auch $P(A_2 | B)$ bestimmt werden.

Kapitel 4

Zufallsgrößen und Verteilungsfunktionen

Bisher wurden zufällige Ereignisse als Ergebnisse zufälliger Versuche untersucht. Die zufälligen Ergebnisse wurden lediglich verbal charakterisiert. Besser wäre eine zahlenmäßige Charakterisierung der Ergebnisse zufälliger Versuche, das heißt, interessierende Merkmale sollen mit sogenannten Zufallsgrößen (Zufallsvariablen) beschrieben werden. Eine derartige Beschreibung, kann als eine Abbildung des Sachverhaltes in den Bereich der reellen Zahlen aufgefasst werden. Genauer entspricht dies Abbildungen von einem zugrundeliegenden Wahrscheinlichkeitsraum in einen neuen Wahrscheinlichkeitsraum, der einerseits eine einfachere Struktur hat und andererseits eine „gute“ Beschreibung des interessierenden Merkmals erlaubt.

Beispiel 4.1. Zufallsexperiment: Werfen einer Münze

zufällige Ereignisse: A - „Wappen liegt oben“ und B - „Zahl liegt oben“

zahlenmäßige Beschreibung: $A \rightarrow 1$ und $B \rightarrow 0$

$P(X = 1) = \frac{1}{2}$ und $P(X = 0) = \frac{1}{2}$

Beispiel 4.2. Zufallsexperiment: Werfen zweier Münzen

$\Omega = \{KK, KZ, ZK, ZZ\} = \{\omega_0, \omega_1, \omega_2, \omega_3\}$

Die Zufallsgröße X ordnet jedem dieser Ergebnisse die Anzahl der auftretenden "Kopf-Würfe" zu, das heißt $X(\omega_0) = 2$, $X(\omega_1) = 1$, $X(\omega_2) = 1$, $X(\omega_3) = 0$. Folglich kann X nur die Werte 0, 1 und 2 annehmen.

$P(X = 0) = \frac{1}{4}$, $P(X = 1) = \frac{2}{4} = \frac{1}{2}$, $P(X = 2) = \frac{1}{4}$

4.1 Definition einer Zufallsgröße und deren Verteilungsfunktion

Das Wahrscheinlichkeitsmaßes P , welches auf dem messbaren Raum (Ω, \mathbb{E}) definiert ist, soll nun auf reelle Zahlen zu übertragen werden.

Definition 4.3. (Variante 1) Eine *Zufallsgröße* X mit $X : \Omega \rightarrow \mathbb{R}$ ist eine Abbildung, die jedem Elementarereignis $\omega \in \Omega$ eine reelle Zahl $x = X(\omega)$ zuordnet.

Definition 4.4. Ein fester Wert $x = X(\omega)$, $\omega \in \Omega$ heißt *Realisierung* der Zufallsgröße X .

Im Beispiel 4.1 nimmt die Zufallsgröße X die Realisierungen 1 und 0 an und im Beispiel 4.2 nimmt die Zufallsgröße X die Realisierungen 0, 1 und 2 an.

Definition 4.5. (Variante 2) Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ auf dem messbaren Raum (Ω, \mathbb{E}) heißt *messbar*, falls für beliebiges $x \in \mathbb{R}$ das Urbild des Intervalls $(-\infty, x]$ einer messbaren Menge ist, dass heißt $X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathbb{E}, \forall x \in \mathbb{R}$.

Dabei gilt, dass jedes Intervall der Form $(-\infty, x]$ mit $x \in \mathbb{R}$ aus einem zufälligen Ereignis aus \mathbb{E} hervorgeht, dass heißt das Urbild eines jeden Intervalls $(-\infty, x]$ ist ein zufälliges Ereignis aus \mathbb{E} .

Definition 4.6. (Variante 2) Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ heißt *Zufallsgröße*, falls X eine messbare Funktion ist.

Beispiel 4.7. Für das bereits bekannte Beispiel „Werfen eines Würfels“ mit ω_i – es fällt die Zahl $i, i = 1, \dots, 6$ kann die Definition einer Zufallsgröße wie folgt angegeben werden: $X(\omega_i)$ ist die zufällige Augenzahl, dass heißt $X(\omega_i) = i$ für $i = 1 \dots 6$. Es handelt sich hier um eine spezielle Zufallsgröße, eine sogenannte diskrete Zufallsgröße.

Beispiel 4.8. Für ein weiteres Beispiel „Lebensdauer eines elektronischen Bauteils“ mit ω_t – Lebensdauer beträgt t Zeiteinheiten, $t \in \mathbb{R}^+$ kann die Zufallsgröße wie folgt definiert werden:

$X(\omega_t)$ entspricht hier der zufälligen Lebensdauer und damit ist $X(\omega_t) = t$ mit $t \in \mathbb{R}^+$. Dies ist eine stetige Zufallsgröße.

Es wird nun die wahrscheinlichkeitstheoretische Charakterisierung einer Zufallsgröße definiert.

Definition 4.9. Die Funktion $F_X(x) := P(X \leq x) = P(\{\omega \mid X(\omega) \leq x\})$ heißt *Verteilungsfunktion* der Zufallsgröße X .

Die Verteilungsfunktion $F_X(x)$ einer Zufallsgröße X ordnet jeder Realisierung x die Wahrscheinlichkeit zu, mit der die Zufallsgröße X höchstens den Wert x annimmt. Weiter gibt die Verteilungsfunktion an, wie die gesamte Wahrscheinlichkeitsmasse 1 über dem Wertebereich der Zufallsgröße verteilt ist. Mit der Angabe der Verteilung ist eine Zufallsgröße wahrscheinlichkeitstheoretisch vollständig beschrieben.

Eigenschaften der Verteilung einer Zufallsgröße:

1. Definitionsbereich von $F_X(x)$ ist $-\infty < x < \infty$
2. $F_X(x)$ ist nicht negativ und wegen der Normiertheit von P gilt: $0 \leq F_X(x) \leq 1, \forall x \in \mathbb{R}$
3. $F_X(x)$ ist monoton wachsend dass heißt aus $x_1 < x_2$ folgt $F_X(x_1) \leq F_X(x_2)$
4. $F_X(x)$ ist rechtsseitig stetig, dass heißt $\lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x), \forall x \in \mathbb{R}, h > 0$

Folgerungen aus den Eigenschaften:

1. $P(X > x) = 1 - F_X(x)$

Beweis. $\{X > x\} = \{\overline{X \leq x}\}; P(X > x) = 1 - P(X \leq x) = 1 - F_X(x)$ □

2. $P(a < X \leq b) = F_X(b) - F_X(a)$

Beweis. $\{X \leq b\} = \underbrace{\{X \leq a\} \cup \{a < X \leq b\}}_{\text{disjunkt}}$
 $\Rightarrow \underbrace{P(X \leq b)}_{F(b)} = P(X \leq a \cup a < X \leq b) = \underbrace{P(X \leq a)}_{F(a)} + P(a < X \leq b)$
 $\Rightarrow P(a < X \leq b) = F(b) - F(a)$ □

Bemerkung 4.10. Für die zufälligen Ereignisse wurde eine Abbildung (Zufallsgröße) und deren wahrscheinlichkeitstheoretische Charakterisierung (Verteilungsfunktion) definiert, siehe Abbildung 4.1. Allgemein wird zwischen zwei Arten von Zufallsgrößen unterschieden, die *diskreten Zufallsgrößen* und die *stetigen Zufallsgrößen*.

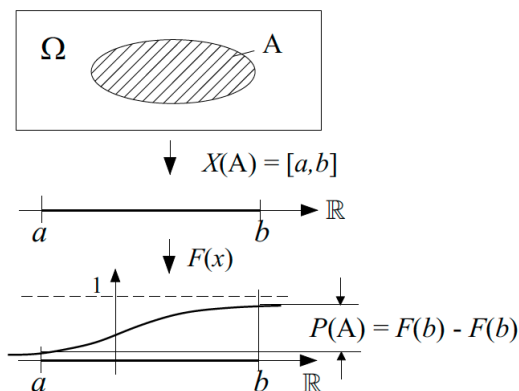


Abbildung 4.1: Schematische Darstellung von der Definition einer Zufallsgröße zu deren Verteilungsfunktion

4.2 Grundlegende Begriffe diskreter Zufallsgrößen

Grob formuliert, ist eine diskrete Zufallsgröße dadurch gekennzeichnet, dass für ihre Realisation in einem vorgegebenen Intervall nur bestimmte Werte in Frage kommen. Anders ausgedrückt: Die Realisation der Zufallsgröße wird durch einen Zählvorgang festgestellt (z.B. Anzahl der richtigen Antworten, Anzahl der Kunden, Absatzmenge).

Definition 4.11. Eine *diskrete Zufallsgröße* nimmt endlich oder abzählbar unendliche viele Realisierungen an.

Dass heißt, der Wertebereich einer diskreten Zufallsgröße ist eine endliche oder abzählbar unendliche Menge.

Beispiel 4.12. Fertigungslos von 15 Stück: Die Zufallsgröße "Anzahl der Ausschusstecke" kann nur die Realisationen $0, 1, 2, \dots, 15$ annehmen. Andere Werte wie 2,4 oder 8,987 sind nicht möglich.

Definition 4.13. Die Einzelwahrscheinlichkeiten sind die Wahrscheinlichkeiten mit denen die Zufallsgröße ihre Werte i annimmt:

$$p_i = P(\{\omega \in \Omega : X(\omega) = i\}), i \in \mathbb{N}_0$$

oder kurz

$$p_i = P(X = i), i \in \mathbb{N}_0.$$

$f_X(i) := P(X = i)$ wird als *Wahrscheinlichkeitsfunktion* der Zufallsgröße X bezeichnet.

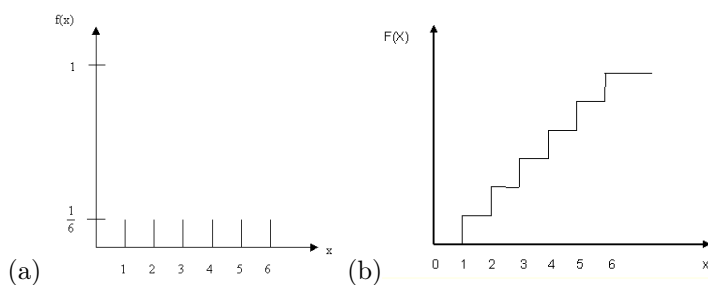


Abbildung 4.2: Für das Beispiel 4.7 sind (a) die Wahrscheinlichkeitsfunktion und (b) die Verteilungsfunktion dargestellt.

Im vorigen Abschnitt wurde bereits erklärt, dass die Gesamtheit der Einzelwahrscheinlichkeiten der Wahrscheinlichkeitsverteilung (oder Verteilung) einer Zufallsgröße entspricht. Für eine diskrete Zufallsgröße ist Verteilung, wie folgt definiert:

Definition 4.14. Die *Verteilung* einer diskreten Zufallsgröße ist definiert durch $F_X(x) = P(X \leq x) = \sum_{i=1}^{i \leq x} p_i, \forall x \in \mathbb{R}$.

Hierbei wird über alle i mit der Eigenschaft $i \leq x$ summiert. Dabei ist $F_X(x)$ rechtsseitig stetig, eine monoton nicht fallende Treppenfunktion mit Sprungstellen der Höhe p_i in i . Außerdem ist $F_X(x)$ normiert, da $\sum_i p_i = 1$ und $P(\Omega) = 1$ gilt.

Beispiel 4.15. Fortsetzung Beispiel 4.7:

Gegeben: Werfen eines Würfels: ω_i – zufällige Zahl $i, i = 1 \dots 6$, $X(\omega_i)$ – zufällige Augenzahl, $X(\omega_i) = i, i = 1 \dots 6$

Gesucht: (a) Verteilung und (b) Wahrscheinlichkeit für eine Augenzahl ≥ 5

$$(a) p_i = \frac{1}{6} \text{ und } F_X(x) = \begin{cases} 0, & x < 1, \text{ da } \emptyset \\ \frac{1}{6}, & 1 \leq x < 2, \text{ da } \omega_1 \\ \frac{1}{3}, & 2 \leq x < 3, \text{ da } \omega_1 + \omega_2 \\ \frac{1}{2}, & 3 \leq x < 4, \text{ da } \omega_1 + \omega_2 + \omega_3 \\ \frac{2}{3}, & 4 \leq x < 5, \text{ da } \omega_1 + \omega_2 + \omega_3 + \omega_4 \\ \frac{5}{6}, & 5 \leq x < 6, \text{ da } \omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 \\ 1, & x \geq 6, \text{ da } \omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 + \omega_6 = \Omega \end{cases}$$

Die Wahrscheinlichkeits- und Verteilungsfunktion sind in Abbildung 4.2 dargestellt. Hier wird deutlich, dass bei der letzten Sprungstelle, also in $x = 6$, $F_X(x)$ den Wert 1 erreicht.

$$(b) P(X \geq 5) = \overline{P(X < 5)} = 1 - P(X < 5) = 1 - \frac{2}{3} = \frac{1}{3}$$

4.3 Momente diskreter Zufallsgrößen

Im letzten Abschnitt wurde deutlich, dass die Verteilung einer diskreten Zufallsgröße durch ihre Verteilungsfunktion bzw. ihre Einzelwahrscheinlichkeiten vollständig bestimmt ist. Wichtige Informationen über eine Verteilung liefern bestimmte Kennwerte (Momente). Dazu gehören der Erwartungswert, die Varianz, die Standardabweichung und der Variationskoeffizient, welche die am häufigsten verwendeten Kennwerte sind.

4.3.1 Das k -te Moment und der Erwartungswert

Grob formuliert ist der Erwartungswert die durchschnittliche Realisation eines Zufallsexperimentes, welche zu erwarten ist.

Definition 4.16. Für eine diskrete Zufallsgröße mit den Einzelwahrscheinlichkeiten p_i ist der *Erwartungswert* definiert als

$$\mu = E(X) := \sum_i ip_i.$$

Interpretation:

- μ ist das gewichtete Mittel über alle Realisierungen einer Zufallsgröße
- Gewichtungsfaktoren bei der Mittelung sind die Einzelwahrscheinlichkeiten p_i
- μ charakterisiert also das "Zentrum" der Verteilung

Bemerkung 4.17. Die Voraussetzung für die Existenz des Erwartungswertes ist die absolute Konvergenz der Reihe, dass heißt $\sum_i |i| p_i < \infty$.¹

Beispiel 4.18. Fortsetzung Beispiel 4.7:

Gegeben: Werfen eines Würfels: ω_i – zufällige Zahl i , $i = 1 \dots 6$, $X(\omega_i)$ – zufällige Augenzahl, $X(\omega_i) = i$, $i = 1 \dots 6$, $p_i = \frac{1}{6}$

Gesucht: Erwartungswert $\rightarrow \mu = E(X) := \sum_{i=1}^6 ip_i = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = 3,5$

Satz 4.19. (*Linearität*) Sei X eine Zufallsgröße mit dem Erwartungswert $E(X)$ und $a, b \in \mathbb{R}$. Dann gilt

$$E(aX + b) = aE(X) + b.$$

Beweis. $E(aX + b) = \sum_i (ai + b) p_i = a \underbrace{\sum_i ip_i}_{EX} + b \underbrace{\sum_i p_i}_1 = aE(X) + b$ □

¹Eine reellwertige oder komplexwertige Reihe $\sum_{n=0}^{\infty} a_n$ heißt absolut konvergent, wenn die Reihe der Absolutbeträge $\sum_{n=0}^{\infty} |a_n| < \infty$ konvergiert.

Folgerung:

- $E(b) = b$
- $E(aX) = aE(X)$

Bemerkung 4.20. Die Linearität lässt sich auch auf endliche Summen von n Zufallsgrößen X_i erweitern, dass heißt es gilt $E(\sum_{i=1}^n aX_i + b) = a \sum_{i=1}^n E(X_i) + b$ mit $a, b \in \mathbb{R}$.

Der Erwartungswert kann auch allgemeiner über das k -te Moment definiert werden.

Definition 4.21. Für eine diskrete Zufallsgröße mit den Einzelwahrscheinlichkeiten p_i ist das k -te *Moment* definiert als

$$\alpha_k = E(X^k) := \sum_i i^k p_i$$

mit $k = 1, 2, \dots$

Folgerung: Für $k = 1$ ergibt sich $\alpha_1 = \mu$, d.h. der Erwartungswert ist das erste Moment.

4.3.2 Das k -te zentrale Moment und die Varianz

Neben den oben definierten Momenten werden auch die zentralen Momente für Zufallsgrößen definiert. Es sind Kenngrößen für die Verteilung der Wahrscheinlichkeitsmasse um den Erwartungswert der Zufallsgröße. Dass heißt, mit Hilfe des Erwartungswertes können weitere Kennwerte einer Verteilung definiert werden. Zunächst wird das k -te zentrale Moment eingeführt, woraus sich die oft verwendete Varianz als Spezialfall ableiten lässt.

Definition 4.22. Für eine diskrete Zufallsgröße mit den Einzelwahrscheinlichkeiten p_i ist das k -te *zentrale Moment* definiert als

$$\beta_k = E((X - \mu)^k) := \sum_i (i - \mu)^k p_i.$$

mit $k = 1, 2, \dots$

β_k zentrale Momente, da sie am Erwartungswert μ zentriert sind.

Folgerung: Das erste zentrale Moment ist immer Null, da $\beta_1 = E((X - \mu)) = E(X) - \mu = 0$.

Das zweite zentrale Moment ist die *Varianz*, welche wie folgt definiert ist.

Definition 4.23. Für eine diskrete Zufallsgröße mit den Einzelwahrscheinlichkeiten p_i ist die *Varianz* (Streuung, Dispersion) definiert als

$$\sigma^2 = D^2(X) = E\left((X - \mu)^2\right).$$

Folgerung: Für $k = 2$ ergibt sich $\beta_k = \sigma^2$, d.h. die Varianz ist das zweite zentrale Moment.

Interpretation der Varianz: Die Varianz einer Wahrscheinlichkeitsverteilung ist ein Maß für die Konzentriertheit, d.h. sie bewertet die Verteilung der Wahrscheinlichkeitsmasse um den Erwartungswert. Ist der Großteil der Masse nahe beim Erwartungswert, so ist die Varianz eher klein. Hier im diskreten Fall dient die Varianz anschaulich zur Bewertung der „Nähe“ um den Erwartungswert bzw. der „Verteilung der Wahrscheinlichkeiten (Gewichte)“. Dies ist nur eine grobe Vorstellung. Die Varianz dient mit dieser Interpretation insbesondere dem Vergleich von Wahrscheinlichkeitsverteilungen. Zusammengefasst kann also gesagt werden:

- σ^2 gibt an, wie stark die Realisierungen um den Erwartungswert μ streuen
- σ^2 ist die mittlere quadratische Abweichung der Zufallsgröße X von μ
- μ und σ^2 sind die beiden wichtigsten Momente einer Zufallsgröße

Der nachfolgende Satz beschreibt eine wichtige Eigenschaft für die Varianz, die oft der Berechnung der Varianz verwendet wird.

Satz 4.24. Sei X eine Zufallsgröße mit dem Erwartungswert $E(X)$. Dann gilt für die Varianz

$$D^2(X) = E(X^2) - (E(X))^2. \quad (4.1)$$

Beweis.

$$\begin{aligned} D^2X &= E\left((X - E(X))^2\right) \\ &= E\left(X^2 - 2XE(X) + (E(X))^2\right) \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - 2(E(X))^2 + (E(X))^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

□

Beispiel 4.25. Fortsetzung Beispiel 4.7:

Gegeben: Werfen eines Würfels: ω_i – zufällige Zahl $i, i = 1 \dots 6$, $X(\omega_i)$ – zufällige Augenzahl, $X(\omega_i) = i, i = 1 \dots 6, p_i = \frac{1}{6}, \mu = E(X) = 3,5$

Gesucht: Varianz: $\sigma^2 = E(X^2) - (E(X))^2 = \sum_{i=1}^6 i^2 p_i - \mu^2 = \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6} - 3,5^2 = 2,91\bar{6}$

Satz 4.26. Sei X eine Zufallsgröße mit dem Erwartungswert $E(X)$ und $a, b \in \mathbb{R}$. Dann gilt

$$D^2(aX + b) = a^2 D^2(X) \quad \forall a, b \in \mathbb{R}.$$

Beweis. Mittels der Eigenschaft 4.1:

$$\begin{aligned} D^2(aX + b) &= E\left((aX + b - E(aX + b))^2\right) \\ &= E\left((aX + b - E(aX) - b)^2\right) \\ &= E\left((aX - E(aX))^2\right) \\ &= E\left(a^2 (X - E(X))^2\right) \\ &= a^2 E\left((X - E(X))^2\right) \\ &= a^2 D^2(X) \end{aligned}$$

□

Folgerung:

1. $D^2(aX) = a^2 D^2(X)$
2. $D^2(b) = 0 \rightarrow$ Eine Konstante variiert (streut) nicht.

Definition 4.27. Für eine Zufallsgröße ist die *Standardabweichung* σ definiert als

$$\sigma = +\sqrt{D^2(X)}$$

Eine weitere Kenngröße ist der Variationskoeffizient v , welcher ein auf den Erwartungswert bezogenes Streuungsmaß ist.

Definition 4.28. Für eine Zufallsgröße ist der *Variationskoeffizient* v definiert als

$$v = \frac{\sigma}{\mu}$$

mit $\mu \neq 0$.

Der Variationskoeffizient v ist also eine Normierung der Standardabweichung σ . Besitzt eine Zufallsgröße einen großen Erwartungswert weist diese im Allgemeinen eine größere Varianz auf als eine Zufallsgröße mit einem kleinen Erwartungswert. Da die Varianz und die daraus abgeleitete Standardabweichung nicht normiert sind, kann ohne Kenntnis des Erwartungswertes nicht beurteilt werden, ob eine Varianz groß oder klein ist.

- $v > 1$ - große Schwankungen
- $v < 1$ - kleine Schwankungen

Beispiel 4.29. Fortsetzung Beispiel 4.7:

Gegeben: Werfen eines Würfels: ω_i - zufällige Zahl $i, i = 1 \dots 6$, $X(\omega_i)$ - zufällige Augenzahl, $X(\omega_i) = i, i = 1 \dots 6, p_i = \frac{1}{6}, \mu = EX = 3,5, \sigma^2 = 2,91\bar{6}$

Gesucht: Standardabweichung und Variationskoeffizient

$$\sigma = \sqrt{2,91\bar{6}} = 1,7076$$

$$v = \frac{1,7076}{3,5} = 0,4879$$

4.4 Diskrete Verteilungsfunktionen

In diesem Abschnitt werden einige spezielle Wahrscheinlichkeitsverteilungen eingeführt, die für die Beschreibung zahlreicher (diskreter) Problemstellungen von Bedeutung sind.

4.4.1 Die Null-Eins-Verteilung

Zufallsgrößen mit einer Null-Eins-Verteilung werden bei Versuchen verwendet, bei denen nur zwei Versuchsausgänge interessieren:

- das Eintreten eines zufälligen Ereignisses A
- das Eintreten des komplementären Ereignisses \bar{A} .

Einfache Beispiele dafür sind „das Werfen einer Münze“, und „das Prüfen eines Produktes auf Funktionalität“.

Allgemeine Bezeichnung: $X \sim \text{Null} - \text{Eins}(p)$

Die hier verwendete diskrete Zufallsgröße ist definiert als $X = \begin{cases} 1, & \text{wenn } A \text{ eintritt} \\ 0, & \text{wenn } \bar{A} \text{ eintritt} \end{cases}$.

Diese Zufallsgröße X besitzt also die Realisierungen 0 und 1.

Satz 4.30. Die Wahrscheinlichkeit für das Eintreten des Ereignisses A sei p . Dann gilt für die Einzelwahrscheinlichkeiten

$$P(X = 1) = p \text{ und } P(X = 0) = 1 - p. \quad (4.2)$$

oder kurz $P(A) = p$ und $P(\bar{A}) = 1 - p$.

Definition 4.31. Eine Zufallsgröße unterliegt einer Null-Eins-Verteilung mit dem Parameter p , wenn sie die Einzelwahrscheinlichkeiten (4.2) besitzt.

Erwartungswert und Varianz der Null-Eins-Verteilung:

1. $E(X) = 0 \cdot (1 - p) + 1 \cdot p = p$
2. $D^2(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p)$

Beispiel 4.32. Aus einer Menge von insgesamt 500 Teilen, unter denen sich 5 Ausschussteile befinden, wird ein Teil entnommen und geprüft. Die Zufallsgröße wird hier beschrieben als X := "zufällige Anzahl der Ausschussteile bei Entnahme eines Teils". Mittels der klassischen Definition der Wahrscheinlichkeit sind die Einzelwahrscheinlichkeiten

$$P(X = 1) = \frac{5}{500} = 0,01 \text{ und } P(X = 0) = \frac{495}{500} = 0,99.$$

Die zugehörige Verteilungsfunktion ist $F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{für } x < 0 \\ 0,99 & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1 \end{cases}.$

Weiter ist $E(X) = 0 \cdot (1 - 0,01) + 1 \cdot 0,01 = 0,01$ und $D^2(X) = 0,01(1 - 0,01) = 0,0099$.

4.4.2 Die Binomialverteilung

Die Binomialverteilung kann als Verallgemeinerung der Null-Eins-Verteilung aufgefasst werden. Sie besitzt 2 Parameter n, p und wird allgemein mit $X \sim B(n, p)$ bezeichnet. Ein typisches Beispiel ist hier die statistische Qualitätskontrolle. Allen Anwendungsbeispielen liegt das sogenannte *BERNOULLI-Schema* zugrunde.

Bernoulli-Schema:

- Durchführung n unabhängiger Versuche, $n \in \mathbb{N}$
- in jedem Versuch gibt es zwei Versuchsausgänge A oder \bar{A} , wobei gilt $P(A) = p$ und $P(\bar{A}) = 1 - p$
- n Versuche sind gleichwahrscheinlich

Ausgehend von dem Bernoulli-Schema betrachtet man folgende Zufallsgröße:

“ X -zufällige Anzahl der Versuche in denen A eintritt (von n Versuchen)“.

Die Zufallsgröße X besitzt die Realisierungen $0, \dots, n$.

Bemerkung 4.33. Für $n = 1$ unterliegt die Zufallsgröße X der Null-Eins-Verteilung.

Definition 4.34. Die Zufallsgröße X heißt binomialverteilt mit den Parametern n und p , wenn für die Einzelwahrscheinlichkeiten gilt

$$p_k = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n, n \in \mathbb{N}$$

mit $n \in \mathbb{N}$.

Definition 4.35. Die Verteilungsfunktion binomialverteilter Zufallsgröße ist definiert als

$$F_X(x) = P(X \leq x) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}.$$

Bemerkungen:

1. $\binom{n}{k}$ gibt die Anzahl der Versuche an, bei denen genau k Versuche mit dem Ereignis A und $n - k$ Versuche mit dem Ereignis \bar{A} enden.
2. Alle Versuche sind unabhängig und jede Versuchsreihe hat die Wahrscheinlichkeit $p^k (1-p)^{n-k}$
3. Es gilt $\sum_{k=0}^n P(X = k) = 1$. (Beweis mittels binomischen Lehrsatz)

Typische Anwendungen der Binomialverteilung:

- Zufällige Anzahl der in einem bestimmten Zeitabschnitt ausfallenden Maschinen von insgesamt n voneinander unabhängig arbeitenden gleichen Maschinen.
- Zufällige Anzahl der Treffer bei n voneinander unabhängigen Schüssen mit gleicher Trefferwahrscheinlichkeit.
- Zufällige Anzahl der Ausschussteile bei einer Stichprobe aus einem Warensortiment

Beispiel 4.36. Abgabe von 4 Schüssen mit der Trefferwahrscheinlichkeit von je $\frac{1}{2}$. Die Zufallsgröße X sei „Anzahl der Treffer bei 4 Schüssen ($n = 4$)“.

$$p_0 = P(X = 0) = \binom{4}{0} \left(\frac{1}{2}\right)^0 \left(1 - \frac{1}{2}\right)^4 = \frac{1}{16}$$

$$p_1 = P(X = 1) = \binom{4}{1} \left(\frac{1}{2}\right)^1 \left(1 - \frac{1}{2}\right)^3 = \frac{4}{16} = \frac{1}{4}$$

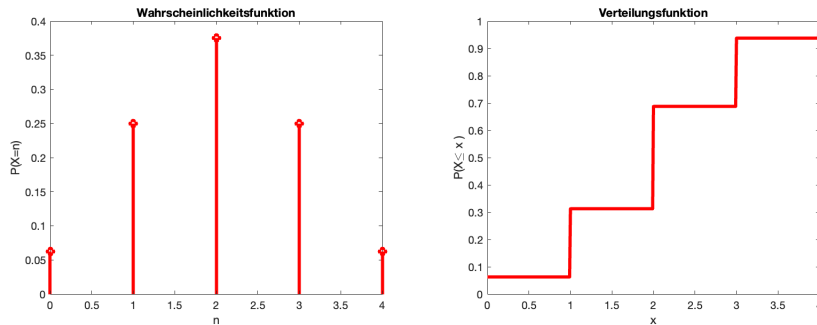


Abbildung 4.3: Die Wahrscheinlichkeitsfunktion und Verteilungsfunktion für eine binomialverteilte Zufallsgröße mit den Parametern $n = 4$ und $p = \frac{1}{4}$.

$$p_2 = P(X = 2) = \binom{4}{2} \frac{1}{2}^2 \left(1 - \frac{1}{2}\right)^2 = \frac{6}{16} = \frac{3}{8}$$

$$p_3 = P(X = 3) = \binom{4}{3} \frac{1}{2}^3 \left(1 - \frac{1}{2}\right)^1 = \frac{4}{16} = \frac{1}{4}$$

$$p_4 = P(X = 4) = \binom{4}{4} \frac{1}{2}^4 \left(1 - \frac{1}{2}\right)^0 = \frac{1}{16}$$

Die zugehörige Wahrscheinlichkeitsfunktion und Verteilungsfunktion ist in Abbildung 4.3 dargestellt.

Erwartungswert und Varianz der Binomialverteilung:

1. $E(X) = np$, da

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \underbrace{\sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}}_1 \\ &= np, \end{aligned}$$

wobei

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

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

2. $D^2(X) = np(1-p)$, da

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^n \binom{n-2}{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} \\
&\quad + np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= n(n-1)p^2 + np \\
&= n^2p^2 - np^2 + np \\
&= np(1-p) + n^2p^2,
\end{aligned}$$

wobei

$$\begin{aligned}
k^2 \binom{n}{k} &= kn \binom{n-1}{k-1} \\
&= kn \binom{n-1}{k-1} - n \binom{n-1}{k-1} + n \binom{n-1}{k-1} \\
&= (k-1)n \binom{n-1}{k-1} + n \binom{n-1}{k-1} \\
&= n(k-1) \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} + n \binom{n-1}{k-1} \\
&= n(n-1) \frac{(n-2)!}{(k-2)!((n-2)-(k-2))!} + n \binom{n-1}{k-1}
\end{aligned}$$

$$= n(n-1) \binom{n-2}{k-2} + n \binom{n-1}{k-1}.$$

Damit folgt $D^2(X) = E(X^2) - (E(X))^2 = np(1-p) + n^2p^2 - (np)^2 = np(1-p)$.

Beispiel 4.37. (Fortsetzung Beispiel 4.36) Abgabe von 4 Schüssen mit der Trefferwahrscheinlichkeit von je $\frac{1}{2}$. Die Zufallsgröße X sei „Anzahl der Treffer bei 4 Schüssen ($n = 4$)“. Gesucht ist die mittlere Anzahl der Treffer und die Varianz.

$E(X) = np = 4 \cdot \frac{1}{2} = 2$ und $D^2(X) = np(1-p) = 4 \cdot \frac{1}{2} \left(1 - \frac{1}{2}\right) = 1$.

4.4.3 Die Poissonverteilung

Die Poissonverteilung ist die Grenzverteilung der Binomialverteilung und sie wird auch die Verteilung der seltenen Ereignisse genannt. Die Grundlage bildet das BERNOULLI-Schema mit nachfolgenden Bedingungen.

Spezielles BERNOULLI-Schema:

1. Die Anzahl n der durchgeführten unabhängigen Versuche ist sehr groß, d.h. $n \rightarrow \infty$.
2. Die Wahrscheinlichkeit p des interessierenden Ereignisses A ist sehr klein, d. h. $p \rightarrow 0$. Deshalb heißt die Poissonverteilung auch „Verteilung der seltenen Ereignisse“.
3. Es gilt jedoch $np = \lambda = \text{const.}$

Allgemein wird sie mit $X \sim \text{Poiss}(\lambda)$ bezeichnet, wobei die Zufallsgröße X auch hier die zufällige Anzahl des Eintretens von A angibt.

Definition 4.38. Die Zufallsgröße X heißt *poissonverteilt* mit dem Parameter λ , wenn für die Einzelwahrscheinlichkeiten gilt

$$p_i = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, \dots, n$$

für $n \rightarrow \infty$, $p \rightarrow 0$ und $np = \lambda = \text{const.}$

Bemerkung 4.39. $\binom{n}{k} p^k (1-p)^{n-k} \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda}$ (Grenzwertsatz von Poisson)

Definition 4.40. Die Verteilungsfunktion poissonverteilten Zufallsgröße ist definiert als

$$F_X(x) = P(X \leq x) = e^{-\lambda} \sum_{k=0}^x \frac{\lambda^k}{k!}.$$

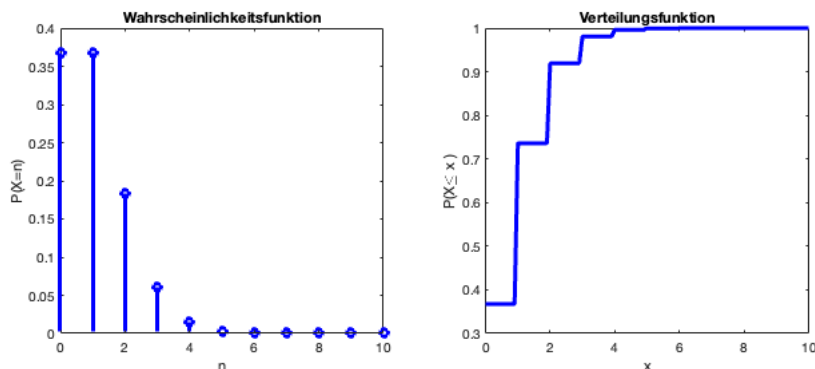


Abbildung 4.4: Darstellung der ersten 10 Funktionswerte der Wahrscheinlichkeitsfunktion und Verteilungsfunktion für eine poissonverteilte Zufallsgröße mit dem Parameter $\lambda = 1$.

Typische Anwendungen der Poissonverteilung:

- Zufällige Anzahl von nicht keimenden Samenkörnern (Packung zu 1000 Stück - wenn durchschnittlich $a\%$ nicht keimen).
- Zufällige Anzahl der Telefonanrufe, die in einem bestimmten Zeitintervall in einer Zentrale ankommen.
- Zufällige Anzahl von α -Teilchen, die von einer radioaktiven Substanz in einem Zeitintervall emittiert werden.

Beispiel 4.41. Die Wahrscheinlichkeit, dass ein Brennelement in einem Kernreaktor einer Qualitätsprüfung nicht standhält beträgt 0,0002. Mit welcher Wahrscheinlichkeit genügen höchstens 2 von 5000 nicht den Qualitätsanforderungen.

X -“Anzahl der Brennelemente, welche nicht den Qualitätsanforderungen genügen“

$X \sim \text{Poiss}(\lambda)$ mit $\lambda = np = 5000 \cdot 0,0002 = 1$

$$P(X \leq 2) = \sum_{k=0}^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^2 \frac{1^k}{k!} e^{-1} = \frac{1^0}{0!} e^{-1} + \frac{1^1}{1!} e^{-1} + \frac{1^2}{2!} e^{-1} = 0,919699$$

Die zugehörige Wahrscheinlichkeitsfunktion und Verteilungsfunktion ist in Abbildung 4.4 dargestellt.

Bemerkung 4.42. Die Einzelwahrscheinlichkeiten der Poissonverteilung können durch den Grenzübergang $n \rightarrow \infty$ aus denen der Binomialverteilung abgeleitet werden, wobei $p = \frac{\lambda}{n}$ ist. Folglich ist die Poissonverteilung für großes n und kleines p und $np = \lambda$ eine gute Näherung für die Binomialverteilung.

Beispiel 4.43. Fortsetzung Beispiel 4.41: Berechnung mittels der Binomialverteilung. $X \sim B(n, p)$ mit $n = 5000$ und $p = 0,0002$.

$$P(X \leq 2) = \sum_{k=0}^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^2 \binom{5000}{k} 0,0002^k 0,9998^{5000-k} = 0,919717$$

Erwartungswert und Varianz der Poissonverteilung:

1. $E(X) = \lambda$, da

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{e^{\lambda}} \\ &= \lambda. \end{aligned}$$

2. $D^2(X) = \lambda$, da

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} \\ &= e^{-\lambda} \left(\sum_{k=1}^{\infty} (k-1) \frac{\lambda^k}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \right) \\ &= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-2)!} + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\ &= e^{-\lambda} \lambda^2 \sum_{k=1}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda^2 + \lambda \end{aligned}$$

Damit folgt $D^2(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Beispiel 4.44. Fortsetzung Beispiel 4.41: $E(X) = \lambda = 1$ und $D^2(X) = \lambda = 1$.

Faustregel für die Anwendung der Poissonverteilung: $np < 10$; $n > 1500p$.

4.4.4 Die Hypergeometrische Verteilung

Die letzte diskrete Verteilung, welche im Rahmen dieser Vorlesung vorgestellt wird, ist die Hypergeometrische Verteilung, welche mit der Binomialverteilung verwandt ist. Die Hypergeometrische Verteilung gibt Auskunft darüber, mit welcher Wahrscheinlichkeit in der Stichprobe eine bestimmte Anzahl von Elementen vorkommt, die ein besonderes Merkmal besitzen.

Das bereits bekannte BERNOULLI-Schema liegt hier in geänderter Form zugrunde.

geändertes Bernoulli-Schema:

- Es liegt eine endliche Grundgesamtheit von N Elementen vor.
- Jeder der n Versuche wird nicht wieder rückgängig gemacht (d.h. zum Beispiel, die Teile einer Stichprobe werden nicht wieder zurückgelegt). Die n Versuche sind nun abhängig.

Das Modell für die Hypergeometrische Verteilung besitzt 3 Parameter N, M und n mit folgender Deutung:

- N ist der Umfang der Grundgesamtheit
- M ist die Anzahl der ausgezeichneten Teile in der Grundgesamtheit (z.B. bei der statistischen Qualitätskontrolle: fehlerhafte Teile), wobei gilt $0 \leq M \leq N$
- n Umfang der Stichprobe, die aus der Grundgesamtheit entnommen wird, wobei gilt $0 \leq n \leq N$

Die Zufallsgröße X beschreibt hier die zufällige Anzahl der sich auszeichnenden Teile in der Stichprobe und wird mit $X \sim Hyp(N, M, n)$ bezeichnet.

Definition 4.45. Die Zufallsgröße X heißt hypergeometrisch verteilt mit den Parametern N, M und n , wenn für die Einzelwahrscheinlichkeiten gilt

$$p_i = P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}},$$

wobei $k = \max(0, n + M - N), \dots, \min(M, n)$.

Bemerkungen:

1. $\binom{N}{n}$ - Anzahl aller möglichen Stichproben mit dem Umfang von genau je n Teilen aus insgesamt N Teilen.

2. $\binom{M}{k}$ - Anzahl aller Möglichkeiten k ausgezeichnete Teile aus den insgesamt M ausgezeichneten Teilen zu wählen.
3. $\binom{N-M}{n-k}$ - Anzahl aller Möglichkeiten $n-k$ nicht ausgezeichnete Teile aus den insgesamt $N-M$ nicht ausgezeichneten Teilen zu wählen.
4. Das Intervall für k ergibt sich aus den Bedingungen:
 - (a) $0 \leq k \leq n$
 - (b) $k \leq M$, d.h. k kann nicht größer als die Gesamtanzahl M der ausgezeichneten Teile sein.
 - (c) $n-k \leq N-M$, d.h. $n-k$ kann nicht größer als die Gesamtanzahl $N-M$ der nicht ausgezeichneten Teile sein $\hookrightarrow k \geq n+M-N$

Definition 4.46. Die Verteilungsfunktion hypergeometrisch verteilten Zufallsgröße ist definiert als

$$F_X(x) = P(X \leq x) = \sum_{k=0}^x \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}.$$

Beispiel 4.47. Eine Lieferung von 500 Bauteilen enthält 10 defekte Teile. Aus der Lieferung wird eine Stichprobe von 50 Teilen entnommen. Mit welcher Wahrscheinlichkeit enthält die Stichprobe kein defektes Bauteil?

Gegeben ist $N = 500$, $M = 10$ und $n = 50$. Die Zufallsgröße ist hier X -“zufällige Anzahl der defekten Bauteile in der Stichprobe“ und unterliegt einer Hypergeometrischen Verteilung.

$$P(X=0) = \frac{\binom{10}{0} \binom{500-10}{50}}{\binom{500}{50}} = \frac{\binom{490}{50}}{\binom{500}{50}} = \frac{\frac{490!}{50!(490-50)!}}{\frac{500!}{50!(500-50)!}} = 0,3452$$

Die zugehörige Wahrscheinlichkeitsfunktion und Verteilungsfunktion ist in Abbildung 4.5 dargestellt.

Erwartungswert und Varianz der Hypergeometrischen Verteilung:

$$1. E(X) = \sum_{k=0}^n k \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = n \frac{M}{N}$$

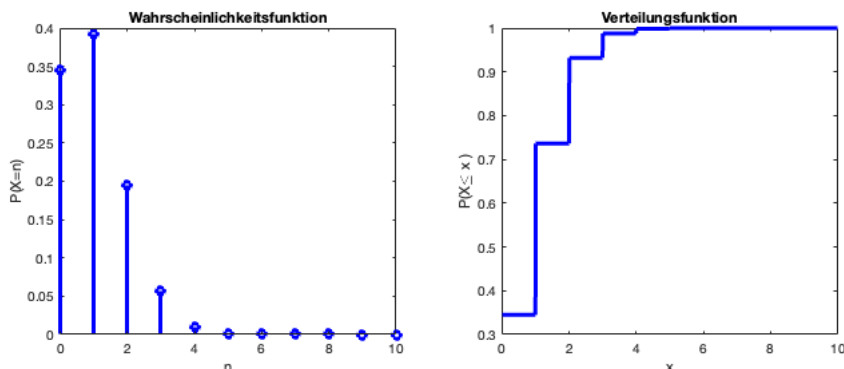


Abbildung 4.5: Darstellung der Wahrscheinlichkeitsfunktion und Verteilungsfunktion für eine hypergeometrisch verteilte Zufallsgröße mit dem Parameter $N = 500$, $M = 10$ und $n = 50$.

$$2. D^2(X) = E(X^2) - (E(X))^2 = \sum_{k=0}^n k^2 \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} - \left(n \frac{M}{N}\right)^2 = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$$

Beispiel 4.48. Fortsetzung Beispiel 4.47: $X \sim \text{Hyp}(N, M, n)$ mit $N = 500$, $M = 10$ und $n = 50$

$$E(X) = n \frac{M}{N} = 50 \frac{10}{500} = 1 \text{ und}$$

$$D^2(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1} = 50 \frac{10}{500} \left(1 - \frac{10}{500}\right) \frac{500-50}{500-1} = 0.8838$$

Hinweis: Wird das gleiche Versuchsschema mit zurücklegen durchgeführt, ergibt sich die Binomialverteilung mit den Parametern n und $p = \frac{M}{N}$.

Beispiel 4.49. Beispiel Fortsetzung 4.47: Ein Bauteil ist mit einer Wahrscheinlichkeit von 0.02 defekt. Es sei die Zufallsgröße nun binomialverteilt mit $n = 50$ und $p = \frac{M}{N} = \frac{10}{500}$.

$$\curvearrowright P(X = 0) = \binom{50}{0} 0.02^0 \cdot (1 - 0.02)^{50} = 0.3642$$

Satz 4.50. Die Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung für $M \rightarrow \infty$ und $N \rightarrow \infty$ geht in die Wahrscheinlichkeitsfunktion der Binomialverteilung über, sofern $\frac{M}{N} \rightarrow p$ konvergiert. 0

Weitere diskrete Verteilungsfunktionen sind die Multinomialverteilung und Geometrische Verteilung, welche hier im Rahmen der Vorlesung entfallen.

4.5 Grundlegende Begriffe stetiger Zufallsgrößen

Was für eine diskrete Zufallsgröße die Wahrscheinlichkeitsverteilung darstellt, ist für eine stetige Zufallsgröße die Dichtefunktion (Dichteverteilung). Der Übergang von einer diskreten zu einer stetigen Zufallsgröße erfolgt durch eine immer feinere Unterteilung der Realisierungen der diskreten Zufallsgröße. Je feiner eine diskrete Zufallsgröße unterteilt ist, desto kleiner wird die Wahrscheinlichkeit für genau eine Realisation. Beim Grenzübergang zu unendlich vielen Realisierungsmöglichkeiten, also dem Übergang von einer diskreten zu einer stetigen Variablen, geht die Wahrscheinlichkeit gegen Null (für eine spezielle Realisierung). Erst die Betrachtung der Wahrscheinlichkeit eines Intervalls $P(a \leq X \leq b)$ mit unendlich vielen Realisierungen liefert eine Wahrscheinlichkeit größer gleich Null.

Zusammengefasst besitzt eine stetige Zufallsgröße überabzählbar viele Realisierungen in einem Intervall, welche durch eine Dichtefunktion beschrieben werden. Die Realisierungen könnten zum Beispiel das Resultat eines Messvorgangs sein.

Definition 4.51. Eine Zufallsgröße heißt *stetig*, wenn es eine integrierbare Funktion $f_X(x) \geq 0$, $-\infty < x < \infty$ derart gibt, dass sich die Verteilungsfunktion $F_X(x) = P(X \leq x) \forall x \in \mathbb{R}$ in der Form

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

darstellen lässt. Die Funktion $f_X(x)$ wird als *Dichtefunktion* bezeichnet.

Eigenschaften der Dichtefunktion $f_X(x)$:

1. $\int_{-\infty}^{\infty} f_X(x) dx = 1$ und
2. $f_X(x) \geq 0, \forall x \in \mathbb{R}$.

Bemerkungen zur Verteilungsfunktion $F_X(x)$:

1. Da $F_X(x)$ monoton nicht fallend ist, muss der Anstieg positiv sein.
2. Wie bereits erwähnt, nimmt die Zufallsgröße X überabzählbar viele Realisierungen in \mathbb{R} an. Ausgehend von der geometrischen Betrachtung des Integralbegriffes ist $F_X(x)$ der Flächeninhalt der Fläche zwischen der Kurve $f_X(t)$ und der Abszissenachse in den Grenzen $-\infty$ bis x . Die Abbildung 4.6 zeigt den Zusammenhang zwischen der Dichte- und der Verteilungsfunktion für eine spezielle stetige Zufallsgröße.

Satz 4.52. Bei einer stetigen Zufallsgröße ist die Verteilung $F_X(x)$ immer stetig, auch wenn die Dichtefunktion nur stückweise stetig ist.

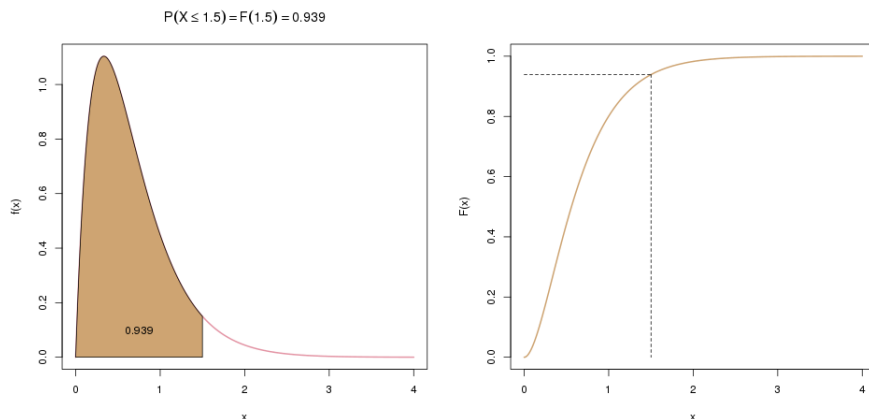


Abbildung 4.6: Links sieht man die Dichtefunktion. Die Wahrscheinlichkeit, dass X kleiner gleich 1.5 ist, entspricht der Fläche unter der Dichte bis zum Wert 1.5 auf der x -Achse. Rechts ist die Verteilungsfunktion $F(x)$ abgebildet, die genau diese Fläche darstellt.

Beispiel 4.53. Fortsetzung Beispiel 4.8: X -zufällige Lebensdauer mit den Realisierungen

$x \in (0, \infty)$. Gegeben ist $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$, wobei $\lambda > 0$ die Materialkonstante darstellt. Gesucht ist die Verteilungsfunktion $F_X(x)$:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda x} + 1 \Rightarrow F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Berechnung von Wahrscheinlichkeiten einer stetigen Zufallsgröße:

1. $P(a \leq X \leq b) = \int_a^b f_X(t) dt = F(b) - F(a)$
2. $P(X > b) = 1 - P(X \leq b) = 1 - \int_{-\infty}^b f_X(t) dt = 1 - F(b)$
3. $P(X = x) = \lim_{h \rightarrow 0^+} P(x \leq X \leq x+h) = \lim_{h \rightarrow 0^+} \int_x^{x+h} f_X(t) dt = 0$

Für eine stetige Zufallsgröße X verschwinden folglich alle Wahrscheinlichkeiten der Form $P(X = x)$, obwohl diese Ereignisse nicht mit dem unmöglichen Ereignis zusammenfallen müssen.

Weiter gilt:

- $P(X \leq a) = P(X < a) = F(a)$

- $P(X > b) = P(X \geq b) = 1 - F(b)$
- $P(a \leq x < b) = P(a < X < b) = P(a < X \leq b) = F(b) - F(a)$

Im stetigen Fall ist es jeweils nicht relevant, ob Intervalle offen, halboffen oder geschlossen sind, da sich die Wahrscheinlichkeiten nicht ändern, weil die einzelnen Punkte in a und b die Wahrscheinlichkeit Null haben, siehe 3. Achtung: Im diskreten Fall spielt dies sehr wohl eine Rolle.

4.6 Momente stetiger Zufallsgrößen

Bereits bekannt ist, dass die Momente wichtige Informationen über eine Verteilung einer Zufallsgröße liefern. Der Erwartungswert, die Varianz, die Standardabweichung und der Variationskoeffizient werden nun für die stetigen Zufallsgrößen eingeführt.

4.6.1 Das k -te Moment und der Erwartungswert

Definition 4.54. Für eine stetige Zufallsgröße X mit der Dichtefunktion $f_X(x)$ ist der Erwartungswert $\mu = E(X)$ von X definiert als

$$\mu = E(X) := \int_{-\infty}^{\infty} x f_X(x) dx, \quad (4.3)$$

falls das uneigentliche Integral absolut konvergent ist, dass heißt es gilt

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty. \quad (4.4)$$

Die Voraussetzung für die Existenz des Erwartungswertes ist also die absolute Konvergenz des Integrals (4.4). Ist diese Bedingung nicht erfüllt, so existiert kein Erwartungswert. μ charakterisiert auch hier das “Zentrum” der Verteilung.

Beispiel 4.55. Fortsetzung Beispiel 4.8: X -zufällige Lebensdauer mit den Realisierungen $x \in (0, \infty)$. Gegeben ist $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$, wobei $\lambda > 0$ die Materialkonstante

darstellt. Gesucht ist $\mu = EX$:

$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \lambda \lim_{b \rightarrow \infty} \int_0^b x e^{-\lambda x} dx = \frac{1}{\lambda} \cdot \frac{1}{\lambda}$ ist hier die mittlere Lebensdauer eines elektronischen Bauteils.

Bemerkung 4.56. Die Allgemeinen Eigenschaften des Erwartungswertes sind auch für stetige Zufallsgrößen erfüllt (zum Beispiel die Linearität des Erwartungswertes $E(aX + b) =$

$aE(X) + b$). Weiter können auch hier endliche Summen von n stetigen Zufallsgrößen X_i betrachtet werden.

Satz 4.57. Für eine nicht negative stetige Zufallsgröße (dass heißt, alle Realisierungen von X sind nicht negativ) gilt

$$\mu = E(X) = \int_0^{\infty} (1 - F(x)) dx.$$

Beispiel 4.58. Fortsetzung Beispiel 4.8: X -zufällige Lebensdauer mit den Realisierungen $x \in (0, \infty)$. Gegeben ist $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$, wobei $\lambda > 0$ die Materialkonstante darstellt.

Die Verteilungsfunktion ist $F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$. Gesucht ist der Erwartungswert unter der Anwendung des vorhergehenden Satzes 4.57.

$$\mu = \int_0^{\infty} (1 - F(x)) dx = \int_0^{\infty} 1 - (1 - e^{-\lambda x}) dx = \int_0^{\infty} e^{-\lambda x} dx = \lim_{b \rightarrow \infty} \int_0^b e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Der Erwartungswert wird nun allgemeiner über das k -te Moment analog zum diskreten Fall definiert.

Definition 4.59. Für eine stetige Zufallsgröße X mit der Dichtefunktion $f_X(x)$ ist das k -te Moment α_k ($k = 1, 2, \dots$) definiert als

$$\alpha_k = E(X^k) := \int_{-\infty}^{\infty} x^k f_X(x) dx.$$

Hinweis: Für $k = 1$ ergibt sich wieder der Erwartungswert $E(X)$.

4.6.2 Moment, zentrales Moment, Varianz

Mit Hilfe des Erwartungswertes können auch hier Kennwerte einer stetigen Verteilung definiert werden. Zunächst wird wieder das k -te zentrale Moment eingeführt, woraus sich die Varianz wieder als Spezialfall ableiten lässt.

Definition 4.60. Für eine stetige Zufallsgröße X mit der Dichtefunktion $f_X(x)$ mit dem Erwartungswert μ ist das k -te zentrale Moment β_k ($k = 1, 2, \dots$) definiert als

$$\beta_k = E(X - \mu)^k := \int_{-\infty}^{\infty} (x - \mu)^k f_X(x) dx.$$

Definition 4.61. Für eine stetige Zufallsgröße X mit der Dichtefunktion $f_X(x)$ mit dem

Erwartungswert μ ist die Varianz σ^2 definiert als zweites zentrales Moment, dass heißt

$$\sigma^2 = D^2(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx.$$

Beispiel 4.62. Fortsetzung Beispiel 4.8: X -zufällige Lebensdauer mit den Realisierungen

$x \in (0, \infty)$. Gegeben ist $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$, wobei $\lambda > 0$ die Materialkonstante

darstellt mit $\mu = \frac{1}{\lambda}$. Gesucht ist σ^2 und σ :

$$\sigma^2 = EX^2 - (EX)^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2 = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Bemerkung 4.63. Alle diese Formeln sind analog zu denen in Abschnitt 4.3 für diskrete Zufallsgrößen. Die Summen werden durch Integrale ersetzt und die Wahrscheinlichkeiten p_i durch $f_X(x) dx$. Es gelten insbesondere die gleichen Eigenschaften wie im diskreten Fall. Auch die Interpretationen bleiben unverändert.

4.7 Stetige Verteilungsfunktionen

Wie bei den diskreten Verteilungen können die meisten Zufallsexperimente mit stetigen Zufallsgrößen einem von wenigen Grundtypen von stetigen Verteilungen zugeordnet werden.

4.7.1 Die Exponentialverteilung

Die Exponentialverteilung ist eine stetige Wahrscheinlichkeitsverteilung, die durch eine Exponentialfunktion gegeben ist. Sie ist verwandt mit der Poisson-Verteilung. Bei der Poisson-Verteilung wurde nach der Wahrscheinlichkeit für das k -malige Eintreffen eines seltenen Ereignisses gefragt. Mit der Exponentialverteilung lässt sich die Wahrscheinlichkeit für eine bestimmte Wartezeit, bis ein seltenes Ereignis eintritt, berechnen. Die Exponentialverteilung wird also als Modell vorrangig bei der Beantwortung der Frage nach der Dauer von zufälligen Zeitintervallen benutzt, wie z. B.

- Dauer von Telefongesprächen,
- Dauer des radioaktiven Zerfalles,
- Zeitdauer von Reparaturen, Wartezeiten, ...

Allgemeine Bezeichnung: $X \sim \text{Exp}(\lambda)$

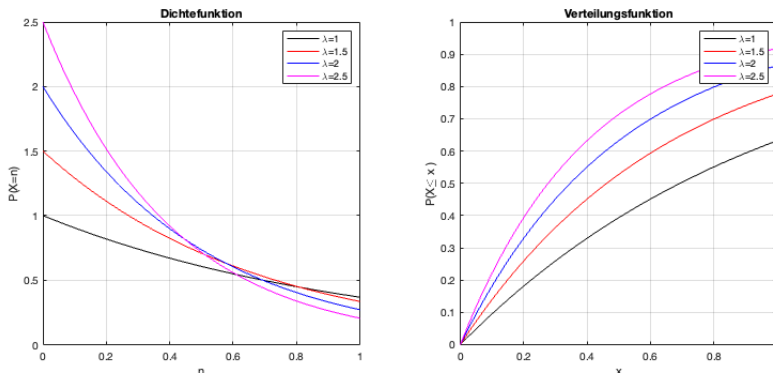


Abbildung 4.7: Links sieht man die Dichtefunktionen für $X \sim \text{Exp}(\lambda)$ mit den Werten $\lambda = 1$, $\lambda = 1.5$, $\lambda = 2$ und $\lambda = 2.5$. Rechts sind die zugehörigen Verteilungsfunktionen $F_X(x)$ abgebildet.

Definition 4.64. Eine stetige Zufallsgröße unterliegt einer Exponentialverteilung mit dem Parameter $\lambda > 0$, wenn sie die Dichtefunktion

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

besitzt.

Definition 4.65. Die Verteilungsfunktion exponentialverteilter Zufallsgröße ist definiert als

$$F_X(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Die Verteilungsfunktion wird durch den Parameter λ geprägt. λ gibt, wie bei der Poissonverteilung, die durchschnittliche Anzahl der Ereignisse pro Einheit an.

Die Abbildung 4.7 zeigt für verschiedene Werte λ mit $\lambda = 1$, $\lambda = 1.5$, $\lambda = 2$ und $\lambda = 2.5$ die Dichtefunktion und die zugehörigen Verteilungsfunktionen $F_X(x)$. Je höher der Wert λ desto größer der Anstieg der jeweiligen Funktionen.

Erwartungswert und Varianz der Exponentialverteilung:

1. $E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = \lambda \lim_{b \rightarrow \infty} \int_0^b x e^{-\lambda x} dx = \frac{1}{\lambda}$
2. $D^2(X) = E(X^2) - (E(X))^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Bemerkung 4.66. $E(X) = \frac{1}{\lambda}$ bedeutet, man kann erwarten, dass zwischen dem Eintreffen zweier Ereignisse durchschnittlich $\frac{1}{\lambda}$ Zeiteinheiten vergehen.

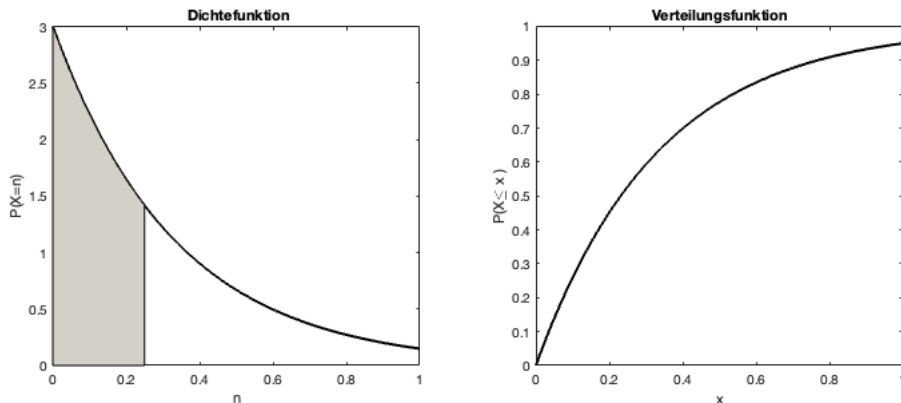


Abbildung 4.8: Links sieht man die Dichtefunktion für $X \sim \text{Exp}(3)$. Die graue Fläche entspricht der Wahrscheinlichkeit, dass X kleiner gleich 0,25 ist. Rechts ist die zugehörige Verteilungsfunktion $F_X(x)$ abgebildet.

Beispiel 4.67. Auf einem regionalen Flughafen landen in einer Stunde im Durchschnitt 3 Flugzeuge. Wie groß ist die Wahrscheinlichkeit, dass die Wartezeit zwischen zwei eintreffenden Flugzeugen weniger als 15 Minuten beträgt?

Wenn die Anzahl der in einer Stunde eintreffenden Flugzeuge als poissonverteilt angenommen werden kann, folgt die Wartezeit zwischen zwei Flugzeugen einer Exponentialverteilung mit dem Parameter $\lambda = 3$. Da 15 Minuten $\frac{1}{4}$ von einer Stunde sind, erhält man mit der Verteilungsfunktion der Exponentialverteilung $P(X \leq 0,25) = 1 - e^{-3 \cdot 0,25} = 1 - e^{-0,75} = 0,5276$. Die durchschnittliche Wartezeit beträgt $E(X) = \frac{1}{\lambda} = \frac{1}{3} h$ bzw. 20 Minuten. Die Dichte- und Verteilungsfunktion für sind in Abbildung 4.8 dargestellt. Die Wahrscheinlichkeit, dass X kleiner gleich 0,25 ist, entspricht der grauen Fläche unter der Dichte bis zum Wert 0,25 auf der x -Achse.

Nachwirkungslosigkeit der Exponentialverteilung:

Mit der Nachwirkungslosigkeit („Gedächtnislosigkeit“): ist gemeint, dass gilt

$$P(X \leq x + x_0 \mid X \geq x_0) = P(X \leq x) = F_X(x).$$

Erklärung: X sei zum Beispiel die Lebensdauer eines elektronischen Bauteils (d.h. „nicht alternden Bauteils“). Das Bauteil hat die Möglichkeit bis zu einem bestimmten Zeitpunkt x_0 zu intakt zu sein. Dann ist das Ausfallverhalten des Bauteils ab x_0 genauso wie das Ausfallverhalten eines neuen Bauteils.

Beispiel 4.68. Die zufällige Wartezeit eines Kunden am Schalter sei exponentialverteilt mit einem Erwartungswert von 10 Minuten.

(a) Wie groß ist die Wahrscheinlichkeit, dass Sie mindestens 15 Minuten warten müssen?

(b) Der Kunde hat schon 10 Minuten gewartet. Wie groß ist die Wahrscheinlichkeit, dass er insgesamt länger als 15 Minuten warten muss ?

Lösung zu (a): $X \sim \text{Exp}(\lambda)$ und $\lambda = \frac{1}{10}$

$$\leadsto P(X > 15) = 1 - P(X \leq 15) = 1 - \left(1 - e^{-\frac{1}{10} \cdot 15}\right) = e^{-\frac{1}{10} \cdot 15} \approx 0.220$$

Lösung zu (b): $P(X > 15 \mid X > 10) = P(X > 5) = e^{-5\lambda} = e^{-0.5} \approx 0.604$

4.7.2 Die Normalverteilung

Die Normalverteilung ist eine der grundlegendsten Verteilungen der Wahrscheinlichkeitstheorie und mathematischen Statistik. In zahlreichen praktischen Problemen beschreibt sie die Häufigkeitsverteilung (Wahrscheinlichkeitsfunktion) von vielen Sachverhalten in Natur, Wirtschaft und Sozialwesen. Wenn eine Häufigkeitsverteilung nicht der Normalverteilung entspricht, so kann oft unter bestimmten Bedingungen der Sachverhalt durch die Normalverteilung angenähert werden.

Allgemeine Bezeichnung: $X \sim N(\mu, \sigma)$

Definition 4.69. Eine stetige Zufallsgröße X unterliegt einer Normalverteilung mit den Parametern μ und $\sigma > 0$, wenn sie die Dichtefunktion

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)}, \quad -\infty < t < \infty.$$

Definition 4.70. Die Verteilungsfunktion normalverteilter Zufallsgröße ist definiert als

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)} dt. \quad (4.5)$$

Eigenschaften der Dichtefunktion $f_X(t)$:

1. Der Parameter μ beschreibt die Lage des „Zentrums“ der Verteilung $f_X(x)$, 4.9.
2. σ steuert die „Breite“ der Verteilung $f_X(x)$, siehe Abbildung 4.9.
3. Die Wendepunkte liegen bei $\mu \pm \sigma$.
4. Die Fläche unter $f_X(t)$ ist 1.

Erwartungswert und Varianz der Normalverteilung:

1. $E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)} dt = \mu$

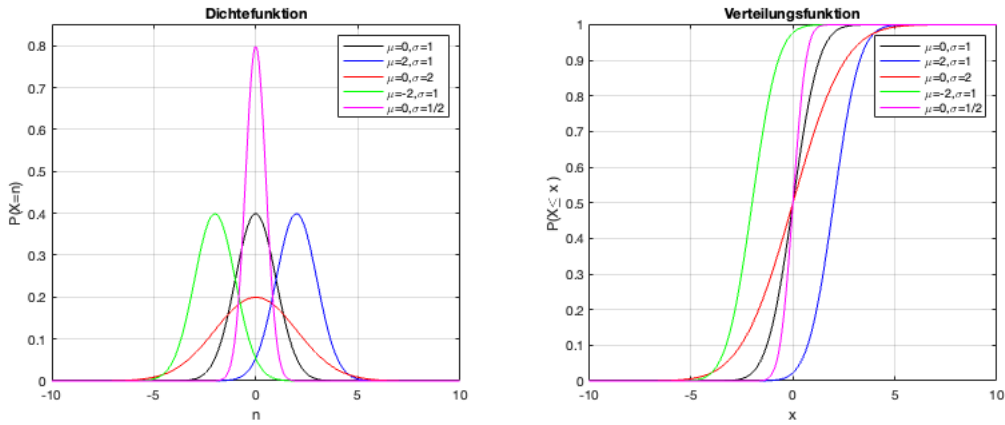


Abbildung 4.9: Links sind die Dichtefunktionen der Normalverteilung für verschieden Parameter μ und σ dargestellt. Rechts sind die zugehörigen Verteilungsfunktionen abgebildet.

Denn für $X \sim N(\mu, \sigma)$ folgt mit $Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$ und

$$E(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} te^{-\frac{t^2}{2}} dt = -\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \Big|_{-\infty}^{\infty} = 0,$$

dass $E(X) = E(\sigma Y + \mu) = \sigma E(Y) + \mu = \mu$.

$$2. D^2(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (t - \mu)^2 e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \sigma^2$$

Denn für $X \sim N(\mu, \sigma)$ folgt mit $Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$ und

$$\begin{aligned} D^2(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \left[\lim_{a, b \rightarrow \infty} -te^{-\frac{t^2}{2}} \Big|_{-a}^b + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dt \right] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \\ &= 1, \end{aligned}$$

dass $D^2(X) = D^2(\sigma Y + \mu) = \sigma^2 D^2(Y) = \sigma^2$.

4.7.2.1 Die Standardisierte Normalverteilung $N(0, 1)$

Problem: Die Verteilungsfunktion $F_X(x)$ in Gleichung 4.5 ist nicht analytisch berechenbar. Genauer besitzt das Integral der Form $\int e^{(-x^2)} dx$ keine Stammfunktion. Allerdings werden die Funktionswerte der Stammfunktion zur Berechnung der Wahrscheinlichkeiten benötigt. Ein Ausweg bietet der Übergang von der Normalverteilung zur ihrer standardisierten Form, so dass mit einer Tabelle der standardisierten Normalverteilung gearbeitet werden kann.

Um mit solch einer Tabelle arbeiten zu können, müssen Sie die Werte für X standardisiert werden. Die Normalverteilung einer Zufallsgrößen mit einem speziellen μ und σ hat im Prinzip immer die gleiche Form. Sie sind nur um den Erwartungswert μ verschoben und um den Faktor $\frac{1}{\sigma}$ in der Breite gestaucht bzw. gestreckt. Dies ist bereits in der Abbildung 4.9 deutlich geworden. Der nachfolgende Satz beschreibt die Umformung von der normalverteilten Zufallsgröße zu ihrer standardisierten Form.

Satz. Es sei X eine normalverteilte Zufallsgröße ($X \sim N(\mu, \sigma)$). Dann ist $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ verteilt.

Allgemeine Bezeichnung: $X \sim N(0, 1)$

Der Erwartungswert der Standardnormalverteilung ist $\mu = 0$ und die Standardabweichung ist $\sigma = 1$.

Definition 4.71. Die *standardisierte Normalverteilung* (Standard Normalverteilung) besitzt den Erwartungswert $\mu = 0$ und die Standardabweichung $\sigma = 1$ und die Dichtefunktion

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{t^2}{2}\right)}, \quad \forall t \in \mathbb{R}$$

bzw. Verteilungsfunktion

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{\left(-\frac{t^2}{2}\right)} dt, \quad \forall t \in \mathbb{R}.$$

Berechnung der Wahrscheinlichkeit einer Zufallsgröße $X \sim N(\mu, \sigma)$ für ...

- $P(X \leq t)$:

$$\begin{aligned} P(X \leq t) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{t - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{t - \mu}{\sigma}\right) \\ &= \phi\left(\frac{t - \mu}{\sigma}\right) \end{aligned}$$

- $P(a \leq X \leq b)$:

$$\begin{aligned}
 P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\
 &= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \\
 &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)
 \end{aligned}$$

Beispiel 4.72. X sei normalverteilt mit $\mu = 6$ und $\sigma = 2$. Bestimmen Sie (a) $P(X < 7)$ und (b) $P(6, 2 \leq X \leq 8)$.

(a) Gesucht ist $P(X < 7) = F(7)$. Mittels Standardisierung ergibt sich

$$P(X < 7) = \Phi\left(\frac{7-6}{2}\right) = \Phi\left(\frac{1}{2}\right) = 0,691462$$

(b) Gesucht ist $P(6, 2 \leq X \leq 8) = F(8) - F(6, 2)$. Mittels Standardisierung ergibt sich

$$P(6, 2 \leq X \leq 8) = \Phi\left(\frac{8-6}{2}\right) - \Phi\left(\frac{6,2-6}{2}\right) = \Phi(1) - \Phi(0,1) = 0,8413 - 0,5398 = 0,3015$$

Die nachfolgenden Sätze beschreiben zwei wichtige Eigenschaften der Normalverteilung.

Satz 4.73. 1. Die Dichtefunktion $f_X(t)$ zur $N(\mu, \sigma)$ Verteilung ist symmetrisch bezüglich μ und es gilt $F_X(\mu) = 0,5$.

Satz 4.74. Ist $X \sim N(0, 1)$ so gilt $\varphi(-x) = \varphi(x)$ und $\Phi(-x) = 1 - \Phi(x)$.

Beispiel 4.75. X sei normalverteilt mit $\mu = 6$ und $\sigma = 2$. Bestimmen Sie $P(X < 3)$.

$$P(X < 3) = F(3)$$

$$\text{Standardisierung: } P(X < 3) = \Phi\left(\frac{3-6}{2}\right) = \Phi\left(-\frac{3}{2}\right) = 1 - \Phi\left(\frac{3}{2}\right) = 1 - 0,933193 = 0,066807$$

4.7.2.2 Die 3σ -Grenze der Normalverteilung

Häufig benötigt man die Wahrscheinlichkeit dafür, dass eine normalverteilte Zufallsgröße Werte in dem symmetrisch zu μ gelegenen Intervall $[\mu - k\sigma, \mu + k\sigma]$ mit $k > 0$ annimmt. Üblicherweise wird die Abweichung von μ in Vielfachheit von σ angegeben, dass heißt in einem sogenannten $k\sigma$ -Intervall.

Satz 4.76. Es sei $X \sim N(\mu, \sigma)$ verteilt und $k \in \mathbb{R}^+$. Die Zufallsgröße Y sei $N(0, 1)$ verteilt, dann gilt

$$P(|X - \mu| < k\sigma) = P(\mu - k\sigma \leq X \leq \mu + k\sigma) = 2\Phi(k) - 1.$$

Beweis.

$$\begin{aligned}
 P(|X - \mu| < k\sigma) &= P(-k\sigma \leq X - \mu \leq k\sigma) \\
 &= P\left(-k \leq \frac{X - \mu}{\sigma} \leq k\right) \\
 &= \phi(k) - \phi(-k) \\
 &= \phi(k) - (1 - \phi(k)) \\
 &= 2\phi(k) - 1
 \end{aligned}$$

□

Beispiel 4.77. X sei der zufällige Messfehler und $N(0, 2)$ verteilt. Gesucht ist die Wahrscheinlichkeit, dass der Messfehler im Bereich $[-3, 3]$ liegt.

$P(|X| \leq 3) = P(|X - 0| < \frac{3}{2} \cdot 2)$ bzw.

$$P(-3 \leq X \leq 3) = \phi\left(\frac{3}{2}\right) - \phi\left(-\frac{3}{2}\right) = \phi\left(\frac{3}{2}\right) - (1 - \phi\left(\frac{3}{2}\right)) = 2\phi\left(\frac{3}{2}\right) - 1 = 0,8664$$

Speziell für $k = 1, 2, 3$ ergeben sich die Werte:

- $P(|X - \mu| < 1\sigma) = P(\mu - 1\sigma \leq X \leq \mu + 1\sigma) = 2\phi(1) - 1 = 2 \cdot 0,8413 - 1 = 0,6826$
- $P(|X - \mu| < 2\sigma) = P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 2\phi(2) - 1 = 2 \cdot 0,9772 - 1 = 0,9544$
- $P(|X - \mu| < 3\sigma) = P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 2\phi(3) - 1 = 2 \cdot 0,9986 - 1 = 0,9972$

Folglich liegen ca. 68% der Werte im Intervall $[\mu - \sigma, \mu + \sigma]$, ca. 95% der Werte im Intervall $[\mu - 2\sigma, \mu + 2\sigma]$ und ca. 99,7% der Werte im Intervall $[\mu - 3\sigma, \mu + 3\sigma]$. Demnach haben Werte, die außerhalb der 3σ -Grenze liegen, insgesamt nur eine Wahrscheinlichkeit von 0,003 (also ein fast unmögliches Ereignis). Solche Werte können in praktischen Anwendungen vernachlässigt werden, da fast alle Realisierungen im 3σ -Intervall liegen. Dieser Sachverhalt wird in Abbildung 4.10 verdeutlicht.

Satz 4.78. (*Lineare Transformation der Normalverteilung*) Es sei $X \sim N(\mu, \sigma)$ verteilt und $a, b \in \mathbb{R}$. Dann ist die Zufallsgröße $Y = aX + b$ eine $N(a\mu + b, |a|\sigma)$ verteilte Zufallsgröße.

Satz 4.79. Die Zufallsgrößen X_i seien $N(\mu_i, \sigma_i)$ und paarweise unabhängig $\forall i = 1, \dots, n$. Dann ist die Zufallsgröße $X = \sum_{i=1}^n X_i$ eine $N\left(\sum_{i=1}^n \mu_i, \sqrt{\sum_{i=1}^n \sigma_i^2}\right)$ verteilte Zufallsgröße.

4.8 Quantile einer Zufallsgröße

Ein Quantil ist ein weiterer Kennwert einer Zufallsgröße, welcher in der Statistik von besonderer Bedeutung ist.

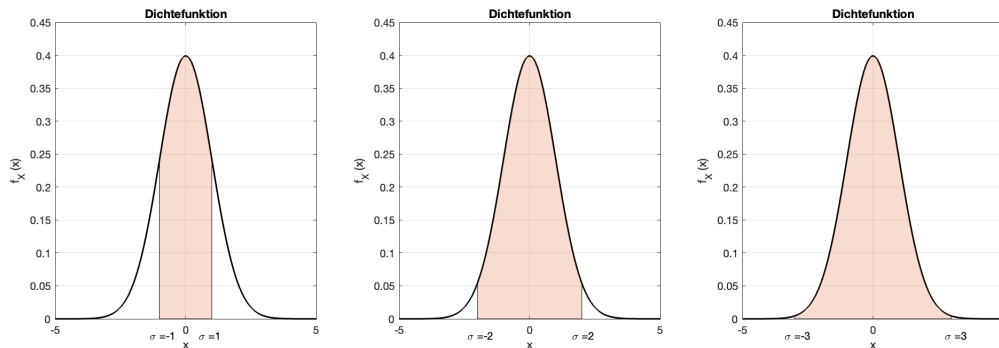


Abbildung 4.10: Die Abbildungen zeigen die Dichtefunktionen der Normalverteilung mit $\mu = 0$ und $\sigma = 1$, wobei in der ersten Abbildung die Fläche unter Kurve im Intervall $[\mu - \sigma, \mu + \sigma]$, in der zweiten Abbildung die Fläche unter Kurve im Intervall $[\mu - 2\sigma, \mu + 2\sigma]$ und in der dritten Abbildung die Fläche unter Kurve im Intervall $[\mu - 3\sigma, \mu + 3\sigma]$ gekennzeichnet ist.

Definition 4.80. Als p -tes Quantil (p -Fraktile, Quantil der Ordnung p) einer Zufallsgröße X bezeichnet man den Funktionswert c_p für den gilt

$$F(c_p - 0) \leq p \leq F(c_p)$$

bzw.

$$F(c_p) \geq p \text{ und } \lim_{t \rightarrow c_p^-} F(t) \leq p.$$

Ist die Verteilungsfunktion der Zufallsgröße stetig, die Verteilung also eine stetige Wahrscheinlichkeitsverteilung, so vereinfacht sich die Definition. Das p -Quantil ist dann eine Lösung der Gleichung

$$F_X(c_p) = P(X \leq c_p) = p.$$

Dies folgt aus der Definition des p -Quantils über die Verteilungsfunktion, da aufgrund der Stetigkeit dann mit dem Funktionswert an der Stelle c_p übereinstimmt.

Für diskrete Zufallsgrößen ist das p -Quantil nicht eindeutig. Ist X eine diskrete Zufallsgröße, so müssen für ein Quantil der Ordnung p die beiden Beziehungen $P(X < c_p) \leq p$ und $P(X \geq c_p) \geq p$ gelten. Bei diskreten Zufallsgrößen kann also der Wert ein ganzes Intervall zwischen zwei Realisierungen annehmen.

Für stetige Zufallsgrößen mit streng wachsender Verteilungsfunktion ist das p -Quantil

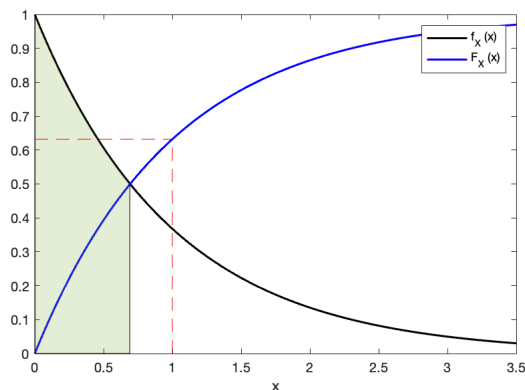


Abbildung 4.11: Die Abbildung zeigt die Dichtefunktion (schwarze Kurve) und die Verteilungsfunktion (blaue Kurve) für das Beispiel 4.81. Die grüne Fläche entspricht 50% der Gesamtfläche unter der Kurve der Dichtefunktion im Intervall $[0, 0.693]$. Die rot gestrichelte Linie ist bei dem Erwartungswert $\frac{1}{\lambda} = 1$ eingezeichnet.

eindeutig aus der Umkehrfunktion F^{-1} ermittelbar, dass heißt

$$F_X(c_p) = p \text{ bzw. } c_p = F_X^{-1}(p).$$

Spezielle Quantile: Für $p = 0,25$, $p = 0,5$ und $p = 0,75$ werden die Quantile besonders bezeichnet.

- $c_{0,5}$ - Median
- $c_{0,25}$ - unteres Quartil
- $c_{0,75}$ - oberes Quartil

Beispiel 4.81. Es sei X die zufällige Lebensdauer eines elektronischen Bauteils mit der Verteilungsfunktion $F_X(x) = 1 - e^{-x}$, $x > 0$. Diese Verteilungsfunktion ist eine Exponentialverteilung mit $\lambda = 1$. Der Erwartungswert liegt hier bei $\mu = \frac{1}{\lambda} = 1$. Gesucht

ist der Median, dass heißt $c_{0,5}$.

$$\begin{aligned} F_X(c_{0,5}) &= 0,5 \\ 1 - e^{-c_{0,5}} &= 0,5 \\ 1 - 0,5 &= e^{-c_{0,5}} \\ \ln(0,5) &= -c_{0,5} \ln(e) \\ \curvearrowright c_{0,5} &= -\ln(0,5) = 0,693 \end{aligned}$$

Wenn die Zufallsgröße X die zufällige Lebensdauer eines elektronischen Bauteils ist, dann kann μ der Erwartungswert interpretiert werden als die mittlere Lebensdauer aller Bauteile. Der Median entspricht dann der Zeitspanne, die im Mittel von der Hälfte der Bauteile „überlebt“ wird. In Abbildung 4.11 entspricht die grüne Fläche 50% der Gesamtfläche unter der Kurve der Dichtefunktion im Intervall $[0, 0.693]$. Die rote gestrichelte Linie entspricht der Interpretation des Erwartungswertes.

Quantile der Standardnormalverteilung Allgemein wird das Quantil für die Standardnormalverteilung mit $c_{N(0,1),p}$ bezeichnet. Die Wahrscheinlichkeiten in den Tabellen für die Standardnormalverteilung beginnen erst bei 0.5. Für kleinere Werte kann unter Ausnutzung der Symmetrieeigenschaft bezüglich $\mu = 0$ folgende Umformung angewendet werden:

$$c_{N(0,1),p} = -c_{N(0,1),1-p}.$$

Beispiel 4.82. $c_{N(0,1),0,98} = 2,05$

Beispiel 4.83. $c_{N(0,1),0,3} = -c_{N(0,1),0,7} = 0,52$

Kapitel 5

Grenzwertsätze

Diese Kapitel umfasst die Untersuchung des Verhaltens einer großen Zahl von zufällig wirkenden Ereignissen. Die Grenzwertsätze bilden hier den Abschluss der Wahrscheinlichkeitsrechnung und sind von zentraler Bedeutung vor allem für die Induktive Statistik. Das *Gesetz der großen Zahlen* macht eine Aussage über die Genauigkeit der Abschätzung eines unbekannten Mittelwertes der Grundgesamtheit durch den Stichprobenmittelwert. Dagegen gibt der *Zentrale Grenzwertsatz* an, gegen welche Verteilung Summen und Durchschnitte beliebig verteilter Zufallsgrößen bei großem Stichprobenumfang tendieren. Die *Tschebyscheffsche Ungleichung* wird einerseits für den Beweis des Gesetzes der großen Zahlen benötigt. Andererseits lässt sie sich aber auch eigenständig zur Abschätzung von Wahrscheinlichkeiten bei einem unbekannten Verteilungstyp verwenden.

5.1 Die Ungleichung von Tschebyscheff

In der angewandten Wahrscheinlichkeitsrechnung interessiert häufig, mit welcher Wahrscheinlichkeit sich die Zufallsvariable X in einem Intervall realisiert, das zentral (symmetrisch) um den Erwartungswert liegt. In manchen Fällen wird die Breite des Intervalls dabei oft als ein Vielfaches k der Standardabweichung ausgedrückt σ .

Sofern für eine Zufallsgröße X die Verteilung, also die Wahrscheinlichkeitsoder Dichtefunktion, bekannt ist, lässt sich die Wahrscheinlichkeit dafür bestimmen, dass X in einem bestimmten Intervall liegt. Wie ist aber eine solche Wahrscheinlichkeit zu ermitteln, wenn der Verteilungstyp unbekannt ist?

Erwartungswert und Varianz sind Parameter, die (auch wenn die Verteilung einer Zufallsvariablen X nicht bekannt ist) Rückschlüsse auf die zugrunde liegende Verteilung erlauben. So lassen sich z. B. mit Hilfe der Varianz, die ja ein Maß für die Streuung der Werte von X

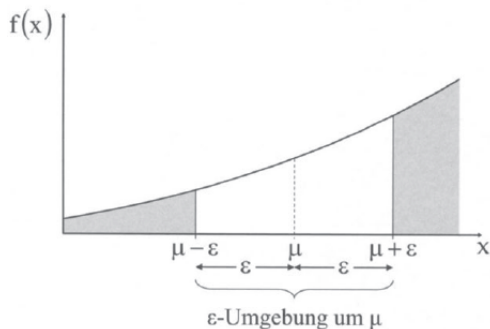


Abbildung 5.1: Die Abbildung zeigt eine beliebige Dichtefunktion einer Zufallsgröße X . ε stellt den Abstand zum Erwartungswert μ dar.

um ihren Erwartungswert $E(X)$ ist, Schranken für die Wahrscheinlichkeit des Abweichens eines Wertes der Zufallsvariablen X von ihrem Erwartungswert berechnen.

Dazu betrachten wir die Abbildung 5.1. Hervorgehoben ist ein Intervall der Breite $\pm\varepsilon$ (mit $\varepsilon \in \mathbb{R}$) um den Erwartungswert μ . Wie groß ist die Wahrscheinlichkeit, dass X außerhalb dieses Intervalls, auch ε -Umgebung genannt, liegt? Die gesuchte Wahrscheinlichkeit entspricht in der Abbildung 5.1 die grau hinterlegte Fläche.

Satz 5.1. (Ungleichung von Tschebyscheff) Es sei X eine Zufallsgröße mit $E(X) = \mu$ und $D^2(X) = \sigma < \infty$. Dann gilt

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \quad \forall \varepsilon > 0 \quad (5.1)$$

beziehungsweise

$$P(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}, \quad \forall \varepsilon > 0.$$

Mit Hilfe dieser Ungleichung kann also unter Verwendung der existierenden ersten beiden Momente die Wahrscheinlichkeit dafür abgeschätzt werden, dass die Zufallsgröße X Werte in gewissen Intervallen der reellen Achse annimmt, ohne die Verteilung von X zu kennen.

Beispiel 5.2. Ein spezielles Medikament wird in Dosen abgefüllt. Das mittlere Gewicht beträgt 10g und die Standardabweichung 1g. Wie groß ist die Wahrscheinlichkeit, dass das Gewicht einer Dose um höchstens 2g vom Sollgewicht abweicht?

Gegeben ist $\mu = 10$, $\sigma = 1$ und $\varepsilon = 2$. Wegen $P(|X - \mu| \leq \varepsilon) = 1 - P(|X - \mu| > \varepsilon) \geq$

$1 - \frac{\sigma^2}{\varepsilon^2}$ folgt

$$\begin{aligned} P(|X - 10| \leq 2) &= 1 - P(|X - 10| > 2) \\ &\geq 1 - \frac{1}{4} \\ &= \frac{3}{4}. \end{aligned}$$

Bemerkung 5.3. Da die Ungleichung von Tschebyscheff für beliebige Zufallsvariablen gilt, ist Abschätzung mit der Tschebyscheff-Ungleichung grob und sollte deshalb nur im Falle einer unbekannten Verteilungsfunktion Anwendung finden. Die Bedeutung der Ungleichung von Tschebyscheff liegt in ihrer allgemeinen Gültigkeit und damit im weiteren Theorieaufbau.

Korollar 5.4. Mit $\varepsilon = k\sigma$ folgt aus Gleichung 5.1

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad k \in \mathbb{N}$$

beziehungsweise

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}, \quad k \in \mathbb{N}.$$

Bemerkungen bezüglich der $k\sigma$ -Grenze:

1. Für $k = 1$ ergibt sich $P(|X - \mu| < 1 \cdot \sigma) \geq 0 \Rightarrow > 0\%$
2. Für $k = 2$ ergibt sich $P(|X - \mu| < 2 \cdot \sigma) \geq 1 - \frac{1}{4} = \frac{3}{4} \Rightarrow > 75\%$
3. Für $k = 3$ ergibt sich $P(|X - \mu| < 3 \cdot \sigma) \geq 1 - \frac{1}{9} = \frac{8}{9} \Rightarrow > 89\%$

Der Vergleich dieser Resultate mit den exakten Werten (unter der Annahme, dass X normalverteilt ist), siehe Abschnitt 4.7.2.2 auf Seite 67, zeigt eine große Differenz. Die Abweichung einer Zufallsgröße von ihrem Erwartungswert mit der Tschebyscheff-Ungleichung ist recht groß und somit eher ungeeignet. Die wesentliche Bedeutung zeigt sich beim Gesetz der großen Zahlen.

Beispiel 5.5. (Fortsetzung) Die Zufallsgröße X sei nun normalverteilt mit $\mu = 10$ und $\sigma = 1$. Gesucht ist

$$\begin{aligned}
P(|X - 10| \leq 2) &= P(8 \leq X \leq 12) \\
&= \Phi\left(\frac{12 - 10}{1}\right) - \Phi\left(\frac{8 - 10}{1}\right) \\
&= 2\Phi(2) - 1 \\
&= 2 \cdot 0,97725 - 1 \\
&= 0,9545.
\end{aligned}$$

Hier zeigt sich, dass die Abschätzung mit der Tschebyscheff-Ungleichung sehr grob ist und die Verteilungsinformationen genutzt werden sollten, wenn sie vorhanden sind.

5.2 Konvergenzarten für Folgen von Zufallsgrößen

In der Analysis wurden die Grenzwerte von Zahlenfolgen, Reihen von Zahlen und Funktionen betrachtet. In der Wahrscheinlichkeitsrechnung werden Grenzwertbetrachtungen an Folgen von Zufallsgrößen beziehungsweise Verteilungsfunktionen durchgeführt. Dabei unterscheidet man 4 Arten von Konvergenz:

1. „fast sicher konvergent“
2. „im quadratischen Mittel konvergent“
3. „stochastisch konvergent“
4. „konvergiert in Verteilung“

Definition 5.6. Eine Folge von Zufallsgrößen (X_n) , $n = 1, 2, \dots$ heißt *fast sicher konvergent* gegen eine Zufallsgröße X , falls

$$P\left(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right) = 1.$$

Kurze Variante: $P(\lim_{n \rightarrow \infty} X_n = X) = 1$

Bezeichnung: $X_n \xrightarrow{f.s.} X$

Interpretation: Die Menge $\Omega_0 \in \Omega$ mit $X_n(\omega) \rightarrow X(\omega) \forall \omega \in \Omega_0$. Ω_0 ist eine p -Nullmenge (Menge mit Wahrscheinlichkeitsmaß 0).

Definition 5.7. Eine Folge von Zufallsgrößen (X_n) , $n = 1, 2, \dots$ heißt *im quadratischen Mittel konvergent* gegen eine Zufallsgröße X , falls

$$\lim_{n \rightarrow \infty} E(|X_n - X|^2) = 0.$$

Bezeichnung: $X_n \xrightarrow{i.q.m.} X$

Interpretation: Die Folge der mittleren quadratischen Abweichungen konvergiert gegen Null.

Definition 5.8. Eine Folge von Zufallsgrößen (X_n) , $n = 1, 2, \dots$ heißt *stochastisch konvergent* gegen eine Zufallsgröße X , falls

$$\lim_{n \rightarrow \infty} P(\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon) = 0, \forall \varepsilon > 0$$

beziehungsweise

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1, \forall \varepsilon > 0.$$

Bezeichnung: $X_n \xrightarrow{p.} X$ (X_n konvergiert stochastisch gegen X)

Interpretation: Im Gegensatz zur fast sicheren Konvergenz braucht bei der Konvergenz der Folge der (X_n) , $n = 1, 2, \dots$ selbst nicht gegen X zu konvergieren, sondern es braucht nur die Folge der angegebenen Wahrscheinlichkeiten gegen Null zu konvergieren.

Definition 5.9. Eine Folge von Zufallsgrößen (X_n) , $n = 1, 2, \dots$ konvergiert in Verteilung gegen eine Zufallsgröße X , falls die Folge der Verteilungen (F_n) , $n = 1, 2, \dots$ der Zufallsgrößen (X_n) gegen die Verteilung F von X konvergiert.

Bezeichnung: $X_n \xrightarrow{\text{in Verteilung}} X$

Interpretation: Die Folge der (X_n) , $n = 1, 2, \dots$ braucht nicht gegen X zu konvergieren, sondern nur die Folge der Verteilungen. Dies ist die schwächste Konvergenzart.

Zusammenhang zwischen den Konvergenzarten:

- $X_n \xrightarrow{f.s.} X \Rightarrow X_n \xrightarrow{p.} X$
- $X_n \xrightarrow{i.q.m.} X \Rightarrow X_n \xrightarrow{p.} X$
- $X_n \xrightarrow{p.} X \Rightarrow X_n \xrightarrow{\text{in Verteilung}} X$

5.3 Gesetz der großen Zahlen

Bei additiver Überlagerung zufälliger Einflüsse zeigen diese mit wachsender Anzahl zunehmend deterministisches (nicht zufälliges) Verhalten. Das Gesetz der großen Zahlen untersucht das Konvergenzverhalten des arithmetischen Mittels einer Folge von Zufallsgrößen gegen den Erwartungswert der Zufallsgröße.

Anwendungen:

1. Grundlage für die Anwendung statistischer Methoden (unbekannte Kenngrößen können durch Stichproben vom Umfang n für $n \rightarrow \infty$ beliebig genau geschätzt werden.)
2. Begründung für Erscheinungen in Natur und Gesellschaft

Zum Beispiel Vielteilchensysteme in der Molekularphysik:

- Einzelnes Molekül beschreibt regellose Bewegung (BROWN'sche Bewegung)
- Anzahl der Moleküle, die je Zeiteinheit die Wand eines Gefäßes treffen sowie der Impuls, der an die Wand abgegeben wird, sind konstante Größen.

Grundanliegen: Für eine Folge (X_i) , $i = 1, 2, \dots$ von Zufallsgrößen werden Grenzaussagen für die Folge der arithmetischen Mittel \bar{X}_n , $n = 1, 2, \dots$ getroffen. Es existieren verschiedene Sätze mit unterschiedlichen Voraussetzungen an die Folge der Zufallsgrößen (X_i) . Dabei unterscheidet man zwischen zwei Arten, dem schwachen und dem starken Gesetz der großen Zahlen, welche nachfolgend grob formuliert sind.

Schwaches Gesetz der großen Zahlen

Es existiert eine Konstante a , so dass gilt $\bar{X}_n \xrightarrow{p} a$.

Starkes Gesetz der großen Zahlen

Es existiert eine Konstante a , so dass gilt $\bar{X}_n \xrightarrow{f.s.} a$.

5.3.1 Das schwache Gesetz der großen Zahlen

Satz 5.10. (SCHWACHES GESETZ DER GROSSEN ZAHLEN (CHINTCHIN)) *Es sei (X_i) , $i = 1, 2, \dots$ eine Folge unabhängig, identisch verteilte Zufallsgrößen mit dem Erwartungswert $E(X_i) = \mu < \infty$ und $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Dann gilt*

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0 \text{ für beliebiges } \varepsilon > 0, \bar{X}_n \xrightarrow{p} \mu$$

beziehungsweise

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1 \text{ für beliebiges } \varepsilon > 0, \bar{X}_n \xrightarrow{p} \mu.$$

Diese Konvergenz ist die stochastische Konvergenz gemäß dem Wahrscheinlichkeitsmaß P (nicht im üblichen Sinne der Analysis), siehe Definition 5.8. Aufgrund des schwachen Gesetzes der großen Zahlen strebt die Wahrscheinlichkeit einer Abweichung des arithmetischen Mittels \bar{X}_n vom Erwartungswert μ von weniger als ε ($\varepsilon > 0$) mit zunehmendem n gegen eins.

Bemerkung 5.11. Die Folge von Zufallsgrößen \bar{X}_n konvergiert stochastisch gegen μ . Diese stochastische Konvergenz bedeutet nicht, dass die Abweichung $\bar{X}_n - \mu$ immer kleiner wird, sondern nur, dass es bei n gegen unendlich fast sicher ist, dass die Zufallsgröße \bar{X}_n Werte in einer ε -Umgebung um μ annimmt. Oder anders ausgedrückt: Die Wahrscheinlichkeit mit der die Zufallsgröße \bar{X}_n in ein beliebig kleines Intervall $[\mu - \varepsilon; \mu + \varepsilon]$ um den Erwartungswert μ der Zufallsvariable X fällt konvergiert bei steigender Anzahl der Durchführungen des Zufallsvorgangs gegen eins. Für große n nimmt \bar{X}_n daher mit hoher Wahrscheinlichkeit Werte im Intervall $[\mu - \varepsilon; \mu + \varepsilon]$ an.

Beispiel 5.12. Ein fairer Würfel wird unendlich oft geworfen. Die Zufallsgröße X_i sei die Augenzahl bei Wurf i . Mit $S_n = X_1 + \dots + X_n$ bezeichnen wir die Summe der Augenzahlen der ersten n Würfe. Der Erwartungswert von X_i ist $E(X_i) = \frac{1}{6}(1 + \dots + 6) = 3.5$. Aus Satz 5.10 folgt nun $\frac{S_n}{n} \xrightarrow{P} 3.5$.

Bemerkung 5.13. Das Gesetz der großen Zahlen nach CHINTCHIN ist unter den Annahmen abgeleitet worden, dass die n Zufallsgrößen X_i stochastisch unabhängig und identisch verteilt sind. Eine spezielle Annahme bezüglich des Typs der Verteilung wurde dabei nicht getroffen. Nun soll ein Spezialfall betrachtet werden. Es seien nun die (X_i) , $i = 1, 2, \dots$ eine Folge binomialverteilt.

Satz 5.14. (SCHWACHES GESETZ DER GROSSEN ZAHLEN (BERNOULLI)) *Es sei (X_i) , $i = 1, 2, \dots$ eine Folge unabhängig, identisch Null-Eins verteilte Zufallsgrößen mit $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p \forall i$. Dann gilt*

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - p| \geq \varepsilon) = 0 \text{ für beliebiges } \varepsilon > 0, \bar{X}_n \xrightarrow{P} p$$

beziehungsweise

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - p| < \varepsilon) = 1 \text{ für beliebiges } \varepsilon > 0, \bar{X}_n \xrightarrow{P} p.$$

Beweis. $E(X_i) = 1 \cdot p + 0 \cdot (1 - p) = p$. Aus Satz 5.10 folgt damit $\bar{X}_n \xrightarrow{P} \mu = p$. □

Aufgrund des schwachen Gesetzes der großen Zahlen nach BERNOULLI beträgt die Wahrscheinlichkeit einer geringeren Abweichung $|\bar{X}_n - p|$ als ein beliebiges $\varepsilon > 0$ eins.

Folgerung: Es sei $H_n(A) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ die relative Häufigkeit. Damit liefert der Satz 5.14 die mathematische Begründung für die statistische Definition der Wahrscheinlichkeit.

Bemerkung 5.15. Das Schwache Gesetz großer Zahlen eröffnet also auch die Möglichkeit, relative Häufigkeiten bei unabhängigen Versuchswiederholungen (z.B. Häufigkeit des Auftretens der Ziffer 6 beim wiederholten Würfelwurf) in Zusammenhang mit den entsprechenden Wahrscheinlichkeiten im zugehörigen stochastischen Modell zu bringen. In einem Würfelexperiment mit einem fairen Würfel bezeichne A das Ereignis eine Sechs zu

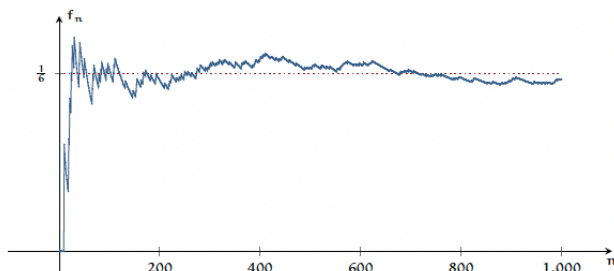


Abbildung 5.2: Die Abbildung zeigt eine 1000-fache Simulation des (unabhängigen) Würfelwurfs .

würfeln, so dass $p = P(A) = \frac{1}{6}$. In Abbildung 5.2 sind die relativen Häufigkeiten einer 1000-fachen Simulation des (unabhängigen) Würfelwurfs dargestellt. Es tritt ein sogenannter Stabilisierungseffekt ein. Aus dieser Beobachtung leitet sich die Aussage ab, dass der Ausgang eines einzelnen Zufallsexperiments zwar nicht vorhersagbar ist, das „Mittel“ des Ausgangs von vielen (unabhängigen) identischen Zufallsexperimenten aber sehr wohl prognostiziert werden kann.

5.3.2 Das starke Gesetz der großen Zahlen

Versionen des Starken Gesetzes großer Zahlen machen ebenfalls eine Aussage über die Konvergenz des arithmetischen Mittels gegen den Erwartungswert der zugrundeliegenden Verteilung, allerdings mit einer anderen (stärkeren) Konvergenz.

Satz 5.16. (ERSTES STARKES GESETZ DER GROSSEN ZAHLEN (KOLMOGOROV)) *Es sei (X_i) , $i = 1, 2, \dots$ eine Folge unabhängig, identisch verteilte Zufallsgrößen mit endlicher Varianz, dass heißt $D^2(X_i) < \infty$. Außerdem sei $\sum_{n=1}^{\infty} \frac{D^2(X_n)}{n^2} < \infty$. Dann gilt für die Summen $S_n = X_1 + \dots + X_n$*

$$\frac{S_n - E(S_n)}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0.$$

Oder anders formuliert: $P\left(\lim_{n \rightarrow \infty} \left|\bar{X}_n - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| = 0\right) = 1$.

Unter den Annahmen mit $E(X_i) = \mu$ und $D^2(X_i) = \sigma^2$ kann folgender Spezialfall abgeleitet werden.

Satz 5.17. (ZWEITES STARKES GESETZ DER GROSSEN ZAHLEN (KOLMOGOROV)) *Es sei (X_i) , $i = 1, 2, \dots$ eine Folge unabhängig, identisch verteilte Zufallsgrößen mit*

endlicher Varianz, dass heißt $D^2(X_i) < \infty$ und es gilt $\sum_{n=1}^{\infty} \frac{D^2(X_n)}{n^2} = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. Dann gilt

$$\frac{S_n - n\mu}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$$

und daraus folgt

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} \mu.$$

Bemerkung 5.18. Die fast sichere Konvergenz impliziert die stochastische Konvergenz (ist also stärker), daher die Bezeichnung starkes Gesetz der großen Zahlen.

5.4 Zentraler Grenzwertsatz

Allgemein wird zwischen globalen, lokalen und zentralen Grenzwertsätzen unterschieden.

Zentraler Grenzwertsatz: Sammelbegriff für eine Gruppe von Grenzwertsätzen, wobei Bedingungen für die Konvergenz einer Summe unabhängiger Zufallsgrößen gegen die Normalverteilung gegeben werden.

Globale Grenzwertsätze: Untersuchung zur Konvergenz von Verteilungsfunktionen.

Lokale Grenzwertsätze: Untersuchungen zur Konvergenz von Einzelwahrscheinlichkeiten beziehungsweise von Dichtefunktionen.

Anwendungen:

1. Näherungsweise Angabe der Verteilung von Stichprobenfunktionen für großen Stichprobenumfang. Zum Beispiel bei der Angabe von Konfidenzschätzungen oder Angabe der Verteilung von Testgrößen.
2. Approximation bekannter Verteilungsfunktionen für großes n durch die Normalverteilung. Zum Beispiel bei der Binomialverteilung oder bei der χ^2 - und t -Verteilung.

5.4.1 Globaler Grenzwertsatz nach LINDBERG/LEVY

Satz 5.19. Es sei (X_i) , $i = 1, 2, \dots$ eine Folge unabhängig, identisch verteilte Zufallsgrößen mit $E(X_i) = \mu$ und endlicher Varianz, dass heißt $D^2(X_i) = \sigma^2 < \infty$. Für die (zentrierten und normierten) Summen

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{\sum_{i=1}^n \sigma^2}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \quad n = 1, 2, \dots$$

mit den Verteilungsfunktionen $F_{Z_n}(x) = P(Z_n \leq x)$ gilt dann

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Bemerkung 5.20. Unter den Voraussetzungen des Satzes gilt näherungsweise

$$\sum_{i=1}^n X_i \stackrel{asympt.}{\sim} N(n\mu, \sqrt{n}\sigma)$$

beziehungsweise

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \stackrel{asympt.}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Insbesondere ist die Abweichung des arithmetischen Mittels \bar{X}_n vom Erwartungswert μ asymptotisch mit $E(\bar{X}_n - \mu) = 0$ normalverteilt:

$$(\bar{X}_n - \mu) \stackrel{asympt.}{\sim} N\left(0, \frac{\sigma}{\sqrt{n}}\right)$$

↪

$$P\left(x_1 \leq \sum_{i=1}^n X_i \leq x_2\right) \approx \phi\left(\frac{x_2 - n\mu}{\sqrt{n}\sigma}\right) - \phi\left(\frac{x_1 - n\mu}{\sqrt{n}\sigma}\right)$$

Beispiel 5.21. (X_i) , $i = 1, 2, \dots$ sei eine Folge unabhängig, identisch poissonverteilter Zufallsgrößen mit dem Parameter $\lambda = 2$:

$$p_k^{(i)} = P(X_i = k) = \frac{2^k}{k!} e^{-2}, \quad k = 0, 1, \dots, \quad i = 1, 2, \dots$$

Gesucht ist die Wahrscheinlichkeit, dass der Wert der Summe $Y_{100} = \sum_{i=1}^{100} X_i$ der Zufallsgröße zwischen 190 und 210 liegt.

Es gilt $\mu = E(X_i) = \lambda = 2$ und $\sigma = \sqrt{D^2(X_i)} = \sqrt{\lambda} = \sqrt{2}$ und $i = 1, 2, \dots$. Aus dem Grenzwertsatz 5.19 folgt Y_n ist näherungsweise $N(n\mu, \sqrt{n}\sigma)$ verteilt. Genauer Y_{100} ist näherungsweise $N(200, 10\sqrt{2})$ verteilt. Damit kann die Wahrscheinlichkeit, dass Y_{100}

zwischen 190 und 210 einen Wert annimmt bestimmt werden:

$$\begin{aligned}
 P(190 \leq Y_{100} \leq 210) &= \phi\left(\frac{210 - 200}{10\sqrt{2}}\right) - \phi\left(\frac{190 - 200}{10\sqrt{2}}\right) \\
 &= \phi\left(\frac{1}{\sqrt{2}}\right) - \phi\left(-\frac{1}{\sqrt{2}}\right) \\
 &= 2\phi\left(\frac{1}{\sqrt{2}}\right) - 1 \\
 &= 0.522.
 \end{aligned}$$

5.4.2 Globaler Grenzwertsatz von MOIVRE/LAPLACE

Satz 5.22. *Es sei (X_i) , $i = 1, 2, \dots$ eine Folge unabhängig, identisch Null-Eins-verteilter Zufallsgrößen, dass heißt $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$ mit $0 < p < 1$. Es sei weiter $Y_n = \sum_{i=1}^n X_i$ dann ist Y_n binomialverteilt mit $E(Y_n) = np$ und $D^2(Y_n) = np(1 - p)$. Für die (zentrierten und normierten) Summen*

$$Z_n = \frac{Y_n - np}{\sqrt{np(1 - p)}} \quad n = 1, 2, \dots$$

mit den Verteilungsfunktionen $F_{Z_n}(x) = P(Z_n \leq x)$. Dann gilt

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Dies ist eine Folgerung aus dem Satz von LINDBERGH/LEVY).

Bemerkung 5.23. Für große Parameter n (praktisch für $np(1 - p) > 9$) kann die Binomialverteilung durch die Normalverteilung approximiert werden. Dass heißt

$$Y_n \stackrel{asympt.}{\sim} N(np, \sqrt{np(1 - p)})$$

↪

$$P(x_1 \leq Y_n \leq x_2) \approx \phi\left(\frac{x_2 - np}{\sqrt{np(1 - p)}}\right) - \phi\left(\frac{x_1 - np}{\sqrt{np(1 - p)}}\right).$$

Bemerkung 5.24. Die Approximation kann durch eine sogenannte Stetigkeitskorrektur verbessert werden, insbesondere dann, wenn n nicht sehr groß ist:

$$P(x_1 \leq Y_n \leq x_2) \approx \phi\left(\frac{x_2 + 0.5 - np}{\sqrt{np(1 - p)}}\right) - \phi\left(\frac{x_1 - 0.5 - np}{\sqrt{np(1 - p)}}\right).$$

Beispiel 5.25. Die Zufallsgröße X sei binomialverteilt mit $n = 20$ und $p = 0.3$. Gesucht ist (a) der exakte Wert und (b) der Näherungswert nach dem Grenzwertsatz von MOIVRE/LAPLACE $P(X \geq 9)$.

$$(a) \text{ Exakte Berechnung: } P(X \geq 9) = \sum_{k=9}^{20} \binom{20}{k} 0.3^k (0.7)^{20-k} = 0.1133$$

$$(b) \text{ Näherungswert nach dem Grenzwertsatz 5.22: } E(Y_n) = np = 6 \text{ und } D^2(Y_n) = np(1-p) = 4.2 \text{ und } \sqrt{4.2} = 2.05$$

Ohne Stetigkeitskorrektur:

$$\begin{aligned} P(X \geq 9) &= P(9 \leq X \leq 20) \\ &\approx \phi\left(\frac{20-6}{2.05}\right) - \phi\left(\frac{9-6}{2.05}\right) \\ &= \phi(6.8293) - \phi(1.4634) \\ &= 0.07214. \end{aligned}$$

Mit Stetigkeitskorrektur:

$$\begin{aligned} P(X \geq 9) &= P(9 \leq X \leq 20) \\ &\approx \phi\left(\frac{20+0.5-6}{2.05}\right) - \phi\left(\frac{9+0.5-6}{2.05}\right) \\ &= \phi(7.073) - \phi(1.220) \\ &= 0.1112. \end{aligned}$$

Wie gut diese Approximation mit einer Normalverteilung für ein gegebenes n ist, hängt von der Verteilung der X_i ab. Zusätzlich gilt natürlich, dass die Approximation mit größer werdendem n immer besser wird. Selbst wenn die Verteilung der X_i nicht bekannt ist, erhält man Informationen von der approximativen Verteilung von Y_n . Der zentrale Grenzwertsatz ist unter anderem ein Grund für die Wichtigkeit der Normalverteilung.

5.4.3 Lokaler Grenzwertsatz von MOIVRE/LAPLACE

Satz 5.26. *Unter den gleichen Bedingungen, wie im obigen Satz 5.22 und mit*

$$\Delta x = \frac{(k+1) - np}{\sqrt{np(1-p)}} - \frac{k - np}{\sqrt{np(1-p)}} = \frac{1}{\sqrt{np(1-p)}}$$

gilt

$$\lim_{n \rightarrow \infty} \frac{p_k^{(n)}}{\varphi\left(\frac{k-np}{\sqrt{np(1-p)}}\right) \cdot \Delta x} = 1.$$

Dabei ist $p_k^{(n)} = P(Y_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, \dots, n$, $n \in \mathbb{N}$ die Einzelwahrscheinlichkeit der Binomialverteilung und $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ die Dichtefunktion der Standard-Normalverteilung.

Bemerkung 5.27. Für große n können die Einzelwahrscheinlichkeiten der Binomialverteilung durch die Dichte der Normalverteilung approximiert werden:

$$P(Y_n = k) \approx \varphi\left(\frac{k-np}{\sqrt{np(1-p)}}\right) \cdot \Delta x.$$

Beispiel 5.28. Die Zufallsgröße X genüge einer Binomialverteilung mit $n = 10$ und $p = 0.5$. Gesucht ist (a) der exakte Wert und (b) der Näherungswert nach dem lokalen Grenzwertsatz von MOIVRE/LAPLACE für $P(X = 2)$.

(a) Exakte Berechnung: $P(X = 2) = \binom{10}{2} 0.5^2 (0.5)^8 = 0.0439$

(b) Näherungswert nach dem Grenzwertsatz 5.26: $E(X) = np = 5$ und $D^2(X) = np(1-p) = 2.5$

$$P(X = 2) \approx \varphi\left(\frac{2-5}{\sqrt{2.5}}\right) \cdot \frac{1}{\sqrt{2.5}} = \varphi(1.8975) \cdot \frac{1}{1.581} = 0.0417.$$

Interpretation: $n = 10$ und $p = 0.5 \Rightarrow np(1-p) = 2.5 < 9 \Rightarrow$ Grenzwertsatz nach MOIVRE/LAPLACE sollte nicht angewendet werden. Da jedoch $p = 0.5$ ist die Binomialverteilung symmetrisch bezüglich $E(X)$, wie die Dichte der Normalverteilung \Rightarrow Die Abweichung ist relativ gering.

Teil III

Statistik

Kapitel 6

Beschreibende Statistik

Die mathematische Statistik untersucht die Eigenschaften zufälliger Erscheinungen auf der Basis von Messreihen zufälliger Variablen. Jede Aussage der mathematischen Statistik beruht auf gewissen Voraussetzungen. Ohne Angabe dieser Voraussetzungen ist die Aussage wertlos.

Aufgabenstellung der mathematischen Statistik: Eine Zufallsgröße X wird in einer Versuchsfolge n mal gemessen. Daraus ergibt sich eine Messreihe oder Urliste (Stichprobe) x_1, x_2, \dots, x_n .

Annahme: Messungen x_i sind voneinander unabhängig. Das Messergebnis variiert von Stichprobe zu Stichprobe. Daraus folgt, dass die Urliste (x_1, x_2, \dots, x_n) eine konkrete Realisierung des Zufallsvektors (X_1, X_2, \dots, X_n) , wobei

- die Zufallsgrößen X_k , ($k = 1, \dots, n$) unabhängig sind und
- alle X_k , haben die gleiche Verteilungsfunktion wie die Zufallsgröße X .

Man unterscheidet in der Statistik zwischen der deskriptiven (beschreibenden) und induktiven (beurteilenden) Statistik.

Aufgabe der deskriptiven Statistik: Das Ziel in diesem Gebiet ist das Sammeln, Ordnen und Darstellen von Daten in Tabellen und Grafiken sowie die Berechnung von Stichprobenkennwerten (Mittelwert, Streuung,...).

Aufgabe der induktiven Statistik: Anhand einer Stichprobe sollen Rückschlüsse über die unbekannten Eigenschaften der zu untersuchenden Objekte gezogen werden.

Dies impliziert die Ermittlung der unbekannten Verteilungsfunktion der Zufallsgröße X bzw. wichtiger Parameter (Mittelwert, Streuung,...) dieser Verteilung.

6.1 Grundlegende Begriffe der deskriptiven Statistik

Merkmalsträger sind die Objekte, die hinsichtlich verschiedener Eigenschaften Gemeinsamkeiten aufweisen und daher zusammen betrachtet werden. Diese Eigenschaften dienen dazu, die Objekte vom Rest der Daten, der nicht betrachtet werden soll, abzugrenzen. Die Abgrenzung von Merkmalsträgern kann sachlich, zeitlich oder räumlich sein.

Beispiel 6.1. Es sollen alle männlichen Studierenden des Fachbereichs Sozialwesen an der Hochschule Niederrhein im Sommersemester 2013 betrachtet werden:

Sachliche Abgrenzung: alle männlichen Studierenden des Fachbereichs Sozialwesen an der Hochschule

Zeitliche Abgrenzung: Sommersemester 2013

Räumliche Abgrenzung: Hochschule in Neustadt

Oft werden die Merkmalsträger durchnummeriert und bekommen den Buchstaben i als Index.

Merkmale bzw. Variablen sind Eigenschaften, die im Blickpunkt einer statistischen Untersuchung stehen und die den Merkmalsträgern zugeordnet werden können. Beispiele wären Haarfarbe oder die BAföG-Höhe. Merkmale werden mit Großbuchstaben gekennzeichnet, z. B. X .

Definition 6.2. Eine *Grundgesamtheit* ist eine Menge von Objekten, für die ein bestimmtes Merkmal X untersucht wird.

Von Grundgesamtheiten spricht man, wenn alle Merkmalsträger betrachtet werden, die aufgrund der sachlichen, zeitlichen und räumlichen Abgrenzung in Frage kommen. Wenn nur ein Teil der Merkmalsträger untersucht wird, spricht man nicht mehr von der Grundgesamtheit, sondern von einer Stichprobe.

Definition 6.3. Zufallsstichprobe (mathematische Stichprobe) vom Umfang n : An n zufällig ausgewählten Objekten einer Grundgesamtheit wird das Merkmal X gemessen bzw. beobachtet. Es ergibt sich ein Beobachtungsvektor (X_1, X_2, \dots, X_n) von unabhängigen und identisch verteilten Zufallsgrößen X_i , welche die gleiche Verteilung wie das Merkmal X besitzen. Die Zufallsgrößen X_i , ($i = 1, \dots, n$) heißen Stichprobenvariable.

Definition 6.4. Eine Realisierung (x_1, x_2, \dots, x_n) der Zufallsstichprobe (X_1, X_2, \dots, X_n) heißt *konkrete Stichprobe*. Jede Realisierung $x_i, i = 1, \dots, n$ heißt ein *Element der Stichprobe*.

Definition 6.5. Eine messbare Funktion $T = T(X_1, X_2, \dots, X_n)$ der zur Stichprobe gehörenden Zufallsgrößen X_1, X_2, \dots, X_n heißt *Stichprobenfunktion*. Dabei ist T selbst eine Zufallsgröße.

Definition 6.6. Eine Realisierung $t = t(x_1, x_2, \dots, x_n)$ der Stichprobenfunktion T für eine konkrete Stichprobe (x_1, x_2, \dots, x_n) heißt *konkrete Stichprobenfunktion*.

6.1.1 Deskriptive Statistik für eindimensionale Daten

Allgemein unterscheidet man zwischen *diskreten* und *stetigen Merkmalen* (analog zu stetigen und diskreten Zufallsgrößen). Ein diskretes Merkmal könnte zum Beispiel die zufällige Note in einer Prüfung sein und ein stetiges Merkmal die Geschwindigkeit von Fahrzeugen, die bei einer Radarkontrolle registriert werden. Diskrete Merkmale können allerdings auch aus stetigen Merkmalen durch Runden der Werte oder durch Klassenbildung entstehen.

Ausgehend von einer Stichprobe (x_1, x_2, \dots, x_n) des Merkmals X können bestimmte statistische Kennwerte ermittelt werden.

6.1.1.1 Häufigkeitstabellen und Häufigkeitsverteilungen

Für die Berechnung bestimmter Kenngrößen werden die Merkmalsausprägungen der Größe nach geordnet. Damit entsteht eine geordnete Stichprobe.

Definition 6.7. Geordnete Stichprobe: Die Merkmalsausprägungen werden der Größe nach geordnet, dass heißt es entsteht $x_1^* < x_2^* < \dots < x_n^*$.

Mit der geordneten Stichprobe können die Häufigkeiten x_i^* bestimmt werden. Dabei unterscheidet man zwischen der absoluten, der relativen und der relativen Summenhäufigkeit.

Definition 6.8. absolute Häufigkeit von x_i^* : $n_i, i = 1, \dots, m$. Dabei gilt $\sum_{i=1}^m n_i = n$.

Definition 6.9. relative Häufigkeit von x_i^* : $h_i = \frac{n_i}{n}, i = 1, \dots, m$, wobei $0 \leq h_i \leq 1, \forall i$.

Definition 6.10. relative Summenhäufigkeit von x_i^* : $\sum_{j=1}^i h_j, i = 1, \dots, m$. Dabei gilt $\sum_{j=1}^m h_j = 1$.

Definition 6.11. empirische Verteilung von X für eine konkrete Stichprobe vom Umfang n :

$$\hat{F}_n(x) = \frac{\text{Anzahl der Merkmalsausprägungen} \leq x}{n} = \sum_{\substack{i \\ x_i^* \leq x}} h_i, \forall x \in \mathbb{R}.$$

Beispiel 6.12. Bei einer Prüfung von 25 Studierenden ergaben sich folgende Noten:

3 5 4 3 2 3 6 4 1 2 3 3 4 5 2 1 3 4 2 4 3 1 2 3 4

Merkmal X : zufällige Note (diskrete Zufallsgröße)

Merkmalsausprägungen: konkrete Noten

Note	absolute Häufigkeit	relative Häufigkeit	relative Summenhäufigkeit
1	3	0,12	0,12
2	5	0,20	0,32
3	8	0,32	0,64
4	6	0,24	0,88
5	2	0,08	0,96
6	1	0,04	1
Σ	25	1	–

$$\text{empirische Verteilungsfunktion: } \hat{F}_{25}(x) = \begin{cases} 0, & x < 1 \\ 0.12, & 1 \leq x < 2 \\ 0.32, & 2 \leq x < 3 \\ 0.64, & 3 \leq x < 4 \\ 0.88, & 4 \leq x < 5 \\ 0.96, & 5 \leq x < 6 \\ 1, & x \geq 6 \end{cases}$$

Die vorliegenden Daten können allerdings auch in Klassen eingeteilt sein. Alle Merkmalsausprägungen der Grundgesamtheit liegen in einem bestimmten Intervall auf der reellen Achse. Zerlegt man nun das Intervall in Teilintervalle, so entstehen sogenannte Klassen. Die Klassen entsprechen also sogenannten Teilintervallen. Eine Klasseneinteilung wird auch bei stetigen Merkmalen vorgenommen. Bei klassierten Daten werden die oben eingeführten Kenngrößen ähnlich definiert.

Gegeben sei eine Zerlegung des Intervalls in k Teilintervalle (Merkmalsklassen). Hierbei gilt die Faustregel $k \approx \sqrt{n}$ bzw. $k \leq 5 \lg(n)$.

Definition 6.13. Klassenbreite: $d_i = x_{ir} - x_{il}$, $i = 1, \dots, k$, wobei x_{ir} der rechte Rand der i -ten Klasse ist und x_{il} der linke Rand.

Definition 6.14. Klassenmitte: $\bar{x}_i = \frac{x_{ir} + x_{il}}{2}$, $i = 1, \dots, k$

Definition 6.15. absolute Klassenhäufigkeit in der Klasse i : n_i , $i = 1, \dots, k$.

Definition 6.16. relative Klassenhäufigkeit in der Klasse i : $h_i = \frac{n_i}{n}$, $i = 1, \dots, k$.

Definition 6.17. relative Summenhäufigkeit bis zur Klasse i : $\sum_{j=1}^i h_j, i = 1, \dots, k$.

Definition 6.18. empirische Verteilungsfunktion von X bei Klasseneinteilung:

$$\hat{F}_n(x) = \sum_{\substack{i \\ x_{ir}^* \leq x}} h_i, \forall x \in \mathbb{R}.$$

Beispiel 6.19. Bei einer Radarkontrolle wurden die Geschwindigkeiten derjenigen Fahrzeuge registriert, welche 55 km/h überschritten. Die ersten 100 Verkehrssünder wurden notiert, wobei keiner schneller als 100 km/h gefahren ist. Für die Höhe des Verwarnungsbetrages sind 3 Klassen zu bilden:

- mehr als 55 km/h , aber höchstens 60 km/h
- mehr als 60 km/h , aber höchstens 70 km/h
- mehr als 70 km/h , aber höchstens 100 km/h.

Merkmal X : zufällige Geschwindigkeit (stetige Zufallsgröße)

Merkmalsausprägungen: konkrete Geschwindigkeiten

Daraus ergeben sich folgende Häufigkeitsverteilungen:

Klasse	Intervalle	\bar{x}_i	d_i	abs. KH	rel. KH	rel. SH
1	(55; 60]	57,5	5	58	0,58	0,58
2	(60; 70]	65	10	30	0,3	0,88
3	(70; 100]	85	30	12	0,12	1
				100	1	

$$\text{empirische Verteilungsfunktion: } \hat{F}_{100}(x) = \begin{cases} 0, & x \leq 55 \\ 0.58, & 55 < x \leq 60 \\ 0.88, & 60 < x \leq 70 \\ 1, & x > 70 \end{cases}. \text{ Da für einen Stichpro-}$$

benumfang $n = 100$ die Anzahl $k = 3$ Klassen zur gering ist, geht viel Information verloren. $\hat{F}_{100}(x)$ ist nur eine sehr grobe Schätzung (Näherung) von $F_X(x)$.

6.1.1.2 Grafische Darstellungsmöglichkeiten für Häufigkeiten

Häufigkeiten können in Form von Histogrammen, Stabdiagrammen, Häufigkeitspolygon selten in Tortendiagrammen dargestellt werden. Die relativen Summenhäufigkeiten werden mittels einer Treppenfunktion visualisiert.

Histogramm (Balkendiagramm): Diese Darstellung ist geeignet für absolute und relative Häufigkeiten. Im Falle eines stetigen Merkmals werden die absoluten bzw. relativen

Häufigkeiten mit einem flächenproportionalen Histogramm, dass heißt über jeder Klasse wird ein Rechteck mit der Höhe $\frac{n_i}{d_i}$ bzw. $\frac{h_i}{d_i}$, $i = 1, \dots, k$ errichtet.

Häufigkeitspolygon: Für die absoluten und relativen Häufigkeiten über x_i^* wird jeweils ein Punkt der Höhe h_i bzw. n_i $i = 1, \dots, m$ errichtet, die dann durch einen Polygonenzüge verbunden werden.

Stabdiagramm: Für die absoluten und relativen Häufigkeiten (insbesondere für ein diskretes Merkmal X) über x_i^* wird jeweils ein Stab der Höhe h_i bzw. n_i $i = 1, \dots, m$ errichtet.

Treppenkurve: Diese Darstellung wird für die relativen Summenhäufigkeiten verwendet. Die Sprungstellen bei diskreten Merkmalen sind x_i^* , $i = 1, \dots, m$ beziehungsweise bei klassierten Daten x_{ir} , $i = 1, \dots, k$. Diese Sprungstellen werden mit horizontalen Linien verbunden, dass heißt $\sum_{j=1}^i h_j$.

Kleine Übungsaufgabe: Visualisieren Sie für die Häufigkeiten der obigen Beispiele 6.12 und 6.19.

6.1.1.3 Mittelwerte (Lageparameter)

Die Schätzung von μ ist mittels verschiedener Lageparameter möglich (arithmetisches Mittel, empirischer Median, empirischer Modalwert).

Definition 6.20. arithmetisches Mittel für diskrete Daten: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ bzw. für klassierten Daten $\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i \bar{x}_i$ mit k Klassen.

Das arithmetische Mittel ist die beste Schätzung für μ , aber Ausreißer empfindlich. Daher kann auch der Median beziehungsweise der Modalwert zur Schätzung von μ verwendet werden, da diese robust gegenüber Ausreißern sind.

Definition 6.21. empirischer Median (Zentralwert): Gegeben sei eine geordnete Stichprobe, dann ist der Median definiert als

$$\bar{x}_{0,5} = \begin{cases} x_{\frac{n+1}{2}}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{falls } n \text{ gerade} \end{cases}$$

Der Modus oder Modalwert ist ein einfach zu berechnender Lageparameter. Es ist der Wert einer Urliste, der am häufigsten vorkommt.

Definition 6.22. empirischer Modalwert: \bar{x}_M ist der Wert mit der größten Häufigkeit Die Aussagekraft von $\bar{x}_{0,5}$ und \bar{x}_M ist geringer als jene von \bar{x} , da nur ein Teil der Information aus der Stichprobe enthalten ist. Gibt es mehrere Merkmalsausprägungen mit der gleichen maximalen Häufigkeit, so existiert kein Modalwert.

Definition 6.23. geometrisches Mittel für diskrete Daten: $\bar{x}_G = \sqrt[n]{\prod_{i=1}^n x_i}, \forall x_i > 0$

Im Gegensatz zum arithmetischen Mittel ist das geometrische Mittel nur für nichtnegative Zahlen definiert und meistens nur für echt positive reelle Zahlen sinnvoll, denn wenn ein Faktor gleich null ist, ist das ganze Produkt gleich null.

Bemerkung 6.24. Die Berechnung des geometrischen Mittels ist sinnvoll wenn relative Änderungen gemittelt werden sollen, dass heißt nicht die Differenz sondern das Verhältnis zwischen den Messwerten wesentlich ist. Angewendet wird dies zum Beispiel zur Berechnung von Wachstumsfaktoren in der Wirtschaftsstatistik.

Beispiel 6.25. Ein Angestellter erhält in 3 aufeinander folgenden Jahren Gehaltssteigerungen von 6%, 10% und 2%. Welche durchschnittliche jährliche Gehaltssteigerung hat er in diesen Jahren?

Lösung: Das Gehalt steigt jeweils auf das 1.06, 1.10 und 1.02 fache des Jahresgehaltes an.

$$\bar{x}_G = \sqrt[3]{1.06 \cdot 1.10 \cdot 1.02} = 1.0595$$

Für das arithmetische Mittel ergibt sich $\bar{x} = \frac{1}{3} (1.06 + 1.10 + 1.02) = 1.06$.

6.1.1.4 Streuungsmaße

Definition 6.26. empirische Streuung für diskrete Daten:

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

bzw. für klassierten Daten

$$s^2 = \frac{1}{n-1} \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^k (n_i \bar{x}_i^2 - n\bar{x}^2)$$

mit k Klassen.

Definition 6.27. empirische Standardabweichung: $s = +\sqrt{s^2}$

Definition 6.28. empirischer Variationskoeffizient: $v = \frac{s}{\bar{x}}, \bar{x} \neq 0$

Der Variationskoeffizient v stellt die relative Standardabweichung dar. Der Variationskoeffizient v dient zum Vergleich von Streuungen, die zu verschiedenen Mittelwerten gehören.

Definition 6.29. mittlere absolute Abweichung: $s_a = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$

Definition 6.30. Spannweite: $s_w = \max_i x_i - \min_i x_i$

Beispiel 6.31. (Fortsetzung 6.12)

Stichprobenmittelwert (arithmetisches Mittel):

$$\bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i = \frac{1}{25} (3 \cdot 1 + 5 \cdot 2 + 8 \cdot 3 + 6 \cdot 4 + 2 \cdot 5 + 1 \cdot 6) = 3.08$$

Stichprobenstreuung (empirische Streuung):

$$s^2 = \frac{1}{24} \sum_{i=1}^{25} (x_i^2 - n\bar{x}^2) = \frac{1}{24} (3 \cdot 1^2 + 5 \cdot 2^2 + 8 \cdot 3^2 + 6 \cdot 4^2 + 2 \cdot 5^2 + 1 \cdot 6^2 - 25 \cdot 3.08^2) = 1.66$$

empirische Standardabweichung: $s = +\sqrt{s^2} = \sqrt{1.66} = 1.28841$

empirischer Variationskoeffizient: $v = \frac{s}{\bar{x}} = \frac{1.28841}{3.08} = 0.418$

Beispiel 6.32. (Fortsetzung 6.19)

Stichprobenmittelwert (arithmetisches Mittel):

$$\bar{x} = \frac{1}{100} \sum_{i=1}^{100} x_i = \frac{1}{100} (58 \cdot 57,5 + 30 \cdot 65 + 12 \cdot 85) = 63.05$$

Stichprobenstreuung (empirische Streuung):

$$s^2 = \frac{1}{99} \sum_{i=1}^{100} (x_i^2 - n\bar{x}^2) = \frac{1}{99} (58 \cdot 57,5^2 + 30 \cdot 65^2 + 12 \cdot 85^2 - 100 \cdot 63,05^2) = 77.598$$

empirische Standardabweichung: $s = +\sqrt{s^2} = \sqrt{77.598} = 8.809$

empirischer Variationskoeffizient: $v = \frac{s}{\bar{x}} = \frac{8.809}{63.05} = 0.14$

6.1.2 Deskriptive Statistik für zweidimensionale Daten

Oft ist man zwei oder mehr Merkmalen gleichzeitig interessiert und möchte Zusammenhänge zwischen den Merkmalen herausbekommen. In diesem Fall beschäftigt man sich mit mehrdimensionalen Daten. Nachfolgend wird das Problem auf zweidimensionale Daten beschränkt. Mehr Dimensionen können analog behandelt werden.

Gegeben ist eine Stichprobe (Urliste) von Messwerten $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ der beiden Merkmale X und Y . Die Wertepaare können mit einem Streuungsdiagramm visualisiert werden. Ein Streuungsdiagramm ist eine zwei dimensionale grafische Darstellung als „Punktwolke“ aller Wertepaare im xy -Koordinatensystem, siehe Abbildung 6.1.

Bei beiden Merkmalen X und Y kann eine Klasseneinteilung vorliegen. Es sei X ein Merkmal mit k Klassen und Y ein Merkmal mit l Klassen mit den Klassenbreiten d_x und d_y sowie den Klassenmitten \bar{x}_i und \bar{y}_j . Nachfolgende Kennwerte können für zweidimensionale klassierte Daten bestimmt werden:

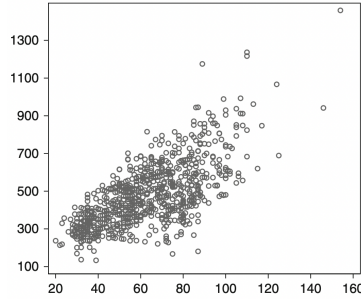


Abbildung 6.1: Diese Abbildung zeigt ein Streuungsdiagramm (Punktwolke) für zwei Merkmale X und Y .

absolute Häufigkeit der Klasse (i, j): n_{ij} mit $i = 1, \dots, k, j = 1, \dots, l$, wobei gilt $\sum_{i=1}^k \sum_{j=1}^l n_{ij} = n$
absolute Häufigkeit von X der Klasse i: $n_{i\bullet} = \sum_{j=1}^l n_{ij}$, wobei gilt $\sum_{i=1}^k n_{i\bullet} = n$
absolute Häufigkeit von Y der Klasse j: $n_{\bullet j} = \sum_{i=1}^k n_{ij}$, wobei gilt $\sum_{j=1}^l n_{\bullet j} = n$
relative Häufigkeit der Klasse (i, j): $h_{ij} = \frac{n_{ij}}{n}$ mit $i = 1, \dots, k, j = 1, \dots, l$, wobei gilt $\sum_{i=1}^k \sum_{j=1}^l h_{ij} = 1$
relative Häufigkeit von X der Klasse i: $h_{i\bullet} = \sum_{j=1}^l h_{ij}$, wobei gilt $\sum_{i=1}^k h_{i\bullet} = 1$
relative Häufigkeit von Y der Klasse j: $h_{\bullet j} = \sum_{i=1}^k h_{ij}$, wobei gilt $\sum_{j=1}^l h_{\bullet j} = 1$

Statistische Maße des Zusammenhanges zweier Merkmale

Die einfachste Form der Abhängigkeit ist die lineare Abhängigkeit. Diese wird durch die empirische Korrelation erfasst, welche auf der empirischen Kovarianz basiert.

Definition 6.33. Empirische Kovarianz:

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right)$$

Der Wert der empirischen Kovarianz kann etwas über die Abhängigkeit zwischen den beiden Merkmalen aussagen. Der Nachteil besteht aber darin, dass die empirische Kovarianz von den Einheiten abhängt, in denen die einzelnen Merkmale gemessen wurde. Daher normiert man die empirische Kovarianz mit den Standardabweichungen der beiden Merkmale und erhält so die empirische Korrelation, auch genannt empirischer Korrelationskoeffizient.

Definition 6.34. empirischer Korrelationskoeffizient nach PEARSON:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2) (\sum_{i=1}^n y_i^2 - n \bar{y}^2)}},$$

wobei s_X Standardabweichung für X und s_Y Standardabweichung für Y ist.

Bemerkung 6.35. Der Korrelationskoeffizient nimmt Werte zwischen ± 1 an, dass heißt $-1 \leq r \leq 1$, welcher wie folgt zu interpretieren ist:

- $r_{XY} > 0$: X und Y sind positiv korreliert, dass heißt mit wachsenden x Werten wachsen auch die y Werte.
- $r_{XY} < 0$: X und Y sind negativ korreliert, dass heißt mit wachsenden x Werten fallen die y Werte.
- $r_{XY} = 0$: X und Y sind unkorreliert, kein linearer Zusammenhang zwischen X und Y erkennbar:
- $|r_{XY}| \approx 1$: starke lineare Abhängigkeit zwischen X und Y

Beispiel 6.36. Von 20 zufällig ausgewählten Personen wurde die Körpergröße X (in cm) und das Körpergewicht (in kg) notiert. Die auf ganze Zahlen gerundeten Messwerte sind in der nachfolgenden Tabelle aufgelistet:

X	170	162	171	178	175	165	169	173	182	176
Y	74	61	68	81	73	62	71	73	83	78
X	160	167	171	163	179	170	173	168	177	166
Y	59	69	72	65	76	75	71	72	75	71

arithmetisches Mittel der x_i : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 170,75 \text{ cm}$

arithmetisches Mittel der y_i : $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 71,45 \text{ kg}$

empirische Standardabweichung der x_i : $s_X = +\sqrt{s^2} \approx 5,964 \text{ cm}$

empirische Standardabweichung der y_i : $s_Y = +\sqrt{s^2} \approx 6,236 \text{ kg}$

empirische Kovarianz:

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{19} \sum_{i=1}^{20} (x_i - 170,75)(y_i - 71,45) = 33,54$$

$$\text{empirischer Korrelationskoeffizient: } r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{33,54}{5,964 \cdot 6,236} = 0,902$$

Bemerkung 6.37. Im Falle $r_{XY} = 0$ sind die Merkmale X und Y unkorreliert. Dies heißt aber nicht, dass sie keinen Zusammenhang hätten, sondern nur dass sie keinen linearen Zusammenhang haben.

Satz 6.38. Für zwei unabhängige Merkmale X und Y gilt $r = 0$. Allerdings folgt aus $r = 0$ nicht die Unabhängigkeit von X und Y .

Hinweis: Ein Zusammenhang zwischen X und Y sollte nur dort betrachten werden, wo zwischen beiden Merkmalen ein kausaler Zusammenhang erkennbar ist. Sonst könnten sogenannte Scheinkorrelationen („Scheinkausalität“) auftreten.

Beispiel 6.39. Es sei X die Anzahl der Neugeborenen einer Region und Y die Anzahl der Störche in einer Region. Der Korrelationskoeffizient könnte hier $r \approx 1$ sein, obwohl kein kausaler Zusammenhang besteht.

6.2 Regressionsanalyse

Eine weitergehende Analyse von Merkmalen, die den Zusammenhang beschreibt, ist die Regressionsanalyse. Genauer, die Regressionsanalyse befasst sich mit der Untersuchung des funktionellen Zusammenhangs zwischen einer abhängigen Variablen und ein oder mehrerer unabhängiger Variablen. Dabei stellt sich die Frage: *Wie stark und auf welcher Art ist der funktionelle Einfluss der unabhängigen Variablen auf die abhängige Variable?* Dies fällt in den Bereich Ursachenanalyse. Eine weitere Frage bezüglich der Wirkungsprognose ist: *Wie verändert sich die abhängige Variable bei der Änderung der unabhängigen Variable?* Und im Kontext der Zeitreihenanalyse ist die Frage: *Wie ändert sich die abhängige Variable im zeitlichen Verlauf und damit in Zukunft?*

6.2.1 Allgemeines Vorgehen bei einer Regressionsanalyse

Gegeben ist eine abhängige Variable Y (Zielvariable, Regressand) und j (mit $j \geq 1$) unabhängige Variablen X_1, \dots, X_j (Einflussvariablen, Regressoren).

Gesucht ist eine Regressionsfunktion g in einer Klasse spezieller vorgegebener Funktionen, die den funktionalen Zusammenhang. Zwischen Y und X_1, \dots, X_j „bestmöglich“ beschreibt im Sinne der kleinsten Quadrate nach GAUSS, dass heißt

$$E(Y - g(X_1, \dots, X_j))^2 \rightarrow \min. \quad (6.1)$$

Je nach Funktionsklasse für g unterscheidet man:

1. Lineare Regression: $Y = g(X_1, \dots, X_j) = b_0 + b_1 X + \dots + b_j X_j$
2. nicht lineare Regression: g beschreibt einen nicht linearen Zusammenhang, zum Beispiel $Y = g(X_1, X_2) = ae^{bX_1 + cX_2}$

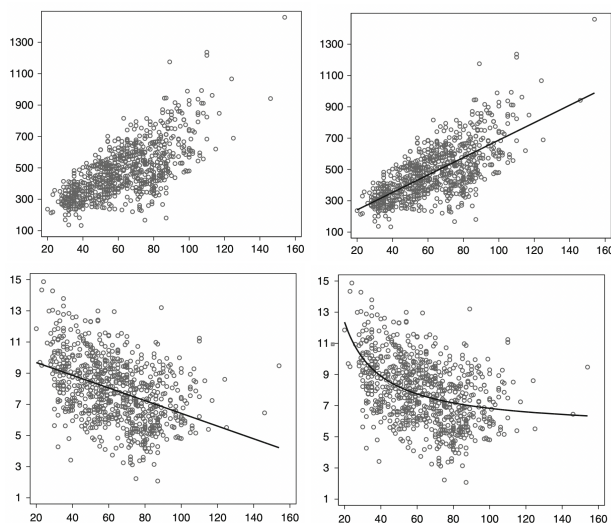


Abbildung 6.2: Die Abbildung zeigt in (a) eine gegebene Datenmenge, (b) eine lineare Regressionsgerade mit positiven Anstieg, (c) eine lineare Regressionsgerade mit negativen Anstieg und (d) eine nicht lineare Regression.

3. einfache Regression: $j = 1$, dass heißt nur eine unabhängige Variable
4. multiple Regression: $j \geq 1$, dass heißt mehrere unabhängige Variablen

Verschiedene Arten der Regression sind in Abbildung 6.2 dargestellt.

Beispiel 6.40. Reicht es für die Absatzsteigerung eines Produktes aus, nur die Kosten für die Werbung zu erhöhen oder haben auch Preis und Zahl der Vertreterbesucher einen Einfluss?

abhängige Variable: Y - Absatzmenge des Produktes

unabhängige Variablen: X_1 - Kosten für Werbung, X_2 - Preis des Produktes, X_3 - Zahl der Vertreterbesucher

vermutete Regressionsfunktion: $y = a_0 + a_1x + a_2x_2 + a_3x_3 \leadsto$ multiple lineare Regression (Ursachenanalyse)

Im Rahmen der Vorlesungen beschränken wir uns auf die einfache lineare Regression.

6.2.2 Einfache lineare Regression

Ziel: Die Angabe einer konkreten, dass heißt auf die konkrete Stichprobe passende, Regressionsgeraden. Diese Regressionsgerade soll die sie Daten „bestmöglich“ repräsentieren.

Das allgemeine Modell einer einfach lineare Regression ist

$$Y = b_0 + b_1 X + U,$$

wobei Y der Regressand, X der Regressor und b_0, b_1 die theoretischen Regressionskoeffizienten sind. Weiter ist U eine sogenannte Störvariable mit der Annahme $U \sim N(0, \sigma)$. Das Verfahren, um die am besten angepasste Gerade zu einer gegebenen Wertepaaren (x_i, y_i) , $i = 1, \dots, n$ zu bestimmen ist die Methode der kleinsten Quadrate (MKQ), siehe Gleichung 6.1. Im Falle der einfachen lineare Regression ist es das Ziel die quadratischen Abweichung von y und einer Geraden $\hat{y} = \hat{b}_0 + \hat{b}_1 x$ zu minimieren, dass heißt

$$\sum_{i=1}^n \left(y_i - \underbrace{\left(\hat{b}_0 + \hat{b}_1 x_i \right)}_{\hat{y}_i} \right)^2 \rightarrow \min. \quad (6.2)$$

Genauer, für die Abweichung der Geraden \hat{y} von einem einzelnen Punkt (x_i, y_i) wählt man das Quadrat des Abstandes von Punkt und Gerade in senkrechter Richtung.

Unter Anwendung der Methode der kleinsten Quadrate (MKQ) ergibt sich dann also die empirische Regressionsgerade

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x. \quad (6.3)$$

Die Parameter \hat{b}_0 und \hat{b}_1 sind die geschätzten Werte für b_0 und b_1 , dass heißt im Sinne der Gleichung 6.2 die optimalen Werte. Die Bestimmung der Parameter \hat{b}_0 und \hat{b}_1 erfolgt mittels der partiellen Ableitungen und Nullsetzen der entstandenen Gleichungen. Für die Schätzung der Parameter b_0 und b_1 ergibt sich

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$$

und

$$\hat{b}_1 = \frac{s_{XY}}{s_X^2} = r \frac{s_Y}{s_X}.$$

Damit kann die empirische Regressionsgerade auch geschrieben werden als

$$\begin{aligned} \hat{y} &= \hat{b}_0 + \hat{b}_1 x \\ &= \bar{y} - \hat{b}_1 \bar{x} + \hat{b}_1 x \\ &= \bar{y} + \hat{b}_1 (x - \bar{x}). \end{aligned}$$

\hat{y} bezeichnet man auch als Prognosewert (geschätzter Wert) von y zu vorgegebenen x auf der Geraden. Die Regressionsgerade verläuft stets durch den Schwerpunkt (\bar{x}, \bar{y}) der Datenwolke (Punktwolke).

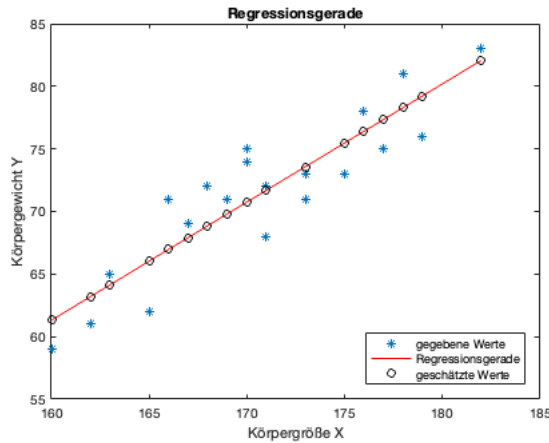


Abbildung 6.3: In dieser Abbildung sind die gegebenen Werte für X und Y mit blauen Sternen, die geschätzten Werte mit schwarzen Kreisen und die Regressionsgerade als rote Linie dargestellt.

Beispiel 6.41. (Fortsetzung des Beispiels 6.36) Folgende Werte wurden bereits für das Merkmal X Körpergröße und das Merkmal Y Körpergewicht berechnet:

$$\bar{x} = 170,75 \text{ cm}, \bar{y} = 71,45 \text{ kg}, s_X \approx 5,964 \text{ cm}, s_Y \approx 6,236 \text{ kg}, s_{XY} \approx 33,54, \\ r_{XY} \approx 0,902$$

Es werden nun die empirischen Regressionskoeffizienten \hat{b}_0 und \hat{b}_1 bestimmt:

$$\hat{b}_1 = \frac{s_{XY}}{s_X^2} = \frac{33,54}{(5,964)^2} = 0.943$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 71,45 - 0.943 \cdot 170.75 = -89,575$$

Damit ergibt sich die Regressionsgerade: $\hat{y} = \hat{b}_1 x + \hat{b}_0 = 0.943x - 89.575$. Die Regressionsgerade ist in Abbildung 6.3 dargestellt.

Gütemaße bei einer Regressionsanalyse

Um, die entstandene empirische Regressionsgerade beurteilen zu können, dass heißt wie gut die Regressionsgerade die gegebenen Daten beschreibt, werden Gütemaße herangezogen. Zuerst werden drei verschiedene Streuungsmaße eingeführt.

Definition 6.42. Die Gesamtstreuung (Gesamtvarianz), die Streuung der y_i -Werte, die zur Zielvariablen Y gehören ist definiert als

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Gesamtvarianz ist also die quadrierte Abweichung aller Werte vom Mittelwert.

Definition 6.43. Die durch die Regressionsgerade erklärte Streuung, die Streuung der geschätzten \hat{y}_i -Werte, ist definiert als

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

Diese wird auch als Regressionsvarianz (auch erklärte Varianz) bezeichnet. Es ist die quadrierte Abweichung der vorhergesagten Werte vom Mittelwert.

Definition 6.44. Die Reststreuung (Fehlervarianz), die durch die Regressionsgerade nicht erklärte Streuung, ist definiert als

$$s_u^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Die Reststreuung (auch nicht erklärte Varianz) ist die quadrierte Abweichung aller Werte von vorhergesagtem Wert. Je kleiner s_u^2 desto enger liegen die Messwerte um die Regressionsgerade gruppiert. s_u^2 kann zum Vergleich mehrerer Regressionsfunktionen verwendet werden.

Satz 6.45. *Es gilt folgende Zerlegung der Gesamtstreuung*

$$(n-1)s_Y^2 = (n-1)s_y^2 + (n-2)s_u^2.$$

Dies bedeutet grob formuliert, dass die Gesamtvarianz sich zusammensetzt aus die durch die Regressionsgerade erklärte Streuung und der Reststreuung, welche nicht durch die Regressionsgerade erklärt wird.

Weitere Gütemaße bei der Regressionsanalyse sind das Bestimmtheitsmaß und das Unbestimmtheitsmaß.

Definition 6.46. Das empirische Bestimmtheitsmaß ist definiert als

$$b = \frac{s_y^2}{s_Y^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Für b gilt $0 \leq b \leq 1$. Das empirische Bestimmtheitsmaß b zeigt an, welcher Anteil an der Gesamtstreuung der abhängigen Variablen Y auf den Regressor X entfällt, das heißt die Varianz die durch die Regressionsfunktion erklärt wird. Beziehungsweise $1 - b$ ist der Anteil der Gesamtvarianz von Y , der auf weitere (im Modell nicht erfasste) Einflüsse zurückführbar ist.

- b nahe 1: Die Regressionsfunktion g beschreibt den Zusammenhang zwischen Y und der unabhängigen Variablen X sehr gut.
- b nahe 0: Die Regressionsfunktion g ist ungeeignet. Der Trend der Messwertpaare wird von g nicht erfasst. \curvearrowright Wahl einer anderen Funktionsklasse für die Regressionsfunktion

Satz 6.47. Bei linearer Regression gilt $b = r^2$.

Definition 6.48. Das empirische Unbestimmtheitsmaß ist definiert als

$$u = \frac{(n-2)}{(n-1)} \cdot \frac{s_u^2}{s_Y^2}.$$

Das Unbestimmtheitsmaß ist das Komplement des Bestimmtheitsmaßes und beschreibt den Anteil der Streuung, der unerklärt bleibt. Das Bestimmtheits- und Unbestimmtheitsmaß addieren sich jeweils zu 1, dass heißt es gilt $b + u = 1$.

Folgerung: Es gilt also folgender Zusammenhang zwischen dem Bestimmtheitsmaß und der Reststreuung $s_u^2 = \frac{n-1}{n-2} s_Y^2 (1 - b)$.

Beispiel 6.49. (Fortsetzung des Beispiels 6.36) Folgende Werte wurden bereits für das Merkmal X Körpergröße und das Merkmal Y Körpergewicht berechnet:

$\bar{x} = 170,75 \text{ cm}$, $\bar{y} = 71,45 \text{ kg}$, $s_X \approx 5,964 \text{ cm}$, $s_Y \approx 6,236 \text{ kg}$, $s_{XY} \approx 33,54$,
 $r_{XY} \approx 0,902$

Regressionsgerade: $\hat{y} = 0.943 \cdot x - 89.575$.

Bestimmtheitsmaß: $b = r^2 = (0.902)^2 \approx 0.814$. Demnach entfällt ungefähr 81% der Gesamtstreuung von Y auf X .

Unbestimmtheitsmaß: $u = 1 - r^2 = 1 - (0.902)^2 \approx 0.1860$. Demnach entfällt ungefähr 19% der Gesamtstreuung von Y auf im Modell nicht erfasste Einflüsse.

Restvarianz: $s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 7,667$

Kennt man die Regressionsgerade \hat{y} , so verwendet man sie zum Interpolieren und Extrapolieren.

Beispiel 6.50. (Fortsetzung des Beispiels 6.36) Zusatzfrage: Welches Körpergewicht ist zu erwarten bei einer Körpergröße von 195 cm ?

$\hat{y} = 0.943 \cdot 195 - 89.575 = 94.31$

Kapitel 7

Beurteilende Statistik

Wie bereits einleitend zur Statistik erwähnt können mit der induktiven Statistik Rückschlüsse über unbekannte Eigenschaften der zu untersuchenden Objekte für eine gegebene Stichprobe gezogen werden. Dies impliziert zum Beispiel die Schätzung unbekannter Parameter einer Verteilung für eine Zufallsgröße.

7.1 Statistische Schätzmethoden

Die Verteilungsfunktion einer Zufallsgröße X ist vom Typ bekannt, enthält aber unbekannte statistische Parameter, zum Beispiel $N \sim (\mu, \sigma)$ verteilt, aber μ und σ sind unbekannt.

Aufgaben der Parameterschätzung:

1. Bestimmung von Schätz- und Näherungswerten für unbekannte Verteilungsparameter einer Zufallsgröße X unter Verwendung einer Stichprobe. Das sind sogenannte Punktschätzungen.
2. Konstruktion von Konfidenzintervallen (Vertrauensintervallen), in denen die unbekannten Parameter mit einer vorgegebenen Wahrscheinlichkeit vermutet werden. Das sind sogenannte Intervall- oder Bereichsschätzungen.

7.1.1 Punktschätzungen

Die Schätzung eines unbekannten Parameters Θ erfolgt anhand einer Schätzfunktion $\Theta = g(X_1, \dots, X_n)$. Dabei sind X_1, \dots, X_n unabhängige Zufallsgrößen mit gleicher Verteilungsfunktion $F(x)$. Für jede konkrete Stichprobe x_1, \dots, x_n erhält man einen Schätzwert $\hat{\Theta} = g(x_1, \dots, x_n)$.

7.1.1.1 Momentenmethode

Die Momentenmethode ist die älteste Schätzmethode und sehr einfach handhabbar. Mittels der Momente einer Zufallsvariable können die Parameter der Verteilung bestimmt werden. Zum Beispiel die Parameter der Normalverteilung werden nun gemäß *Momentenmethode* so geschätzt, dass die Momente der Verteilung gerade den obigen eingeführten empirischen Momenten entsprechen.

Allgemeines Vorgehen:

1. $\Theta = g(\alpha_1, \alpha_2, \dots, \beta_1, \beta_2, \dots)$ - Θ ist als Funktion der Momente darstellbar.
2. $\hat{\Theta} = g(A_1, A_2, \dots, B_1, B_2, \dots)$ - $\hat{\Theta}$ ist Schätzfunktion für Θ mit den empirischen Momenten $A_k = \frac{1}{n} \sum_{i=1}^n x_i^k$ und $B_k = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^k$ (unter der Annahme der Existenz der Momente).

Man erhält also eine Schätzfunktion, indem man in der zu schätzenden Funktion die Momente der Wahrscheinlichkeitsverteilungen durch die Stichprobenmomente ersetzt. Nachfolgend werden die Parameter für die Normalverteilung bestimmt.

Beispiel 7.1. Die Parameter $\alpha_1 = E(X) = \mu$ und $\beta_2 = D^2(X) = \sigma^2$ einer $N(\mu, \sigma)$ -verteilten Grundgesamtheit sind zu schätzen. Mit der Beziehung $D^2(X) = \sigma^2 = E(X^2) - (E(X))^2 \curvearrowright E(X^2) = \sigma^2 + (E(X))^2$ ergibt sich folgendes Gleichungssystem

Das erste Moment: $\frac{1}{n} \sum_{i=1}^n X_i = E(X) = \mu$

Das zweite Moment: $\frac{1}{n} \sum_{i=1}^n X_i^2 = \sigma^2 + (E(X))^2 \curvearrowright$ Umstellen nach σ^2 .

Schätzung für den Mittelwert: Mit $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ und Ersetzung der theoretischen durch die empirischen Momente ergibt sich für die Schätzung des Mittelwertes

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

welcher aus einer konkreten Stichprobe (x_1, \dots, x_n) bestimmt wird.

Schätzung für die Varianz: Für die Schätzung der Varianz verwendet man die Schätzfunktion $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Mittels einer konkreten Stichprobe (x_1, \dots, x_n) ergibt sich der Schätzwert

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Schätzwerte für Parameter spezieller Wahrscheinlichkeitsverteilungen:

Verteilung	Schätzung für:
Binomialverteilung: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$	$\hat{p} = \frac{X}{n} = \frac{k}{n}$
Poisson Verteilung: $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$	$\hat{\lambda} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Exponentialverteilung: $f(x) = \lambda e^{-\lambda x}$	$\hat{\lambda} = \frac{1}{\bar{x}}$
Normalverteilung: $f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	$\hat{\mu} = \bar{x}; \hat{\sigma}^2 = s^2$

Beispiel 7.2. Die Lebensdauer eines elektronischen Bauteils genüge einer Exponentialverteilung mit dem Parameter λ . Es ist ein Schätzwert $\hat{\lambda}$ für λ anhand der folgenden Stichprobe zu bestimmen (Lebensdauer t_i in Stunden h):

i	1	2	3	4	5	6	7	8
t_i	950	980	1150	770	1230	1210	990	1120

$$\bar{t} = \frac{1}{8} (950 + 980 + 1150 + 770 + 1230 + 1210 + 990 + 1120) = 1050 \text{ h}$$

$$\hat{\lambda} = \frac{1}{\bar{t}} = \frac{1}{1050} \text{ h}^{-1} = 0.00095238 \text{ h}^{-1}$$

\curvearrowright Die mittlere Lebensdauer \bar{t} der elektronischen Bauteile beträgt 1050 h.

7.1.1.2 Maximum-Likelihood-Methode

Die Maximum-Likelihood-Methode ist eine allgemeine Methode zur Schätzung von Parametern für Zufallsgrößen. Gegenüber der Momentenmethode hat die Maximum-Likelihood-Methode den Vorteil, dass sie im Allgemeinen „effizienter“ ist (d. h. genauere Schätzungen liefert) und zusätzlich mit weiteren theoretischen Überlegungen die Genauigkeit der Schätzer angegeben werden kann (zum Beispiel die Fisher-Information, welche hier nicht behandelt wird). Ein Nachteil ist jedoch, dass man manchmal die log- Likelihood-Funktion mit numerischen (iterativen) Verfahren maximieren muss.

Für stetige Zufallsgrößen:

Stetige Zufallsgrößen X mit der Wahrscheinlichkeitsdichte $f(x, \beta_1, \dots, \beta_n)$ mit den zu schätzenden Parametern β_1, \dots, β_n .

1. Schritt: Zur Stichprobe x_1, \dots, x_n wird die Likelihood-Funktion gebildet:

$$L = \prod_{i=1}^n f(x, \beta_1, \dots, \beta_n).$$

Basierend auf der Likelihood-Funktion ermittelt man nun den sogenannten Maximum-Likelihood-Schätzer. Die Maximum-Likelihood-Methode besteht nun darin, denjenigen Parameterwert Θ zu finden, für den die Wahrscheinlichkeit, dass die gegebene Stichprobe x_1, \dots, x_n eintritt, am größten ist. Oder anders formuliert: Θ soll so gewählt werden, dass die Likelihood $L(\Theta)$ maximal wird. Daher der Name „Maximum-Likelihood-Methode“.

2. Schritt: Die optimalen Parameter $\beta_1^*, \dots, \beta_n^*$ werden aus der Bedingung $L \rightarrow \max$ bestimmt. Es entsteht ein Gleichungssystem zur Berechnung von $\beta_1^*, \dots, \beta_n^*$:

$$\frac{1}{L} \frac{\partial L}{\partial \beta_j} = \sum_{i=1}^n \left(\frac{1}{f(x_i, \beta_1, \dots, \beta_k)} \frac{\partial f(x_i, \beta_1, \dots, \beta_k)}{\partial \beta_j} \right) = 0 \quad j = 1, \dots, k.$$

Beispiel 7.3. Die Zufallsgröße X sei $N(\mu, \sigma)$ -verteilt. Zu schätzen sind μ und σ .

$$f(x, \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Likelihood-Funktion:

$$L = \frac{1}{(\sigma \sqrt{2\pi})^n} e^{\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)}$$

Partielle Ableitungen nach μ und σ :

$$\frac{\partial L}{\partial \mu} = \frac{L}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \stackrel{!}{=} 0$$

\hookrightarrow

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

$$\frac{\partial L}{\partial \sigma} = L \left(-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 \right) \stackrel{!}{=} 0$$

\hookrightarrow

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2.$$

Dies ist derselbe Schätzer wie bei der Momentenmethode. Die Maximum-Likelihood-Methode liefert aber nicht zwangsläufig bei allen Verteilungen das gleiche Resultat wie die Momentenmethode.

Für diskrete Zufallsgrößen: Der diskrete Fall verläuft im Wesentlichen analog. Man braucht nur f durch p also die Wahrscheinlichkeitsdichte durch Wahrscheinlichkeitsfunktion ersetzen.

7.2 Konfidenzschätzungen

Die Wahrscheinlichkeit für eine Realisation einer stetigen Variablen, genau einen bestimmten Wert zu treffen, ist Null. Für eine endliche Wahrscheinlichkeit (größer Null) muss ein Intervall um den Wert angegeben sein. Kennt man die Größe des Intervalls, so kann auch die Wahrscheinlichkeit $P(a \leq X \leq b)$ dafür angegeben, dass eine Realisation der Zufallsgröße X in das Intervall $[a, b]$ fällt. Dieses Vorgehen kann auch andersherum betrachtet werden. Eine Wahrscheinlichkeit wird festgelegt und berechnet dann, wie groß das Intervall sein muss, damit eine Realisation mit der gegebenen Wahrscheinlichkeit in dieses Intervall fällt. Wenn man den Mittelwert \bar{x} einer Stichprobe kennt, so kann man mit der gleichen Überlegung ermitteln, wie groß das Intervall um den wahren Mittelwert μ sein muss, damit der Mittelwert \bar{x} bei einer festgelegten Wahrscheinlichkeit in das Intervall fällt. Aus Symmetriegründen ist es die gleiche Fragestellung, ob der wahre Mittelwert im Intervall um den Stichprobenmittelwert liegt. Dieses Intervall heißt *Konfidenzintervall* oder *Vertrauensbereich*.

Mathematische Einführung:

Mit einer Punktschätzung, bei der ein unbekannter Parameter durch einen einzigen aus einer Stichprobe ermittelten Wert geschätzt wird, gewinnt man keine Aussage über die Genauigkeit der einer solchen Schätzung. Die Abweichungen einzelner Punktschätzwerte vom Wert des Parameters können erheblich sein. Das ist besonders dann der Fall, wenn der Stichprobenumfang klein ist. Um eine Vorstellung über die Genauigkeit einer Schätzung zu erhalten, kann man *Konfidenzschätzungen*, eine spezielle Form von *Bereichsschätzungen*, betrachten. Bei dieser wird für einen unbekannten Parameter der Grundgesamtheit mit Hilfe einer Stichprobe ein Intervall in gewissen Grenzen gesucht, so dass der unbekannte Parameter mit einer vorgegebenen Wahrscheinlichkeit überdeckt wird.

Im Anwendungsfall für den Erwartungswert bedeutet dies: Wenn der wahre Mittelwert μ nicht bekannt ist, so schätzt man diesen durch den Stichprobenmittelwert \bar{x} . \bar{x} ist der sogenannte Schätzer (Punktschätzwert) für den wahren Mittelwert μ , wie bereits im letzten Abschnitt kennengelernt. Die Wahrscheinlichkeit, dass mit $\bar{x} = \mu$ der Schätzer den wahren Mittelwert genau trifft, ist Null. Aber es kann ein Intervall um den Punktschätzwert angegeben werden, in dem der wahre Mittelwert mit einer festgelegten Wahrscheinlichkeit liegen wird. Dieses Intervall ist das Konfidenzintervall für den Erwartungswert.

Grundanliegen der Konfidenzschätzungen:

Für eine mathematische Stichprobe $X = (X_1, X_2, \dots, X_n)$ ist ein möglichst kleines Intervall $I = [G_u(X), G_o(X)]$ bestimmen, so dass für eine vorgegebenen Wahrscheinlichkeit $1 - \alpha$ der unbekannte Parameter überdeckt wird, siehe Abbildung 7.1. Der Abstand der unteren Konfidenzgrenze $G_u = G_u(X)$ bzw. der oberen Konfidenzgrenze $G_o = G_o(X)$ vom Schätzwert heißt *maximaler Schätzfehler* ε .

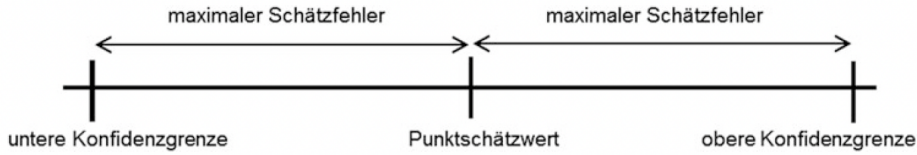


Abbildung 7.1: Allgemeine Darstellung eines Konfidenzintervalls.

Definition 7.4. Ein Intervall $I = [G_u(X), G_o(X)]$ heißt *zweiseitiges Konfidenzintervall* (KI) für den unbekannten Parameter Θ zum Konfidenzniveau $1 - \alpha$, falls

$$P(G_u \leq \Theta \leq G_o) = 1 - \alpha.$$

- $G_u = G_u(X)$ und $G_o = G_o(X)$ sind Stichprobenfunktionen.
- $g_u = G_u(x)$ und $g_o = G_o(x)$ sind für eine konkrete Stichprobe die Realisierungen von G_u und G_o
- $[g_u, g_o]$ ist ein konkretes KI.
- α ist die Irrtumswahrscheinlichkeit.

Bemerkung 7.5. Die Länge des Konfidenzintervalls $L = G_o - G_u$ hängt ab vom Verteilungstyp F , der Stichprobe, der Stichprobengröße n und der Irrtumswahrscheinlichkeit α .

Erläuterung zur Irrtumswahrscheinlichkeit α :

Die ermittelten KI zum Konfidenzniveau $1 - \alpha$ werden im Mittel in $(1 - \alpha) \%$ aller Fälle Θ überdecken und in $\alpha \cdot 100\%$ aller Fälle nicht. α ist ein Ausdruck für das Risiko, dass der Anwender von der Schätzung festlegen muss. Typische Werte für α in der praktischen Anwendung sind 0.05, 0.01 und 0.001.

Konfidenzintervalle können auch nur einseitig betrachtet werden, dass heißt der Vertrauensbereich liegt linksseitig oder rechtsseitig.

Definition 7.6. Ein Intervall $I = (-\infty, G_o(X)]$ heißt *linksseitiges Konfidenzintervall* (KI) für den unbekannten Parameter Θ zum Konfidenzniveau $1 - \alpha$, falls

$$P(-\infty < \Theta \leq G_o) = 1 - \alpha.$$

Definition 7.7. Ein Intervall $I = [G_u(X), \infty)$ heißt *rechtsseitiges Konfidenzintervall* (KI) für den unbekannten Parameter Θ zum Konfidenzniveau $1 - \alpha$, falls

$$P(G_u \leq \Theta < \infty) = 1 - \alpha.$$

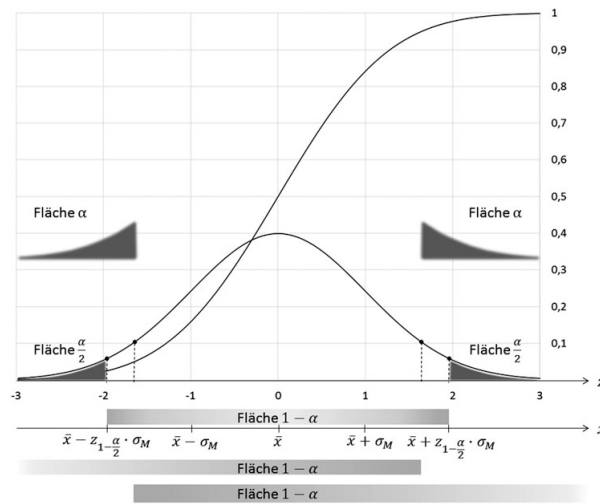


Abbildung 7.2: Darstellung der einseitigen und zweiseitigen Konfidenzintervalle für den Erwartungswert.

In der Abbildung 7.2 werden die einseitigen und zweiseitigen Konfidenzintervalle dargestellt.

7.2.1 Konstruktion eines Konfidenzintervalls

Ein Konfidenzintervall wird nach folgendem Prinzip konstruiert:

1. Ausgangspunkt ist eine (möglichst gute) Punktschätzfunktion für $\hat{\Theta}$ für Θ .
2. $\hat{\Theta}$ wird solange umgeformt zu $\tilde{\Theta}$, so dass die Verteilungsfunktion von $\tilde{\Theta}$ wenigstens asymptotisch bekannt ist
3. Ansatz für ein zweiseitiges Konfidenzintervall aufstellen
4. Umstellen des Ansatzes nach Θ .

Nachfolgend werden die Konfidenzintervalle für den Erwartungswert, die Varianz und für einen Anteilswert hergeleitet.

7.2.1.1 Das zweiseitige Konfidenzintervall für den Erwartungswert

Die Aufstellung des Konfidenzintervalls unterscheidet sich nach den gegebenen Voraussetzungen. Dabei unterscheidet man allgemein vier Fälle.

- Fall 1:** Die Grundgesamtheit X ist normalverteilt mit bekannter Varianz σ^2 .
Fall 2: Die Grundgesamtheit X ist normalverteilt mit unbekannter Varianz σ^2 .
Fall 3: Die Grundgesamtheit X ist beliebig verteilt mit bekannter Varianz σ^2 .
Fall 4: Die Grundgesamtheit X ist beliebig verteilt mit unbekannter Varianz σ^2 .

Konfidenzintervall für den ersten Fall:

Gegeben: $X \sim N(\mu, \sigma)$, σ^2 bekannt, $\Theta = \mu$

(1) $\hat{\Theta} = \bar{X}$

(2) $\hat{\Theta} = \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \leadsto \tilde{\Theta} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0, 1)$

(3) Ansatz für ein Konfidenzintervall:

$$\begin{aligned} P\left(c_{N(0,1), \frac{\alpha}{2}} \leq \tilde{\Theta} \leq c_{N(0,1), 1-\frac{\alpha}{2}}\right) &= 1 - \alpha \\ P\left(-c_{N(0,1), 1-\frac{\alpha}{2}} \leq \tilde{\Theta} \leq c_{N(0,1), 1-\frac{\alpha}{2}}\right) &= 1 - \alpha \end{aligned}$$

(4) Umstellen des Ansatzes:

$$\begin{aligned} P\left(c_{N(0,1), \frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\sigma} \cdot \sqrt{n} \leq c_{N(0,1), 1-\frac{\alpha}{2}}\right) &= 1 - \alpha \\ P\left(-c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} - \bar{x} \leq -\mu \leq c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} - \bar{x}\right) &= 1 - \alpha \\ P\left(c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x} \geq \mu \geq -c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}\right) &= 1 - \alpha \\ P\left(\underbrace{-c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}}_{G_u} \leq \mu \leq \underbrace{c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}}_{G_o}\right) &= 1 - \alpha \end{aligned}$$

Das Konfidenzintervall ist symmetrisch zu μ . Es handelt sich hier um ein sogenanntes zentrales Konfidenzintervall, siehe Abbildung 7.2.

Satz 7.8. Die Länge des Konfidenzintervalls ist

$$\begin{aligned} L &= G_o - G_u \\ &= c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x} - \left(-c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}\right) \\ &= 2c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}. \end{aligned}$$

Beispiel 7.9. Es soll eine neue Schraubensorte eingeführt werden. Dazu wurde unter anderem die Länge der Schrauben gemessen. Für eine kleine Stichprobe von 10 Schrauben

ergaben sich folgende Werte in $[mm]$:

10 8 9 10 11 11 9 12 8 12

Der Hersteller gibt einen Mittelwert von $\mu = 11 \text{ mm}$ und eine Varianz von $\sigma^2 = 4 \text{ mm}$ an. Gesucht ist ein Intervall für den Erwartungswert der Längen, so dass die Schrauben mit 95%-iger Wahrscheinlichkeit in dieses Intervall fallen. Kann den Angaben des Herstellers für μ getraut werden?

Lösung: Gesucht ist ein Konfidenzintervall für μ zum Konfidenzniveau $1 - 0.05 = 0.95$. Dafür benötigen wir den Stichprobenmittelwert: $\bar{x} = 10 \text{ mm}$.

$$\begin{aligned} P\left(-c_{N(0,1),1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x} \leq \mu \leq c_{N(0,1),1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}\right) &= 1 - \alpha \\ P\left(-c_{N(0,1),0.975} \cdot \frac{2}{\sqrt{10}} + 10 \leq \mu \leq c_{N(0,1),0.975} \cdot \frac{2}{\sqrt{10}} + 10\right) &= 0.95 \\ P\left(-1.96 \cdot \frac{2}{\sqrt{10}} + 10 \leq \mu \leq 1.96 \cdot \frac{2}{\sqrt{10}} + 10\right) &= 0.95 \\ P(8.76 \leq \mu \leq 11.24) &= 0.95 \end{aligned}$$

Das Konfidenzintervall für den Erwartungswert lautet also $KI = [8.76, 11.24]$. Den Angaben des Herstellers mit $\mu = 11 \text{ mm}$ kann getraut werden, da diese innerhalb des Intervalls liegen.

Konfidenzintervall für den zweiten Fall:

Gegeben: $X \sim N(\mu, \sigma)$, σ^2 unbekannt, $\Theta = \mu$

(1) $\hat{\Theta} = \bar{X}$

(2) $\tilde{\Theta} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$, wobei $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$. Der Parameter $\tilde{\Theta}$ ist t -verteilt, dass heißt $\tilde{\Theta} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \cdot \sqrt{n} \sim t(n-1)$

(3) Ansatz für ein Konfidenzintervall:

$$P\left(c_{t(n-1),\frac{\alpha}{2}} \leq \tilde{\Theta} \leq c_{t(n-1),1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Aufgrund der Symmetrie der t -Verteilung gilt $c_{t(n-1),\frac{\alpha}{2}} = -c_{t(n-1),1-\frac{\alpha}{2}}$. Damit ergibt sich für den Ansatz

$$P\left(-c_{t(n-1),1-\frac{\alpha}{2}} \leq \tilde{\Theta} \leq c_{t(n-1),1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

(4) Umstellen des Ansatzes:

$$P\left(-c_{t(n-1), 1-\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{s} \cdot \sqrt{n} \leq c_{t(n-1), 1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(\underbrace{-c_{t(n-1), 1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} + \bar{x}}_{G_u} \leq \mu \leq \underbrace{c_{t(n-1), 1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} + \bar{x}}_{G_o}\right) = 1 - \alpha$$

Bemerkung 7.10. Für eine große Stichprobengröße n gilt $c_{t(n-1), 1-\frac{\alpha}{2}} \approx c_{N(0,1), 1-\frac{\alpha}{2}}$.

Bemerkung 7.11. Die Länge des Konfidenzintervalls hängt wegen der Schätzung der Standardabweichung von der konkreten Stichprobe ab.

Beispiel 7.12. (Fortsetzung Schraubensorte 7.9) Es sei nun die Varianz unbekannt. Gesucht ist nun ein Intervall für die Längen, wobei σ^2 unbekannt ist, so dass die Schrauben mit 95%-iger Wahrscheinlichkeit in dieses Intervall fallen. Kann den Angaben des Herstellers für μ getraut werden?

Lösung: Gesucht ist ein Konfidenzintervall für μ zum Konfidenzniveau $1 - 0.05 = 0.95$. Dafür benötigen wir den Stichprobenmittelwert: $\bar{x} = 10 \text{ mm}$ und die Stichprobenvarianz (die Schätzung der Varianz) $s^2 = 2.22$

$$P\left(-c_{t(n-1), 1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} + \bar{x} \leq \mu \leq c_{t(n-1), 1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} + \bar{x}\right) = 1 - \alpha$$

$$P\left(-c_{t(9), 0.975} \cdot \frac{\sqrt{2.22}}{\sqrt{10}} + 10 \leq \mu \leq c_{t(9), 0.975} \cdot \frac{\sqrt{2.22}}{\sqrt{10}} + 10\right) = 0.95$$

$$P\left(-2,262 \cdot \frac{\sqrt{2.22}}{\sqrt{10}} + 10 \leq \mu \leq 2,262 \cdot \frac{\sqrt{2.22}}{\sqrt{10}} + 10\right) = 0.95$$

$$P(8.9337 \leq \mu \leq 11.0663) = 0.95$$

Das Konfidenzintervall für den Erwartungswert lautet also $KI = [8.9337, 11.0663]$. Den Angaben des Herstellers mit $\mu = 11 \text{ mm}$ kann getraut werden, da diese innerhalb des Intervalls liegen.

Konfidenzintervall für den dritten Fall:

Gegeben: X beliebig verteilt, σ^2 bekannt, $\Theta = \mu$

$$(1) \hat{\Theta} = \bar{X}$$

$$(2) \tilde{\Theta} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \stackrel{n \rightarrow \infty}{\sim} N(0, 1)$$

Für $n > 30$ ist in diesem Fall wie im ersten Fall vorzugehen. Es wird hier ein asymptotisches Konfidenzintervall bestimmt.

Konfidenzintervall für den vierten Fall:

Gegeben: X beliebig verteilt, σ^2 unbekannt, $\Theta = \mu$

Für hinreichend großes n wird $s^2 = \sigma^2$ gesetzt. Der vierte Fall für großes n wie der erste Fall zu behandeln. Es ist hier ebenfalls ein asymptotisches Konfidenzintervall.

$$(1) \hat{\Theta} = \bar{X}$$

$$(2) \tilde{\Theta} = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \frac{\bar{X} - s}{\sigma} \cdot \sqrt{n} \stackrel{n \rightarrow \infty}{\sim} N(0, 1)$$

7.2.1.2 Einseitige Konfidenzintervalle für den Erwartungswert

Für das links- beziehungsweise rechtsseitige Konfidenzintervall bezüglich des ersten Falles ergibt sich

$$KI = \left(-\infty, \bar{x} + c_{N(0,1), 1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right] \text{ bzw. } KI = \left[\bar{x} - c_{N(0,1), 1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}, \infty \right).$$

Die anderen Fälle 2-3 können analog erstellt werden.

7.2.1.3 Das Konfidenzintervall für die Varianz

Bei einem Konfidenzintervall für die Varianz wird die wahre Varianz σ^2 , wenn diese nicht bekannt ist, geschätzt durch den Stichprobenvarianz s^2 . s^2 ist hier der sogenannte Schätzer (Punktschätzwert) für die wahre Varianz σ^2 . Es kann ein Intervall um den Punktschätzwert angegeben werden, in dem die wahre Varianz mit einer festgelegten Wahrscheinlichkeit liegen wird. Dieses Intervall ist das Konfidenzintervall für den Varianz. Es werden hier nur zwei Fälle betrachtet für μ bekannt und μ unbekannt. Die entsprechenden Konfidenzintervalle können analog zu den vom Erwartungswert hergeleitet werden.

Konfidenzintervall für den ersten Fall:

Gegeben: $X \sim N(\mu, \sigma)$, μ bekannt, $\Theta = \sigma^2$

$$(1) \hat{\Theta} = S^2, S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$(2) \tilde{\Theta} = \frac{nS^2}{\sigma^2} \sim \chi^2(n)$$

(3) Ansatz für ein Konfidenzintervall:

$$P\left(c_{\chi^2(n), \frac{\alpha}{2}} \leq \tilde{\Theta} \leq c_{\chi^2(n), 1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

(4) Umstellen des Ansatzes:

$$\begin{aligned}
 P\left(c_{\chi^2(n), \frac{\alpha}{2}} \leq \frac{ns^2}{\sigma^2} \leq c_{\chi^2(n), 1-\frac{\alpha}{2}}\right) &= 1 - \alpha \\
 P\left(\frac{c_{\chi^2(n), \frac{\alpha}{2}}}{ns^2} \leq \frac{1}{\sigma^2} \leq \frac{c_{\chi^2(n), 1-\frac{\alpha}{2}}}{ns^2}\right) &= 1 - \alpha \\
 P\left(\underbrace{\frac{ns^2}{c_{\chi^2(n), 1-\frac{\alpha}{2}}}}_{G_u} \leq \sigma^2 \leq \underbrace{\frac{ns^2}{c_{\chi^2(n), \frac{\alpha}{2}}}}_{G_o}\right) &= 1 - \alpha
 \end{aligned}$$

Beispiel 7.13. (Fortsetzung Schraubensorte 7.9) Der Hersteller gibt einen Mittelwert von $\mu = 11 \text{ mm}$ und eine Varianz von $\sigma^2 = 4 \text{ mm}$ an. Gesucht ist ein Intervall für die Varianz der Längen, so dass die Schrauben mit 95%-iger Wahrscheinlichkeit in dieses Intervall fallen. Kann den Angaben des Herstellers für σ^2 getraut werden?

Lösung: Zu bestimmen ist das das Konfidenzintervall für σ^2 bei bekannten μ . Die Stichprobenvarianz beträgt $s^2 = 2.22$.

$$\begin{aligned}
 P\left(\frac{10 \cdot 2.22}{c_{\chi^2(10), 0.975}} \leq \sigma^2 \leq \frac{10 \cdot 2.22}{c_{\chi^2(10), 0.025}}\right) &= 0.95 \\
 P\left(\frac{10 \cdot 2.22}{20.5} \leq \sigma^2 \leq \frac{10 \cdot 2.22}{3.25}\right) &= 0.95 \\
 P(1.0829 \leq \sigma^2 \leq 6.8308) &= 0.95
 \end{aligned}$$

Das Konfidenzintervall für die Varianz ist $KI = [1.0829, 6.8308]$. Den Angaben des Herstellers mit $\sigma^2 = 4 \text{ mm}$ kann getraut werden, da diese innerhalb des Intervalls liegen.

Konfidenzintervall für den zweiten Fall:

Gegeben: $X \sim N(\mu, \sigma)$, μ unbekannt, $\Theta = \sigma^2$

$$(1) \hat{\Theta} = S^2, S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$(2) \tilde{\Theta} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

(3) Ansatz für ein Konfidenzintervall:

$$P\left(c_{\chi^2(n-1), \frac{\alpha}{2}} \leq \tilde{\Theta} \leq c_{\chi^2(n-1), 1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

(4) Umstellen des Ansatzes:

$$\begin{aligned}
 P\left(c_{\chi^2(n-1), \frac{\alpha}{2}} \leq \frac{(n-1)s^2}{\sigma^2} \leq c_{\chi^2(n-1), 1-\frac{\alpha}{2}}\right) &= 1 - \alpha \\
 P\left(\frac{c_{\chi^2(n-1), \frac{\alpha}{2}}}{(n-1)s^2} \leq \frac{1}{\sigma^2} \leq \frac{c_{\chi^2(n-1), 1-\frac{\alpha}{2}}}{(n-1)s^2}\right) &= 1 - \alpha \\
 P\left(\underbrace{\frac{(n-1)s^2}{c_{\chi^2(n-1), 1-\frac{\alpha}{2}}}}_{G_u} \leq \sigma^2 \leq \underbrace{\frac{(n-1)s^2}{c_{\chi^2(n-1), \frac{\alpha}{2}}}}_{G_o}\right) &= 1 - \alpha
 \end{aligned}$$

Beispiel 7.14. (Fortsetzung Schraubensorte 7.9) Der Hersteller gibt eine Varianz von $\sigma^2 = 4 \text{ mm}$ an. Gesucht ist ein Intervall für die Varianz der Längen, so dass die Schrauben mit 95%-iger Wahrscheinlichkeit in dieses Intervall fallen. Kann den Angaben des Herstellers für σ^2 getraut werden?

Lösung: Zu bestimmen ist das das Konfidenzintervall für σ^2 bei unbekannten μ . Die Stichprobenvarianz beträgt $s^2 = 2.22$.

$$\begin{aligned}
 P\left(\frac{9 \cdot 2.22}{c_{\chi^2(9), 0.975}} \leq \sigma^2 \leq \frac{9 \cdot 2.22}{c_{\chi^2(9), 0.025}}\right) &= 0.95 \\
 P\left(\frac{9 \cdot 2.22}{19} \leq \sigma^2 \leq \frac{9 \cdot 2.22}{2.70}\right) &= 0.95 \\
 P(1.168 \leq \sigma^2 \leq 8.22) &= 0.95
 \end{aligned}$$

Das Konfidenzintervall für die Varianz ist $KI = [1.168, 8.22]$. Den Angaben des Herstellers mit $\sigma^2 = 4 \text{ mm}$ kann getraut werden, da diese innerhalb des Intervalls liegen.

7.2.1.4 Einseitige Konfidenzintervalle für die Varianz

Für das links- beziehungsweise rechtsseitige Konfidenzintervall für die Varianz bezüglich des ersten Falles (μ bekannt) ergibt sich

$$KI = \left[0, \frac{ns^2}{c_{\chi^2(n), \alpha}}\right] \text{ bzw. } KI = \left[\frac{ns^2}{c_{\chi^2(n), 1-\alpha}}, \infty\right).$$

Der andere Fall, wenn μ unbekannt ist, kann analog erstellt werden, dass heißt

$$KI = \left[0, \frac{(n-1)s^2}{c_{\chi^2(n-1), \alpha}}\right] \text{ bzw. } KI = \left[\frac{(n-1)s^2}{c_{\chi^2(n-1), 1-\alpha}}, \infty\right).$$

7.2.1.5 Das Konfidenzintervall für einen Anteilswert

In vielen Fällen fragt man nicht nach einem Mittelwert, sondern nach einem Prozentsatz in der Grundgesamtheit. Das kann zum Beispiel die Ausschussquote in der Produktion oder die Zuverlässigkeit von Zustelldiensten sein. Auch hier ist man auf die Schätzung des wahren Anteils mit Hilfe einer Stichprobe angewiesen. In der Stichprobe können Merkmalsträger eine Eigenschaft haben oder nicht, nach dem Anteil derjenigen, die die Eigenschaft haben, ist gefragt. Der Anteil entspricht dann der klassischen Definition der Wahrscheinlichkeit p . Folglich ist die Wahrscheinlichkeit $1 - p$, dass ein Merkmalsträger die Eigenschaft nicht hat. Wenn eine große Grundgesamtheit vorliegt, wird sich die Wahrscheinlichkeit durch Entnahme einer zufälligen Stichprobe nicht ändern. Damit liegt ein Bernoulli-Schema vor.

Gegeben sei $X \sim \text{Null} - \text{Eins}(p)$. Der unbekannte Parameter ist hier $\Theta = p$. Für ein Konfidenzintervall bezüglich $\Theta = p$ wird $E(X) = \mu = p$ und $D^2(X) = \sigma^2 = p(1 - p)$ verwendet. Die Erstellung eines Konfidenzintervalls für $\Theta = p$ erfolgt unter Zurückführung auf ein Konfidenzintervall für den Erwartungswert einer beliebig verteilten Grundgesamtheit mit unbekannter Varianz (Fall 3). Dabei ist $\hat{\mu} = \bar{x}$ und $\hat{\sigma}^2 = s^2$. Speziell für die Null-Eins Verteilung gilt $\hat{\mu} = \hat{p}$ und $\hat{\sigma}^2 = \hat{p}(1 - \hat{p})$. Die Approximationsbedingung ist $n \cdot \hat{p}(1 - \hat{p}) > 9$. Nach dem Grenzwertsatz von MOIVRE/LAPLACE gilt $\bar{X}_n \stackrel{\text{asympt.}}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ ergibt sich das zweiseitige Konfidenzintervall

$$KI = \left[\hat{p} - c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + c_{N(0,1), 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right].$$

Die einseitigen Bereiche sind

$$KI_{links} = \left(-\infty; \hat{p} + c_{N(0,1), 1-\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

und

$$KI_{rechts} = \left[\hat{p} - c_{N(0,1), 1-\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \infty \right).$$

Beispiel 7.15. Für eine neue Apfelsorte soll herausgefunden werden, welcher Anteil der Äpfel mit 95%-iger Wahrscheinlichkeit ein Gewicht von größer als 200 g hat. Der Gesamternte wird eine Stichprobe von 100 Äpfeln entnommen. Es stellt sich heraus, dass 20 Äpfel haben ein Gewicht von über 200 g, 80 Äpfel wiegen 200 g oder weniger. Der Schätzer für den Anteil ist $\hat{p} = \frac{20}{100} = 0.2$. Die Approximationsbedingung mit $n \cdot \hat{p}(1 - \hat{p}) = 100 \cdot 0.2 \cdot 0.8 = 16 > 9$ ist erfüllt.

$$\begin{aligned}
KI &= \left[0.2 - c_{N(0,1),0.975} \cdot \sqrt{\frac{0.2(1-0.2)}{100}}; 0.2 + c_{N(0,1),0.975} \cdot \sqrt{\frac{0.2(1-0.2)}{100}} \right] \\
&= \left[0.2 - 1.96 \cdot \sqrt{\frac{0.2 \cdot 0.8}{100}}; 0.2 + 1.96 \cdot \sqrt{\frac{0.2 \cdot 0.8}{100}} \right] \\
&= [0.2 - 1.96 \cdot 0.04; 0.2 + 1.96 \cdot 0.04] \\
&= [0.1216; 0.2784].
\end{aligned}$$

Mit 95%-iger Wahrscheinlichkeit liegt der Anteil der Äpfel mit einem Gewicht von größer als 200 g zwischen 12.16% und 27.8%.

7.3 Statistische Tests

Die Testtheorie wurde erstmals 1930 von NEUMANN und PEARSON entwickelt. Statistische Tests werden zur Überprüfung einer statistischen Hypothese über eine oder mehrere Grundgesamtheiten auf der Grundlage einer Stichprobe durchgeführt.

Beispiel 7.16. Ein Tankstellenbesitzer hat die Oktanzahl des gelieferten Treibstoffes gemessen. 11 Messungen ergaben ein arithmetisches Mittel von 94,0. Der Vertrag mit dem Benzinhersteller zieht jedoch eine mittlere Oktanzahl von 95,0 vor.

Frage: Liegen systematische Qualitätsmängel vor die zur Vertragskündigung berechtigen? Diese Frage zielt darauf ab, ob die Abweichung signifikant ist. Beziehungsweise kann auch die Frage gestellt werden, ob die gemessene Abweichung einen rein zufälligen Charakter hat. Hier bezeichnet man dies als nicht signifikante Abweichung.

Signifikanztests geben Aussagen über solche oder ähnliche Fragen. Bei Signifikanztests gibt es 2 Hypothesen:

- H_0 - Nullhypothese
- H_1 - Alternativhypothese (Negation von H_0)

Mittels einer konkreten Stichprobe ist zu überprüfen, ob H_0 mit der Stichprobe verträglich ist. Wenn Verträglichkeit vorliegt, dann wird H_0 nicht abgelehnt, d.h. es liegt kein signifikanter Widerspruch vor.

Es gibt weitere Tests, welche hier nicht behandelt werden, zum Beispiel Bayes-Tests, Sequentielle Tests und randomisierte Tests.

Die Signifikanztests können nach ihrer Problemstellung klassifiziert werden.

Problemstellungen bei Signifikanztests:

1. *Signifikanztests für eine einfache Stichprobe:* In einer Grundgesamt ist ein Merkmal relevant. Die Hypothese bezieht sich auf die Verteilung dieses Merkmals. Zum Beispiel für das Gewicht eines produzierten Teils wird geprüft, ob dies signifikant vom Sollgewicht abweicht.
2. *Signifikanztests für 2 oder mehrere unabhängige Stichproben:* In zwei oder mehreren Grundgesamten ist ein Merkmal relevant. Die Hypothese beinhaltet einen Vergleich der Merkmalsverteilungen. Zum Beispiel kann geprüft werden, ob verschiedene PKW-Typen im Mittel den gleichen Benzinverbrauch haben.
3. *Signifikanztests für 2 oder mehrere verbundene Stichproben:* In einer Grundgesamt sind zwei oder mehrere Merkmale relevant. Die Hypothese bezieht sich auf die gemeinsame Verteilung der Merkmale. Zum Beispiel kann geprüft werden, ob bei einem Sortiment Bolzen die Merkmale Schaftdurchmesser und -länge voneinander unabhängig sind.

Allgemeines Vorgehen bei Signifikanztests:

1. Aufstellen der Hypothesen H_0 und H_1
2. Wahl des Signifikanzniveaus α
3. Berechnung der Realisierung t der Testfunktion T für eine konkrete Stichprobe
4. Bestimmung des kritischen Bereiches K zum Signifikanzniveau α
5. Testentscheidung

Bemerkung 7.17. Die Testgröße T muss bekannt sein. Die Funktion $T = g(X_1, \dots, X_n)$ heißt Testfunktion.

Beispiel 7.18. $X \sim N(\mu, \sigma)$ und $T = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Nachfolgend werden Signifikanztests vorerst für eine einfache Stichprobe betrachtet.

7.3.1 Parametertests für eine einfache Stichprobe

Für eine Grundgesamtheit G sei ein Merkmal X relevant. Es wird eine Stichprobe vom Umfang n aus G entnommen. X sei entsprechend einer Verteilung F verteilt mit $\mu = E(X)$ und $\sigma^2 = D^2(X)$. F_0 , μ_0 und σ_0^2 seien die entsprechenden hypothetischen Werte. Ähnlich zu den Konfidenzintervallen sei ein Parameter Θ unbekannt. Für den Parametertest wird ein Bereich, der sogenannte Parameterbereich Γ angegeben. Der Parameterbereich ist die Menge aller möglichen Werte, die ein unbekannter Parameter annehmen kann, wobei $\Gamma = \mathbb{R}$ im Allgemeinen ist. Es wird nun eine parametrische Hypothese aufgestellt, dass heißt Γ wird in zwei disjunkte Teilbereiche (Teilintervalle) mittels der Hypothesen H_0 und H_1 zerlegt:

- $H_0: \Theta \in \Gamma_0$

- $H_1: \Theta \in \Gamma_1 = \Gamma \setminus \Gamma_0$, wobei $\Gamma_0 \cup \Gamma_1 = \Gamma$ und $\Gamma_0 \cap \Gamma_1 = \emptyset$

Ein einseitiger Test wird durchgeführt, wenn behauptet wird, der Parameter Θ sei kleiner gleich oder größer gleich einem bestimmten Wert Θ_0 .

Definition 7.19. linksseitiger Test: $H_0: \Theta \geq \Theta_0$ und $H_1: \Theta < \Theta_0$, das heißt $\Gamma_0 = [\Theta_0, \infty)$ und $\Gamma_1 = (-\infty, \Theta_0)$

Definition 7.20. rechtsseitiger Test: $H_0: \Theta \leq \Theta_0$ und $H_1: \Theta > \Theta_0$, das heißt $\Gamma_0 = (-\infty, \Theta_0]$ und $\Gamma_1 = (\Theta_0, \infty)$

Ein zweiseitiger Test wird durchgeführt, wenn behauptet wird, der Parameter Θ besäße einen ganz bestimmten Wert Θ_0 .

Definition 7.21. zweiseitiger Test: $H_0: \Theta = \Theta_0$ und $H_1: \Theta \neq \Theta_0$, das heißt $\Gamma_0 = \{\Theta_0\}$ und $\Gamma_1 = \mathbb{R} \setminus \{\Theta_0\}$

Allgemeines Vorgehen bei einem Parametertest:

1. Aufstellen der Hypothesen $H_0: \Theta = \Theta_0$ und $H_1: \Theta \neq \Theta_0$. Θ_0 ist ein angegebener Wert für Θ . Zum Beispiel gibt der Hersteller einen Mittelwert einer Eigenschaft eines Ereignisses an.
2. Wahl des Signifikanzniveaus α , wobei $0 < \alpha < 1$ und meistens wird $\alpha = 0.05$ bzw. $\alpha = 0.01$ gesetzt.
3. Bestimmung der Testfunktion $T = g(X_1, \dots, X_n)$ die dem Problem angepasst ist. Berechnung der Realisierung t für eine konkrete Stichprobe $g(x_1, \dots, x_n)$.
4. Bestimmung des kritischen Bereiches K zum Signifikanzniveau α . Bei einem zweiseitigen Test werden die untere und obere Grenze c_u und c_o bestimmt, so dass gilt $P(c_u \leq T \leq c_o) = 1 - \alpha$. Der kritische Bereich ist damit $K = (-\infty, c_u) \cup (c_o, \infty)$, siehe Abbildung 7.3.
5. Testentscheidung:

- $t \in K \leadsto$ Ablehnung von H_0 und Annahme von H_1
- $t \notin K \leadsto$ keine Ablehnung von H_0 . **Achtung: H_0 darf auch nicht angenommen werden. Nur weil H_0 nicht abgelehnt wird, heißt das nicht, dass H_0 angenommen werden darf.**

Bemerkung 7.22. Fällt ein Testfunktionswert t in den Ablehnungsbereich, dann wird die Abweichung vom zu prüfenden Wert als signifikant (bedeutsam) angesehen, d.h. die Abweichung erscheint zu groß, um noch als zufallsbedingtes Abweichen erklärt werden zu können.

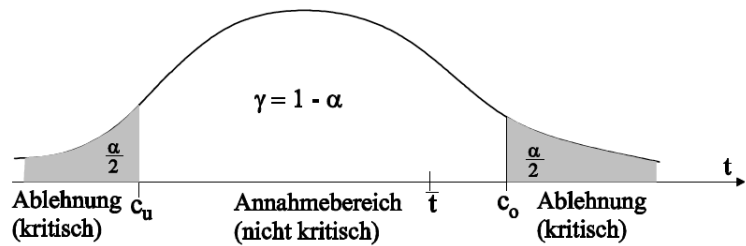


Abbildung 7.3: Darstellung des Annahmebereiches (nicht kritischer Bereich) und der Ablehnungsbereiche (kritische Bereiche) bei einem zweiseitigen Test.

Fehlerarten beim Signifikanztest: Beim Prüfen statistischer Hypothesen sind immer Fehlentscheidungen möglich (unabhängig vom Entscheidungsausgang, kann die getroffene Entscheidung richtig oder falsch sein). Es treten folgende Fehlerarten auf:

	H_0 richtig	H_0 falsch
Nichtablehnung von H_0	richtige Entscheidung	Fehler 2. Art: $P(\text{Fehler 2. Art}) = \beta$
Ablehnung von H_0	Fehler 1. Art: $P(\text{Fehler 1. Art}) = \alpha$	richtige Entscheidung

Ein Fehler 1. Art (α -Fehler) bedeutet, dass eine richtige Nullhypothese abgelehnt wird. Die Wahrscheinlichkeit dafür ist die Irrtumswahrscheinlichkeit α . Hingegen bedeutet der Fehler 2. Art (β -Fehler), dass eine falsche Nullhypothese nicht abgelehnt wird. Die Wahrscheinlichkeit dafür ist β . Der Zusammenhang zwischen α und β besteht darin, dass bei gleichem Stichprobenumfang eine Verkleinerung α von zu einer Vergrößerung von β führt (und umgekehrt). Weiter kann das Risiko für einen Fehler 1. und 2. Art bei gleichbleibendem α verringert werden, indem der Stichprobenumfang n vergrößert wird, siehe Abbildung 7.4. Die Berechnung des β -Fehlers ist nur dann möglich, wenn der wahre Wert bekannt ist. Dies ist in der Praxis nur selten der Fall.

7.3.2 Test für den Erwartungswert

Ähnlich zu den Konfidenzintervalle wird auch in der Testtheorie nach den Vorraussetzun-gen unterschieden.

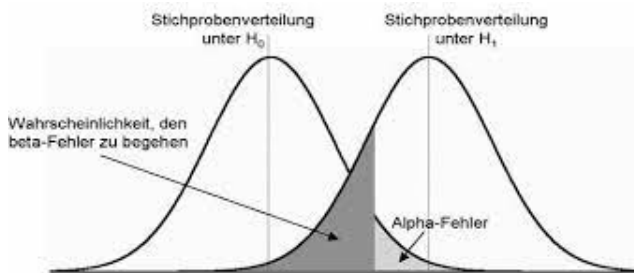


Abbildung 7.4: Diese Abbildung stellt den Zusammenhang zwischen den α - und β -Fehler dar. Die hellgraue Fläche repräsentiert den α -Fehler und die dunkelgraue Fläche den β -Fehler.

7.3.2.1 Der Einstichproben-Gauß-Test

Der Einstichproben-Gauß-Test ist ein spezieller Parametertest. Es ist der Test für den Erwartungswert μ einer Normalverteilten Grundgesamtheit, dass heißt $X \sim N(\mu, \sigma)$ bei bekannter Varianz σ^2 . Für die Hypothesen gibt es drei Möglichkeiten:

- $H_0 : \mu = \mu_0 \leadsto$ zweiseitiger Test
- $H_0 : \mu \geq \mu_0 \leadsto$ einseitiger Test (linksseitiger Fall)
- $H_0 : \mu \leq \mu_0 \leadsto$ einseitiger Test (rechtsseitiger Fall)

Die 5 Schritte des Einstichproben-Gauß-Test für den zweiseitigen Fall:

Ein zweiseitiger Test für den Erwartungswert wird durchgeführt, wenn behauptet wird, der Parameter μ besäße einen ganz bestimmten Wert μ_0 .

1. $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
2. Signifikanzniveau α festlegen $\leadsto \frac{\alpha}{2}$ wird für die kritischen Bereiche angesetzt
3. $T = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right) \leadsto T = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} \sim N(0, 1)$ und damit ist die Realisierung $t = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$
4. $K = \left(-\infty, c_{N(0,1), \frac{\alpha}{2}}\right) \cup \left(c_{N(0,1), 1 - \frac{\alpha}{2}}, \infty\right)$
5. Testentscheidung:

- $t \in K \leadsto$ Ablehnung von H_0 und Annahme von H_1 mit einer Irrtumswahrscheinlichkeit von α
- $t \notin K \leadsto$ keine Ablehnung von H_0

Beispiel 7.23. Das Gewicht von Brötchen sei normalverteilt mit $\sigma = 6g$. Für 81 zufällig ausgewählte Brötchen ergab sich ein Durchschnittsgewicht von $\bar{x} = 37g$. Überprüfen Sie

die Behauptung des Bäckers „Das Durchschnittsgewicht der Brötchen beträgt 38g“ zum Signifikanzniveau $\alpha = 5\%$.

$\mu_0 = 38g$ ist der hypothetische Wert

1. $H_0 : \mu = 38g$ und $H_1 : \mu \neq 38g$
2. $\alpha = 0.05 \leadsto \frac{0.05}{2}$ wird für die kritischen Bereiche angesetzt
3. $t = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} = \frac{37-38}{6} \sqrt{81} = -\frac{3}{2}$
- 4.

$$\begin{aligned}
 K &= \left(-\infty, c_{N(0,1), \frac{\alpha}{2}}\right) \cup \left(c_{N(0,1), 1-\frac{\alpha}{2}}, \infty\right) \\
 &= \left(-\infty, -c_{N(0,1), 1-\frac{\alpha}{2}}\right) \cup \left(c_{N(0,1), 1-\frac{\alpha}{2}}, \infty\right) \\
 &= \left(-\infty, -c_{N(0,1), 0.975}\right) \cup \left(c_{N(0,1), 0.975}, \infty\right) \\
 &= \left(-\infty, -1.96\right) \cup \left(1.96, \infty\right)
 \end{aligned}$$

5. $t \notin K \leadsto$ keine Ablehnung von $H_0 \leadsto$ Abweichung zwischen \bar{x} und μ_0 nicht signifikant.

Einseitiger Einstichproben-Gauß-Test: Ein einseitiger Test für den Erwartungswert wird durchgeführt, wenn behauptet wird, der Parameter μ sei kleiner gleich oder größer gleich einem bestimmten Wert μ_0 . Bei einseitigen Tests verhält sich der Ablauf bis auf die Schritte 1 und 5 analog.

- linksseitiger Test: $H_0 : \mu \geq \mu_0$ und $H_1 : \mu < \mu_0$ und $K = (-\infty, c_{N(0,1), \alpha})$ (mindestens)
- rechtsseitiger Test: $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$ und $K = (c_{N(0,1), 1-\alpha}, \infty)$ (höchstens)

Bemerkung 7.24. Bei einseitigen Tests ist stets zu überlegen, welcher der beiden Tests auf Grund der Stichprobe überhaupt sinnvoll ist. Eine Richtung ist durch die Stichprobe offensichtlich und nicht zu testen. Der Stichprobenwert ist somit Grundlage für die Alternativhypothese.

Beispiel 7.25. (Fortsetzung 7.23) Überprüfen Sie die Behauptung des Bäckers „Das Durchschnittsgewicht der Brötchen beträgt mindestens 38g“ zum Signifikanzniveau $\alpha = 5\%$.

linksseitiger Test: $H_0 : \mu \geq 38g$ und $H_1 : \mu < 38g$ und

$$\begin{aligned}
 K &= \left(-\infty, c_{N(0,1), 0.05}\right) \\
 &= \left(-\infty, -c_{N(0,1), 0.95}\right) \\
 &= \left(-\infty, -1.645\right)
 \end{aligned}$$

Auch hier ist $t \notin K \leadsto$ keine Ablehnung von $H_0 \leadsto$ Abweichung zwischen \bar{x} und μ_0 nicht signifikant.

7.3.2.2 Der Einstichproben- t -Test

Der Einstichproben- t -Test für den Erwartungswert μ einer Normalverteilten Grundgesamtheit, dass heißt $X \sim N(\mu, \sigma)$ allerdings bei unbekannter Varianz σ^2 . Die Vorgehensweise dieses Test verhält sich analog zum obigen Gauß-Test. Lediglich die unbekannte Varianz muss berücksichtigt werden, welche zu geänderten Verteilung der Testfunktion führt. Die Varianz wird mittels der Stichprobenvarianz geschätzt. Es ergibt sich damit die Testfunktion

$$T = \frac{\bar{X} - \mu_0}{S} \sqrt{n} \sim t(n-1),$$

welche t -verteilt mit $n-1$ Freiheitsgraden ist. Die kritischen Bereiche für die verschiedenen Testsituationen sind

- $K = \left(-\infty, -c_{t(n-1), 1-\frac{\alpha}{2}}\right) \cup \left(c_{t(n-1), 1-\frac{\alpha}{2}}, \infty\right)$ für den zweiseitigen Test
- $K = \left(-\infty, c_{t(n-1), \alpha}\right)$ für den linksseitigen Test und
- $K = \left(c_{t(n-1), 1-\alpha}, \infty\right)$ für den rechtsseitigen Test.

Beispiel 7.26. Ein Hersteller produziert Zylinderscheiben mit dem Solldurchmesser 20.2 mm . Der Produktion wird eine Stichprobe von $n = 16$ entnommen. Man erhält einen mittleren Durchmesser $\bar{x} = 20.6\text{ mm}$ und eine empirische Standardabweichung $s = 0.5\text{ mm}$. Es ist die Nullhypothese $H_0 : \mu = \mu_0 = 20.2\text{ mm}$ gegen Alternativhypothese $H_1 : \mu \neq \mu_0$ zu einem Signifikanzniveau $\alpha = 0.05$ zu testen. Vorausgesetzt wird, dass der Durchmesser X der Zylinderscheiben eine normalverteilte Zustandsgröße ist.

1. $H_0 : \mu = 20.2\text{ mm}$ und $H_1 : \mu \neq 20.2\text{ mm}$
2. $\alpha = 0.05 \leadsto \frac{0.05}{2}$ wird für die kritischen Bereiche angesetzt
3. $T = \frac{\bar{X} - \mu_0}{S} \sqrt{n} \sim t(15)$ und $t = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{20.6 - 20.2}{0.5} \sqrt{16} = 3.2$
- 4.

$$\begin{aligned} K &= \left(-\infty, -c_{t(n-1), 1-\frac{\alpha}{2}}\right) \cup \left(c_{t(n-1), 1-\frac{\alpha}{2}}, \infty\right) \\ &= \left(-\infty, -c_{t(15), 0.975}\right) \cup \left(c_{t(15), 0.975}, \infty\right) \\ &= \left(-\infty, -2.131\right) \cup \left(2.131, \infty\right) \end{aligned}$$

5. $t \in K \leadsto$ Ablehnung von $H_0 \leadsto$ Annahme von $H_1 \leadsto$ Abweichung zwischen \bar{x} vom Sollwert μ_0 ist signifikant (Der Produktionsprozess muss unterbrochen werden).

7.3.2.3 Asymptotischer Einstichproben-Gauß-Test

Ein asymptotischer Einstichproben-Gauß-Test wird angewendet, wenn die Grundgesamtheit X beliebig verteilt und σ^2 bekannt oder unbekannt ist mit einer Stichprobengröße $n > 30$.

1. X beliebig verteilt und σ^2 bekannt: $T = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} \stackrel{n \rightarrow \infty}{\sim} N(0, 1)$
2. X beliebig verteilt und σ^2 unbekannt: $T = \frac{\bar{X} - \mu_0}{s} \sqrt{n} \stackrel{n \rightarrow \infty}{\sim} N(0, 1)$

Die Vorgehensweise ist analog zum Einstichproben-Gauß-Test.

7.3.3 Test für die Varianz - Der χ^2 -Streuungstest

Der Hypothesentest für die Varianz funktioniert im Prinzip genau so wie der Test auf den Erwartungswert. Genauer, der χ^2 -Streuungstest ist ein Test für die Varianz σ^2 einer Normalverteilten Grundgesamtheit, dass heißt $X \sim N(\mu, \sigma)$ bei unbekanntem Erwartungswert μ . Für die Hypothesen gibt es drei Möglichkeiten:

- $H_0 : \sigma^2 = \sigma_0^2 \leadsto$ zweiseitiger Test
- $H_0 : \sigma^2 \geq \sigma_0^2 \leadsto$ einseitiger Test (linksseitiger Fall)
- $H_0 : \sigma^2 \leq \sigma_0^2 \leadsto$ einseitiger Test (rechtsseitiger Fall)

Die 5 Schritte des χ^2 -Streuungstests für den zweiseitigen Fall: Ein zweiseitiger Test für die Varianz wird durchgeführt, wenn behauptet wird, der Parameter σ^2 besäße einen ganz bestimmten Wert σ_0^2 .

1. $H_0 : \sigma^2 = \sigma_0^2$ und $H_1 : \sigma^2 \neq \sigma_0^2$
2. Signifikanzniveau α festlegen $\leadsto \frac{\alpha}{2}$ wird für die kritischen Bereiche angesetzt
3. $T = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \leadsto$ Realisierung $t = \frac{(n-1)s^2}{\sigma_0^2}$
4. $K = \left(0, c_{\chi^2(n-1), \frac{\alpha}{2}}\right) \cup \left(c_{\chi^2(n-1), 1-\frac{\alpha}{2}}, \infty\right)$
5. Testentscheidung:
 - $t \in K \leadsto$ Ablehnung von H_0 und Annahme von H_1 mit einer Irrtumswahrscheinlichkeit von α
 - $t \notin K \leadsto$ keine Ablehnung von H_0

Einseitiger χ^2 -Streuungstests : Ein einseitiger Test für die Varianz wird durchgeführt, wenn behauptet wird, der Parameter σ^2 sei kleiner gleich oder größer gleich einem bestimmten Wert σ_0^2 . Bei einseitigen Tests verhält sich der Ablauf bis auf die Schritte 1 und 5 analog.

- linksseitiger Test: $H_0 : \sigma^2 \geq \sigma_0^2$ und $H_1 : \sigma^2 < \sigma_0^2$ und $K = (0, c_{\chi^2(n-1), \alpha})$
- rechtsseitiger Test: $H_0 : \sigma^2 \leq \sigma_0^2$ und $H_1 : \sigma^2 > \sigma_0^2$ und $K = (c_{\chi^2(n-1), 1-\alpha}, \infty)$

Beispiel 7.27. Die Zufallsgröße X beschreibe die zufällige Länge von Schrauben, wobei $\sigma_0 = 1.2 \text{ mm}$. Eine Stichprobe vom Umfang $n = 25$ ergab die empirische Standardabweichung $s = 1.5 \text{ mm}$. Es ist ein einseitiger Test der folgenden Form durchzuführen: Nullhypothese $H_0 : \sigma^2 \leq \sigma_0^2 = 1.44 \text{ mm}^2$ gegen die Alternativhypothese $H_1 : \sigma^2 > \sigma_0^2$ bei $\alpha = 0.01$.

1. $H_0 : \sigma^2 \leq 1.44 \text{ mm}^2$ und $H_1 : \sigma^2 > 1.44 \text{ mm}^2$
2. Signifikanzniveau $\alpha = 0.01$
3. $T = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(24) \curvearrowright$ Realisierung $t = \frac{24 \cdot 2.25}{1.44} = 37.5$
- 4.

$$\begin{aligned} K &= (c_{\chi^2(n-1), 1-\alpha}, \infty) \\ &= (c_{\chi^2(24), 0.99}, \infty) \\ &= (42.98, \infty) \end{aligned}$$

5. Testentscheidung: $t \notin K \curvearrowright$ keine Ablehnung von H_0

7.3.4 Test für den Anteilswert - Asymptotischer Einstichproben-Gauß-Test für den Anteilswert

Das Prinzip des asymptotischen Einstichproben-Gauß-Test für den Anteilswert ist analog zu den zuvor eingeführten Tests. Der Ausgangspunkt ist eine Null-Eins verteilte Grundgesamtheit X , das heißt $X \sim \text{Null-Eins}(p)$ mit $n > \frac{9}{p_0(1-p_0)}$. Für die Hypothesen gibt es drei Möglichkeiten:

- $H_0 : p = p_0 \curvearrowright$ zweiseitiger Test
- $H_0 : p \geq p_0 \curvearrowright$ einseitiger Test (linksseitiger Fall)
- $H_0 : p \leq p_0 \curvearrowright$ einseitiger Test (rechtsseitiger Fall)

Für die Null-Eins Verteilung wird dies wieder auf einen bekannten Fall zurückgeführt, das heißt $\hat{\mu} = \hat{p}$ und $\hat{\sigma}^2 = \hat{p}(1 - \hat{p})$. Es ergibt sich damit die Testfunktion

$$T = \frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \stackrel{n \rightarrow \infty}{\rightsquigarrow} N(0, 1).$$

Die Vorgehensweise ist analog zum Einstichproben-Gauß-Test.

Beispiel 7.28. Der Ausschussanteil einer Produktion elektronischer Bauteile sei laut Hersteller höchstens 4%. Die Entnahme einer Stichprobe vom Umfang $n = 300$ Bauelementen enthielt $k = 15$ funktionsuntüchtige Bauelemente. Die Angabe des Herstellers ist mit einem Signifikanzniveau $\alpha = 0.01$ zu überprüfen. Es wird die Nullhypothese $H_0 : p \leq p_0$ gegen die Alternativhypothese $H_1 : p > p_0$ geprüft. \curvearrowright rechtsseitiger Test

Es ist $\hat{p} = \frac{15}{300} = 0.05$. Wegen $300 > \frac{9}{p_0(1-p_0)} = \frac{9}{0.04(1-0.04)} = 234.375$ ist die Voraussetzung erfüllt.

1. $H_0 : p \leq 0.04$ und $H_1 : p > 0.04$

2. Signifikanzniveau $\alpha = 0.01$

3. $T = \frac{\hat{P}-p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \stackrel{n \rightarrow \infty}{\sim} N(0,1)$ und $t = \frac{0.05-0.04}{\sqrt{0.04(1-0.04)}}\sqrt{300} = 0.8839$

4. $K = (c_{N(0,1),1-\alpha}, \infty) = (c_{N(0,1),0.99}, \infty) = (2.326, \infty)$

5. $t \notin K \curvearrowright$ keine Ablehnung von H_0