Rémi Georgiou

ZhAW | HS 2015

Mathematik Stochastik

MANIT3 – Zusammenfassung

Inhaltsverzeichnis

[Deskriptive Statistik 3](#_Toc441649082)

[Skalierung (Messniveau) 3](#_Toc441649083)

[Kategorielle Daten 3](#_Toc441649084)

[Stetige Merkmale, Dichtebegriff 4](#_Toc441649085)

[Empirische Verteilungsfunktion 5](#_Toc441649086)

[Quantile 6](#_Toc441649087)

[Klassierte Daten 6](#_Toc441649088)

[Boxplot-Darstellung 7](#_Toc441649089)

[Streumasse 8](#_Toc441649090)

[Lineare Regression 10](#_Toc441649091)

[Bestimmen der Geradengleichung 11](#_Toc441649092)

[Varianz-Zerlegung und Bestimmtheitsmass 13](#_Toc441649093)

[Kombinatorische Formeln 15](#_Toc441649094)

[Produktregel 15](#_Toc441649095)

[Anordnungen 15](#_Toc441649096)

[Permutation (Vertauschung) – Anordnungsproblem ohne Wiederholungen 15](#_Toc441649097)

[Permutation (Vertauschung) – Anordnungsproblem mit Wiederholungen 15](#_Toc441649098)

[Anwendung bei Binomial-Koeffizienten 15](#_Toc441649099)

[Auswahlen 16](#_Toc441649100)

[Übersicht 17](#_Toc441649101)

[Wahrscheinlichkeiten 17](#_Toc441649102)

[Klassische Definition nach Laplace 17](#_Toc441649103)

[Axiome der Wahrscheinlichkeit 19](#_Toc441649104)

[Aussagenlogik 19](#_Toc441649105)

[Mehrstufige Versuche 20](#_Toc441649106)

[Ereignisbaum 20](#_Toc441649107)

[Anwendung 1: Zuverlässigkeit von (komplexen) Systemen 22](#_Toc441649108)

[Anwendung 2: Kryptographische Hashfunktion, Kollisionswahrscheinlichkeit 23](#_Toc441649109)

[Bedingte Wahrscheinlichkeit 24](#_Toc441649110)

[Verteilungen 25](#_Toc441649111)

[Zufallsvariable 25](#_Toc441649112)

[Diskrete und stetige Zufallsvariablen 25](#_Toc441649113)

[Verteilungsfunktion 26](#_Toc441649114)

[Hypergeometrische Verteilung 26](#_Toc441649115)

[Binomialverteilung 27](#_Toc441649116)

[Geometrische Verteilung 29](#_Toc441649117)

[Vergleich der Binomialverteilung mit der Hypergeometrischen 29](#_Toc441649118)

[Poisson-Verteilung 30](#_Toc441649119)

[Grenzmodell 30](#_Toc441649120)

[Schätzen 32](#_Toc441649121)

[Punktschätzung 32](#_Toc441649122)

[Intervallschätzung 32](#_Toc441649123)

[Vertrauensintervall 33](#_Toc441649124)

# Deskriptive Statistik

## Skalierung (Messniveau)

|  |  |  |
| --- | --- | --- |
| Messniveau | Beschreibung | Beispiele |
| Nominal | reine Kategorisierung  keine Grössenordnung  keine Unterschiedsmessung | Wohnort,  Studiengang: IT, ET, ST  Telefonnummern |
| Ordinal | Grössenordnung vorhanden, Rangierung möglich  Unterschiedsmessung ist nicht messbar | Kleidergrösse: s, m, l  Qualifikation: genügend, gut, … |
| Metrisch  (intervall skaliert)   1. Metrisch stetig 2. Metrisch diskret | Zahlenskala, insbesondere Grössenunterschiede messbar  Grössen sind beliebig genau messbar  Es sind nur bestimmte Werte vorgesehen | Gewicht, Einkommen, Anzahl  V(t): 574.8 km/h  Mathenote, Kinderzahl |

Nominal und Ordinal wird zu Kategoriell zusammengefasst.

**Graphische Darstellung der Daten:**

Kategoriell 🡪 Balkendiagramm

Metrisch 🡪 Klassenbildung, Histogramm

## Kategorielle Daten

|  |  |
| --- | --- |
| : Anzahl | : relative Häufigkeit |
| : Merkmalswert | : kumulierte Häufigkeit (absolut) |
| : Häufigkeit (absolut) | : kumulierte relative Häufigkeit |

|  |  |
| --- | --- |
|  | Darstellung der Häufigkeitsverteilung mit einer Häufigkeitstabelle und einem **Stabdiagramm**. Häufigkeiten oder . |

Die kumulierten Häufigkeiten werden als (empirische) **Summenkurve**  dargestellt. 🡪 unstetige Sprungfunktion

|  |  |
| --- | --- |
|  | Matlab:  s = cumsum(hi)/sum(x) |

## Stetige Merkmale, Dichtebegriff

Einteilung der Daten in **Klassen** (-Werte von … bis).

|  |  |
| --- | --- |
|  | Die Häufigkeit h lässt sich **nicht als Stab** über der Klasse darstellen, sondern man stellt die Häufigkeit als Rechteckfläche über der Klassenbreite dar: Die Höhe des Rechtecks ist die (Häufigkeits-) **Dichte D**.  Die Dichte ist eine spezifische Häufigkeit, sie gibt die Häufigkeit pro Einheit von an. |

Absolute Häufigkeit im Diagramm eingezeichnet.

|  |  |
| --- | --- |
|  | pro Einheit von in dieser Klasse |

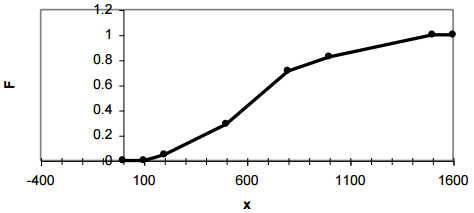
|  |  |
| --- | --- |
| Absolute Dichte | Relative Dichte |

Alle Rechtecke zusammen bilden das Histogramm:

|  |  |
| --- | --- |
|  | MATLAB:  bar(x-edges,y-values,‘histc‘); |

|  |
| --- |
| * Das Histogramm ist ein **Flächendiagramm**. Die Flächen sind die Häufigkeiten. Die Rechtecke werden nahtlos aneinander gefügt. * Bei äquidistanten Klassenbreiten sind die Dichten proportional zu den Häufigkeiten. |

Die **Summenkurve (Verteilungsfunktion)** stellt die Summenhäufigkeiten dar. Sie ist eine stückweise lineare Funktion 🡪 innerhalb der Klasse steigt sie um den Betrag .



MATLAB:  
x = 0:length(hi);  
s = cumsum(hi)/sum(x);  
stairs(x,s);

## Empirische Verteilungsfunktion

Stellt die **relative** kumulierte Summenhäufigkeit dar. In MATLAB mit cdfplot(x) realisierbar. Besonders wichtig, weil man Teilhäufigkeiten darin ablesen kann und mit wenig Aufwand einen Überblick über die Daten bekommt.

|  |  |
| --- | --- |
| **Symmetrisch 🡪 normrnd(10,3,1,150);** | **Asymmetrisch 🡪 exprnd(10,1,150);** |
| Symmetrische Verteilung Etwa 50 % (Median liegen unter 10. | Schiefe Verteilung  Median |
| Median 10 (x-Wert) | Median 6 (x-Wert) |

Der Median ist der **Mittelwert einer Verteilung** und teilt einen Datensatz in zwei Hälften.

## Quantile

Das -Quantil mit ist der Messwert , für den gilt. Es wird als bezeichnet.

Quartile:

1. Quartil 2. Quartil **Median** 3. Quartil

Zur Bestimmung des -Quantils werden die Werte aufsteigend sortiert. Die zu zugehörige Ordnungsnummer ist ungefähr . ist aber nicht immer ganzzahlig!

|  |
| --- |
| Definition -Quantil:   * ganzzahlig: * nicht ganzzahlig ganzzahlig aufrunden: ceil() |

Beispiel:

### Klassierte Daten

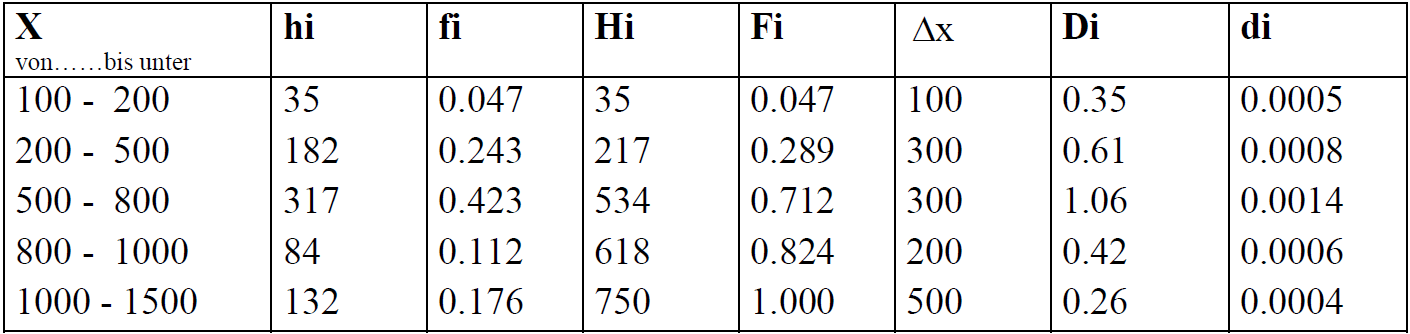
**Median durch Interpolation**

Fi

Bsp:

|  |  |
| --- | --- |
| Dichte  2000\*0.12=240   400-240=160   1260-400=860   1620-1260=360   1840-1620=220   2000-1840=160  hi | **Hi**  2000\*0.12=240 2000\*0.2=400 2000\*0.63=1260 2000\*0.81=1620 2000\*0.92=1840 2000\*1.00=2000 |

Weiteres Beispiel:



**Mittelwert**

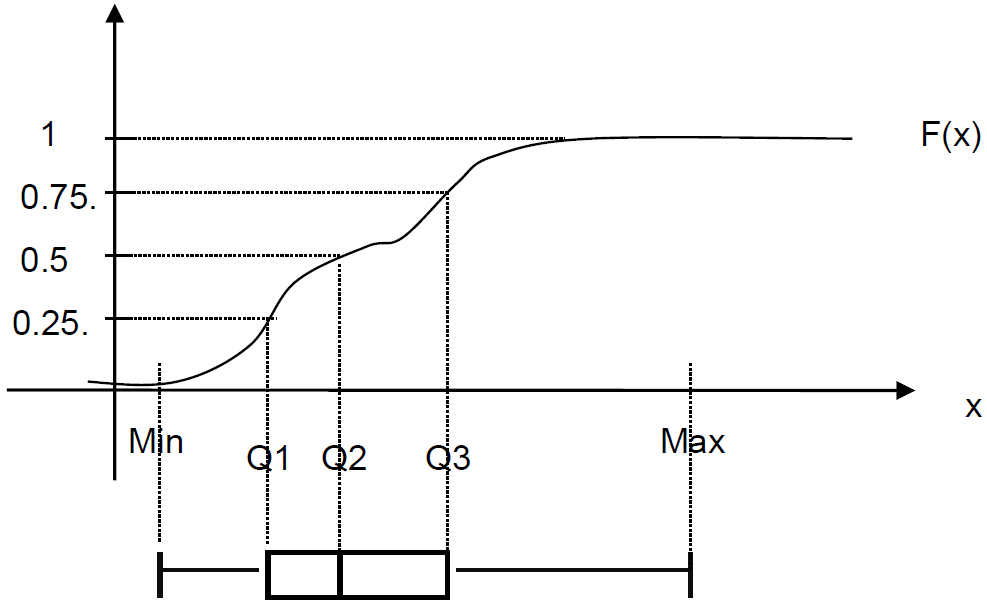
Den Mittelwert berechnet man mit Hilfe der mit den Häufigkeiten gewichteten Klassenmitten:

Als **Modus** kann man die Mitte der Klasse mit der grössten Dichte nehmen, hier: .

### Boxplot-Darstellung

Der Median ist rot dargestellt. Der untere Box-Rand ist das 1. Quartil (). Der obere Box-Rand ist das 3. Quartil ().

|  |  |
| --- | --- |
| **Symmetrisch 🡪 normrnd(10,3,1,150);** | **Asymmetrisch 🡪 exprnd(10,1,150);** |
|  |  |



## Streumasse

|  |
| --- |
| **Quadrierte Summen :** |

|  |
| --- |
| **Stichprobenvarianz (Schätztheorie):**  Bei mehreren Variablen:  Und somit |

|  |
| --- |
| **Standardabweichung :** |

|  |
| --- |
| **Varianz :** |

Für praktische Berechnungen kann man die Umformung der Stichprobenvarianzgleichung verwenden. Herleitung der Formel für :

„n mal Mittelwert“

**Berechnung der Summe (aus Standardabweichung und Mittelwert)**

|  |
| --- |
| **Kovarianz** **:** Ein Zusammenhangsmass zwischen zugeordneten Datenpaaren ()  Sie stellt ein mittleres Abweichungsrechteck vom Punkt dar. |

# Lineare Regression

Wir beobachten zwei Variablen und . Man interessiert sich dafür, ob zwischen den beiden Variablen eine grössenmässige Beziehung besteht und von welcher Art diese ist. Man fasst die -Werte als Abhängige von auf.

z. b. Dichte und Eisen-Gehalt (%)

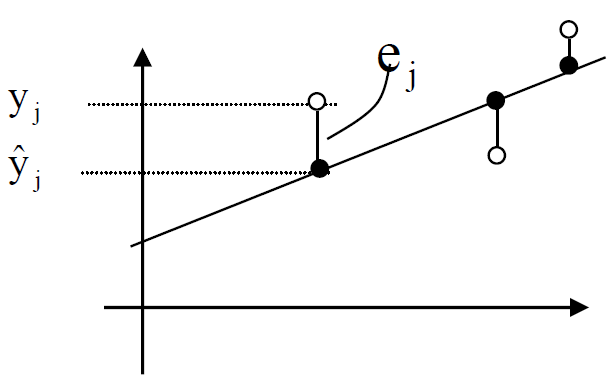
|  |  |
| --- | --- |
|  | Es ist naheliegend einen linearen Zusammenhang zwischen der Dichte und dem Eisen-Gehalt zu vermuten. Das Bestimmen einer besten Geraden nennt **man lineare Regression**. Diese Gerade heisst **Regressionsgerade** (alter Begriff) oder auch **Ausgleichsgerade** (in der Technik gebraucht).  **Regressionsgleichung**  + Zufallsfehler Modell  Mit der Geradengleichung berechnet man ideale y-Werte aus den x-Werten. |

**Messwerte oder beobachtete Werte:**

**Die berechneten Werte werden als erklärte Werte bezeichnet:**

Die Schätzung wird notiert als

Die Differenzen sind **Residuen** (Fehler).



|  |  |
| --- | --- |
|  | Man bestimmt die Gerade so, dass die Summe der **Residuenquadrate** möglichst klein wird: |

In Matrixform:

**Lineares Modell: linear in den Parametern und .**

Gesucht sind Y-Achsenabschnitt und Steigung so, dass

## Bestimmen der Geradengleichung

Für die Geradengleichung gilt es, und so zu bestimmen, dass die Summe

minimal wird.

In einem ersten Schritt fassen wir die Terme in der Klammer zusammen und setzen

und schreiben die Summe neu:

Diese Summe ist minimal, wenn ist. Damit ist .

Mittelwert von

Mittelwert xon

Diese Formel deutet bereits die spezielle Lage der Geraden an, denn es folgt daraus:

Das bedeutet, dass der Punkt (der Schwerpunkt der Punktewolke) die Geradengleichung  
 erfüllt, d.h. liegt auf der Regressionsgeraden.

Die Gleichung der Regressionsgeraden kann man deshalb auch Punktsteigungsform formulieren:

Jetzt muss noch die Steigung berechnet werden.

Der Ausdruck für b wird in die Summe eingesetzt.

🡪 Die rechte Seite ist ein quadratischer Ausdruck in , den man in die Form bringt.

Die Kurve ist eine Parabel, wird am kleinsten im Scheitelpunkt, wir schreiben daher die Scheitelpunktform:

wird minimal, wenn gesetzt wird.

|  |
| --- |
| Für die Gleichung der Regressionsgeraden gilt: |

|  |
| --- |
| Die Summe der Residuenquadrate hat den Wert |

|  |
| --- |
| Regressionsgerade : |

Die **Kovarianz**  kann man über die Steigung bestimmen:

Mittels ***plot(x,y)*** kann in MATLAB eine Scatterplot gezeichnet werden. Die Regressiongerade wird mit dem Befehl ***lsline*** eingezeichnet.

Für die Regression ist eine Matrix X vorzubereiten mit Vektoren als Kolonnen.

n = length(x);

X = [ones(n,1) x‘];

B = regress(y‘, X);

Der Vektor B enthält den Y-Achsenabschnitt und die Steigung .

## Varianz-Zerlegung und Bestimmtheitsmass

Die erklärten -Werte sind definiert durch und die Residuen durch .

Beispiel:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **3** | **5** | **6** | **10** |  |
|  | **5** | **4** | **8** | **11** |  |
|  | 4.11 | 6.04 | 7 | 10.85 |  |
|  | 0.89 | -2.04 | 1 | 0.15 |  |

Offensichtlich gilt: und

Diese Eigenschaft ergibt sich aus der kleinste Quadrate Methode (kQM).

Die Residuenvarianz ist deshalb:

Man erkennt die Zerlegung

|  |
| --- |
| Allgemein gilt die sogenannte **Quadratesummen-Zerlegung**:  Teilt man diese Summe durch , so erhält man die **Varianz-Zerlegung**:  **totale Varianz = erklärte Varianz + Fehler-Varianz** |

Aus diesem Zusammenhang wird eine Vergleichszahl eingeführt, das **Bestimmtheitsmass** .

Das Bestimmtheitsmass ist der Anteil der erklärten Varianz an der gesamten Varianz. bedeutet, dass 73 % der gesamten Varianz durch die Regression erklärt ist, der Rest ist Zufallsstreuung.

Mit Hilfe der Varianzzerlegung und der Minimumformel der Residuenquadrate erhält man zwei Ausdrücke für :

|  |
| --- |
|  |

Bedeutung von bezüglich der Regressionsgeraden.

Dann ist , die Punkte liegen exakt auf der Geraden.

, dies ist dann der Fall, wenn die Regressionsgerade **die Steigung**  hat. Dann ist auch , d.h. die - und -Werte korrelieren nicht.

* lässt sich nicht als Funktion von modellieren.

|  |
| --- |
| **Normierung der Kovarianz :**  heisst **Korrelation** zwischen und (genauer: Produktmomenten-Korrelation nach Pearson) |

Die **Korrelation**  hat dasselbe Vorzeichen wie die Kovarianz und die Steigung der Regressionsgeraden und es gilt respektive .

Die **Kovarianz**  kann man über die Varianzen und die **Korrelation**  bestimmen:

Wenn die Steigung der Regressionsgeraden ist, , dann erklärt nicht !

Berechnung der Residuen:

Bei standardisierten Daten gilt:

# Kombinatorische Formeln

Für die Wahrscheinlichkeitsrechnung kann man die Kombinatorik eingrenzen auf die Fragen nach der Anzahl **Anordnungen** oder der Anzahl **Auswahlen**. Dabei kann geht man von einem Grundprinzip aus, der sog. **Produktregel**.

## Produktregel

Das Produkt der Möglichkeiten pro Stufe.

Beispiel 1: Bilden von vierstelligen Passwörtern mit verschiedenen Ziffern; erste Ziffer .

Beispiel 2: Kontrolle von 2 Glühbirnen aus einer 12er-Schachtel Glühbirnen.

## Anordnungen

### Permutation (Vertauschung) – Anordnungsproblem ohne Wiederholungen

Auf wie viele Arten kann man verschiedene Elemente anordnen?

Gesucht ist die Anzahl aller Anordnungen 🡪 Produktregel

|  |
| --- |
| verschiedene Elemente kann man auf Arten anordnen. |

### Permutation (Vertauschung) – Anordnungsproblem mit Wiederholungen

Beispiel: Elemente

### Anwendung bei Binomial-Koeffizienten

**Koeffizienten direkt berechnen**

Diese Produkte kommen so häufig vor, wie man die Buchstabenfolgen bestehend aus und anordnen kann.

Beispiel:

Für das Produkt erhält man:

35 ist der Binomialkoeffizient bei .

|  |
| --- |
| **Definition der Binomialkoeffizienten:**  „n tief k“: |

|  |
| --- |
| **Binomische Formel**: |

Beispiel: wird ausmultipliziert. Welcher Faktor steht bei ?

Zeilensumme im Pascal’schen Dreieck:

Merke:

Allgemein:

## Auswahlen

Aus einer Stichprobe will man einen Eindruck über die Gesamtheit gewinnen.

|  |
| --- |
| Ziehen von Eementen aus verschiedenen Elementen ohne Berücksichtigung der Reihenfolge:  verschiedene Auswahlen |

## Übersicht

Es sind vier Ziehungsarten möglich.

|  |  |  |  |
| --- | --- | --- | --- |
| **Ohne zurücklegen:** | | **Mit zurücklegen: beliebig** | |
| Variation -ter Ordnung **ohne** Wiederholung | Kombination -ter Ordnung **ohne** Wiederholung | Variation -ter Ordnung **mit** Wiederholung | Kombination -ter Ordnung **mit** Wiederholung |
| Reihenfolge wird berücksichtigt | Reihenfolge spielt keine Rolle | Reihenfolge wird berücksichtigt | Reihenfolge spielt keine Rolle |
| mit | mit | möglich | möglich |

Beispiel 1: Anzahl Siegerlisten der ersten 3 Plätze bei einem Pferderennen mit 7 Pferden.

Beispiel 2: Anzahl achtstellige Zahlen bestehend aus lauter ungeraden Ziffern (1,3,5,7,9).

# Wahrscheinlichkeiten

## Klassische Definition nach Laplace

Wurf eines Würfels. Jede Zahl hat dieselbe Wahrscheinlichkeit .

|  |
| --- |
| Kurz:  Es gilt immer: |

Beispiel 1: Münze sechsmal werfen. OOOOOO

1. Der dritte Wurf zeigt Zahl  
   ,
2. Genau dreimal erscheint Zahl  
    „3 aus 6 Plätzen“
3. Mindestens viermal erscheint Zahl  
   4x Zahl + 5x Zahl + 6x Zahl

„Entweder, oder“

Beispiel 2: Wurfbilder beim Werfen von zwei Würfeln.

1. Die Augensumme ist 9  
   , Diagonale: (6|3), (5|4), (4|5), (3|6)
2. Mindestens eine der Augenzahlen ist eine Primzahl (2,3,5)  
   Gegenteil von mindestens EINE ist KEINE: Gegenereignis

Gegenwahrscheinlichkeit

1. Die Differenz der Augenzahlen ist höchstens 2  
   Diagonalen

Beispiel 3: Schachtel mit Glühbirnen. 4 defekt, 8 intakt, Stichprobe , Ziehen ohne zurücklegen

1. Die Glühbirnen in der Stichprobe sind alle intakt.  
   ,
2. Höchstens zwei defekte Glühbirnen in der Stichprobe
3. Mindestens ein defekt in der Stichprobe

Gegenwahrscheinlichkeit

Beispiel Zahlenlotto: Jede Zahl hat die gleiche Chance gezogen zu werden. „Ziehe 6 aus 45“.

Anzahl mögliche Tipps:

Wahrscheinlichkeit für 0 Richtige:

Allgemeine Formel zur Berechnung von k Richtigen im Lotto:

Ein Ereignis ist eine Ergebnis-Konstellation. Ereignis

Probabilität des Ereignis . Die Menge der Ereignisse heisst **Ereignisraum** oder **Wahrscheinlichkeitsraum** .

Das sichere Ereignis

Das unmögliche Ereignis

## Axiome der Wahrscheinlichkeit

|  |
| --- |
| Jedem Ereignis wird eine Zahl zugeordnet mit den Eigenschaften:  A1:  A2  A3 falls , dann  (gemäss A. Kolmogoroff)  Wahrscheinlichkeitsfunktion |

Diese drei Axiome genügen, wenn der Ereignisraum endlich ist. Für den Falls, dass es unendlich viele Teilmengen gibt, ist noch ein weiteres Axiom notwendig (🡪 sogenannte Ereignisalgebra).

**Zu A3: Prinzip von Inklusion und Exklusion**

Additionssatz:

**Folgerungen aus den Axiomen:**

1. und

ist die **Gegenwahrscheinlichkeit** von .

Speziell :

2. ist echte Teilmenge von

3. **Additionssatz**

### Aussagenlogik

|  |  |  |
| --- | --- | --- |
| **Distributivität:** | **Regeln von de Morgan:** |  |

**Beispiel Kontraposition**

4. Die Wahrscheinlichkeitsdefinition nach Laplace ergibt sich ebenfalls aus den Axiomen A1,A2,A3.

Die Grundereignisse werden als gleichwahrscheinlich vorausgesetzt.

Wenn , so hat jedes Grundereignis die Wahrscheinlichkeit .

Mit ergibt sich

## Mehrstufige Versuche

Beispiel: In einer Urne liegen 5 weisse und 3 schwarze Kugeln.

Man zieht zwei Kugeln nacheinander ohne zurücklegen. Alle Möglichkeiten:

weiss,weiss:

weiss,schwarz:

schwarz,weiss:

schwarz,schwarz:

### Ereignisbaum

|  |  |
| --- | --- |
|  | Mehrstufige Versuche kann mal durch einen Ereignisbaum darstellen.  Summer der Äste = 1  **Pfadregeln:**   * entlang eines Pfades werden die Stufenwahrscheinlichkeiten multipliziert. * Die Pfade sind disjunkte Ereignisse. Ihre Wahrscheinlichkeiten werden addiert. |

Dieselbe Rechnung diesmal aber mit dem Gegenereignis:

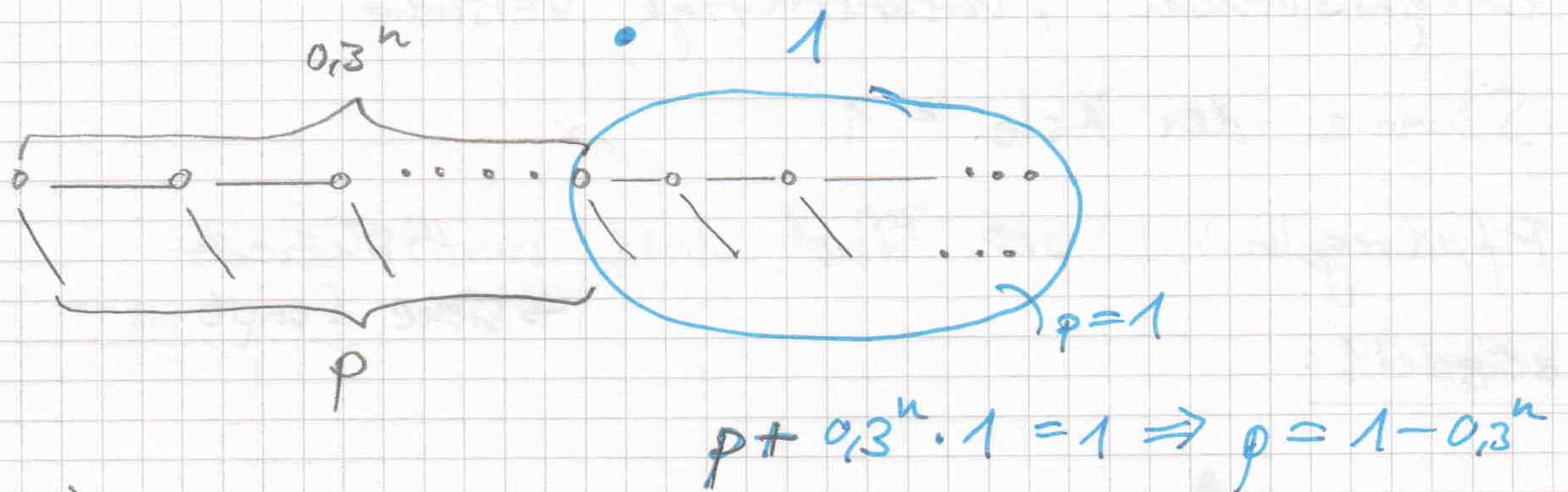
Aufgabe 1:

|  |  |
| --- | --- |
|  | Günstige Pfade: |

Aufgabe 2:

|  |  |
| --- | --- |
|  |  |
|  | Geometrische Reihe: |

Direkt berechnet:



### Anwendung 1: Zuverlässigkeit von (komplexen) Systemen

**Ausfallswahrscheinlichkeit** der -ten Komponente

**Zuverlässigkeit** der -ten Komponente

Ausfallswahrscheinlichkeit des Systems

Zuverlässigkeit des Systems

Man stelle sich einen Signalfluss von A nach B vor. Die Komponenten fallen aus (stören den Signalfluss) oder sind intakt. Einfache Systeme bestehen aus zwei Komponenten und diese kann man grundsätzlich auf 2 Arten zusammenschalten:

|  |  |
| --- | --- |
| Serie-Schaltung | Parallel-Schaltung |

|  |  |
| --- | --- |
|  | Teilsystem :  Teilsystem :  Teilsystem : |

Beispiel: Ein 4-motoriges Flugzeug kann sich in der Luft halten, wenn

1. Auf jeder Seite mindestens ein Motor intakt ist (Serie-Schaltung)
2. Entweder beide inneren oder beide äusseren Motoren intakt sind (Parallel-Schaltung)

|  |  |
| --- | --- |
|  |  |

### Anwendung 2: Kryptographische Hashfunktion, Kollisionswahrscheinlichkeit

Man hat Hashwerte und Versuche. Mit welcher Wahrscheinlichkeit stimmen bei mindestens zwei Versuchen die Hashwerte überein (sogenannte Kollision)? Oder Wie viele Versuche braucht es, damit die Kollisionswahrscheinlichkeit 0.5 ist?

Anschauliche Demonstration mit dem Geburtstagsproblem: In einer Gruppe von Personen wettet jemand, dass mindestens 2 Personen den gleichen Geburtstag haben. Wie gross muss n sein, damit die Gewinnchance ist?

Wahrscheinlichkeit für eine Kollision sei .

Tage

Kollision

:

Lösung:

Allgemein:

Für kleine :

nach auflösen

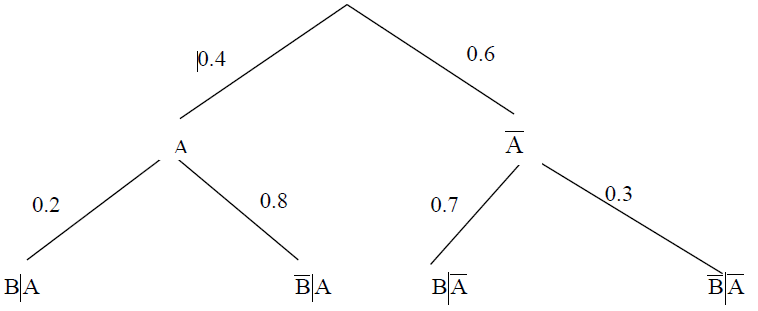
Aufgabe : Kollision mit einem bestimmten Hashwert

## Bedingte Wahrscheinlichkeit

Verhältnis von Wahrscheinlichkeiten. Die Bedingung steht im Nenner.

: Wahrscheinlichkeit für unter der Bedingung .

Siehe **Ereignisbaum** 🡪 Multiplikation entlang eines Pfades.



|  |
| --- |
| **Totale Wahrscheinlichkeit:** |

Beispiel:

|  |
| --- |
| **Satz von Bayes:** |

# Verteilungen

## Zufallsvariable

In statistischen Fragen werden die Ereignisse als Zahlvariable defineirt. Ordnet man jedem Ereignis einen Zahlenwert zu, so nimmt die Variable diese Werte zufällig an. Z.B. Augensumme beim Werfen von zwei Würfeln. ist eine **Zufallsvariable** und die Wahrscheinlichkeiten sind als auszudrücken.

**Wahrscheinlichkeitsverteilung** von :

Beispiel 1:

Man wirft eine Münze (Kopf, Zahl) drei Mal. sei die Anzahl ‚Kopf‘. nimmt die Werte oder an. Zugehörige Wahrscheinlichkeiten:

z.B. :

Wahrscheinlichkeiten werden als Tabelle oder graphische als Stabdiagramm dargestellt.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  |  | | 0 | 0.125 | | 1 | 0.375 | | 2 | 0.375 | | 3 | 0.125 | |  |

### Diskrete und stetige Zufallsvariablen

|  |  |
| --- | --- |
| **Diskret** | **Stetig** |
| * Endlich viele Werte * Stabdiagramm * Stabhöhe = Wahrscheinlichkeit * Summe der Wahrscheinlichkeiten = 1 * Zufallsvariable als Zählvariable  zählt z.B. die Anzahl Erfolge bei einem Zufallsexperiment | * Alle Werte eines Intervalls aus * Eine stetige Zufallsvariable misst eine Grösse 🡪 Messdaten   **Wahrscheinlichkeits-Dichtefunktion**    definiert für Die gesamte Fläche unter der Dichtekurve ist |

## Verteilungsfunktion

Grundlegend für jede Zufallsvariable (diskret oder stetig). Berechnete **kumulierte** Wahrscheinlichkeiten .

**Die Verteilungsfunktion ist der grundlegende Begriff zur Beschreibung der Wahrscheinlichkeits-verteilung.**

|  |
| --- |
| (kumulierte) **Verteilungsfunktion** der Zufallsvariablen :    ist für jedes reelle definiert. |

Für diskrete Zufallsvariable gilt:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | | 0 | 0.125 | 0.125 | | 1 | 0.375 | 0.5 | | 2 | 0.375 | 0.875 | | 3 | 0.125 | 1 | | .000  .000 |  |

ist eine monoton wachsende Funktion. Die Wahrscheinlichkeiten nehmen sprunghaft zu.

## Hypergeometrische Verteilung

## Binomialverteilung

Beschreibt die Wiederholung eines elementaren Versuchs. Anzahl Schritte vorgegeben. Man führt den Versuch mal durch. Es kann nur zwei Ergebnisse haben: Erfolg oder Misserfolg.

|  |  |  |
| --- | --- | --- |
|  | Indikatorvariable | P |
| Erfolg |  |  |
| Misserfolg |  |  |

: Erfolg

Beispiel: ,

Die Sequenz hat die Wahrscheinlichkeit .

Es gibt günstige Sequenzen. Alle haben diese Wahrscheinlichkeit.

Somit ist

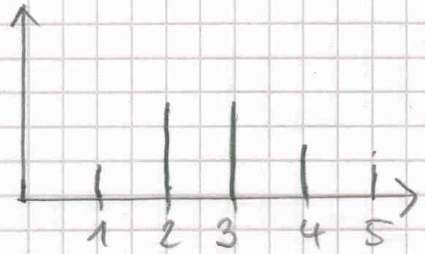
Allgemein: Führt man einen Versuch mit der Erfolgswahrscheinlichkeit mal durch, so ist die Wahrscheinlichkeit für Erfolge.

|  |
| --- |
| Die Zufallsvariable ist **binomial** verteilt: |

|  |  |
| --- | --- |
| MATLAB Beispiel:  x = 0:20;  y = binopdf(x,20,0.3);  bar(x,y);  Die Wahrscheinlichkeit ist am grössten (Modus) bei .  D.h. die Wahrscheinlichkeiten sind gross in der Nähe von . |  |

Beispiel 8:

5% der Maroni eines Händlers sind schlecht. Jemand kauft 50 Maroni. ist die Anzahl schlechte Maroni. Modell: grösste Wahrscheinlichkeit



1. Wahrscheinlichkeit, dass höchstens 4 schlecht sind:
2. Wahrscheinlichkeit, dass mindestens eine schlecht ist:
3. Wie viele Maroni müsste man kaufen, damit mit Wahrscheinlichkeit 0.99 mindestens eine schlechte dabei ist?  
    Auflösen nach mit Solver 🡪

## Geometrische Verteilung

Warten (Versuche) bis der Erfolg eintritt. ist die Anzahl Misserfolge (Warteschritt), bis der erste Erfolg eintritt. Misserfolgswahrscheinlichkeit.

Beispiel: Sequenz:

geometrische Reihe

|  |
| --- |
| Die Zufallsvariable ist **geometrisch** verteilt: |

Die Wahrscheinlichkeiten entsprechen der Exponentialfunktion.

Beispiel 9:

Man setzt beim Roulette auf Rot. ist die Anzahl Misserfolge bis Rot erscheint. Bestimme die Verteilung von . .

1. Mit welcher Wahrscheinlichkeit erscheint genau beim 3. Versuch zum ersten mal ROT?
2. Mit welcher Wahrscheinlichkeit erscheint spätestens beim 4. Versuch zum ersten mal ROT?
3. Mit welcher Wahrscheinlichkeit erscheint frühestens beim 5. Versuch zum ersten mal ROT?

## Vergleich der Binomialverteilung mit der Hypergeometrischen

Da die binomialen Wahrscheinlichkeiten einfacher zu berechnen sind, kann man die hypergeometrische Verteilung approximativ durch die binomiale Verteilung berechnen.

Beispiel 10:

(Binomialverteilung)

(Hypergeometrische Verteilung)

# Poisson-Verteilung

Bekannt ist der Mittelwert der Verteilung. Gleiche mittlere Anzahl Treffer .

Innerhalb von Intervallen der Länge hat es im Mittel Treffer.

D.h. die Anzahl Treffer in ist binomial verteilt:

Für Treffer in gilt die Wahrscheinlichkeit

Der Erwartungswert für jedes ist demnach .

## Grenzmodell

Demonstration mit .

Die Wahrscheinlichkeiten stabilisieren sich für .

Die Varianz hat den Grenzwert strebt gegen

Poisson-Verteilung mit dem Paramter :

Für grosse und kleine und wenn nicht zu gross ist, ist die Poisson-Verteilung eine gute Näherung für die Binomialverteilung mit .

Aus kleinen Trefferwahrscheinlichkeiten ergibt sich die Interpretation, dass die Poisson-Verteilung seltene Ereignisse gut beschreibt. (Waldbrände, Schadenfälle).

Beispiele :

, Erwartungswert ist . Ist die Summe = 1 ?

Reihe für

Beispiel 1 :



Beispiel 2 :



# Schätzen

Anwendung des zentralen Grenzwertsatzes

Geschätzte Grössen symbolisiert man mit ^.

## Punktschätzung

Intuitiv wird der Durchschnitt (= Schätzfunktion) berechnet.

|  |
| --- |
| Parameter: Mittelwert und Anteil  ist der relative Anteil der Anzahl Erfolge. |

Beispiel:

Stichprobe von 120 Marroni (Gewichte in Gramm)

* Mittleres Gewicht aller Marroni des Händlers geschätzt:
* Anteil wurmstichiger Marroni aller Marroni des Händlers

## Intervallschätzung

Eine Angabe für die **Genauigkeit der Punktschätzung**.

Man verwendet den zentralen Grenzwertsatz, für grosse Stichproben gilt:

**Stichprobenvarianz**

**Standardfehler** der Stichprobe

## Vertrauensintervall

Die Zufallsvariable liegt mit Wahrscheinlichkeit zwischen den Quantilen der Verteilung und

|  |  |
| --- | --- |
|  | Durch Umformung erhält man: |

Da eine Zufallsvariable ist, so ist auch das Intervall variabel.

Es überdeckt mit Wahrscheinlichkeit den festen aber unbekannten Wert .

|  |
| --- |
| **Vertrauensintervall** für  mit |