

Skriptum

107.A04 Wahrscheinlichkeitstheorie und stochastische Prozesse für Informatik 4.0

Karl Grill

24.1.2014

Institut für Statistik und Wahrscheinlichkeitstheorie TU Wien

Unter Creative Commons Attribution Sharealike Lizenz

©2013, 2014 Karl Grill

Inhaltsverzeichnis

1. Grundlagen der Wahrscheinlichkeitstheorie	7
1.1. Die Axiome von Kolmogorov	7
1.2. Bedingte Wahrscheinlichkeiten	13
1.3. Stochastische Unabhängigkeit	15
2. Zufallsvariable	19
2.1. Motivation und Einführung	19
2.2. Mathematische Definition	20
2.3. Diskrete Zufallsvariable	21
2.3.1. Verteilungsfunktion einer diskreten Zufallsvariable	21
2.3.2. Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariable	23
2.3.3. Gemeinsame Verteilung zweier diskreter Zufallsvariablen	23
2.4. Stetige Zufallsvariable	24
2.4.1. Verteilungsfunktion einer stetigen Zufallsvariable	24
2.4.2. Dichte einer stetigen Zufallsvariable	25
2.4.3. Gemeinsame Verteilung zweier stetiger Zufallsvariablen	25
2.4.4. Bedingte Dichte von X unter Y	25
2.4.5. Transformationssatz für Dichten	25
2.5. Gemischte Zufallsvariable	26
2.6. Eigenschaften von Verteilungsfunktionen	26
2.6.1. Wahrscheinlichkeiten mithilfe der Verteilungsfunktion	27
2.7. Unabhängigkeit von Zufallsvariablen	27
2.8. Erwartungswert und Varianz	28
2.8.1. Erwartungswert $\mathbb{E}(X)$ bzw. μ_X	28
Eigenschaften des Erwartungswertes	29
2.8.2. Varianz	30
Verschiebungssatz	31
Eigenschaften der Varianz	31
Kovarianz	31
Ungleichung von Markov	32
Ungleichung von Chebychev	32
Ungleichung von Kolmogorov	32
2.9. Folgen von Zufallsvariablen	32
2.9.1. Schwaches Gesetz der großen Zahlen	33
2.9.2. Starkes Gesetz der großen Zahlen	34
2.9.3. Zentraler Grenzwertsatz	34
2.10. Diskrete Verteilungen	35
2.10.1. Binomialverteilung	35
2.10.2. Gleichverteilung	36
2.10.3. Geometrische Verteilung	37
2.10.4. Poissonverteilung	38

2.10.5. Hypergeometrische Verteilung	40
2.11. Stetige Verteilungen	41
2.11.1. Gleichverteilung	41
2.11.2. Exponentialverteilung	42
2.11.3. Gammaverteilung	44
2.11.4. Cauchy-Verteilung	45
2.11.5. Normalverteilung	46
2.12. Verteilungen Zusammenfassung	48
3. Markovketten	49
3.1. Stochastische Prozesse	50
3.2. Markovketten in diskreter Zeit	51
3.2.1. Übergangswahrscheinlichkeiten	51
3.2.2. Klasseneigenschaften	52
3.2.3. Markov Chain Monte Carlo	55
4. Statistik	57
4.1. Grundlagen der Statistik	57
4.1.1. Modelle	57
4.1.2. Aufgaben der Statistik	59
Die Statistik besitzt 2 Grundaufgaben	59
4.2. Schätztheorie	59
4.2.1. Punktschätzung	60
Momentenmethode	61
Maximum Likelihood Methode	62
4.2.2. Intervallschätzung	68
Konfidenzintervall für μ einer Normalverteilung mit bekannter Va-	
rianz σ^2 :	70
Konfidenzintervall für μ einer Normalverteilung mit unbekannter	
Varianz σ^2 :	71
Konfidenzintervall für σ^2 einer Normalverteilung:	72
approximatives Konfidenzintervall für Anteilswerte	73
4.3. Tests	73
4.3.1. Grundlagen	73
4.3.2. Spezielle Tests	77
Für μ , wenn σ^2 bekannt (Normalverteilt):	77
Für μ , wenn σ^2 unbekannt (Normalverteilt):	77
Für die Varianz einer Normalverteilung:	78
Für Anteilswerte	78
4.3.3. Anpassungstests	78
Chi-Quadrat-Anpassungstest	79
4.3.4. Tests und Konfidenzintervalle	81
5. Informationstheorie	83
5.1. Information und Entropie	83
5.1.1. Bedingte Entropie und die Kullback-Leibler Divergenz	85
5.2. Codes	88
5.2.1. Der Huffman-Code	88
5.2.2. Der Shannon-Code	89
5.2.3. Der Fanu-Code	89

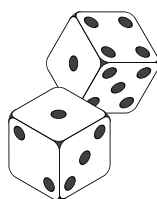
5.2.4. Weitere Codierungsverfahren	89
5.3. Informationsquellen	90
5.4. Blockcodes	91
5.5. Kanalcodierung	91
5.6. Natürliche Sprachen als Informationsquellen	91
A. Tabellen	93
A.1. Die Verteilungsfunktion der Standardnormalverteilung	93
A.2. Quantile z_p der Standardnormalverteilung	95
A.3. Quantile $t_{n;p}$ der t -Verteilung mit n Freiheitsgraden:	96
A.4. Quantile $\chi^2_{n;p}$ der χ^2 -Verteilung mit n Freiheitsgraden:	97
B. Mathematische Hintergründe	99
B.1. Matrizen	99
B.1.1. Determinante	99
Determinantenrechenregeln	99
Die Determinante einer 2x2 Matrix	99
Die Determinante einer 3x3 Matrix	99
B.1.2. Matrixpotenz	100
B.1.3. Diagonalmatrix D_x , bzw. Transformationsmatrix T bestimmen . . .	100
B.1.4. Matrix-Vector-Multiplikation	101
B.1.5. Matrix-Matrix-Multiplikation	101
B.1.6. Einheitsmatrix	101
B.1.7. Inverse Matrix	101
B.1.8. Elementare Zeilen/Spaltenoperationen	101
B.2. Quadratische Lösungsformel	102
B.3. Mengenlehre	102
B.4. Logarithmus-Rechenregeln	102
B.4.1. Produkte	102
B.4.2. Quotienten	102
B.4.3. Potenzen	102
B.4.4. Basisumrechnung	103
B.5. Ungleichungen	103
B.5.1. Umkehrbarkeit	103
B.5.2. Addition+Subtraktion	103
B.5.3. Multiplikation+Division	103
B.5.4. Potenzieren	103
B.6. Differenzengleichungen	103
B.6.1. lineare Differenzengleichungen 2.Ordnung	103
B.7. Binomischer Lehrsatz	104
B.8. Kombinatorik	104
B.8.1. Permutationen(Anordnungen)	104
Ohne Wiederholung	104
Mit Wiederholung	104
B.8.2. Kombinationen	105
Ohne Wiederholung	105
Mit Wiederholung	105
B.8.3. Variationen	105
Ohne Wiederholung	105
Mit Wiederholung	106

C. Glossar	107
Symbolverzeichnis	107

1. Grundlagen der Wahrscheinlichkeitstheorie

1.1. Die Axiome von Kolmogorov

Für das erste Beispiel im Gebiet der Wahrscheinlichkeitstheorie rollen wir zwei Würfel.



Steaphan Greene, CC BY-SA 3.0

Definition 1.1 *Alle möglichen Ausgänge kombiniert, nennen wir **Grundmenge**.*

Grundmenge M bzw. Ω

Die Grundmenge M bzw. Ω besteht aus dem Produkt der Ausgänge von Würfel A und der Ausgänge von Würfel B:

Würfel A: $A = \{1, 2, 3, 4, 5, 6\}$

Würfel B: $B = \{1, 2, 3, 4, 5, 6\}$

$$M = \Omega = A \times B = \{(1, 1), (1, 2), \dots, (3, 4), \dots, (6, 6)\}$$

Egal was gewürfelt wird, unser Ergebnis ist bereits in der Grundmenge M . Die Wahrscheinlichkeit \mathbb{P} dass das Ergebnis in der Grundmenge liegt, ist 100%.

$$\mathbb{P}(M) = \mathbb{P}(\Omega) = 1$$

Elementarereignis

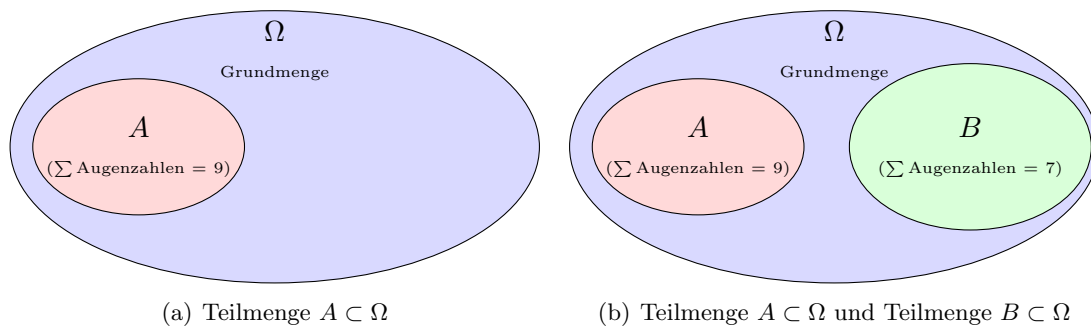
Ein einzelner Ausgang, also z.B. $(2, 5)$, wird Elementarereignis genannt.

Leere Menge \emptyset

Eine leere Menge enthält kein Element. Findet jedoch ein Zug statt (z.B. Würfel fallen), entsteht mit Gewissheit ein Ergebnis.

$$\mathbb{P}(\emptyset) = 0$$

Um Aussagen über bestimmte Ereignisse treffen zu können, können wir nun „Ergebnismengen“ konstruieren, die nur für uns interessante Elementarereignisse enthalten. Beispielsweise interessieren uns alle Würfe, bei denen die Summe der Augenzahlen 9 ergibt. Nun nehmen wir uns genau diese Elementarereignisse aus der Grundmenge: (siehe Abbildung 1.1(a))


Abbildung 1.1.: Teilmengen der Grundmenge Ω

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

Wir arbeiten also mit Mengen, die eine bestimmte Wahrscheinlichkeit besitzen. Im besten Fall enthält die Menge alle Elemente, im schlechtesten Fall keine. Es gilt:

$$0 \leq \mathbb{P}(A) \leq 1$$

Zusätzlich geben wir uns auch zufrieden, wenn die Summe der Augenzahlen 7 ergibt. (siehe Abbildung 1.1(b))

$$B = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

Diese Mengen sind **disjunkt**, besitzen also keine gleichen Elemente. Die Wahrscheinlichkeit, dass A oder B eintritt, kann durch die einfache Addition der einzelnen Wahrscheinlichkeiten berechnet werden:

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) \\ 0 \leq \mathbb{P}(A) \leq \mathbb{P}(B) \leq \mathbb{P}(A + B) \leq 1 \end{aligned}$$

Definition 1.2 (σ -Algebra) :

Da wir später öfters noch den Begriff σ -Algebra brauchen werden, möchte ich sie zuerst allgemein definieren. Diese Algebra kommt aus der Mengenlehre und besitzt drei wichtige Eigenschaften:

- $\Omega \in \mathcal{S}$
- Ist $A \in \mathcal{S}$ dann ist auch $A^C \in \mathcal{S}$
- Sind $A_i \in \mathcal{S}$, $i \in \mathbb{N}$, so ist auch $\cup_{i=1}^{\infty} A_i \in \mathcal{S}$

Angenommen wir werfen eine Münze. Dann ist die Grundmenge Ω (manchmal auch **Obermenge** bezeichnet):

$$\Omega = \{K, Z\}$$

und die Elementarereignisse $\{K\}$ und $\{Z\}$ Nur was ist jetzt \mathcal{S} ? Ohne weitere Angaben gilt:

$$\mathcal{S} = \{\emptyset, \{K\}, \{Z\}, \{K, Z\}\}$$

Beispiel 1.1 Angenommen wir werfen einen Würfel:

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

und die Elementarereignisse $\{1\}, \{2\}, \dots, \{6\}$. Ohne weitere Einschränkungen gilt

$$\mathcal{S} = \{\emptyset, \{1\}, \dots, \{1, 2\}, \dots, \{2, 4, 5\}, \dots, \{1, 2, 5, 6\}, \dots, \{1, 2, 3, 4, 6\}, \dots, \{1, 2, 3, 4, 5, 6\}\}$$

\mathcal{S} muss jedoch nicht immer die gesamte Potenzmenge sein. Möchten wir nur folgenden Fälle betrachten:

$$\begin{aligned} A &= \{2, 4, 6\} = \text{Gerade Zahl wird geworfen} \\ B &= \{1, 3, 5\} = \text{Ungerade Zahl wird geworfen} \\ C &= \{6\} \end{aligned}$$

Diese Algebra besteht aus 2^3 Elementen:

$$\mathcal{S} = \{\emptyset, \{6\}, \{2, 4, 6\}, \{1, 3, 5\}, \{1, 3, 5, 6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 5, 6\}\}$$

Die einzelnen Elemente lassen sich direkt aus den Eigenschaften einer σ -Algebra erstellen.

Mit diesem Wissen können wir uns nun den Axiomen von Kolmogorov zuwenden.

Definition 1.3 (Axiome von Kolmogorov) : Ist Ω eine Grundmenge und \mathcal{S} eine σ -Algebra von Ω . Eine Abbildung

$$\mathbb{P} : \mathcal{S} \rightarrow \mathbb{R}$$

heißt **Wahrscheinlichkeit** oder **Wahrscheinlichkeitsmaß**, wenn \mathbb{P} die folgenden drei Eigenschaften besitzt:

1. Wahrscheinlichkeit einer beliebigen Menge zwischen 0 und 1:

$$\forall A \in \mathcal{S} : 0 \leq \mathbb{P}(A) \leq 1$$

2. Wahrscheinlichkeit der leeren Menge ist 0:

$$\mathbb{P}(\emptyset) = 0$$

3. Ergebnis ist immer in der Grundmenge:

$$\mathbb{P}(\Omega) = 1$$

4. Addition von disjunkten Mengen:

Wenn $A_n, n \in \mathbb{N}$ disjunkte Mengen sind, dann gilt:

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

Das Tripel $(\Omega, \mathcal{S}, \mathbb{P})$ heißt **Wahrscheinlichkeitsraum**.

Bei unserem Würfelspiel (siehe Beispiel 1.1) mit den zwei geworfenen Würfeln können wir noch weitere Eigenschaften feststellen. Einerseits gibt es endlich viele mögliche Ausgänge ($|\Omega| < \infty$), andererseits besitzt jede Kombination k die gleiche Wahrscheinlichkeit ($|\Omega| = 36 \Rightarrow$

Würfel 1 \ Würfel 2	1	2	3	4	5	6
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
5	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
6	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Tabelle 1.1.: Würfelergebnisse und Wahrscheinlichkeiten mit 2 Würfeln (Beispiel 1.1)

$\mathbb{P}(k) = \frac{1}{|\Omega|} = \frac{1}{36}$) und schlussendlich schließen sich die Ereignisse aus. In Tabelle 1.1 sieht man alle möglichen Kombinationen sowie jede Einzelwahrscheinlichkeit.

Das führt uns zur Definition eines Laplace-Experiments:

Definition 1.4 (Laplace-Experiment) :

Im Laplace-Experiment mit endlich vielen Ausgängen besitzt jeder Ausgang die gleiche Wahrscheinlichkeit.

$$\mathbb{P}(A) = \frac{|A|}{|M|}$$

Definition 1.5 (Weitere Eigenschaften aus den Axiomen von Kolmogorov)

:

1. Wahrscheinlichkeit, dass A nicht auftritt:

$$\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$$

2. A ist eine Teilmenge von B ($A \subseteq B$)

Wenn $A \subseteq B$, dann gilt: $\mathbb{P}(A) \leq \mathbb{P}(B)$

3. Vereinigung von A und B :

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

4. Vereinigung ist kleiner, gleich der Summe

$$\text{Für } A_n \subseteq A_{n+1} : \mathbb{P}\left(\bigcup_n A_n\right) = \lim_n \mathbb{P}(A_n)$$

$$\text{Für } A_n \subseteq A_{n+1} : \mathbb{P}\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}(A_n)$$

Additionstheorem

Als Anwendung des Satzes 1.1 berechnen wir die Wahrscheinlichkeit, dass eine zufällig gewählte Permutation von n Elementen keinen Fixpunkt hat. Das Additionstheorem scheint auf den ersten Blick etwas kompliziert zu sein, ist aber ein einfaches Schema, das nur wiederholt werden muss.

Satz 1.1 Additionstheorem

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

$$\text{mit } S_i = \sum_{i \leq j_1 \leq j_2 \leq \dots \leq j_i \leq n} \mathbb{P}(A_{j_1} \cap \dots \cap A_{j_i})$$

Also somit:

$$\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) - + \dots$$

Vorher haben wir gesagt, dass zwei disjunkte Mengen einfach addiert werden können. Sind diese jedoch nicht vollständig disjunkt, es existieren also Elemente, die in beiden Mengen enthalten sind, so würden manche Elemente mehrmals gezählt werden. Das Beispiel 1.2 verdeutlicht dieses Problem genauer.

Formale Berechnung: Wahrscheinlichkeit, dass kein Fixpunkt existiert

Im Beispiel 1.2 ist ein Fixpunkt ein Ehepaar, das nach dem Signal zusammen bleibt. Wir benutzen in solchen Beispielen die Gegenwahrscheinlichkeit, betrachten den Fall, dass mindestens ein Fixpunkt A existiert.

Vereinigung aller Wahrscheinlichkeiten, dass ein Fixpunkt existiert:

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} S_i$$

Dafür müssen wir die Summanden S_k berechnen. Das Ereignis $A_1 \cap \dots \cap A_k$ tritt ein, wenn j_1, \dots, j_n Fixpunkte sind, die anderen $n-k$ Elemente können beliebig vertauscht werden. Es gibt $(n-k)!$ Möglichkeiten, diese anzuordnen. Insgesamt existieren $n!$ Möglichkeiten, die Wahrscheinlichkeit ist also $\frac{(n-k)!}{n!}$. Außerdem existieren $\binom{n}{k}$ Möglichkeiten, aus n Paaren, k auszuwählen. Wir schreiben für S_k :

$$S_k = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}$$

Somit können wir $\mathbb{P}(A)$ und die eigentlich wichtigere Wahrscheinlichkeit $\mathbb{P}(A^C)$ berechnen:

$$\mathbb{P}(A) = \sum_{k=1}^n (-1)^{k-1} \frac{1}{k!} \Rightarrow \mathbb{P}(A^C) = 1 - \mathbb{P}(A) = \sum_{k=0}^n (-1)^k \frac{1}{k!}$$

Die Umformung kann überprüft werden, indem die ersten paar Summanden aufgeschrieben werden. Für große n konvergiert die Reihe gegen $\frac{1}{e}$. Die Näherung ist so gut, dass sich die Anzahl der Permutationen ohne Fixpunkt (für $n \geq 1$) bestimmen lässt, indem man $\frac{n!}{e}$ auf die nächste ganze Zahl rundet.

Beispiel 1.2 (Ehepaare tanzen (Additionstheorem)) Nun stellen wir uns folgende Situation vor: Drei Ehepaare tanzen zusammen. Nach einer Weile ertönt ein Signal, die Namen aller Männer werden in ein Topf geworfen und jede Frau darf sich einen Namen ziehen. Nun stellt sich uns die Frage, wie hoch die Wahrscheinlichkeit ist, dass kein Ehepaar miteinander tanzt.

Dafür benutzen wir die Gegenwahrscheinlichkeit, $\mathbb{P}(\text{kein Ehepaar tanzt})$ ergibt sich durch $1 - \mathbb{P}(\text{mind. ein Ehepaar tanzt})$.

Ich definiere:

$$\mathbb{P}(E_i) = \text{Ehepaar } i \text{ tanzt nach dem Signal weiterhin miteinander}$$

Der schwarze Rahmen in Abbildung 1.2(a) steht für alle Ausgangsmöglichkeiten nach dem Signal, er ist die Grundmenge Ω . Darin befinden sich drei Kreise, jeder Kreis beinhaltet die Wahrscheinlichkeit, dass das entsprechende Ehepaar weiterhin miteinander tanzt.

Wie bereits erwähnt, benutzen wir die Gegenwahrscheinlichkeit. Dafür müssen wir die orange Fläche in Abbildung 1.2(b) berechnen.

Instinktiv würden wir jetzt folgende Gleichung aufschreiben:

$$1 - \mathbb{P}(\text{min. ein Ehepaar tanzt}) = 1 - \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3)$$

Jedoch wurden einige Flächenelemente doppelt gezählt. Diese sind in Abbildung 1.2(c) orange gekennzeichnet:

Damit unsere Gleichung wieder stimmt, müssen wir diese abziehen:

$$\begin{aligned} 1 - \mathbb{P}(\text{min. ein Ehepaar tanzt}) &= \\ &= 1 - \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) - \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1 \cap E_3) - \mathbb{P}(E_2 \cap E_3) \end{aligned}$$

Nun fehlt uns nur noch eine kleine Korrektur, denn durch den letzten Schritt haben wir das kleine Mittelstück einmal zu viel abgezogen.

$$\begin{aligned} 1 - \mathbb{P}(\text{min. ein Ehepaar tanzt}) &= \\ &= 1 - \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) - \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1 \cap E_3) \\ &\quad - \mathbb{P}(E_2 \cap E_3) + \mathbb{P}(E_1 \cap E_2 \cap E_3) \end{aligned}$$

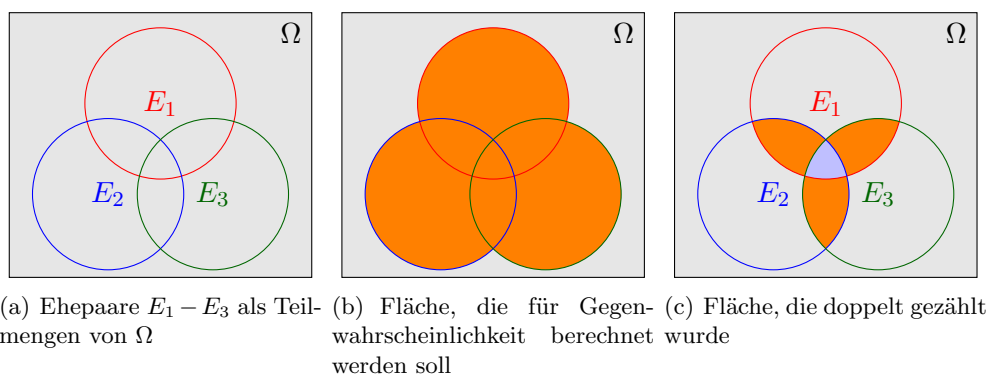


Abbildung 1.2.: Der Kreis E_1 gibt z.B. die Wahrscheinlichkeit an, dass das Ehepaar E_1 nach dem Ziehen wieder miteinander tanzt. Zu Beispiel 1.2

1.2. Bedingte Wahrscheinlichkeiten

Definition 1.6 Die Bedingte Wahrscheinlichkeit

Sie gibt an, wie wahrscheinlich ein Ereignis ist, wenn ein anderes Ereignis schon eingetreten ist.

$$\underbrace{\mathbb{P}(A|B)}_{\text{Wahrscheinlichkeit von } A \text{ unter der Bedingung } B} = \frac{\underbrace{\mathbb{P}(A \cap B)}_{\text{Wahrscheinlichkeit, dass } A \text{ und } B \text{ gleichzeitig eintreten.}}}{\underbrace{\mathbb{P}(B)}_{\text{Wahrscheinlichkeit, dass } B \text{ eintritt}}}$$

Diese Formel lässt sich am besten graphisch veranschaulichen. Dafür unterscheiden wir zwei Fälle:

Fall 1: disjunkte Mengen A, B
Siehe Abbildung 1.3(a)

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0}{\mathbb{P}(B)} = 0$$

Sobald wir wissen, dass B eingetreten ist, können wir mit Sicherheit behaupten, dass A nicht eingetreten ist.

Fall 2: nicht disjunkt. Mengen A, B
Siehe Abbildung 1.3(b)

Nun wissen wir, dass B eingetreten ist, unser Punkt liegt irgendwo in der blauen Fläche $\mathbb{P}(B)$.

Die Wahrscheinlichkeit, dass nun auch die rote Fläche getroffen wurde, ist das Verhältnis der blauen zur grünen Fläche.

Das wiederum ist

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

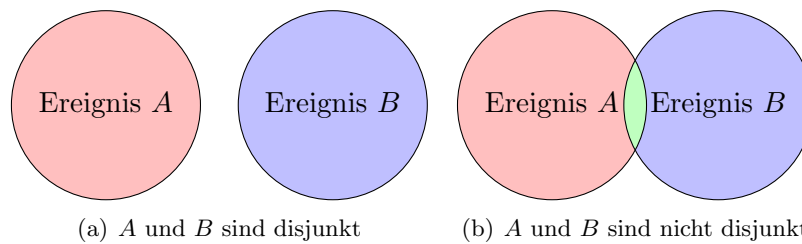


Abbildung 1.3.: Disjunkte und nicht Disjunkte Mengen erklärt

Satz 1.2 (Multiplikationssatz(Multiplikationstheorem)) :

Mit dem Multiplikationssatz wird die Wahrscheinlichkeit berechnet, dass die Ereignisse A_1, A_2, \dots, A_k gleichzeitig eintreten.

Dafür drehen wir die Formel der bedingten Wahrscheinlichkeit um und erweitern sie:

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

Die mehrfache Anwendung dieser Formel liefert:

$$\mathbb{P}(A_1 \cap \dots \cap A_k) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_k|A_1 \cap \dots \cap A_{k-1})$$

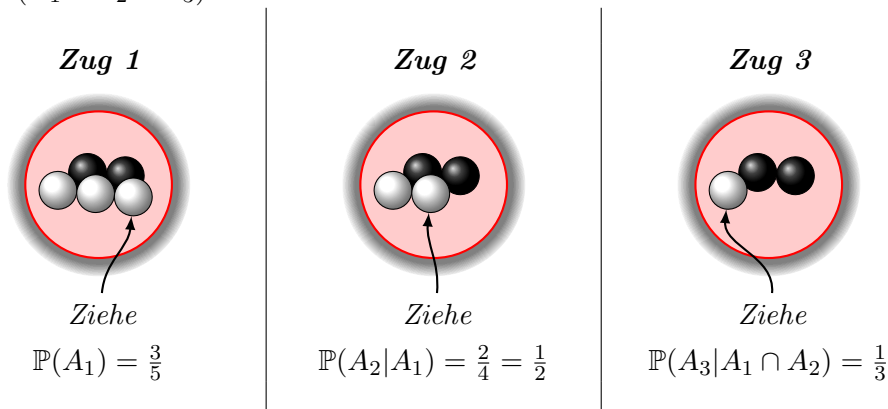
Die Ungleichung von Bonferroni kann das etwa Abschätzen.

$$\mathbb{P}\left(\bigcap_i A_i\right) = 1 - \sum_i \mathbb{P}(A_i^c)$$

Beispiel 1.3 (Ziehen ohne Zurücklegen (Multiplikationstheorem)) :

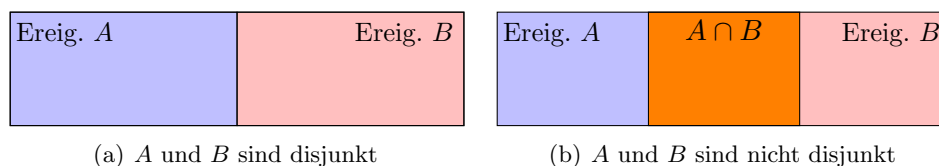
In einer Urne sind zwei schwarze und drei weiße Kugeln. Es wird dreimal ohne Zurücklegen gezogen und wir wollen die Wahrscheinlichkeit bestimmen, dass alle gezogenen Kugeln weiß sind.

Wir setzen also A_i gleich dem Ereignis, dass die i -te gezogene Kugel weiß ist und suchen $\mathbb{P}(A_1 \cap A_2 \cap A_3)$.



Nun wenden wir das Multiplikationstheorem an:

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) = \frac{3}{5} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{10}$$



Abbildungung 1.4.: Disjunkte und nicht disjunkte Tafeln

1.3. Stochastische Unabhängigkeit

Zwei Ereignisse heißen stochastisch unabhängig, wenn gilt:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \quad \text{bzw.} \quad \mathbb{P}(A|B) = \mathbb{P}(A) = \mathbb{P}(A|B^C)$$

Die Wahrscheinlichkeit des Eintretens von A hängt nicht davon ab, ob das Ereignis B oder B^C eintritt.

Beispiel 1.4 zur Erklärung:

Sie erhalten 10€, falls ein geworfenes Kreidestück in der oberen Hälfte der Tafel landet, ansonsten verlieren Sie 10€. Dabei stehen Sie in einem Nebenraum und überlegen sich, ob Sie die Wette annehmen sollen. In der Zwischenzeit wird die Kreide geworfen. Bevor Sie ihre Wette platzieren, kommt ein „Spion“ und erzählt Ihnen, dass er für einen gewissen Betrag verrät, ob die Kreide in der linken oder rechten Seite der Tafel gelandet ist.

Wie Wertvoll ist diese Information für Sie?

Das Angebot ist für Sie wertlos, denn die Information ändert nichts an Ihrem Wissen über den Wettausgang.

Zur Verdeutlichung siehe Abbildung 1.4: Die Grundmenge Ω wird durch den Rahmen dargestellt, darin befinden sich die Mengen A und B (disjunkt und überlappend). Es gilt $\mathbb{P}(M) = 1$, $\mathbb{P}(A) = 0,5$ und $\mathbb{P}(B) = 0,5$.

Siehe Abbildung 1.4(a)

Siehe Abbildung 1.4(b)

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \\ 0 &\neq 0.25 \end{aligned}$$

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \\ 0.25 &= 0.25 \end{aligned}$$

Wird B getroffen, ist

$$\mathbb{P}(A|B) = 0 \neq \mathbb{P}(A)$$

Die Ereignisse sind stochastisch abhängig, dass B getroffen wurde gibt uns viel Auskunft über die Wahrscheinlichkeit von A .

Wird B getroffen, ist

$$\mathbb{P}(A|B) = 0.5 = \mathbb{P}(A)$$

Die Ereignisse sind stochastisch unabhängig, dass B getroffen wurde, gibt uns keine Auskunft über die Wahrscheinlichkeit von A .

Definition 1.7 Totale Unabhängigkeit

Die n Ereignisse A_1, A_2, \dots, A_n heißen *total unabhängig*, falls für jede Auswahl A_1, \dots, A_k von k Ereignissen gilt:

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_k) = \mathbb{P}(A_1) \mathbb{P}(A_2) \dots \mathbb{P}(A_k)$$

Definition 1.8 Paarweise Unabhängigkeit

Die n Ereignisse A_1, A_2, \dots, A_n heißen *paarweise unabhängig*, wenn für alle $1 \leq i < j \leq n$ gilt:

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \mathbb{P}(A_j)$$

Beispiel 1.5 (aus Wikipedia [3]) :

In einer Schachtel befinden sich 4 Zettel mit folgenden Zahlenkombinationen: 112, 121, 211, 222. Einer der Zettel wird zufällig (je mit Wahrscheinlichkeit $1/4$) gezogen. Wir betrachten dann folgende drei Ereignisse:

$$A_1 = \{1 \text{ an erster Stelle}\} \text{ mit } \mathbb{P}(A_1) = \frac{1}{2}$$

$$A_2 = \{1 \text{ an zweiter Stelle}\} \text{ mit } \mathbb{P}(A_2) = \frac{1}{2}$$

$$A_3 = \{1 \text{ an dritter Stelle}\} \text{ mit } \mathbb{P}(A_3) = \frac{1}{2}$$

Offensichtlich sind die drei Ereignisse paarweise unabhängig, da gilt

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

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

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

Die drei Ereignisse sind jedoch nicht (gemeinsam) unabhängig, da gilt

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

Des Weiteren kann aus $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3)$ nicht geschlossen werden, dass die drei Ereignisse paarweise unabhängig sind.

Satz 1.3 (von der vollständigen Wahrscheinlichkeit) :

Für den Satz der Vollständigen bzw. totalen Wahrscheinlichkeit brauchen wir folgende Bedingungen:

- Disjunkte Ereignismengen B_i von denen $\mathbb{P}(B_i)$ bekannt ist
- $\mathbb{P}(A|B_i)$ muss bekannt bzw. leicht berechenbar sein
- $\sum_i \mathbb{P}(B_i) = 1$, somit wird die gesamte Grundmenge auf i Flächen aufgeteilt.

Wir möchten uns also die vollständige Wahrscheinlichkeit berechnen, dass A eintritt. Unser Problem liegt darin, dass wir nur wissen, wie sicher A eintritt, wenn bereits ein anderes Ereignis B_i eingetroffen ist (bedingte Wahrscheinlichkeit!). Somit ergibt sich:

$$\mathbb{P}(A) = \sum_i \mathbb{P}(B_i) \mathbb{P}(A|B_i)$$

Satz 1.4 (von Bayes) :

$$\mathbb{P}(A|B) \iff \mathbb{P}(B|A)$$

Ist die bedingte Wahrscheinlichkeit von B unter A , also $\mathbb{P}(B|A)$ bekannt, kann mit dem Satz von Bayes die Wahrscheinlichkeit von A unter B , also $\mathbb{P}(A|B)$ berechnet werden.

Dafür definieren wir die bedingte Wahrscheinlichkeit für beide Fälle, stellen um und erhalten:

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(B)} \Rightarrow \mathbb{P}(B \cap A) = \mathbb{P}(B) \mathbb{P}(A|B) \\ \mathbb{P}(B|A) &= \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \Rightarrow \mathbb{P}(B \cap A) = \mathbb{P}(A) \mathbb{P}(B|A) \\ \mathbb{P}(B) \mathbb{P}(A|B) &= \mathbb{P}(B \cap A) = \mathbb{P}(A) \mathbb{P}(B|A) \end{aligned}$$

Somit erhalten wir:

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

Beispiel 1.6 Krankheit (Vollständige Wahrscheinlichkeit & Satz von Bayes)

Von einer Krankheit sind 2% der Bevölkerung betroffen. Ein Test gibt bei einem Kranken mit Wahrscheinlichkeit 0.99 ein positives Ergebnis, bei einem Gesunden mit Wahrscheinlichkeit 0.01.

- a) Bestimmen Sie die Wahrscheinlichkeit, dass eine zufällig gewählte Person positiv getestet wird.

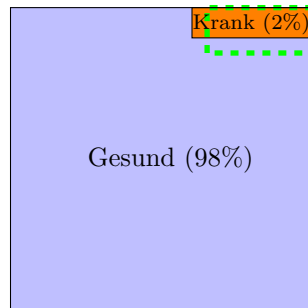


Abbildung 1.5.: Krankheit in Bezug auf Testergebnis (grün Strichliert) siehe Beispiel 1.6, (Achtung, übertrieben gezeichnet)

Abbildung 1.5 beschreibt die Situation. Der grüne Rahmen kennzeichnet den Wahrscheinlichkeitsbereich für ein positives Testergebnis. Es ist klar ersichtlich, dass die Trefferquote bei kranken Personen viel höher ist, als bei gesunden. Nun möchten wir die Wahrscheinlichkeit bestimmen, dass eine zufällig gewählte Person positiv getestet ist, also im grün-strichlierten Bereich liegt.

$$\mathbb{P}(A) = \text{Positiv getestet}$$

$$\mathbb{P}(B) = \text{Negativ getestet}$$

$$\mathbb{P}(A) = \mathbb{P}(B) \mathbb{P}(A|B) + \mathbb{P}(B^C) \mathbb{P}(A|B^C) = 0.98 \cdot 0.01 + 0.02 \cdot 0.99 = 0.0296 \approx 3\%$$

Hier wurde $\mathbb{P}(A)$ mit dem Satz der Vollständigen Wahrscheinlichkeit berechnet.

- b) Bestimmen Sie die bedingte Wahrscheinlichkeit dafür, dass eine zufällig gewählte Person krank ist, wenn das Testergebnis positiv ist.

Mit dem Wissen aus a) kann nun $\mathbb{P}(B^C|A)$ berechnet werden.

$$\mathbb{P}(B^C|A) = \frac{\mathbb{P}(B^C \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B^C) \mathbb{P}(A|B^C)}{\mathbb{P}(A)} = \frac{0.02 \cdot 0.99}{0.0296} = 0.6689 \approx 67\%$$

Hier wurde $\mathbb{P}(B^C \cap A)$ mit dem Satz von Bayes berechnet.

2. Zufallsvariable

2.1. Motivation und Einführung

Bisher hat uns immer direkt das Ergebnis eines Elementarereignisses interessiert, zum Beispiel haben wir uns gefragt, wie wahrscheinlich es ist, dass eine Münze Kopf zeigt oder eine Drei gewürfelt wird. Nun interessiert uns aber nicht mehr das Elementarereignis, sondern die „Informationen“, die aus mehreren Elementarereignissen entstehen. Am Einfachsten ist das mit einem Beispiel verständlich.

Beispiel 2.1 *Der Dozent bietet uns folgendes Spiel an: Jeder Student zahlt dem Dozenten 20 Cent Einsatz und wirft dann 3 Münzen.*

Bei den Münzen wird „Kopf“ $\triangleq K$ oder „Zahl“ $\triangleq Z$ registriert. Je nach der Anzahl der geworfenen „Köpfe“ zahlt der Dozent anschließend die Beträge in Tabelle 2.1

Wie wahrscheinlich sind die einzelnen Auszahlungen? Was muss der Dozent im Schnitt zahlen? Ist das Spiel fair?

Für die kürzere Schreibweise führen wir ein:

X_i ...Anzahl der vom i -ten Studenten bei 3 Versuchen geworfene Köpfe

Y_i ...Auszahlung an den i -ten Studenten

Wir betrachten den ersten Studenten:

Dieser wirft 2 „Köpfe“, also ist $X_1 = 2$.

Aber was heißt das und was ist der Unterschied zwischen X_1 und $X_1 = 2$?

Zufallsvariable	Realisation von X_1
X_1	$X_1 = 2$
Symbolische Kurzbeschreibung des Münzspiels	Die Realisation 2 ist eingetreten.

Nun stellen wir uns natürlich die Frage, wie Wahrscheinlich der Student zweimal Kopf würfelt. Dafür betrachten wir zuerst die Grundmenge Ω , also die Menge, die alle möglichen Ausgänge des Spiels beinhaltet:

$$\Omega = \{ZZZ, KZZ, ZKZ, ZZK, KKZ, KZK, ZKK, KKK\}$$

Wir gehen von einer fairen Münze aus, es gilt: $\mathbb{P}(Z) = \mathbb{P}(K) = \frac{1}{2}$. Zusätzlich wissen wir aus dem vorherigen Kapitel, dass wenn die Ereignisse total unabhängig sind, sich jede Wahrscheinlichkeit, zum Beispiel $\mathbb{P}(ZKZ)$, leicht berechnen lässt:

$$\mathbb{P}(ZKZ) = \mathbb{P}(Z)\mathbb{P}(K)\mathbb{P}(Z) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Anzahl der Köpfe	Auszahlung in Cent
0	0
1	0
2	20
3	100

Tabelle 2.1.: Auszahlungsbeträge bei 3 Münzwürfen

Ereignisse	Realisationen von X_1	Wahrscheinlichkeit P_1
$\{ZZZ\}$	0	$\mathbb{P}(X_1 = 0) = \frac{1}{8}$
$\{KZZ, ZKZ, KZZ\}$	1	$\mathbb{P}(X_1 = 0) = \frac{3}{8}$
$\{KKZ, KZK, ZKK\}$	2	$\mathbb{P}(X_1 = 0) = \frac{3}{8}$
$\{KKK\}$	3	$\mathbb{P}(X_1 = 0) = \frac{1}{8}$

Tabelle 2.2.: Abbildungen der Ereignisse in Wahrscheinlichkeiten ($\Omega \rightarrow \mathbb{R}$)

Das gilt trivialerweise auch für die anderen Ausgänge des Spiels. Und was macht die Zufallsvariable X_1 genau? Sie „bewertet“ die Elemente der Grundmenge Ω anhand der Anzahl von „Köpfen“:

$$\begin{aligned}
 X_1(ZZZ) &= 0 \\
 X_1(KZZ) &= X_1(ZKZ) = X_1(ZZK) = 1 \\
 X_1(KKZ) &= X_1(KZK) = X_1(ZKK) = 2 \\
 X_1(KKK) &= 3
 \end{aligned}$$

Mathematisch ausgedrückt ist eine Zufallsvariable eine Abbildung der Grundmenge Ω in die reellen Zahlen:

$$X_1 : \Omega \rightarrow \mathbb{R}$$

Diese Abbildung ist nicht bijektiv, wie sich mit Tabelle 2.2 leicht nachprüfen lässt:

Offensichtlich existieren jeweils drei Ereignisse, die auf 1 bzw. 2 abbilden und somit ist die Abbildung nicht eindeutig umkehrbar.

Anmerkung: für $\mathbb{P}(X_1 = 1)$ bzw. $\mathbb{P}(X_1 = 2)$ wurden die einzelnen Wahrscheinlichkeiten einfach addiert, da sich die Ereignisse paarweise ausschließen.

2.2. Mathematische Definition

Die Wahrscheinlichkeit des im vorherigen Kapitel (Siehe Definition 1.3) beschriebenen Wahrscheinlichkeitsraumes $(\Omega, \mathcal{S}, \mathbb{P})$ wird mithilfe von X auf einen neuen Wahrscheinlichkeitsraum abgebildet.

$$(\Omega, \mathcal{S}, \mathbb{P}) \Rightarrow X \Rightarrow (\mathcal{B}, \mathcal{F}, \mathbb{P}_X)$$

Dem Ereignis $F \in \mathcal{F}$ ordnen wir nun die Wahrscheinlichkeit seiner „Verursacher“ zu:

$$\mathbb{P}_X(F) = \mathbb{P}(X \in F) = \mathbb{P}(X^{-1}(F)) = \mathbb{P}(\{\omega : X(\omega) \in F\})$$

Definition 2.1 Das **Wahrscheinlichkeitsmaß** \mathbb{P}_X heißt das zu X gehörige **Bildmaß**. Vereinfacht kann X auch als Abbildung

$$X : \Omega \rightarrow \mathbb{R}^d$$

definiert werden. Dabei gilt:

- wenn Ω überabzählbar ist, muss man von X die Eigenschaft „Messbarkeit“ verlangen.
- meistens ist $d = 1$, bei $d > 1$ wird $X = (X_1, \dots, X_d)$ ein Vektor von reellen Zufallsvariablen.

Beispiel 2.2 anhand des Münzspiels (siehe Beispiel 2.1)

- $\Omega \Rightarrow \mathcal{B}$

$$\{ZZZ, KZZ, ZKZ, ZKK, KKZ, KZK, ZKK, KKK\} \Rightarrow \{0, 1, 2, 3\}$$

- $\mathcal{S} \Rightarrow \mathcal{F}$

Abhängig von der Anforderung

- $\mathbb{P} \Rightarrow \mathbb{P}_X$

$$\forall \omega \in \Omega : \mathbb{P}(\omega) = \frac{1}{8} \Rightarrow \begin{cases} \mathbb{P}_{X_1}(0) = \mathbb{P}_{X_1}(3) = \frac{1}{8} \\ \mathbb{P}_{X_1}(1) = \mathbb{P}_{X_1}(2) = \frac{3}{8} \end{cases}$$

Graphische Erklärung

Alle Elementarereignisse $\omega \in \Omega$, also die durch Zufall eingetretene Ereignisse (z.B. $\{ZZZ\}, \{KZZ\}, \{KKK\}$), werden über die Abbildung der Zufallsvariable X auf die reellen Zahlen und in die Menge \mathcal{B} abgebildet, dem Ereignis wird eine Realisation zugeordnet.

Abbildung 2.1 beschreibt das komplette System anschaulich:

Die blauen Bereiche repräsentieren eine Abbildung, die Quadrate einen vollständigen Wahrscheinlichkeitsbereich.

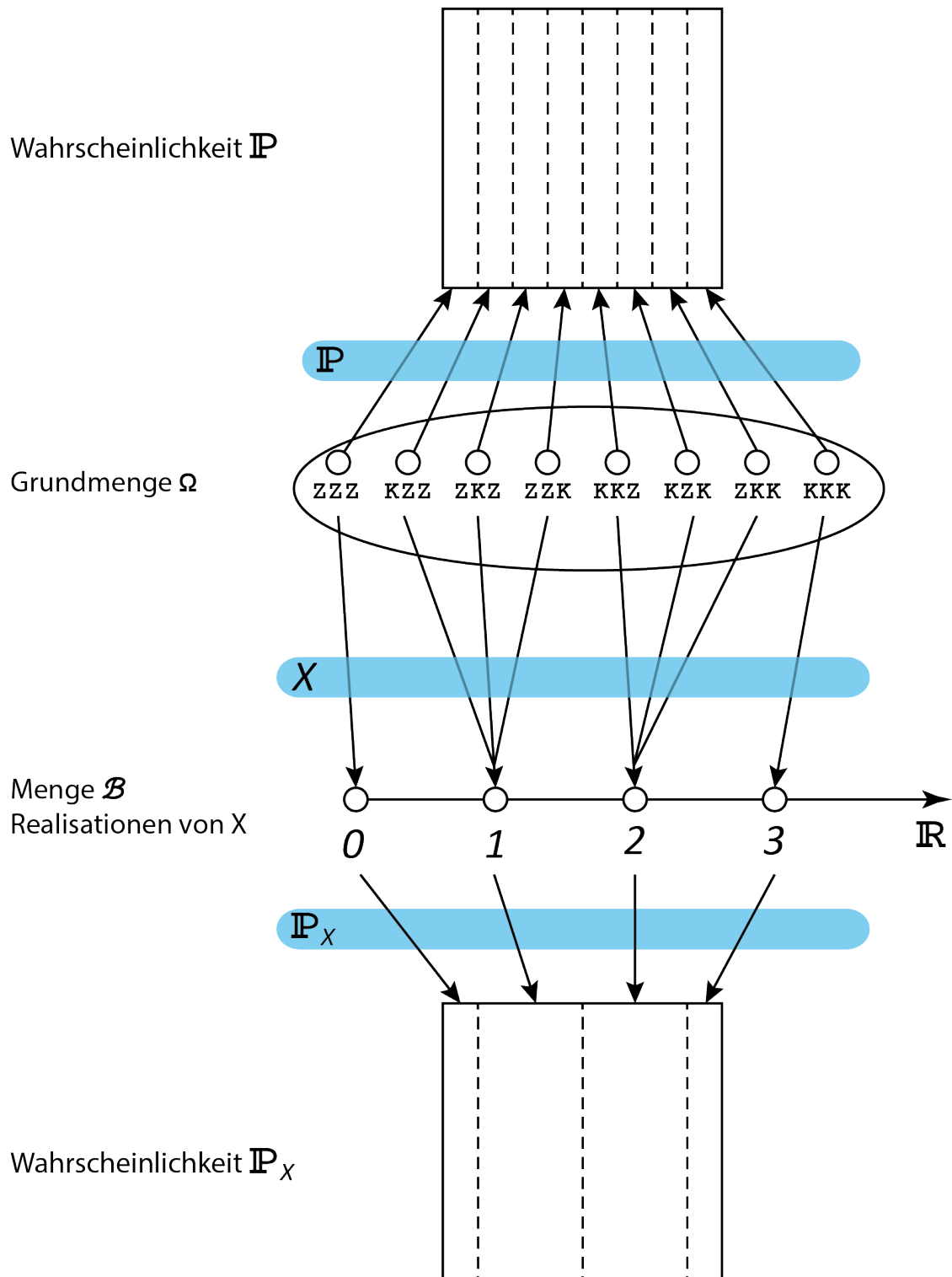
2.3. Diskrete Zufallsvariable

2.3.1. Verteilungsfunktion einer diskreten Zufallsvariable

Eine diskrete Zufallsvariable X besitzt endlich oder abzählbar unendlich viele Realisationen x_i , die mit Wahrscheinlichkeit $p_i = \mathbb{P}(X = x_i) > 0$ angenommen werden. Für jedes $x \in \mathbb{R}$ ist die Verteilungsfunktion (siehe Abbildung 2.2 die blaue Line) $F_X : \mathbb{R} \rightarrow [0, 1]$ der Zufallsvariablen X definiert durch:

$$F_X(x) = \mathbb{P}(X \leq x)$$

Die Verteilungsfunktion ist somit monoton steigend und immer zwischen 0 und 1. Für jeden Wert $F_X(x_i)$ an der Stelle i gilt:

Abbildung 2.1.: Abbildung des Wahrscheinlichkeitsraumes von X (siehe Beispiel 2.2)

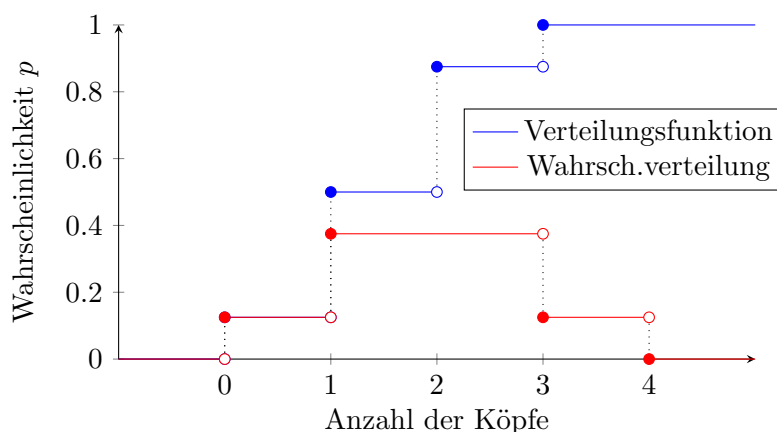


Abbildung 2.2.: die diskrete Verteilungsfunktion und die diskrete Wahrscheinlichkeitsverteilung von Beispiel 2.1

$$F_X(x_i) = \sum_{j=1}^i p_j = \sum_{j=1}^i \mathbb{P}(X = x_j)$$

2.3.2. Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariable

Siehe Abbildung 2.2 die rote Line.

Eine diskrete Zufallsvariable X besitzt endlich oder abzählbar unendlich viele Realisationen x_i , die mit Wahrscheinlichkeit $p_i = \mathbb{P}(X = x_i) > 0$ angenommen werden. Für diese x_i gilt:

$$\sum_{i=1}^{\infty} \mathbb{P}(X = x_i) = 1$$

Die Angabe aller $p_i, i = 1, \dots, \infty$ heißt die Dichte von X .

2.3.3. Gemeinsame Verteilung zweier diskreter Zufallsvariablen

Beispiel 2.3 Dafür erweitern wir unser Münzspiel(Beispiel 2.1) um folgende Annahme:

Es würfelt ein zweiter Student und wir führen dafür die Zufallsvariable X_2 ein. Den Dozent interessiert aber nur, wie viel er insgesamt auszahlen muss. Zu Beginn haben wir gesagt:

Kopfanzahl	0	1	2	3
Auszahlung [Cent]	0	0	20	100

Da uns nur noch der Auszahlungsbetrag interessiert führen wir eine neue Zufallsvariable Y ein.

Y_i ...Auszahlungsbetrag an den i -ten Studenten.

Summe S_2	$Y_1 = 0$	$Y_2 = 20$	$Y_3 = 100$
$Y_2 = 0$	0	20	100
$Y_2 = 20$	20	40	120
$Y_2 = 100$	100	120	200

Tabelle 2.3.: Auszahlungsbeträge bei 2 Studenten (siehe Beispiel 2.3)

$\mathbb{P}_{Y_1 Y_2}(y_1, y_2)$	$Y_1 = 0$	$Y_2 = 20$	$Y_3 = 100$
$Y_2 = 0$	$\frac{16}{64}$	$\frac{12}{64}$	$\frac{4}{64}$
$Y_2 = 20$	$\frac{12}{64}$	$\frac{9}{64}$	$\frac{3}{64}$
$Y_2 = 100$	$\frac{4}{64}$	$\frac{3}{64}$	$\frac{1}{64}$

Tabelle 2.4.: Auszahlungswahrscheinlichkeit bei 2 Studenten (siehe Beispiel 2.3)

Nun können wir für zwei Studenten mit Y_1 und Y_2 die Auszahlungssumme $S_2 = Y_1 + Y_2$ definieren (siehe Tabelle 2.3)

Anstatt den Auszahlungsbeträgen ist es für den Dozenten viel wichtiger, wie wahrscheinlich er diese Beträge auszahlen muss. Wir haben also eine Auszahlungswahrscheinlichkeit von Paaren:

$$\mathbb{P}_{Y_1 Y_2}(y_1, y_2) = \mathbb{P}(Y_1 = y_1, Y_2 = y_2)$$

Da diese zwei Zufallsvariablen voneinander unabhängig sind (genauere Definition folgt später), kann die Tabelle 2.4 mit der Eigenschaft $\mathbb{P}_{Y_1 Y_2}(y_1, y_2) = \mathbb{P}_1(y_1) \mathbb{P}_2(y_2)$ berechnet werden.

Hat man nur diese Tabelle und man möchte die Verteilung $\mathbb{P}_1(y_1)$ berechnen, dann geschieht das mit:

$$\mathbb{P}_1(y_1) = \sum_{y_2} \mathbb{P}_{Y_1 Y_2}(y_1, y_2)$$

Das ist nichts anderes als die Spalten- bzw. Zeilensumme.

2.4. Stetige Zufallsvariable

2.4.1. Verteilungsfunktion einer stetigen Zufallsvariable

In anderen Situationen haben wir keine diskreten Wahrscheinlichkeiten, sondern eine stetige Verteilung (siehe Abbildung 2.3, rote Linie). Für die Verteilungsfunktion wird die Summe zum Integral:

$$F_X(x) = \int_{-\infty}^x f_X(u) du$$

bzw.

$$F_X(x) = \int_{-\infty}^{x_d} \dots \int_{-\infty}^{x_1} f_X(u_1, \dots, u_d) dx_1 \dots dx_d$$

falls $X = (X_1, \dots, X_d)$ mehrdimensional ist.

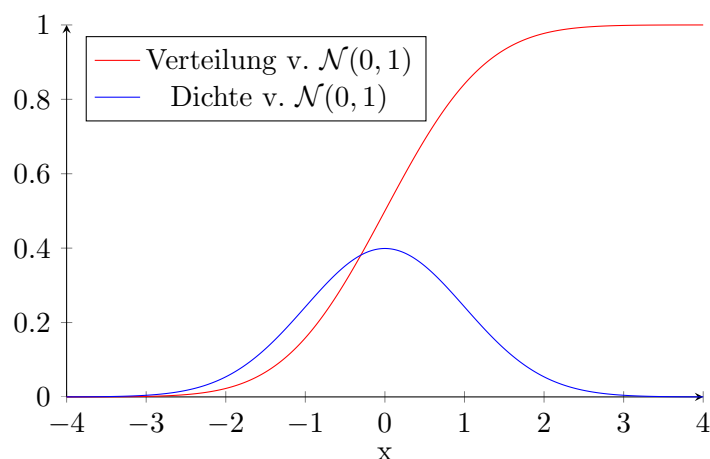


Abbildung 2.3.: die stetige Verteilungsfunktion und die Dichte der Standardnormalverteilung.

2.4.2. Dichte einer stetigen Zufallsvariable

Siehe Abbildung 2.3, blaue Linie.

f_X ist dann die Dichte der Verteilung von X , und wir nennen X stetig (verteilt).

2.4.3. Gemeinsame Verteilung zweier stetiger Zufallsvariablen

Für den stetigen Fall gilt das gleiche wie für diskrete Zufallsvariablen, jedoch benutzen wir für die Berechnung einer einzelnen Dichte das Integral:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

2.4.4. Bedingte Dichte von X unter Y

Definition 2.2 Gleich wie bei Ereignissen definieren wir:

X und Y seien stetig verteilt mit Dichte $f_{X,Y}$. Die bedingte Dichte von X unter $Y = y$ ist

$$f_X(x|Y=y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Damit erhalten wir die **Bedingte Wahrscheinlichkeit** als

$$\mathbb{P}(X \leq a | Y \leq y) = \int_{-\infty}^a f_X(x|Y=y) dx$$

2.4.5. Transformationssatz für Dichten

Definition 2.3 Es sei X eine stetige zufällige Variable mit der Verteilung F_X , bzw. der Dichte f_X und ψ eine streng monoton wachsende Funktion. Ist $Y = \psi(X)$, so ist

$$F_X(x) = F_Y(\psi(x))$$

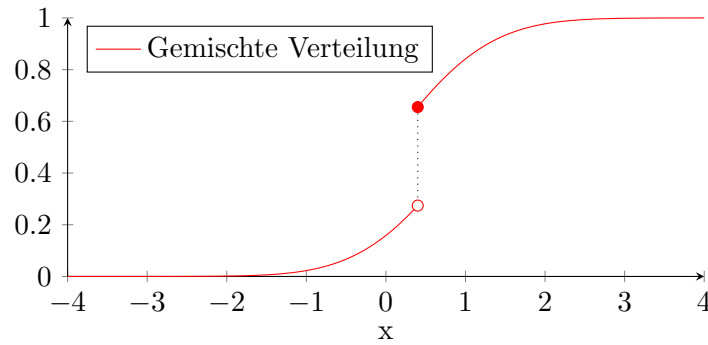


Abbildung 2.4.: Verteilungsfunktion einer Gemischten Zufallsvariable

Ist ψ streng monoton fallend, so ist

$$F_X(x) = 1 - F_Y(\psi(x))$$

In beiden Fällen ist

$$f_X(x) = f_Y(\psi(x)) \left| \frac{d\psi(x)}{dx} \right|$$

2.5. Gemischte Zufallsvariable

Wenn F_X sowohl Sprünge als auch eine nichtverschwindende Ableitung hat, dann nennen wir X gemischt verteilt. (Siehe Abbildung 2.4) In diesem Fall gibt es sowohl eine Wahrscheinlichkeitsfunktion als auch eine Dichte. Es gilt:

$$\mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x) + \int_A f_X(x) dx$$

2.6. Eigenschaften von Verteilungsfunktionen

Definition 2.4 (Eigenschaften der Verteilungsfunktion $F : \mathbb{R} \rightarrow \mathbb{R}$) :

- 1) $0 \leq F(x) \leq 1 \quad \forall x$
- 2) F ist monoton steigend
- 3) F ist rechtsstetig
- 4) $\lim_{x \rightarrow -\infty} F(x) = 0$
- 5) $\lim_{x \rightarrow \infty} F(x) = 1$

Definition 2.5 (Eigenschaften der Verteilungsfunktion $F : \mathbb{R}^2 \rightarrow \mathbb{R}$) :

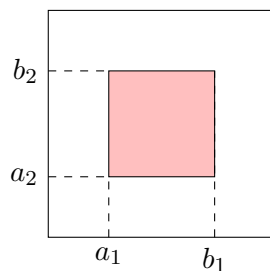


Abbildung 2.5.: 6. Eigenschaft der Verteilungsfunktion (rote Fläche):
 $\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2)$

- 1) $0 \leq F(x_1, x_2) \leq 1 \quad \forall x$
- 2) F ist monoton steigend (in jeder Argumentvariable)
- 3) F ist rechtsstetig
- 4) $\lim_{x_1 \rightarrow -\infty} F(x_1, x_2) = \lim_{x_2 \rightarrow -\infty} F(x_1, x_2) = 0$
- 5) $\lim_{x_1 \rightarrow \infty} F(x_1, x_2) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2) = 1$
- 6) Für $a_1 < b_1, a_2 < b_2$ gilt:
 $F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$

Die 6. Eigenschaft beschreibt die Wahrscheinlichkeit:
 $\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2)$
 Siehe Abbildung 2.5.

2.6.1. Wahrscheinlichkeiten mithilfe der Verteilungsfunktion

$$\begin{aligned} \mathbb{P}(X \leq a) &= F_X(a), \\ \mathbb{P}(X < a) &= F_X(a - 0), \\ \mathbb{P}(a < X \leq b) &= F_X(b) - F_X(a), \\ \mathbb{P}(a < X < b) &= F_X(b - 0) - F_X(a), \\ \mathbb{P}(a \leq X \leq b) &= F_X(b) - F_X(a - 0), \\ \mathbb{P}(a \leq X < b) &= F_X(b - 0) - F_X(a - 0), \\ \mathbb{P}(X = a) &= F_X(a) - F_X(a - 0). \end{aligned}$$

Dabei ist $F(x - 0) = \lim_{h \rightarrow 0} F(x - h)$ der linksseitige Grenzwert von F in x .

2.7. Unabhängigkeit von Zufallsvariablen

Definition 2.6 Die Zufallsvariablen (X, Y) heißen unabhängig, wenn für alle x_1, \dots, x_n

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i).$$

Die unendliche Folge $(X_n, n \in \mathbb{N})$ heißt unabhängig, wenn jede endliche Teilfolge unabhängig ist.

Wenn die gemeinsame Verteilung diskret bzw. stetig ist, kann man in dieser Definition die Verteilungsfunktion durch die Wahrscheinlichkeits- bzw. Dichtefunktion ersetzen.

Dafür müssen wir nur die Eigenschaften der Ereignisse auf Zufallsvariablen übertragen:

Definition 2.7 X und Y heißen unabhängig, wenn $\forall x \in \mathbb{R}$ und $y \in \mathbb{R}$ die Ereignisse $X \leq x$ und $Y \leq y$ unabhängig sind.

Das heißt, es muss gelten

$$\mathbb{P}(X \leq x | Y \leq y) = \mathbb{P}(X \leq x)$$

oder gleichwertig

$$\mathbb{P}(X \leq x \cap Y \leq y) = \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y)$$

Mit mehreren Zufallsvariablen muss man wieder zwischen paarweiser und totaler Unabhängigkeit unterscheiden. Die n zufälligen Variablen X_1, \dots, X_n heißen total unabhängig, wenn für alle x_1, \dots, x_n gilt:

$$\mathbb{P}_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n \mathbb{P}_i(x_i)$$

Satz 2.1 (Faltung von Dichten) X und Y seien unabhängig mit Dichte f_X und f_Y . Dann ist die Dichte von $X + Y$ die Faltung von f_X und f_Y :

$$f_{X+Y}(z) = f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx.$$

2.8. Erwartungswert und Varianz

2.8.1. Erwartungswert $\mathbb{E}(X)$ bzw. μ_X

Im folgenden werden die Symbole \mathbb{E} und μ für den Erwartungswert verwendet.

Definition 2.8 (von Erwartungswert \mathbb{E} bzw. μ_X von X) :

Für eine mit Wahrscheinlichkeitsverteilung $\mathbb{P}(X = x_j), j = 1, 2, \dots, \infty$ verteilte **diskrete Zufallsvariable** Y gilt:

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} x_j \cdot \mathbb{P}(X = x_j)$$

sofern die unendliche Reihe absolut konvergiert. Ansonsten existiert kein Erwartungswert.

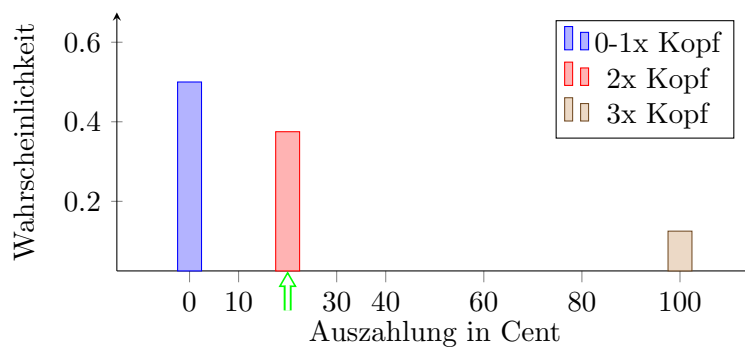


Abbildung 2.6.: Wahrscheinlichkeitsverteilung für Auszahlung (grüner Pfeil ist Schwerpunkt) siehe Tabelle 2.1

Für *stetige Zufallsvariablen* gilt:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Falls Y *gemischt verteilt* ist, gilt:

$$\mathbb{E}(X) = \sum_x x \cdot \mathbb{P}(X = x_j) + \int_{-\infty}^{\infty} x f_X(x) dx$$

Betrachten wir das anhand unseres Münzspiels (siehe Tabelle 2.1).

Mit der Wahrscheinlichkeitsverteilung für die Auszahlungen (siehe Abbildung 2.6):
Für den Schwerpunkt (grüner Pfeil in Abbildung 2.6) berechnen wir:

$$\mathbb{E}(X) = 0 \cdot \frac{4}{8} + 20 \cdot \frac{3}{8} + 100 \cdot \frac{1}{8} = 20$$

ACHTUNG: Der Erwartungswert einer Zufallsvariablen ist der Schwerpunkt (arithmetisches Mittel) ihrer Verteilung.

Eigenschaften des Erwartungswertes

- 1) **Additivität** Sind X, Y zwei Zufallsvariablen, egal ob abhängig oder nicht, mit den Erwartungswerten $\mathbb{E}(X), \mathbb{E}(Y)$, dann gilt:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

- 2) **Linearität**

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

- 3) **Monotonie**

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

- 4) **Unabhängigkeit:** Wenn X und Y unabhängig sind, dann gilt:

$$\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y)$$

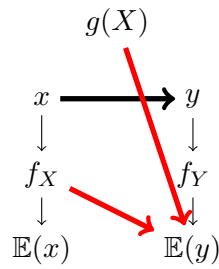


Abbildung 2.7.: Satz des unachtsamen Statistikers (siehe Satz 2.2)

Satz 2.2 des unachtsamen Statistikers Siehe Abbildung 2.7.

X sei stetig verteilt mit der Dichte f_X , $Y = g(X)$. Um den Erwartungswert von Y zu berechnen, müssen wir nicht die Verteilung von Y berechnen. Wir können direkt folgendes Integral verwenden:

$$\mathbb{E}(Y) = \int g(x) f_X(x) dx$$

2.8.2. Varianz

Definition 2.9 Die Varianz $\mathbb{V}(X)$ ist der Mittelwert der quadrierten Abweichungen der einzelnen Realisationen vom Schwerpunkt der Verteilung.

$$\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X - \mu)^2 = \sigma^2$$

Für die Berechnung benutzen wir den Erwartungswert (diskret bzw. stetig):

$$\mathbb{V}(X) = \sum_{j=1}^{\infty} (x_j - \mu)^2 \cdot \mathbb{P}(X = x_j) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx$$

ACHTUNG: Die Varianz \mathbb{V} ist ein quadratisches Streuungsmaß

Definition 2.10 Stichprobenvarianz s_n^2 mit ML-Schätzer

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

Definition 2.11 Korrigierte Stichprobenvarianz s_n^2

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

Verschiebungssatz

Wir betrachten die Varianz $\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$, formen um und wenden die Eigenschaften des Erwartungswertes an:

$$\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\right)$$

Anwendung der Additivität:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(2X\mathbb{E}(X)) + \mathbb{E}(\mathbb{E}(X)^2)$$

Anwendung der Linearität und der Eigenschaft $\mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X)$:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2$$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Satz 2.3 Steinerscher Verschiebungssatz

$$\mathbb{E}\left((X - a)^2\right) = \mathbb{V}(X) + (\mathbb{E}(X) - a)^2$$

Eigenschaften der Varianz

Definition 2.12 (Eigenschaften der Varianz) :

- 1) $\mathbb{V}(X) \geq 0$
- 2) $\mathbb{V}(X) = 0$, genau dann, wenn $\mathbb{P}(X = \mathbb{E}(X)) = 1$
- 3) $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$
- 4) Wenn X und Y unabhängig sind, dann ist $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$

Kovarianz

Definition 2.13 (Kovarianz) :

Die Kovarianz $\text{Cov}(X, Y)$ gibt Aussage darüber, ob hohe Werte von X mit hohen Werten von Y oder eher mit niedrigen Werten von Y einhergehen.

- Die Kovarianz ist positiv, wenn X und Y einen monotonen Zusammenhang besitzen, d.h. hohe (niedrige) Werte von X gehen mit hohen (niedrigen) Werten von Y einher.
- Ist die Kovarianz null, so besteht kein monotoner Zusammenhang zwischen X und Y .

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Damit können wir die Varianz berechnen:

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X, Y)$$

Wenn die Kovarianz von zwei Zufallsvariablen gleich 0 ist, dann nennen wir sie unkorreliert.

Ungleichung von Markov**Definition 2.14 (Ungleichung von Markov) :***X sei eine nichtnegative Zufallsvariable und $\lambda > 0$. Dann ist*

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda} \mathbb{E}(X)$$

Ungleichung von Chebychev**Definition 2.15 (Ungleichung von Chebychev) :***Diese Ungleichung gibt eine allgemeine Schranke an, wie weit eine Zufallsvariable X um ihren Mittelwert streut. Für jede Variable X mit existierender Varianz σ^2 und jede beliebige positive Zahl λ gilt:*

$$\mathbb{P}(|X - \mu| \geq \lambda) \leq \frac{\mathbb{V}(X)}{\lambda^2}$$

$$\mathbb{P}(|X - \mu| < \lambda) \geq 1 - \frac{\mathbb{V}(X)}{\lambda^2}$$

Beispiel 2.4 (aus Wikipedia [1]) Nehmen an, dass Wikipedia-Artikel einen Erwartungswert der Länge von $\mathbb{E}(X) = 1000$ Zeichen mit einer Standardabweichung von $\sigma_X = \sqrt{\mathbb{V}(X)} = 200$ Zeichen besitzen.

Die Wahrscheinlichkeit, dass ein Wikipedia-Artikel eine Länge zwischen 600 und 1400 ($\Rightarrow \lambda = 400$) Zeichen hat, berechne sich mit:

$$\mathbb{P}(|X - 1000| < 400) = 1 - \frac{200^2}{400^2} = 1 - \frac{1}{4} = 75$$

Ungleichung von Kolmogorov

Definition 2.16 Kolmogorov-Ungleichung Die Kolmogorov-Ungleichung ist eine sogenannte Maximal-Ungleichung aus der Stochastik. Seien X_1, \dots, X_n unabhängige Zufallsvariablen mit Erwartungswert 0, und zusätzlich $S_0 = 0$, und $S_n = X_1 + \dots + X_n$. Dann ist

$$\mathbb{P}\left(\max_{k \leq n} (S_k) \geq \lambda\right) \leq \frac{\mathbb{V}(S_n)}{\lambda^2}$$

2.9. Folgen von Zufallsvariablen

Nun untersuchen wir Folgen von Zufallsvariablen X_1, \dots, X_n , die unabhängig und identisch verteilt sind, abgekürzt i.i.d. (independent and identically distributed). Einfachstes Beispiel sind drei unabhängige Münzwürfe.

2.9.1. Schwaches Gesetz der großen Zahlen

Bei geringem Stichprobenumfang n ist die Chebychev-Ungleichung sehr ungenau, sie wird aber genauer, je größer n wird. Wir gehen von n Zufallsvariablen aus und bilden uns den Mittelwert:

$$\bar{X}^{(n)} = \frac{1}{n} \sum_{i=1}^n X_i$$

Der Erwartungswert von i.i.d Zufallsvariablen ist immer gleich, somit gilt:

$$\mathbb{E}(\bar{X}^{(n)}) = \mu$$

Da die Varianz ein quadratisches Streuungsmaß ist, folgt:

$$\mathbb{V}(\bar{X}^{(n)}) = \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^n X_i\right)$$

denn beim Ausklammern der Konstante $\frac{1}{n}$ wird diese quadriert.

$$\frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$

ACHTUNG: Die letzte Umformung kommt aus der Summenregel für unkorrelierte Zufallsvariablen: $\mathbb{V}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{V}(X_i)$

Nun benutzen wir die Ungleichung von Chebychev und passen sie an:

$$\mathbb{P}\left(\left|\bar{X}^{(n)} - \mu\right| < \lambda\right) \geq 1 - \frac{\frac{\sigma^2}{n}}{\lambda^2}$$
$$\mathbb{P}\left(\left|\bar{X}^{(n)} - \mu\right| < \lambda\right) \geq 1 - \frac{\sigma^2}{n\lambda^2}$$

Schlussendlich schicken wir n gegen unendlich und beachten, dass die Wahrscheinlichkeit nie größer als Eins werden kann:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\bar{X}^{(n)} - \mu\right| < \lambda\right) \geq 1$$

Definition 2.17 Schwaches Gesetz der großen Zahlen

Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von i.i.d.-zufälligen Variablen mit $\mathbb{E}(X_n) = \mu$. Dann gilt $\forall \epsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\bar{X}^{(n)} - \mu\right| < \epsilon\right) = 1$$

Das schwache Gesetz der großen Zahlen sagt uns also, dass mit steigender Anzahl von Zufallsvariablen große Abweichungen immer unwahrscheinlicher werden.

2.9.2. Starkes Gesetz der großen Zahlen

Das starke Gesetz der großen Zahlen bedeutet, dass

$$\lim_{n \rightarrow \infty} \bar{X}^{(n)} = \mu$$

nicht immer, aber „so gut wie immer“ gilt.

Definition 2.18 Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von i.i.d.-zufälligen Variablen mit $\mathbb{E}(X_i) = \mu$. Dann gilt:

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \bar{X}^{(n)} = \mu\right) = 1$$

Man sagt: $\bar{X}^{(n)}$ konvergiert fast sicher gegen μ .

2.9.3. Zentraler Grenzwertsatz

Dafür führen wir zuerst den Begriff der standardisierten Zufallsvariable ein.

Definition 2.19 Standardisierte Zufallsvariable

Eine Zufallsvariable X mit Erwartungswert gleich null und Varianz gleich eins heißt standardisierte Zufallsvariable. Ist X eine beliebige Zufallsvariable mit $\mathbb{E}(X) = \mu$ und $\mathbb{V}(X) = \sigma^2$, dann heißt

$$X^* = \frac{X - \mu}{\sigma}$$

σ bezeichnet dabei die Standardabweichung $\sigma = \sqrt{\mathbb{V}(X)}$

die Standardisierte von X . Der Vorgang selbst heißt Standardisierung. Für die standardisierte Variable X^* gilt:

$$\mathbb{E}(X^*) = 0$$

$$\mathbb{V}(X^*) = 1$$

Beim zentralen Grenzwertsatz berechnen wir zuerst die Summe von n i.i.d. Zufallsvariablen mit $\mathbb{E}(X) = \mu$ und $\mathbb{V}(X) = \sigma^2$:

$$S_n = X_1 + \dots + X_n$$

$$\mathbb{E}(S_n) = n\mu$$

$$\mathbb{V}(S_n) = n\sigma^2$$

Wir standardisieren S_n :

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

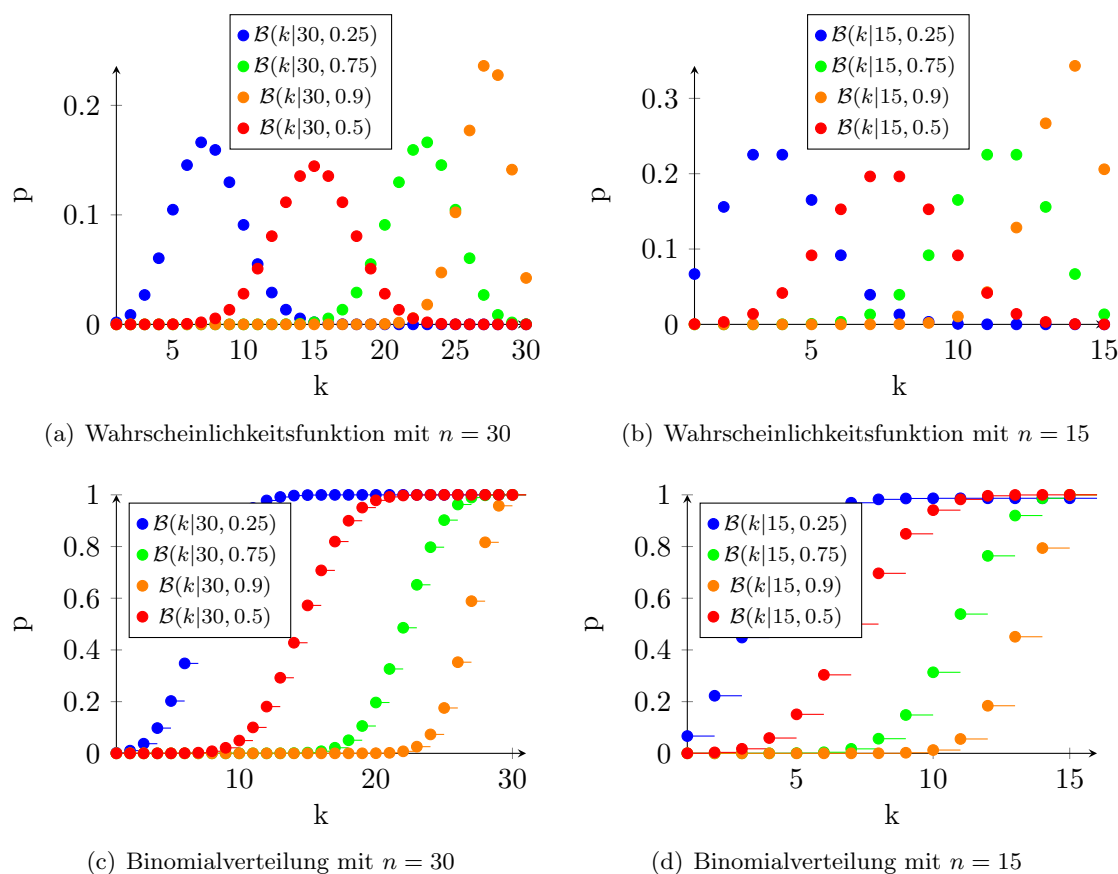


Abbildung 2.8.: Die Binomialverteilung

Satz 2.4 Zentraler Grenzwertsatz:

Der zentrale Grenzwertsatz besagt, dass die Verteilungsfunktion S_n^* für $n \rightarrow \infty$ punktweise gegen die Verteilungsfunktion der Standardnormalverteilung $\mathcal{N}(0, 1)$ konvergiert. Ist $\Phi(z)$ die Verteilungsfunktion von $\mathcal{N}(0, 1)$, dann bedeutet dies, dass für jedes reelle z gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}(S_n^* \leq z) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq z\right) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

2.10. Diskrete Verteilungen**2.10.1. Binomialverteilung****Beispiel 2.5 Münzwurf:**

Bezeichnen wir die Wahrscheinlichkeit von **Kopf** mit p , so ergibt sich die Wahrscheinlichkeit von **Zahl** mit $1 - p$.

Ist X die Anzahl der Kopfwürfe bei n Würfeln, so tritt das Ereignis $X = k$ genau dann auf, wenn in einer Wurfsequenz (...K...Z...Z...) genau k -mal K auftritt und daher auch $(n - k)$ mal Z. Die Wahrscheinlichkeit einer solchen Sequenz ist

$$p^k (1 - p)^{n-k}$$

Dabei kann die Reihenfolge egal sein. Wir müssen also noch wissen, wie viele Möglichkeiten es gibt, aus n Würfeln, k auszuwählen:

$$\binom{n}{k}$$

Sie beschreibt die Anzahl der Erfolge in einer Serie von gleichartigen und unabhängigen Versuchen, die jeweils genau zwei mögliche Ergebnisse haben („Erfolg“ oder „Misserfolg“). (Wikipedia [2])

Damit haben wir die Binomialverteilung (siehe Abbildung 2.8) definiert:

Definition 2.20 Binomialverteilung

$$\mathcal{B}(k|n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, \dots, n$$

$$\mathbb{E}(X) = np$$

$$\mathbb{V}(X) = np(1 - p)$$

Für die Binomialverteilung existiert ein **Additionstheorem**:

$$\text{Aus } \begin{cases} X \sim \mathcal{B}_n(\theta) \\ Y \sim \mathcal{B}_m(\theta) \end{cases} \quad \text{folgt } X + Y \sim \mathcal{B}_{n+m}(\theta)$$

Existieren nicht nur zwei Möglichkeiten, kann die Binomialverteilung mit der Polynomverteilung vertauscht werden. Es sind r, w, \dots, g die Anzahl der einzelnen Elemente (z.B. rote, weiße und grüne) und $\theta_r, \theta_w, \dots, \theta_g$ die entsprechenden Wahrscheinlichkeiten:

$$\mathbb{P}(r, w, \dots, g) = \frac{n!}{r!w! \dots g!} (\theta_r)^r (\theta_w)^w \dots (\theta_g)^g$$

2.10.2. Gleichverteilung

Beispiel 2.6 Die Gleichverteilung beschreibt den Wurf eines idealen n -Seiten Würfels. (siehe Abbildung 2.9)

Definition 2.21

$$\mathbb{P}(X = k) = \frac{1}{n}$$

$$\mathbb{E}(X) = \frac{n+1}{2}$$

$$\mathbb{V}(X) = \frac{n^2 - 1}{12}$$

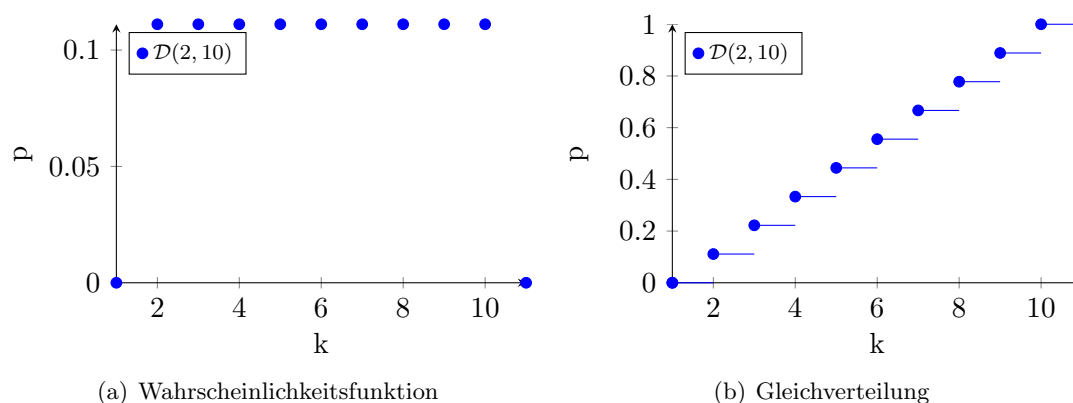


Abbildung 2.9.: Die Gleichverteilung

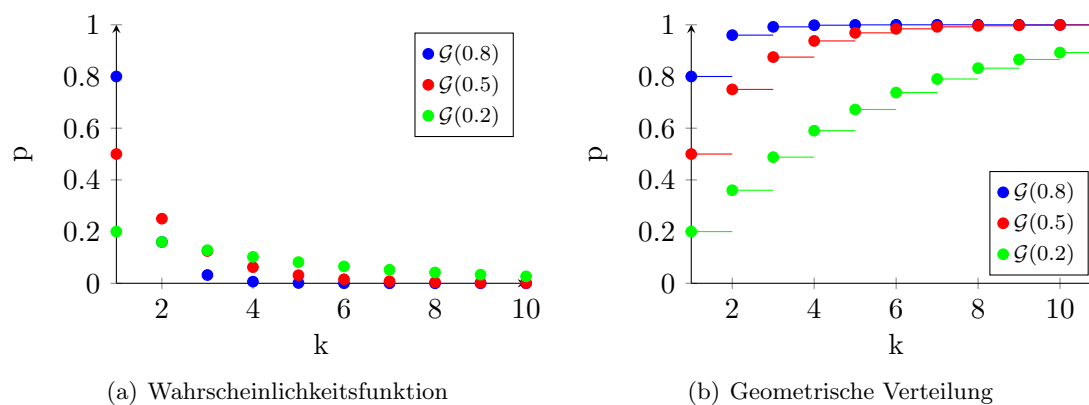


Abbildung 2.10.: Die Geometrische Verteilung

Befindet sich k in einem beliebigen Intervall $a \leq k \leq b$, so gilt analog:

$$\mathbb{P}(X = k) = \frac{1}{b - a + 1}$$

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

$$\mathbb{V}(X) = \frac{(b - a)^2 - 1}{12}$$

2.10.3. Geometrische Verteilung

Siehe Abbildung 2.10

Beispiel 2.7 Die geometrische Verteilung beschreibt die Anzahl der Versuche bis zum ersten Erfolg (z.B. wie oft muss ich beim „Mensch ärgere dich nicht“ würfeln, bis ich einen 6er würfle). Wir fragen uns also, wie hoch die Wahrscheinlichkeit ist, erst nach k -Würfen ein Ereignis mit Wahrscheinlichkeit θ eintritt.

Nachdem ich aber weiß, dass vor k -Würfen dieses Ereignis nicht eintritt, kann ich definieren:

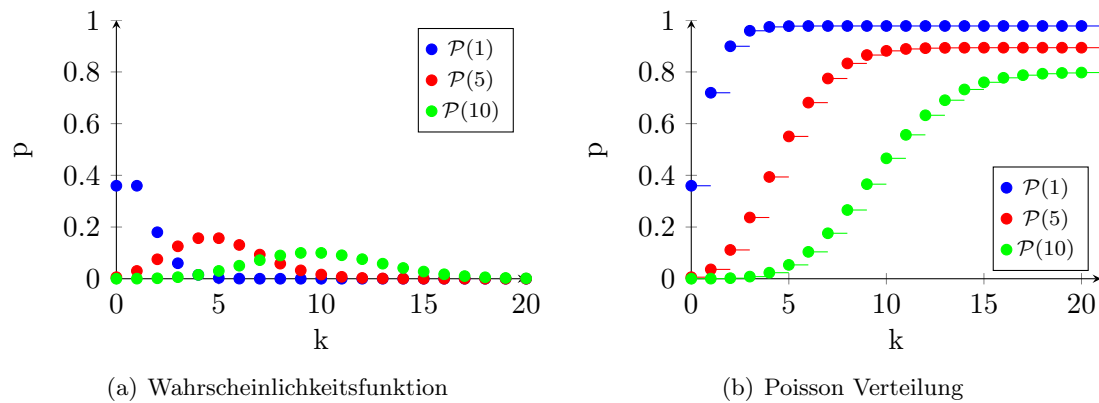


Abbildung 2.11.: Die Poisson Verteilung

Definition 2.22 Eine Zufallsvariable mit der Verteilung

$$\mathbb{P}(X = k) = \theta(1 - \theta)^{k-1}$$

heißt geometrisch verteilt mit dem Parameter θ .

Erwartungswert und Varianz ergeben sich aus der Potenzreihenentwicklung. Zusammengefasst:

$$\mathbb{E}(X) = \frac{1}{\theta}$$

$$\mathbb{V}(X) = \frac{1 - \theta}{\theta^2}$$

Eine wichtige Eigenschaft der geometrischen Verteilung ist, dass sie kein Gedächtnis hat. Folgende Situation: Ich habe bereits 20 Mal „Kopf“ geworfen. Die Wahrscheinlichkeit, dass ich nun beim 25. Wurf „Zahl“ werfe ist gleich hoch, wie das ich beim 5. Wurf „Zahl“ werfe. Anders ausgedrückt:

Satz 2.5 Die Wahrscheinlichkeit, dass Sie erst beim $(k + j)$ -ten Wurf Erfolg haben unter der Bedingung, dass die ersten j Versuche Misserfolge waren, ist gerade die Wahrscheinlichkeit, dass Sie erst beim k -ten Versuch Erfolg haben,

$$\mathbb{P}(X = k + j | X > j) = \mathbb{P}(X = k)$$

2.10.4. Poissonverteilung

Bei der Poissonverteilung (siehe Abbildung 2.11) wissen wir relativ wenig über die Auftrittswahrscheinlichkeit eines Ereignisses. Aus früheren Messungen kennen wir aber $\mathbb{E}(X) = \lambda$, also der Erwartungswert pro Einheit eines Kontinuums.

Die Idee dahinter ist, dass wir das Kontinuum in n Miniintervalle unterteilen. Die Wahrscheinlichkeit, dass ein Ereignis in einem Miniintervall auftritt ist θ_n . Betrachten wir nun n -Intervalle und sei $X = \sum_{i=1}^n X_i$ die Gesamtanzahl aller eingetretenen Ereignisse, so entspricht dieses Modell der Binomialverteilung:

$$\mathbb{P}(X = k) = \binom{n}{k} (\theta_n)^k (1 - \theta_n)^{n-k}$$

Nun verringern wir die Größe dieser Intervalle, wir kommen zu dem Grenzübergang $n \rightarrow \infty$ und $\theta_n \rightarrow 0$. Dazu wird die Binomialverteilung in vier Teile unterteilt (ohne Beweis):

$$\underbrace{\binom{n}{k}}_{\frac{1}{k!}} \cdot \underbrace{\frac{1}{n^k}}_{\lambda^k} \cdot \underbrace{(n\theta_n)^k}_{e^{-\lambda}} \cdot \underbrace{(1 - \theta_n)^n}_{1} \cdot \underbrace{(1 - \theta_n)^{-k}}_1$$

Definition 2.23 Wahrscheinlichkeitsfunktion

Eine Zufallsvariable mit der Verteilung

$$\mathbb{P}(X = k) = \frac{1}{k!} \lambda^k e^{-\lambda}$$

heißt Poisson-verteilt mit dem Parameter λ , geschrieben $\mathcal{P}(\lambda)$. Ist $X \sim \mathcal{P}(\lambda)$, dann ist

$$\mathbb{E}(X) = \mathbb{V}(X) = \lambda$$

Beispiel 2.8 Wir betrachten die Telefonzentrale bei einer großen Feuerwehration. Sei X die Anzahl der Alarmmeldungen pro Stunde und $\lambda = 10$ die durchschnittliche Zahl von Alarmmeldungen pro Stunde, dann kann $X \sim \mathcal{P}(10)$ modelliert werden. Die punktförmigen Ereignisse sind die Augenblicke, in denen die Anrufe eintreffen. Nun können wir die Verteilungsfunktion für $\mathcal{P}(10)$ berechnen:

k	$\mathbb{P}(X \leq k)$	k	$\mathbb{P}(X \leq k)$
0	$4.5400 \cdot 10^{-5}$	11	0.69678
1	$4.9940 \cdot 10^{-4}$	12	0.79156
2	$2.7694 \cdot 10^{-3}$	13	0.86446
3	$1.0336 \cdot 10^{-2}$	14	0.91654
4	$2.9253 \cdot 10^{-2}$	15	0.95126
5	$6.7086 \cdot 10^{-2}$	16	0.97296
6	0.13014	17	0.98572
7	0.22022	18	0.99281
8	0.33282	19	0.99655
9	0.45793	20	0.99841
10	0.58304		

Zum Beispiel ist die Wahrscheinlichkeit, dass höchstens 10 Anrufe pro Stunde eingehen 58,3 und die Wahrscheinlichkeit, dass genau 12 Anrufe stattfinden 79,2 – 69,7 = 9,5.

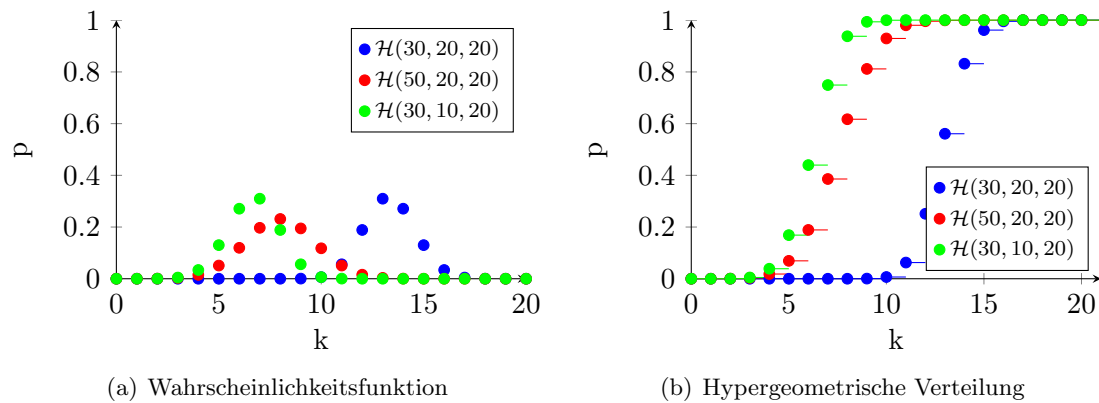


Abbildung 2.12.: Die Hypergeometrische Verteilung

Satz 2.6 Die Poisson-Verteilung eignet sich zur Approximation der Binomialverteilung $\mathcal{B}_n(\theta)$, wenn n groß und θ klein ist. Eine Faustregel fordert dazu:

$$n \geq 50, \theta \leq 0.1 \text{ und } n\theta \leq 10$$

2.10.5. Hypergeometrische Verteilung

Die hypergeometrische Verteilung (siehe Abbildung 2.12) wenden wir in Situationen an, in denen wir Stichproben entnehmen. Diese Verteilung basiert auf „Ziehen ohne Zurücklegen“.

Beispiel 2.9 Urne mit 45 Kugeln, 20 davon sind gelb, es werden 10 gezogen. Wie groß ist die Wahrscheinlichkeit, dass genau 4 Gelbe dabei sind?

Nun definieren wir die Situation:

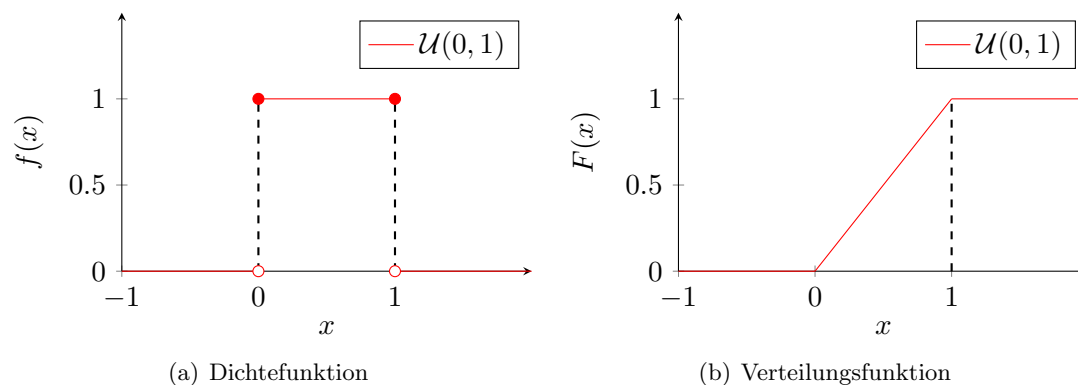
- $N \dots$ Gesamtanzahl der Elemente
- $n \dots$ Anzahl der Elemente einer Stichprobe
- $M \dots$ Elemente mit einer bestimmten Eigenschaft
- $k \dots$ Wahrscheinlichkeit, k Elemente zu ziehen

Definition 2.24 Eine Zufallsvariable X ist hypergeometrisch verteilt: $X \sim \mathcal{H}(N, M, n)$

$$h(k, N; M; n) := \mathbb{P}(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

Die Wahrscheinlichkeitsfunktion beinhaltet folgende Teile:

- $\binom{M}{k} \dots$ Anzahl der Möglichkeiten, aus M Elementen k auszuwählen
- $\binom{N-M}{n-k} \dots$ Anzahl, aus den verbleibenden $N-M$ Möglichkeiten, $n-k$ auszuwählen

Abbildung 2.13.: Die stetige Gleichverteilung im Intervall $[0, 1]$

$\binom{N}{n}$... Gesamtanzahl an Möglichkeiten, aus N Objekten n auszuwählen

Jede Zufallsvariable X_i besitzt die gleiche Verteilung und somit den gleichen Erwartungswert $\mathbb{E}(X_i) = \theta$.

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X_i) = n\theta$$

Für die Varianz gilt das nicht so einfach, da die Zufallsvariablen nicht unabhängig sind. Ohne Beweis gilt:

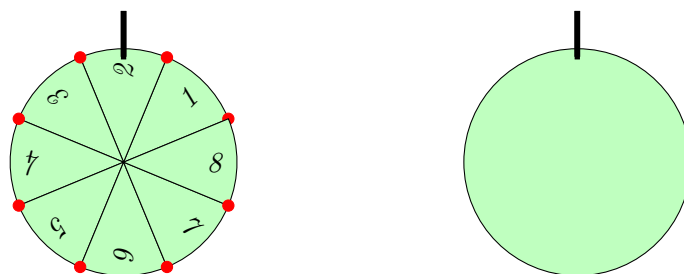
$$\mathbb{V}(X) = n\theta(1-\theta) \frac{N-n}{N-1}$$

Der Faktor $\frac{N-n}{N-1}$ wird auch Korrekturfaktor für die endliche Gesamtheit bezeichnet.

2.11. Stetige Verteilungen

2.11.1. Gleichverteilung

Beispiel 2.10 Für die stetige Gleichverteilung betrachten wir am besten folgende zwei Glücksräder:



Während beim linken Glücksrad die Wahrscheinlichkeit für jedes „Pizzastück“ $\frac{1}{n} = \frac{1}{8}$ beträgt, ist diese Unterscheidung auf der rechten Seite nicht mehr möglich.

Oft wird die stetige Gleichverteilung im Intervall $[0, 1]$ angegeben – in unserem Beispiel müssten wir nur den Umfang als 1 Einheit betrachten (siehe Abbildung 2.13). Dann gilt:

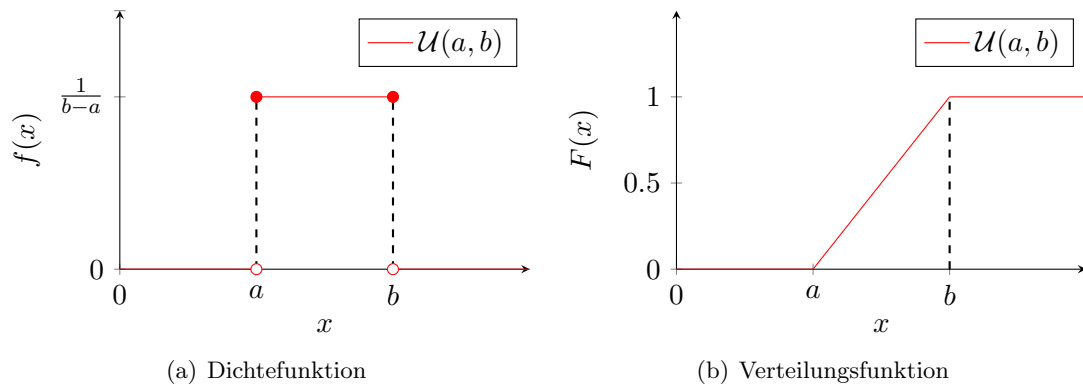


Abbildung 2.14.: Die stetige Gleichverteilung

Definition 2.25 Eine Zufallsvariable U mit der Dichte

$$f(u) = \begin{cases} 1, & \text{für } 0 \leq u \leq 1, \\ 0, & \text{sonst,} \end{cases}$$

besitzt eine stetige Gleichverteilung auf dem Intervall $[0, 1]$. Erwartungswert und Varianz sind

$$\mathbb{E}(U) = \frac{1}{2}$$

$$\mathbb{V}(U) = \frac{1}{12}$$

Wenn die stetige Gleichverteilung nicht auf $[0, 1]$ sondern auf einem beliebigen Intervall $[a, b]$ definiert ist (siehe Abbildung 2.14)

Definition 2.26 Ist U im Intervall $[0, 1]$ gleichverteilt und $Y = a + bU$, dann ist Y im Intervall $[a, b]$ gleichverteilt mit der Dichte

$$f_Y(y) = \begin{cases} \frac{1}{b}, & \text{für } a \leq y \leq a + b, \\ 0, & \text{sonst,} \end{cases}$$

(siehe Abbildung 2.14). Außerdem gilt:

$$\mathbb{E}(Y) = a + \frac{b}{2}$$

$$\mathbb{V}(Y) = \frac{b^2}{12}$$

2.11.2. Exponentialverteilung

Für die Einführung der Exponentialverteilung starten wir beim Beispiel von der Poissonverteilung. Wir sitzen wieder in der Feuerwache, dieses Mal interessiert uns jedoch nicht die Anzahl der Anrufe, sondern die Wartezeit bis zum nächsten Anruf. Es sei X die Anzahl der Anrufe in einer Stunde und X_t die Anzahl der Anrufe in t -Stunden. Nach unserer Voraussetzung ist $\mathbb{E}(X_t) = \lambda t$, also $X_t \sim \mathcal{PV}(\lambda t)$. Die Wahrscheinlichkeit, dass in t -Stunden k -Anrufe kommen, ist

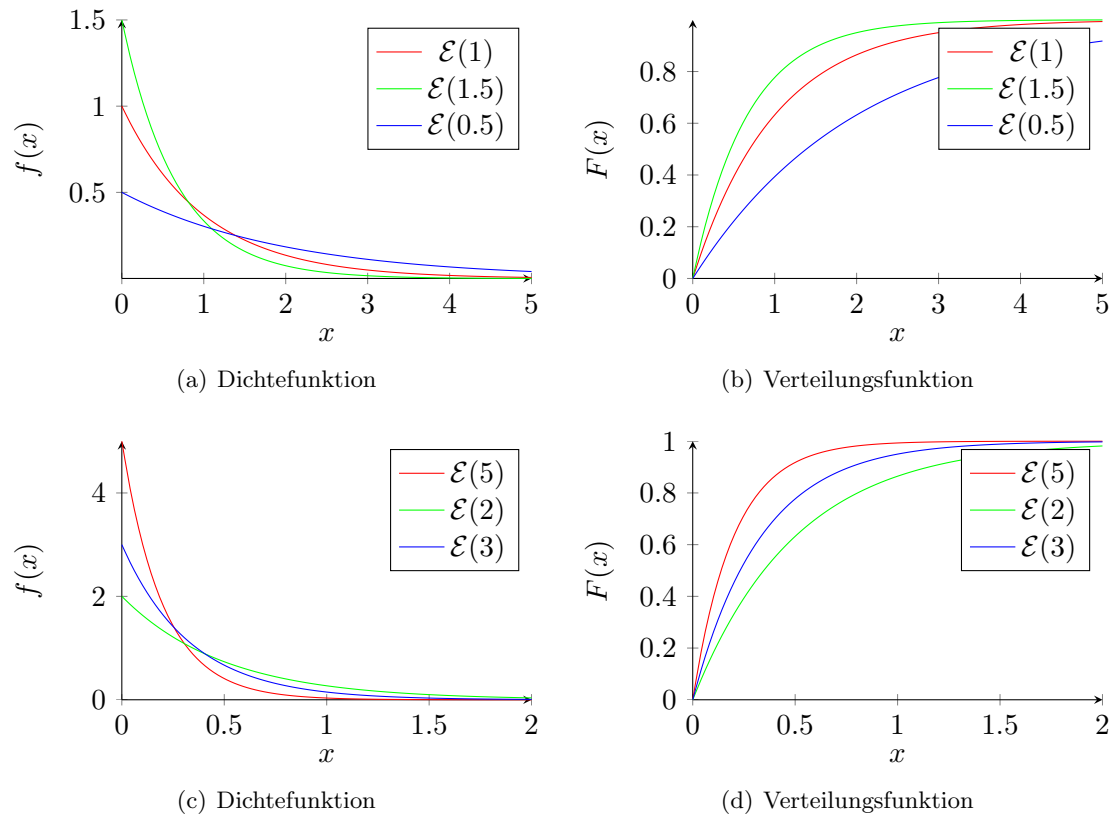


Abbildung 2.15.: Die Exponentialverteilung

$$\mathbb{P}(X_t = k) = \frac{1}{k!}(\lambda t)^k e^{-\lambda t}$$

und dass nach t -Stunden kein Anruf kommt

$$\mathbb{P}(X_t = 0) = e^{-\lambda t}$$

Diese Funktion mit $\lambda = 5$ zeigt die Wahrscheinlichkeit, dass nach der Zeit t kein Anruf eingegangen ist. Analog dazu die Funktion $1 - e^{-\lambda t}$, dass nach der Zeit t mindestens ein Anruf eingegangen ist.

Aus Abbildung 2.15 kann die Wahrscheinlichkeit direkt herausgelesen werden (man muss nur den gelesenen Wert durch 5 dividieren). Zum Beispiel ist die Wahrscheinlichkeit, dass zwischen den ersten 10 Minuten und den ersten 20 Minuten mindestens ein Anruf kommt:

$$(1 - e^{-\lambda t_2}) - (1 - e^{-\lambda t_1}) = e^{-\lambda t_1} - e^{-\lambda t_2} = e^{-5 \cdot 0.1} - e^{-5 \cdot 0.2} = 0,607 - 0,368 = 0,239 \approx 24$$

Allgemein definieren wir für die Exponentialfunktion:

Definition 2.27 Die Verteilung mit der Dichte ($\lambda > 0$)

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

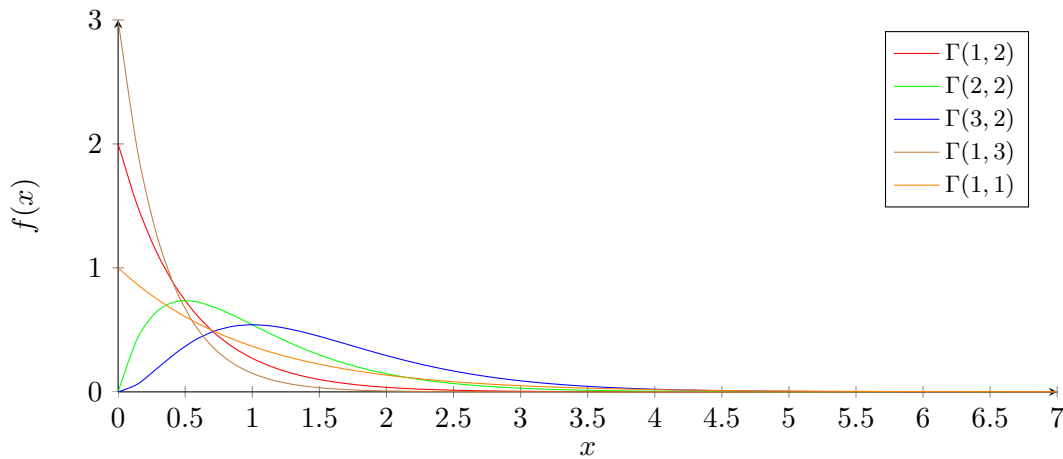


Abbildung 2.16.: Dichtefunktion der Gammaverteilung

heißt *Exponentialverteilung* und wird mit $\text{ExpV}(\lambda)$ notiert. Dabei ist

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

$$\mathbb{V}(X) = \frac{1}{\lambda^2}$$

$$\text{Med}(X) = \frac{\ln 2}{\lambda}$$

2.11.3. Gammaverteilung

Mit der Gammaverteilung (siehe Abbildung 2.16) ist es möglich, die Fakultät jeder positiven reell wertigen (im weiteren Sinne auch komplexen) Zahl zu berechnen. Dabei nimmt man sich das Euler'sche Integral zweiter Art zur Hilfe.

Definition 2.28 (Gammaverteilung) :

Die Gammaverteilung $\Gamma(\alpha, \lambda)$ ist definiert mit:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x > 0 \\ 0 & \text{sonst} \end{cases}$$

Die **Gammafunktion** Γ ist für beliebige $x \in \mathbb{R}_{>0}$ definiert als

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

und erfüllt die Funktionalgleichung

$$\Gamma(x+1) = x\Gamma(x) \Rightarrow \Gamma(x+1) = (x)! \Rightarrow \Gamma(x) = (x-1)!$$

Für die Gammafunktion gilt der Ergänzungssatz

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$$

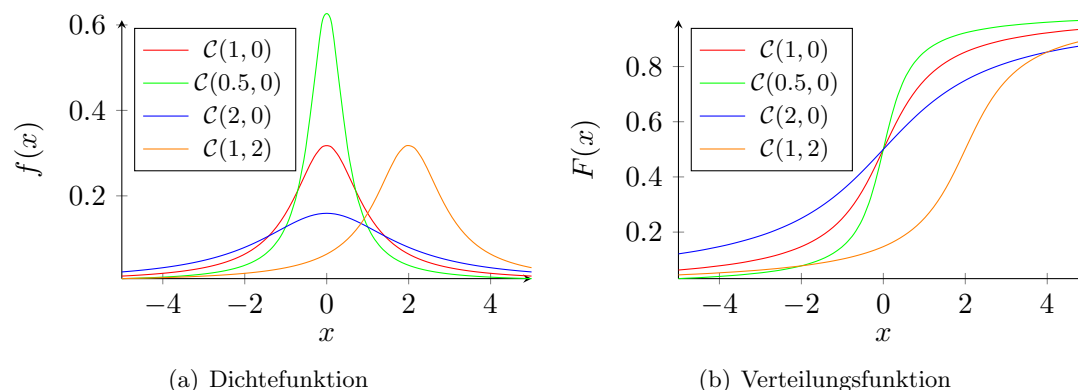


Abbildung 2.17.: Die Cauchyverteilung

2.11.4. Cauchy-Verteilung

Die Cauchy-Verteilung (siehe Abbildung 2.17) wird oft verwendet, wenn man die Auswirkung von Ausreißern in Daten auf statistische Verfahren untersuchen möchte. Allgemein gilt:

Definition 2.29 (Cauchy-Verteilung) :

Die Cauchy-Verteilung ist eine stetige Wahrscheinlichkeitsverteilung, die die Wahrscheinlichkeitsdichte

$$f(x) = \frac{1}{\pi} \frac{s}{s^2 + (x - t)^2}$$

mit $s > 0$ und $-\infty < t < \infty$ besitzt.

Die Verteilungsfunktion der Cauchy-Verteilung ist

$$F(X) = \mathbb{P}(X < x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x - t}{s}\right)$$

ACHTUNG: Die Cauchy-Verteilung besitzt weder Erwartungswert noch Varianz.

Zusätzlich existiert noch eine Standard-Cauchy-Verteilung (= t-Verteilung) mit Zentrum $t = 0$ und dem Breitenparameter $s = 1$.

$$f(x) = \frac{1}{\pi(1 + x^2)}$$

Es folgt ein Beweis dafür, dass kein Erwartungswert existiert. Man sagt, dass bei stetigen Verteilungen der Erwartungswert $\mathbb{E}(X)$ existiert, wenn $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$ gilt.

$$\begin{aligned} \int_{-\infty}^{\infty} |x| \frac{1}{\pi(1 + x^2)} dx &= 2 \int_0^{\infty} x \frac{1}{\pi(1 + x^2)} dx = \frac{1}{\pi} \int_0^{\infty} \frac{2x}{1 + x^2} dx = \frac{1}{\pi} \ln(1 + x^2) \Big|_0^{\infty} \\ &= \frac{1}{\pi} (\infty - 0) = \infty \end{aligned}$$

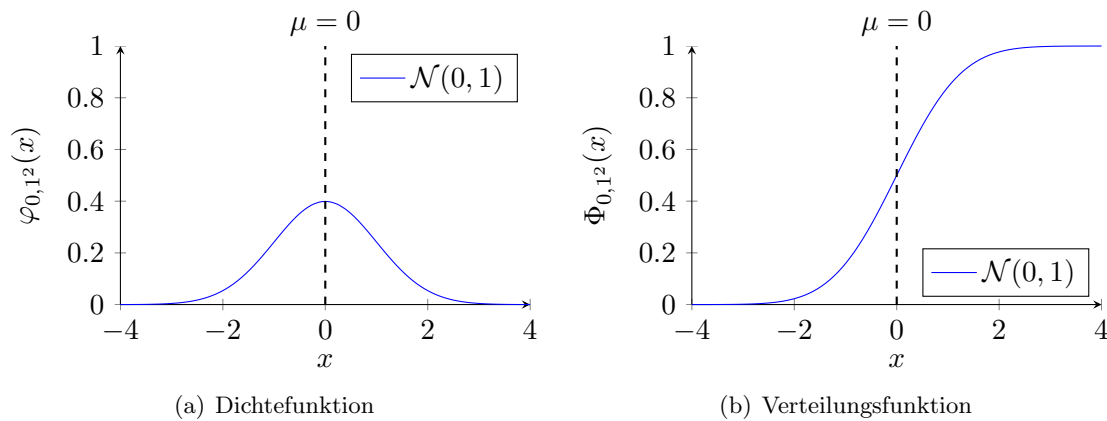


Abbildung 2.18.: Die Standardnormalverteilung

2.11.5. Normalverteilung

Die Gauß-Verteilung aus der Normalverteilungsfamilie gehört zu den theoretisch und praktisch wichtigsten Wahrscheinlichkeitsverteilungen.

Definition 2.30 Eine Zufallsvariable X mit der Dichte

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

heißt **standardnormalverteilt**. Wir schreiben

$$X \sim \mathcal{N}(0; 1)$$

Für Erwartungswert und Varianz gilt

$$\mathbb{E}(X) = 0$$

$$\mathbb{V}(X) = 1$$

Oft wird die Dichte (siehe Abbildung 2.18(a)) $f(x)$ von $\mathcal{N}(0; 1)$ auch mit φ , die Verteilungsfunktion (siehe Abbildung 2.18(b)) mit Φ bezeichnet. Die Funktion $\Phi(x)$ ist tabelliert und findet sich unter Anhang A.1.

Alle Verteilungen, die aus der Standardnormalverteilung durch lineare Transformationen hervorgehen, bilden die Familie der Normalverteilungen (siehe Abbildung 2.19)

Definition 2.31 Ist $X \sim \mathcal{N}(0; 1)$ und wird X **linear transformiert** zu $Y = \mu + \sigma X$ dann hat Y den Erwartungswert μ , die Varianz σ^2 und die Dichte

$$f_Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

Y heißt **normalverteilt**, geschrieben

$$Y \sim \mathcal{N}(\mu; \sigma^2)$$

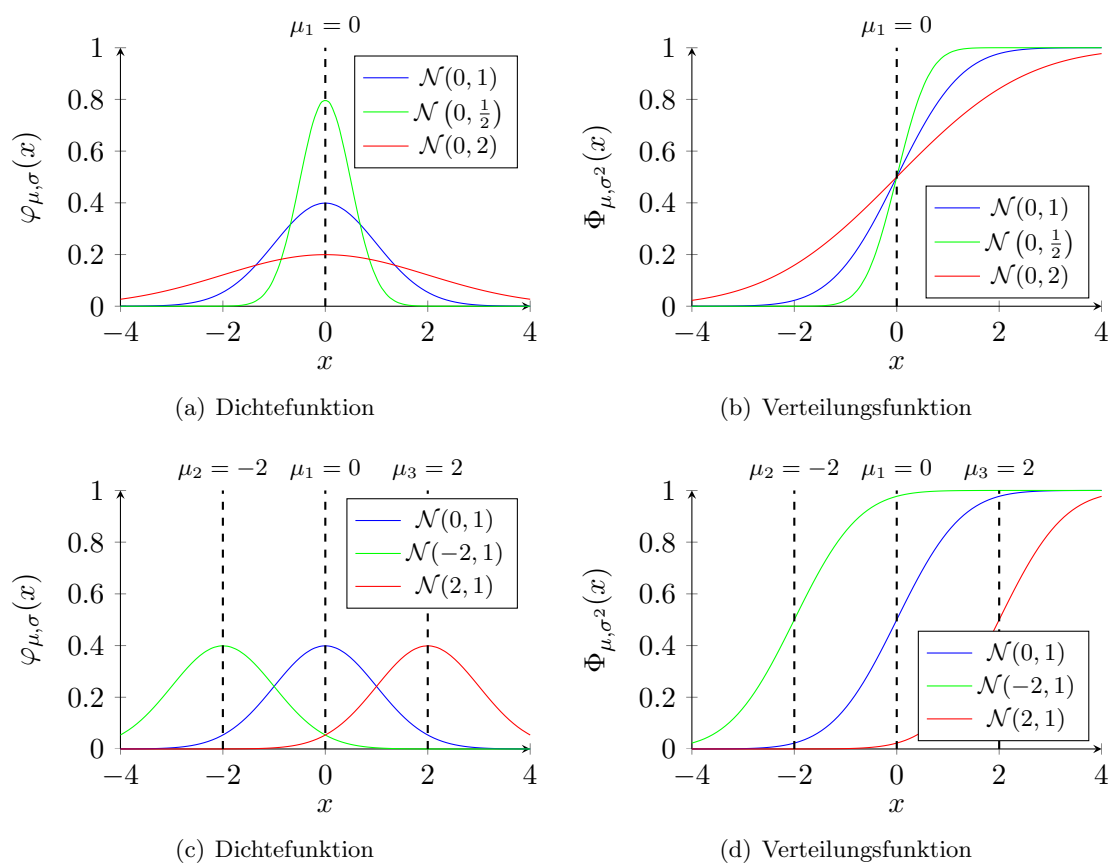


Abbildung 2.19.: Die Normalverteilung

2.12. Verteilungen Zusammenfassung

Diskrete Verteilungen

Name	Symbol	$p(x)$	$\mathbb{E}(X)$	$\mathbb{V}(X)$
Binomial	$\mathcal{B}(n, p)$	$\binom{n}{x} p^x (1-p)^{n-x} (0 \leq x \leq n)$	np	$np(1-p)$
Gleichverteilung	$\mathcal{D}(a, b)$	$\frac{1}{b-a+1} (a \leq x \leq b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2-1}{12}$
Geometrisch	$\mathcal{G}(p)$	$p(1-p)^x, (x \geq 0)$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson	$\mathcal{P}(\lambda)$	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ
Hypergeometrisch	$\mathcal{H}(N, A, n)$	$\frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}} (0 \leq x \leq n)$	$\frac{nA}{N}$	$\frac{nA(N-A)(N-n)}{N(N-1)}$

Stetige Verteilungen

Name	Symbol	$f(x)$	$\mathbb{E}(X)$	$\mathbb{V}(X)$
Gleichverteilung	$\mathcal{U}(a, b)$	$\frac{1}{b-a} (a \leq x \leq b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$\mathcal{E}(\lambda)$	$\lambda e^{-\lambda x} (x \geq 0)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$\Gamma(\alpha, \lambda)$	$\frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} (x \geq 0)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Cauchy	$\mathcal{C}(a)$	$\frac{a}{\pi(x^2+a^2)}$	N.A.	N.A.
Normal	$\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2
Beta 1. Art	$\beta_1(\alpha, \lambda)$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} (0 \leq x \leq 1)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$
Beta 2. Art	$\beta_1(\alpha, \lambda)$	$\frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha, \beta)} (0 \leq x)$	$\frac{\alpha}{\beta-1} (\beta > 1)$	$\frac{\alpha(\alpha+\beta-1)}{(\beta-2)(\beta-1)^2} (\beta > 2)$

3. Markovketten

Vorbemerkungen

Um es gleich vorweg zu nehmen: wir betrachten hier Markovketten. Das sind Markovprozesse (ein Spezialfall von stochastischen Prozessen) mit diskretem Zustandsraum. D.h.: Die Zustände in unserer Markovkette können wir als Folgen von Zufallsvariablen betrachten oder auch anschaulicher als Knoten eines Graphen die wir über einen Namen in Form einer ganzen Zahl identifizieren können. Beispielsweise könnte unsere Markovkette die Zustände $\{-3\}, \{-2\}, \{-1\}, \{0\}, \{1\}, \{2\}$ enthalten. Außerdem beschäftigen wir uns hier vorerst mit Markovketten nur in diskreter Zeit, was zu einer weiteren Vereinfachung führt. Wir prüfen also nur zu diskreten Zeitpunkten ($0 \leq t \leq \infty$) bzw. ($0 \leq n \leq \infty$) in welchem Zustand sich die Markovkette befindet.

Beispiel 3.1 *Ein Beispiel könnte eine Spielfigur sein, welche jede Runde am Spielbrett (beispielsweise dem von DKT) entlang zieht. Dies geschieht abhängig vom Würfel-ergebnis. Wir nehmen an wir würfeln mit nur einem (fairen) Würfel. Wenn wir jede Runde prüfen auf welchem Feld die Spielfigur zu stehen kommt, entsprechen die Runden unseren diskreten Zeitpunkten und die Spielfelder (welche wir über eine Nummer identifizieren könnten) entsprechen unseren diskreten Zuständen.*

Wir machen Momentaufnahmen, in welchem Zustand sich die Markovkette gerade befindet und bezeichnen diese mit X_n . Dabei handelt es sich um Zufallsvariablen, denn Zustände werden von anderen Zuständen aus, nur mit gewissen Wahrscheinlichkeiten erreicht. $X_3 = 12$ Könnte bedeuten, dass sich die Markovkette zum Zeitpunkt 3 im Zustand 12 befindet. So könnte etwa die Spielfigur auf Feld 12 stehen.

Wie wir erkennen können, sind für Spielfigur aus nicht alle Spielfelder erreichbar sondern nur die (von der aktuellen Position aus gesehen) nächsten 6 Felder. Die Wahrscheinlichkeit für den Übergang zu diesem Feld bezeichnen wir als **Übergangswahrscheinlichkeiten** $p_{ij} = P(X_{n+1} = j | X_n = i)$

Also die Wahrscheinlichkeit, im nächsten Zug auf Feld/im Zustand j zu landen wenn wir uns jetzt auf Feld/im Zustand i befinden. In unserem Beispiel könnten dies lauten:

$$p_{12,13} = \frac{1}{6} \quad p_{12,14} = \frac{1}{6} \quad p_{12,15} = \frac{1}{6} \quad p_{12,16} = \frac{1}{6} \quad p_{12,17} = \frac{1}{6} \quad p_{12,18} = \frac{1}{6}$$

Alle anderen Zustände wären nicht erreichbar, d.h.: die entsprechenden Übergangswahrscheinlichkeiten wären 0. Wir können uns vorstellen, dass auch die **Markoveigenschaft** zu trifft. Diese besagt, dass die Wahrscheinlichkeit, einen bestimmten Zustand zu erreichen, nur vom aktuellen Zustand abhängt und nicht von den Zuständen die zuvor angenommen wurden. Auch für den nächsten Zug der Spielfigur wird es unerheblich sein, welche Zustände/Felder sie zuvor passiert hat – erreichbar werden immer nur die von der aktuellen Position aus gesehenen 6 Felder sein.

Wir können auch t-stufige Übergangswahrscheinlichkeiten definieren: $p_{ij}(t) = P(X_{n+t} = j | X_n = i)$

Diese sind Wahrscheinlichkeiten, einen Zustand in t -Schritten bzw. Zügen anzunehmen.

Wir fassen unsere Übergangswahrscheinlichkeiten zu Matrizen P bzw. $P(t)$ zusammen. $P(t)$ ist die **t -stufige Übergangsmatrix** und kann aus den Potenzen $P(t) = P^t$ berechnet werden.

Ein weiterer wichtiger Begriff ist die **Wahrscheinlichkeitsverteilung**. Als einfaches Beispiel nehmen wir hier die **Anfangsverteilung** her. Diese gibt für jeden Zustand i die Wahrscheinlichkeit an, dass sich die Kette zu Beginn - also zum Zeitpunkt $n=0$ - in Zustand i befindet.

Diese (wie auch die anderen Verteilungen) wird in Form eines Wahrscheinlichkeitsvektors angegeben, welcher als Zeilenvektor aufgefasst wird. Nennen wir unsere Anfangsverteilung $\underline{\alpha}$ hier z.B.: $\underline{\alpha} = (\alpha_0 \quad \alpha_1 \quad \dots \quad \alpha_k)$

Dann ist diese in unserem Spiel-Beispiel:

$$\underline{\alpha} = (1 \quad 0 \quad \dots \quad 0)$$

Wenn wir davon ausgehen, dass die Figur definitiv am Start-Feld startet und dieses als Zustand 0 bezeichnet wird.

Im folgenden wird auch von der **stationären Verteilung** $\underline{\pi}$ die Rede sein, welche inkonsistenter Weise auch oft ohne den Vektor-Unterstrich angegeben wird. Wichtig ist zu wissen, dass diese ebenfalls durch einen Wahrscheinlichkeitsvektor in Form eines Zeilenvektors beschrieben wird und π_i die einzelnen Einträge bezeichnet.

Eine Bemerkung ist noch zu machen was Formeln mit Summen betrifft. In den meisten Fällen sind darin auch Übergangswahrscheinlichkeiten enthalten und wie wir sehen werden sind dieses sehr oft 0. D.h.: Es fallen ohnehin die meisten Glieder der Reihe weg und es bleibt meistens eine mehr oder weniger einfache Differenzengleichung für die gesuchte Unbekannte übrig. Mehr dazu in den Beispielen, welche am Ende des Kapitels folgen.

Mit diesem Wissen, sollte es möglich sein, das nachfolgende Skriptum sowie die Beispiele einigermaßen zu verstehen, ohne dass einen die Symbole all zu sehr irritieren.

3.1. Stochastische Prozesse

Definition 3.1 *Ein stochastischer Prozess ist eine Familie $(X_t, t \in T)$ von Zufallsvariablen. Die Indexmenge T wird Parameterraum genannt und soll eine Teilmenge der reellen Zahlen sein. Der Wertebereich M_X von X_t heißt Zustandsraum oder Phasenraum. Wenn T endlich oder abzählbar (etwa \mathbb{N}) ist, sprechen wir von einem Prozess in diskreter Zeit, wenn T ein ganzes (endliches oder unendliches) Intervall ist, von einem Prozess in stetiger Zeit,*

Stochastische Prozesse in diskreter Zeit sind einfach Folgen von Zufallsvariablen. Der Unterschied zu unseren früheren Überlegungen besteht darin, dass wir nicht mehr annehmen, dass die einzelnen Zufallsvariablen unabhängig sind. Wir müssen also die Abhängigkeiten zwischen den einzelnen Zufallsvariablen festlegen, das heißt, wir müssen gewisse Annahmen über die gemeinsame Verteilung von $(X_{t_1}, \dots, X_{t_n})$ mit $t_1 < \dots < t_n$ treffen (Ein berühmter Satz von Kolmogorov besagt, dass durch die Angabe dieser "endlichdimensionalen Randverteilungen" ein stochastischer Prozess festgelegt wird). Einige Möglichkeiten zählen wir jetzt auf:

Definition 3.2 Der Prozess $(X_t, t \in T)$ heißt Prozess mit unabhängigen Zuwächsen, wenn für $t_1 < t_2 < \dots < t_n$ die Zufallsvariablen

$$X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_n) - X(t_{n-1})$$

unabhängig sind.

Hier verwenden wir die Notation $X(t)$ für X_t , damit die Indizes nicht zu sehr überladen werden. Prozesse mit unabhängigen Zuwächsen in diskreter Zeit sind einfach Summen von unabhängigen Zufallsvariablen, wie wir sie im letzten Kapitel untersucht haben.

Definition 3.3 Der Prozess $X(t)$ heißt Markovprozess, wenn für $t_1 < t_2 < \dots < t_n$

$$\mathbb{P}(X(t_n) \leq x_n | X(t_1) = x_1, \dots, X(t_{n-1}) = x_{n-1}) = \mathbb{P}(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}).$$

Die Zukunft hängt also von der Vergangenheit nur über den letzten Wert $X(t_{n-1})$ ab. Ein Beispiel für Markovprozesse sind Prozesse mit unabhängigen Zuwächsen.

Definition 3.4 Der Prozess $X_t, t \in T$ heißt stationär, wenn für $t_1 < t_2 < \dots < t_n$ und $h > 0$ die gemeinsame Verteilung von $(X(t_1), \dots, X(t_n))$ mit der von $(X(t_1 + h), \dots, X(t_n + h))$ übereinstimmt.

3.2. Markovketten in diskreter Zeit

3.2.1. Übergangswahrscheinlichkeiten

Markovprozesse mit diskretem Zustandsraum nennen wir Markovketten. Wir können noch zwischen Markovketten in diskreter und in stetiger Zeit unterscheiden. Die Diskussion von Markovketten in stetiger Zeit werden wir auf später verschieben. In beiden Fällen kann man die diskrete Verteilung der einzelnen Zufallsvariablen durch ihre Wahrscheinlichkeitsfunktion beschreiben, deshalb erhält die Markoveigenschaft die besonders einfache Form

$$\mathbb{P}(X_n = x_n | X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}).$$

Die Wahrscheinlichkeiten

$$\mathbb{P}(X_{n+1} = j | X_n = i)$$

nennen wir die Übergangswahrscheinlichkeiten der Markovkette. Wenn diese nicht von n abhängen, sprechen wir von einer homogenen Markovkette und setzen

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$$

Definition 3.5 Die Wahrscheinlichkeiten

$$p_{ij}(t) = \mathbb{P}(X_{n+t} = j | X_n = i)$$

nennen wir die *t-stufigen Übergangswahrscheinlichkeiten*. Aus dem Satz von der vollständigen Wahrscheinlichkeit erhalten wir die **Chapman-Kolmogorov Gleichungen**

$$p_{ij}(s+t) = \sum_{k \in M_X} p_{ik}(s)p_{kj}(t).$$

in Matrixnotation mit

$$P(t) = (p_{ij}(t))_{M_X \times M_X}$$

lauten die Chapman-Kolmogorov Gleichungen

$$P(t+s) = P(t)P(s),$$

und mit $P = P(1)$

$$P(t) = P^t.$$

Wir nennen P die **Übergangsmatrix** und $P(t)$ die **t-stufige Übergangsmatrix**.

Wir setzen zusätzlich $p_i(t) = \mathbb{P}(X_t = i)$ und $p(t) = (p_i(t), i \in M_X)$ (als Zeilenvektor). Wieder mit dem Satz von der vollständigen Wahrscheinlichkeit erhalten wir

$$p(t) = p(0)P^t.$$

Wie man am besten Matrixpotenzen berechnet, siehe Anhang B.1.2. Durch $p(0)$ und P werden alle endlichdimensionalen Verteilungen festgelegt.

In Hinkunft verwenden wir zur Abkürzung die Notationen

$$\mathbb{P}_i(A) = \mathbb{P}(A|X_0 = i)$$

und

$$\mathbb{E}_i(Y) = \mathbb{E}(Y|X_0 = i).$$

3.2.2. Klasseneigenschaften

Wir definieren

Definition 3.6 Der Zustand j heißt **Nachfolger** von i ($i \rightarrow j$), wenn es ein $t \geq 0$ gibt, sodass $p_{ij}(t) > 0$.

Wenn sowohl $i \rightarrow j$ als auch $j \rightarrow i$ gilt, dann heißen i und j **verbunden** oder **kommunizierend**.

Das Kommunizieren ist eine Äquivalenzrelation, wir können daher den Phasenraum in die Äquivalenzklassen zerlegen, die wir Rekurrenzklassen oder kurz Klassen nennen. Gibt es nur eine Klasse (wenn also alle Zustände miteinander kommunizieren), heißt die Markovkette **irreduzibel**. Ein Zustand mit $p_{ii} = 1$ heißt **absorbierender Zustand**. Ein solcher Zustand ist offensichtlich eine Klasse für sich (no pun intended).

Definition 3.7 Eine Eigenschaft heißt **Klasseneigenschaft**, wenn sie entweder für alle Zustände einer Klasse oder für keinen gilt.

Ein einfaches Beispiel einer Klasseneigenschaft ist die Periode:

Definition 3.8 Die **Periode** eines Zustandes ist

$$d(i) = \text{ggT}\{t \geq 0 : p_{ii}(t) > 0\}.$$

Sie definiert also den kleinsten möglichen Weg, um von Zustand i wieder zum Zustand i zurückzukehren.

Etwas spannender ist die **Rekurrenz**: dazu definieren wir zuerst

$$\tau_i = \inf\{t > 0 : X_t = i\},$$

die **Übergangszeit** bzw. **Rückkehrzeit** (je nachdem, ob $X_0 \neq i$ ist oder nicht) nach i , und

$$\nu_i = \#\{t > 0 : X_t = i\},$$

die Anzahl der Besuche in i .

Satz 3.1 *Die folgenden Bedingungen sind äquivalent:*

1. $\mathbb{P}_i(\tau_i < \infty) = 1$,
2. $\mathbb{P}_i(\nu_i = \infty) = 1$,
3. $\mathbb{E}_i(\nu_i) = \infty$,
4. $\sum_t p_{ii}(t) = \infty$.

Wenn diese Bedingungen erfüllt sind, dann heißt i rekurrent, sonst transient. Rekurrenz und Transienz sind Klasseigenschaften.

Bei der Rekurrenz kann man weiter unterscheiden:

Definition 3.9 *i sei ein rekurrenter Zustand. Wenn*

$$\mathbb{E}_i(\tau_i) < \infty$$

*gilt, dann heißt i **positiv rekurrent**, sonst **nullrekurrent**.*

Definition 3.10 *$(\pi_i, i \in M_X)$ heißt **stationäre Verteilung**, wenn*

$$\pi_i \geq 0,$$

$$\sum_i \pi_i = 1$$

und

$$\sum_i \pi_i p_{ij} = \pi_j.$$

Die stationäre Verteilung ist die Verteilung, die man für die Zustände einer Markovkette erhält, wenn man sie nach einer sehr langen Zeit betrachtet. Dann sollten die Anfangseffekte (= Anfangswahrscheinlichkeit) nicht mehr sichtbar sein. Dazu müssen wir erst die Wahrscheinlichkeit für einen Zustand i zum Zeitpunkt t berechnen.

Satz 3.2 Wenn (X_n) irreduzibel und aperiodisch ist, dann existieren die Grenzwerte

$$\lim_{n \rightarrow \infty} p_{ij}(t) = \pi_j = \frac{1}{\mathbb{E}_j(\tau_j)}.$$

Im periodischen Fall gilt

$$\pi_j = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n p_{ij}(t).$$

Wenn (π_i) nicht identisch verschwindet (also wenn die Kette positiv rekurrent ist), dann ist es eine **stationäre Verteilung**. Umgekehrt folgt aus der Existenz einer stationären Verteilung die positive Rekurrenz. Die positive Rekurrenz bzw. Nullrekurrenz ist ebenfalls eine Klasseneigenschaft.

Für einen absorbierenden Zustand i_0 definieren wir die **Absorptionswahrscheinlichkeit** a_i als

$$\mathbb{P}_i(\tau_{i_0} < \infty) = \mathbb{P}_i(X \text{ wird in } i_0 \text{ absorbiert}).$$

Satz 3.3 Die Absorptionswahrscheinlichkeiten sind die kleinste nichtnegative Lösung des Gleichungssystems

$$\begin{aligned} a_{i_0} &= 1, \\ a_i &= \sum_j p_{ij} a_j, i \neq i_0. \end{aligned}$$

Das gibt uns eine Möglichkeit, die Transienz oder Rekurrenz einer irreduziblen Markovkette zu entscheiden. Wir wählen einen Zustand und machen ihn zu einem absorbierenden Zustand und bestimmen für die modifizierte Übergangsmatrix die Absorptionswahrscheinlichkeiten. Sind diese gleich 1, ist die Kette rekurrent.

Wir nehmen an, dass i_0 der einzige absorbierende Zustand ist und alle anderen Zustände kommunizieren und die Absorptionswahrscheinlichkeiten 1 sind. Dann erhalten wir für die mittlere Zeit bis zur Absorption (**mittlere Absorptionszeit**)

$$m_i = \mathbb{E}_i(\tau_{i_0})$$

eine ähnliche Gleichung:

$$\begin{aligned} m_{i_0} &= 0, \\ m_i &= 1 + \sum_j p_{ij} m_j, i \neq i_0. \end{aligned}$$

Allgemeiner kann man eine Zufallsvariable $S = \sum_{t=1}^{\tau-1} x_{t,t+1}$ betrachten (d.h., wir “sammeln” auf unserem Weg zur Absorption Beträge x_{ij} , die von dem jeweiligen Übergang abhängen dürfen).

Wir setzen

$$m_i = \mathbb{E}_i(S)$$

und

$$v_i = \mathbb{V}_i(S).$$

Dann sind m_i und v_i die Lösungen der Gleichungen

$$\begin{aligned}
 m_{i_0} &= 0, v_{i_0} = 0, \\
 m_i &= \sum_j p_{ij} x_{ij} + \sum_j p_{ij} m_j, i \neq i_0. \\
 v_i &= \sum_j p_{ij} (x_{ij} i + m_j - m_i)^2 + \sum_j p_{ij} v_j, i \neq i_0.
 \end{aligned}$$

Wenn es mehrere absorbierende Zustände gibt, $A = \{i_1, \dots, i_k\}$, dann kann man die Wahrscheinlichkeiten der Absorption $a_i(i_j)$ in dem absorbierenden Zustand i_j oder allgemeiner in einem der Zustände in der Menge $B \subseteq A$ ($a_i = a_i(B)$). Das gibt die Gleichungen

$$\begin{aligned}
 a_i &= \sum_j p_{ij} a_j, i \notin A \\
 a_i &= 1, i \in B \\
 a_i &= 0, i \in A \setminus B.
 \end{aligned}$$

Manchmal ist es interessant (wieviel Geld werde ich haben, wenn ich nicht bankrott gehe?), den Erwartungswert von S unter der Bedingung, dass die Absorption in einem bestimmten Zustand oder in einer bestimmten Teilmenge der Menge der absorbierenden Zustände erfolgt. Da hilft es, dass $X(t)$ unter der Bedingung der Absorption in B wieder eine Markovkette bildet, mit den Übergangswahrscheinlichkeiten

$$p_{ij}^B = \frac{p_{ij} a_j(B)}{a_i(B)}.$$

Mit diesen modifizierten Übergangswahrscheinlichkeiten (wobei die absorbierenden Zustände $\notin B$ weggelassen werden) kann man die Gleichungen für m_i v_i lösen und erhält so die bedingte Erwartung bzw. Varianz.

3.2.3. Markov Chain Monte Carlo

4. Statistik

4.1. Grundlagen der Statistik

Wenn klar ist, worüber wir eine Aussage machen wollen, kann man sich auf die Einzelbeobachtungen stürzen:

Grundgesamtheit Ω Ist die Menge aller statistischen Einheiten (Untersuchungseinheiten) mit übereinstimmenden Identifikationskriterien (sachlich, räumlich, zeitlich).

Untersuchungseinheit ω je nach Kontext auch Elemente, Objekte oder Individuen.

Merkmal müssen alle Elemente der Grundgesamtheit besitzen, dieses kann anschließend verglichen werden.

Ausprägung eines Merkmals ist dann der Wert, den die Untersuchungseinheit für ein Merkmal besitzt.

Beispiel 4.1 Eine Melone, die am 04.03.2013 von einer Obstverkäuferin angeboten wird, ist eine **Untersuchungseinheit** ω aus der **Grundgesamtheit** Ω der Melonen, die am selben Tag von dieser Obstverkäuferin verkauft werden.
Das Gewicht einer Melone ist das **Merkmal**, das jede Melone (die am 04.03.2013 von der Obstverkäuferin verkauft wird) besitzen muss. Dass die vorher beschriebene Melone das Gewicht 1kg besitzt, ist die **Ausprägung** dieses **Merkmals**.

Somit muss für jedes ω klar sein, ob $\omega \in \Omega$ oder $\omega \notin \Omega$.

4.1.1. Modelle

Definition 4.1 (Statistisches Modell) : Ein statistisches Modell ist eine Menge \mathcal{P} von Verteilungen (siehe 2.10 bzw. 2.11), die für das Problem in Frage kommen.
Es wird unterschieden zwischen **parametrischen Modellen** und **nicht parametrischen Modellen**.

parametrisches Modell Der Verteilungstyp/die Form (z.B. Alternativverteilung, Normalverteilung, usw.) ist festgelegt, jedoch die vollständige Wahrscheinlichkeitsverteilung noch nicht.

Das bedeutet, einige Eigenschaften der Verteilung können noch bestimmt und verändert werden. (z.B.: die Parameter p, μ und $\sigma^2 \dots$)

$$\mathcal{P} = \{P_\theta, \theta \in \Theta\} \quad \text{mit } \Theta \subseteq \mathbb{R}^d$$

Wobei $\Theta \dots$ Parameterraum, $\theta \dots$ Parameter (kann Mehrdimensional sein, z.B. $\theta = \mathcal{N}(\mu, \sigma^2)$)

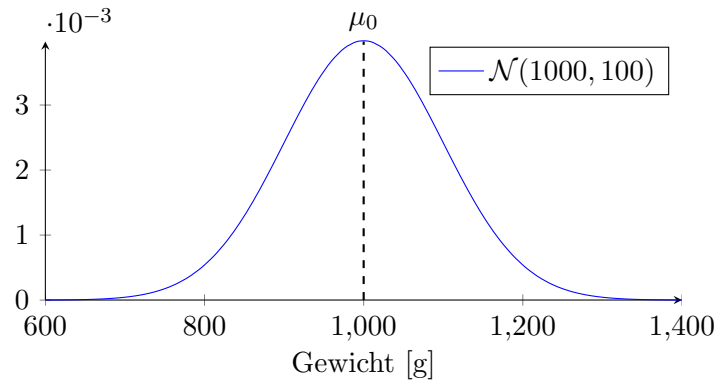


Abbildung 4.1.: Zu Beispiel 4.2: Verteilung des Gewichtes von Melonen. ($\mu_0 = 1000$, $\sigma_0 = 100$)

ACHTUNG: Es gibt nicht das eine "wahre" Modell, es gibt nur brauchbare(plausible) oder unbrauchbare Modelle. Jedoch gibt es sehr wohl die "wahre" Verteilung mit dem "wahren" Parameter. Denn es wird gefordert, dass alle Parameter eindeutig festgelegte Werte besitzen.

Beispiel 4.2 Das Gewicht Y der Melonen einer Obstverkäuferin ist Normalverteilt, $Y \sim \mathcal{N}(\mu, \sigma^2)$ mit unbekanntem μ und σ .

Der Parameter ist 2-dimensional: $\theta = (\mu, \sigma)$ und der Parameterraum $\Theta = \mathbb{R} \times \mathbb{R}_+$.

Ein Beispiel wäre z.B.: $\mu_0 = 1000$ und $\sigma_0 = 100$. (siehe Abbildung 4.1)

nicht parametrisches Modell Die Modellstruktur wird nicht von vornherein festgelegt, sondern aus den Daten bestimmt. Es bedeutet nicht, dass dieses Modell keine Parameter besitzt, sondern dass die Art und die Anzahl der Parameter flexibel und nicht von vornherein festgelegt ist. "alle" Verteilungen sind möglich.

Definition 4.2 (Stichprobe: (X_1, \dots, X_n)) :

Eine Stichprobe ist eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit einer (unbekannten) Verteilung \mathcal{P} . Sie ist sozusagen eine Teilmenge der Grundgesamtheit. Der **Stichprobenumfang** n ist die Anzahl der Proben einer Grundgesamtheit.

Die Unabhängigkeit der einzelnen Stichproben erreicht man beispielsweise durch Ziehungen mit zurücklegen.

Definition 4.3 (Das Stichprobenmittel) :

$$\bar{X}_n = \sum_{i=1}^n X_i$$

ist eine erwartungstreue Schätzfunktion mit $\mathbb{E}(\bar{X}_n) = \mu$ und $\mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n}$

Beispiel 4.3 In einer Firma arbeiten 100 Personen (Grundgesamtheit=100). Davon sind 30 weiblich und 70 Männlich. Es gibt 5 Abteilungen in der Firma, in denen die Personen arbeiten.

Wie kann man nun Stichproben aus diesen Personen auswählen?

Zufällige Auswahlen :

- i) Arbeiter durchnummerieren (1-100), dann 20 Zahlen zufällig ziehen.
- ii) Jeder Arbeiter würfelt, bei einer 6 wird er gezogen.
- iii) Eine der 5 Abteilungen wird zufällig gewählt.

Bewusste Auswahlen :

- i) Bei der Betriebsversammlung die Personen wählen, auf die zufällig der Blick vom Chef fällt.
- ii) Arbeiter wählen, die als erste oder letzte zur Arbeit kommen.
- iii) besonders nette Arbeiter wählen.

4.1.2. Aufgaben der Statistik

Die Aufgabe der (induktiven/schließenden) Statistik ist es nun, auf Grund der Stichprobe (X_1, \dots, X_n) Aussagen über die unbekannte Verteilung bzw. der unbekannten Parameter der Grundgesamtheit Ω zu treffen.

Definition 4.4 eine Statistik (auch Teststatistik) T ist eine Zufallsvariable, die aus einer Stichprobe berechnet wird: $T = T(X_1, \dots, X_n)$

Die Statistik besitzt 2 Grundaufgaben

Schätzen: Aus einer Stichprobe einen Schätzwert für θ bestimmen. (+ evtl. die Genauigkeit bestimmen)

Testen: Wenn eine Aussage über die Verteilung existiert, muss festgestellt werden, ob diese Aussage zutrifft. (gut/schlecht)

4.2. Schätztheorie

Hierbei unterscheiden wir zwischen **Punktschätzung** und **Intervallschätzung**.

Beispiel 4.4 Greifen wir nochmal das Beispiel mit der Firma auf: Beim Zählen aller Personen aus einer Abteilung (20 Personen gesamt, 6 davon sind Frauen) kommt der Statistiker auf einen Anteil der Frauen von $\theta = p = \frac{6}{20} = 0.3$.

Nimmt der Statistiker diesen beobachteten Wert (0.3) als $\hat{\theta}_n$, so hat man ein Beispiel für eine Punktschätzung.

Ist der Statistiker jedoch vorsichtiger, so könnte er diesen Wert durch ein Intervall abschätzen: $0.2 \leq \hat{\theta}_n \leq 0.4$. Oder, er sagt, dass der Frauenanteil $\hat{\theta}_n \geq 0.25$

4.2.1. Punktschätzung

Die erste Aufgabe, mit der wir uns beschäftigen, besteht darin, aus einer Stichprobe Schätzwerte für den unbekannten Parameter zu bestimmen. Wir definieren

Definition 4.5 Ein Schätzer $\hat{\theta}_n$ ist eine Folge von Statistiken:

$$(\hat{\theta}_n, n \in \mathbb{N}) : \hat{\theta}_n = \hat{\theta}_n(\underbrace{X_1, \dots, X_n}_{\text{Stichprobe}}) \Rightarrow \text{Schätzfunktion}$$

Wobei hier $\hat{\theta}_n$ der Schätzwert für den Parameter θ ist.

Diese Definition lässt recht dumme Schätzer zu (z.B. 42, vgl. Douglas Adams), deshalb definieren wir einige wünschenswerte Eigenschaften, die wir von Schätzern erwarten:

Definition 4.6 (Eigenschaften von Schätzern) :

i) **Konsistenz:** $\hat{\theta}_n \rightarrow \theta$ für $n \rightarrow \infty$:

Das bedeutet, sie ist eine Eigenschaft, die nur für große Stichprobenumfänge gilt. (D.h. eine konsistente Schätzfunktion kann für endliche Stichprobenumfänge eine große Varianz (Verzerrung) besitzen.)

a) Schätzer $\hat{\theta}_n$ heißt schwach konsistent, wenn $\hat{\theta}_n \rightarrow \theta$ in Wahrscheinlichkeit konvergiert (also $\lim_{n \rightarrow \infty} \mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \epsilon) = 0 \quad \forall \epsilon > 0 \quad \forall \theta \in \Theta$).

Das bedeutet, die Wahrscheinlichkeit, dass die Schätzfunktion $\hat{\theta}_n$ Werte außerhalb eines beliebig kleinen Intervalls ϵ um den wahren Wert θ annimmt, mit steigendem n gegen 0 geht.

(Siehe Abbildung 4.2)

b) Schätzer $\hat{\theta}_n$ heißt stark konsistent, wenn $\hat{\theta}_n \rightarrow \theta$ mit Wahrscheinlichkeit 1 für steigende n konvergiert (also $\mathbb{P}_\theta(\lim_{n \rightarrow \infty} |\hat{\theta}_n - \theta| = 0) = 1 \quad \forall \theta \in \Theta$).

ii) **Erwartungstreue, Unverzerrtheit:**

Schätzer $\hat{\theta}_n$ heißt erwartungstreu(unverzerrt), wenn

$$\mathbb{E}_\theta(\hat{\theta}_n) = \theta.$$

Also die Schätzfunktion soll gleich dem wahren Parameter θ sein.

Verzerrung(Bias): ist definiert als $\text{bias}(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n) - \theta$.

(\Rightarrow wenn $\text{bias}(\hat{\theta}_n) = 0$, dann ist $\hat{\theta}_n$ erwartungstreu.)

iii) **Effizienz:**

Schätzer $\hat{\theta}_n$ heißt effizient, wenn er erwartungstreu ist und die kleinste Varianz unter allen erwartungstreuen Schätzern hat.

z.B: 2 Schätzer($\hat{\theta}_n$ und $\tilde{\theta}_n$): dann ist $\hat{\theta}_n$ effizienter, als $\tilde{\theta}_n$, wenn gilt:

$$\mathbb{V}(\hat{\theta}_n) \leq \mathbb{V}(\tilde{\theta}_n) \quad \forall \theta$$

ACHTUNG: Die Eigenschaft "erwartungstreu" bedeutet nicht, dass man erwarten kann, dass der Schätzer den wahren Wert liefert. Es sagt nur, dass der Ver-

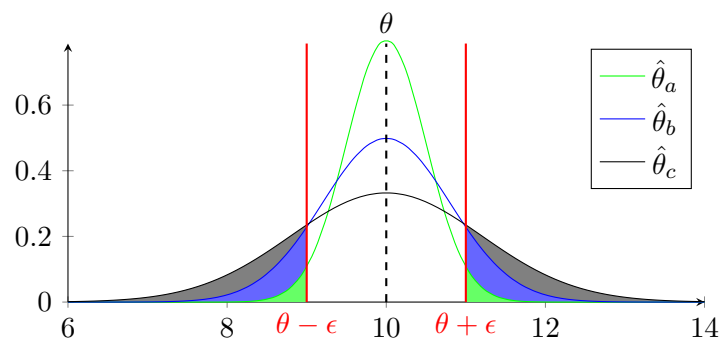


Abbildung 4.2.: Konsistenz der einzelnen Schätzer $\hat{\theta}_a$, $\hat{\theta}_b$ und $\hat{\theta}_c$, wobei für die Stichprobenumfänge gilt: $a > b > c$

teilungsschwerpunkt der Schätzwerte der wahre Wert θ ist und nicht systematisch abweicht.

Momentenmethode

Eine einfache Methode, um Schätzer zu konstruieren, ist die Momentenmethode. Mit ihr kann man die unbekannten Komponenten des Parametervektors $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ bestimmen.

Ihr liegt das Gesetz der großen Zahlen zugrunde. Diese Methode ist sehr einfach, jedoch sind die resultierenden Schätzer nicht immer Erwartungstreu.

Definition 4.7 (Moment) :

Das Moment der Ordnung k von X (wobei X eine Zufallsvariable ist) ist wie folgt definiert:

$$m_k = \mathbb{E}(X^k) \quad \forall k \in \mathbb{N}$$

Das k -te Moment von X ist sozusagen der Erwartungswert der k -ten Potenz von X .

Vorgangsweise, um k unbekannte Parameter $\vec{\theta}$ zu berechnen: Die ersten k -Momente der Verteilung können als Funktion der Parameter $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ ausgedrückt werden:

$$\mathbb{E}(X^j) = m_j = g_j(\theta_1, \theta_2, \dots, \theta_k) \quad \forall j = 1, \dots, k \text{ bzw.}$$

$$\mathbb{E}(X^j) = m_j = g_j(\vec{\theta}) \quad \forall j = 1, \dots, k$$

Die **empirischen Momente**

$$\hat{m}_j = \frac{1}{n} \sum_{i=1}^n x_i^j \quad \forall j = 1, \dots, k$$

können anschließend für die Momente $m_j \quad \forall j = 1, \dots, k$ eingesetzt werden: (dabei wird aus dem Parametervektor $\vec{\theta}$ der Schätzvektor $\vec{\theta}_n$.)

$$g_j(\vec{\theta}_n) = \frac{1}{n} \sum_{i=1}^n x_i^j = \mathbb{E}(X^j) \quad \forall j = 1, \dots, k$$

$$\vec{\theta}_n = g_j^{-1} \left(\frac{1}{n} \sum_{i=1}^n x_i^j \right) = g_j^{-1}(\mathbb{E}(X^j)) \quad \forall j = 1, \dots, k$$

Damit hat man ein Gleichungssystem mit k -Gleichungen, die gelöst werden müssen.

Beispiel 4.5 Angenommen, Ω ist Poissonverteilt ($\mathcal{P}(\lambda)$): $\Rightarrow \mathbb{E}(X) = \lambda$.

D.h. nur ein Parameter λ ist zu berechnen, folglich ist unser $k = 1$:

$$\theta_1 = \lambda = \mathbb{E}(X) = m_1 \quad \text{und} \quad \hat{m}_1 = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\Rightarrow \hat{\theta}_n = \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$$

Beispiel 4.6 Angenommen, Ω ist Normalverteilt ($N(\mu, \sigma^2)$):

$\Rightarrow \mathbb{E}(X) = \mu$ und $\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Da die Normalverteilung 2 Parameter (μ und σ) besitzt, ist unser $k = 2$:

$$\text{Mit } \theta_1 = \mu = \mathbb{E}(X) = m_1 \quad \text{und} \quad \hat{m}_1 = \frac{1}{n} \sum_{i=1}^n x_i \quad \Rightarrow \quad \hat{\theta}_1 = \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{Mit } \theta_2 = \mathbb{E}(X^2) = m_2 \quad \text{und} \quad \hat{m}_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 \quad \Rightarrow \quad \hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n x_i^2$$

$$\Rightarrow \hat{\sigma}^2 = \hat{\theta}_2 + (\hat{\theta}_1)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 + \underbrace{\left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2}_{\bar{x}_n^2} = \frac{1}{n} \sum_{i=1}^n x_i^2 + \bar{x}_n^2$$

Maximum Likelihood Methode

Eine andere Methode um Schätzer zu konstruieren, ist die Maximum Likelihood Methode:

Beispiel 4.7 Fortsetzung zu Beispiel 4.4 nehmen wir an, die Anzahl der weiblichen Arbeiter Y ist Binomialverteilt ($Y \sim B_n(\theta)$) mit Parameter θ .

Wie wahrscheinlich ist es nun, im Modell der Binomialverteilung, den beobachteten Wert $k = 6$ zu erhalten?

$$P_\theta(Y = k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} \quad \Rightarrow \quad P_\theta(Y = 6) = \binom{20}{6} \theta^6 (1 - \theta)^{20-6}$$

Siehe Abbildung 4.3, hier sieht man, dass die Wahrscheinlichkeit, dass der Anteil der Frauen $\theta = 0.3$ beträgt, am höchsten ist (rote Linie) mit $P_{0.3}(Y = 6) = 0.19$.

Bei $\theta = 0.4$ (grün) ist $P_{0.4}(Y = 6) = 0.12$ und bei $\theta = 0.5$ ist $P_{0.5}(Y = 6) = 0.04$, d.h. die Wahrscheinlichkeit ist immer noch relativ groß.

Erst bei $\theta = 0.6$ ist die Wahrscheinlichkeit $P_{0.6}(Y = 6) = 0.0049$ ein kleiner Bruchteil von $P_{0.3}$.

Wenn die Wahrscheinlichkeit, die aktuelle Zufallsvariable(Ereignis) zu erhalten, für einen Parameter sehr klein ist, spricht das gegen diesen Parameter. Umgekehrt heißt das, dass

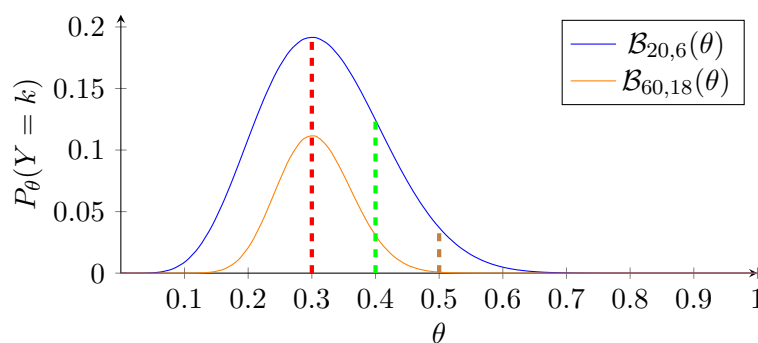


Abbildung 4.3.: Zu Beispiel 4.7: Zeigt die Wahrscheinlichkeiten $P_\theta(Y = 6)$ für die einzelnen θ -Werte. ($\mathcal{B}_{20,6}$). Außerdem sieht man, dass die Wahrscheinlichkeit bei einer 3x größeren Stichprobe ($n = 20 \cdot 3$ und $k = 6 \cdot 3$) nicht mehr so stark streut. ($\mathcal{B}_{20 \cdot 3, 6 \cdot 3}$)

ein Parameterwert umso plausibler ist, je größer die Wahrscheinlichkeit der Zufallsvariable unter diesem Parameter ist.

Hier kommt nun die Likelihood-Funktion ins Spiel:

Definition 4.8 Die Likelihoodfunktion für eine diskrete Zufallsvariable (Ereignis) X , mit gegebenem Modell (Verteilung), abhängig von Parameter $\theta \in \Theta$ ist definiert als:

$$L(X; \theta) = c(X) \cdot P_\theta(X)$$

und für stetige Zufallsvariablen X mit der Dichte $f_\theta(X)$ folgt:

$$L(X; \theta) = c(X) \cdot f_\theta(X)$$

Das **Likelihood-Prinzip**: Berechne die Wahrscheinlichkeit, genau die Stichprobe ω aus Ω zu ziehen.

Wobei $c(X)$ ein konstanter Faktor ist, der NICHT von θ abhängt. \Rightarrow die Likelihood-Funktion ist mehrdeutig.

Definition 4.9 Die Log-Likelihood-Funktion ist wie folgt definiert:

$$l(X; \theta) = \ln L(X; \theta)$$

sie wird häufig verwendet, denn sie ist meist viel handlicher.

Definition 4.10 Multiplikationssatz für Likelihood-Funktionen:

Die Stichprobe $\vec{X} = (X_1, X_2, \dots, X_n)$ setzt sich aus n unabhängigen Zufallsvariablen (Ereignissen) zusammen:

$$L(\vec{X}; \theta) = L(X_1, X_2, \dots, X_n; \theta) = \prod_{i=1}^n L(X_i; \theta)$$

somit gilt für die Log-Likelihood-Funktion:

$$l(\vec{X}; \theta) = l(X_1, X_2, \dots, X_n; \theta) = \sum_{i=1}^n l(X_i; \theta)$$

Beispiel 4.8 Ein Mitarbeiter wurde beauftragt, den Anteil θ der weiblichen Arbeiter zu schätzen. Dazu schreibt er sich zu jeder Person X_i auf, ob diese männlich ($X_i = 0$), oder weiblich ($X_i = 1$) war.

Um den Anteil nun zu bestimmen, hat er sich die 20 Personen aus der Selben Abteilung wie in Beispiel 4.7 angesehen und folgende Überlegung gemacht: Da jede Person die gleiche Wahrscheinlichkeit besitzt, gilt folgendes:

$$L(X_i = 1; \theta) = P(X_i = 1; \theta) = \theta$$

$$L(X_i = 0; \theta) = P(X_i = 0; \theta) = 1 - \theta$$

Die Stichprobe lautet wie folgt:

$$\begin{aligned} \vec{\omega} = \{ & X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0, X_7 = 1, X_8 = 1, \\ & X_9 = 0, X_{10} = 0, X_{11} = 0, X_{12} = 0, X_{13} = 0, X_{14} = 1, X_{15} = 0, \\ & X_{16} = 0, X_{17} = 0, X_{18} = 0, X_{19} = 1, X_{20} = 0 \} \end{aligned}$$

Mit dem Multiplikationssatz (siehe Definition 4.10) gilt nun:

$$\begin{aligned} L(\vec{\omega}; \theta) &= L(X_1 = 1; \theta) \cdot L(X_2 = 0; \theta) \cdot \dots \cdot L(X_{20} = 0; \theta) \\ L(\vec{\omega}; \theta) &= (1 - \theta) \cdot \theta \cdot (1 - \theta) \cdot \dots \cdot \theta \cdot (1 - \theta) = \theta^6 \cdot (1 - \theta)^{14} \end{aligned}$$

Das ist aber wiederum das selbe Ergebnis, wie das aus Beispiel 4.7, bis auf den konstanten Faktor $\binom{20}{6}$, dieser ist aber nicht so wichtig, denn die Likelihood-Funktion ist, wie bereits erwähnt mehrdeutig.

Und nun kommen wir zum Maximum-Likelihood-Schätzer:

Definition 4.11 Der Maximum-Likelihood (ML-) Schätzer $\hat{\theta}_n$ ist der Wert von θ , der die Likelihoodfunktion im Parameterraum Θ maximiert:

$$L(X, \hat{\theta}_n) \geq L(X, \theta) \quad \forall \theta \in \Theta$$

Der ML benötigt das Maximum der Likelihood-Funktion, somit müssen wir $L(X; \theta)$ einmal ableiten, um das Maximum zu erhalten. Meist ist es aber einfacher, statt der Likelihood-Funktion die Log-Likelihood-Funktion $l(X; \theta)$ zu verwenden.

Beispiel 4.9 Fortsetzung von Beispiel 4.8: (Log-Produktregeln siehe Anhang B.4.1)

$$L(\vec{\omega}; \theta) = \theta^6 \cdot (1 - \theta)^{14} \Rightarrow l(\vec{\omega}; \theta) = \ln(\theta^6 \cdot (1 - \theta)^{14}) = 6 \cdot \ln(\theta) + 14 \cdot \ln(1 - \theta)$$

$$\frac{dl(\vec{\omega}; \theta)}{d\theta} = l'(\vec{\omega}; \theta) = \frac{6}{\theta} + \frac{14}{1 - \theta}(-1)$$

Da $l(\vec{\omega}; \theta)$ am Maximum ($\theta = \hat{\theta}_n$) die Steigung=0 besitzt, folgt:

$$l'(\vec{\omega}; \hat{\theta}_n) = 0 = \frac{6}{\hat{\theta}_n} + \frac{14}{1 - \hat{\theta}_n}(-1) \Rightarrow \frac{6}{\hat{\theta}_n} = \frac{14}{1 - \hat{\theta}_n} \Rightarrow 6 - 6\hat{\theta}_n = 14\hat{\theta}_n \Rightarrow 20\hat{\theta}_n = 6$$

$$\hat{\theta}_{20} = \frac{6}{20} = 0.3$$

Das ist auch genau das, was man erwartet, wenn in einer Stichprobe mit dem Umfang 20 sechs Elemente die Ausprägung weiblich besitzen.

Beispiel 4.10 Gesucht sind μ und σ der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ mittels Maximum-Likelihood-Schätzer:

Die Dichte der Normalverteilung ist ja wie folgt definiert, als

$$f_{\mu, \sigma}(x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \mu)^2}{2\sigma^2}}$$

ohne der Konstante $\frac{1}{\sqrt{2\pi}}$ erhalten wir für den Log-Likelihood:

$$l(x_i; \mu, \sigma) = \ln \left(\frac{1}{\sigma} e^{-\frac{(y_i - \mu)^2}{2\sigma^2}} \right) = \ln(\sigma^{-1}) + \ln \left(e^{-\frac{(y_i - \mu)^2}{2\sigma^2}} \right) = -\ln \sigma - \frac{(y_i - \mu)^2}{2\sigma^2}$$

Somit gilt für $\vec{x} = (x_1, \dots, x_n)$:

$$l(\vec{x}; \mu, \sigma) = \sum_{i=1}^n l(x_i; \mu, \sigma) = \sum_{i=1}^n \left(-\ln \sigma - \frac{(y_i - \mu)^2}{2\sigma^2} \right) = -n \cdot \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2$$

Mit dem Steinerschen Verschiebungssatz (siehe Satz 2.3) folgt:

$$\mathbb{E}((X - a)^2) = \mathbb{V}(X) + (\mathbb{E}(X) - a)^2 \Rightarrow \frac{\sum_{i=1}^n ((X - a)^2)}{n} = \mathbb{V}(X) + \left(\frac{\sum_{i=1}^n (X)}{n} - a \right)^2$$

$$\sum_{i=1}^n ((y_i - \mu)^2) = n \cdot \mathbb{V}(\vec{y}) + n \cdot \left(\frac{\sum_{i=1}^n (y_i)}{n} - \mu \right)^2 = n \cdot \mathbb{V}(\vec{y}) + n \cdot (\bar{y} - \mu)^2$$

Somit erhalten wir:

$$l(\vec{x}; \mu, \sigma) = -n \cdot \ln \sigma - \frac{n}{2\sigma^2} (\mathbb{V}(\vec{y}) + (\bar{y} - \mu)^2)$$

Zum bestimmen des ML-Schätzers müssen wir Differenzieren:

Bestimmen von $\hat{\mu}$:

$$\frac{\partial l(\vec{x}; \mu, \sigma)}{\partial \mu} = \frac{\partial \left(-n \cdot \ln \sigma - \frac{n \mathbb{V}(\vec{y})}{2\sigma^2} + \frac{n \cdot (\bar{y} - \mu)^2}{2\sigma^2} \right)}{\partial \mu} = \frac{2 \cdot n \cdot (\bar{y} - \mu)}{2\sigma^2} = 0$$

Umformen auf μ :

$$\bar{y} - \mu = 0 \Rightarrow \bar{y} = \hat{\mu}$$

Bestimmen von $\hat{\sigma}$:

$$\frac{\partial l(\vec{x}; \mu, \sigma)}{\partial \sigma} = \frac{\partial \left(-n \cdot \ln \sigma - \frac{n \mathbb{V}(\vec{y})}{2\sigma^2} + \frac{n(\bar{y} - \mu)^2}{2\sigma^2} \right)}{\partial \sigma} = \frac{-n}{\sigma} + \frac{n}{2} 2\sigma^{-3} (\mathbb{V}(\vec{y}) + (\bar{y} - \mu)^2) = 0$$

Umformen auf σ :

$$\frac{\mathcal{K}}{\sigma^3} (\mathbb{V}(\vec{y}) + (\bar{y} - \mu)^2) = \frac{\mathcal{K}}{\sigma} \Rightarrow (\mathbb{V}(\vec{y}) + (\bar{y} - \mu)^2) = \sigma^2 \Rightarrow \hat{\sigma} = \sqrt{(\mathbb{V}(\vec{y}) + (\bar{y} - \mu)^2)}$$

Eingesetzt für $\mu = \hat{\mu} = \bar{y}$ folgt:

$$\hat{\sigma} = \sqrt{(\mathbb{V}(\vec{y}) + (\bar{y} - \bar{y})^2)} = \sqrt{(\mathbb{V}(\vec{y}))}$$

Somit sind bei der Normalverteilung der Mittelwert und die Wurzel aus der Varianz die Maximum Likelihood Schätzer für μ und σ .

$$\hat{\mu} = \bar{y} \text{ und } \hat{\sigma} = \sqrt{\mathbb{V}(\vec{y})}$$

Damit der ML-Schätzer konsistent ist, müssen die Dichten gewisse Regularitätsvoraussetzungen erfüllen. Bei der Suche nach einem effizienten Schätzer hilft der

Satz 4.1 (Cramér-Rao) Wenn p_θ bzw. f_θ zweimal nach θ differenzierbar ist und zusätzliche Regularitätsvoraussetzungen erfüllt, dann gilt für jeden **Erwartungstreuen** Schätzer $\hat{\theta}_n$

$$\mathbb{V}(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}.$$

D.h. sie liefert eine untere Schranke für die Varianz eines Schätzers $\hat{\theta}_n$ bzgl. des zu schätzenden Parameters θ . Dabei ist

$$I_n(\theta) = -\mathbb{E}_\theta \left(\frac{\partial^2}{\partial \theta^2} \log(L(X_1, \dots, X_n; \theta)) \right)$$

die **Fisher-Information**.

Wenn die Cramér-Rao-Schranke angenommen wird, ist der ML-Schätzer effizient.

Definition 4.12 Die **Fischerinformation** ist für stetige Zufallsvariablen X mit der Dichte $f_\theta(X)$ definiert ist, als

$$I(\theta) = -\mathbb{E}_\theta \left(\frac{\partial^2}{\partial \theta^2} \log(f_\theta(X)) \right)$$

bzw. für diskrete Zufallsvariablen X definiert ist, als

$$I(\theta) = -\mathbb{E}_\theta \left(\frac{\partial^2}{\partial \theta^2} \log(p_\theta(X)) \right).$$

Da die Log-Likelihood wie folgt definiert ist,

$$\ln L(\vec{X}; \theta) = \sum_{i=1}^n \ln(p_\theta(x))$$

folgt außerdem:

$$\begin{aligned}
 I_n(\theta) &= -\mathbb{E}_\theta \left(\frac{\partial^2}{\partial \theta^2} \ln(L(\vec{X}; \theta)) \right) = -\mathbb{E}_\theta \left(\frac{\partial^2}{\partial \theta^2} \sum_{i=1}^n \ln(p_\theta(X)) \right) = \\
 I_n(\theta) &= -\sum_{i=1}^n \mathbb{E}_\theta \left(\frac{\partial^2}{\partial \theta^2} \ln(p_\theta(X)) \right) = -n \cdot \underbrace{\mathbb{E}_\theta \left(\frac{\partial^2}{\partial \theta^2} \ln(p_\theta(X)) \right)}_{I(\theta)} = -n \cdot I(\theta)
 \end{aligned}$$

Wenn wir einen erwartungstreuen Schätzer finden können, dessen Varianz mit der Cramér-Rao Schranke übereinstimmt, dann können wir sicher sein, dass er effizient ist.

Allerdings gibt es einen solchen Schätzer nicht immer, die Dichte bzw. Wahrscheinlichkeitsfunktion muss dazu von einer bestimmten Form sein. Wenn es einen solchen Schätzer gibt, dann stimmt er mit dem Maximum-Likelihood Schätzer überein (überhaupt ist der ML-Schätzer unter gewissen Regularitätsvoraussetzungen asymptotisch normalverteilt mit Mittel θ und Varianz $1/I_n(\theta)$, also gewissermaßen “asymptotisch effizient”).

Beispiel 4.11 Gehen wir wieder zurück zur Obstverkäuferin von Beispiel 4.2

Hier hatten wir das Gewicht Y der Melonen Normalverteilt. ($Y = \mathcal{N}(\mu, \sigma^2)$) Gehen wir davon aus, dass wir das σ^2 bereits kennen und $\hat{\mu}_n$ mit $\hat{\mu}_n = \bar{X}_n$ abschätzen.

Wir sollen nun herausfinden, ob der Schätzer effizient ist: (Siehe Normalverteilung, Dichtefunktion 2.11.5)

Lt. Definition 4.3 ist das Stichprobenmittel ein erwartungstreuer Schätzer, daher können wir den Cramér-Rao-Satz anwenden.

$$f(x, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow$$

$$\ln f(x, \mu) = \ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) = \ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) + \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \underbrace{\ln(e)}_{=1}$$

Bildung der 2 Ableitungen:

$$\frac{\partial \ln f(x, \mu)}{\partial \mu} = -\frac{2(x-\mu)(-1)}{2\sigma^2} = \frac{x-\mu}{\sigma^2}$$

$$\frac{\partial^2 \ln f(x, \mu)}{\partial \mu^2} = -\frac{1}{\sigma^2} \Rightarrow I(\mu) = \frac{1}{\sigma^2}$$

Folglich ist die Cramér-Rao-Schranke definiert als: $\mathbb{V}(\hat{\mu}_n) \geq \frac{\sigma^2}{n}$.

Da $\mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n}$ in Definition 4.3 definiert wurde, folgt:

$$\mathbb{V}(\hat{\mu}_n) = \mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n} \Rightarrow \hat{\mu}_n \text{ ist effizient für } \mu$$

Definition 4.13 Eine Statistik $T = T(X_1, \dots, X_n)$ heißt *suffizient*, wenn die bedingte Verteilung von (X_1, \dots, X_n) unter T nicht von θ abhängt.

Wenn die Likelihood-Funktion in der Form

$$L(X_1, \dots, X_n; \theta) = g(T, \theta) \cdot h(X_1, \dots, X_n)$$

geschrieben werden kann, dann ist T suffizient.

Beispiel 4.12 Nehmen wir wieder das Beispiel mit den weiblichen Arbeitern 4.9: diese sind, wie bereits erwähnt, Binomialverteilt (siehe Kapitel 2.10.1), d.h. Die Wahrscheinlichkeit k weibliche Arbeiter bei einer Stichprobengröße von n zu erhalten beträgt (wie ebenfalls bereits erwähnt):

$$p_{\theta}(X = k) = \binom{n}{k} \cdot \theta^k (1-\theta)^{n-k} \Rightarrow L(\vec{X}; \theta) = \theta^k (1-\theta)^{n-k} \dots (p_{\theta} \text{ ohne konst. Faktoren})$$

Da die Zufallsvariablen X_1, \dots, X_n nur 1 (weiblich), oder 0 (männlich) sein können, folgt:

$$L(\vec{X}; \theta) = \theta^{\sum_{i=1}^n X_i} (1-\theta)^{n - \sum_{i=1}^n X_i}$$

Setzt man nun $T = T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$, folgt:

$$L(\vec{X}; \theta) = \underbrace{\theta^T (1-\theta)^{n-T}}_{g(T, \theta)}$$

It. definition sollte die funktion $T(X_1, \dots, X_n)$ ja eig. auch eine h -Funktion besitzen, ist dieser hier 1? man weiß es nicht...? Daher ist $T = T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ suffizient für θ .

Beispiel 4.13 Unsere Stichprobe $\vec{\omega} = X_1, \dots, X_n$ ist Poissonverteilt, existiert eine Wahrscheinlichkeitsfunktion, die in $h(X_1, \dots, X_n)$ und $g(T, \theta)$ zerlegt werden kann, so dass $T(X_1, \dots, X_n)$ eine suffiziente Statistik bildet?

$$p_{\theta}(x_i) = \frac{\theta^{x_i} \cdot e^{-\theta}}{x_i!} \Rightarrow L(\vec{\omega}; \theta) = \prod_{i=1}^n p_{\theta}(x_i) = \prod_{i=1}^n \frac{\theta^{x_i} \cdot e^{-\theta}}{x_i!} = \underbrace{\prod_{i=1}^n \frac{1}{x_i!}}_{h(X_1, \dots, X_n)} \cdot \underbrace{\theta^{\sum_{i=1}^n x_i} \cdot e^{-\theta}}_{g(\sum_{i=1}^n x_i, \theta)}$$

Somit ist $T = \sum_{i=1}^n x_i$ suffizient.

Satz 4.2 wenn die Dichten-/Wahrscheinlichkeitsfunktionen "brav" sind, dann ist Maximum-Likelihood-Schätzer konsistent. (Wenn noch etwas "braver", dann gilt:

$$\hat{\theta}_n \approx \mathcal{N}\left(\theta, \frac{1}{nI(\theta)}\right)$$

d.h. für große n ist der ML-Schätzer approximativ normalverteilt.

Für wichtige Verteilungen kann man mit den genannten Kriterien gute/optimale Schätzer generieren, allerdings passen sie nur für diese Verteilungsmodelle, auf die sie zugeschnitten sind. In der Praxis sind die Modelle aber immer etwas anders. Folglich nimmt man oft weniger effiziente Schätzer, aber dafür sind die auch nicht so anfällig bei anderen Modellen. Sie werden als **robuste Schätzer** bezeichnet.

4.2.2. Intervallschätzung

Die Punktschätzer sind präzise, aber auch meist falsch. Wenn z.B.: der Schätzer $\hat{\theta}_n$ eine stetige Zufallsvariable ist, so ist $P(\hat{\theta}_n = \theta) = 0$. D.h. man verliert jede Sicherheit.

Manchmal möchte man aber auch Angaben über die Genauigkeit eines Schätzwertes machen können, also ein Intervall bestimmen, in dem der gesuchte Parameter liegt. Leider

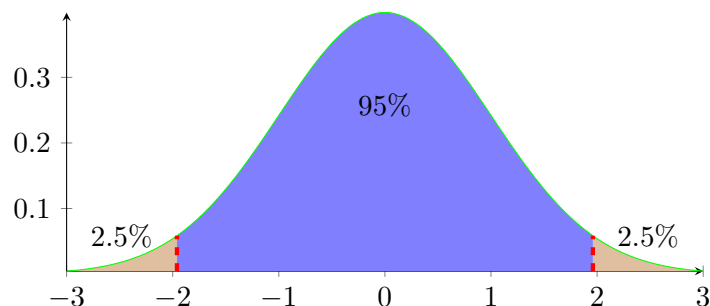


Abbildung 4.4.: Zu Beispiel 4.14: der blau eingefärbte Bereich hat die Wahrscheinlichkeit 95% (95% Prognoseintervall), die braun eingefärbten Bereiche besitzen zusammen 5%.

kann so etwas nicht mit absoluter Sicherheit geschehen, weil in den meisten Fällen für jeden Wert des Parameters jede Stichprobe positive (wenn auch sehr kleine) Wahrscheinlichkeit hat, gezogen zu werden.

Wir werden uns also damit begnügen müssen, ein Intervall zu bestimmen, das den gesuchten Parameter mit einer gewissen Wahrscheinlichkeit enthält.

Doch bevor wir zu den Konfidenzintervallen kommen, müssen wir noch die Quantile der Normalverteilung definieren:

Definition 4.14 *Quantil (Standard-)Normalverteilung*

Ist $X \sim \mathcal{N}(\mu, \sigma^2)$ (Normalverteilt siehe Kapitel 2.11.5), bzw. $X^* \sim \mathcal{N}(0, 1)$ (Standardnormalverteilt), dann sind die α -Quantile τ_α bzw. $\tau_\alpha^* = z_\alpha$ definiert durch:

$$\mathbb{P}(X \leq \tau_\alpha) = \mathbb{P}(X^* \leq \tau_\alpha^*) = \frac{\alpha}{2}.$$

Wobei $\tau_\alpha = \mu + \sigma\tau_\alpha^*$.

Die Quantile der Standardnormalverteilung werden auch noch als z_α bezeichnet.

Da die Dichte symmetrisch ist, gilt ebenso: $\mathbb{P}(X \leq \tau_{1-\alpha}) = \mathbb{P}(X^* \leq \tau_{1-\alpha}^*) = \frac{\alpha}{2}$

Weiters gilt:

$$\tau_{1-\alpha}^* = -\tau_\alpha^*$$

und

$$\mathbb{P}(|X^*| < \tau_{1-\alpha}^*) = \mathbb{P}(\tau_\alpha^* < X^* < \tau_{1-\alpha}^*) = 1 - 2\frac{\alpha}{2} = 1 - \alpha$$

Beispiel 4.14 Die Zufallsvariable X ist Standardnormalverteilt, wie hoch ist die Wahrscheinlichkeit, in welchem Bereich liegen 95% ($\Rightarrow \alpha = 0.05$, denn $\alpha + 0.95 = 1$) aller möglichen Werte für X ?

Das gesuchte Quantil findet man im Anhang A.1, dort findet man für $1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975$ den Wert $\tau_{1-\alpha}^* = 1.96$.

Somit befindet sich mit 95%-iger Wahrscheinlichkeit eine Realisation der zufälligen Variablen Y in dem Intervall $-1.96 \leq x \leq 1.96$.

Siehe Abbildung 4.4.

Definition 4.15 Konfidenzintervall $a = a(X_1, \dots, X_n) \leq b = b(X_1, \dots, X_n)$ seien zwei Statistiken. Das Intervall $[a, b]$ heißt **Konfidenzintervall** für θ mit **Überdeckungswahrscheinlichkeit** $\gamma = 1 - \alpha$ (meist setzt man $\gamma = 0.95$, oder $\gamma = 0.99$), wenn

$$\mathbb{P}_\theta(a \leq \theta \leq b) \geq \gamma.$$

D.h. der Parameter θ liegt mit Wahrscheinlichkeit von mind. γ in $[a, b]$.

Wenn in dieser Ungleichung Gleichheit gilt, sprechen wir von einem exakten Konfidenzintervall.

ACHTUNG: Vor einer Beobachtung macht man eine Wahrscheinlichkeits-Aussage, danach eine Behauptung. Diese Behauptung kann nur noch wahr sein, wenn Sie im Annahmehereich liegt, oder falsch, wenn sie außerhalb liegt. Sie besitzt keine Wahrscheinlichkeit mehr, das heißt, sie ist nicht mit Wahrscheinlichkeit von beispielsweise 95% wahr.

Ein möglicher Ausgangspunkt für die Konstruktion von Konfidenzintervallen ist ein Schätzer für θ . Unter sehr günstigen Bedingungen kann die Verteilung dieses Schätzers exakt bestimmt werden, in anderen Fällen (immer noch günstig) ist er zumindest asymptotisch normalverteilt. In diesem Fall können wir ein approximatives Konfidenzintervall in der Form

$$\left[\hat{\theta}_n - z_{\frac{1+\gamma}{2}} \sigma_n, \hat{\theta}_n + z_{\frac{1+\gamma}{2}} \sigma_n \right],$$

wobei σ_n^2 die Varianz von $\hat{\theta}_n$ ist. Diese hängt in den meisten Fällen vom unbekannten θ ab, das wir durch $\hat{\theta}_n$ ersetzen.

Für den Spezialfall einer Normalverteilung können wir exakte Konfidenzintervalle angeben:

Konfidenzintervall für μ einer Normalverteilung mit bekannter Varianz σ^2 :

mit **Ansatz:** $a = \bar{X}_n - c$ und $b = \bar{X}_n + c$ (d.h. $[\bar{X}_n - c, \bar{X}_n + c]$) und Satz B.3 folgt:

$$\gamma = \mathbb{P}_\mu(\bar{X}_n - c \leq \mu \leq \bar{X}_n + c) = \mathbb{P}_\mu(\underbrace{-c \leq \mu - \bar{X}_n \leq +c}_{+c \geq -\mu + \bar{X}_n \geq -c}) = \mathbb{P}_\mu(-c \leq \bar{X}_n - \mu \leq +c)$$

dividiert man nun die gesamte Ungleichung durch $\sqrt{\frac{\sigma^2}{n}}$ bekommt man: (wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist)

$$\begin{aligned} \gamma &= \mathbb{P}_\mu \left(\frac{-c}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{\overbrace{\bar{X}_n - \mu}^{\sim \mathcal{N}(0,1)}}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{c}{\sqrt{\frac{\sigma^2}{n}}} \right) = \Phi \left(\frac{c}{\sqrt{\frac{\sigma^2}{n}}} \right) - \overbrace{\Phi \left(-\frac{c}{\sqrt{\frac{\sigma^2}{n}}} \right)}^{\Phi(-x)=1-\Phi(+x)} = 2\Phi \left(\frac{c}{\sqrt{\frac{\sigma^2}{n}}} \right) - 1 \\ \frac{\gamma+1}{2} &= \Phi \left(\frac{c}{\sqrt{\frac{\sigma^2}{n}}} \right) \Rightarrow \frac{c}{\sqrt{\frac{\sigma^2}{n}}} = \Phi^{-1} \left(\frac{\gamma+1}{2} \right) = z_{\frac{1+\gamma}{2}} \end{aligned}$$

Wobei $z_{\frac{1+\gamma}{2}}$ das $\frac{1+\gamma}{2}$ -Quantil der Standardnormalverteilung ist.

$$c = z_{\frac{1+\gamma}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}$$

Definition 4.16 Das Konfidenzintervall für μ bei bekanntem σ^2 :

$$KI = \left[\bar{X}_n - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\sigma^2}{n}} \right],$$

Konfidenzintervall für μ einer Normalverteilung mit unbekannter Varianz σ^2 :

hier gibt es 2 Möglichkeiten:

- i) Schätzen von $\hat{\sigma}^2 = s_n^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ und in der KI -Formel von Definition 4.16 einfach σ^2 durch s_n^2 ersetzen. das ist aber nur eine Approximation.
- ii) Ein genaueres Ergebnis bekommt man mit folgendem **Ansatz**:

$$\left[\bar{X}_n - c \sqrt{\frac{s_n^2}{n}}, \bar{X}_n + c \sqrt{\frac{s_n^2}{n}} \right]$$

$$\gamma = \mathbb{P} \left(\bar{X}_n - c \sqrt{\frac{S_n^2}{n}} \leq \mu \leq \bar{X}_n + c \sqrt{\frac{S_n^2}{n}} \right) = \mathbb{P} \left(-c \leq \underbrace{\frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}}}_{=T} \leq c \right)$$

Gesucht ist nun die Verteilung, von $T = \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}}$

Definition 4.17 Students-T-Verteilung

Wenn $U \sim \mathcal{N}(0, 1)$ und $V \sim \chi_n^2$ verteilt sind, dann heißt die Verteilung von

$$T_n = \frac{U}{\sqrt{V/n}}$$

Students-T-Verteilung mit n -Freiheitsgraden.

Wobei t_n die Dichte

$$f_n(x) = C_n \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{n}}}$$

beitzt.

Somit folgt:

$$\gamma = \mathbb{P}(-c \leq \underbrace{T}_{\sim t_{n-1}} \leq c) = \mathbb{P}(|t_{n-1}| \leq c) = \mathbb{P}(t_{n-1} \leq c) - \mathbb{P}(-(t_{n-1} \leq c)) = 2\mathbb{P}(t_{n-1} \leq c) - 1$$

$$\mathbb{P}(t_{n-1} \leq c) = \frac{1+\gamma}{2} \Rightarrow c = \mathbb{P}_{t_{n-1}}^{-1}\left(\frac{1+\gamma}{2}\right) = t_{n-1; \frac{1+\gamma}{2}}$$

Definition 4.18 Das Konfidenzintervall für μ bei unbekanntem σ^2 :

$$KI = \left[\bar{X}_n - t_{n-1; \frac{1+\gamma}{2}} \sqrt{\frac{S_n^2}{n}}, \bar{X}_n + t_{n-1; \frac{1+\gamma}{2}} \sqrt{\frac{S_n^2}{n}} \right],$$

Beispiel 4.15 Gehen wir wieder auf das Beispiel der Obstverkäuferin zurück: Die Verkäuferin verkauft eine Palette Melonen mit $n = 51$ Stück. Wir finden aber, dass das Gewicht viel aussagekräftiger wäre. Auf Nachfrage erhalten wir die Information, dass ihre Melonen im Mittel $\mu = 1\text{kg}$ wiegen.

Da wir aber ein bisschen in WSP aufgepasst haben, zweifeln wir diese Aussage an und wollen nun überprüfen, ob diese Aussage stimmt:

Wir kaufen eine Palette, wiegen jede Melone einzeln und erhalten als (empirischen) Mittelwert $\bar{X} = 0.89$ und für die korrigierte Stichprobenvarianz erhalten wir $S_n = 0.26$. Hat uns die Verkäuferin nun belogen?

Das Konfidenzintervall für den wahren Mittelwert μ zum Irrtumsniveau 5% lautet:

$$\begin{aligned} KI &= \left[\bar{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \right] \\ KI &= \left[0.89 - t_{51-1, 1-\frac{5\%}{2}} \frac{0.26}{\sqrt{51}}, 0.89 + t_{51-1, 1-\frac{5\%}{2}} \frac{0.26}{\sqrt{51}} \right] \\ KI &= [0.89 - t_{50, 0.975} 0.0364, 0.89 + t_{50, 0.975} 0.0364] \\ KI &= [0.89 - 2.01 \cdot 0.0364, 0.89 + 2.01 \cdot 0.0364] \\ KI &= [0.89 - 0.07317, 0.89 + 0.07317] = [0.81683, 0.96317] \end{aligned}$$

Die korrigierte Stichprobenvarianz (siehe Definition 2.11) ist definiert als:

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

Somit hat sie uns mit großer Wahrscheinlichkeit (wissentlich oder unwissentlich) belogen.

Konfidenzintervall für σ^2 einer Normalverteilung:

mit **Ansatz:** $[a \cdot S_n^3, b \cdot S_n^2]$ und den Rechenregeln der Ungleichungen (siehe Anhang) folgt:

$$\gamma = \mathbb{P}(a \cdot S_n^3 \leq \sigma^2 \leq b \cdot S_n^2) = \mathbb{P}\left(\frac{1}{b \cdot S_n^2} \leq \frac{1}{\sigma^2} \leq \frac{1}{a \cdot S_n^3}\right) = \mathbb{P}\left(\frac{1}{b} \leq \frac{S_n^2}{\sigma^2} \leq \frac{1}{a}\right)$$

Anschließend wird die Ungleichung mit $n-1$ Multipliziert:

$$\gamma = \mathbb{P}\left(\frac{n-1}{b} \leq (n-1) \cdot \frac{S_n^2}{\sigma^2} \leq \frac{n-1}{a}\right) = \mathbb{P}\left(\chi_{n-1}^2 \leq \frac{n-1}{a}\right) - \mathbb{P}\left(\chi_{n-1}^2 \leq \frac{n-1}{b}\right) \Rightarrow b = \frac{n-1}{\chi_{n-1; \frac{1-\gamma}{2}}^2}$$

Nun müssen wir a und b so wählen, dass $\mathbb{P}(\chi_{n-1}^2 \geq \frac{n-1}{a}) = \frac{1-\gamma}{2} = \mathbb{P}(\chi_{n-1}^2 \leq \frac{n-1}{b})$

$$\Rightarrow \frac{n-1}{b} = \chi_{n-1; \frac{1-\gamma}{2}}^2$$

Nun steht aber nicht $\mathbb{P}(\chi_{n-1}^2 \leq \frac{n-1}{a}) = \frac{1-\gamma}{2}$, sondern $\mathbb{P}(\chi_{n-1}^2 \geq \frac{n-1}{a}) = \frac{1-\gamma}{2}$. Daher müssen wir die Wahrscheinlichkeit mit a noch anpassen:

$$\mathbb{P}(\chi_{n-1}^2 \leq \frac{n-1}{a}) = \frac{1+\gamma}{2} \Rightarrow \frac{n-1}{a} = \chi_{n-1; \frac{1+\gamma}{2}}^2 \Rightarrow a = \frac{n-1}{\chi_{n-1; \frac{1+\gamma}{2}}^2}$$

Definition 4.19 Das Konfidenzintervall für σ^2 :

$$KI = \left[\frac{(n-1)S_n^2}{\chi_{n-1; \frac{1+\gamma}{2}}^2}, \frac{(n-1)S_n^2}{\chi_{n-1; \frac{1-\gamma}{2}}^2} \right].$$

approximatives Konfidenzintervall für Anteilswerte

Der Anteil \hat{p} ist entweder $\hat{p} = \frac{N}{n}$, wobei N die Anzahl der günstigen Fälle angibt, oder $\hat{p} = \bar{X}_n$, wenn $X_i = \begin{cases} 1 : \text{günstig} \\ 0 : \text{sonst} \end{cases}$ gilt.

Somit ist $\hat{p} \sim \mathcal{N}(p, \frac{p(1-p)}{n})$. Mit dem Ansatz $[\hat{p} - c; \hat{p} + c]$ kommt man anschließend auf

Definition 4.20 Das approximative Konfidenzintervall für Anteilswerte:

$$KI = \left[\hat{p} - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right].$$

Ein anderer Weg, um Konfidenzintervalle zu erhalten, führt über die Theorie der statistischen Tests.

4.3. Tests

Beispiel 4.16 Für ein Fest haben wir 50 Flaschen Bier mit einer Nennfüllmenge von 500ml, beim Leeren der Flaschen hat sich aber $\bar{X}_{50} = 490$ herausgestellt. ($S_{50} = 800$)

Frage: Ist das ein genügender Beweis, dass $\mu < 500$?

Dazu nehmen wir an, dass $X \sim \mathcal{N}(\mu, \sigma^2)$. Wie hoch ist nun die Wahrscheinlichkeit, dass so etwas passiert, wenn $\mu = 500$:

$$\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) = \mathcal{N}\left(\mu, \frac{S_n^2}{n}\right)$$

$$\mathbb{P}(\bar{X}_n \leq 490) = \Phi\left(\frac{490 - 500}{\sqrt{800/50}}\right) = \Phi(-2.5) \approx 0.01$$

4.3.1. Grundlagen

Bei dieser Art von Problemen geht es nicht mehr darum, einen Näherungswert für einen unbekannten Parameter zu bestimmen, sondern es soll eine Aussage über den Parameter überprüft werden, etwa “der Ausschussanteil ist kleiner als 1%” oder “das mittlere Gewicht ist 1kg”.

Wir definieren zuerst

Definition 4.21 Eine **Hypothese** H ist eine Teilmenge des Parameterraums Θ . Sie ist somit eine Menge von Verteilungen, bzw. Parametern.

Wir schreiben Hypothesen meistens nicht in Mengennotation, sondern als Aussage (meistens eine Gleichung oder Ungleichung) für den Parameter. Wir unterscheiden:

einfache Hypothese $H = \{P\}$ bzw. $H = \{\theta\}$ enthält nur einen Parameterwert d.h. $|H| = 1$ (Mächtigkeit=1). Für den Begriff Mächtigkeit, siehe Anhang B.3.

zusammengesetzte Hypothese : $|H| > 1$

und im parametrischen Fall:

einseitige Hypothese $H = [\theta_0, \infty)$ oder $H = (-\infty, \theta_0]$, mit Schreibweise $H : [\theta \leq \theta_0]$ oder $H : [\theta \geq \theta_0]$ etc.) hier wird die Richtung eines Unterschieds/Zusammenhang angegeben.

z.B: Männer sind größer als Frauen.

zweiseitige Hypothese $H : [\theta \neq \theta_0]$ mit Schreibweise: $H : \theta < \theta_0$ oder $H : \theta > \theta_0$ hier wird nur behauptet, dass ein Unterschied besteht, aber nicht, in welche Richtung der Unterschied geht.

z.B: Männer und Frauen sind nicht gleich groß.

Definition 4.22 Ein Test wird als Entscheidung zwischen zwei Hypothesen formuliert:

- i) der **Nullhypothese** H_0 und
- ii) der **Gegenhypothese** oder **Alternative** H_1 .

Wobei gelten muss, dass $H_0 \cap H_1 = \emptyset$

Die Rollen der beiden Hypothesen sind nicht symmetrisch — der übliche Sprachgebrauch ist “die Nullhypothese wird angenommen” oder “die Nullhypothese wird verworfen”.

Die Nullhypothese H_0 ist die Präzisierung der angezweifelte Aussage, über deren Richtigkeit eine Entscheidung zu fällen ist. Die Alternativhypothese H_1 sagt, was gilt, wenn H_0 falsch ist.

Beispiel 4.17 In unserer Firma ist der Kohlenmonoxid-Anteil der Luft zu hoch (behauptet zumindest der Betriebsrat), dieser will nun beweisen, dass dieser wirklich viel zu hoch ist (über 30ppm), daher werden die Hypothesen lauten:

$$H_0: \mu \leq 30 \text{ ppm vs. } H_1: \mu > 30 \text{ ppm}$$

Der Chef der Firma hingegen ist überzeugt, dass die Grenzwerte eingehalten werden, daher werden seine Hypothesen lauten:

$$H_0: \mu \geq 30 \text{ ppm vs. } H_1: \mu < 30 \text{ ppm}$$

ACHTUNG: wenn immer möglich, sollte man das, was zu beweisen ist, als Gegenhypothese formulieren.

Ein Test kann auf mehrere Arten angegeben werden:

- i) durch die Menge der möglichen Stichprobenwerte, bei denen die Nullhypothese angenommen wird: den **Annahmebereich** A , bzw. durch die Menge der möglichen

Entscheidung	Realität	
	H_0 wahr	H_1 wahr
H_0 Annehmen	richtig	Fehler 2.Art
H_0 Ablehnen	Fehler 1.Art	richtig

Tabelle 4.1.: Entscheidungen für Fehler

Stichprobenwerte, bei denen die Gegenhypothese angenommen wird: den **Verwerfungsbereich** V :

$$\begin{aligned} A &= (X_1, \dots, X_n) : \text{entscheide für } H_0, \text{ wenn } (X_1, \dots, X_n) = (x_1, \dots, x_n) \\ V &= A^c \end{aligned}$$

- ii) Oft ist es jedoch einfacher, eine **Teststatistik** T anzugeben, und die Nullhypothese zu verwerfen, wenn diese Statistik einen kritischen Wert t_c überschreitet (oder unterschreitet).

$$T = T(X_1, \dots, X_n) : \text{nehme } H_0 \text{ an, wenn } T \leq t_c \text{ (natürlich auch } T \geq t_c)$$

Beim Testen kann man zwei Arten von Fehlern begehen (siehe Tabelle 4.1):

- 1.) **Fehler 1.Art**(die Nullhypothese wird verworfen, obwohl sie zutrifft) und
- 2.) **Fehler 2.Art**(die Nullhypothese wird angenommen, obwohl sie nicht zutrifft).

Man möchte natürlich die Wahrscheinlichkeit für beide Fehler möglichst klein halten. Leider geht das nicht gleichzeitig — die Wahrscheinlichkeit für einen Fehler 1.Art kann (zumindest ab einem gewissen Punkt) nur Verkleinert werden, indem der Annahmebereich vergrößert wird, und dadurch wächst die Wahrscheinlichkeit für einen Fehler 2. Art. In der Statistik wird dieses Dilemma gelöst, indem man eine Schranke für die Wahrscheinlichkeit eines Fehlers 1.Art angibt:

Definition 4.23 Ein Test heißt vom **Niveau (Signifikanzniveau)** α , wenn die Wahrscheinlichkeit für einen Fehler 1.Art (die bei zusammengesetzten Hypothesen eine Funktion von θ ist) nicht größer als α ist. Das Niveau wird manchmal auch Irrtumswahrscheinlichkeit bezeichnet. Gängige Werte sind beispielsweise $\alpha = 0.05, 0.01, \dots$

$$\mathbb{P}_\theta(\text{Fehler 1.Art}) = \mathbb{P}_\theta(H_0 \text{ wird verworfen}) \leq \alpha \quad \forall \theta \in H_0$$

Für ein Beispiel, siehe Abbildung 4.5.

Der Vollständigkeit halber sei hier erwähnt, dass β einen Fehler 2. Art bezeichnet.

ACHTUNG: Ein **optimaler Test** für ein gegebenes Niveau α ist die kleinste Wahrscheinlichkeit für einen Fehler 2. Art. (Für fixes α ist β minimal)

Eine Möglichkeit, einen Test zu konstruieren, liefert uns die Likelihood-Methode: die Grundidee besteht darin, sich für die Hypothese zu entscheiden, für die die aktuelle Stichprobe die größere Wahrscheinlichkeit hat. Damit man das Niveau α einstellen kann, wird noch ein zusätzlicher Faktor eingeführt:

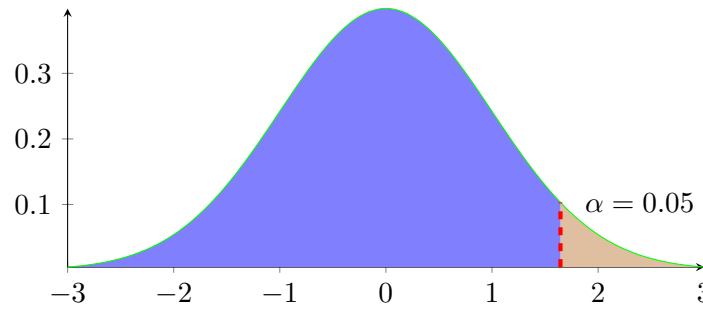


Abbildung 4.5.: Zu Definition 4.23: Bei $\alpha = 0.05$, werden alle Stichprobenergebnisse > 1.645 (braun eingefärbt) als Unwahrscheinlich bezeichnet.

Definition 4.24 Die **Likelihoodquotientenstatistik** für einfache Hypothesen: $H_0 : \{\theta_0\}$ und $H_1 : \{\theta_1\}$, außerdem müssen wir λ_c (Kritischer Wert) so wählen, dass der Test das Niveau α hat.

$$\ell(X_1, \dots, X_n) = \frac{L(X_1, \dots, X_n, \theta)}{L(X_1, \dots, X_n, \theta)}$$

H_0 annehmen, wenn $\ell(X_1, \dots, X_n) \geq \lambda_c$
 H_0 verwerfen, wenn $\ell(X_1, \dots, X_n) < \lambda_c$.

Bei der Modifikation **Randomisierter Likelihoodquotiententest** werden die Regeln, wann verworfen und wann angenommen wird etwas verändert:

H_0 annehmen, wenn $\ell(X_1, \dots, X_n) > \lambda_c$
 H_0 verwerfen, mit Wahrscheinlichkeit p , wenn $\ell(X_1, \dots, X_n) = \lambda_c$
 H_0 verwerfen, wenn $\ell(X_1, \dots, X_n) < \lambda_c$.

Für den randomisierten Likelihoodquotiententest gilt somit:

$$\mathbb{P}_{\theta_0}(\text{Fehler 1. Art}) = \mathbb{P}_{\theta_0}(\lambda < \lambda_c) + p \cdot \mathbb{P}_{\theta_0}(\lambda = \lambda_c)$$

Satz 4.3 Für einfache Hypothesen ist der randomisierte Likelihoodquotiententest optimal.

Definition 4.25 Die **Likelihoodquotientenstatistik** für zusammengesetzte Hypothesen ist

$$\ell(X_1, \dots, X_n) = \frac{\sup_{\theta \in H_0} L(X_1, \dots, X_n, \theta)}{\sup_{\theta \in H_1} L(X_1, \dots, X_n, \theta)}$$

bzw.

$$\ell(X_1, \dots, X_n) = \frac{\sup_{\theta \in H_0} L(X_1, \dots, X_n, \theta)}{\sup_{\theta \in \Theta} L(X_1, \dots, X_n, \theta)}$$

(die zweite Form ist oft einfacher zu berechnen, und für große Stichprobenumfänge sind die Tests identisch).

Der **Likelihoodquotiententest** verwirft die Nullhypothese, wenn $\ell(X_1, \dots, X_n)$ kleiner ist als ein kritischer Wert.

In einem speziellen Fall ist der Likelihoodquotiententest optimal:

Satz 4.4 (Neyman-Pearson) Falls sowohl H_0 als auch H_1 einfach ist, dann ist der Likelihoodquotiententest optimal, d.h., er hat unter allen Tests mit demselben Niveau die minimale Wahrscheinlichkeit für einen Fehler zweiter Art (in diesem Fall wird der Likelihoodquotient einfach als $L(X_1, \dots, X_n, \theta_0)/L(X_1, \dots, X_n, \theta_1)$ berechnet)

4.3.2. Spezielle Tests

Für μ , wenn σ^2 bekannt (Normalverteilt):

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{\sigma^2/n}}$$

- (I) $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$: verwerfen H_0 , wenn $T > z_{1-\alpha}$.
- (II) $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$: verwerfen H_0 , wenn $T < z_\alpha = -z_{1-\alpha}$.
- (III) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$: verwerfen H_0 , wenn $|T| > z_{1-\frac{\alpha}{2}}$.

Für μ , wenn σ^2 unbekannt (Normalverteilt):

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2/n}}$$

wobei hier das S_n^2 die korrigierte Stichprobenvarianz (siehe Definition 2.11) ist:

- (I) $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$: verwerfen H_0 , wenn $T > t_{n-1;1-\alpha}$.
- (II) $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$: verwerfen H_0 , wenn $T < -t_{n-1;1-\alpha}$.
- (III) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$: verwerfen H_0 , wenn $|T| > t_{n-1;1-\frac{\alpha}{2}}$.

Beispiel 4.18 Berechnen wir nun das Beispiel mit der Obstverkäuferin 4.15, die uns belogen hat. Nachdem wir uns die Palette Melonen gekauft haben, behauptet einen anderer Obstverkäufer, dass die Melonen unserer Obstverkäuferin im Mittel nur $\frac{3}{4}$ kg schwer sind. Ein anderer Kunde behauptet allerdings, dass sie im Mittel ca. 0.9 kg schwer sind.

Somit haben wir 3 Werte, die es zu testen gilt: $\mu_0 = 1$, $\mu_0 = 0.9$ und $\mu_0 = 0.75$. Welche der Werte können wir verwerfen und welche annehmen?

Wir holen uns den Wert von $t_{n-1;1-\alpha}$ aus der Tabelle:

$$t_{n-1;1-\alpha} = t_{50;0.975} = 2.01$$

Wir behaupten jetzt, dass der wahre Mittelwert μ der Melonen der Obstverkäuferin kleiner als μ_0 ist:

$$H_0 : \mu \leq \mu_0 \quad H_1 : \mu > \mu_0$$

$$\mu_0 = 1 : T = \frac{|\bar{X}_n - \mu_0|}{\sqrt{S_n^2/n}} = \frac{|0.89 - 1|}{\sqrt{0.26^2/51}} = \frac{|-0.11|}{0.0364} = 3.022 > 2.01 \Rightarrow \text{verwerfen}$$

$$\mu_0 = 0.9 : T = \frac{|\bar{X}_n - \mu_0|}{\sqrt{S_n^2/n}} = \frac{|0.89 - 0.9|}{\sqrt{0.26^2/51}} = \frac{|-0.01|}{0.0364} = 0.27457 < 2.01 \Rightarrow \text{anneh.}$$

$$\mu_0 = 0.75 : T = \frac{|\bar{X}_n - \mu_0|}{\sqrt{S_n^2/n}} = \frac{|0.89 - 0.75|}{\sqrt{0.26^2/51}} = \frac{|0.14|}{0.0364} = 3.84615 < 2.01 \Rightarrow \text{verwerf.}$$

Für die Varianz einer Normalverteilung:

$$T = \frac{(n-1)S_n^2}{\sigma_0^2}$$

- (I) $H_0 : \sigma^2 \geq \sigma_0^2$ gegen $H_1 : \sigma^2 < \sigma_0^2$: verwerfen H_0 , wenn $T < \chi_{n-1;\alpha}^2$.
- (II) $H_0 : \sigma^2 = \sigma_0^2$ gegen $H_1 : \sigma^2 \neq \sigma_0^2$: verwerfen H_0 , wenn $T > \chi_{n-1;1-\frac{\alpha}{2}}^2$ oder $T < \chi_{n-1;\frac{\alpha}{2}}^2$.
- (III) $H_0 : \sigma^2 \leq \sigma_0^2$ gegen $H_1 : \sigma^2 > \sigma_0^2$: verwerfen H_0 , wenn $T > \chi_{n-1;1-\alpha}^2$.

Für Anteilswerte

$$T = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

- (I) $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$: verwerfen H_0 , wenn $T > z_{1-\alpha}$.
- (II) $H_0 : p \geq p_0$ gegen $H_1 : p < p_0$: verwerfen H_0 , wenn $T < -z_{1-\alpha}$.
- (III) $H_0 : p = p_0$ gegen $H_1 : p \neq p_0$: verwerfen H_0 , wenn $|T| > z_{1-\frac{\alpha}{2}}$.

4.3.3. Anpassungstests

Ein Anpassungstest ist ein Hypothesentest, der die unbekannte Wahrscheinlichkeitsverteilung einer Zufallsvariablen auf (annäherndes) Folgen eines Verteilungsmodells prüfen soll.

Beispiel 4.19 Man nimmt einen k -Seitigen Würfel, mit diesem wird n -Mal (Stichprobe (X_1, \dots, X_n)) gewürfelt, wobei jeweils diskrete Werte zwischen $1, \dots, k$ angenommen werden können. Anschließend wird überprüft, ob es sich um einen fairen Würfel mit $p_1, \dots, p_k = \frac{1}{k}$ handelt.

Die Frage, die wir hier somit untersuchen, ist, ob eine Stichprobe aus einer gegebenen Verteilung stammt. Das geht am einfachsten, wenn es sich bei der hypothetischen Verteilung um eine diskrete Verteilung mit endlich vielen Werten $1, \dots, k$ handelt. Wir testen also

$$H_0 : X \sim P = (p_1, \dots, p_k)$$

gegen die Alternative

$$H_1 : X \not\sim P$$

Diese Frage kann man mit der Likelihoodquotientenmethode behandeln, oder man nimmt als Näherung für den LQ-Test den

1	2	3	4	5	6
20	16	23	24	18	19

Tabelle 4.2.: Ergebnis nach $n = 120$ Würfeln (siehe Beispiel 4.20)**Chi-Quadrat-Anpassungstest**

Wir führen die Häufigkeiten Y_1, \dots, Y_k (die Häufigkeiten von $1, \dots, k$ in der Stichprobe) ein:

$$Y_i = \#\{X_j = i \mid \forall j \leq n\}$$

Für großes n ist $Y_i \approx np_i$ (und approximativ normalverteilt), wenn H_0 zutrifft. Wir wollen alle Differenzen zwischen Y_i und np_i gleichzeitig überprüfen. Dazu bilden wir eine gewichtete Quadratsumme:

$$T = \sum_{i=1}^k \frac{(Y_i - np_i)^2}{np_i}.$$

Diese Statistik ist asymptotisch χ^2 -verteilt mit $k - 1$ Freiheitsgraden. Die Nullhypothese wird abgelehnt, wenn

$$T > \chi_{k-1; 1-\alpha}^2.$$

Da diese Aussagen nur asymptotisch gelten, muss n hinreichend groß sein.

ACHTUNG: Die übliche Faustregel ist, dass $np_i \geq 5$ sein soll.

Beispiel 4.20 Wir nehmen nochmal das Beispiel 4.19 und verwenden nun konkret einen Würfel mit 6 Seiten ($k = 6$), außerdem würfeln wir $n = 120$ -Mal mit den Ergebnissen in Tabelle 4.2.

Als erstes, checken wir, ob die Faustregel erfüllt ist: $n \cdot p_i > 5$:

$$n \cdot p_i = \frac{120}{6} = 20 > 5$$

Somit ist die Faustregel erfüllt.

Das Ergebnis für $T = \sum_{j=1}^k \frac{(y_i - np_i)^2}{np_i}$ aus den Würfeln lautet somit: $T = 2.3$ (Berechnung siehe Tabelle 4.3). Das Ergebnis für T muss nun mit $\chi_{5;0.95}^2$ verglichen werden:

$$T = 2.3 < \chi_{5;0.95}^2 = 11.07$$

Somit wird die Nullhypothese nicht verworfen. Es handelt sich somit um einen fairen Würfel, für den gilt: $p_1, \dots, p_6 = \frac{1}{6}$, also alle Augenzahlen sind gleich Wahrscheinlich.

Es kann bei diesem Verfahren jedoch zu Komplikationen kommen:

- i) Wenn $n \cdot p_i < 5$:
- ii) Wenn es unendlich viele Werte gibt:
- iii) Wenn es sich um eine stetige Verteilung handelt (überabzählbar).

Abhilfe in diesen Fällen schafft eine Klasseneinteilung.

Bei stetigen Verteilungen kann man die Klassengrenzen so setzen, dass alle Klassen gleiche Wahrscheinlichkeit haben, was die Rechnung vereinfachen kann.

i	y_i	np_i	$\frac{(y_i - np_i)^2}{np_i}$
1	20	$\frac{120}{6} = 20$	$\frac{(20-20)^2}{20} = 0$
2	16	$\frac{120}{6} = 20$	$\frac{(16-20)^2}{20} = \frac{16}{20}$
3	23	$\frac{120}{6} = 20$	$\frac{(23-20)^2}{20} = \frac{9}{20}$
4	24	$\frac{120}{6} = 20$	$\frac{(24-20)^2}{20} = \frac{16}{20}$
5	18	$\frac{120}{6} = 20$	$\frac{(18-20)^2}{20} = \frac{4}{20}$
6	19	$\frac{120}{6} = 20$	$\frac{(19-20)^2}{20} = \frac{1}{20}$
		$\sum_{j=1}^k$	$\frac{16+9+16+4+1}{20} = \frac{46}{20} = \frac{23}{10} = 2.3$

Tabelle 4.3.: Berechnung von $T = \sum_{j=1}^k \frac{(y_j - np_j)^2}{np_j}$ (siehe Beispiel 4.20)

Satz 4.5 Wenn die Verteilung, auf die wir testen, nicht vollständig spezifiziert ist (einige Parameter fehlen), etwa, wenn man testen will, ob eine Normalverteilung vorliegt, von der wir Mittel und Varianz nicht kennen, dann müssen die Parameter nach der Maximum-Likelihood Methode geschätzt werden.

Mit diesen geschätzten Parametern können dann die Wahrscheinlichkeiten berechnet werden. Die Anzahl der Freiheitsgrade muss dann korrigiert werden, indem die Anzahl der geschätzten Parameter abgezogen wird, es sind also statt $k - 1$ $k - 1 - d$ Freiheitsgrade ($\chi_{k-1-d,1-\alpha}^2$ anstatt $\chi_{k-1,1-\alpha}^2$), wobei d die Anzahl der geschätzten Parameter ist.

Im Beispiel einer Normalverteilung ist $d = 2$ und somit das gesuchte χ^2 definiert als $\chi_{k-3,1-\alpha}^2$.

Beispiel 4.21 Wir haben $n = 100$ Messwerte (x_1, \dots, x_n) (Verteilungsfunktion siehe Abbildung 4.6), von denen behauptet wird, dass sie Normalverteilt wären. Jedoch gibt es keine Varianz und keinen Mittelwert.

Als erstes gilt es, die Varianz und den Mittelwert aus den Daten zu schätzen. (mit ML-Schätzer) Das haben wir aber bereits in Beispiel 4.10 gemacht, das Ergebnis war folgendes:

$$\hat{\mu} = \bar{y} \text{ und } \hat{\sigma} = \sqrt{\mathbb{V}(y)}$$

Aus der Formel $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ erhalten wir für $\bar{y} = 4.85678$.

Aus der Formel $\mathbb{V}(y) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ erhalten wir für $\mathbb{V}(y) = 2.11263$.

Anschließend sollten wir eine Klasseneinteilung (mit $k = 5$ Klassen) machen, dabei sollte aber jede Klasse die gleiche Wahrscheinlichkeit (20%) besitzen (mit 20 ist auch gleich die Faustregel erfüllt, denn $20 > 5$):

$$k_1 = [0 - 20\%], k_2 = [20 - 40\%], k_3 = [40 - 60\%], k_4 = [60 - 80\%], k_5 = [80 - 100\%]$$

Um nun die Intervalle für die Werte zu bekommen, gehen wir davon aus, dass die Werte nicht Normalverteilt, sondern Standardnormalverteilt sind: Denn so können wir uns die Werte $z_{20\%}$, $z_{40\%}$, $z_{60\%}$ und $z_{80\%}$ von der Tabelle der Standardnormalverteilungs-Quantile (Siehe Anhang A.2) heraussuchen.

Somit erhalten wir für $z_{20\%} = -0.842$, $z_{40\%} = -0.253$, $z_{60\%} = 0.253$ und $z_{80\%} = 0.842$.

Um nun die Intervalle für die Normalverteilung zu bekommen, müssen nur in die Formel $\tau_p = \mu + \sigma \cdot z_p$ die Quantile z_p eingesetzt werden (siehe Definition 4.14):

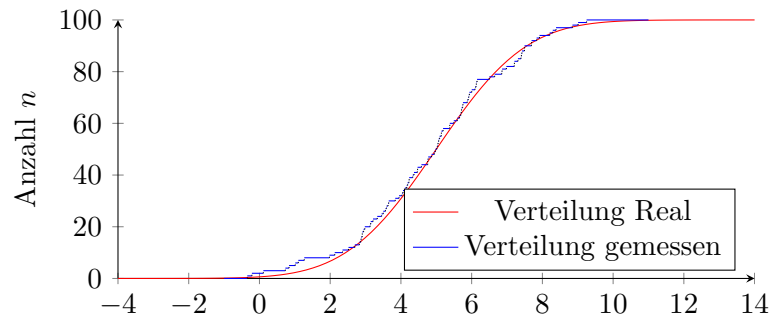


Abbildung 4.6.: Zu Beispiel 4.21: Verteilung der Messwerte.

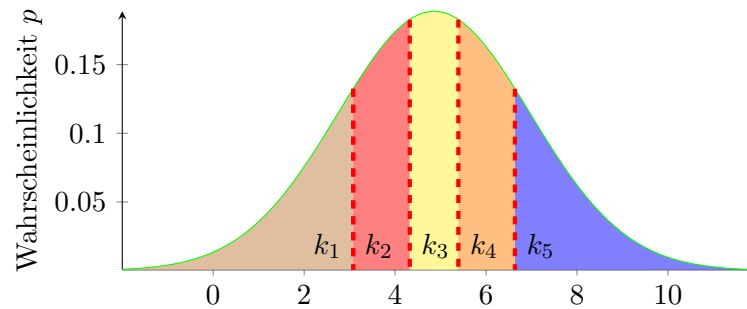


Abbildung 4.7.: Zu Beispiel 4.21: Verteilung der Klassen, jede hat 20% Wahrscheinlichkeit.

$$\tau_{20} = \mu + \sigma \cdot z_{20} = 4.85678 + 2.11263 \cdot (-0.842) = 3.07797$$

$$\tau_{40} = \mu + \sigma \cdot z_{40} = 4.85678 + 2.11263 \cdot (-0.253) = 4.32228$$

$$\tau_{60} = \mu + \sigma \cdot z_{60} = 4.85678 + 2.11263 \cdot (0.253) = 5.39128$$

$$\tau_{80} = \mu + \sigma \cdot z_{80} = 4.85678 + 2.11263 \cdot (0.842) = 6.63559$$

Somit lauten die Klassenintervalle: (Siehe Abbildung 4.7)

$$\underbrace{(-\infty, 3.07797]}_{k_1}, \underbrace{(3.07797, 4.32228]}_{k_2}, \underbrace{(4.32228, 5.39128]}_{k_3}, \underbrace{(5.39128, 6.63559]}_{k_4}, \underbrace{(6.63559, \infty)}_{k_5}$$

Das Ergebnis für $T = \sum_{j=1}^k \frac{(y_i - np_i)^2}{np_i}$ aus den Würfeln lautet somit: $T = 0.5$ (Berechnung siehe Tabelle 4.4). Das Ergebnis für T muss nun mit $\chi_{k-1-d;0.95}^2 = \chi_{5-1-2;0.95}^2 = \chi_{2;0.95}^2$ verglichen werden:

$$T = 0.5 < \chi_{2;0.95}^2 = 5.991$$

Somit wird die Nullhypothese nicht verworfen und die Messwerte sind wohl $\mathcal{N}(4.85678, 2.11263)$.

4.3.4. Tests und Konfidenzintervalle

Es gibt einen interessanten Zusammenhang zwischen Tests und Konfidenzintervallen:

i	y_i	np_i	$\frac{(y_i - np_i)^2}{np_i}$
1	20	$\frac{100}{5} = 20$	$\frac{(20-20)^2}{20} = \frac{0}{20}$
2	19	$\frac{100}{5} = 20$	$\frac{(19-20)^2}{20} = \frac{1}{20}$
3	21	$\frac{100}{5} = 20$	$\frac{(21-20)^2}{20} = \frac{1}{20}$
4	18	$\frac{100}{5} = 20$	$\frac{(18-20)^2}{20} = \frac{4}{20}$
5	22	$\frac{100}{5} = 20$	$\frac{(22-20)^2}{20} = \frac{4}{20}$
		$\sum_{j=1}^k$	$\frac{0+1+1+4+4}{20} = \frac{10}{20} = 0.5$

Tabelle 4.4.: Berechnung von $T = \sum_{j=1}^k \frac{(y_i - np_i)^2}{np_i}$ (siehe Beispiel 4.21)

Satz 4.6 Dualität von Konfidenzintervall und Tests

Wenn $I = [a(X_1, \dots, X_n), b(X_1, \dots, X_n)]$ ein Konfidenzintervall (mit Übergangswahrscheinlichkeit $\gamma = 1 - \alpha$) für einen Parameter θ ist, dann erhält man einen Test mit Signifikanzniveau α , wenn man die Nullhypothese $H_0 : \theta = \theta_0$ genau dann verwirft, wenn $\theta_0 \notin I$.

Ist umgekehrt für jedes θ_0 ein Test mit Niveau α für die Nullhypothese $H_0 : \theta = \theta_0$ gegeben, dann ist die Menge aller θ_0 , für die $H_0 : \theta = \theta_0$ nicht verworfen wird, ein Konfidenzintervall (mit Überdeckungswahrscheinlichkeit $\gamma = 1 - \alpha$).

Beispiel 4.22 Rufen wir uns nochmal das Beispiel mit der lügenden Obstverkäuferin ins Gedächtnis: 4.15 bzw. die Fortsetzung 4.18, denn beide sind ein Beispiel für die Dualität von Konfidenzintervall und Test.

Wir haben beispielsweise $\mu_0 = 1$ und $\mu_0 = 0.75$ verworfen, aber $\mu_0 = 0.9$ angenommen. Genau dieses Verhalten spiegelt sich auch im Konfidenzintervall wieder, denn es lautete: $KI = [0.81683, 0.96317]$.

5. Informationstheorie

5.1. Information und Entropie

Beispiel 5.1 Aus einem Stoß mit 16 Schnapskarten zieht Spieler A zufällig eine Karte, Spieler B stellt nun Fragen über die Karte, die A mit "ja" oder "nein" beantworten darf.

Wie viele Fragen sind (bei optimaler Spielweise) maximal notwendig, um die zufällig gezogene Karte zu erraten?

Mit der optimalen Strategie, der binäre Suche (Menge der noch möglichen Karten in gleichgroße disjunkte Teilmengen aufspalten und Fragen ob x in der ersten Teilmenge enthalten ist), lässt sich die maximale Spieldauer bestimmen:

Da wir wissen, dass unsere Schnapskarte X mit einer diskreten Gleichverteilung $p_x = \frac{1}{16}$ verteilt ist, können wir ihren Informationsgehalt bestimmen:

Definition 5.1 (Informationsgehalt) Der Informationsgehalt I einer Zufallsvariable X mit der Wahrscheinlichkeit p_x ist definiert als:

$$I_x = I(p_x) = \log_2\left(\frac{1}{p_x}\right) = -\log_2(p_x) \quad (5.1)$$

Ergibt in unserem Fall

$$I_x = \log_2(16) = 4 \dots \text{bit (ja-nein Fragen)}$$

Interessanter wird es, wenn Spieler A die Karte nicht mehr zieht, sondern zufällig mit der Verteilung $P = (p_1, \dots, p_m)$ wählt.

Da die Wahrscheinlichkeiten der Karten jetzt nicht mehr alle gleich sind und wir somit die Wahrscheinlichkeit der verdeckten Karte nicht kennen, können wir nicht so einfach den Informationsgehalt der gewählten Karte bestimmen.

Da wir jedoch die Verteilung P kennen, können wir damit den **mittleren Informationsgehalt** bzw. die **Informationsdichte** bestimmen:

Definition 5.2 (Entropie) Die Entropie H ist ein Maß für den mittleren Informationsgehalt bzw. der Informationsdichte einer Verteilung.

Sie ist definiert als **Erwartungswert** des Informationsgehalts:

$$H(P) = \sum_{i=1}^m I(p_i) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (5.2)$$

Für eine Zufallsvariable $X \sim P$ gilt:

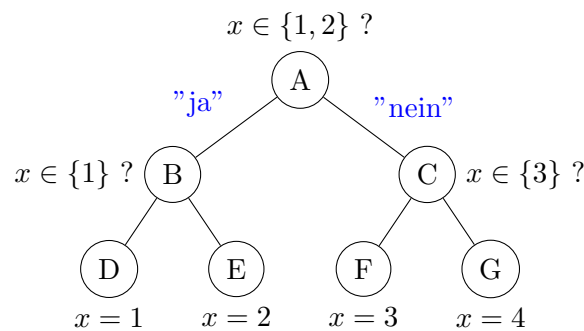
$$H(X) = H(P) \quad (5.3)$$

In unserem Spiel müssen jetzt eine neue Strategie finden, da die einfache binäre Suche nicht mehr eindeutig definiert ist.

Eine Strategie kann als Binärbaum dargestellt werden und umgekehrt repräsentiert ein Binärbaum eine Spielstrategie:

Jeder innere Knoten entspricht einer Frage, jede linke ausgehende Kante einer Antwort "ja" und jede rechte ausgehende Kante einem "nein". Die Endknoten (Blätter) stellen einen Wert X der Zufallsvariable dar.

Es gibt genau m -Blätter, denn es existieren ja auch m Zufallsvariablen. Beispiel mit $x \in \{1, 2, 3, 4\}$:



Die Suche nach der optimalen Strategie lässt sich so formulieren, dass wir unter allen Binärbäumen denjenigen suchen, für den **mittlere Unbestimmtheit** minimal ist:

Definition 5.3 (Mittlere Unbestimmtheit) Als *mittlere Unbestimmtheit* H^* bzw. *mittlere Codewortlänge* oder auch *mittlere Blattlänge* definiert ist:

$$H^*(P) = \sum_{i=1}^m p_i l_i \quad (5.4)$$

wobei l_i die Blattlänge des Blatts mit dem Index i (also seine Entfernung von der Wurzel)

Für diese Aufgabe ist es nützlich zu wissen, wann ein Binärbaum mit den Blattlängen l_1, \dots, l_m existiert.

Satz 5.1 (Ungleichung von Kraft) Ein Binärbaum mit den Blattlängen l_1, \dots, l_m existiert genau dann, wenn

$$\sum_{i=1}^m 2^{-l_i} \leq 1.$$

Gleichheit gilt genau dann, wenn der Baum vollständig ist.

Die Suche nach der optimalen Strategie besteht also darin,

$$\sum p_i l_i$$

unter der Nebenbedingung

$$\sum 2^{-l_i} \leq 1$$

zu minimieren. Es ist leicht einzusehen, dass im optimalen Fall in der letzten Ungleichung Gleichheit gelten muss (sonst könnten wir einfach das größte l_i verkleinern). Wenn wir die Forderung, dass l_i ganzzahlig sein muss, beiseite lassen, lässt sich das Minimum mit der Lagrange-Methode bestimmen und hat den Wert

$$\sum p_i \log_2\left(\frac{1}{p_i}\right).$$

Dies entspricht der **Entropie**, die minimale Spieldauer bei optimaler Strategie ist also durch die Entropie bestimmt.

Definition 5.4 (Huffman-Algorithmus) Die optimale Spielweise für eine Verteilung $P = (p_1, \dots, p_m)$ ist der sogenannte **Huffman-Algorithmus**, der folgendermaßen vorgeht:

1. Wenn $m = 1$, dann besteht der Baum nur aus der Wurzel, Ende.
2. Ordne die Wahrscheinlichkeiten: $p_1 \geq \dots \geq p_m$.
3. Fasse die zwei kleinsten Wahrscheinlichkeiten zusammen: $p_{m-1}^* = p_{m-1} + p_m$,
4. Konstruiere den optimalen Baum für $P^* = (p_1, \dots, p_{m-1}, p_{m-1}^*)$
5. Ersetze Blatt $m-1$ durch einen inneren Knoten mit den Blättern $m-1$ und m .

Für den Huffman-Algorithmus gilt:

Satz 5.2 $H(P) \leq H^*(P) \leq H(P) + 1$

Der Summand 1 in der oberen Abschätzung stört ein wenig; man kann diese Differenz verringern, indem man statt eines einzelnen Elements mehrere (unabhängige und identisch verteilte) errät, sagen wir n . Die Entropie der gemeinsamen Verteilung ist (wie weiter unten gezeigt wird) Entropie $nH(P)$, damit lässt sich die mittlere Anzahl der Fragen, um den gesamten Block zu erraten, mit $nH(P) + 1$ abschätzen, pro Element ergibt das $H(P) + 1/n$, was beliebig nahe an die Entropie herankommt.

5.1.1. Bedingte Entropie und die Kullback-Leibler Divergenz

Wenn X und Y zwei Zufallsvariable sind, können wir (X, Y) als eine neue Zufallsvariable mit Wahrscheinlichkeit $p(x, y) = \mathbb{P}(X = x, Y = y)$ betrachten.

Die Entropie, die diese neuen Zufallsvariable besitzt, heißt **gemeinsame Entropie** von X und Y :

$$H(X, Y) = H((X, Y))$$

Nehmen wir die bedingte Wahrscheinlichkeit $p(x|y) = \mathbb{P}(X = x|Y = y)$ und definieren

Definition 5.5 (Bedingte Entropie)

$$H(X|Y = y) = - \sum_x p(x|y) \log_2(p(x|y))$$

und nennen

$$H(X|Y) = \sum_y p_Y(y) H(X|Y = y)$$

die **bedingte Entropie** von X unter Y .

Die bedingte Entropie ist ein Maß für die "Ungewissheit" die über einer Variable (X) verbleibt, wenn eine andere Variable (Y) bekannt ist.

Der Wertebereich für die bedingte Entropie ist folgendermaßen beschränkt:

$$0 \leq H(X|Y) \leq H(X)$$

Genauer gilt:

Satz 5.3

$$H(X|Y) = H(X)$$

falls X und Y stochastisch unabhängig sind und

$$H(X|Y) = 0$$

falls X aus Y funktionell bestimmt werden kann, das heißt es existiert eine Funktion g mit $X = g(Y)$

Aus diesem Satz ergeben sich weitere Eigenschaften für die **bedingte** bzw. die **gemeinsame Entropie**.

Satz 5.4

$$H(X, Y) = H(Y) + H(X|Y),$$

$$\max(H(X), H(Y)) \leq H(X, Y) \leq H(X) + H(Y),$$

$$H(X|Y) \leq H(X),$$

$$H(X|Y, Z) \leq H(X|Y).$$

$$H(X, Y) = H(X) + H(Y)$$

gilt genau dann, wenn X und Y unabhängig sind.

$$H(X, Y) = H(X)$$

gilt genau dann, wenn es eine Funktion g gibt, sodass $Y = g(X)$.

Mit Hilfe der bedingten Entropie können wir den Informationsbegriff erweitern:

Definition 5.6 (Information zwischen 2 Zufallsvariablen)

$$I(X, Y) = H(X) + H(Y) - H(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

heißt die Information zwischen X und Y .

Der letzte Satz impliziert

Satz 5.5

$$0 \leq I(X, Y) \leq \min(H(X), H(Y)).$$

Die Information ist genau dann 0, wenn X und Y unabhängig sind. $I(X, Y) = H(X)$ gilt genau dann, wenn $X = g(Y)$ gilt, also wenn X eindeutig aus Y bestimmt werden kann.

Wenn X aus Y zwar nicht mit absoluter Sicherheit, aber mit großer Wahrscheinlichkeit bestimmt werden kann, dann unterscheidet sich die Information nur wenig von der Entropie von X :

Satz 5.6 Wenn $\mathbb{P}(X \neq Y) \leq \epsilon$ ist, dann gilt

$$I(X, Y) \geq H(X) - H(\epsilon, 1 - \epsilon) - \epsilon \log_2(m).$$

Wir betrachten nun den Fall, dass wir eine Zufallsvariable X mit Verteilung P erraten müssen, aber die optimale Strategie für eine andere Verteilung Q verwenden. Wenn wir annehmen, dass die Anzahl der Fragen für Ausgang i gleich $\log_2(1/q_i)$ ist (das stimmt nicht genau, aber wir können etwa eine fast optimale Strategie finden, in der sich die Anzahl der Fragen um weniger als 1 davon unterscheidet), dann brauchen wir im Mittel

$$\sum p_i \log_2(1/q_i)$$

Fragen statt

$$\sum p_i \log_2(1/p_i),$$

also um

$$\sum p_i \log_2(p_i/q_i)$$

Fragen zuviel.

Das führt uns zu der Definition

Definition 5.7 (Kullback-Leibler-Divergenz)

$$D(P, Q) = \sum_i p_i \log_2(p_i/q_i).$$

heißt die Informationsdivergenz (I -divergenz, Kullback-Leibler Distanz, relative Entropie, Strafe des Irrtums) zwischen P und Q .

Die Kullback-Leibler-Divergenz ist ein Maß für die "Unterschiedlichkeit" zweier Verteilungen P und Q .

Es gilt:

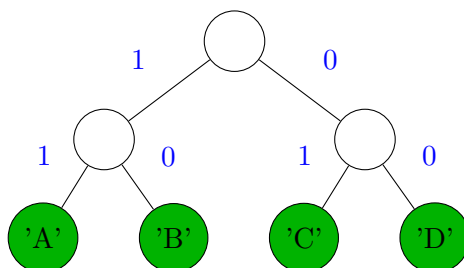
$$D(P, Q) = 0$$

falls die Verteilungen P und Q identisch sind.

5.2. Codes

Eine Fragestrategie, wie wir sie im vorherigen Kapitel kennen gelernt haben, kann auch unter einem anderen Gesichtspunkt gesehen werden, nämlich indem für jede Frage die Antwort "nein" mit einer 0, die Antwort "ja" mit einer 1 codiert wird.

Sprich jeder linke Knoten des Binärbaums wird mit einer 1 und jeder rechte Knoten des Binärbaums wird mit einer 0 gekennzeichnet. Als Beispiel nehmen wir folgenden Code bzw. Binärbaum an:



Definition 5.8 (Code) Ein Code C über den Alphabeten A und B , ist eine Abbildung (= Codierung):

$$C : A \rightarrow B \quad (5.5)$$

In unserem Fall ist die Menge $A = \{00, 01, 10, 11\}$ und $B = \{ 'A', 'B', 'C', 'D' \}$, die Codierung ist folgendermaßen definiert:

1. $C(11) = 'A'$
2. $C(10) = 'B'$
3. $C(01) = 'C'$
4. $C(00) = 'D'$

Codes, die auf diese Weise gewonnen werden, haben eine wichtige Eigenschaft — sie sind präfixfrei.

Definition 5.9 (Präfixfrei) Ein Code C mit den Codewörtern $(c_1, c_2 \dots c_n)$ heißt präfixfrei oder auch fortlaufende Entzifferbar, wenn kein Codewort c_i präfix eines anderen Codeworts c_j ist.
Mit anderen Worten darf kein Codewort am Beginn eines anderen Codeworts stehen.

Fortlaufende Entzifferbarkeit heißt, dass an jeder Stelle der codierten Nachricht festgestellt werden kann, ob dort ein Codewort endet, ohne dass die nachfolgenden Zeichen bekannt sind.

5.2.1. Der Huffman-Code

Transformiert man, die Spielestrategie des Huffman-Algorithmus(siehe Definition 5.4) in einen Code, erhält man den **Huffman-Code**. Der Huffmancode ist der optimale präfixfreie Code, also der mit der kleinsten mittleren Codewortlänge (siehe Definition 5.3).

5.2.2. Der Shannon-Code

Einen anderen Code mit fast optimaler Codewortlänge erhalten wir, wenn wir von unserer oberen Abschätzung für die Entropie ausgehen. Wir wollen also einen Code mit Codewortlängen $l_i = \lceil \log_2(1/p_i) \rceil$ explizit angeben.

Dazu ordnen wir die Wahrscheinlichkeiten absteigend ($p_1 \geq \dots \geq p_m$) und setzen $f_i = \sum_{j=1}^i p_j$.

Das Codewort c_i erhalten wir, indem wir f_{i-1} als Binaärzahl darstellen und die ersten l_i Nachkommastellen als Code verwenden.

ACHTUNG: f_{i-1} ist kleiner als 1, das erste Bit des Codeworts ist also durch 2^{-1} bestimmt

Beispiel 5.2 (Umwandeln von Dezimal in Binär) :

Wir wollen die Zahl 0.1_d in eine Binärzahl umwandeln:

$$0.1 \cdot 2 = 0.2 \Rightarrow 0$$

$$0.2 \cdot 2 = 0.4 \Rightarrow 0$$

$$0.4 \cdot 2 = 0.8 \Rightarrow 0$$

$$0.8 \cdot 2 = 1.6 \Rightarrow 1$$

$$0.6 \cdot 2 = 1.2 \Rightarrow 1$$

$$0.2 \cdot 2 = 0.4 \Rightarrow 0$$

⋮

Somit lautet die Zahl umgewandelt:

$$0.1_d = 0.00110_b$$

Es ist nicht schwer einzusehen, dass dadurch ein präfixfreier Code definiert wird, der Shannon-Code. Anschaulich lässt sich das Verfahren an einem Beispiel darstellen:

5.2.3. Der Fano-Code

Der einzige Schönheitsfehler dabei ist, dass die Wahrscheinlichkeiten geordnet werden müssen.

Diesen Schönheitsfehler behebt der Fano-Code. Hier verwendet man dieselben Notationen wie beim Shannon-Code, nur codiert man $\frac{(f_{i-1}+f_i)}{2}$ mit diesmal $\lceil \log_2(\frac{1}{p_i}) \rceil + 1$ Bits.

Wieder mit einem Beispiel:

5.2.4. Weitere Codierungsverfahren

Diese Idee kann man auf das Codieren von ganzen Blöcken anwenden, im Extremfall wird die ganze Nachricht als einzelner Block kodiert. Im Vergleich zum Huffman-Code ist das hier möglich, weil nicht der ganze Code generiert werden muss, sondern nur das eine, das die Nachricht kodiert. Verfahren, die auf dieser Idee beruhen, werden als arithmetische Codes bezeichnet.

Außer den präfixfreien Codes gibt es auch noch andere, die eindeutig entziffert werden können. Wir definieren

- Definition 5.10**
1. Ein Code heißt endlich **endlich eindeutig entzifferbar**, wenn jede endliche Aneinanderreihung von Codewörtern eindeutig in Codewörter zerlegt werden kann.
 2. Ein Code heißt **eindeutig entzifferbar** (manchmal zur Unterscheidung von 1. **unendlich eindeutig entzifferbar**), wenn jede endliche oder unendliche Aneinanderreihung von Codewörtern eindeutig zerlegt werden kann.

Der Code $\{0, 01\}$ ist offensichtlich nicht präfixfrei, aber trotzdem eindeutig entzifferbar, für die korrekte Zerlegung muss man allerdings das nachfolgende Zeichen kennen. Der Code $\{0, 01, 11\}$ ist endlich eindeutig entzifferbar, aber nicht eindeutig entzifferbar.

Wir können nun die Frage stellen, ob durch den Verzicht auf die fortlaufende Entzifferbarkeit etwas gewonnen werden kann, also ob es einen endlich eindeutig entzifferbaren Code gibt, der kleiner mittlere Codewortlänge hat als der Huffman-Code. Diese Hoffnung ist allerdings vergebens, denn es gilt

Satz 5.7 Die Codewortlängen in einem endlich eindeutig entzifferbaren Code erfüllen die Kraft'sche Ungleichung.

Aus diesem Satz folgt, dass es zu jedem endlich eindeutige entzifferbaren Code einen präfixfreien Code mit denselben Codewortlängen (und damit mit derselben mittleren Codewortlänge) gibt.

(universelle Codes)

5.3. Informationsquellen

Definition 5.11 Eine **Informationsquelle** ist eine Folge $\mathcal{X} = (X_1, \dots)$ von Zufallsvariablen.

Die verschiedenen Möglichkeiten für die Abhängigkeitsstruktur dieser Folge ergeben die Definitionen

Definition 5.12 Wenn die Zufallsvariablen X_n unabhängig und identisch verteilt sind, dann heißt \mathcal{X} **gedächtnislos**.

Wenn (X_n) eine Markovkette bilden oder stationär sind, dann heißt die Quelle Markovquelle bzw. stationäre Quelle. Eine irreduzible Markovquelle nennen wir **ergodisch**.

Eine wichtige Größe ist die Entropie einer Quelle. Im Sinne der Idee, die wir bei der Einführung der Entropie verwendet haben, definieren wir sie über die mittlere Codewortlänge beim optimalen Kodieren von langen Blöcken:

Definition 5.13 Die **Entropie der Quelle** \mathcal{X} ist

$$H(\mathcal{X}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, \dots, X_n),$$

wenn dieser Grenzwert existiert (wenn nicht, dann hat \mathcal{X} keine Entropie.)

Für eine gedächtnislose Quelle gilt

$$H(\mathcal{X}) = H(X_1).$$

Für eine (irreduzible) Markovquelle mit Übergangsmatrix P ergibt sich

$$H(X) = \sum \pi_i H(P_i),$$

wobei P_i die i -te Zeile von P und π die stationäre Verteilung ist.

Für eine stationäre Quelle existiert die Entropie.

Satz 5.8 (Shannon-MacMillan) Für eine stationäre Quelle gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 p(X_1, \dots, X_n) = -H(\mathcal{X})$$

in Wahrscheinlichkeit.

Das kann man so sehen, dass mit hoher Wahrscheinlichkeit gilt, dass die Wahrscheinlichkeit, genau die Folge (X_1, \dots, X_n) zu ziehen $\approx 2^{-nH(\mathcal{X})}$ ist.

5.4. Blockcodes

5.5. Kanalcodierung

5.6. Natürliche Sprachen als Informationsquellen

A. Tabellen

A.1. Die Verteilungsfunktion der Standardnormalverteilung

$$\Phi(x) = \int_{-\infty}^x e^{-\frac{x^2}{2}}$$

Negative Werte werden aus Gründen der Symmetrie nicht angegeben: denn es gilt:

$$\Phi(-x) = 1 - \Phi(x)$$

Die Transformation von der Normalverteilung zur Standardnormalverteilung:

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

A.1. Die Verteilungsfunktion der Standardnormalverteilung

$x \backslash *$	0	1	2	3	4	5	6	7	8	9
0.0*	.500	.504	.508	.512	.516	.520	.524	.528	.532	.536
0.1*	.540	.544	.548	.552	.556	.560	.564	.567	.571	.575
0.2*	.579	.583	.587	.591	.595	.599	.603	.606	.610	.614
0.3*	.618	.622	.626	.629	.633	.637	.641	.644	.648	.652
0.4*	.655	.659	.663	.666	.670	.674	.677	.681	.684	.688
0.5*	.691	.695	.698	.702	.705	.709	.712	.716	.719	.722
0.6*	.726	.729	.732	.736	.739	.742	.745	.749	.752	.755
0.7*	.758	.761	.764	.767	.770	.773	.776	.779	.782	.785
0.8*	.788	.791	.794	.797	.800	.802	.805	.808	.811	.813
0.9*	.816	.819	.821	.824	.826	.829	.831	.834	.836	.839
1.0*	.841	.844	.846	.848	.851	.853	.855	.858	.860	.862
1.1*	.864	.867	.869	.871	.873	.875	.877	.879	.881	.883
1.2*	.885	.887	.889	.891	.893	.894	.896	.898	.900	.901
1.3*	.903	.905	.907	.908	.910	.911	.913	.915	.916	.918
1.4*	.919	.921	.922	.924	.925	.926	.928	.929	.931	.932
1.5*	.933	.934	.936	.937	.938	.939	.941	.942	.943	.944
1.6*	.945	.946	.947	.948	.949	.951	.952	.953	.954	.954
1.7*	.955	.956	.957	.958	.959	.960	.961	.962	.962	.963
1.8*	.964	.965	.966	.966	.967	.968	.969	.969	.970	.971
1.9*	.971	.972	.973	.973	.974	.974	.975	.976	.976	.977
2.0*	.977	.978	.978	.979	.979	.980	.980	.981	.981	.982
2.1*	.982	.983	.983	.983	.984	.984	.985	.985	.985	.986
2.2*	.986	.986	.987	.987	.987	.988	.988	.988	.989	.989
2.3*	.989	.990	.990	.990	.990	.991	.991	.991	.991	.992
2.4*	.992	.992	.992	.992	.993	.993	.993	.993	.993	.994
2.5*	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995
2.6*	.995	.995	.996	.996	.996	.996	.996	.996	.996	.996
2.7*	.997	.997	.997	.997	.997	.997	.997	.997	.997	.997
2.8*	.997	.998	.998	.998	.998	.998	.998	.998	.998	.998
2.9*	.998	.998	.998	.998	.998	.998	.998	.999	.999	.999

A.2. Quantile z_p der Standardnormalverteilung

p	z_p	p	z_p	p	z_p
.51	.025	.71	.553	.91	1.341
.52	.050	.72	.583	.92	1.405
.53	.075	.73	.613	.93	1.476
.54	.100	.74	.643	.94	1.555
.55	.126	.75	.674	.95	1.645
.56	.151	.76	.706	.96	1.751
.57	.176	.77	.739	.97	1.881
.58	.202	.78	.772	.975	1.960
.59	.228	.79	.806	.98	2.054
.60	.253	.80	.842	.99	2.326
.61	.279	.81	.878	.991	2.366
.62	.305	.82	.915	.992	2.409
.63	.332	.83	.954	.993	2.457
.64	.358	.84	.994	.994	2.512
.65	.385	.85	1.036	.995	2.576
.66	.412	.86	1.080	.996	2.652
.67	.440	.87	1.126	.997	2.748
.68	.468	.88	1.175	.998	2.878
.69	.496	.89	1.227	.999	3.090
.70	.524	.90	1.282	.9999	3.719

A.3. Quantile $t_{n;p}$ der t -Verteilung mit n Freiheitsgraden:

n	.9	.95	.975	.99	.995	n	.9	.95	.975	.99	.995
1	3.078	6.314	12.706	31.821	63.675	26	1.316	1.706	2.056	2.479	2.779
2	1.886	2.920	4.303	6.965	9.725	27	1.314	1.703	2.052	2.473	2.467
3	1.638	2.353	3.183	4.541	5.841	28	1.313	1.701	2.048	2.467	2.763
4	1.533	2.132	2.776	3.747	4.604	29	1.311	1.699	2.045	2.462	2.756
5	1.476	2.015	2.571	3.365	4.032	30	1.310	1.697	2.042	2.457	2.750
6	1.440	1.943	2.447	3.143	3.707	31	1.309	1.696	2.040	2.453	2.744
7	1.415	1.895	2.365	2.998	3.499	32	1.309	1.694	2.037	2.449	2.738
8	1.397	1.860	2.306	2.896	3.355	33	1.308	1.692	2.035	2.445	2.733
9	1.383	1.833	2.262	2.821	3.250	34	1.307	1.691	2.032	2.441	2.728
10	1.372	1.812	2.228	2.764	3.169	35	1.306	1.690	2.030	2.438	2.724
11	1.363	1.796	2.201	2.718	3.106	40	1.303	1.684	2.021	2.423	2.704
12	1.356	1.782	2.179	2.681	3.055	45	1.301	1.679	2.014	2.412	2.690
13	1.350	1.771	2.160	2.650	3.012	50	1.299	1.676	2.009	2.403	2.678
14	1.345	1.761	2.145	2.624	2.977	55	1.297	1.673	2.004	2.396	2.668
15	1.341	1.753	2.131	2.602	2.947	60	1.296	1.671	2.000	2.390	2.660
16	1.337	1.746	2.120	2.583	2.921	65	1.295	1.669	1.997	2.385	2.654
17	1.333	1.740	2.110	2.567	2.898	70	1.294	1.667	1.994	2.381	2.648
18	1.330	1.734	2.101	2.552	2.878	75	1.293	1.665	1.992	2.377	2.643
19	1.328	1.729	2.093	2.539	2.861	80	1.292	1.664	1.990	2.374	2.639
20	1.325	1.725	2.086	2.528	2.845	85	1.292	1.663	1.988	2.371	2.635
21	1.323	1.721	2.080	2.518	2.831	90	1.291	1.662	1.987	2.368	2.632
22	1.321	1.717	2.074	2.508	2.819	95	1.291	1.661	1.985	2.366	2.629
23	1.319	1.714	2.069	2.500	2.807	100	1.290	1.660	1.984	2.364	2.626
24	1.318	1.711	2.064	2.492	2.797	105	1.290	1.659	1.983	2.362	2.623
25	1.316	1.708	2.060	2.485	2.787	∞	1.282	1.645	1.960	2.326	2.576

A.4. Quantile $\chi^2_{n;p}$ der χ^2 -Verteilung mit n Freiheitsgraden:

n	.005	.01	.02	.025	.05	.1	.5	.9	.95	.975	.98	.99	.995
1	.000	.000	.001	.001	.004	.016	.455	2.706	3.841	5.024	5.412	6.635	7.879
2	.010	.020	.040	.051	.103	.211	1.386	4.605	5.991	7.378	7.824	9.210	10.597
3	.072	.115	.185	.216	.352	.584	2.366	6.251	7.815	9.348	9.837	11.345	12.838
4	.207	.297	.429	.484	.711	1.064	3.357	7.779	9.488	11.143	11.668	13.277	14.860
5	.412	.554	.752	.831	1.145	1.610	4.351	9.236	11.070	12.832	13.308	15.086	16.750
6	.676	.872	1.134	1.237	1.635	2.204	5.348	10.645	12.592	14.449	15.033	16.812	18.548
7	.989	1.239	1.564	1.690	2.167	2.833	6.346	12.017	14.067	16.013	16.622	18.475	20.278
8	1.344	1.646	2.032	2.180	2.733	3.490	7.344	13.362	15.507	17.535	18.168	20.090	21.955
9	1.735	2.088	2.532	2.700	3.325	4.168	8.343	14.684	16.919	19.023	19.679	21.666	23.589
10	2.156	2.558	3.059	3.247	3.940	4.865	9.342	15.987	18.307	20.483	21.161	23.209	25.188
11	2.603	3.053	3.609	3.816	4.575	5.578	10.341	17.275	19.675	21.920	22.618	24.725	26.757
12	3.074	3.571	4.178	4.404	5.226	6.304	11.340	18.549	21.026	23.336	24.054	26.217	28.300
13	3.565	4.107	4.765	5.009	5.892	7.042	12.340	19.812	22.362	24.736	25.472	27.688	29.819
14	4.075	4.660	5.368	5.629	6.571	7.790	13.339	21.064	23.685	26.119	26.873	29.141	31.319
15	4.601	5.229	5.985	6.262	7.261	8.547	14.339	22.307	24.996	27.488	28.259	30.578	32.801
16	5.142	5.812	6.614	6.908	7.962	9.312	15.338	23.542	26.269	28.845	29.633	32.000	34.267
17	5.697	6.408	7.255	7.564	8.672	10.085	16.338	24.769	27.587	30.191	30.995	33.409	35.718
18	6.265	7.015	7.906	8.231	9.390	10.835	17.338	25.909	28.869	31.526	32.346	34.805	37.156
19	6.844	7.633	8.567	8.907	10.117	11.651	18.338	27.204	30.144	32.852	33.687	36.191	38.582
20	7.434	8.260	9.237	9.591	10.851	12.443	19.337	28.412	31.410	34.170	35.020	37.566	39.997
21	8.034	8.897	9.915	10.283	11.591	13.240	20.337	29.615	32.671	35.479	36.343	38.932	41.401
22	8.643	9.542	10.600	10.982	12.338	14.041	21.337	30.813	33.924	36.781	37.659	40.289	42.796
23	9.260	10.196	11.293	11.689	13.091	14.848	22.337	32.007	35.172	38.076	38.968	41.638	44.181
24	9.886	10.856	11.992	12.401	13.848	15.659	23.337	33.196	36.415	39.364	40.270	42.980	45.559
25	10.520	11.524	12.697	13.120	14.611	16.473	24.337	34.382	37.652	40.646	41.566	44.324	46.928
26	11.160	12.198	13.409	13.844	15.379	17.292	25.336	35.563	38.885	41.923	42.856	45.642	48.290
27	11.808	12.879	14.125	14.573	16.151	18.114	26.336	36.741	40.113	43.194	44.140	46.963	49.645
28	12.461	13.565	14.847	15.308	16.928	18.939	27.336	37.916	41.337	44.461	45.419	48.278	50.993
29	13.121	14.256	15.574	16.047	17.708	19.768	28.336	39.087	42.557	45.722	46.693	49.588	52.336
30	13.787	14.953	16.306	16.791	18.493	20.599	29.336	40.256	43.773	46.979	47.962	50.892	53.672
31	14.458	15.655	17.042	17.539	19.281	21.434	30.336	41.422	44.985	48.232	49.226	52.191	55.003
32	15.134	16.362	17.783	18.291	20.072	22.271	31.336	42.585	46.194	49.480	50.487	53.486	56.328
33	15.815	17.074	18.527	19.047	20.867	23.110	32.336	43.745	47.400	50.725	51.743	54.776	57.648
34	16.501	17.789	19.275	19.806	21.664	23.952	33.336	44.903	48.602	51.966	52.995	56.061	58.964
35	17.192	18.509	20.027	20.569	22.465	24.797	34.336	46.059	49.802	53.203	54.244	57.342	60.275
40	20.707	22.164	23.838	24.433	26.509	29.051	39.335	51.805	55.758	59.342	60.436	63.691	66.766
45	24.311	25.901	27.720	28.366	30.612	33.350	44.335	57.505	61.656	65.410	66.555	69.957	73.166
50	27.991	29.707	31.664	32.357	34.764	37.689	49.335	63.167	67.505	71.420	72.613	76.154	79.490
55	31.735	33.570	35.659	36.398	38.958	42.060	54.335	68.796	73.311	77.380	78.619	82.292	85.749
60	35.534	37.485	39.699	40.482	43.188	46.459	59.335	74.397	79.082	83.298	84.580	88.397	91.952
65	39.383	41.444	43.779	44.603	47.450	50.883	64.335	79.973	84.821	89.177	90.501	94.422	98.105
70	43.275	45.442	47.893	48.758	51.739	55.329	69.334	85.527	90.531	95.023	96.388	100.425	104.215
75	47.206	49.475	52.039	52.942	56.054	59.795	74.334	91.061	96.217	100.839	102.243	106.393	110.286
80	51.172	53.540	56.213	57.153	60.391	64.278	79.334	96.578	101.879	106.629	108.069	112.329	116.321
85	55.170	57.634	60.412	61.389	64.749	68.777	84.334	102.079	107.522	112.393	113.871	118.236	122.325
90	59.196	61.754	64.635	65.647	69.126	73.291	89.334	107.565	113.145	118.136	119.649	124.116	128.299
95	63.250	65.898	68.879	69.925	73.520	77.818	94.334	113.038	118.752	123.858	125.405	129.973	134.247
100	67.328	70.065	73.142	74.222	77.929	82.358	99.334	118.498	124.342	129.561	131.142	135.806	140.169

B. Mathematische Hintergründe

B.1. Matrizen

B.1.1. Determinante

Determinantenrechenregeln

Satz B.1 Für die Determinante einer Matrix gilt:

- i) Multipliziert man die Spalte/Zeile einer Matrix A mit einem Faktor $\lambda \in K$, so ist die Determinante der neuen Matrix $\det A' = \lambda \cdot \det A$.
- ii) Addiert man zu einer Spalte/Zeile einer Matrix das Vielfache einer anderen Spalte/Zeile, so verändert sich der Wert der Determinante nicht.
- iii) Vertauscht man in einer Matrix A zwei Spalten/Zeilen, so ist die Determinante der neuen Matrix $\det A' = -\det A$.

Formt man i um, so erhält man die Regel, wie man aus einer Zeile/Spalte einer Matrix einen Wert herausheben kann:

Beispiel B.1 Herausheben von Werten einer Spalte, um die Determinantenberechnung zu vereinfachen.

$$\det \begin{pmatrix} \frac{1}{2} - \lambda & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} - \lambda & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} - \lambda \end{pmatrix} \xrightarrow{\text{mit Regel ii folgt:}} \det \begin{pmatrix} 1 - \lambda & \frac{1}{4} & \frac{1}{4} \\ 1 - \lambda & \frac{1}{2} - \lambda & \frac{1}{4} \\ 1 - \lambda & \frac{1}{4} & \frac{1}{2} - \lambda \end{pmatrix} =$$

$$\xrightarrow{\text{mit Herausheben (Regel i)}} = (1 - \lambda) \det \begin{pmatrix} 1 & \frac{1}{4} & \frac{1}{4} \\ 1 & \frac{1}{2} - \lambda & \frac{1}{4} \\ 1 & \frac{1}{4} & \frac{1}{2} - \lambda \end{pmatrix} = (1 - \lambda) \left(\lambda^2 - \frac{\lambda}{2} + \frac{1}{16} \right)$$

Die Determinante einer 2x2 Matrix

$$\det \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (a \cdot d) - (b \cdot c)$$

Die Determinante einer 3x3 Matrix

$$\det \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = (a \cdot e \cdot i + b \cdot f \cdot g + c \cdot d \cdot h) - (c \cdot e \cdot g + f \cdot h \cdot a + i \cdot b \cdot d)$$

B.1.2. Matrixpotenz

Sie ist das Ergebnis einer wiederholten Matrixmultiplikation.

Satz B.2 Lässt sich eine Matrix diagonalisieren, dann existiert eine reguläre **Transformationsmatrix** T und eine Diagonalmatrix D , sodass gilt:

$$A^n = T \cdot D^n \cdot T^{-1} \text{ wobei } T^{-1} \text{ die Inverse Matrix zu } T \text{ ist.}$$

B.1.3. Diagonalmatrix D_x , bzw. Transformationsmatrix T bestimmen

Folgende Schritte sind notwendig:

- 1) Eigenwerte λ_i der Matrix A bestimmen:

$$\det(A - \lambda I) = 0 \text{ wobei } I \text{ die Einheitsmatrix}$$

Hier bekommt man dann bei einer 3x3 Matrix 3 Werte λ_1, λ_2 und λ_3 , bei einer 2x2 Matrix 2 Werte: λ_1 bzw. λ_2 . (also einfach Nullstellen finden...)

- 2) Eigenvektoren ($\vec{e}(\lambda)$) bestimmen:

$$(A - \lambda_i I) \cdot \vec{e} = (A - \lambda_i I) \cdot \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \vec{0}$$

Für jedes vorher bestimmte $\lambda_1, \dots, \lambda_n$ in die Gleichung statt λ einsetzen. Anschließend die Matrix mit dem Gaußschen Eliminationsverfahren in eine äquivalente Matrix umformen und den **Rang** (Die Anzahl der Zeilenvektoren, die ungleich 0 sind, ist der Rang der Matrix.) bestimmen.

Je nachdem, wie groß der Rang im Gegensatz zur Anzahl der Zeilen ist, so viele Variablen einführen. (z.B: $r = 1, n = 3 \Rightarrow 2$ Variablen einführen.)

ACHTUNG: Matrix-Vektor-Multiplikation

Anschließend die ausmultiplizierten Zeilen jeweils jede Zeile als eigene Gleichung betrachten und das Gleichungssystem lösen.

- 3) Diagonalmatrix D ist dann:

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

- 4) Eigenvektormatrix berechnen: $T = \begin{pmatrix} \vec{e}_1(\lambda_1) & \dots & \vec{e}_n(\lambda_n) \end{pmatrix}$ wobei $\vec{e}_x(\lambda_x)$...Eigenvektor zu λ_x

B.1.4. Matrix-Vector-Multiplikation

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a \cdot x + b \cdot y + c \cdot z \\ d \cdot x + e \cdot y + f \cdot z \end{pmatrix}$$

B.1.5. Matrix-Matrix-Multiplikation

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \cdot \begin{pmatrix} u & x \\ v & y \\ w & z \end{pmatrix} = \begin{pmatrix} a \cdot u + b \cdot v + c \cdot w & a \cdot x + b \cdot y + c \cdot z \\ d \cdot u + e \cdot v + f \cdot w & d \cdot x + e \cdot y + f \cdot z \end{pmatrix}$$

B.1.6. Einheitsmatrix

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

B.1.7. Inverse Matrix

Definition B.1 Für die Einheitsmatrix I gilt:

$$A \cdot A^{-1} = I$$

Beispiel B.2 Bestimmen der Inversen Matrix:

Aus

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

soll die inverse Matrix A^{-1} bestimmt werden:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot A^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Wir versuchen nun mittels geeigneter elementarer Zeilen- und Spaltenumformungen (siehe Anhang B.1.8, jedoch ohne dem Vertauscheniii) die Einheitsmatrix auf die linke Seite zu bekommen.

$$\begin{aligned} \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array} \right) &\xrightarrow{Z_2=Z_2-3 \cdot Z_1} \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array} \right) \xrightarrow{Z_1=Z_1+Z_2} \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & -2 & -3 & 1 \end{array} \right) \xrightarrow{Z_2=\frac{Z_2}{-2}} \\ &\Rightarrow \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array} \right) \Rightarrow A^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \end{aligned}$$

B.1.8. Elementare Zeilen/Spaltenoperationen

- i) Multiplikation einer Spalte/Zeile a_i mit einem Skalar $\lambda \in K \setminus \{0\}$.
- ii) Addieren eines Vielfachen einer Spalte/Zeile a_i zu einer Spalte a_j .
- iii) Vertauschen zweier Spalten/Zeilen a_i, a_j .

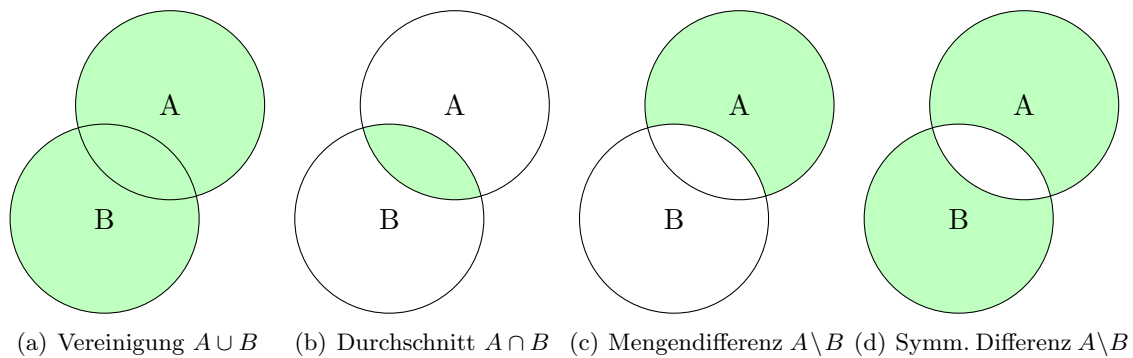


Abbildung B.1.: Die einzelnen Mengenoperationen

B.2. Quadratische Lösungsformel

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

B.3. Mengenlehre

Vereinigung $A \cup B$ (OR): Umfasst alle Elemente, die in A oder B vorkommen. (siehe Abbildung B.1(a))

Durchschnitt $A \cap B$ (AND): Umfasst nur die Elemente, die in A und B vorkommen. (siehe Abbildung B.1(b))

Mengendifferenz $A \setminus B$: Umfasst alle Elemente von A , die nicht in B liegen, also A ohne B . (siehe Abbildung B.1(c))

Symmetrische Differenz $A \Delta B$: Umfasst alle Elemente die in A und B , jedoch nicht im Durchschnitt von A und B liegen. (siehe Abbildung B.1(d))

Komplement A' (NOT): Umfasst alle Elemente, die nicht in A liegen.

Mächtigkeit/Kardinalität $|A|$: Anzahl der Elemente einer Menge.

B.4. Logarithmus-Rechenregeln

B.4.1. Produkte

$$\log_b(x \cdot y) = \log_b(x) + \log_b(y) \Rightarrow \log_b \prod_{i=1}^n x_i = \sum_{i=1}^n \log_b x_i$$

B.4.2. Quotienten

$$\log_b \left(\frac{x}{y} \right) = \log_b(x) - \log_b(y)$$

B.4.3. Potenzen

$$\log_b(x^r) = r \cdot \log_b(x)$$

B.4.4. Basisumrechnung

$$\log_b x = \frac{\log_a x}{\log_a b}$$

B.5. Ungleichungen

B.5.1. Umkehrbarkeit

$$(T_1 \leq T_2) \Leftrightarrow (T_2 \geq T_1)$$

B.5.2. Addition+Subtraktion

Wenn $T_1 < T_2$ dann gilt: $T_1 + T_3 < T_2 + T_3$

Wenn $T_1 < T_2$ dann gilt: $T_1 - T_3 < T_2 - T_3$

B.5.3. Multiplikation+Division

Aus $T_1 < T_2$ folgt: $-T_1 > -T_2$

Aus $T_1 < T_2$ folgt: $0 < \frac{1}{T_2} < \frac{1}{T_1}$

Aus $T_3 > 0$ und $T_1 < T_2$ folgt: $T_1 T_3 < T_2 T_3$ und $\frac{T_1}{T_3} < \frac{T_2}{T_3}$

Aus $T_3 < 0$ und $T_1 < T_2$ folgt: $T_1 T_3 > T_2 T_3$ und $\frac{T_1}{T_3} > \frac{T_2}{T_3}$

Satz B.3 Bei Punktrechnung mit Zahl $z > 0$ ($z \in \mathbb{R}$) bleiben die Vergleichszeichen erhalten, während sie sich bei Punktrechnung mit Zahl $z < 0$ ($z \in \mathbb{R}$) umkehren.

B.5.4. Potenzieren

$$T_1(x) \leq T_2(x) \Leftrightarrow (T_1(x))^a \leq (T_2(x))^a \quad \forall a > 0$$

$$T_1(x) \leq T_2(x) \Leftrightarrow (T_1(x))^a \geq (T_2(x))^a \quad \forall a < 0$$

B.6. Differenzengleichungen

B.6.1. lineare Differenzengleichungen 2.Ordnung

Haben die Form:

$$x_{n+2} + ax_{n+1} + bx_n = s_n \quad \forall n \in \mathbb{N}$$

Folgende Schritte sind notwendig:

- 1) Bestimmung der allgemeinen Lösung $x_n^{(h)}$ der homogenen Gleichung:
Mit Ansatz $x_n^{(h)} = \lambda^n$ eingesetzt: $\lambda^{n+2} + a\lambda^{n+1} + b\lambda^n = 0$ erhält man die charakteristische Gleichung:

$$\lambda^2 + a\lambda + b = 0$$

Löst man diese mit der Quadratischen Lösungsformel muss man 3 Fälle unterscheiden:

$$x_n^{(h)} = \begin{cases} C_1 \lambda_1^n + C_2 \lambda_2^n & \text{falls } \lambda_1 \neq \lambda_2 \text{ reell} \\ r^n (C_1 \cos n\varphi + C_2 \sin n\varphi) & \text{falls } \lambda_{1,2} = r(\cos \varphi \pm i \sin \varphi) \text{ konjugiert komplex} \\ (C_1 + C_2 n) \lambda_1^n & \text{falls } \lambda_1 = \lambda_2 \text{ reell} \end{cases}$$

Die Berechnung der Konstanten C_1 bzw. C_2 kann nun mit den Anfangsbedingungen erfolgen.

- 2) Bestimmung der partikulären Lösung $x_n^{(p)}$. Sollte s_n nicht 0 sein, muss man die partikuläre Lösung bestimmen:

Störfunktion s_n	Versuchslösung $x_n^{(p)}$
1	A
r^n	$A r^n$
$\sin(rn)$ oder $\cos(rn)$	$A \sin(rn) + B \cos(rn)$
n^k (oder Polynom vom Grad k)	$A_0 + A_1 n + A_2 n^2 + \dots + A_k n^k$
$n^k \cdot r^n$	$(A_0 + A_1 n + A_2 n^2 + \dots + A_k n^k) r^n$

- 3) Ermittlung der Lösungsgesamtheit gemäß $x_n = x_n^{(h)} + x_n^{(p)}$

B.7. Binomischer Lehrsatz

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

B.8. Kombinatorik

B.8.1. Permutationen(Anordnungen)

Anordnungen von n -Objekten, Reihenfolge wichtig.

Ohne Wiederholung

Anordnung von n Objekten, die alle unterscheidbar sind. Anfangs n -Elemente, dann $n-1$, $n-2$, usw. zur Auswahl.

$$n! = n \cdot (n-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1$$

Beispiel B.3 Beispielsweise gibt es $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$ mögliche Anordnungen von vier verschiedenfarbigen Kugeln in einer Reihe.

Mit Wiederholung

i -tes Objekt n_i -fach wiederholt:

Anordnung von n Objekten, manche nicht unterscheidbar. Sind k -Objekte ident, sind diese auf den Plätzen vertauschbar, ohne neue Reihenfolge.

$$\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

Beispiel B.4 Gibt es $\frac{4!}{2!1!1!} = \frac{24}{2} = 12$ mögliche Anordnungen von vier farbigen Kugeln in einer Reihe, wenn genau zwei der Kugeln die gleiche Farbe aufweisen, und $\frac{4!}{2!2!} = \frac{24}{4} = 6$ mögliche Anordnungen, wenn jeweils zwei Kugeln gleichfarbig sind.

B.8.2. Kombinationen

Auswahlen von k aus n unterscheidbaren Objekten, Reihenfolge unwichtig:

Ohne Wiederholung

auch bekannt als **Binomialkoeffizient**.

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

Beispiel B.5 Lotto! Wenn aus 45 Zahlen nur 6 ohne Wiederholung und ohne Beachtung der Reihenfolge ausgewählt werden.

Mit Wiederholung

$$\frac{(n+k-1)!}{k!(n-1)!} = \binom{n-1+k}{k}$$

Beispiel B.6 Aus einer Urne mit fünf nummerierten Kugeln ($n = 5$) wird dreimal eine Kugel gezogen ($k = 3$) und jeweils wieder zurückgelegt. Man kann also bei allen drei Ziehungen immer aus fünf Kugeln auswählen. Wenn man die Reihenfolge der gezogenen Zahlen nicht berücksichtigt, gibt es

$$\binom{\binom{5}{3}}{3} = \binom{10}{3} = \frac{7!}{4!3!} = 35$$

verschiedene Kombinationen.

B.8.3. Variationen

Auswahlen von k aus n unterscheidbaren Objekten, Reihenfolge ist wichtig.

Ohne Wiederholung

Bei einer Variation ohne Wiederholung sollen k von n Objekten (mit $k \leq n$) auf k verfügbare Plätze platziert werden, wobei jedes Objekt nur höchstens einen Platz einnehmen darf.

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

Beispiel B.7 Wenn aus einer Urne mit fünf verschiedenen Kugeln dreimal ohne Zurücklegen gezogen wird, sind $5 \cdot 4 \cdot 3 = 60$ verschiedene Auswahlen möglich: bei

der ersten Ziehung noch fünf Möglichkeiten, dann nur noch vier und für die dritte Ziehung schließlich nur noch drei Möglichkeiten.

Mit Wiederholung

Bei einer Variation mit Wiederholung werden aus n Objekten k Objekte unter Beachtung der Reihenfolge ausgewählt, wobei Objekte auch mehrfach ausgewählt werden können.

$$n^k$$

Beispiel B.8 *Wenn aus einer Urne mit fünf verschiedenen Kugeln dreimal mit Zurücklegen gezogen wird, dann sind $5 \cdot 5 \cdot 5 = 5^3 = 125$ verschiedene Auswahlen möglich.*

C. Glossar

Symbolverzeichnis

α	Wahrscheinlichkeit des α -Quantils 67
α	Wahrscheinlichkeit für einen Fehler 1.Art 73
$\underline{\alpha}$	Anfangsverteilung 48
β	Wahrscheinlichkeit für einen Fehler 2.Art 73
\mathbb{E}	Erwartungswert 26
\emptyset	Die leere Menge 5
γ	Überdeckungswahrscheinlichkeit eines Konfidenzintervalls 68
H	$H(P)$ Entropie 81
H^*	$H^*(P)$ Mittlere Unbestimmtheit 82
I	$I_x = I(p_x)$ Informationsgehalt 81
i.i.d.	independent and identically distributed 30
m	Anzahl der Blätter im Baum/die Anzahl der Zufallsvariablen 82
μ	Erwartungswert 26
Ω	Grundgesamtheit (auch Grundmenge M), die Menge ihrer Untersuchungseinheiten ω . 55
ω	Untersuchungseinheit (Element) der Grundgesamtheit Ω 55
π	stationäre Verteilung 48
\mathbb{P}	Wahrscheinlichkeit 5
s_n^2	(korrigierte) Stichprobenvarianz 28
Φ	Verteilungsfunktion der Standardnormalverteilung 44
φ	Dichte der Standardnormalverteilung 44
Θ	Der Parameterraum $\theta \in \Theta$. 55
θ	Der Parameter eines parametrischen Modells (kann Mehrdimensional sein) $\theta \in \Theta$ 55
$\hat{\theta}_n$	Schätzer, eine Folge von Statistiken 58
\mathbb{V}	Varianz 28
z_α	α -Quantil der Standardnormalverteilung 67

Literaturverzeichnis

- [1] Wikipedia. Tschebyscheff-ungleichung — wikipedia, die freie enzyklopädie, 2013. [Online; Stand 8. März 2014].
- [2] Wikipedia. Binomialverteilung — wikipedia, die freie enzyklopädie, 2014. [Online; Stand 8. März 2014].
- [3] Wikipedia. Stochastische unabhängigkeit — wikipedia, die freie enzyklopädie, 2014. [Online; Stand 7. März 2014].

Index

- absorbierender Zustand, 50
- Absorptionswahrscheinlichkeit, 52
- Absorptionszeit
 - mittlere -, siehe mittlere Absorptionszeit
- Additionstheorem, 9
- Algorithmus
 - Huffman, siehe Huffman Algorithmus
- Alternative, siehe Gegenhypothese
- Anfangsverteilung, 48
- Annahmebereich, 72
- Anordnung, 102
- Anpassungstests, 76–79
- Ausprägung, 55
- Axiome von Kolmogorov, 5–10
 - weitere Eigenschaften, 8
- Bayes
 - Satz, siehe Satz von Bayes
- Bedingte Entropie, 84
- Bedingte Wahrscheinlichkeit, 11–12
 - stetig, 23
- Betaverteilung (1.Art), 46
- Betaverteilung (2.Art), 46
- Bias, 58
- Bildmaß, 19
- Binär
 - von Dezimal, 87
- Binomialkoeffizient, 103
- Binomialverteilung, 33, 46
- Binomischer Lehrsatz, 102
- Blattlänge
 - Mittlere-, siehe Mittlere Unbestimmtheit
- Bonferroni Ungleichung, siehe Ungleichung von Bonferroni
- Cauchy-Verteilung, 43
- Cauchyverteilung, 46
- Chapman-Kolmogorov Gleichungen, 49
- Chi-Quadrat-Test, 77
- Code
 - Definition, 86
- Codes, 86–88
- Codewortlänge
 - Mittlere-, siehe Mittlere Unbestimmtheit
- Cramér-Rao, Satz, 64
- Determinante, 97
 - 2x2-Matrix, 97
 - 3x3-Matrix, 97
 - Rechenregeln, 97
- Dezimal in Binär, 87
- Diagonalmatrix, 98
- Dichte, 23
- Dichten
 - Faltung, 26
- Differenz
 - Mengen-, siehe Mengendifferenz
 - Symmetrische-, siehe Symmetrische Differenz
- Differenzengleichungen
 - lineare 2.Ordnung, 101
- Differenzengleichungen, 101
- disjunkte Mengen, 6, 11
 - nicht disjunkte Mengen, 11
- Diskrete Verteilung, 33–39
- diskrete Verteilungsfunktion, 19
- diskrete Wahrscheinlichkeitsverteilung, 21
- Diskrete Zufallsvariable, 19–22
- Dualität Konfidenzintervall/Tests, 80
- Durchschnitt, 100
- Effizienz, 58
- Eigenschaften des Erwartungswertes, 27
- eindeutig entzifferbar, 88
- einfache Hypothese, 72
- Einheitsmatrix, 99
- einseitige Hypothese, 72
- Elementare Zeilen/Spaltenoperationen, 99
- Elementarereignis, 5
- empirisches Moment, 59
- Entropie, 81

-
- Bedingte-, siehe Bedingte Entropie der Quelle, 88
 - gemeinsame-, siehe gemeinsame Entropie
 - entzifferbar
 - eindeutig (unendlich), siehe eindeutig entzifferbar
 - ergodisch, 88
 - Erwartungstreue, 58
 - Erwartungswert, 26–28
 - Eigenschaften, siehe Eigenschaften des Erwartungswertes
 - Exponentialverteilung, 40, 46
 - Faltung von Dichten, 26
 - Fanu-Code, 87
 - Fehler
 1. Art, 73
 2. Art, 73
 - Fisher-Information, 64
 - Folgen von Zufallsvariablen, 30–33
 - Gammafunktion, 42
 - Gammaverteilung, 42, 46
 - gedächtnislos, 88
 - Gegenhypothese, 72
 - gemeinsame Entropie, 83, 84
 - Gemeinsame Verteilung
 - diskreter Zufallsvariablen, 21
 - Gemeinsame Verteilung
 - stetiger Zufallsvariablen, 23
 - Gemischte Zufallsvariable, 24
 - Geometrische Verteilung, 35, 46
 - Gesetz
 - Schwaches G. der gr. Z., siehe Schwaches G. der gr. Z.
 - Starkes G. der gr. Z., siehe Starkes G. der gr. Z.
 - Gleichverteilung, 46
 - diskret, 34
 - stetig, 39
 - Grenzwertsatz
 - Zentraler-, siehe Zentraler Grenzwertsatz
 - Grundgesamtheit, 55
 - Grundmenge, 5
 - Huffman-Algorithmus, 83
 - Huffman-Code, 86
 - Hypergeometrische Verteilung, 38, 46
 - Hypothese, 71
 - einfache, siehe einfache Hypothese
 - einseitige, siehe einseitige Hypothese
 - Gegen-, siehe Gegenhypothese
 - Null-, siehe Nullhypothese
 - zusammengesetzte, siehe zusammengesetzte Hypothese
 - zweiseitige, siehe zweiseitige Hypothese
 - Information, 84
 - Information und Entropie, 81–85
 - Informationsgehalt, 81
 - Informationsquelle, 88
 - Informationsquellen, 88–89
 - Informationstheorie, 81–89
 - Intervallschätzung, 57, 66–71
 - Inverse Matrix, 99
 - irreduzibel, 50
 - Kardinalität, 100
 - Klasseneigenschaft, 50
 - Kolmogorov Axiome, siehe Axiome von Kolmogorov
 - Kolmogorov-Chapman Gleichungen, siehe Chapman-Kolmogorov Gleichungen
 - Kombinationen, 103
 - Kombinatorik, 102
 - kommunizierend, 50
 - Konfidenzintervall, 68
 - Dualität mit Tests, siehe Dualität Konfidenzintervall/Tests
 - für μ , bekanntes σ^2 von \mathcal{N} , 69
 - für μ , unbekanntes σ^2 von \mathcal{N} , 69
 - für σ^2 von \mathcal{N} , 71
 - für Anteilswerte, 71
 - Tests, 79
 - Konsistenz, 58
 - Korrigierte Stichprobenvarianz, 28
 - Kovarianz, 29
 - Kraft Ungleichung, siehe Ungleichung von Kraft
 - Kullback-Leibler-Divergenz, 85
 - Laplace-Experiment, siehe Laplacescher Wahrscheinlichkeitsraum
 - Laplacescher Wahrscheinlichkeitsraum, 8
 - Leere Menge, 5
 - Likelihood-Funktion, 61

- Log-, siehe Log-Likelihood-Funktion
- Multiplikationssatz, siehe Multiplikationssatz für Likelihood-Funktionen
- Likelihood-Prinzip, 61
- Likelihoodquotientenstatistik, siehe Likelihoodquotiententest
- Likelihoodquotiententest, 74
 - Randomisierter, siehe Randomisierter Likelihoodquotiententest
- Log-Likelihood-Funktion, 61
- Logarithmus-Rechenregeln, 100
 - Basisumrechnung, 101
 - Potenzen, 100
 - Produkte, 100
 - Quotienten, 100
- Mächtigkeit, 100
- Markoveigenschaft, 47
- Matrix, 97
 - potenz, 98
 - Determinante, siehe Determinante
 - Diagonal, siehe Diagonalmatrix
 - Einheits-, siehe Einheitsmatrix
 - Inverse-, siehe Inverse Matrix
 - Rang, siehe Rang
 - Transformations-, siehe Transformationsmatrix
- Matrix-Matrix-Multiplikation, 99
- Matrix-Vector-Multiplikation, 99
- Matrizen, 97
- Maximum Likelihood Methode, 60–64
- Maximum-Likelihood-Schätzer, 62
- Mengendifferenz, 100
- Mengenlehre, 100
- Merkmal, 55
- Mittelwert, siehe Stichprobenmittel
- mittlere Absorptionszeit, 52
- Mittlere Blattlänge, siehe Mittlere Unbestimmtheit
- Mittlere Codewortlänge, siehe Mittlere Unbestimmtheit
- Mittlere Unbestimmtheit, 82
- ML-Schätzer, siehe Maximum-Likelihood-Schätzer
- Modell
 - nicht parametrisches, siehe nicht parametrisches Modell
 - parametrisches, siehe parametrisches Modell
 - statistisches, siehe statistisches Modell
- Moment, 59
 - empirisches, 59
- Momentenmethode, 59
- Multiplikationssatz, 12
- Multiplikationssatz für Likelihood-Funktionen, 61
- Multiplikationstheorem, siehe Multiplikationssatz
- Nachfolger, 50
- Neyman Pearson-Satz, 75
- nicht parametrisches Modell, 56
- Niveau, 73
- Normalverteilung, 44, 46
 - Quantil, siehe Quantil, (Standard-) Normalverteilung
 - Standard-, siehe Standardnormalverteilung
 - transformation Standardnormalverteilung, 44
- Nullhypothese, 72
- Nullrekurrent, 51
- Obermenge, 6
- optimaler Test, 73
- Paarweise Unabhängigkeit, 14
- parametrisches Modell, 55
- Periode, 50
- Permutationen, 102
- Poissonverteilung, 36, 46
- positiv Rekurrent, 51
- Präfixfrei, 86
- Punktschätzung, 57–66
- Quadratische Lösungsformel, 100
- Quantil, (Standard-)Normalverteilung, 67
- Quelle
 - Entropie der-, siehe Entropie der Quelle
- Randomisierter Likelihoodquotiententest, 74
- Rang, 98
- Rekurrenz, 51
 - null-, siehe Nullrekurrent
 - positiv-, siehe positiv Rekurrent
- robuster Schätzer, 66
- Rückkehrzeit, 51

- Satz
 Neyman Pearson, siehe Neyman Pearson-Satz
 Satz des unachtsamen Statistiker, 28
 Satz von Bayes, 15
 Satz von Cramér-Rao, siehe Cramér-Rao-Satz
 Schätzer
 Effizienz, siehe Effizienz
 Eigenschaften, 58
 Erwartungstreue, siehe Erwartungstreue
 Intervall-, siehe Intervallschätzung
 Konsistenz, siehe Konsistenz
 Maximum-Likelihood-, siehe Maximum-Likelihood-Schätzer
 Momentenschätzer, siehe Momentenmethode
 Punkt-, siehe Punktschätzung
 robust, siehe robuste Schätzer
 Unverzerrtheit, siehe Unverzerrtheit
 Schätztheorie, 57
 Schwaches G. der gr. Z., 31
 Shannon-Code, 87
 Siebformel, 9
 σ -Algebra, 6
 Signifikanzniveau, siehe Niveau
 Spezielle Tests, 75
 Standardisierte Zufallsvariable, 32
 Standardnormalverteilung, 44
 Quantil, siehe Quantil, (Standard-)Normalverteilung
 transformation Normalverteilung, 44
 Starkes G. der gr. Z., 32
 stationäre Verteilung, 48, 51, 52
 Statistik, 55–80
 Aufgaben, 57
 Definition, 57
 statistisches Modell, 55
 Steinerscher Verschiebungssatz, 29
 Stetige Verteilung, 39–44
 stetige Verteilungsfunktion, 22
 Stetige Zufallsvariable, 22–24
 Stichprobe, 56
 Stichprobenmittel, 56
 Stichprobenumfang, 56
 Stichprobenvarianz
 korrigiert, siehe Korrigierte Stichprobenvarianz
 mit ML-Schätzer, 28
 stochastische Unabhängigkeit, 13–16
 Student-T-Verteilung, 69
 suffizient, 65
 Symmetrische Differenz, 100
 T-Verteilung, Students-, siehe Student-T-Verteilung
 Test
 Likelihoodquotienten-, siehe Likelihoodquotiententest
 Tests, 71–80
 Anpassungs-, siehe Anpassungstests
 Chi-Quadrat-, 77
 Dualität mit Konfidenzintervall, siehe Dualität Konfidenzintervall/Tests
 für μ , σ^2 bekannt von \mathcal{N} , 75
 für μ , σ^2 unbekannt von \mathcal{N} , 75
 für σ von \mathcal{N} , 76
 Für Anteilswerte, 76
 Likelihoodquotienten-, siehe Likelihoodquotiententest
 Spezielle, siehe Spezielle Tests
 Teststatistik, 57, 73
 Totale Unabhängigkeit, 14
 Totale Wahrscheinlichkeit, siehe Vollständige Wahrscheinlichkeit
 Transformationsmatrix, 98
 Transformationssatz für Dichten, 23
 Überdeckungswahrscheinlichkeit, 68
 Übergangsmatrix, 50
 Übergangswahrscheinlichkeit, 47, 49
 Übergangszeit, 51
 Unabhängigkeit
 Paarweise-, siehe Paarweise Unabhängigkeit
 stochastische, siehe stochastische Unabhängigkeit
 Totale-, siehe Totale Unabhängigkeit
 Verteilung, 25
 Unabhängigkeit von Zufallsvariablen, 25–26
 unachtsamer Statistiker
 Satz-, siehe Satz des unachtsamen Statistiker
 Unbestimmtheit
 Mittlere-, siehe Mittlere Unbestimmtheit
 Ungleichung

- von Bonferroni, 12
- von Chebychev, 30
- von Kolmogorov, 30
- von Kraft, 82
- von Markov, 30
- Ungleichungen, 101
 - Addition/Subtraktion, 101
 - Multiplikation/Division, 101
 - Potenzieren, 101
 - Umkehrbarkeit, 101
- Untersuchungseinheit, 55
- Unverzerrt, 58
- Varianz, 28–30
 - Eigenschaften, 29
- Variationen, 103
- verbunden, 50
- Vereinigung, 100
- Verschiebungssatz
 - Steiner-, siehe Steinerscher Verschiebungssatz
- Verteilung
 - Binomial-, siehe Binomialverteilung
 - Cauchy-, siehe Cauchyverteilung
 - Diskret, siehe Diskrete Verteilung
 - Exponential-, siehe Exponentialverteilung
 - Gamma-, siehe Gammaverteilung
 - Geometrische-, siehe Geometrische Verteilung
 - Hypergeometrische-, siehe Hypergeometrische Verteilung
 - Normal-, siehe Normalverteilung
 - Poisson-, siehe Poissonverteilung
 - Standardnormal-, siehe Standardnormalverteilung
 - Stetig, siehe Stetige Verteilung
 - Zusammenfassung, 46
- Verteilungsfunktion
 - Eigenschaften, 24–25
 - diskret, siehe diskrete Verteilungsfunktion
 - stetig, siehe stetige Verteilungsfunktion
 - Wahrscheinlichkeiten, siehe Wahrscheinlichkeiten mit Verteilungsfunktion
- Verwerfungsbereich, 73
- Verzerrung, 58
- Vollständige Wahrscheinlichkeit, 15
- Wahrscheinlichkeit, 7
 - Bedingte -, siehe Bedingte Wahrscheinlichkeit
 - Vollständige-, siehe Vollständige Wahrscheinlichkeit
- Wahrscheinlichkeiten mit Verteilungsfunktion, 25
- Wahrscheinlichkeitsmaß, 7, 19
- Wahrscheinlichkeitsraum, 7
- Wahrscheinlichkeitstheorie, 5–16
- Wahrscheinlichkeitsverteilung, 48
 - diskret, siehe diskrete Wahrscheinlichkeitsverteilung
- Zentraler Grenzwertsatz, 32–33
- Zufallsvariable, 17–46
 - Unabhängigkeit, siehe Unabhängigkeit von Zufallsvariablen
 - Beschreibung, 17–18
 - diskret, siehe Diskrete Zufallsvariable
 - Folgen, siehe Folgen von Zufallsvariablen
 - gemischt, siehe Gemischte Zufallsvariable
 - Mathematische Definition, 18–19
 - standardisierte, 32
 - stetig, siehe Stetige Zufallsvariable
 - zusammengesetzte Hypothese, 72
- Zustand, absorbierend, siehe absorbierender Zustand
- zweiseitige Hypothese, 72