STATISTIK II

WAHRSCHEINLICHKEIT

von

Alex Keel

15. Auflage

2004

Verlag Wilhelm Surbir Wittenbach/St. Gallen

Vorlesung an der Universität St. Gallen - Hochschule für Wirtschafts-, Rechts- und Sozialwissenschaften (HSG)

Prof. Dr. Alex Keel, Bodanstrasse 6, CH-9000 St. Gallen,
Tel. ++41 / 71 / 224 24 31, Fax. ++41 / 71 / 224 28 94

Verlag Wilhelm Surbir, Betten 10, CH-9303 Wittenbach,
Tel. und Fax. ++41 / 71 / 298 36 16

Vorwort

Mit dem Begleittext Statistik II *Wahrscheinlichkeit* ist die Absicht einer vorlesungsbegleitenden Einführung in die Wahrscheinlichkeitsrechnung verbunden. Der Text wendet sich an Studierende der Wirtschaftswissenschaften in den ersten Semestern. Die mathematisch-formalen Voraussetzungen wurden entsprechend tief angesetzt. Ebenso wurde aus didaktischen Überlegungen anstelle mathematischer Korrektheit das Schwergewicht auf heuristisch motivierte Intuition gelegt. Bei der Stoffauswahl wurde darauf geachtet, dass der Leser bereits im Rahmen dieser Einführung mit möglichen Anwendungen vertraut wird. Naturgemäss konzentrieren sich die praktischen Applikationen auf Bereiche aus dem Operations Research, der Ökonometrie, der Finance sowie den empirisch ausgelegten Wirtschaftswissenschaften.

Der aktuelle Wissensstand kann anhand eines Übungsskriptes jederzeit festgestellt werden.

Gerne benutze ich die Gelegenheit, meinen Mitarbeiterinnen und Mitarbeitern zu danken. Der vorliegende Band ist in kooperativer Zusammenarbeit entstanden. Die Herren Stefan Ott, Dipl.-Math.oec., Anton Vucurovic, lic.oec. und Reto Leibundgut, lic.oec. standen mir als aktive Diskussionspartner stets zur Verfügung. Ebenso danke ich ihnen für die Mithilfe bei der Gestaltung und Durchsicht des Manuskriptes.

Frau M.-C. Baumann hat den Text in gewohnter Souveränität zu Papier gebracht. Ihre Nachsicht und Geduld gegenüber immer neuen Wünschen verdienen besondere Erwähnung. Herzlichen Dank.

I N H A L T S V E R Z E I C H N I S

1. EINLEITUNG 1

2. ZUFALLSEXPERIMENTE, ZUFALLSEREIGNISSE 3
   2.1 Zufallsexperiment 3
   2.2 Ereignisraum 3
   2.3 Ereignisse und Operationen mit Ereignissen 5

3. WAHRSCHEINLICHKEITSMODELLE 10
   3.1 Die klassische Wahrscheinlichkeit 10
   3.2 Häufigkeitswahrscheinlichkeit 13

4. WAHRSCHEINLICHKEITSFUNKTION 15
   4.1 Axiome der Wahrscheinlichkeitsrechnung 15
   4.2 Eigenschaften der Wahrscheinlichkeitsfunktion P 17
      4.2.1 Elementare Sätze 17
      4.2.2 Bedingte Wahrscheinlichkeit 21
      4.2.3 Multiplikationssatz 24
      4.2.4 Satz von der totalen Wahrscheinlichkeit 25
      4.2.5 Bayes'sche Formel 27
      4.2.6 Unabhängige Ereignisse 30

5. (EINDIMENSIONALE) ZUFALLSVARIABLE 37
   5.1 Diskrete Zufallsvariable 37
      5.1.1 Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen 39
      5.1.2 Verteilungsfunktion einer diskreten Zufallsvariablen 41
   5.2 Stetige Zufallsvariable 44
   5.3 Erwartungswert und Varianz von Zufallsvariablen 50
      5.3.1 Mittelwert einer Zufallsvariablen 50
      5.3.2 Varianz einer Zufallsvariablen 51
   5.4 Funktionen von Zufallsvariablen 55
   5.5 Mathematische Erwartung - Erwartungswert von Funktionen zufälliger Variablen 61
      5.5.1 Eigenschaften mathematischer Erwartungen 64
      5.5.2 Lineare Transformation, standardisierte Zufallsvariablen 66
      5.5.3 Ungleichung von Tschebyscheff 67
      5.5.4 Jensen'sche Ungleichung 69
      5.5.5 Momente von Zufallsvariablen 71

5.5.6 Die momenterzeugende Funktion 73
5.5.7 Approximation von Momenten 77
5.5.8 Median und Quantile von Zufallsvariablen 78

6. SPEZIELLE DISKRETE VERTEILUNGEN 81
6.1 Diskrete Gleichverteilung 81
6.2 Verteilungen im Zusammenhang mit Bernoulliexperimenten 82
6.2.1 Bernoulliexperiment 82
6.2.2 Bernoulliverteilung 83
6.2.3 Binomialverteilung 84
6.2.4 Die Verteilung von Anteilen 92
6.2.5 Hypergeometrische Verteilung 94
6.2.6 Poissonverteilung 99
6.2.7 Approximation der Binomialverteilung für kleine Erfolgswahrscheinlichkeiten 101
6.2.8 Der Poisson-Prozess 104
6.2.9 Geometrische Verteilung 108
6.2.10 Negative Binomialverteilung 111
6.2.11 Zusammenhang zwischen der Binomialverteilung und der negativen Binomialverteilung 113

7. SPEZIELLE STETIGE VERTEILUNGEN 115
7.1 Stetige Gleichverteilung 115
7.2 Exponentialverteilung 117
7.3 Normalverteilung 120
7.4 Die Lognormalverteilung 129
7.5 Pareto-Verteilung 134
7.6 Gammafunktion 136
7.7 Betafunktion 137
7.8 Die Chi-quadratverteilung 138
7.9 Die Gammaverteilung 142
7.10 Die t-Verteilung (Student-Verteilung) 144
7.11 Die F-Verteilung 149
7.12 Betaverteilung 151

8. APPROXIMATION DISKRETER VERTEILUNGEN DURCH STETIGE VERTEILUNGEN 155
8.1 Kontinuitätskorrektur 155

8.2 Approximation der Binomialverteilung durch die Normalverteilung . . . 156
8.3 Approximation der hypergeometrischen Verteilung durch die Normalverteilung . . . 163
8.4 Approximation der Poissonverteilung durch die Normalverteilung . . . 165
8.5 Schematische Zusammenfassung der Approximationsmöglichkeiten . . . 168

9. MEHRDIMENSIONALE ZUFALLSVARIABLEN . . . 169
9.1 Diskrete 2-dimensionale Zufallsvariablen . . . 169
9.1.1 Randverteilungen . . . 172
9.1.2 Bedingte Verteilungen . . . 173
9.1.3 Unabhängige Zufallsvariablen . . . 174
9.2 Stetige 2-dimensionale Zufallsvariablen . . . 176
9.2.1 Randverteilungen . . . 179
9.2.2 Bedingte Verteilungen . . . 180
9.2.3 Unabhängige Zufallsvariablen . . . 181

10. FUNKTIONEN VON ZUFALLSVARIABLEN . . . 183
10.1 Gemeinsame Verteilung . . . 183
10.2 Erwartungswert einer Funktion von Zufallsvariablen . . . 187
10.2.1 Bedingte Erwartung einer zweidimensionalen Zufallsvariablen . . . 188
10.3 Summe von Zufallsvariablen . . . 190
10.3.1 Mittelwert und Varianz einer Summe von Zufallsvariablen . 190
10.3.2 Kovarianz und Korrelationskoeffizient . . . 192
10.3.3 Gesetz der grossen Zahlen . . . 197
10.3.4 Momenterzeugende Funktion einer Summe von unabhängigen Zufallsvariablen . . . 198
10.3.5 Summe unabhängiger Binomialvariablen . . . 200
10.3.6 Summe unabhängiger Poissonvariablen . . . 201
10.3.7 Summe normalverteilter Zufallsvariablen . . . 201
10.3.8 Summe unabhängiger Chi-quadratvariablen . . . 204
10.4 Produkt von Zufallsvariablen . . . 204
10.4.1 Erwartungswert eines Produktes von zwei Zufallsvariablen . 204
10.4.2 Produkt lognormalverteilter Zufallsvariablen . . . 205
10.5 Der zentrale Grenzwertsatz . . . 207
10.6 Spezielle mehrdimensionale Verteilungen . . . 212
10.6.1 Multinomialverteilung . . . 212

10.6.2 Die zweidimensionale Normalverteilung 216

LITERATURVERZEICHNIS 224

INDEX 225

2. TEIL - WAHRSCHEINLICHKEITSRECHNUNG

1. EINLEITUNG

Die Wahrscheinlichkeitsrechnung gehört zu den fundamentalen Bausteinen der gesamten statistischen Methodenlehre. Wie viele mathematische Disziplinen hat auch die Wahrscheinlichkeitsrechnung ihren Ursprung in rein praktischen Fragestellungen, speziell bei den Glücksspielen.

Wir haben es dem passionierten Spieler Chevalier de Méré zu verdanken, dass er sich bei der richtigen Person, nämlich bei Blaise Pascal, über das Auseinanderklaffen "mathematischer Erkenntnisse" mit den Erfahrungen des praktischen Lebens beklagt hat.

De Méré wusste aufgrund langer Erfahrung, dass es bei viermaligem Würfeln für die Bank günstig ist, auf das Ereignis "mindestens eine 6" zu wetten. De Méré, als Leiter einer Spielbank, hat sich eine Variation dieses Spiels mit einem Doppelwürfel ausgedacht. Ein Spieler gewinnt, falls er in x Würfen keine "Doppelsechs" erreicht. Der Arzt Geronimo Cardano versuchte, das Problem zu lösen, indem er von der Wahrscheinlichkeit von 1/36 für den Wurf einer Doppelsechs ausging, sodass man nach 18 Würfen mit Wahrscheinlichkeit 1/2 erwarten könne, dass eine Doppelsechs auftrete. Somit sei es vorteilhaft, auf das Ereignis „mindestens eine Doppelsechs in 19 Würfen" zu wetten.

Als erfahrener Spieler wusste de Méré, dass der Wert $x = 19$ offensichtlich zu klein ist. Er ging im Gegensatz dazu vom Falle des (bekanntermassen vorteilhaften) Spiels mit *einem* Würfel aus und bildete das Verhältnis aus der Anzahl Würfe (4) sowie der möglichen Ausgänge (6), also 2/3. In der Annahme, dass diese Proportionalität die Chancen des Spiels bestimme, legte er im Falle des Doppelwürfels (36 Ausgänge) den Wert x fest.

Es stellte sich jedoch heraus, dass die Bank beim Spiel mit dem Doppelwürfel langfristig nicht auf ihre Rechnung kam. In dieser Verzweiflung wandte sich de Méré an Pascal, der das Problem auch löste und in einem Brief an Fermat über de Méré meinte: "Il est très bon esprit, mais il n'est pas géomètre."

Glücksspiele wie Würfeln, Roulette, Ziehen eines Loses oder das Lottospiel sind dadurch gekennzeichnet, dass ihr Ausgang ungewiss ist. Bekannt ist in der Regel der Realisationsbereich der Ausgänge, unbekannt ist jedoch, welches Resultat im Einzelfall eintritt. Man sagt, dass der Zufall über den konkreten Ausgang entscheidet. An dieser Stelle setzt die Wahrscheinlichkeitstheorie an. Sie lehrt, wie der intuitiv zwar fühlbare Begriff des Zufalls mathematisiert werden kann. Mit Hilfe der Wahrscheinlichkeitstheorie kann die Unsicherheit nicht beseitigt, wohl aber formalisiert werden.

Obschon sich der Zufall am augenfälligsten bei Glücksspielen manifestiert, ist damit nur ein marginaler Bereich seiner Einflusssphäre angeschnitten. Die Realität ist in vielfältigster Weise dadurch gekennzeichnet, dass sich ihre künftige Entwicklung nicht voraussagen lässt. Diese Tatsache wird selbst durch die stets fortschreitende wissenschaftliche Entwicklung in keiner Weise entschärft. Die Wissenschaft vermag zwar das Niveau des Vorhersehbaren zu verändern, das Phänomen selber jedoch nicht zu beseitigen. Im Gegenteil lebt die Forschung und Entwicklung geradezu vom Zufall und zwar sowohl in der Makro- wie auch in der Mikrosphäre. Die Evolution, die sich permanent in der Pflanzen- und Tierwelt vollzieht, mag als weiteres Beispiel für die positive Beeinflussung durch den Zufall dienen. Schliesslich liefert auch die Ökonomie eindrückliche Beispiele, bei denen der Zufall die Szenerie beherrscht. Wie könnten beispielsweise Finanzmärkte auf Dauer existieren, wenn heute bekannt wäre, was morgen passiert. Es fällt schwer, sich eine Welt vorzustellen, welche ohne den Zufall als realen Faktor existieren könnte. In einer solchen Welt wären sämtliche Vorgänge vorherbestimmt, zumal der aktuelle Stand eindeutig aus der Vergangenheit folgen müsste und somit auch kausal auf die Zukunft einwirken müsste.

Diese einfachen Überlegungen zeigen, dass sich eine Auseinandersetzung mit dem Zufall lohnt. Die Wahrscheinlichkeitstheorie bietet dazu ein effizientes Hilfsmittel. Sie vermag zumindest jene Gesetzmässigkeiten zu formulieren, welche letztlich als Folge einer Menge von Zufälligkeiten entstanden sind. Wird beispielsweise eine symmetrische Münze sehr oft geworfen, so ist zwar der Ausgang eines Einzelwurfs ungewiss, hingegen zeigt sich eine auffallende Übereinstimmung zwischen der Anzahl der beiden möglichen Ausgänge “Kopf” und “Zahl”.

Solche langfristigen Regelmässigkeiten findet man aber nicht nur bei Glücksspielen. Vor der Befruchtung lässt sich das Geschlecht der Nachkommenschaft nicht voraussagen, hingegen ist bekannt, dass unter sehr vielen Geburten etwa die Hälfte weiblichen Geschlechts ist. Eine Versicherungsgesellschaft weiss nicht, ob ein bestimmter Kunde im Verlaufe des nächsten Jahres stirbt; sie besitzt hingegen Anhaltspunkte über die Gesamtzahl der Todesfälle innerhalb einzelner Altersstufen. Bei einem Produktionsprogramm ist unbekannt, ob das nächste produzierte Stück defekt sein wird; man weiss hingegen, dass der Ausschussanteil etwa 5% beträgt.

All diese Beispiele sind durch Ungewissheit im konkreten Einzelfall und durch eine offensichtliche Gesetzmässigkeit in langen Versuchsfolgen gekennzeichnet. Solche Experimente werden als *Zufallsexperimente* bezeichnet. Unsere Aufgabe besteht darin, mathematische Modelle zur Beschreibung derartiger Zufallsexperimente zu finden.

2. ZUFALLSEXPERIMENTE, ZUFALLSEREIGNISSE

2.1 Zufallsexperiment

Ein Experiment ε heisst zufällig oder stochastisch, wenn die Gesamtheit der möglichen Ausgänge von ε bekannt ist, aber ein konkretes Ereignis nicht vorausgesagt werden kann. Kann das Experiment unter identischen Bedingungen wiederholt werden, so sind bei jedem Ausgang prinzipiell alle möglichen Ausgänge realisierbar.

<u>Beispiel</u>

- Augenzahl beim Würfeln.
- Eine Münze wird zweimal geworfen.
- Aus einer grossen Serie wird eine Lampe zufällig ausgewählt und die Lebensdauer bestimmt.
- 1000 zufällig ausgewählte Versicherte werden ein Jahr lang beobachtet und die Anzahl Todesfälle festgestellt.
- Der Kurs eines Wertpapiers wird an einem Handelstag beobachtet und festgestellt, wie oft der Kurs eine Zu- resp. Abnahme erfahren hat.

2.2 Ereignisraum

Def. 2.1

Ein Zufallsexperiment ε besitze die möglichen Ausgänge $\omega_1, \omega_2, \ldots, \omega_n$. Der Ereignisraum Ω ist dann die *Menge* aller möglichen Ausgänge

$$\begin{aligned} &\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\} \\ &\omega_i,\ i = 1, \ldots, n : \textit{Elementarereignis} \end{aligned} \tag{2.1}$$

<u>Beispiele</u>

1.) ε : 2 Münzen werden geworfen

$\Omega_1 = \{(K,K), (K,Z), (Z,K), (Z,Z)\}$
falls die Ausgänge „Kopf“ (K) bzw. „Zahl“ (Z) auf beiden Münzen interessieren

$\Omega_2 = \{(2,0), (1,1), (0,2)\}$
falls die Anzahl K und die Anzahl Z interessieren

$\Omega_3 = \{(g), (u)\}$
falls interessiert, ob die Bilder gleich (g) oder verschieden (u) sind.

Ein Zufallsexperiment kann durch mehrere Ereignisräume beschrieben werden. Zweckmässigerweise wählt man aber eine Form, die möglichst viele Einzelheiten enthält, sie soll erschöpfend und vollständig sein.

2.) Würfel, Augenzahl
$\Omega = \{1,2,3,4,5,6\}$

3.) Doppelwürfel (2 unterscheidbare Würfel)
$\Omega = \{(1,1)\ (1,2)\ \ldots\ (1,6)\ (2,1)\ (2,2),\ \ldots\ (6,6)\}$

4.) Sport-Toto mit 13 Spielen
$\Omega =$
Anzahl Elemente?

5.) Patienten werden bezüglich 3er Symptome befragt, ob sie vorhanden sind oder nicht
$\Omega = \{(111)\ (110)\ (101)\ (100)\ (011)\ (010)\ (001)\ (000)\}$

6.) Eine Münze wird solange geworfen, bis erstmals “Kopf” erscheint
$\Omega =$
Anzahl Elemente?

7.) Eine Person wird nach ihrem Geburtstag befragt
$\Omega_1 = \{1,2,\ldots,365\}$
k Personen werden nach ihrem Geburtstag befragt
$\Omega_k =$

8.) Position eines Zeigers im Einheitskreis

$$\Omega = \left\{ x \middle| 0 \leq x \leq 2\pi \right\}$$

Beachte: Ω ist nur dann ein Ereignisraum, wenn

- jedes Element von Ω einem Ausgang des Experimentes entspricht und
- jedem Ausgang des Experimentes genau ein Element von Ω entspricht.

Beispiel

Ω = {0 Sechsen, 2 Sechsen} ist kein Ereignisraum für den Doppelwürfel.
Abhängig von der Art des Ereignisses und der Festlegung des Ereignisraumes kann Ω sein:

- endlich (Beispiele 1,2,3,4,5,7)
- abzählbar unendlich (Beispiel 6)
- überabzählbar unendlich (Beispiel 8)

2.3 Ereignisse und Operationen mit Ereignissen

Der Ereignisraum enthält alle wesentlichen Informationen eines Zufallsexperimentes.

Oft interessieren aber nicht nur die Elementarausgänge ω_i eines Zufallsexperimentes sondern eine ganze Menge von Ausgängen, die jedoch hinsichtlich einer bestimmten Eigenschaft als äquivalent zu betrachten sind. Äquivalent bedeutet, dass alle derart zusammengefassten Ausgänge eine wohldefinierte charakteristische Eigenschaft erfüllen. So zusammengefasste Elementarausgänge gestatten die Beschreibung allgemeinerer *Ereignisse*; gleichzeitig bilden sie aber auch eine *Teilmenge* von Ω.

Wir betrachten als Beispiel den Doppelwürfel. Die charakterisierenden Eigenschaften $i > j$, $i + j = 7$ oder $i = 1$ führen z.B. zu folgenden Teilmengen

$$\begin{aligned} A &= \left\{ (i, j) \,\middle|\, i > j \right\} \\ B &= \left\{ (i, j) \,\middle|\, i + j = 7 \right\} \\ C &= \left\{ (i, j) \,\middle|\, i = 1 \right\} \end{aligned} \tag{2.2}$$

Das Ereignis A tritt genau dann ein, wenn der erste Würfel eine höhere Augenzahl als der zweite besitzt, wenn also der konkrete Ausgang eines Zufallsexperimentes in der Teilmenge A liegt.

Def. 2.2

ε sei ein Zufallsexperiment mit dem Ereignisraum Ω. Ein Ereignis E ist eine Teilmenge von Ω. E tritt genau dann ein, wenn der Ausgang von ε in der Teilmenge E liegt.

Ereignis und Teilmenge sind äquivalente Begriffe.

In Analogie zu den Operationen mit Mengen definieren wir Relationen und Verknüpfungen für Ereignisse.

Def. 2.3

Das zum Ereignis A gehörige *komplementäre* Ereignis A^c besteht darin, dass A nicht eintritt.

$$A^c = \left\{ \omega \,\middle|\, \omega \notin A \right\} \tag{2.3}$$

Für das oben definierte Beispiel B (Doppelwürfel) gilt

$$\begin{aligned} B &= \left\{ (i,j) \,\middle|\, i+j=7 \right\} \\ B^c &= \left\{ (i,j) \,\middle|\, i+j \neq 7 \right\} \end{aligned} \tag{2.4}$$

Def. 2.4

Das Ereignis B *umfasst* das Ereignis A, $A \subset B$, wenn bei jedem Eintritt des Ereignisses A auch das Ereignis B eintritt.

$$\omega \in A \;\Rightarrow\; \omega \in B \quad \text{für alle} \quad \omega \in A \tag{2.5}$$

<u>Beispiel</u>

B Studierende der HSG

A Studierende der Grundstufe an der HSG

Jeder Grundstufenstudent ist ebenfalls Student an der HSG. (Die Umkehrung gilt nicht.)

Def. 2.5

Die Summe $A \cup B$ zweier Ereignisse A und B entspricht jenem Ereignis, bei welchem mindestens eines der beiden Ereignisse eintritt.

$$\omega \in (A \cup B) \Leftrightarrow \omega \in A \quad \text{oder} \quad \omega \in B \tag{2.6}$$

<u>Beispiel</u>

Wurf eines regelmässigen Würfels

A Augenzahl kleiner als 4

B Augenzahl ungerade

$A \cup B = \{1,2,3,5\}$

Def. 2.6

Das Produkt $A \cap B$ zweier Ereignisse A und B ist jenes Ereignis, bei welchem A und B gleichzeitig eintreten.

$$\omega \in (A \cap B) \Leftrightarrow \omega \in A \text{ und } \omega \in B \tag{2.7}$$

<u>Beispiel</u>

A Augenzahl kleiner als 4

B Augenzahl ungerade

$A \cap B = \{1,2,3\} \cap \{1,3,5\} = \{1,3\}$

Def. 2.7

Ω heisst sicheres Ereignis

$\varnothing = \Omega^c$ heisst unmögliches Ereignis

Beispiel

$A = \{(x, y) \mid ax + by + c = 0\}$

$B = \{(x, y) \mid ax + by + d = 0\}$

$A \cap B = \varnothing$, falls $c \neq d$ (2 parallele Geraden)

Def. 2.8

Zwei Ereignisse A und B heissen *unvereinbar*, wenn sie nicht gleichzeitig eintreten können, falls also

$$A \cap B = \varnothing \tag{2.8}$$

Beispiel

$A = \{x \mid x \text{ gerade Zahl}\}$

$B = \{y \mid y \text{ ungerade Zahl}\}$

$A \cap B = \varnothing$

Def. 2.9

Eine Menge von Ereignissen $A_1, A_2, \ldots, A_n$ wird als vollständiges System (Partition) bezeichnet, falls

$$\begin{aligned} &1.)\ \ A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i = \Omega \\ &2.)\ \ A_i \cap A_j = \varnothing \quad \forall\ i \neq j \end{aligned} \tag{2.9}$$

Beispiel

1.) A und A^c bilden ein vollständiges System.

2.) Eine Münze wird solange geworfen, bis erstmals “Kopf” erscheint, höchstens jedoch 10mal. Bezeichnet

A_i: Spiel dauert genau i Würfe

so ist

$$A_1 \cup A_2 \cup \ldots \cup A_{10}$$

ein sicheres Ereignis für die Beendigung des Spiels und

$$A_i \cap A_j = \varnothing, \quad i \neq j,$$

denn das Spiel kann nicht gleichzeitig beim i-ten und j-ten Ausgang (i ℧j) aufhören.

Merke: Die Anzahl möglicher Ereignisse im Zusammenhang mit ε hängt direkt von der Konstruktion von Ω ab. Sie bilden in ihrer Gesamtheit die Potenzmenge von Ω.

3. WAHRSCHEINLICHKEITSMODELLE

3.1 Die klassische Wahrscheinlichkeit

Die klassische Definition der Wahrscheinlichkeit ist aus der engen Verknüpfung mit Glücksspielen hervorgegangen. Wir werfen eine faire (symmetrische) Münze und fragen nach der Wahrscheinlichkeit für das Ereignis “Zahl”. Die Versuchsanordnung (faire Münze) gibt keinen Anlass, einen der beiden möglichen Ausgänge zu bevorzugen. Man ordnet deshalb beiden gleichwahrscheinlichen Ausgängen dasselbe Ungewissheitsmass, nämlich 1/2 zu.

Def. 3.1

Ein Zufallsexperiment besitze n gleichwahrscheinliche Ausgänge. n_A sei die Anzahl Ausgänge, bei denen das Ereignis A eintritt. Die Wahrscheinlichkeit für das Ereignis A ist dann

$$P(A) = \frac{n_A}{n} \tag{3.1}$$

n_A bezeichnet man als die Anzahl der (für das Ereignis A) *günstigen* Fälle und n als die Anzahl der *möglichen* Fälle.

Beim Werfen eines Würfels sind 6 Ausgänge möglich. Ist der Würfel symmetrisch, so haben wir keine Veranlassung, einen Ausgang zu bevorzugen. Wir ordnen daher jedem Ausgang dieselbe Wahrscheinlichkeit 1/6 zu. Fragen wir nach dem Ereignis “A: gerade Zahl würfeln”, so sind von 6 möglichen Ausgängen 3 günstig.
Die einfache und intuitiv verständliche Form der klassischen Wahrscheinlichkeit kann

$$P\left(\text{gerade Zahl}\right) = \frac{3}{6} = \frac{1}{2} \tag{3.2}$$

nicht darüber hinwegtäuschen, dass deren Anwendung oft mit Schwierigkeiten verbunden ist. Besondere Vorsicht ist bei den Prämissen ‘unvereinbar’, ‘gleichwahrscheinlich’ und ‘zufällig’ bezüglich der Ausgänge von Zufallsexperimenten geboten.

Die der klassischen Definition der Wahrscheinlichkeit unterstellten Voraussetzungen sind bedeutend anspruchsvoller als man auf den ersten Blick glauben könnte. Insbesondere macht der Begriff der Gleichwahrscheinlichkeit über der Menge der Elementarausgänge bei der Konkretisierung grössere Schwierigkeiten. Abgesehen vom Hinweis auf eine mögliche Tautologie fehlen oft auch Anweisungen, wie im Einzelfall die stipulierte Gleichwahrscheinlichkeit überprüft werden soll. In der Regel wird sie wie bei den Glücksspielen einfach unterstellt.

Tröstlich ist jedoch, dass man sich bei irrtümlicher Annahme der Gleichwahrscheinlichkeit in kompetenter Umgebung weiss. D'Alembert (1717-1783), Mathematiker, Philosoph und Enzyklopädist argumentierte beim Werfen zweier symmetrischer Münzen mit 3 möglichen Fällen, nämlich

(1) zweimal Kopf
(2) zweimal Wappen
(3) einmal Kopf und einmal Wappen

Konsequenterweise ordnete er den 3 Ausgängen je das Mass 1/3 zu. Die Fehlüberlegung ist offensichtlich, indem D'Alembert die Reihenfolge der Würfe nicht berücksichtigt hat. Einmal Kopf und einmal Wappen kann nämlich auf zwei Arten, nämlich WK und KW entstehen, während es für zweimal Kopf und zweimal Wappen nur eine Möglichkeit gibt. Es gibt somit 4 gleichwahrscheinliche Ausgänge

KK KW WK WW

je mit der Wahrscheinlichkeit 1/4.

Ein zweites Beispiel geht auf De Méré zurück. Er beobachtete, dass beim gleichzeitigen Werfen von 3 symmetrischen Würfeln die Augensumme 11 häufiger als 12 auftrat, obschon sie seiner Ansicht nach gleichwahrscheinlich sein sollten. Er überlegte wie folgt: Es gibt 6 verschiedene Kombinationen, sowohl die Augensumme 11 als auch 12 zu realisieren nämlich
11: (146) (155) (236) (245) (335) (344)
12: (156) (246) (255) (336) (345) (444)

Tatsächlich gilt aber

$$\begin{aligned} P\left(Augensumme = 11\right) &= \frac{27}{216} \\ P\left(Augensumme = 12\right) &= \frac{25}{216} \end{aligned} \tag{3.3}$$

Begründung?

Wir betrachten noch einmal das Experiment mit einer symmetrischen Münze. Zur intuitiven Begründung, dass die Wahrscheinlichkeit für das Ereignis “K” 1/2 beträgt, betrachten wir folgende Tabelle. Sie enthält die Ergebnisse von 100 Versuchsreihen, wobei jedesmal eine Münze $n = 100$ mal geworfen wurde. Festgehalten wurde jeweils die Anzahl n_A des Ereignisses A: „Kopf“.

48	52	52	57	49	44	47	56	55	58
56	55	47	51	52	49	54	45	50	51
53	52	54	57	50	37	46	44	49	51
53	48	55	45	50	50	55	47	49	53
48	53	52	51	53	49	42	51	56	44
57	56	51	45	57	43	50	55	40	40
46	54	47	48	56	52	55	51	54	58
63	46	51	56	54	54	51	45	53	47
46	44	50	62	49	49	45	53	38	46
50	48	54	39	51	55	45	60	49	44
520	508	513	511	521	482	490	507	493	492

Man stellt fest, dass sich die relativen Häufigkeiten $\frac{n_A}{n}$ in jeder Serie von 100 Versuchen bemerkenswert wenig von $P(A) = 1/2$ unterscheiden. $P(A)$ ist offenbar ein “Ideal mass”, um das sich die relativen Häufigkeiten scharen. Addiert man alle Häufigkeiten innerhalb der einzelnen Spalten, so erhält man Versuchsserien vom Umfang $n = 1000$. Auch hier fällt eine eher noch ausgeprägtere Massierung der relativen Häufigkeiten um das Idealmass von 1/2 auf.

So ideal sich diese klassischen a priori-Wahrscheinlichkeiten darstellen, so schnell stösst man auch an die Grenzen ihrer praktischen Einsatzmöglichkeiten. Eine erste Modifikation ist für den Fall unendlicher Ereignisräume nötig. Wie gross ist z.B. die Wahrscheinlichkeit, aus der Menge der natürlichen Zahlen eine gerade Zahl zu ziehen?

Die klassische Wahrscheinlichkeit versagt auch dann, wenn die Ausgänge des Zufallsexperimentes nicht alle dieselbe Wahrscheinlichkeit besitzen. Mit welcher Wahrscheinlichkeit ist z.B. ein Element einer Produktionsserie defekt oder wie gross ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person das nächste Jahr nicht überlebt? In beiden Situationen ist die klassische Wahrscheinlichkeit nicht anwendbar, denn die möglichen Ausgänge dürfen nicht als gleichwahrscheinlich vorausgesetzt werden. Wir benötigen ein neues Konzept.

3.2 Häufigkeitswahrscheinlichkeit

Erfahrungen aus der Praxis zeigen, dass genügend lange Beobachtungsreihen von Zufallsexperimenten bemerkenswerte Gesetzmässigkeiten aufweisen. Eine spezielle Situation haben wir schon für den Fall gleichwahrscheinlicher Ausgänge beim Münzenexperiment festgestellt. Analoge Phänomene betrachtet man aber auch bei nicht gleichwahrscheinlichen Ausgängen. Betrachten wir dazu ein Zufallsexperiment, bestehend aus n unabhängigen Versuchen, die alle unter den gleichen Bedingungen wiederholt werden. Ein Produktionsprozess erfülle diese Bedingungen. Das interessierende Ereignis A heisse „defektes Element“. Die nachfolgende Tabelle enthält die Anzahl defekter Elemente in 100 Serien mit je 100 Produktionseinheiten.

11	12	11	12	8	5	8	17	13	12
10	15	13	14	15	9	15	12	8	9
10	10	12	5	11	11	12	10	9	7
14	13	10	11	4	9	14	6	11	11
16	8	12	10	9	13	10	10	16	9
13	12	11	6	11	9	10	12	5	7
11	7	9	14	14	10	10	12	6	6
13	9	11	11	14	18	10	10	6	10
10	5	4	9	13	8	7	11	9	10
10	13	8	5	7	10	3	10	6	13
118	104	101	97	106	102	99	110	89	94

Die relativen Häufigkeiten scharen sich auch hier um einen unbekannten Wert, den man analog wie im Falle von gleichwahrscheinlichen Ausgängen als Wahrscheinlichkeit für das Ereignis A bezeichnen kann.

Wahrscheinlichkeiten sind in diesem Sinne wieder Idealmasse für die relativen Häufigkeiten in langen Versuchsserien von unabhängigen Versuchen unter konstanten Experimentierbedingungen.

Addiert man die Anzahl der defekten Elemente innerhalb der Spalten, so erhält man Versuchsserien vom Umfang 1000. Wie früher stellt man auch hier wieder fest, dass sich die relativen Häufigkeiten mit wachsender Anzahl Versuche tendenziell enger um den (unbekannten) Wert der Wahrscheinlichkeit scharen. Die Wahrscheinlichkeit $P(A)$

für ein Ereignis A lässt sich deshalb durch die relative Häufigkeit $\frac{n_A}{n}$ aus einer unabhängigen Versuchsserie vom Umfang n approximieren.

4. WAHRSCHEINLICHKEITSFUNKTION

Wir haben bisher Zufallsexperimente ε mit den dazugehörigen Ereignissen *A*,*B*, ... betrachtet. Im folgenden geht es darum, den Ereignissen *A*, *B*, ... Masse *P*(*A*), *P*(*B*) derart zuzuordnen, dass aus ihnen die "Wahrscheinlichkeit" ihres Eintretens bei der Durchführung des Experimentes zum Ausdruck kommt. Dabei sollen gewisse intuitive Eigenschaften erfüllt sein. Je eher wir mit einem Ereignis *A* rechnen können, desto grösser soll das Mass *P*(*A*) sein. Ebenso soll das Mass auf ein bestimmtes Intervall normiert sein.

4.1 Axiome der Wahrscheinlichkeitsrechnung

Ω sei der Ereignisraum eines Zufallsexperimentes ε und A, B seien beliebige Ereignisse im Zusammenhang mit ε.

$$\begin{aligned}
&\Omega = \{\omega_1, \omega_2, ..., \omega_n\} \\
&\omega_i,\ i = 1, \ldots, n: Elementarereignisse \\
&A \subset \Omega \\
&B \subset \Omega
\end{aligned} \tag{4.1}$$

Eine Funktion *P*, die *jedem* Ereignis $A \subset \Omega$ eine reelle Zahl *P*(*A*) zuordnet, heisst Wahrscheinlichkeitsfunktion, falls sie folgende Axiome erfüllt:

(A1)	$P(A)$	≥ 0,	für alle $A \subset \Omega$
(A2)	$P(\Omega)$	$= 1$	(Normierung)
(A3)	$P(A \cup B)$	$= P(A) + P(B)$,	falls $A, B \subset \Omega$ und $A \cap B = \varnothing$

P ist eine sog. Mengenfunktion; der Definitionsbereich von *P* ist die Potenzmenge von Ω und der Wertebereich entspricht den reellen Zahlen im Intervall [0,1].

Die Axiome (A1), (A2) und (A3) wurden 1933 vom russischen Mathematiker Andrei Nikolajewitsch Kolmogoroff (1903-1987) formuliert. Kolmogoroff ist in diesem Sinne der Begründer der modernen axiomatischen Wahrscheinlichkeitstheorie.

Diese Axiome sind sowohl mit der klassischen Wahrscheinlichkeit als auch mit der Häufigkeitsinterpretation von Wahrscheinlichkeiten verträglich. Die Axiome 1 und 2 sind intuitiv verständlich. Axiom 3 lässt sich häufigkeitstheoretisch motivieren.

$$P(A \cup B) \approx \frac{n(A \cup B)}{n} = \frac{n_A}{n} + \frac{n_B}{n} = P(A) + P(B) \tag{4.2}$$

Beispiele

1.) ε : symmetrischer Würfel

ω_i	1	2	3	4	5	6
$P(\omega_i)$	1/6	1/6	1/6	1/6	1/6	1/6

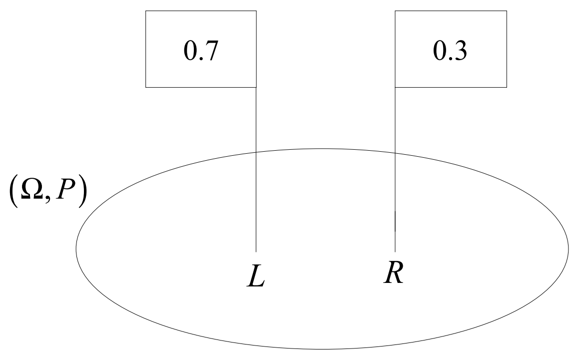
A: ungerade Zahl
B: gerade Zahl
$P(A) = 1/2$ $P(B) = 1/2$ $P(A \cup B) = P(\Omega) = 1$

2.) ε : asymmetrischer Würfel

ω_i	1	2	3	4	5	6
$P(\omega_i)$	0.2	0.1	0.1	0.2	0.3	0.1

A: ungerade Zahl
$P(A) = P(1) + P(3) + P(5) = 0.6$

3.) Eine Ratte durchläuft ein Labyrinth. Sie trifft ihre Entscheide, nach links (L) resp. nach rechts (R) zu gehen mit Wahrscheinlichkeit 0.7 resp. 0.3. Für jeden Entscheid existiert der Ereignisraum



Beachte:
Jede Abbildung der Form

$$P(L) = p, \qquad P(R) = 1 - p, \qquad 0 \le p \le 1$$

wäre eine zulässige Zuordnung von Wahrscheinlichkeiten. Abgesehen von der Restriktion $P(\Omega) = 1$ sagen die Axiome 1-3 über die zahlenmässigen Werte der Wahrscheinlichkeiten nichts aus. Diese werden im Rahmen der Wahrscheinlichkeitsrechnung als bekannt vorausgesetzt. Letztere lehrt nur, wie man mit Wahrscheinlichkeiten operiert und nicht wie man sie bestimmt. Dies ist Aufgabe der induktiven Statistik.

Der Ereignisraum Ω und die Wahrscheinlichkeitsfunktion P werden zusammen als Wahrscheinlichkeitsraum (Ω, P) bezeichnet.

4.2 Eigenschaften der Wahrscheinlichkeitsfunktion *P*

4.2.1 Elementare Sätze

Satz 4.1

Ist $\varnothing$ das unmögliche Ereignis, so ist $P(\varnothing) = 0$.

Beweis: $A = A \cup \varnothing$ für ein beliebiges Ereignis A

$A \cap \varnothing = \varnothing$

Nach (A3) $P(A) = P(A \cup \varnothing) = P(A) + P(\varnothing)$

$\Rightarrow P(\varnothing) = 0$

Satz 4.2 (Endliche Additivität)

Sind $A_1, A_2, \dots, A_n$ n paarweise unvereinbare (disjunkte) Ereignisse, so gilt

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^{n} P(A_i) \tag{4.3}$$

Satz 4.3

Bezeichnet A^c das Komplementärereignis zu A, so gilt

$$P(A^c) = 1 - P(A) \tag{4.4}$$

Beweis: (A2), (A3)

Satz 4.4
Sind A und B beliebige Ereignisse über Ω, so gilt

$$P(A) = P(A \cap B) + P(A \cap B^c) \tag{4.5}$$

Beweis:

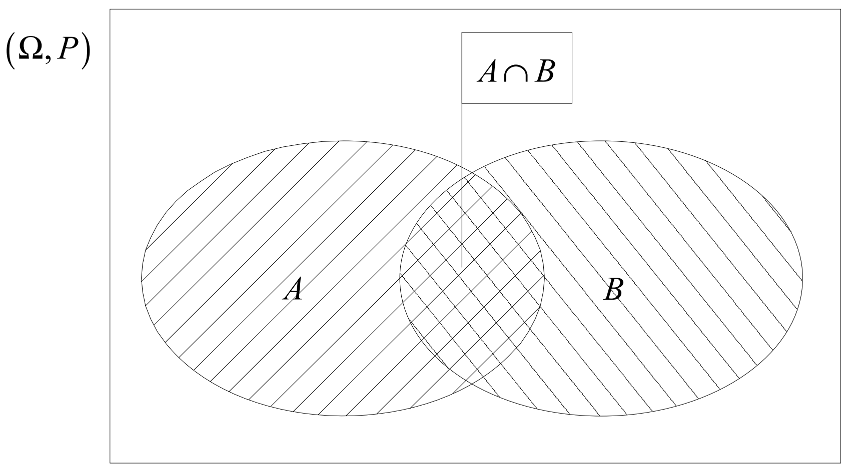
$$A = (A \cap B) \cup (A \cap B^c) \text{ und } (A \cap B) \cap (A \cap B^c) = \varnothing \tag{4.6}$$

Satz 4.5 (*Additionssatz* für beliebige Ereignisse)
Sind A und B beliebige Ereignisse im Zusammenhang mit ε, so gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{4.7}$$

Beweis:

$$\begin{aligned} &A \cup B = A \cup (A^c \cap B) \quad \text{mit } A \cap (A^c \cap B) = \varnothing \\ &P(A \cup B) = P(A) + P(A^c \cap B) = P(A) + P(B) - P(A \cap B) \end{aligned} \tag{4.8}$$



Satz 4.6 (Ein-Ausschaltformel)

$$A_1, A_2, \ldots, A_n \subset \Omega$$

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum P(A_i) - \sum_{i<j}\sum P(A_i \cap A_j) + \sum \sum_{i<j<k} \sum P(A_i \cap A_j \cap A_k) - \ldots. + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n) \tag{4.9}$$

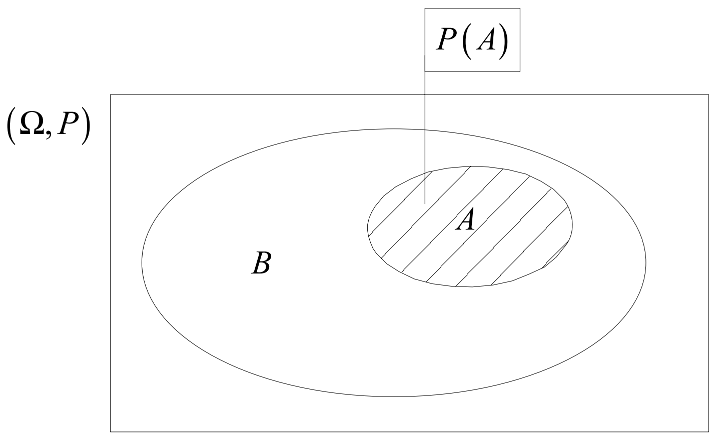
<u>Beweis:</u> vollständige Induktion. Verankerung für $n = 2$ in Satz 4.5

Satz 4.7

Seien $A, B \subset \Omega$ und $A \subset B$. Dann gilt $P(A) \leq P(B)$.

<u>Beweis:</u>

$$B = A \cup (B \cap A^c) \tag{4.10}$$



Satz 4.8 (Boole'sche Ungleichung)

$$A_1, A_2, \ldots, A_n \subset \Omega$$

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) \leq P(A_1) + P(A_2) + \ldots + P(A_n) \tag{4.11}$$

Beweis: Prinzip der vollständigen Induktion

- Verankerung:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2) \tag{4.12}$$

- Induktionsschritt

Enthält Ω (abzählbar) unendlich viele Elemente, so benötigt man noch ein zusätzliches Axiom, nämlich

$$\text{(A4)} \qquad P(A_1 \cup A_2 \cup \ldots) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \tag{4.13}$$

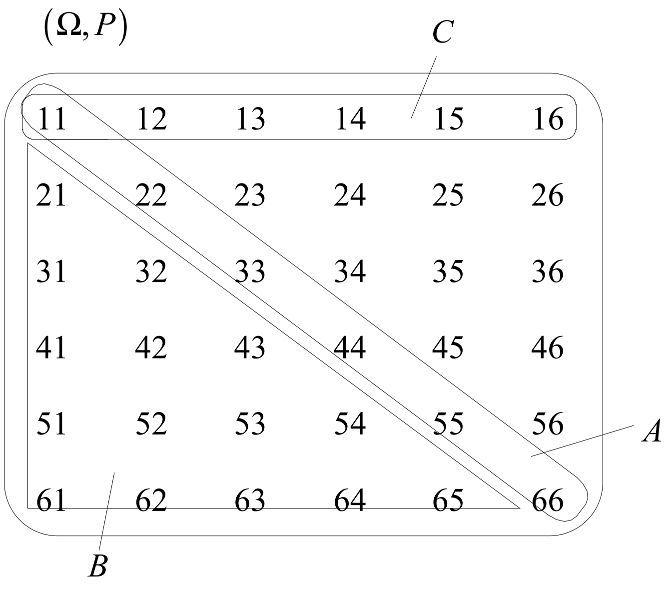
für sämtliche unendlichen Folgen paarweise disjunkter Teilmengen.

Ist Ω eine endliche Menge, so reduziert sich (A4) auf (A2). Der Satz 4.2 folgt dann automatisch. Aus Satz 4.2 kann hingegen (A4) nicht abgeleitet werden!

Ist die Menge der möglichen Ergebnisse überabzählbar unendlich, so kann nicht garantiert werden, dass das Wahrscheinlichkeitsmass P für sämtliche Teilmengen definiert werden kann. Um auch für derartige Fälle eine praktikable Lösung zu ermöglichen, konstruiert man eine reduzierte Menge von Teilmengen, bis P wieder axiomatskonform und umfassend definiert ist. Die reduzierte Familie von Teilmengen wird als σ–Algebra bezeichnet.

Beispiele:

ε : symmetrischer Doppelwürfel



$A = \{(i,j) \mid i = j\}$

$B = \{(i,j) \mid i > j\}$

$C = \{(i,j) \mid i = 1\}$

$P(A) = P(1,1) + P(2,2) + P(3,3) + P(4,4) + P(5,5) + P(6,6) = 6/36$
$P(B) = 15/36$
$P(C) = 6/36$
$P(A \cup B) = P(A) + P(B) = 21/36$
$P(A \cap B) = 0$
$P(A \cup C) = 11/36$
$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= 6/36 + 15/36 + 6/36 - 0 - 1/36 - 0 + 0 \\ &= 26/36 \end{aligned}$$

4.2.2 Bedingte Wahrscheinlichkeit

Bei der Analyse von Beobachtungen stellt sich oft die Frage, inwiefern das Eintreten eines Ereignisses A davon beeinflusst wird, ob ein anderes Ereignis B bereits eingetreten ist. Beim Lottospiel interessiert etwa die Wahrscheinlichkeit für 5 richtig vorausgesagte Zahlen (A), falls bekannt ist, dass der Spieler bereits zu den Gewinnern (mindestens drei richtige Voraussagen) gehört. Dabei interessiert vor allem der Zusammenhang von Ereignissen; beispielsweise kann das Eintreten von B dasjenige von A nach sich ziehen oder umgekehrt ausschliessen.

Beispiele:

- Wie gross ist die Wahrscheinlichkeit, dass eine Maschine in den nächsten 24 Stunden (A) ausfällt, wenn bekannt ist, dass sie schon 6 Tage fehlerfrei funktionierte (B)?

- Mit welcher Wahrscheinlichkeit besitzt eine zufällig ausgewählte Person blaue Augen (A), wenn bekannt ist, dass die Person blond ist (B)?

- Mit welcher Wahrscheinlichkeit ist eine zufällig befragte Person farbenblind (A), falls bekannt ist, dass es sich um eine Frau handelt (B)?

Es interessiert also, wie sich das Eintreten eines Ereignisses B auf die Wahrscheinlichkeit eines Ereignisses A auswirkt, oder allgemeiner, wie verschiedene Ereignisse voneinander abhängen.

Dieser Zusammenhang wird in der Wahrscheinlichkeitsrechnung durch die sogenannte "bedingte Wahrscheinlichkeit von A unter der Voraussetzung B" $P(A \mid B)$ zum Ausdruck gebracht.

Def. 4.1

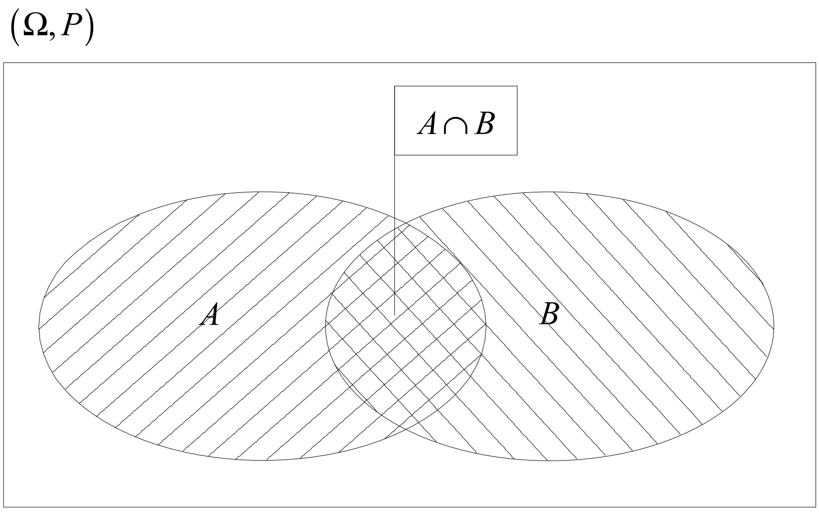
Die bedingte Wahrscheinlichkeit für A, falls B schon eingetreten ist, beträgt

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0 \tag{4.14}$$

Statistische Motivation:
Ein Experiment werde n-mal unter denselben Bedingungen wiederholt.
$n(B)$ und $n(A \cap B)$ seien die Anzahl Versuche, bei denen B resp. A und B gleichzeitig eingetreten sind. Dann beträgt die bedingte relative Häufigkeit

$$\frac{n(A \cap B)}{n(B)} = \frac{\frac{n(A \cap B)}{n}}{\frac{n(B)}{n}} \approx \frac{P(A \cap B)}{P(B)} = P(A|B) \tag{4.15}$$

Eine anschauliche Motivation gestatten auch Venn-Diagramme.



Das Wahrscheinlichkeitsmass über dem Durchschnitt $A \cap B$ wird an jenem über B (der Bedingung) gemessen.

Bei festem B definieren die bedingten Wahrscheinlichkeiten ein neues Wahrscheinlichkeitsmass über dem Wahrscheinlichkeitsraum $\Omega^* = B$.

Es gilt

$$\begin{aligned}
&1.)\ \ P(A|B)=\frac{P(A\cap B)}{P(B)}\geq 0\\
&2.)\ \ P(B|B)=1\\
&3.)\ \ P((A\cup C)|B)=P(A|B)+P(C|B),\quad \text{falls}\ \ A\cap C=\varnothing
\end{aligned}\tag{4.16}$$

denn

$$\begin{aligned}
P((A\cup C)|B)&=\frac{P((A\cup C)\cap B)}{P(B)}=\frac{P((A\cap B)\cup(C\cap B))}{P(B)}\\
&=P(A|B)+P(C|B)
\end{aligned}\tag{4.17}$$

<u>Beispiel 1</u>

Ein Kartenspiel besteht aus 4 · 9 = 36 Karten. Wir betrachten die Ereignisse

A: As ziehen

B: Pic ziehen

$$\begin{aligned}
&P(A)=\frac{4}{36},\quad P(B)=\frac{9}{36},\quad P(A\cap B)=\frac{1}{36}\\
&P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{36}}{\frac{9}{36}}=\frac{1}{9}
\end{aligned}\tag{4.18}$$

<u>Beispiel 2</u>

Wir betrachten alle Familien mit 2 Kindern. Die Wahrscheinlichkeit einer Knabengeburt sei 0.5. Von einer zufällig ausgewählten Familie sei bekannt, dass sie mindestens einen Knaben hat. Mit welcher Wahrscheinlichkeit besitzt die Familie genau 2 Knaben?

A: Familie hat 2 Knaben

B: Familie hat Knaben

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)}{P(B)}=\frac{\frac{1}{4}}{\frac{3}{4}}=\frac{1}{3}\tag{4.19}$$

4.2.3 Multiplikationssatz

Aus der Formel der bedingten Wahrscheinlichkeit erhält man

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \tag{4.20}$$

und es folgt die Verallgemeinerung für das gleichzeitige Eintreten von n beliebigen Ereignissen $A_1, A_2, \dots, A_n$

$$\begin{aligned} &P(A_1 \cap A_2 \cap \ldots \cap A_n) \\ &= P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdot \ldots . \cdot P(A_n | A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \end{aligned} \tag{4.21}$$

<u>Beweis:</u> mittels vollständiger Induktion
für $n = 3$ gilt

$$\begin{aligned} &P(A_1 \cap A_2 \cap A_3) = P((A_1 \cap A_2) \cap A_3) = P(A_1 \cap A_2) \cdot P(A_3 | A_1 \cap A_2) \\ &= P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \end{aligned} \tag{4.22}$$

<u>Beispiel 1</u>

Beim Schweizer Zahlenlotto werden aus 45 Kugeln zufällig 6 ohne Zurücklegen gezogen. Wie gross ist die Wahrscheinlichkeit für den ersten Gewinnrang? (alle 6 Zahlen richtig vorausgesagt)

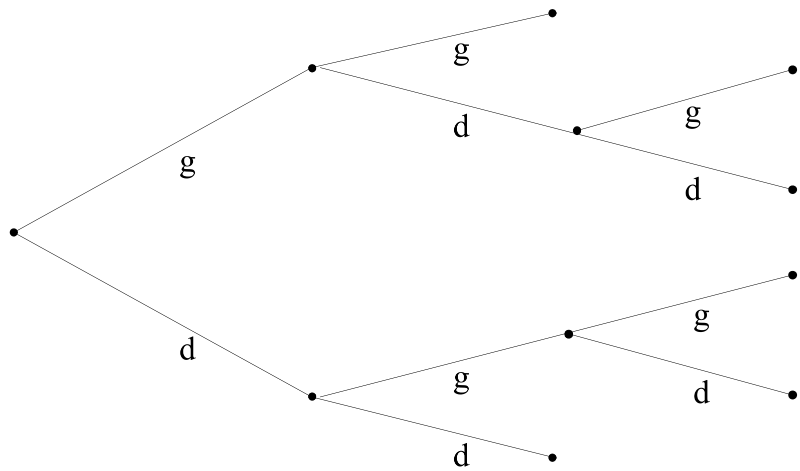
A_i: i-te Kugel stimmt mit einer angekreuzten Zahl überein.

$$\begin{aligned} &P(A_1 \cap A_2 \ldots . \cap A_6) = P(A_1) \cdot P(A_2 | A_1) \cdot \ldots . \cdot P(A_6 | A_1 \cap \ldots \cap A_5) \\ &= \frac{6}{45} \cdot \frac{5}{44} \cdot \frac{4}{43} \cdot \frac{3}{42} \cdot \frac{2}{41} \cdot \frac{1}{40} = \frac{1}{\binom{45}{6}} = \frac{1}{8'145'060} = 0.1227738 \cdot 10^{-6} \end{aligned} \tag{4.23}$$

<u>Beispiel 2</u>

2 defekte Bauelemente werden mit 2 guten zuammengepackt. Die Elemente sollen solange getestet werden, bis die beiden defekten bekannt sind.

Mit welcher Wahrscheinlichkeit kann man die defekten Elemente nach dem 2. resp. 3. Versuch bestimmen?



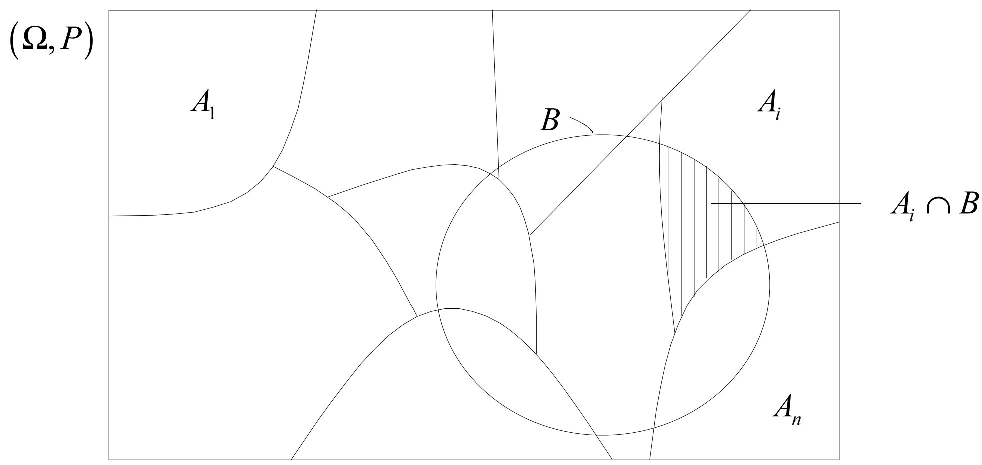
4.2.4 Satz von der totalen Wahrscheinlichkeit

Die Ereignisse $A_1, A_2, \ldots, A_n$ seien ein vollständiges System, d.h.

$$\bigcup_{i=1}^{n} A_i = \Omega \,; \quad A_i \cap A_j = \emptyset, \quad \text{für } i \neq j \tag{4.24}$$

(Jedes Element von Ω ist in genau einem A_i enthalten)

B sei ferner ein beliebiges Ereignis



Dann gilt

$$\begin{aligned} B &= (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_n \cap B) \\ P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B) \\ &= P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + \ldots + P(B|A_n) \cdot P(A_n) \end{aligned} \tag{4.25}$$

Satz 4.9

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

(4.26)

$$\text{für} \quad n = 2: \quad P(B) = P(B|A)\,P(A) + P(B|A^c)\,P(A^c)$$

Beispiel 1:
Eine Firma kauft ein Gut von 3 Lieferanten ein. Lieferant 1 produziert defekte Elemente mit einer Wahrscheinlichkeit von 0.06, Lieferant 2 mit einer Wahrscheinlichkeit 0.1 und Lieferant 3 mit einer solchen von 0.05. Das Gesamtlager besteht zu 40% von Lieferant 1, zu 25% von Lieferant 2 und zu 35% von Lieferant 3.

Wie gross ist die Wahrscheinlichkeit, dass ein zufällig dem Lager entnommenes Element defekt ist?

B: Element ist defekt.
A_i: Element stammt vom i-ten Lieferanten.
Lösung: $P(B) = 0.0665$

Beispiel 2:
Farbenblindheit ist ein geschlechtsspezifisches Merkmal. Die Wahrscheinlichkeit für Farbenblindheit sei 1% bei den Frauen und 8% bei den Männern. Wie gross ist die Wahrscheinlichkeit für Farbenblindheit in einer Population, die zu 52% aus Frauen und zu 48% aus Männern besteht?

F: zufällig ausgewählte Person ist farbenblind
Lösung: $P(F) = 0.0436$

4.2.5 Bayes'sche Formel

Gegeben seien 2 Urnen U_1 und U_2. U_1 enthalte 5 weisse und 7 rote Kugeln, U_2 1 weisse und 5 rote. Aus einer zufällig ausgewählten Urne wird ebenfalls zufällig eine Kugel entnommen. Diese sei rot. Mit welcher Wahrscheinlichkeit wurde U_1 gewählt?

	U_1 (5w, 7r)	U_2 (1w, 5r)
a priori $P(U_i)$	$\frac{1}{2}$	$\frac{1}{2}$
bedingte WS für rote Kugel $P(R\|U_i)$	$\frac{7}{12}$	$\frac{5}{6}$
$P(U_i \cap R) = P(R\|U_i) \cdot P(U_i)$	$\frac{7}{24}$	$\frac{5}{12}$
$P(R) = P(U_1 \cap R) + P(U_2 \cap R)$	$\frac{7}{24} + \frac{5}{12} = \frac{17}{24}$	
a posteriori $P(U_i \| R) = \frac{P(U_i \cap R)}{P(R)}$	$\frac{7}{17}$	$\frac{10}{17}$

Die Wahrscheinlichkeit, Urne 1 zu ziehen, hat sich durch die Zusatzinformation, dass eine rote Kugel gezogen wurde, von $\frac{1}{2}$ auf $\frac{7}{17}$ reduziert. Im Gegensatz zu den ursprünglichen a priori–Wahrscheinlichkeiten nennt man die durch die Beobachtung modifizierte Wahrscheinlichkeit die sogenannte a posteriori–Wahrscheinlichkeit. Das obige Beispiel motiviert den Satz von Bayes.

Satz 4.10

$A_1, A_2, \ldots, A_n$ bilden ein vollständiges System von Ω und $B \subset \Omega$ sei ein beliebiges Ereignis. Dann gilt

$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(B|A_j) \cdot P(A_j)} \tag{4.27}$$

Der Beweis des Satzes von Bayes folgt unmittelbar aus der Definition der bedingten Wahrscheinlichkeit sowie aus dem Satz von der totalen Wahrscheinlichkeit.

Praktisch bedeutsam ist die Tatsache, dass man bekannte Wahrscheinlichkeiten aufgrund zusätzlicher Beobachtungen anpassen kann. Auffallend und zugleich bemerkenswert am Satz von Bayes ist sicher die Umkehrung von Ereignis und Bedingung. Die limitierte praktische Anwendung des Satzes von Bayes liegt primär in der Unkenntnis der a priori–Wahrscheinlichkeiten.

Die Bayes'sche Formel kann wie folgt interpretiert werden: Es sei bekannt, dass ein Ereignis B (Wirkung) eingetreten ist, das nur gemeinsam mit einem Ereignis eines vollständigen Systems $A_1, A_2, \ldots, A_n$ (Ursachen) auftreten kann. Welche der Ursachen A_i das Ereignis B hervorgerufen hat, ist jedoch unbekannt. Kennt man aber die Wahrscheinlichkeiten, mit denen jede der Ursachen A_i das Ereignis B bewirkt, so gestattet der Satz von Bayes die Berechnung der Wahrscheinlichkeit, dass B durch A_i bedingt wurde. Damit können die Alternativen, dass B durch A_i bewirkt wurde, objektiv bemessen werden.

Beispiel 1:
Ein Autopark bestehe aus drei Typen A, B und C und zwar zu 20%, 30% resp. 50%. Die Wahrscheinlichkeiten für einen Defekt seien 0.1, 0.08 und 0.05. Für die Reparatur ist es wichtig zu wissen, um welchen Typ es sich handelt. Es trifft eine Defektmeldung D ein. Wie gross sind die Typenwahrscheinlichkeiten?

$$\begin{aligned} P(A|D) &= \frac{P(D|A)\cdot P(A)}{P(D|A)\cdot P(A)+P(D|B)\cdot P(B)+P(D|C)\cdot P(C)} \\ &= \frac{0.1\cdot 0.2}{0.1\cdot 0.2+0.08\cdot 0.3+0.05\cdot 0.5} = 0.290 \end{aligned} \tag{4.28}$$

$P(B \mid D) = 0.348$ $\qquad$ $P(C \mid D) = 0.362$

Beachte: Die Summe der Bayeswahrscheinlichkeiten beträgt 1.

Beispiel 2:
In einer grossen Bevölkerung leide 1‰ unter einer ansteckenden Krankheit. Die Diagnose der Krankheit erfolge unter folgenden Bedingungen. Von den Kranken werden 90% und von den Gesunden 99% nach demselben Prozedere richtig diagnostiziert.

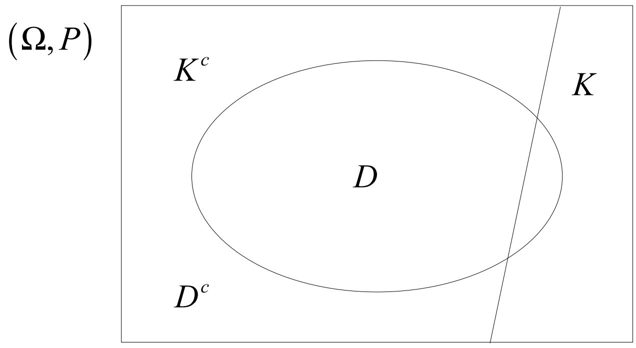
Eine zufällig ausgewählte Person wird als krank diagnostiziert. Mit welcher Wahrscheinlichkeit ist sie tatsächlich krank?

Mit den Ereignissen

K: krank
D: Diagnose positiv

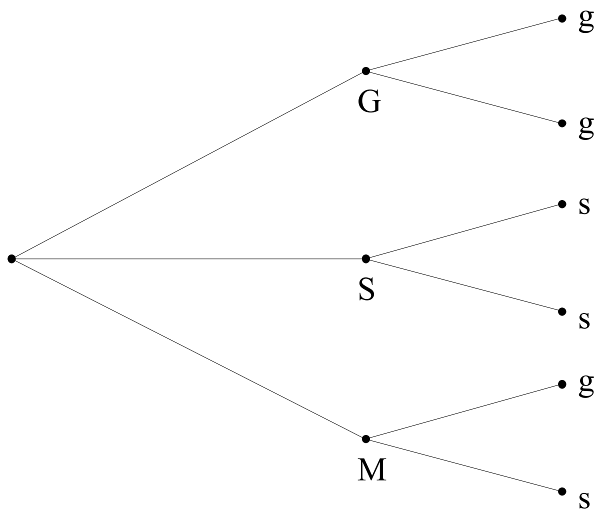
findet man nach dem Satz von Bayes

$$P(K|D) = \frac{P(D|K)\cdot P(K)}{P(D|K)\cdot P(K) + P(D|K^c)\cdot P(K^c)} = 0.083 \tag{4.29}$$



Bertrand's Paradoxon

Gegeben sind 3 Schatzkästlein mit je 2 Schubladen. 1 Kästlein hat in beiden Schubladen Goldmünzen, ein Kästlein in beiden Schubladen Silbermünzen und ein Kästlein in der einen Schublade Silber- und in der anderen Goldmünzen. Es wird zufällig ein Kästlein ausgewählt und eine ebenfalls zufällig ausgewählte Schublade geöffnet. Das Resultat ist Gold. Mit welcher Wahrscheinlichkeit birgt die andere Schublade auch einen Goldschatz?



G	Goldkästlein
S	Silberkästlein
M	Gold-Silberkästlein
g	Goldschublade

Gesucht: P(G|g)

4.2.6 Unabhängige Ereignisse

Sind A und B zwei unvereinbare Ereignisse ($A \cap B = \varnothing$), so gilt

$$P(A|B) = 0 \tag{4.30}$$

Gilt hingegen $B \subset A$, so folgt

$$P(A|B) = 1 \tag{4.31}$$

In beiden Fällen können mit der Information B feste Aussagen über die bedingten Wahrscheinlichkeiten gemacht werden. Wie das nachfolgende Beispiel zeigt, gibt es aber auch Situationen, bei denen die Tatsache, dass ein Ereignis B bereits eingetreten ist, die Wahrscheinlichkeit für das Ereignis A nicht beeinflusst.

ε : Doppelwürfel
A: Augenzahl auf Würfel 1 ungerade
B: Augenzahl auf Würfel 2 grösser als 4

Die beiden Ereignisse A und B stehen insofern nicht in Beziehung zueinander, als die Information über das Ereignis B die Wahrscheinlichkeit für das Ereignis A nicht beeinflusst.

$$
\begin{gathered}
P(A) = \frac{1}{2} \qquad P(B) = \frac{1}{3} \qquad P(A \cap B) = \frac{1}{6} \\
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2} = P(A)
\end{gathered}
\tag{4.32}
$$

Die bedingte und die unbedingte Wahrscheinlichkeit von A stimmen überein. Ebenfalls gilt

$$
P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3} = P(B)
\tag{4.33}
$$

Sind $P(A) > 0$ und $P(B) > 0$, so folgt aus den letzten beiden Beziehungen

$$
P(A \cap B) = P(A) \cdot P(B)
\tag{4.34}
$$

genau dann, wenn die bedingten und unbedingten Wahrscheinlichkeiten der Ereignisse A und B übereinstimmen, die Ereignisse also unabhängig voneinander sind.

Def. 4.2

Zwei Ereignisse A und B sind genau dann unabhängig, wenn gilt

$$
P(A \cap B) = P(A) \cdot P(B)
\tag{4.35}
$$

Satz 4.11

Die Ereignisse A und B mit $P(A) > 0$ und $P(B) > 0$ sind genau dann unabhängig, wenn eine der Bedingungen

$$\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B) \end{aligned} \tag{4.36}$$

erfüllt ist.

Beispiel:

ε : Doppelwürfel

A: Augensumme gerade

B: Augenzahl auf dem ersten Würfel 4

C: Augensumme 8

$$\begin{array}{lll} P(A) = \dfrac{1}{2} & P(B) = \dfrac{1}{6} & P(C) = \dfrac{5}{36} \\ P(A \cap B) = \dfrac{1}{12} & P(A \cap C) = \dfrac{5}{36} & P(B \cap C) = \dfrac{1}{36} \end{array} \tag{4.37}$$

Beachte: Unabhängigkeit und Unvereinbarkeit sind zwei völlig verschiedene Konzepte, die allerdings in Beziehung zueinander stehen. Unvereinbare Ereignisse mit positiven Wahrscheinlichkeiten sind nämlich stets abhängig. (Nachweis?)

Satz 4.12

Sind A und B unabhängige Ereignisse, so sind auch A und B^c, A^c und B, sowie A^c und B^c unabhängig.

	B	B^c	
A	$P(A) \cdot P(B)$	$P(A) \cdot P(B^c)$	$P(A)$
A^c	$P(A^c) \cdot P(B)$	$P(A^c) \cdot P(B^c)$	$P(A^c)$
	$P(B)$	$P(B^c)$	1

Beweis für die Ereignisse A und B^c

$$A = (A \cap B) \cup (A \cap B^c)$$
$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A) \cdot P(B) + P(A \cap B^c) \quad (4.38)$$
$$P(A \cap B^c) = P(A) - P(A) \cdot P(B) = P(A)(1 - P(B)) = P(A) \cdot P(B^c)$$

Es ist sehr schwierig, die Frage nach Unabhängigkeit rein gefühlsmässig zu beantworten. Die Unabhängigkeit ist nämlich keine Eigenschaft der Ereignisse selber, sondern leitet sich aus den Wahrscheinlichkeiten der Ereignisse ab.

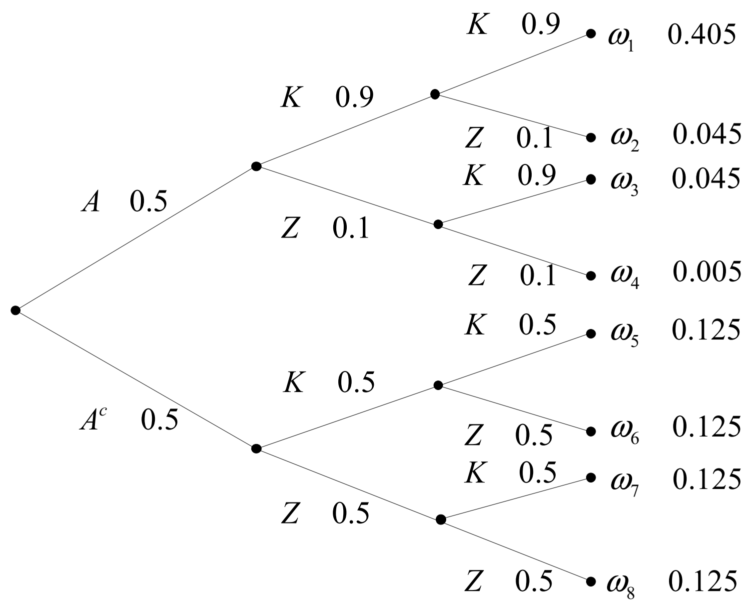
Beispiel: (Bühlmann)
Eine Urne enthält 2 Sorten Münzen, gleichviele von jeder Sorte. Die Münzen der ersten Sorte fallen mit Wahrscheinlichkeit 0.9 auf Kopf, jene der zweiten Sorte sind symmetrisch.

ε : Eine zufällig gezogene Münze wird zweimal geworfen.

Ereignisse:
K_1: Kopf beim ersten Wurf
K_2: Kopf beim zweiten Wurf
A: Münze der ersten Sorte gezogen
A^c : Münze der zweiten Sorte gezogen

Sind die Ereignisse K_1 und K_2 unabhängig?
Intuitive Antwort?

Wir stellen ε in einem Baumdiagramm dar.



$$\begin{aligned}
&K_1 = \{\omega_1, \omega_2, \omega_5, \omega_6\} && P(K_1) = 0.7 \\
&K_2 = \{\omega_1, \omega_3\ \omega_5, \omega_7\} && P(K_2) = 0.7 \\
&K_1 \cap K_2 = \{\omega_1, \omega_5\} && P(K_1 \cap K_2) = 0.53
\end{aligned} \tag{4.39}$$

$$P(K_1 \cap K_2) \neq P(K_1) \cdot P(K_2)$$

$\Rightarrow K_1$ und K_2 abhängig!

Unabhängigkeit von mehreren Ereignissen

Def. 4.3

Die Ereignisse $A_1, A_2 \ldots, A_n$ heissen genau dann gemeinsam unabhängig, wenn

$$
\begin{array}{lll}
P\left(A_i \cap A_j\right) & = P\left(A_i\right) \cdot P\left(A_j\right) & i \neq j \\
P\left(A_i \cap A_j \cap A_k\right) & = P\left(A_i\right) \cdot P\left(A_j\right) \cdot P\left(A_k\right) & i \neq j, \quad j \neq k, \quad i \neq k \\
\cdot & & \\
\cdot & & \\
\cdot & & \\
P\left(A_1 \cap A_2 \cap \ldots \cap A_n\right) & = P\left(A_1\right) \cdot P\left(A_2\right) \cdot \ldots \cdot P\left(A_n\right) &
\end{array}
\tag{4.40}
$$

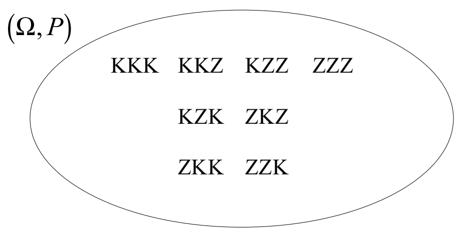
Insgesamt werden $2^n - n - 1$ Gleichungen benötigt.

Beispiel

Drei Personen vereinbaren folgendes Spiel. Eine symmetrische Münze wird 3mal geworfen.

A: Spieler *A* gewinnt Fr. 1.-, falls 1. Wurf Kopf
B: Spieler *B* gewinnt Fr. 1.-, falls 2. Wurf Kopf
C: Spieler *C* gewinnt Fr. 1.-, falls 3. Wurf Kopf

Man überprüfe, ob die Ereignisse *A*, *B* und *C* voneinander unabhängig sind.



$$P(A) = P(B) = P(C) = \frac{4}{8} = \frac{1}{2}$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{2}{8} = \frac{1}{4}$$

$$P(A \cap B \cap C) = \frac{1}{8}$$

A, *B*, *C* sind gemeinsam unabhängig.

Beachte:
Es müssen stets alle in der Definition aufgeführten Bedingungen erfüllt sein. Für 3 Ereignisse folgt z.B. aus der paarweisen Unabhängigkeit nicht automatisch die gemeinsame Unabhängigkeit.

<u>Beispiel:</u>
ε : Eine symmetrische Münze wird 3 mal geworfen.

A: Münze zeigt immer dasselbe Bild
B: 2. Wurf Kopf
C: 1. Wurf Zahl

$$P(A) = \frac{1}{4} \qquad P(B) = P(C) = \frac{1}{2}$$

$$P(A \cap B) = \frac{1}{8} = P(A) \cdot P(B) \qquad A, B \text{ unabhängig}$$

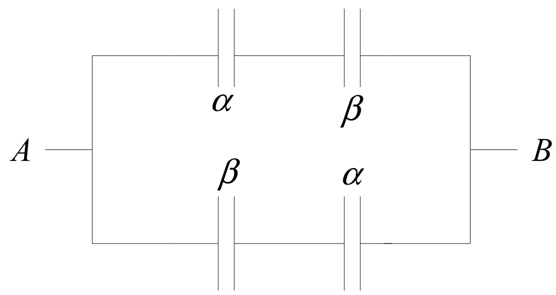
$$P(A \cap C) = \frac{1}{8} = P(A) \cdot P(C) \qquad A, C \text{ unabhängig}$$

$$P(B \cap C) = \frac{1}{4} = P(B) \cdot P(C) \qquad B, C \text{ unabhängig}$$

aber $\quad A \cap B \cap \mathrm{C} = \varnothing$

<u>Beispiel:</u>
Die Übertragung einer Nachricht von A nach B erfolgt über 2 voneinander unabhängige Kanäle. Diese sind je zuammengesetzt aus 2 unabhängigen Aggregaten, die mit Wahrscheinlichkeit α resp. β ausfallen können.

Man berechne die Wahrscheinlichkeit, dass die Übertragung zustandekommt.



5. (EINDIMENSIONALE) ZUFALLSVARIABLEN

5.1 Diskrete Zufallsvariable

Gegenstand der bisherigen Betrachtungen waren Wahrscheinlichkeitsräume (Ω, P) im Zusammenhang mit Zufallsexperimenten ε. Man hat es aber häufig mit Experimenten zu tun, bei denen primär nicht der Ausgang an sich, sondern Zahlenwerte, die aus den jeweiligen Ausgängen resultieren, interessieren. Wirft man z.B. eine Münze 3mal, so kann man sich dafür interessieren, wie oft "Kopf" (unter Vernachlässigung der Reihenfolge) in der Wurfsequenz vorkommt. Man gibt (i.a. überflüssige) Informationen preis.

ε: 3 symmetrische Münzen werden geworfen.

Ω	Anzahl K
ZZZ	0
ZZK	1
ZKZ	1
KZZ	1
ZKK	2
KZK	2
KKZ	2
KKK	3

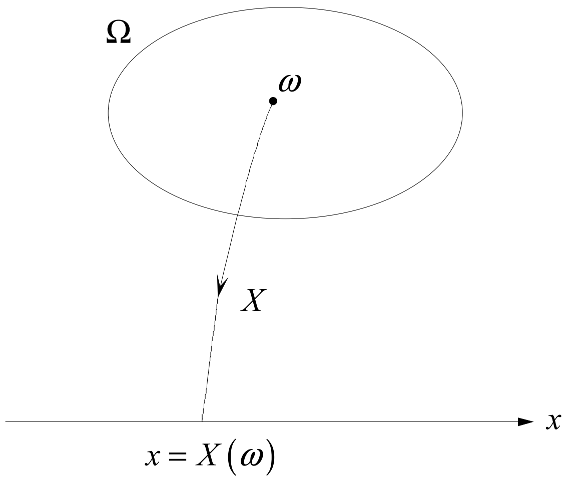
Jedem Ausgang von ε wird genau eine Zahl zugeordnet. Der Wahrscheinlichkeitsraum (Ω, P) wird auf die reelle Achse abgebildet.

Def. 5.1

Die Zufallsvariable X ist eine *Abbildung*, die jedem Element $\omega \in \Omega$ des Wahrscheinlichkeitsraumes eine reelle Zahl $x \in \mathbb{R}$ zuordnet.

Die Abbildung X des Wahrscheinlichkeitsraumes impliziert, dass neben der Zuordnung von Zahlen $x \in \mathbb{R}$ zu Ereignissen $\omega \in \Omega$ gleichzeitig auch das Wahrscheinlichkeitsmass $P(\omega)$ auf die reelle Achse übertragen wird.

Eine Zufallsvariable heisst *diskret*, wenn die Wertemenge der Abbildung abzählbar viele Werte enthält.



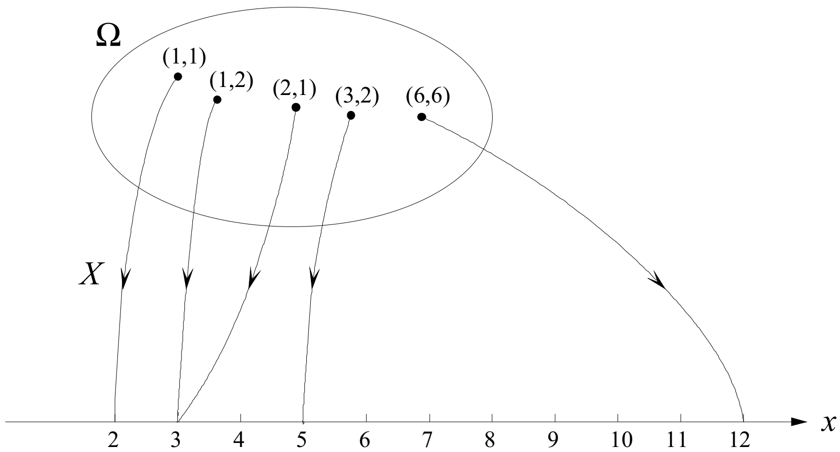
<u>Beispiel:</u>

ε : Doppelwürfel

Ω: $\{(i,j) \mid i,j \in \{1,2,\ldots,6\}\}$

X: Summe der beiden Augenzahlen

X: $(i,j) \to i+j$ Jedem Elementarereignis wird nach der Vorschrift X genau eine reelle Zahl zugeordnet.



Bezeichnung: $X: \omega \to x$ oder $X(\omega) = x$

X ist eine über dem Wahrscheinlichkeitsraum definierte Funktion eines Zufallsexperimentes. Durch sie wird festgelegt, welche reellen Zahlen den möglichen Ausgängen des Zufallsexperimentes zugeordnet werden.

$W(X) = (x_1, x_2, \ldots, x_n)$ heisst Wertebereich der Funktion *X*. Wir werden künftig voraussetzen, dass $x_1 < x_2 < \ldots < x_n$.

Die Bezeichnung einer Funktion als Zufallsvariable ist nicht besonders glücklich. Die ältere und auch eingebürgerte Bezeichnung "Zufallsvariable" ist mit der Vorstellung verbunden, dass eine Grösse verschiedene (variable) Werte annehmen kann, wobei im Einzelfall der Zufall entscheidet, welchen Wert sie annimmt.

Bei Zufallsexperimenten interessieren allgemein die Wahrscheinlichkeiten, mit denen bestimmte Ereignisse eintreten. Durch die Abbildung *X* betrachten wir nicht mehr die Ausgänge selber, sondern damit in Zusammenhang stehende reelle Zahlen. Wir müssen daher Wahrscheinlichkeiten in der Menge der reellen Zahlen festlegen.

5.1.1 Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen

X sei eine diskrete Zufallsvariable über dem Wahrscheinlichkeitsraum (Ω,*P*). Zu bestimmen sind die Wahrscheinlichkeiten, mit denen *X* die Realisationen $(x_1, x_2, \ldots, x_n)$ annimmt.

$A_i \subset \Omega$ umfasse *genau* jene Elemente von Ω mit $X(\omega) = x_i$

$$A_i = \{\omega | X(\omega) = x_i\} \qquad i = 1, 2, ..., n \tag{5.1}$$

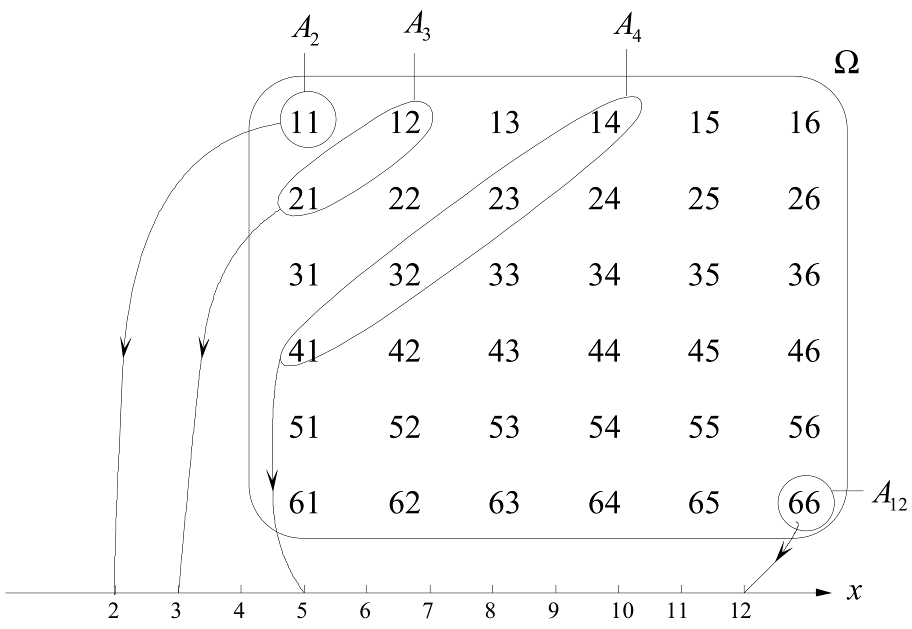
Die Frage nach der Wahrscheinlichkeit, dass die Zufallsvariable *X* in der Realisation x_i vorliegt, ist dann identisch mit der Frage der Eintretenswahrscheinlichkeit für das Ereignis A_i.

Beispiel:

ε : Doppelwürfel

X: $(i, j) \rightarrow i + j$

$A_k = \{(i, j) \mid i + j = k\}$



Durch die Funktion X wird das Wahrscheinlichkeitsmass P von Ω auf die reelle Achse übertragen.

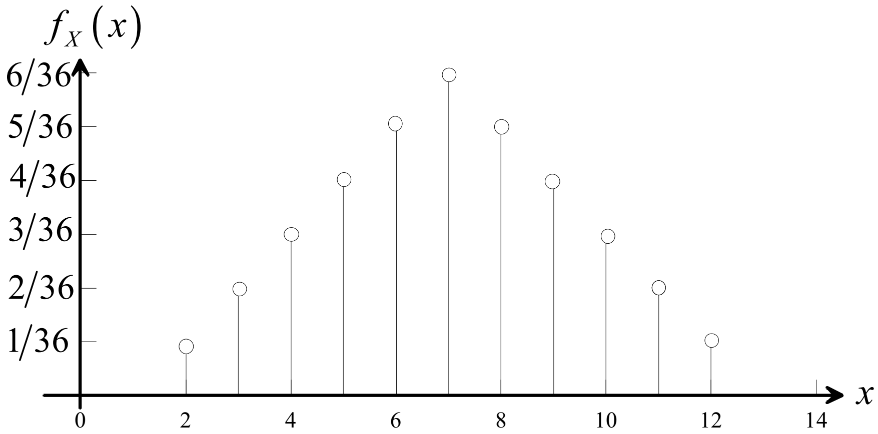
Die Wahrscheinlichkeitsfunktion einer Zufallsvariablen X beschreibt die Wahrscheinlichkeit, dass X einen bestimmten Wert x_i annimmt: $P(X = x_i)$.

Def. 5.2

$$f_X(x) = \begin{cases} P(X = x_i) = P(A_i) & \text{falls } x = x_i \\ \quad \text{mit } A_i = \{\omega | X(\omega) = x_i\} & \\ 0 & \text{falls } x \neq x_i \end{cases} \tag{5.2}$$

Für das Beispiel des Doppelwürfels erhält man

x	2	3	4	5	6	7	8	9	10	11	12
$f_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



Beispiel:
Eine Maschine produziert defekte Elemente mit einer Wahrscheinlichkeit von 0.05. Der Produktion werden zufällig 4 Elemente entnommen.

X sei die Anzahl defekter Elemente in der Stichprobe vom Umfang $n = 4$.

Ω	x	$f_X(x)$
GGGG	0	

5.1.2 Verteilungsfunktion einer diskreten Zufallsvariablen

Die Verteilungsfunktion $F_X(x)$ beschreibt die Wahrscheinlichkeit, dass eine Zufallsvariable X einen vorgegebenen Wert x nicht übersteigt.

$$F_X(x) = P(X \le x) = \sum_{x_i \le x} f_X(x_i) \tag{5.3}$$

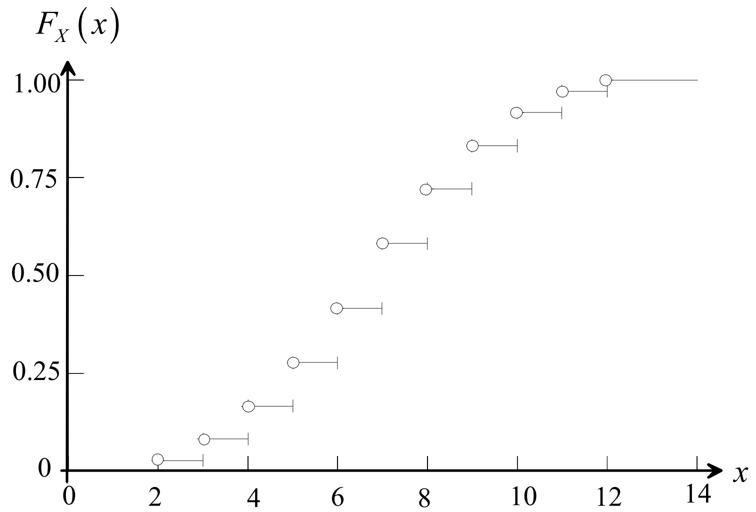
Summiert werden jene Wahrscheinlichkeiten $f_X(x_i)$, welche die Bedingung $x_i \le x$ erfüllen.

Beispiel:

ε: symmetrischer Doppelwürfel

X: Summe der Augenzahlen

x	2	3	4	5	6	7	8	9	10	11	12
$f_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F_X(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$



Eigenschaften der Verteilungsfunktion

1.) $0 \le F_X(x) \le 1$

2.) $F_X(x) = 0$ für $x < x_{\min}$

$F_X(x) = 1 \qquad \text{für} \;\; x \geq x_{\max}$

3.) $\quad F_X(x_{i-1}) + f_X(x_i) = F_X(x_i)$

4.) $\quad P(a < X \leq b) = F_X(b) - F_X(a) \qquad \text{für} \;\; a \leq b$

Beweis von 4.)

$$\begin{aligned} P(X \leq b) &= P(X \leq a) + P(a < X \leq b) \\ F_X(b) &= F_X(a) + P(a < X \leq b) \\ P(a < X \leq b) &= F_X(b) - F_X(a) \qquad \text{q.e.d} \end{aligned} \tag{5.4}$$

5.) $F_X(x)$ ist eine stückweise konstante, monoton steigende, rechtsstetige Treppenfunktion mit Sprüngen an den Stellen x_i von der Höhe $f_X(x_i)$.

<u>Beispiel:</u>
Eine Zufallsvariable besitze die Wahrscheinlichkeitsfunktion

$$f_X(x) = \begin{cases} k & \text{für} \;\; x = 0 \\ 2k & \text{für} \;\; x = 1 \\ 3k & \text{für} \;\; x = 2 \\ 5k & \text{für} \;\; x = 3 \\ 0 & \text{sonst} \end{cases} \tag{5.5}$$

1.) Man bestimme k

2.) $P(1 < X \leq 3),\, P(X > 1),\, P(X = 1)$

3.) Welches ist der kleinste Wert von X, für den gilt:

$$P(X \leq x) = F_X(x) > 0.5 \tag{5.6}$$

5.2 Stetige Zufallsvariable

Jemand möchte mit dem Bus zum Bahnhof fahren. Er kennt jedoch den Fahrplan nicht, weiss aber, dass die Verkehrsbetriebe einen 10-Minutentaktfahrplan aufrechterhalten. Die betreffende Person kommt zur Haltestelle und interessiert sich für die Wartezeit. Letztere kann als Zufallsvariable X mit folgenden Eigenschaften aufgefasst werden.

1.) X kann (zumindest theoretisch) im Intervall [0,10] jeden Wert annehmen.

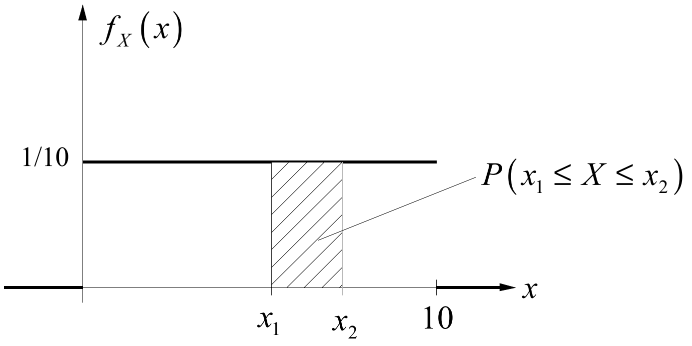
2.) $P(X = x) = 0$ Wahrscheinlichkeit einer ganz bestimmten Realisation ist Null (aber kein unmögliches Ereignis)

3.) $P(0 \leq X \leq 10) = 1$ Sicheres Ereignis

4.) $P(x \leq X \leq x + d) =$ konst. Die Wahrscheinlichkeit einer Wartezeit der Länge d ist für alle Zeitpunkte x im Intervall $[0, 10 - d]$ konstant und hängt nur von d ab.

Diese Forderungen sind erfüllt, wenn wir Wahrscheinlichkeiten als Flächen (vgl. Histogramm) unter einer nicht negativen Funktion f betrachten. Im vorliegenden Fall erfüllt folgende Funktion die obigen Bedingungen.

$$f_X(x) = \begin{cases} \dfrac{1}{10} & 0 \leq x \leq 10 \\ 0 & \text{sonst} \end{cases} \tag{5.7}$$



$P(5 \leq X \leq 7)$
$= P(8 \leq X \leq 10) = 0.2$
$P(X \leq 3) = 0.3$
$P(X \leq 6) = 0.6$

$f_X(x)$ heisst Dichtefunktion einer stetig gleichverteilten Zufallsvariablen X.

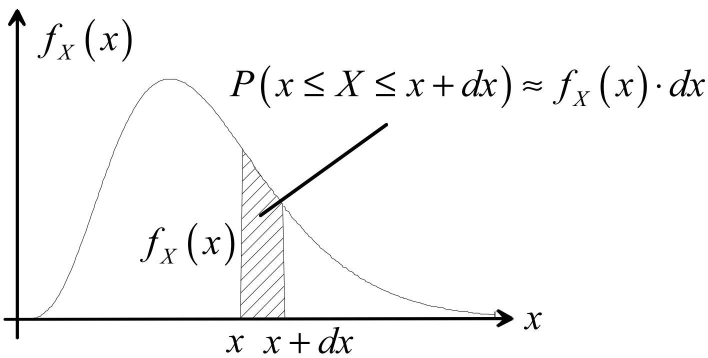
Die folgende Definition verallgemeinert das im vorgängigen Beispiel beschriebene Konzept.

Def. 5.3

Eine Zufallsvariable X besitzt eine stetige Wahrscheinlichkeitsverteilung, wenn eine auf der Menge der reellen Zahlen integrierbare Funktion f so existiert, dass für alle x gilt:

$$\begin{aligned} &1.\quad f_X(x) \geq 0 \\ &2.\quad P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x)\,dx \\ &3.\quad \int_{-\infty}^{+\infty} f_X(x)\,dx = 1 \end{aligned} \tag{5.8}$$

$f_X(x)$ heisst Wahrscheinlichkeits*dichte*funktion.



Def. 5.4

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt \tag{5.9}$$

heisst Verteilungsfunktion der stetigen Zufallsvariablen X.

Verteilungs- und Dichtefunktionen einer Zufallsvariablen lassen sich wie im diskreten Fall auseinander ableiten. Es gilt

$$\frac{dF_X(x)}{dx} = F_X'(x) = f_X(x) \tag{5.10}$$

Beachte:

- Wahrscheinlichkeitsdichten f_X sind *keine* Wahrscheinlichkeiten. Nur *Flächen* unter der Dichtefunktion sind Wahrscheinlichkeiten.

- Die Wahrscheinlichkeit eines unmöglichen Ereignisses ist Null.
  Ist die Wahrscheinlichkeit eines Ereignisses Null, so folgt daraus nicht, dass es sich um ein unmögliches Ereignis handelt.
  Ist x_0 eine mögliche Ausprägung einer stetigen Zufallsvariablen, so gilt

$$P(X = x_0) = \int_{x_0}^{x_0} f_X(x)\, dx = 0 \tag{5.11}$$

- Aus $P(X = a) = P(X = b) = 0$ folgt

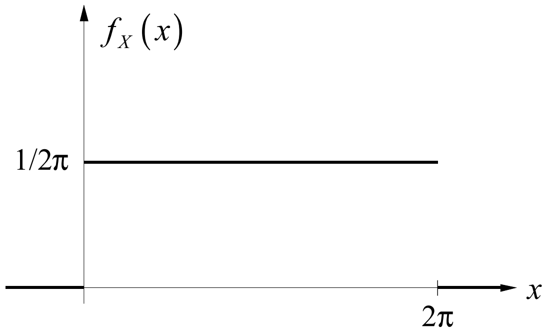
$$\begin{aligned} P(a < X \le b) &= P(a \le X \le b) = P(a \le X < b) \\ &= P(a < X < b) = F(b) - F(a) \end{aligned} \tag{5.12}$$

Beispiel:
Die Stellung des Zeigers im Einheitskreis lässt sich durch eine Zufallsvariable

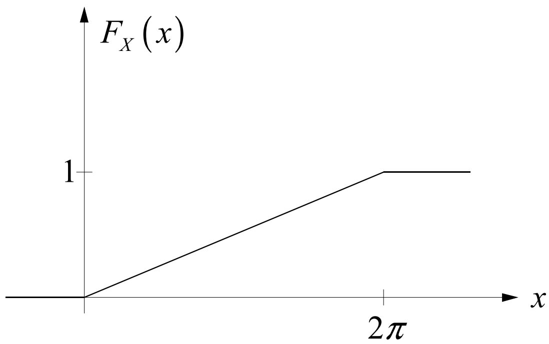
X: Winkel des Zeigers mit einer festen Koordinatenachse

beschreiben. Ist der Drehmechanismus fair, so wird keine Position bevorzugt, die Dichte ist für alle x, $0 \le x \le 2\pi$, konstant. X heisst dann eine *gleichverteilte Zufallsvariable* über dem Intervall $[0, 2\pi]$.

$$f_X(x) = \begin{cases} \frac{1}{2\pi} & 0 \le x \le 2\pi \\ 0 & \text{sonst} \end{cases} \tag{5.13}$$



$$F_X(x) = \begin{cases} 0 & x < 0 \\ \int_0^x \frac{1}{2\pi}\,dt = \frac{x}{2\pi} & 0 \le x \le 2\pi \\ 1 & x > 2\pi \end{cases} \qquad (5.14)$$



Wir haben bis anhin Dichtefunktionen auf mehrheitlich intuitiver Basis kennengelernt. Das nachfolgende Beispiel soll zeigen, dass Dichten als natürliches Resultat eines Experimentes unter zwar idealisierten, jedoch realistischen Bedingungen hervorgehen können. Wir betrachten ein Experiment, bei dem sich der Bestand mit konstanter Zerfallsrate auflöst, d.h. bei dem der relative Abbau pro Zeiteinheit α konstant bleibt.

$y(t)$ sei der Bestand zum Zeitpunkt t.

$$\frac{\frac{\Delta y(t)}{y(t)}}{\Delta t} = \frac{\frac{\Delta y(t)}{\Delta t}}{y(t)} = -\alpha$$

$$\Rightarrow \quad \frac{\frac{dy(t)}{dt}}{y(t)} = -\alpha \tag{5.15}$$

$$\Rightarrow \quad \frac{dy(t)}{y(t)} = -\alpha\, dt$$

Als Lösung dieser Differentialgleichung findet man

$$\begin{aligned} \ln\, y(t) &= -\alpha t + c \\ \text{resp.} \quad y(t) &= e^{-\alpha t + c} = ke^{-\alpha t} \end{aligned} \tag{5.16}$$

Definieren wir die (stetige) Zufallsvariable T als Zeit bis zum Zerfall, so gilt für alle Elemente, welche zum Zeitpunkt t dem Bestand angehören

$$T > t$$

Der *Anteil* dieser Elemente bezogen auf den Anfangsbestand beträgt

$$\frac{y(t)}{y(0)} = \frac{ke^{-\alpha t}}{k} = e^{-\alpha t} \tag{5.17}$$

und man schliesst

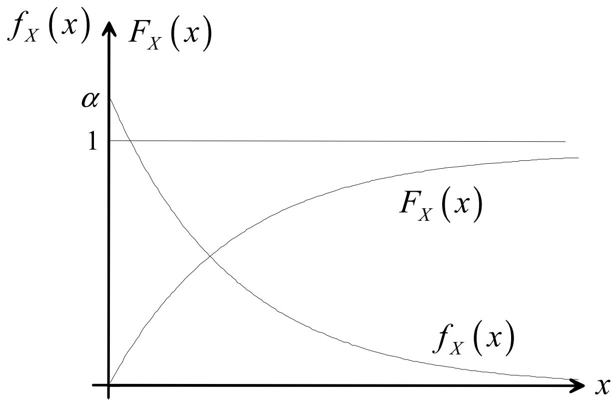
$$\begin{aligned} P(T > t) &= e^{-\alpha t} \\ P(T \le t) &= F_T(t) = 1 - e^{-\alpha t} \\ \textit{resp.} \quad f_T(t) &= \frac{dF_T(t)}{dt} = \alpha e^{-\alpha t} \end{aligned} \tag{5.18}$$

Mit der üblichen Notation ($T \to X$) heissen

$$\begin{aligned} f_X(x) &= \alpha e^{-\alpha x} \qquad \alpha > 0, \quad x \geq 0 \\ F_X(x) &= 1 - e^{-\alpha x} \end{aligned} \tag{5.19}$$

Dichte resp. Verteilungsfunktionen einer *exponential*-verteilten Zufallsvariablen.

Man bestätigt sofort, dass sämtliche Voraussetzungen an eine Dichte erfüllt sind.



Die Zeit, bei der die Zerfallswahrscheinlichkeit z.B. eines radioaktiven Atoms gerade 0.5 beträgt, nennt man die sog. Halbwertzeit T_h.

Man bestimmt T_h so, dass

$$\begin{aligned} P(X \leq T_h) = F_X(T_h) = 1 - e^{-\alpha T_h} &= 0.5 \\ T_h = \frac{-1}{\alpha} \ln 0.5 &= \frac{\ln 2}{\alpha} \end{aligned} \tag{5.20}$$

Von m Atomen sind nach der Zeit T_h im Mittel $\dfrac{m}{2}$ zerfallen.

5.3 Erwartungswert und Varianz von Zufallsvariablen

Zufallsvariablen werden umfassend durch ihre Verteilungsfunktion $F_X(\cdot)$ beschrieben. Zur konzisen Beurteilung stochastischer Eigenschaften ist es oft hilfreich und zweckmässig, aussagekräftige Verteilungsparameter zu kennen. Zu den wichtigsten zählen zweifelsohne Mittelwert und Varianz.

5.3.1 Mittelwert einer Zufallsvariablen

X sei eine *diskrete* Zufallsvariable mit den Ausprägungen $x_i\,(i=1,\ldots,n)$ und den dazugehörigen Wahrscheinlichkeiten $f_X(x_i)$.

$$\mu = E(X) = \sum_{i=1}^{n} x_i \; f_X(x_i) \tag{5.21}$$

heisst Mittelwert der (diskreten) Zufallsvariablen X.

Bei unendlich vielen Ausprägungen von X existiert der Mittelwert nur, falls

$$\sum_{i=1}^{\infty} | x_i | \, f_X(x_i) < \infty \tag{5.22}$$

Beispiel:
X sei die Augenzahl eines symmetrischen Würfels.

$$\mu = E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3.5 \tag{5.23}$$

- Der Mittelwert einer Zufallsvariablen braucht nicht mit einer der möglichen Ausprägungen zusammenzufallen.

- Das empirische Analogon zu μ ist das arithmetische Mittel $\bar{x}$, welches sich als Summe der mit den relativen Häufigkeiten gewichteten Merkmalsausprägungen ergibt. Wir werden später begründen, weshalb sich arithmetische Mittel um den Erwartungswert μ verteilen.

Ist X eine *stetige* Zufallsvariable mit der Dichte $f_X(\cdot)$, so heisst

$$\mu = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx \tag{5.24}$$

Mittelwert von X.

Auch stetige Zufallsvariablen besitzen nur dann einen Mittelwert, wenn

$$\int_{-\infty}^{+\infty} |x|\, f_X(x)\, dx < \infty \tag{5.25}$$

<u>Beispiel:</u>
Sei X exponentialverteilt mit

$$f_X(x) = \begin{cases} \alpha\, e^{-\alpha\, x} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{5.26}$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \int_0^{\infty} x\, \alpha e^{-\alpha x}\, dx \\ &= \alpha \left\{ \frac{-x}{\alpha}\, e^{-\alpha x} \Big|_0^{\infty} + \frac{1}{\alpha} \int_0^{\infty} e^{-\alpha x}\, dx \right\} \\ &= -\, x e^{-\alpha x}\; e^{-\alpha x} \Big|_0^{\infty} - \frac{1}{\alpha}\, e^{-\alpha x} \Big|_0^{\infty} = -\, e^{-\alpha x} \left(x + \frac{1}{\alpha} \right) \Big|_0^{\infty} = \frac{1}{\alpha} \end{aligned} \tag{5.27}$$

5.3.2 Varianz einer Zufallsvariablen

Mittelwerte beschreiben im oben definierten Sinn die Lage oder das Zentrum der Realisationen von Zufallsvariablen. Die Varianz einer Zufallsvariablen ist ein Streuungsparameter und bringt die Abweichung ihrer Ausprägungen vom Mittelwert zum Ausdruck.

Def. 5.5

X sei eine diskrete Zufallsvariable mit der Wahrscheinlichkeitsfunktion $f_X(\cdot)$ und dem Mittelwert μ. Die Varianz $V(X) = \sigma^2$ ist dann

$$V(X) = \sigma^2 = \sum_i (x_i - \mu)^2 \; f_X(x_i) \tag{5.28}$$

Analog zum Mittelwert werden auch bei der Varianz die Wahrscheinlichkeiten als Gewichtungsmuster für die Abweichungsquadrate verwendet.

X sei eine stetige Zufallsvariable mit der Dichte $f_X(x)$ und dem Mittelwert μ. Die Varianz $V(X) = \sigma^2$ ist dann

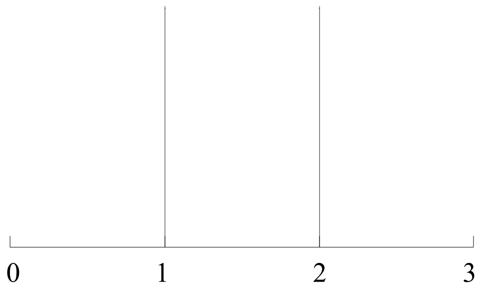
$$V(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 \; f_X(x)\, dx \tag{5.29}$$

Def. 5.6

$\sigma = \sqrt{V(X)}$ heisst Standardabweichung der Zufallsvariablen X.

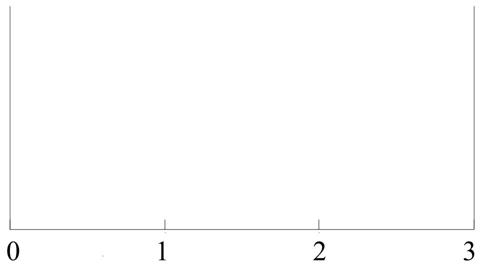
Die Varianz ist ein Mass für die "mittlere" Abweichung einer Zufallsvariablen von ihrem Mittelwert. Sie besitzt im Gegensatz zum Mittelwert keine anschauliche Interpretation. Das Verhalten der Varianz lässt sich am einfachsten durch einige Beispiele motivieren.

x	1	2
$f_X(x)$	$\frac{1}{2}$	$\frac{1}{2}$



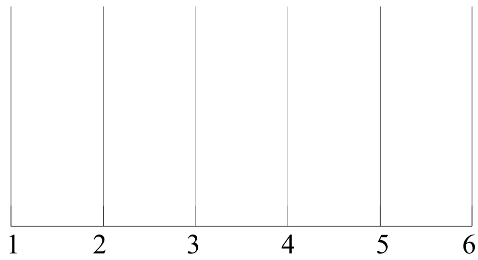
$\mu = 1.5$
$\sigma^2 = 0.25$

x	0	3
$f_X(x)$	$\frac{1}{2}$	$\frac{1}{2}$



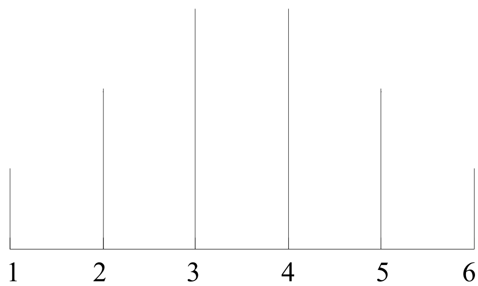
$\mu = 1.5$
$\sigma^2 = 2.25$

x	1	2	3	4	5	6
$f_X(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



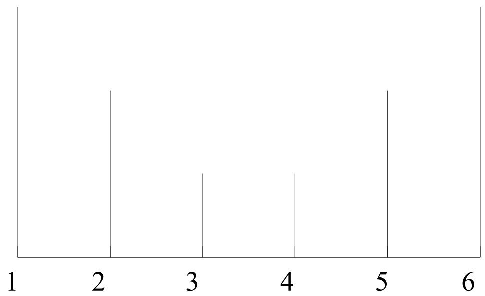
$\mu = 3.5$
$\sigma^2 = 2.92$

x	1	2	3	4	5	6
$f_X(x)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{3}{12}$	$\frac{3}{12}$	$\frac{1}{6}$	$\frac{1}{12}$



$\mu = 3.5$
$\sigma^2 = 1.92$

x	1	2	3	4	5	6
$f_X(x)$	$\frac{3}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{3}{12}$



$\mu = 3.5$
$\sigma^2 = 3.92$

Die Varianz besitzt nicht mehr dieselbe Masseinheit wie die Zufallsvariablen selber, die Standardabweichung hingegen schon.

Beispiel:
Sei X eine exponentialverteilte Zufallsvariable mit

$$
\begin{aligned}
f_X(x) &= \alpha e^{-\alpha x} \qquad x \geq 0 \\
\mu &= \frac{1}{\alpha} \\
\sigma^2 &= \int_0^\infty \left(x - \frac{1}{\alpha} \right)^2 \alpha e^{-\alpha x} \, dx \\
&= \alpha \int_0^\infty x^2 \, e^{-\alpha x} \, dx - 2 \frac{1}{\alpha} \int_0^\infty \alpha \, x e^{-\alpha x} \, dx + \left(\frac{1}{\alpha} \right)^2 \int_0^\infty \alpha e^{-\alpha x} \, dx \\
&= \alpha \int_0^\infty x^2 \, e^{-\alpha x} \, dx - \left(\frac{1}{\alpha} \right)^2 \\
&= \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \left(\frac{1}{\alpha} \right)^2
\end{aligned}
\tag{5.30}
$$

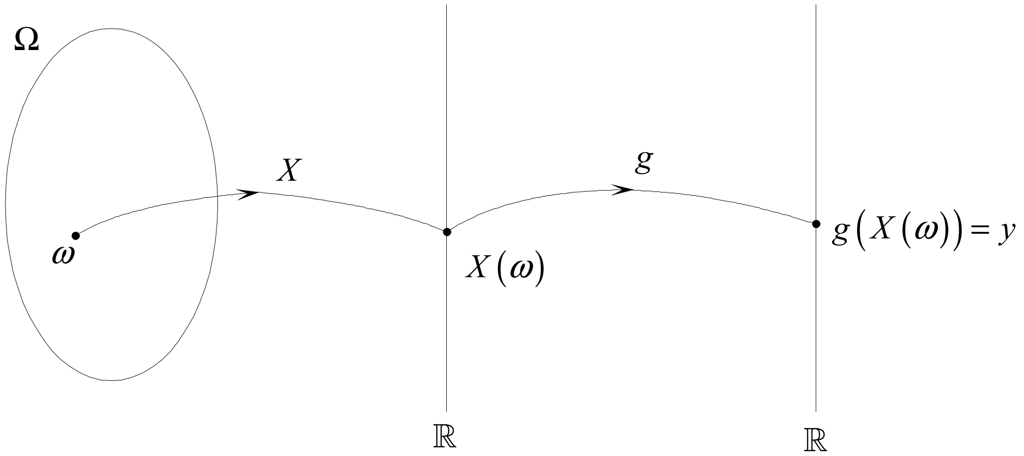
Beispiel:
Sei X eine gleichverteilte Zufallsvariable mit

$$
\begin{aligned}
f_X(x) &= \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\[2em] 0 & \text{sonst} \end{cases} \\
E(X) &= \int_a^b \frac{x}{b-a} \, dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{a+b}{2} \\
V(X) &= \int_a^b \left(x - \frac{a+b}{2} \right)^2 \frac{1}{b-a} \, dx = \frac{(b-a)^2}{12}
\end{aligned}
\tag{5.31}
$$

5.4 Funktionen von Zufallsvariablen

Zufallsvariablen bilden den Wahrscheinlichkeitsraum auf die reelle Achse ab. Wir bilden X ein zweites Mal auf $\forall$ ab und betrachten die zusammengesetzte Funktion

$$Y = g(X)$$



Wir interessieren uns für die Wahrscheinlichkeitsverteilung von Y, wobei jene von X als bekannt vorausgesetzt wird.

<u>Beispiel:</u>
X sei eine diskrete Zufallsvariable und $Y = 2X + 10$.

x	1	2	3
$f_X(x)$ $f_Y(y)$	0.2	0.3	0.5
y	12	14	16

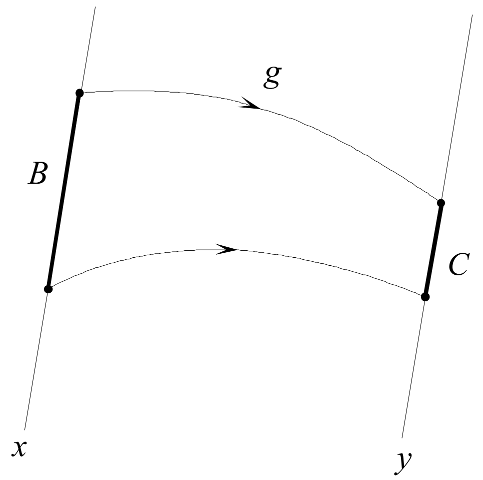
Durch die Abbildung g werden die Wahrscheinlichkeiten von X auf Y übertragen. Die Verteilung von Y lässt sich aus derjenigen von X ableiten. Bezeichnen F_X resp. F_Y die jeweiligen Verteilungsfunktionen, so gilt

$$F_Y(y) = P(Y \le y) = P(2X + 10 \le y)$$
$$= P\left(X \le \frac{y}{2} - 5 \right) = F_X\left(\frac{y}{2} - 5 \right) \tag{5.32}$$

Satz 5.1

Ist $Y = g(X)$ eine Funktion einer Zufallsvariablen X, dann ist auch Y eine Zufallsvariable, deren Verteilungsfunktion sich aus jener von X ableiten lässt.

Ist C ein beliebiges Ereignis in Zusammenhang mit Y und B jenes Ereignis in Zusammenhang mit X, das auf C abgebildet wird, so nennt man B und C *äquivalente Ereignisse*.



$$B = \{x \mid g(x) \in C\}$$

Beispiel:

$$Y = g(X) = X^2$$
$$C = \{y \mid 0 \le y \le 2\}, \quad B = \{x \mid -\sqrt{2} \le x \le \sqrt{2}\} \tag{5.33}$$

C und B sind äquivalente Ereignisse, denn $0 \le y \le 2$ genau dann, wenn

$$-\sqrt{2} \le x \le \sqrt{2} \tag{5.34}$$

Def. 5.7

g sei eine reellwertige Funktion einer Zufallsvariablen X mit bekannter Wahrscheinlichkeitsverteilung. Dann ist auch $Y = g(X)$ eine Zufallsvariable, deren Wahrscheinlichkeitsverteilung sich gemäss

$$P(C) = P\big(B \mid g(B) = C\big) \tag{5.35}$$

ergibt, wobei B und C äquivalente Ereignisse von X resp. Y sind.

Sei X eine diskrete Zufallsvariable mit der Wahrscheinlichkeitsfunktion $f_X(\cdot)$ und den Ausprägungen $x_1, ..., x_n$ sowie $Y = g(X)$. Dann gilt für die Wahrscheinlichkeitsfunktion $f_Y(\cdot)$ von Y

$$f_Y\left(y_j\right) = \sum_{\{i:g(x_i)=y_j\}} f_X(x_i) \tag{5.36}$$

Ist X eine stetige Zufallsvariable mit der Dichte $f_X(\cdot)$ und $Y = g(X)$, so findet man die Verteilungsfunktion $F_Y(\cdot)$ von Y, indem man $f_X(\cdot)$ über den jeweils äquivalenten Bereich integriert.

$$F_Y(y) = P(Y \leq y) = P\big(g(X) \leq y\big) = \int_{\{x:g(x)\leq y\}} f_X(x)\, dx \tag{5.37}$$

<u>Beispiel:</u>

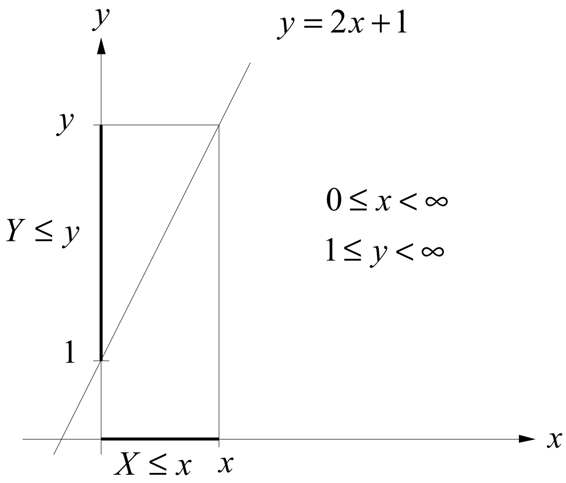
x	0	1	2	3
$f_X(x)$	0.1	0.3	0.2	0.4

$Y = (X - 2)^2$

x	0	1	2	3
$f_X(x)$	0.1	0.3	0.2	0.4
y	4	1	0	1

y	0	1	4
$f_Y(y)$	0.2	0.7	0.1

Beispiel:
Sei X exponentialverteilt mit $\alpha = 1$
$f_X(x) = e^{-x} \qquad x \geq 0$
$Y = 2X + 1$



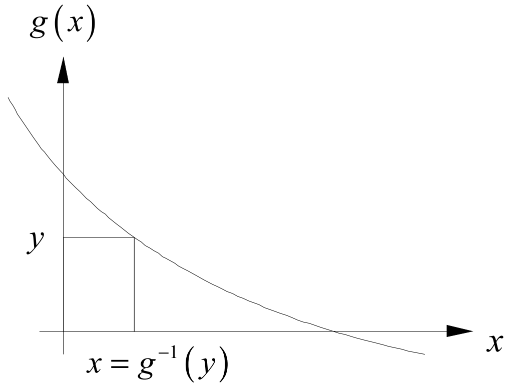
$$P(Y \leq 3) = P(2X + 1 \leq 3) = P(X \leq 1)$$
$$= \int_0^1 e^{-x}\, dx = -e^{-x} \Big|_0^1 = 1 - e^{-1} = 0.6321 \tag{5.38}$$

Für stetige Zufallsvariablen X mit der Dichte $f_X(\cdot)$ und streng monotonen Abbildungen g existiert eine direkte Methode zur Generierung der Dichte $f_Y(\cdot)$ von $Y = g(X)$.

Satz 5.2

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \left|\frac{dg^{-1}(y)}{dy}\right| \tag{5.39}$$

Beweis für streng monoton fallende Funktionen g:



$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = P\big(g(X) \leq y\big) \\
&= P\big(X \geq g^{-1}(y)\big) \qquad \begin{pmatrix} \text{da } g \text{ streng monoton fallend} \\ \text{und stetig} \end{pmatrix} \\
&= 1 - P\big(X \leq g^{-1}(y)\big) \\
&= 1 - F_X\big(g^{-1}(y)\big)
\end{aligned}
\tag{5.40}
$$

$$
\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{d\big(1 - F_X\big(g^{-1}(y)\big)\big)}{dg^{-1}(y)} \frac{dg^{-1}(y)}{dy} \\
&= -f_X\big(g^{-1}(y)\big) \frac{dg^{-1}(y)}{dy} \\
&= f_X\big(g^{-1}(y)\big) \left| \frac{dg^{-1}(y)}{dy} \right|
\end{aligned}
\tag{5.41}
$$

da $\dfrac{dg^{-1}(y)}{dy} < 0$ $\quad$ ($g(X)$ streng monoton fallend)

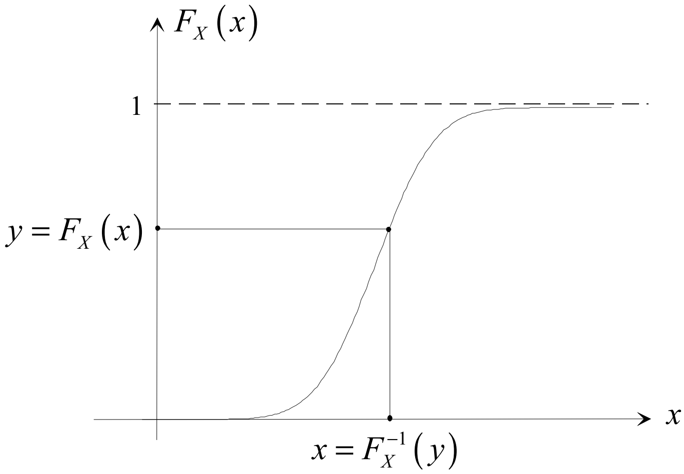
Beispiel:
X sei exponentialverteilt mit

$$f_X(x) = \alpha\, e^{-\alpha x} \qquad \alpha > 0 \quad x \geq 0$$
$$Y = g\left(X\right) = 2X + 1, \quad g^{-1}\left(Y\right) = \frac{Y-1}{2} \tag{5.42}$$

$$\frac{dg^{-1}(y)}{dy} = \frac{d}{dy}\left(\frac{y-1}{2}\right) = \frac{1}{2} \tag{5.43}$$

$$f_Y(y) = \alpha\, e^{-\alpha\frac{y-1}{2}} \cdot \frac{1}{2} = \frac{\alpha}{2}\, e^{-\frac{\alpha}{2}(y-1)} \qquad \alpha > 0 \quad y \geq 1 \tag{5.44}$$

Beispiel:
X sei eine beliebig stetig verteilte Zufallsvariable mit der Verteilungsfunktion $F_X(\cdot)$.
Dann ist die Zufallsvariable $Y = F_X\left(X\right)$ über dem Intervall [0,1] stetig gleichverteilt.



$$Y = F_X(X)$$
$$f_Y(y) = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

$$Y = F_X(X) \quad resp. \quad X = F_X^{-1}\left(Y\right) = g^{-1}\left(Y\right) \tag{5.45}$$

$$\frac{dg^{-1}\left(y\right)}{dy} = \frac{dF_X^{-1}\left(y\right)}{dy} = \frac{1}{F_X'\left(x\right)} = \frac{1}{f_X(x)} \tag{5.46}$$

mit $x = F_X^{-1}(y)$ und Satz 5.2 gilt

$$f_Y(y) = f_X(x)\,\frac{1}{f_X(x)} = 1 \tag{5.47}$$

Beispiel:

X sei exponentialverteilt mit

$$\begin{aligned}
&f_X(x) = \alpha\, e^{-\alpha x} \qquad \alpha > 0 \quad x \geq 0\\
&F_X(x) = 1 - e^{-\alpha x}\\
&\quad Y = F_X(X) = 1 - e^{-\alpha X} \qquad 0 \leq Y \leq 1\\
&\quad X = F_X^{-1}(Y) = -\frac{1}{\alpha}\ln(1-Y)
\end{aligned} \tag{5.48}$$

$$\begin{aligned}
\frac{dF_X^{-1}(y)}{dy} &= \frac{1}{\alpha(1-y)}\\
f_Y(y) &= \alpha\, e^{-\alpha\left(\frac{-1}{\alpha}\ln(1-y)\right)}\,\frac{1}{\alpha(1-y)}\\
&= e^{\ln(1-y)}\,\frac{1}{1-y}\\
&= 1 \qquad\qquad 0 \leq y \leq 1
\end{aligned} \tag{5.49}$$

5.5 Mathematische Erwartung - Erwartungswert von Funktionen zufälliger Variablen

Im vorhergehenden Abschnitt wurde gezeigt, dass Funktionen von Zufallsvariablen wieder Zufallsvariablen mit einem dazugehörigen Verteilungsgesetz sind. $Y = g(X)$ ist eine Zufallsvariable, welche auf derselben Ereignismenge wie X definiert ist und dem Elementarausgang ω das Bild $y = g(X(\omega))$ zuordnet.

Für den Erwartungswert von Y gilt definitionsgemäss

$$\mu_Y = E(Y) = \sum_i y_i\ f_Y(y_i) \quad \text{resp.} \quad \int_{-\infty}^{+\infty} y\ f_Y(y)\ dy \tag{5.50}$$

wobei allerdings vorausgesetzt wird, dass $f_Y(\cdot)$ bekannt ist. Ist man lediglich am Erwartungswert von $Y = g(X)$ interessiert, so kann man den Zwischenschritt umgehen und die Berechnung auf der Basis von $f_X(\cdot)$ vollziehen.

Satz 5.3

Sei X eine Zufallsvariable mit bekannter Wahrscheinlichkeitsverteilung $f_X(\cdot)$ und $Y = g(X)$. Dann ist der Erwartungswert von $Y = g(X)$

$$\begin{aligned} E(Y) = E\big(g(X)\big) &= \sum_{i=1}^{\infty} g(x_i)\ f_X(x_i) \\ &= \sum_x g(x)\ f_X(x) \end{aligned} \tag{5.51}$$

für diskrete Zufallsvariablen X und

$$E(Y) = E\big(g(X)\big) = \int_{-\infty}^{+\infty} g(x)\ f_X(x)\ dx \tag{5.52}$$

für stetige Zufallsvariablen X.

Für diskrete Zufallsvariablen ist der Beweis des obigen Satzes sehr einfach. Er wird am nachfolgenden Beispiel illustriert. Für stetige Zufallsvariablen ist der Beweis anspruchsvoller.

Beispiel:
Die Verteilung von X sei

X	-2	-1	0	1	2
f_X	0.1	0.15	0.2	0.25	0.3

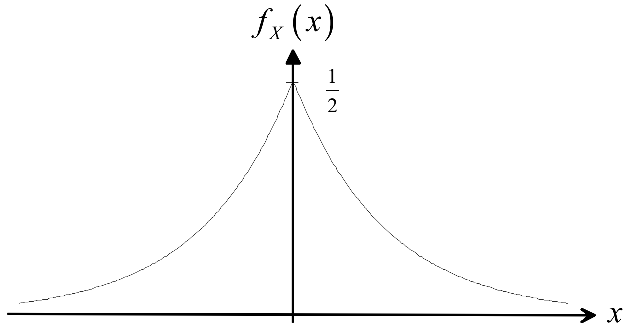
und $y = g(X) = X^2$

Y	0	1	4
f_Y	0.2	0.4	0.4

$$
\begin{aligned}
E(Y) &= 0\cdot 0.2 + 1\cdot 0.4 + 4\cdot 0.4 = 2 \\
E(g(X)) &= (-2)^2\cdot 0.1 + (-1)^2\cdot 0.15 + 0^2\cdot 0.2 + 1^2\cdot 0.25 + 2^2\cdot 0.3 \\
&= 0^2\cdot 0.2 + 1^2\cdot(0.15+0.25) + 2^2\cdot(0.1+0.3) = 2
\end{aligned}
\tag{5.53}
$$

<u>Beispiel:</u>

$$
f_X(x) = \begin{cases} \dfrac{1}{2}e^{-x} & x \ge 0 \\[2ex] \dfrac{1}{2}e^{x} & x < 0 \end{cases}
$$



$$
\begin{aligned}
Y &= |X| \\
E(Y) = E(|X|) &= \int_{-\infty}^{+\infty} |x|\, f_X(x)\, dx \\
&= \frac{1}{2}\int_{-\infty}^{0}(-x)e^{x}\, dx + \frac{1}{2}\int_{0}^{\infty} xe^{-x} dx = \frac{1}{2}(1+1) = 1
\end{aligned}
\tag{5.54}
$$

Beachte:

Falls $g(X) = X$ gilt $E(g(X)) = E(X) = \mu$

Falls $g(X) = (X-\mu)^2$ gilt $E(g(X)) = E\left\{(X-\mu)^2\right\} = \sigma^2$

5.5.1 Eigenschaften mathematischer Erwartungen

1.) $E(c) = c,$ für c konstant

2.) $E\left(cg(X)\right) = cE\left(g(X)\right),$ für c konstant

3.) $E\left\{c_1 g_1(X) + c_2 g_2(X)\right\} = c_1 E\left(g_1(X)\right) + c_2 E\left(g_2(X)\right),$ für c_1, c_2 konstant

<u>Beweis</u>

1.)

$$g(X) = c$$

$$E\left(g(X)\right) = E(c) = \int_{-\infty}^{+\infty} c\, f_X(x)\, dx = c \int_{-\infty}^{+\infty} f_X(x)\, dx = c \cdot 1 = c \tag{5.55}$$

2.)

$$E\left(cg(X)\right) = \int_{-\infty}^{+\infty} cg(x)\, f_X(x)\, dx = c\int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx = cE\left(g(X)\right) \tag{5.56}$$

3.)

$$\begin{aligned} E\left(c_1 g_1(X) + c_2 g_2(X)\right) &= \int_{-\infty}^{+\infty} \left(c_1 g_1(x) + c_2 g_2(x)\right) f_X(x)\, dx \\ &= \int_{-\infty}^{+\infty} c_1 g_1(x)\, f_X(x)\, dx + \int_{-\infty}^{+\infty} c_2 g_2(x)\, f_X(x)\, dx \\ &= c_1 E\left(g_1(X)\right) + c_2 E\left(g_2(X)\right) \end{aligned} \tag{5.57}$$

Satz 5.4

Für die Varianz einer Zufallsvariablen X mit dem Mittelwert $\mu = E(X)$ gilt

$$\sigma^2 = V(X) = E\left(X^2\right) - \left[E(X)\right]^2 \tag{5.58}$$

Beweis

$$V(X) = E\left\{(X-\mu)^2\right\} = E\left\{\left(X^2 - 2\mu X + \mu^2\right)\right\}$$
$$= E\left(X^2\right) - 2\mu E(X) + \mu^2 = E\left(X^2\right) - \mu^2 \tag{5.59}$$

(Wir haben vorausgesetzt, dass $E\left(X^2\right)$ existiert!)

Beispiel:
Man bestimme den Mittelwert und die Varianz der Zufallsvariablen

$$Z = 3X + 2$$

wobei X die Wahrscheinlichkeitsverteilung

x	1	2	5
$P(X=x)$	0.2	0.3	0.5

besitzt.

$$E(Z) = E(3X+2) = 3E(X) + 2$$
$$= 3(1\cdot 0.2 + 2\cdot 0.3 + 5\cdot 0.5) + 2 = 11.9$$

(5.60)

$$V(Z) = V(3X+2) = E\left[\left(3X+2-3E(X)-2\right)^2\right] = E\left[9\left(X-E(X)\right)^2\right] = 9V(X)$$
$$= 9\left[E\left(X^2\right) - 3.3^2\right] = 9(13.9 - 10.89) = 27.09$$

Beachte:

$$V(c) = 0$$
$$V(X+c) = V(X) \quad c = \text{konstant}, \qquad \text{Translationsinvarianz} \tag{5.61}$$

5.5.2 Lineare Transformation, standardisierte Zufallsvariablen

Wir setzen für $g(X)$ speziell

$$\begin{aligned}&Y = g(X) = c_1 X + c_2 \qquad \text{lineare Transformation}\\ &c_1\,,c_2 \quad \text{konstant}\end{aligned} \tag{5.62}$$

und erhalten

$$\begin{aligned}E(Y) &= E\left(c_1 X + c_2\right) = c_1 E(X) + c_2 = c_1 \mu + c_2\\ V(Y) &= V\left(c_1 X + c_2\right) = c_1^2\, V(X) = c_1^2\, \sigma^2\end{aligned} \tag{5.63}$$

Spezielle Bedeutung besitzt die lineare Transformation, die einer Zufallsvariablen X die zugehörige *standardisierte Variable* Z zuordnet.

Gegeben: Zufallsvariable X mit $E(X) = \mu,\ V(X) = \sigma^2$

Def. 5.8

$Z = \dfrac{X - \mu}{\sigma}$ heisst die zu X gehörige standardisierte Variable.

Satz 5.5

$E(Z) = 0$
$V(Z) = 1$

<u>Beweis</u>

In der Terminologie der allgemeinen linearen Transformation setzen wir

$$c_1 = \frac{1}{\sigma} \qquad c_2 = \frac{-\mu}{\sigma}$$

$$\begin{aligned}E(Z) &= c_1\, E(X) + c_2 = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0\\ V(Z) &= c_1^2\, V(X) = \frac{1}{\sigma^2}\cdot\sigma^2 = 1\end{aligned} \tag{5.64}$$

5.5.3 Ungleichung von Tschebyscheff

Die Ungleichung von Tschebyscheff ermöglicht Wahrscheinlichkeitsaussagen für Abweichungen beliebig verteilter Zufallsvariablen von ihrem Erwartungswert.

Bezeichnet X eine Zufallsvariable mit dem Verteilungsgesetz $f_X(\cdot)$ und $c > 0$ eine beliebige reelle Konstante, so gilt für jedes $\varepsilon > 0$ die Ungleichung von Tschebyscheff

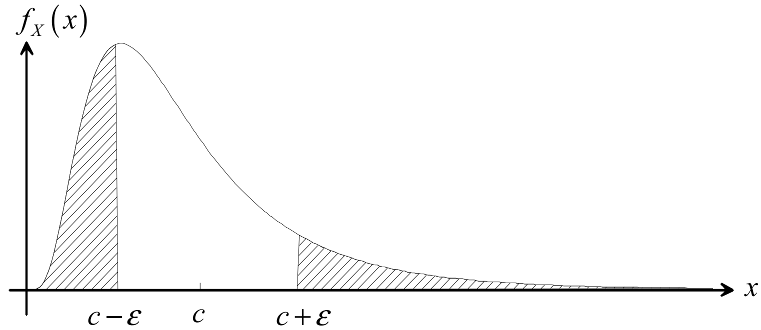
$$P\left(|X-c| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\, E\left[\left(X-c\right)^2\right] \tag{5.65}$$

Es wird vorausgesetzt, dass $E\left((X-c)^2\right)$ existiert und endlich bleibt.

Der Beweis wird für eine stetige Zufallsvariable mit der Dichte f_X geführt.

$$P\left(|X-c| \geq \varepsilon\right) = \int\limits_{|x-c|\geq\varepsilon} f_X(x)\,dx$$

$$= \int\limits_{-\infty}^{c-\varepsilon} f_X(x)\,dx + \int\limits_{c+\varepsilon}^{\infty} f_X(x)\,dx$$



Im gesamten Integrationsbereich gilt

$$|x-c| \geq \varepsilon \quad resp. \quad \frac{(x-c)^2}{\varepsilon^2} \geq 1 \tag{5.66}$$

woraus die Abschätzung folgt

$$P\left(|X-c|\geq\varepsilon\right)=\int\limits_{|x-c|\geq\varepsilon} f_X(x)\,dx \leq \int\limits_{|x-c|\geq\varepsilon} \frac{(x-c)^2}{\varepsilon^2} f_X(x)\,dx$$
$$\leq \frac{1}{\varepsilon^2}\int\limits_{-\infty}^{+\infty}(x-c)^2 f_X(x)\,dx \tag{5.67}$$
$$=\frac{1}{\varepsilon^2} E\left((X-c)^2\right)$$

Wählt man speziell $c=\mu$

$$P\left(|X-\mu|\geq\varepsilon\right)\leq\frac{V(X)}{\varepsilon^2}=\frac{\sigma^2}{\varepsilon^2} \tag{5.68}$$

Setzt man ferner $\varepsilon=t\sigma$

$$P\left(|X-\mu|\geq t\sigma\right)\leq\frac{1}{t^2} \tag{5.69}$$

Die Wahrscheinlichkeit, dass eine beliebig verteilte Zufallsvariable (mit endlicher Varianz) Werte annimmt, welche mehr als zwei Standardabweichungen von ihrem Erwartungswert abweichen, beträgt höchstens 25%.

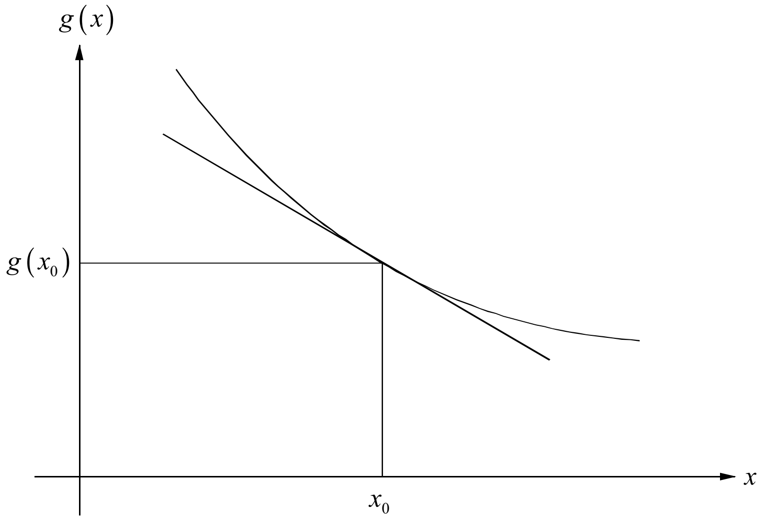
Ist X speziell normalverteilt $\left(X\sim N(\mu,\sigma)\right)$, so gilt

$$P\left(|X-\mu|\geq 2\sigma\right)=0.0465<0.25 \tag{5.70}$$

Die Aussagen aus der Tschebyscheff'schen Ungleichung sind in diesem Spezialfall bedeutend unschärfer; sie gelten hingegen für beliebige Verteilungen.

5.5.4 Jensen'sche Ungleichung

Es sei $g(\cdot)$ eine über dem gesamten Definitionsbereich konvexe Funktion.



Die Konvexität von g garantiert die Existenz einer Geraden durch den Punkt $\left(x_0, g\left(x_0\right)\right)$, welche strikt unterhalb des Graphen von $g(x)$ liegt.

Ungleichung von Jensen

Sei X eine Zufallsvariable mit dem Erwartungswert $E(X)$ und g eine konvexe Funktion. Dann liefert die Ungleichung von Jensen eine Abschätzung nach unten für den Erwartungswert $E\left(g\left(X\right)\right)$:

$$E\left(g\left(X\right)\right) \geq g\left(E\left(X\right)\right) \tag{5.71}$$

Beweis:

Sei die Funktion g gegeben. Dann existiert eine Gerade

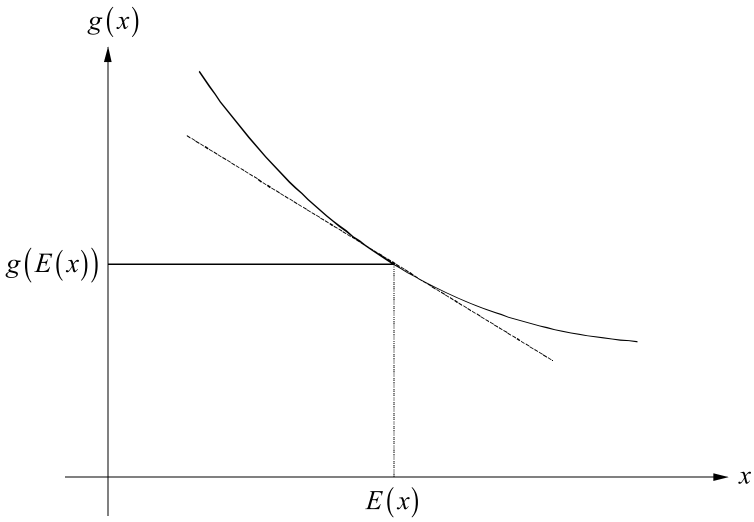
$$h(x) = c + mx \tag{5.72}$$

für die aufgrund der Konvexität von g gilt

$$h(x) = c + mx \leq g(x) \tag{5.73}$$

Wählt man die Parameter c und m so, dass h den Graphen von g im Punkt $\left(E\left(X\right), g\left(E\left(X\right)\right)\right)$ berührt, so gilt ferner

$$h\big(E(X)\big) = g\big(E(X)\big)$$



Der Erwartungswertoperator ist linear, woraus folgt

$$E\big(h(X)\big) = E(c + mX) = c + mE(X) = h\big(E(X)\big) \tag{5.74}$$

Somit folgt schliesslich

$$g\big(E(X)\big) = h\big(E(X)\big) = E\big(h(X)\big) \leq E\big(g(X)\big) \tag{5.75}$$

weil $h(X) \leq g(X)$.

<u>Beispiel:</u>

$$\begin{aligned} X &\sim \exp\left(\frac{1}{5}\right) \\ E(X) &= 5 \\ V(X) &= 25 \\ g(X) &= X^2 \qquad \text{konvex} \end{aligned} \tag{5.76}$$

$$\begin{aligned} E\big(g(X)\big) &= E\left(X^2\right) = V(X) + E(X)^2 = 25 + 25 \\ &\geq 25 = E(X)^2 = g\big(E(X)\big) \end{aligned} \tag{5.77}$$

5.5.5 Momente von Zufallsvariablen

Momente von Zufallsvariablen sind Masse für Abweichungen der Zufallsvariablen von einer beliebigen Konstante *a*.

Wir betrachten die Funktion $Y = (X - a)^k$ (a = konstant, $k = 0,1,2, \ldots$). Erwartungswerte von *Y* nennt man *Momente* von *X* bezüglich der Konstanten *a*. Ist $a = 0$, so spricht man von Momenten schlechthin, für $a = \mu$ spricht man von sog. *zentralen* Momenten.

Def. 5.9

Das *k*-te Moment μ_k' einer Zufallsvariablen *X* ist definiert als $\mu_k' = E(X^k)$.

Das *k*-te zentrale Moment μ_k einer Zufallsvariablen *X* ist definiert als

$$\mu_k = E\left\{(X - \mu)^k\right\} \qquad (5.78)$$

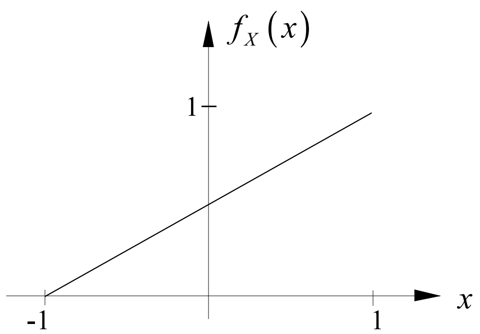
Beachte:

- Das erste Moment μ_1' ist der Mittelwert μ von *X*, $\mu_1' = \mu$
- $\mu_1 = E(X - \mu) = 0$
- $\mu_2 = E(X - \mu)^2 = \sigma^2$
- Für symmetrische Verteilungen verschwinden alle ungeraden zentralen Momente.

Beispiel:

X sei eine stetige Zufallsvariable mit der Dichte

$$f_X(x) = \begin{cases} 0 & x < -1 \\ \dfrac{1}{2}(x+1) & -1 \le x \le 1 \\ 0 & x > 1 \end{cases}$$



$$\begin{aligned}\mu_3 &= E(X-\mu)^3 = E\left(X^3 - 3\mu X^2 + 3\mu^2 X - \mu^3\right)\\ &= E\left(X^3\right) - 3E(X)E\left(X^2\right) + 3\left[E(X)\right]^2 E(X) - \left[E(X)\right]^3\\ &= E\left(X^3\right) - 3E(X)E\left(X^2\right) + 2\left[E(X)\right]^3\end{aligned} \tag{5.79}$$

Das dritte zentrale Moment wird auch als *Mass für die Asymmetrie einer Verteilung* verwendet.

$$\begin{aligned}E\left(X^3\right) &= \frac{1}{2}\int_{-1}^{1} x^3(x+1)\,dx = \frac{1}{5}\\ E\left(X^2\right) &= \frac{1}{2}\int_{-1}^{1} x^2(x+1)\,dx = \frac{1}{3}\\ E(X) &= \frac{1}{2}\int_{-1}^{1} x(x+1)\,dx = \frac{1}{3}\\ \mu_3 &= \frac{1}{5} - 3\cdot\frac{1}{3}\cdot\frac{1}{3} + 3\cdot\frac{1}{9}\cdot\frac{1}{3} - \frac{1}{27} = -\frac{8}{135} = -0.059\end{aligned} \tag{5.80}$$

Def. 5.10

$$\gamma = \frac{1}{\sigma^3} E\left\{(X-\mu)^3\right\} = \frac{\mu_3}{\sigma^3} \tag{5.81}$$

heisst Schiefe der Verteilung von *X*.

Für das obige Beispiel erhält man mit $V(X) = \dfrac{2}{9}$

$$\gamma = \frac{1}{\sigma^3} E(X-\mu)^3 = \frac{-\dfrac{8}{135}}{\left(\dfrac{\sqrt{2}}{3}\right)^3} = -0.57 \tag{5.82}$$

5.5.6 Die momenterzeugende Funktion

Von besonderer Bedeutung zur Charakterisierung und zur Berechnung der Momente einer Zufallsvariablen ist die sog. momenterzeugende Funktion.

Def. 5.11

X sei eine Zufallsvariable mit der Wahrscheinlichkeitsfunktion $f_X(x)$. Die momenterzeugende Funktion $m_X(t)$ von $f_X(x)$ ist der Erwartungswert von e^{tX}

$$\begin{aligned} m_X(t) &= E\left(e^{tX}\right) = \sum_{i=1}^{\infty} e^{tx_i} f_X(x_i) \qquad (X \text{ diskret}) \\ m_X(t) &= E\left(e^{tX}\right) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\,dx \qquad (X \text{ stetig}) \end{aligned} \tag{5.83}$$

Beispiel:

Für die exponentialverteilte Zufallsvariable X mit der Dichte

$$f_X(x) = \alpha e^{-\alpha x} \qquad x \geq 0 \tag{5.84}$$

erhält man die momenterzeugende Funktion

$$\begin{aligned} m_X(t) &= E\left(e^{tX}\right) = \int_0^{+\infty} e^{tx}\,\alpha e^{-\alpha x}\,dx = \alpha \int_0^{\infty} e^{-x(\alpha-t)}dx \\ &= \frac{-\alpha}{\alpha-t} e^{-x(\alpha-t)} \Big|_0^{\infty} = \frac{\alpha}{\alpha-t} \qquad t < \alpha \end{aligned} \tag{5.85}$$

Existiert die momenterzeugende Funktion, dann ist $m_X(t)$ differenzierbar in t. Für den diskreten Fall erhält man

$$
\begin{aligned}
m_X(t) &= \sum_i e^{tx_i}\, f_X(x_i) \\
\frac{dm_X(t)}{dt} &= m'_X(t) = \sum_i x_i\, e^{tx_i}\, f_X(x_i) \\
\frac{dm'_X(t)}{dt} &= m''_X(t) = \sum_i x_i^2\, e^{tx_i}\, f_X(x_i) \\
m_X^{(k)}(t) &= \sum_i x_i^k\, e^{tx_i}\, f_X(x_i)
\end{aligned}
\tag{5.86}
$$

woraus folgt

$$
m_X^{(k)}(0) = \sum_i x_i^k\, f_X(x_i) = E\left(X^k\right) \tag{5.87}
$$

d.h. die k-te Ableitung der momenterzeugenden Funktion an der Stelle $t = 0$ ergibt gerade den Erwartungswert von X^k.

<u>Beispiel:</u>
Für die exponentialverteilte Zufallsvariable X gilt

$$
\begin{aligned}
m_X(t) &= \frac{\alpha}{\alpha - t} \\
m'_X(t) &= \frac{\alpha}{(\alpha - t)^2} \qquad m'_X(0) = \frac{1}{\alpha} = E(X) \\
m''_X(t) &= \frac{2\alpha}{(\alpha - t)^3} \qquad m''_X(0) = \frac{2}{\alpha^2} = E\left(X^2\right)
\end{aligned}
\tag{5.88}
$$

und man erhält Mittelwert und Varianz

$$
\begin{aligned}
\mu &= E(X) = \frac{1}{\alpha} \\
\sigma^2 &= E\left(X^2\right) - \left(E(X)\right)^2 = \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}
\end{aligned}
\tag{5.89}
$$

Satz 5.6 (ohne Beweis)

X und Y seien zwei Zufallsvariablen mit den Wahrscheinlichkeits- resp. Dichtefunktionen $f_X(x)$ und $f_Y(y)$ und existierenden momenterzeugenden Funktionen $m_X(t)$ und $m_Y(t)$. X und Y besitzen genau dann identische Verteilungsfunktionen, wenn

$$m_X(t) = m_Y(t) \qquad \text{für } |t| < h \tag{5.90}$$

Momenterzeugende Funktionen dienen nicht ausschliesslich der Berechnung von Momenten. Dank ihrer eineindeutigen Verknüpfung mit dem Verteilungsgesetz einer Zufallsvariablen entfalten sie ein weit grösseres Anwendungsfeld. Einer bestimmten Verteilung ist eine bestimmte momenterzeugende Funktion zugeordnet, und umgekehrt lässt sich aus der Kenntnis der momenterzeugenden Funktion eindeutig auf das Verteilungsgesetz schliessen.

Beispiel:

Ein Ziel wird beschossen. Die Trefferwahrscheinlichkeit sei p. Die Zufallsvariable X sei die Anzahl Versuche, die dem ersten Treffer vorangehen. Erfolgen die einzelnen Versuche unabhängig voneinander, so besitzt X die Wahrscheinlichkeitsfunktion

$$f_X(x) = (1-p)^x\, p = q^x p \qquad x = 0,1,2,\ldots \tag{5.91}$$

In diesem Fall heisst X *geometrisch* verteilt.

Wir berechnen mit Hilfe der momenterzeugenden Funktion Mittelwert und Varianz.

$$\begin{aligned} m_X(t) &= E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx}\, q^x p \\ &= p \sum_{x=0}^{\infty} \left(e^t q\right)^x = \frac{p}{1-qe^t} \qquad \text{falls } qe^t < 1, \text{ also } t < -\ln q \end{aligned} \tag{5.92}$$

$$
\begin{aligned}
m_X'(t) &= pqe^t\left(1-qe^t\right)^{-2}\\
m_X''(t) &= pqe^t\left[\left(1-qe^t\right)^{-2}+2qe^t\left(1-qe^t\right)^{-3}\right]\\
m_X'(0) &= \frac{pq}{p^2}=\frac{q}{p}\\
m_X''(0) &= pq\left(p^{-2}+2qp^{-3}\right)=\frac{q}{p^2}(1+q)\\
\mu &= E(X)=m_X'(0)=\frac{q}{p}\\
\sigma^2 &= V(X)=m_X''(0)-\left(m_X'(0)\right)^2=\frac{q}{p^2}(1+q)-\frac{q^2}{p^2}=\frac{q}{p^2}
\end{aligned}
\tag{5.93}
$$

Ist X eine Zufallsvariable mit der momenterzeugenden Funktion $m_X(t)$ und $Y = aX + b$ eine lineare Funktion von X, so gilt für die momenterzeugende Funktion von Y

$$m_Y(t) = e^{tb}\, m_X(at) \tag{5.94}$$

Denn:

$$
\begin{aligned}
m_Y(t) &= m_{aX+b}(t) = E\left(e^{t(aX+b)}\right)\\
&= e^{tb}\, E\left(e^{taX}\right) = e^{tb} m_X(at)
\end{aligned}
\tag{5.96}
$$

Die momenterzeugende Funktion ist somit das e^{tb}-fache jener von X, allerdings nicht an der Stelle t sondern an der Stelle at.

<u>Beispiel:</u>
Setzt man speziell $a = 1$ und $b = -\mu$ so gilt $Y = X - \mu$.

$$
\begin{aligned}
m_Y(t) &= m_{X-\mu}(t) = e^{-t\mu} m_X(t)\\
\frac{dm_Y(t)}{dt} &= m_Y'(t) = -\mu\, e^{-t\mu}\, m_X(t) + e^{-t\mu}\, m_X'(t)\\
&= e^{-t\mu}\left(m_X'(t) - \mu\, m_X(t)\right)
\end{aligned}
\tag{5.97}
$$

5.5.7 Approximation von Momenten

Ist das Verteilungsgesetz $f_Y(\cdot)$ einer Funktion $Y = g(X)$ der Zufallsvariablen X bekannt, so lassen sich die Momente von Y exakt bestimmen, wobei vorausgesetzt wird, dass es sich bei $g(\cdot)$ um eine beliebig oft stetig differenzierbare reellwertige Funktion handelt. Bei unbekannter Verteilung von Y erlaubt der Satz von Taylor die Berechnung approximativer Momente.

Entwickelt man $g(X)$ um den Erwartungswert μ_X von X, so gilt

$$Y = g(X) = \sum_{i=0}^{\infty} \frac{(X-\mu_X)^i}{i!} \frac{d^{(i)}g(X)}{dX^i} \tag{5.98}$$

wobei die Ableitungen an der Stelle $X = \mu_X$ zu verwenden sind.

Bricht man die Entwicklung nach dem quadratischen Glied ab, so gilt

$$Y = g(X) \approx g(\mu_X) + (X-\mu_X)\, g'(\mu_X) + \frac{(X-\mu_X)^2}{2}\, g''(\mu_X) \tag{5.99}$$

woraus nach einigen Umformungen die Approximationen für den Mittelwert und die Varianz folgen.

$$E(Y) = \mu_Y \approx g(\mu_X) + \frac{\sigma_X^2}{2}\, g''(\mu_X) \tag{5.100}$$

$$\begin{aligned} V(Y) = \sigma_Y^2 &\approx (\sigma_X\, g'(\mu_X))^2 + \left(E(X-\mu_X)^4 - \sigma_X^4\right)\left(\frac{g''(\mu_X)}{2}\right)^2 \\ &+ E(X-\mu_X)^3\, g'(\mu_X)\, g''(\mu_X) \end{aligned} \tag{5.101}$$

Bricht man die Taylorentwicklung bereits nach dem linearen Glied ab, so verschwinden in den Approximationen einfach die zweiten Terme.

Beispiel:

$$Y = g(X) = \ln X$$

$$E(Y) = \mu_Y \approx \ln \mu_X - \frac{\sigma_X^2}{2\mu_X^2} \tag{5.102}$$

$$\begin{aligned} V(Y) &\approx \left(\frac{\sigma_X}{\mu_X}\right)^2 + \left(\mu_{X_4} - \sigma_X^4\right)\frac{1}{4\mu_X^4} + \mu_{X_3}\frac{1}{\mu_X}\left(-\frac{1}{\mu_X^2}\right) \\ &= \left(\frac{\sigma_X}{\mu_X}\right)^2 + \frac{\mu_{X_4} - \sigma_X^4}{4\mu_X^4} - \frac{\mu_{X_3}}{\mu_X^3} \end{aligned} \tag{5.103}$$

wobei μ_{X_i} das i-te zentrale Moment von X bezeichnet.

5.5.8 Median und Quantile von Zufallsvariablen

Def. 5.12

$x_p \ (0 \le p \le 1)$ heisst p-Quantil einer Zufallsvariablen X, falls

$$P(X < x_p) \le p \quad \text{und} \quad P(X > x_p) \le 1 - p \tag{5.104}$$

Gibt es mehrere Werte x_p, die die obigen Bedingungen erfüllen, so wählen wir für x_p das arithmetische Mittel des grössten und kleinsten Wertes.

Def. 5.13

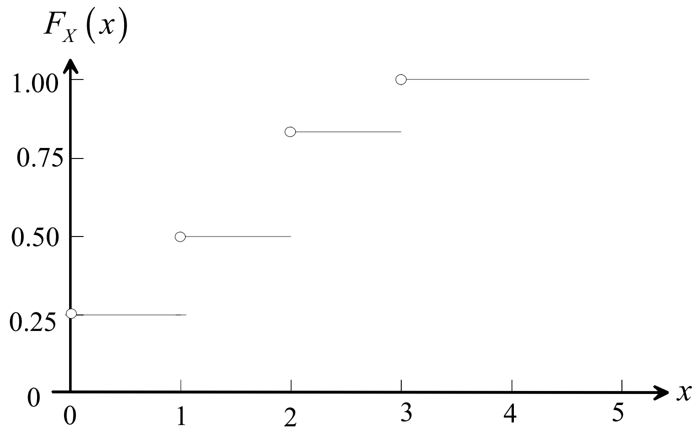
Das 0.5-Quantil einer Zufallsvariablen X heisst Median von X

$$x_{0.5} = Me \tag{5.105}$$

Die Definition der Quantile wird am besten durch ein Beispiel veranschaulicht.

Beispiel:
Die Wahrscheinlichkeitsfunktion einer Zufallsvariablen X sei gegeben durch

$$
\begin{aligned}
P(X=0) &= \frac{1}{4} \\
P(X=1) &= \frac{1}{4} \\
P(X=2) &= \frac{1}{3} \\
P(X=3) &= \frac{1}{6}
\end{aligned}
\tag{5.106}
$$



$x_{0.75}$: oberes Quartil

$x_{0.75} = 2$, denn

$$
\begin{aligned}
P(X<2) &= 0.5 \leq 0.75 \\
P(X>2) &= \frac{1}{6} \leq 0.25
\end{aligned}
\tag{5.107}
$$

$x_{0.5}$: Median

alle Werte x im Intervall [1,2] erfüllen die Bedingung

$$
\begin{aligned}
P(X<x) &\leq 0.5 \\
P(X>x) &\leq 0.5
\end{aligned}
\tag{5.108}
$$

Wir wählen deshalb

$$x_{0.5} = \frac{1}{2}(1+2) = 1.5 \tag{5.109}$$

Ist das p-Quantil eindeutig bestimmt, so lässt sich x_p auch dadurch bestimmen, dass man den kleinsten Wert von X sucht, der die Bedingung

$$F_X(x) \geq p \tag{5.110}$$

erfüllt.

Für stetige Zufallsvariablen bestimmt sich das p-Quantil allgemein aus der Bedingung:

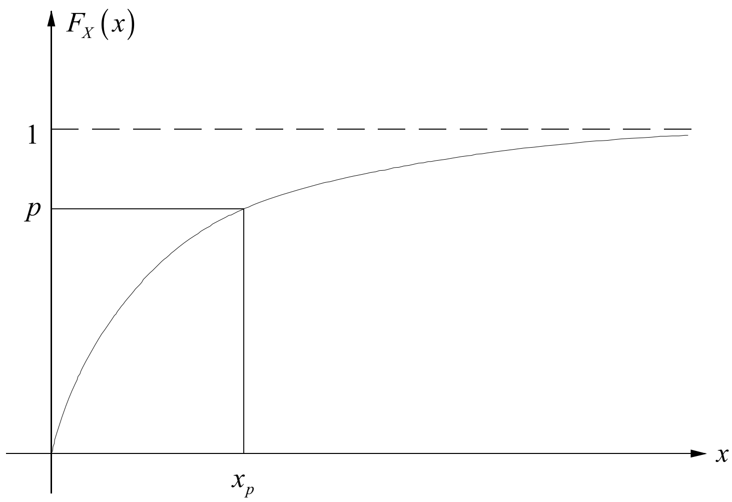
$$x_p \quad \text{so, dass} \quad F_X(x_p) = p$$

<u>Beispiel:</u>
X sei eine exponentialverteilte Zufallsvariable mit

$$\begin{aligned} f_X(x) &= \alpha e^{-\alpha x} \qquad \alpha > 0, \quad x \geq 0 \\ F_X(x) &= 1 - e^{-\alpha x} \end{aligned} \tag{5.111}$$

Für das p-Quantil x_p bestimmt man x so, dass

$$\begin{aligned} F_X(x) &= 1 - e^{-x\alpha} = p \\ x_p &= \frac{-1}{\alpha} \ln(1-p) \end{aligned} \tag{5.112}$$



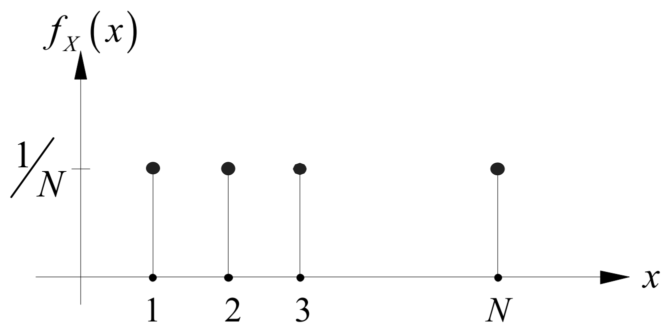
6. SPEZIELLE DISKRETE VERTEILUNGEN

In der praktischen Anwendung der Wahrscheinlichkeitsrechnung stösst man immer wieder auf Situationen, deren wahrscheinlichkeitstheoretische Beschreibung mit den dazugehörigen Zufallsvariablen prinzipiell dieselbe ist. Wir werden im folgenden eine Auswahl wichtiger Modelle darstellen. Die Kunst besteht dann darin, den konkreten Einzelfall dem richtigen, problemadäquaten Modell zuzuordnen.

6.1 Diskrete Gleichverteilung

Eine Zufallsvariable X kann die Werte $1,2,...,N$ mit gleicher Wahrscheinlichkeit annehmen. Die Wahrscheinlichkeitsfunktion ist dann

$$f_X(x) = \begin{cases} \dfrac{1}{N} & x = 1, 2, \ldots, N \\ 0 & sonst \end{cases} \tag{6.1}$$



Satz 6.1

Mittelwert, Varianz, momenterzeugende Funktion

$$\begin{aligned} E(X) &= \mu = \frac{N+1}{2} \\ V(X) &= \sigma^2 = \frac{N^2-1}{12} \\ m_X(t) &= \sum_{x=1}^{N} e^{xt}\,\frac{1}{N} \end{aligned} \tag{6.2}$$

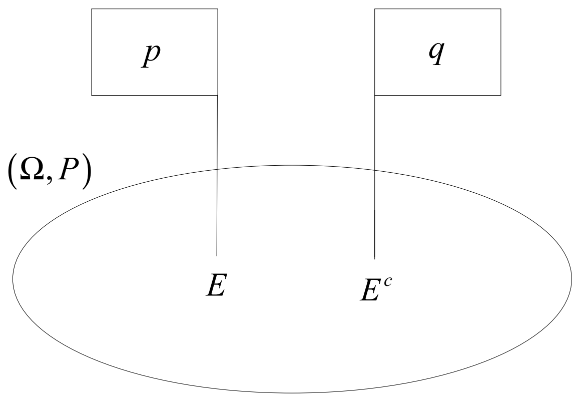
Beweis für die Varianz

$$
\begin{aligned}
\sigma^2 &= E\left(X^2\right) - \left(E\left(X\right)\right)^2 = \sum_{x=1}^{N} x^2 \cdot \frac{1}{N} - \left(\frac{N+1}{2}\right)^2 \\
&= \frac{1}{N}\,\frac{N\left(N+1\right)\left(2N+1\right)}{6} - \frac{\left(N+1\right)^2}{4} = \left(\frac{N+1}{12}\right) \cdot \left(N-1\right) \\
&= \frac{N^2-1}{12}
\end{aligned}
\tag{6.3}
$$

6.2 Verteilungen im Zusammenhang mit Bernoulliexperimenten

6.2.1 Bernoulliexperiment

Ein Zufallsexperiment ε heisst Bernoulliexperiment, wenn es genau 2 Ausgänge besitzt. Im Hinblick auf eine einheitliche Notation bezeichnet man die möglichen Ausgänge mit Erfolg E und Misserfolg E^c. Die Wahrscheinlichkeit für Erfolg sei p und jene für Misserfolg $q = 1 - p$.



Die über Ω definierte Zufallsvariable X mit

$X(E) = 1$
$X(E^c) = 0$

heisst Bernoullivariable.

Trotz oder gerade wegen ihrer einfachen Struktur besitzen Bernoullivariablen ein weites Einsatzfeld; es wird lediglich eine Dichotomisierung vorausgesetzt.

Typische Beispiele sind etwa der Münzwurf, „gut“ oder „defekt“ in der Qualitätskontrolle, Überleben oder Tod in der Versicherung, Gewinn oder Nichtgewinn beim Spiel etc.

6.2.2 Bernoulliverteilung

Eine Bernoullivariable X wird einmal beobachtet. Die Wahrscheinlichkeitsfunktion $f_X(\cdot)$ besitzt dann die Form

$$f_X(x) = \begin{cases} p^x (1-p)^{1-x} & x = 0,1 \\ 0 & \text{sonst} \end{cases} \tag{6.4}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ q & 0 \le x < 1 \\ 1 & \ge 1 \end{cases} \tag{6.5}$$

Der einfache Wertebereich für X hat dazu geführt, dass X auch als (0,1)-Variable bezeichnet wird.

Satz 6.2

Mittelwert, Varianz, momenterzeugende Funktion

$$\begin{aligned} \mu &= E(X) = p \\ \sigma^2 &= V(X) = p \cdot q \\ m_X(t) &= q + pe^t \end{aligned} \tag{6.6}$$

Beweis

$$\begin{aligned} \mu &= E(X) = 0 \cdot q + 1 \cdot p = p \\ \sigma^2 &= E(X^2) - \mu^2 = 0^2 \cdot q + 1^2 \cdot p - p^2 = p(1-p) = pq \\ m_X(t) &= E(e^{tX}) = e^{t \cdot 0} \cdot q + e^{t \cdot 1} \cdot p = q + pe^t \end{aligned} \tag{6.7}$$

6.2.3 Binomialverteilung

Wir betrachten eine Folge von *n unabhängigen Bernoullivariablen bei konstanter Erfolgswahrscheinlichkeit*.

Beispiele:

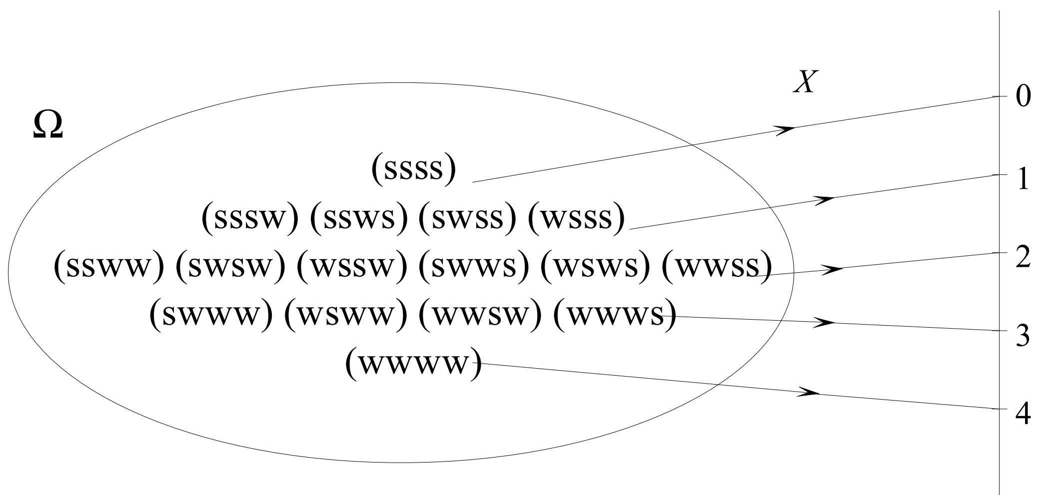
1.) Jemand spielt Sport Toto und lässt für jedes Spiel den Zufall entscheiden, welches der Zeichen 1,2 oder x angekreuzt werden soll. Insgesamt sind 13 Spiele vorauszusagen, wobei jedes Spiel mit Wahrscheinlichkeit 1/3 richtig vorausgesagt (E) wird.

2.) Eine Urne enthält 20 Kugeln, von denen 15 weiss und 5 schwarz sind. Es werden nacheinander 4 Kugeln gezogen, wobei die entnommene Kugel wieder zurückgelegt wird. Die Züge sollen unabhängig voneinander erfolgen.

Wir definieren die Zufallsvariable

X = Anzahl Bernoullivariablen mit dem Ausgang E

Urnenbeispiel: E: weiss
E^c: schwarz



Eine bestimmte Folge von n unabhängigen Bernoullivariablen, bei welcher der Ausgang E x-mal und E^c (n-x)-mal vorkommt

$$\begin{array}{l} E\ E\ E^c\ E\ E^c \ldots\ldots E^c\ E \\ p\ p\ q\quad p\ q \ldots\ldots q\quad p \end{array} \tag{6.8}$$

tritt mit Wahrscheinlichkeit $p^x\ q^{n-x}$ ein. Wenn wir nur nach dem Ereignis x Erfolge fragen, spielt die Reihenfolge der Ausgänge keine Rolle; wichtig ist nur, dass x E- und (n - x) E^c-Elemente vorhanden sind.

Insgesamt gibt es $\binom{n}{x}$ Möglichkeiten, x E-Elemente und (n - x) E^c-Elemente unter Berücksichtigung der Reihenfolge anzuordnen (Permutationen mit Wiederholungen). Alle paarweise unvereinbaren Anordnungen besitzen dieselbe Wahrscheinlichkeit

$$p^x\,(1-p)^{n-x} \tag{6.9}$$

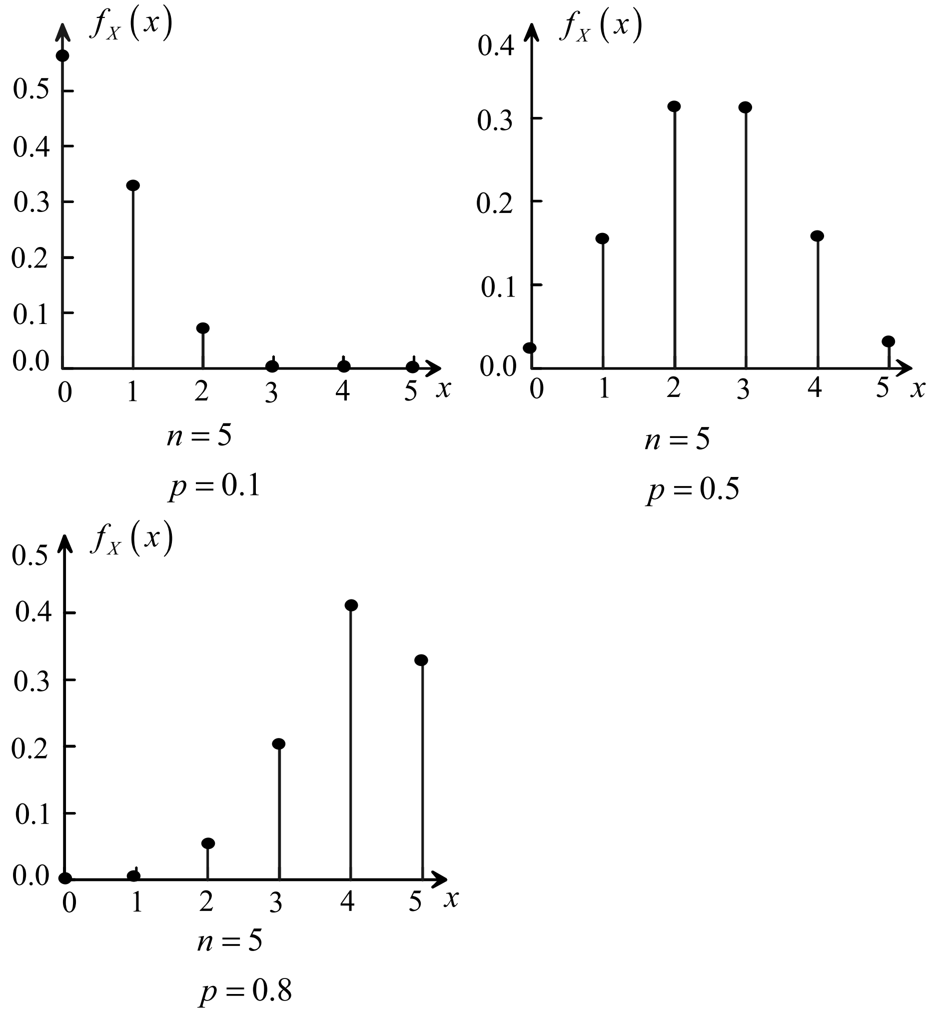
Für die Zufallsvariable

X: Anzahl Erfolge in n unabhängigen Bernoulliversuchen mit konstanter Erfolgswahrscheinlichkeit p

erhält man die Wahrscheinlichkeitsverteilung

$$f_X(x)\begin{cases} \binom{n}{x} p^x\ q^{n-x} & x = 0,1,2,\ldots,n \\ 0 & \text{sonst} \end{cases} \tag{6.10}$$

$$F_X(x) = P(X \le x) = \sum_{k=0}^{x} \binom{n}{k} p^k\ q^{n-k} \tag{6.11}$$



$f_X(x)$ erfüllt die Voraussetzungen einer Wahrscheinlichkeitsfunktion. Es gilt

$$\sum_{x=0}^{n} f_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x \, q^{n-x} = (p+q)^n = 1 \tag{6.12}$$

(Daher der Name Binomialverteilung!)

X ist abhängig von den beiden Parametern n und p.

Wir verwenden deshalb die Bezeichnung

$$X \sim B(n, p) \tag{6.13}$$

für die Aussage, dass X binomialverteilt ist mit den Parametern n und p.

Satz 6.3

Mittelwert, Varianz, momenterzeugende Funktion

$$\begin{aligned} \mu &= E(X) = np \\ \sigma^2 &= V(X) = npq \\ m_X(t) &= \left(q + pe^t\right)^n \end{aligned} \tag{6.14}$$

<u>Beweis</u>

$$\begin{aligned} m_X(t) &= E\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x\, q^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} \left(e^t p\right)^x q^{n-x} = \left(q + pe^t\right)^n \\ m_X'(t) &= npe^t\, \left(q + pe^t\right)^{n-1} \\ m_X''(t) &= npe^t\, \left(q + pe^t\right)^{n-1} + n(n-1)pe^t\left(q + pe^t\right)^{n-2} pe^t \\ m_X'(0) &= np \\ m_X''(0) &= np + n(n-1)p^2 \\ E(X) &= m_X'(0) = np \\ V(X) &= m_X''(0) - (np)^2 = np + n^2 p^2 - np^2 - n^2 p^2 = np(1-p) = npq \end{aligned} \tag{6.15}$$

Beachte:

- Für $n = 1$ ist die Binomialverteilung identisch mit der Bernoulliverteilung. (Vergleiche die momenterzeugenden Funktionen)

- Die binomialverteilte Zufallsvariable X entspricht der Summe von n unabhängigen Bernoullivariablen.

- Ist $X \sim B(n, p)$, so gilt für $Y = n - X \sim B(n, 1-p)$.

Wir berechnen die Wahrscheinlichkeiten für die richtigen Voraussagen X beim Sport-Toto Beispiel.

$$X \sim B\left(13, \frac{1}{3}\right)$$
$$f_X(x) = \binom{13}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{13-x} \tag{6.16}$$

x	$f_X(x)$	$F_X(x)$
0	0.00514	0.00514
1	0.03340	0.03854
2	0.10020	0.13873
3	0.18369	0.32242
4	0.22962	0.55204
5	0.20665	0.75869
6	0.13777	0.89646
7	0.06888	0.96535
8	0.02583	0.99118
9	0.00718	0.99835
10	0.00144	0.99979
11	0.00020	0.99998
12	0.00002	0.99999
13	0.00000	1.00000

$$E(X) = np = \frac{13}{3} = 4\frac{1}{3} \tag{6.17}$$

d.h. wenn jemand sehr oft einen Totoschein nach der obigen Methode ausfüllt, wird er im Mittel den Ausgang von 4 1/3 Partien richtig voraussagen. Wird ein Gewinn für 10 oder mehr richtig vorausgesagte Spiele ausbezahlt, so beträgt seine Gewinnwahrscheinlichkeit

$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.99835 = 0.00165.$

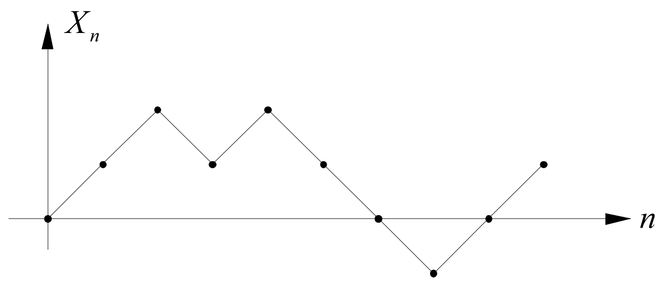
Die Berechnung binomialer Wahrscheinlichkeiten ist vor allem für grosse Werte von n sehr mühsam. Für grössere Werte von n verwendet man mit Vorzug die folgende *Rekursionsbeziehung* binomialverteilter Zufallsvariablen

$$f_X(0) = (1-p)^n$$
$$f_X(x+1) = f_X(x)\,\frac{(n-x)}{(x+1)}\,\frac{p}{q} \qquad x = 0,1,\ldots,n-1 \tag{6.18}$$

<u>Beweis</u>

$$\begin{aligned} f_X(x+1) &= \binom{n}{x+1} \cdot p^{x+1}\, q^{n-x-1} = \frac{p}{q}\binom{n}{x+1} p^x\, q^{n-x} \\ &= \frac{p}{q}\,\frac{n!}{(x+1)!\,(n-x-1)!}\, p^x\, q^{n-x} \\ &= \frac{p}{q}\,\frac{(n-x)}{(x+1)}\,\frac{n!}{x!\,(n-x)!}\, p^x\, q^{n-x} \\ &= \frac{p}{q}\,\frac{(n-x)}{(x+1)}\, f_X(x) \end{aligned} \tag{6.19}$$

Ein wichtiges Anwendungsfeld der Binomialverteilung findet sich in der Theorie der sog. "Random walks". Im einfachsten Fall wird angenommen, dass sich ein Partikel pro Zeiteinheit längs einer Geraden um eine Einheitsstrecke nach rechts mit Wahrscheinlichkeit p und nach links mit Wahrscheinlichkeit q bewege. X_n bezeichne die Position des Partikels nach n Schritten. Die folgende Graphik zeigt einen möglichen Weg, wenn X_n als Funktion der Zeit aufgetragen wird.



Formal lässt sich der obige Prozess auf der Basis eines Bernoulliexperimentes beschreiben. Mit Erfolg bezeichnen wir einen Schritt nach rechts und mit Misserfolg einen nach links. Für den i-ten Schritt definieren wir die Zufallsvariable

$$Y_i = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p \\ -1 & \text{mit Wahrscheinlichkeit } q \end{cases} \tag{6.20}$$

und es folgt für die Position nach n Schritten

$$X_n = \sum_{i=1}^{n} Y_i \tag{6.21}$$

mit folgenden möglichen Ausprägungen

$$-n,\ -n+2,\ -n+4\ ,....,n-4,\ n-2,\ n \tag{6.22}$$

wobei unterstellt wird, dass der Prozess im Ursprung beginnt.

Die Verbindung zur Binomialverteilung liefert die X_n zugeordnete Binomialvariable W, welche die Anzahl Erfolge misst. Es gilt

$$\begin{array}{lllllll} X_n = -n, & -n+2, & -n+4, & \dots, & n-4, & n-2, & n \\ W = 0, & 1, & 2, & \dots, & n-2, & n-1, & n \end{array} \tag{6.23}$$

Allgemein

$$\begin{aligned} X_n &= 2W - n \qquad \text{mit} \quad W \sim B(n,p) \\ W &= \frac{X_n + n}{2} \end{aligned} \tag{6.24}$$

Für die Verteilung von X_n gilt somit

$$\begin{gathered} P(X_n = x) = f_W\left(\frac{x+n}{2}\right) = \binom{n}{\frac{x+n}{2}} p^{\frac{x+n}{2}} q^{\frac{n-x}{2}} \\ x = -n, -n+2, ..., n-2, n \end{gathered} \tag{6.25}$$

$$
\begin{aligned}
E(X_n) &= E(2W-n) = 2E(W) - n = 2np - n = n(2p-1) \\
&= n(p-q) \\
V(X_n) &= V(2W-n) = 4V(W) = 4npq
\end{aligned}
\tag{6.26}
$$

Anmerkung: $X_n = Y_1 + Y_2 + \ldots + Y_n$ wird auch als binomialer Random walk bezeichnet.

Anstelle eines Partikels, welches sich längs einer Achse bewegt, kann man sich auch einen Aktienkurs vorstellen, welcher zu äquidistanten Zeitpunkten entweder um einen bestimmten Betrag nach oben resp. nach unten variiert.

Die Bewegungen müssen nicht notwendigerweise eine Einheit betragen; der Hinweis auf die Modellierung von Kursverläufen legt vielmehr Änderungen um beliebige Beträge *d* nach oben und unten nahe. Damit ändern sich allerdings die prinzipiellen Überlegungen nicht.

Mit

$$
Y_i = \begin{cases} d & \text{mit Wahrscheinlichkeit } p \\ -d & \text{mit Wahrscheinlichkeit } q \end{cases}
\tag{6.27}
$$

folgt

$$
\begin{aligned}
E(Y_i) &= d(p-q) \\
V(Y_i) &= 4d^2\, pq
\end{aligned}
\tag{6.28}
$$

Die Zufallsvariable

$$
X_n = \sum_{i=1}^{n} Y_i
\tag{6.29}
$$

lässt sich wieder durch eine Binomialverteilung beschreiben.

$$
P(X_n = x) = \binom{n}{w} p^w\, q^{n-w} \qquad w = 0,1,2,\ldots,n
\tag{6.30}
$$

mit $x = d\,(2w-n)$

$$E(X_n) = nd(p-q)$$
$$V(X_n) = 4nd^2 \; pq \tag{6.31}$$

6.2.4 Die Verteilung von Anteilen

Ist X binomialverteilt mit den Parametern n und p, so beschreibt die Zufallsvariable

$$P_E = \frac{X}{n} \tag{6.32}$$

den *Anteil* der Erfolge in n Versuchen. P_E nimmt die Werte

$$0, \frac{1}{n}, \frac{2}{n}, \ldots., \frac{n-1}{n}, 1 \tag{6.33}$$

mit den Wahrscheinlichkeiten

$$P(P_E = p_E) = P\left(\frac{X}{n} = p_E\right) = P(X = np_E) = \binom{n}{np_E} \cdot p^{np_E} \; q^{n(1-p_E)}$$
$$p_E = 0, \frac{1}{n}, \ldots., 1 \tag{6.34}$$

an.

Nach den allgemeinen Sätzen über Mittelwert und Varianz transformierter Zufallsvariablen erhält man

$$E(P_E) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} np = p$$
$$V(P_E) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} npq = \frac{pq}{n} \tag{6.35}$$

Beispiel:

Aus langjähriger Erfahrung ist bekannt, dass bei einer Abstimmung 40% der Berechtigten an die Urne gehen. Aus einer grossen Bevölkerungsgruppe werden 20 zufällig ausgewählte Personen über ihr Stimmverhalten befragt.

Mit welcher Wahrscheinlichkeit beträgt unter der obigen Hypothese der Anteil der Stimmenden in der Stichprobe

1.) genau 40%
2.) mehr als 50%

X: Anzahl Stimmende in der Stichprobe
P_E: Anteil der Stimmenden in der Stichprobe

1.)

$$P(P_E = 0.4) = P\left(\frac{X}{20} = 0.4\right) = P(X = 8) = \binom{20}{8} 0.4^8 \cdot 0.6^{12} = 0.1797 \quad (6.36)$$

2.)

$$P(P_E > 0.5) = P\left(\frac{X}{20} > 0.5\right) = 1 - P(X \le 10) = 1 - \sum_{x=0}^{10} \binom{20}{x} 0.4^x \cdot 0.6^{20-x} = 0.1275 \quad (6.37)$$

Wir betrachten noch einmal das Urnenbeispiel von S. 84.

Gegeben sind 20 Kugeln, von denen 15 weiss und 5 schwarz sind. Es werden 4 Kugeln gezogen, wobei nach jedem Zug die Kugel wieder zurückgelegt wird, nachdem die Farbe notiert wurde.

X: Anzahl weisse Kugeln
$X \sim B(4, 0.75)$

x	0	1	2	3	4
$f_X(x)$	0.0039	0.0469	0.2109	0.4219	0.3164
$F_X(x)$	0.0039	0.0508	0.2617	0.6836	1

Durch das Zurücklegen der Kugeln garantieren wir eine konstante Erfolgswahrscheinlichkeit. In der Praxis kann man aber häufig die ausgewählten Elemente nicht mehr zurücklegen (z.B. bei der Überprüfung von Nahrungsmittelkonserven). Diese Situation entspricht dem sog. hypergeometrischen Modell.

6.2.5 Hypergeometrische Verteilung

Einer Urne mit K weissen (E) und (N - K) roten Kugeln (E^c) wird eine Stichprobe vom Umfang n entnommen, wobei die Kugeln, im Gegensatz zum Binomialmodell, nicht mehr zurückgelegt werden. Man spricht daher von einer *Stichprobe ohne Zurücklegen*. Mit jedem Zug wird die Zusammensetzung der Urne und damit auch die Erfolgswahrscheinlichkeit verändert.

Zur Ermittlung der Wahrscheinlichkeitsverteilung der Zufallsvariablen

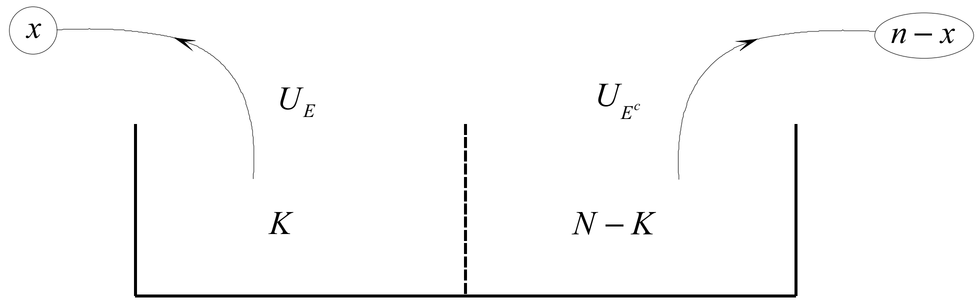
X: Anzahl Erfolgselemente (weisse Kugeln) in einer Stichprobe ohne Zurücklegen vom Umfang n

benützen wir die klassische Definition der Wahrscheinlichkeit

$$P(X=x)=\frac{\text{Anzahl Stichproben mit } x \text{ Erfolgselementen}}{\text{Gesamtzahl der möglichen Stichproben}}=\frac{g}{m} \qquad (6.38)$$

m: Aus N Elementen kann man auf $\binom{N}{n}$ verschiedene Arten n Elemente auswählen, wenn die Reihenfolge keine Rolle spielt.

g: Zur Ermittlung der günstigen Fälle denken wir uns die Urne aufgeteilt in eine Erfolgsurne U_E (mit K Elementen) und in eine Nichterfolgsurne U_{E^c} (mit (N - K) Elementen).



x Erfolge werden realisiert, wenn x Elemente aus U_E - dies ist auf $\binom{K}{x}$ Arten möglich - und (n - x) Elemente aus U_{E^c} - dies ist auf $\binom{N-K}{n-x}$ Arten möglich - gezogen werden.

Jede Teilstichprobe aus U_E kann mit allen Teilstichproben aus U_{E^c} kombiniert werden. Nach dem allgemeinen Abzählprinzip erhält man für g:

$$g = \binom{K}{x}\binom{N-K}{n-x} \tag{6.39}$$

womit für die hypergeometrisch verteilte Zufallsvariable X die Wahrscheinlichkeitsfunktion folgt.

$$f_X(x) = \begin{cases} \dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} & \max\left[0, n-(N-K)\right] \le x \le \min\left[n, K\right] \\ 0 & \text{sonst} \end{cases} \tag{6.40}$$

obere Grenze: Es gibt höchstens K Erfolgselemente

untere Grenze: Es gibt höchstens (N - K) E^c-Elemente resp. mindestens n - (N - K) E-Elemente.

Die Werte N, K und n sind die drei expliziten Parameter der hypergeometrischen Verteilung. In der Terminologie der Binomialverteilung gilt

$$p = \frac{K}{N} \qquad q = 1 - \frac{K}{N} \tag{6.41}$$

Satz 6.4

Ist X hypergeometrisch verteilt mit den Parametern N, K und n, so gilt

$$\begin{aligned} E(X) &= n\frac{K}{N} \\ V(X) &= n\frac{K}{N}\left(1-\frac{K}{N}\right)\left(\frac{N-n}{N-1}\right) \end{aligned} \tag{6.42}$$

Die Beweise können mittels entsprechender Umformung der Binomialkoeffizienten geführt werden.

Vergleicht man Mittelwerte und Varianzen der Binomial- und der hypergeometrischen Verteilung, so stellt man fest, dass

- die Mittelwerte dieselbe Struktur aufweisen
- die Varianzen sich durch den Faktor $\frac{N-n}{N-1}$, den sog. Korrekturfaktor für Stichproben ohne Zurücklegen, voneinander unterscheiden. Die Varianzen stimmen überein für den Fall $n = 1$, und für festes n gilt

$$\lim_{N\to\infty} \frac{N-n}{N-1} = 1 \tag{6.43}$$

- $V(X) = 0$ falls $n = N$ (vollständige Information)

Beispiel:
Beim Schweizer Zahlenlotto werden aus 45 Kugeln 6 ohne Zurücklegen ausgewählt. Mit welcher Wahrscheinlichkeit werden 0,1,2,....,6 Zahlen richtig vorausgesagt, wenn man nicht über hellseherische Fähigkeiten verfügt?

X = Anzahl richtige Voraussagen
$N = 45$ $K = 6$ $n = 6$

x	$f_X(x)$	$F_X(x)$
0	0.40056	0.40056
1	0.42413	0.82469
2	0.15147	0.97616
3	0.02244	0.99861
4	0.001365	0.99997
5	0.0000287	0.99999987
6	0.0000001228	1

Ein Gewinn wird ausbezahlt, falls $X \geq 3$. Die Gewinnwahrscheinlichkeit beträgt daher $P(X \geq 3) = 0.02384$.

Die Berechnung hypergeometrischer Wahrscheinlichkeiten ist bereits für moderate Werte von N, K und n sehr aufwendig. Man behilft sich mit Tabellen, Computerprogrammen oder mit Approximationen durch andere Verteilungen.

Für die praktische Berechnung erweist sich folgende *Rekursionsformel* als zweckmässig. Nach einigen Umformungen der Binomialkoeffizienten erhält man

$$f_X(x+1) = f_X(x)\,\frac{(K-x)}{(x+1)}\,\frac{(n-x)}{(N-K-n+x+1)} \tag{6.44}$$

Dem hypergeometrischen Modell liegt im Gegensatz zum Binomialmodell eine Stichprobe ohne Zurücklegen zugrunde. Ist der Stichprobenumfang n sehr klein gegenüber N und K, so ist zu vermuten, dass der Unterschied zwischen Stichproben “mit” und “ohne” Zurücklegen verschwindet.

Ist X hypergeometrisch verteilt, so gilt

$$\begin{aligned}
f_X(x) &= \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} = \frac{\dfrac{K!}{x!\,(K-x)!}\,\dfrac{(N-K)!}{(n-x)!\,(N-K-(n-x))!}}{\dfrac{N!}{n!\,(N-n)!}} \\
&= \frac{n!}{x!\,(n-x)!}\,\frac{\dfrac{K!}{(K-x)!}\,\dfrac{(N-k)!}{(N-K-n+x)!}}{N(N-1)....(N-n+1)} \\
&= \binom{n}{x}\frac{K(K-1)(...)(K-x+1)(N-K)(N-K-1)(...)(N-K-(n-x)+1)}{N(N-1)(...)(N-n+1)} \\
&\approx \binom{n}{x}\left(\frac{K}{N}\right)^x\left(\frac{N-K}{N}\right)^{n-x}
\end{aligned} \tag{6.45}$$

Wir halten fest

Hypergeometrische Verteilung	$\approx$	Binomialverteilung
$n << N, K, N-K$		$n, p = \dfrac{K}{N}$

Als *Faustregel* für die Approximation der hypergeometrischen Verteilung durch die Binomialverteilung verwenden wir

$$\text{Auswahlsatz } = \frac{n}{N} < 0.05 = 5\% \tag{6.46}$$

<u>Beispiel:</u>
$N = 1000$ $K = 300$ $n = 10$ $p = 0.3$

x	$f_X(x)$	
	hypergeometrisch	binomial
0	0.0277	0.0282
1	0.1203	0.1211
2	0.2339	0.2335
3	0.2682	0.2668
4	0.2008	0.2001
5	0.1026	0.1030
6	0.0363	0.0367
7	0.0087	0.0090
8	0.0014	0.0015
9	0.0001	0.0001
10	0.0000	0.0000

<u>Beispiel:</u>
Ein Warenposten von 20 Elementen werde liquidiert, wobei bekannt ist, dass 4 Elemente defekt sind. Es werden 5 zufällig ausgewählte Elemente gekauft. Man diskutiere die Zufallsvariable

X: Anzahl defekte Elemente in der Auswahl

Wahrscheinlichkeitsverteilung, Mittelwert, Varianz?

6.2.6 Poissonverteilung

Def. 6.1

Eine Zufallsvariable X heisst poissonverteilt, falls deren Wahrscheinlichkeitsfunktion gegeben ist durch

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad x = 0,1,2,\ldots \qquad \lambda > 0 \tag{6.47}$$

$f_X(x)$ erfüllt die Voraussetzungen einer Wahrscheinlichkeitsfunktion. Es gilt

$$\sum_{x=0}^{\infty} f_X(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1 \tag{6.48}$$

Beachte:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \tag{6.49}$$

Satz 6.5

Mittelwert, Varianz, momenterzeugende Funktion

$$\begin{aligned} E(X) &= \lambda \\ V(X) &= \lambda \\ m_X(t) &= e^{\lambda(e^t - 1)} \end{aligned} \tag{6.50}$$

Beweis

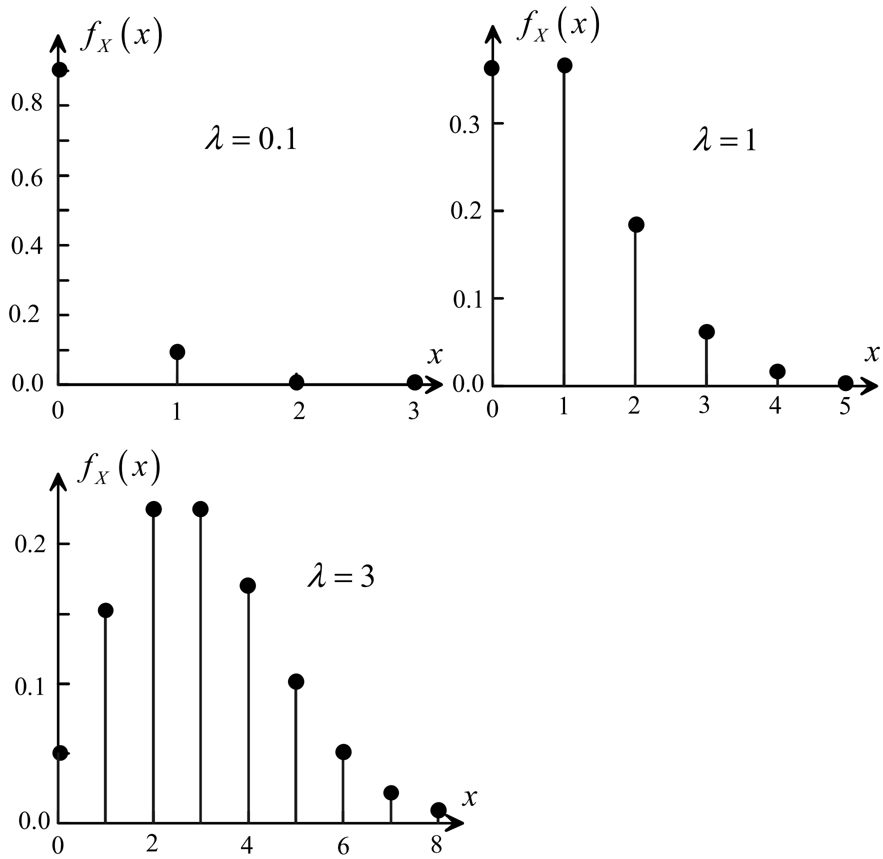
$$\begin{aligned} m_X(t) &= E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \\ m'_X(t) &= e^{\lambda(e^t - 1)} \lambda e^t = \lambda e^t \, m_X(t) \\ m''_X(t) &= \lambda e^t m_X(t) + \lambda e^t \, m'_X(t) \\ &= \lambda e^t m_X(t)\left(1 + \lambda e^t\right) \end{aligned} \tag{6.51}$$

$$
\begin{aligned}
&m_X'(0) = \lambda \\
&m_X''(0) = \lambda^2 + \lambda \\
\\
&E(X) = m_X'(0) = \lambda \\
&V(X) = m_X''(0) - \left(m_X'(0)\right)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}
\tag{6.52}
$$

Beachte: Mittelwert und Varianz stimmen überein! λ ist der einzige Parameter der Poissonverteilung.

Die Berechnung der Wahrscheinlichkeiten poissonverteilter Zufallsvariablen erfolgt zweckmässigerweise über folgende *Rekursionsbeziehung*

$$
\begin{aligned}
f_X(0) &= e^{-\lambda} \\
f_X(x+1) &= f_X(x)\frac{\lambda}{x+1}
\end{aligned}
\tag{6.53}
$$



Die Poissonverteilung eignet sich in verschiedensten Bereichen zur Darstellung zufälliger Phänomene. Der Realisationsbereich poissonverteilter Zufallsvariablen umfasst die Menge der nicht negativen ganzen Zahlen. Diese Eigenschaft macht die Poissonverteilung zum Kandidaten für die Beschreibung von Zählvariablen - z.B. die Anzahl Kollisionen innerhalb einer Woche auf einer bestimmten Kreuzung, die Pannen in Atomkraftwerken innerhalb eines Jahres, die Anzahl Tippfehler pro Seite.

6.2.7 Approximation der Binomialverteilung für kleine Erfolgswahrscheinlichkeiten

Die Praxis zeigt oft Binomialmodelle mit kleinen (grossen) Erfolgswahrscheinlichkeiten. Typische Beispiele finden sich etwa in der Qualitätskontrolle oder in der Versicherung. In diesen Fällen wird die Behandlung mit dem Binomialmodell bereits für moderat grosse Werte von n sehr aufwendig und fehleranfällig. Es stellt sich daher die Frage nach einer allfälligen Grenzverteilung, wobei folgende Annahmen getroffen werden.

$$\begin{aligned}
&X \sim B(n,p) \\
&p \to 0, \quad n \to \infty, \quad np = \lambda \,(\text{konstant})
\end{aligned} \tag{6.54}$$

In idealer Weise sind die obigen Voraussetzungen bei folgendem Spiel erfüllt. 2 identische Spiele mit n Karten werden untereinander aufgereiht, und es wird nach der Anzahl übereinstimmender Karten gefragt. Mit wachsendem n nimmt die Wahrscheinlichkeit für Übereinstimmung $\frac{1}{n}$ ab. Gleichzeitig nimmt die Anzahl Positionen für mögliche Übereinstimmungen zu und zwar so, dass der Mittelwert konstant, nämlich 1 bleibt.

Für den Grenzübergang gilt

$$\begin{aligned}
f_X(x) &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n(n-1)(\ldots)(n-x+1)}{x!}\,\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{1}{n}\right)(\ldots)\left(1-\frac{x-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{-x}
\end{aligned} \tag{6.55}$$

mit $n \to \infty$ und endlichem x gilt

$$\begin{aligned}
&\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda} \\
&\lim_{n\to\infty}\left(1-\frac{1}{n}\right)(\ldots)\left(1-\frac{x-1}{n}\right) = 1 \\
&\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = 1
\end{aligned} \tag{6.56}$$

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \to \frac{\lambda^x}{x!} e^{-\lambda} \tag{6.57}$$

x	Binomial				Poisson
	$n = 5$ $p = 0.2$	$n = 10$ $p = 0.1$	$n = 50$ $p = 0.02$	$n = 100$ $p = 0.01$	$\lambda = 1$
0	0.3277	0.3487	0.3642	0.3660	0.3679
1	0.4096	0.3874	0.3716	0.3697	0.3679
2	0.2048	0.1937	0.1858	0.1849	0.1839
3	0.0512	0.0574	0.0607	0.0610	0.0613
4	0.0064	0.0112	0.0145	0.0149	0.0153
5	0.0003	0.0015	0.0027	0.0029	0.0031
6		0.0001	0.0004	0.0005	0.0005
7		0.0000	0.0001	0.0001	0.0001
8			0.0000	0.0000	0.0000

Wegen der kleinen Erfolgswahrscheinlichkeit der Binomialverteilung, die sich durch die Poissonverteilung approximieren lässt, nennt man letztere auch die Verteilung seltener Ereignisse.

<u>Beispiel</u>
Mit welcher Wahrscheinlichkeit erzielt man beim Lotto mindestens 1 mal 6 richtige Zahlen, wenn 1000 Tips zufällig und unabhängig voneinander ausgefüllt werden.

X: Anzahl Tips mit 6 richtigen Zahlen

$$X \sim B\left(1000, \frac{1}{\binom{45}{6}}\right) \tag{6.58}$$

$$\begin{aligned} \lambda &= np = 1000 \cdot \frac{1}{\binom{45}{6}} = 0.000122774 \\ P(X \geq 1) &= 1 - P(X = 0) = 1 - e^{-0.000122774} = 0.000122766 \end{aligned} \tag{6.59}$$

Beispiel
Ein Versichertenbestand enthält 1000 aktive Frauen mit dem Alter 35. Die jährliche Invalidierungswahrscheinlichkeit betrage für jede Versicherte 0.00046.

Mit welcher Wahrscheinlichkeit invalidieren pro Jahr
1.) genau 3 Versicherte
2.) keine Versicherte
3.) mindestens 2 Versicherte

falls angenommen werden kann, dass die Invalidierungen unabhängig voneinander erfolgen.

$$\begin{aligned} X &\sim B(1000, 0.00046) \\ \lambda &= np = 0.46 \end{aligned} \tag{6.60}$$

1.) $P(X = 3) \approx \frac{0.46^3}{3!} e^{-0.46} = 0.01024$

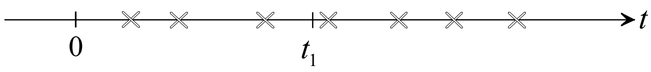
2) $P(X = 0) = e^{-0.46} = 0.63128$

3) $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 0.07833$

6.2.8 Der Poisson-Prozess

Die Poissonverteilung - die wir als Approximation der Binomialverteilung kennengelernt haben - besitzt Eigenschaften, wie sie von einer Vielzahl von sog. stochastischen Prozessen vorausgesetzt werden. Wir werden im folgenden einige grundsätzliche Aspekte eines derartigen Prozesses betrachten, ohne jedoch die letzten Details ausleuchten zu wollen.

Gegeben sei ein Prozess, bei dem im Verlaufe der Zeit bestimmte Ereignisse (z.B. Ankunft einer Person an einem Schalter, Ausfall einer Maschine, Telefonanrufe, Unfall auf einer Kreuzung) *unabhängig von einander* und *unabhängig vom jeweiligen Zeitpunkt t* eintreten.



Die Markierungen auf der Zeitachse beschreiben den Eintritt der Ereignisse. Im Intervall $[0, t_1]$ sind insgesamt 4 Ereignisse (4 Ankünfte am Schalter) eingetreten.

Wir nehmen nun an, dass eine nichtnegative Grösse α so existiert, dass die nachfolgenden sog. *Poisson-Annahmen* erfüllt sind.

1.) Die Wahrscheinlichkeit, dass in einem kleinen Intervall der Länge h genau ein Ereignis eintritt, ist approximativ $\alpha \cdot h$:
$P(\text{genau ein Ereignis im Intervall der Länge } h) = \alpha \cdot h + O_1(h)$

2.) Die Wahrscheinlichkeit, dass in einem kleinen Intervall der Länge h mehr als ein Ereignis eintritt, ist vernachlässigbar klein (verglichen mit der Wahrscheinlichkeit für genau ein Ereignis):
$P(\text{zwei oder mehr Ereignisse im Intervall } h) = O_2(h)$

3.) Die Anzahl Ereignisse innerhalb des Intervalls $[t_0, t_1]$ ist unabhängig von der Anzahl Ereignisse im Intervall $[t_2, t_3]$ $(t_1 < t_2)$.

Der Ausdruck $O(h)$ bedeutet $\lim\limits_{h \to O} \frac{O(h)}{h} = 0$, d.h. $O(h)$ strebt schneller gegen 0 als h;

z.B. $O(h) = h^2$.

Die Grösse α kann als mittlere Rate, mit der die Ereignisse pro Zeiteinheit eintreffen, interpretiert werden. Man bezeichnet sie deshalb auch als mittlere *Ankunftsrate*.

Zusammengefasst besagen die Poissonannahmen, dass die Wahrscheinlichkeit für den Eintritt eines Ereignisses in kleinen Zeitintervallen proportional zur Intervalllänge ist, dass ein mehrmaliges Eintreten des Ereignisses in kleinen Zeitintervallen praktisch ausgeschlossen ist, und, dass - implizit in 1.) und 3.) - ein stationäres Verhalten vorausgesetzt wird, d.h. dass die Eintretenswahrscheinlichkeit eines Ereignisses nur von der Länge des Intervalls, nicht aber von dessen Lage abhängig ist.

Satz 6.6

Unter den obigen Voraussetzungen 1 - 3 genügt die Zufallsvariable

$X(t)$ = Anzahl Ereignisse in einem Intervall der Länge t

einer Poissonverteilung mit dem Parameter $\lambda = \alpha \cdot t$

$$P\left(X(t) = x\right) = e^{-\alpha t} \frac{(\alpha t)^x}{x!} \qquad x = 0, 1, \ldots \tag{6.61}$$

Beweis

Wir verwenden für die Wahrscheinlichkeit von genau x Ereignissen E in einem Intervall der Länge t die Notation

$$P\left(X(t) = x\right) = p_x(t) \tag{6.62}$$

und betrachten die Intervalle $(0,t)$ resp. $(0, t+h)$

$$\begin{aligned} p_0(t+h) &= P\left(\text{kein } E \text{ in } (0,t) \text{ und kein } E \text{ in } (t, t+h)\right) \\ &= P\left(\text{kein } E \text{ in } (0,t)\right) \cdot P\left(\text{kein } E \text{ in } (t,t+h)\right) \\ &= p_0(t) \cdot p_0(h) \end{aligned} \tag{6.63}$$

Mit

$$p_0(h) = 1 - \alpha h - O_1(h) - O_2(h) \tag{6.64}$$

folgt

$$p_0(t+h) = p_0(t)\left(1-\alpha h - O_1(h) - O_2(h)\right)$$
$$\frac{p_0(t+h)-p_0(t)}{h} = -\alpha\, p_0(t) - \frac{O(h)}{h} \tag{6.65}$$

für $h \to 0$ erhält man

$$p_0'(t) = -\alpha\, p_0(t)$$
$$\text{resp.} \quad p_0(t) = e^{-\alpha t} \tag{6.66}$$

für $x \geq 1$ erhält man analog

$$p_x(t+h) = p_x(t)p_0(h) + p_{x-1}(t)p_1(h) \tag{6.67}$$

Aus

$$p_0(h) = 1-\alpha h - O_1(h) - O_2(h)$$
$$p_1(h) = 1 - p_0(h) - O_2(h) = \alpha h + O_1(h) \tag{6.68}$$

resultiert

$$p_x(t+h) = p_x(t)\left(1-\alpha h - O(h)\right) + p_{x-1}(t)\left(\alpha h + O_1(h)\right)$$
$$\frac{p_x(t+h)-p_x(t)}{h} = -\alpha\, p_x(t) + \alpha\, p_{x-1}(t) + \frac{O(h)}{h} \tag{6.69}$$

Für $h \to 0$

$$p_x'(t) = \frac{dp_x(t)}{dt} = -\alpha\, p_x(t) + \alpha\, p_{x-1}(t) \tag{6.70}$$

Man erhält also ein Differentialgleichungssystem

$$p_0'(t) = -\alpha\, p_0(t)$$
$$p_x'(t) = -\alpha\, p_x(t) + \alpha\, p_{x-1}(t) \tag{6.71}$$

Mit den Anfangsbedingungen

$$
\begin{aligned}
p_0(0) &= 1 \\
p_1(0) &= 0
\end{aligned}
\tag{6.72}
$$

bestätigt man einfach die nachfolgende Lösung des Systems

$$
p_x(t) = P\left(X(t) = x\right) = e^{-\alpha t} \frac{(\alpha t)^x}{x!} \qquad x = 0,1,\ldots
\tag{6.73}
$$

Beispiel
Die Telefonzentrale eines Betriebes vermittelt durchschnittlich 120 Anrufe pro Stunde. Man unterstelle, dass die Anzahl der Anrufe für beliebige Zeitspannen einer Poissonverteilung genügt, und dass als Masseinheit für die Zeit Minuten verwendet werden.

Mit welcher Wahrscheinlichkeit treffen

1.) keine Anrufe während einer 1-Minuten Periode ein
2.) mehr als 15 Anrufe in einer 5-Minuten Periode ein?

120 Anrufe pro Stunde entsprechen 2 Anrufen pro Minute.
α = 2 Anrufe/Minute

1.)

$$
P\left(X(1) = 0\right) = e^{-2\cdot 1} = 0.1353
\tag{6.74}
$$

2.)

$$
\begin{aligned}
P\left(X(5) > 15\right) &= 1 - P\left(X(5) \le 15\right) \\
&= 1 - \sum_{x=0}^{15} e^{-10} \frac{10^x}{x!} = 1 - 0.9513 = 0.0487
\end{aligned}
\tag{6.75}
$$

Beispiel
In 100 cm^3 eines Impfstoffes seien 100 Viren enthalten. Der Impfstoff werde in 100 Ampullen à je 1 cm^3 abgefüllt. Bei der Impfung ist mit Nebenreaktionen zu rechnen, sobald eine Ampulle mehr als einen Virus enthält.

Mit welcher Wahrscheinlichkeit sind bei einer Impfung Nebenreaktionen zu erwarten?

X bezeichne die Anzahl Viren pro cm^3 und genüge einer Poissonverteilung.

α bezeichnet die "Rate", mit welcher Viren pro Volumeneinheit vorkommen.

$$\alpha = \frac{100\,Viren}{100\text{ cm}^3} = 1\frac{V}{\text{cm}^3} \tag{6.76}$$

t entspricht dem Volumen der Ampullen (t = 1 cm^3) womit αt = 1Virus

$$P(X = x) = \frac{1^x}{x!} e^{-1} = \frac{e^{-1}}{x!} \tag{6.77}$$

$$P(X > 1) = 1 - P(X \le 1) = 1 - e^{-1} - e^{-1}\frac{1}{1!} = 1 - 2e^{-1} = 0.2642 \tag{6.78}$$

6.2.9 Geometrische Verteilung

Wir betrachten wieder eine Folge unabhängiger Bernoulliversuche. Die konstante Erfolgswahrscheinlichkeit sei p.

Mit welcher Wahrscheinlichkeit erscheint das Ereignis E erstmals *nach* x unabhängigen Versuchen, d.h. beim (x + 1)-ten Versuch?

X bezeichne die Anzahl Versuche, die dem ersten Erfolg vorangehen. Dann gilt

$$P(X = x) = q^x p \qquad x = 0,1,\ldots \tag{6.79}$$

Def. 6.2

Eine Zufallsvariable mit der Wahrscheinlichkeitsfunktion

$$P(X = x) = f_X(x) = \begin{cases} q^x p & x = 0,1,.... \\ 0 & \text{sonst} \end{cases} \tag{6.80}$$

heisst geometrisch verteilt (q = 1 - p).

$$P(X \le x) = F_X(x) = 1 - q^{x+1} \tag{6.81}$$

Satz 6.7

Mittelwert, Varianz, momenterzeugende Funktion

$$E(X) = \frac{q}{p}$$
$$V(X) = \frac{q}{p^2} \tag{6.82}$$
$$m_X(t) = \frac{p}{1 - qe^t}$$

Im Gegensatz zur Binomialverteilung mit einer fest vorgegebenen Anzahl Versuche wird bei der geometrischen Verteilung der Stichprobenumfang zur Zufallsvariablen.

<u>Beispiel</u>

Ein Billardspieler kann eine Serie bis zum ersten Fehlversuch fortsetzen. Letzterer treffe mit Wahrscheinlichkeit 0.25 zu. Mit welcher Wahrscheinlichkeit dauert das Spiel

1.) genau 6 Stösse
2.) höchstens 5 Stösse
3.) mindestens 7 Stösse

X: Anzahl erfolgreicher Stösse

1.) $P(X = 5) = 0.75^5 \cdot 0.25 = 0.0593$

2.) $P(X \leq 4) = 1 - 0.75^5 = 0.7627$

3.) $P(X \geq 6) = 1 - P(X \leq 5) = 1 - (1 - 0.75^6) = 0.1780$

Satz 6.8

Für geometrisch verteilte Zufallsvariablen X gilt

$$P(X \geq i + j | X \geq i) = P(X \geq j) \qquad i, j = 0,1,\ldots \tag{6.83}$$

Beweis

$$P(X \geq i+j | X \geq i) = \frac{P(X \geq i+j)}{P(X \geq i)} = \frac{1-P(X < i+j)}{1-P(X < i)}$$
$$= \frac{1-(1-q^{i+j})}{1-(1-q^{i})} = q^{j} = P(X \geq j) \tag{6.84}$$

Bei geometrisch verteilten Zufallsvariablen ist die bedingte Wahrscheinlichkeit, dass der erste Erfolg nach $i+j$ Versuchen erfolgt, falls schon i Misserfolge vorangegangen sind, identisch mit der unbedingten Wahrscheinlichkeit, dass der erste Erfolg nach dem j-ten Versuch eintritt.

Die i vorangehenden Misserfolge haben keinen Einfluss auf die nachfolgenden Versuche. Man nennt dies einen *Prozess ohne Gedächtnis*.

Beispiel
Beim Spiel “Eile mit Weile” darf erst begonnen werden, nachdem erstmals eine “5" gewürfelt wurde

1.) Mit welcher Wahrscheinlichkeit kann jemand erst nach dem 5. Wurf beginnen?

2.) Mit welcher Wahrscheinlichkeit muss ein Spieler nach 10 Runden immer noch warten, wenn bekannt ist, dass er schon 5 Runden gewartet hat?

3.) Nach wievielen Zügen kann man im Mittel das Spiel beginnen?

X = Anzahl Versuche vor der ersten “5"

Wir nehmen an, X wäre geometrisch verteilt mit $p = \frac{1}{6}$. (Unter welchen Annahmen trifft dies zu?)

1.)

$$P(X = 4) = \left(\frac{5}{6}\right)^{4} \frac{1}{6} = 0.0804 \tag{6.85}$$

2.)

$$P(X \geq 10 | X \geq 5) = \frac{P(X \geq 10)}{P(X \geq 5)} = \frac{1 - P(X \leq 9)}{1 - P(X \leq 4)}$$

$$= \frac{1 - \left(1 - \left(\frac{5}{6}\right)^{10}\right)}{1 - \left(1 - \left(\frac{5}{6}\right)^{5}\right)} = \left(\frac{5}{6}\right)^{5} = 0.4019 \tag{6.86}$$

3.)

$$E(X) = \frac{q}{p} = \frac{\frac{5}{6}}{\frac{1}{6}} = 5 \tag{6.87}$$

6.2.10 Negative Binomialverteilung

Wir betrachten eine Folge von unabhängigen Bernoulliversuchen mit einer konstanten Erfolgswahrscheinlichkeit p.

Die Zufallsvariable X bezeichne die Anzahl Misserfolge, welche dem r-ten Erfolg vorangehen. Für r Erfolge sind somit genau $x + r$ Versuche notwendig. r ist eine natürliche Zahl und x nimmt die Werte 0,1,2,... an. Zur Bestimmung der Wahrscheinlichkeitsfunktion von X überlegt man sich, dass im letzten ($x + r$)-ten Versuch Erfolg und in $x + r - 1$ vorangehenden Versuchen genau (r -1) Erfolge und x Misserfolge in beliebiger Reihenfolge vorkommen müssen. Die Kombination dieser Ereignisse führt zur Wahrscheinlichkeitsfunktion

$$f_X(x) = p\binom{x+r-1}{r-1} p^{r-1}(1-p)^x = \binom{x+r-1}{r-1} p^r (1-p)^x \qquad x = 0,1,2,... \tag{6.88}$$

Die negative Binomialverteilung heisst auch Pascal'sche Verteilung. Definiert man einen verallgemeinerten Binomialkoeffizienten

$$\binom{-r}{x} = \frac{(-r)(-r-1)(...)(-r-x+1)}{x!} = (-1)^x \binom{r+x-1}{x} = (-1)^x \binom{r+x-1}{r-1} \tag{6.89}$$

so lässt sich die Wahrscheinlichkeitsfunktion der negativen Binomialverteilung vereinfachen

$$P(X = x) = f_X(x) = (-1)^x \binom{-r}{x} p^r (1-p)^x \tag{6.90}$$

Die Bezeichnung “negative Binomialverteilung” geht auf den Ausdruck $\binom{-r}{x}$ in der Wahrscheinlichkeitsfunktion zurück.

Satz 6.9

Mittelwert, Varianz, momenterzeugende Funktion

$$\begin{aligned} E(X) &= \frac{rq}{p} \\ V(X) &= \frac{rq}{p^2} \\ m_X(t) &= \left(\frac{p}{1-qe^t} \right)^r \end{aligned} \tag{6.91}$$

Beachte:
Falls $r = 1$ entspricht die negative Binomialverteilung der geometrischen Verteilung.

Negativ binomialverteilte Zufallsvariablen sind bekannt zur Beschreibung von *Wartezeiten* (bis zum r-ten Erfolg).

Falls $r \to \infty$ und $p \to 0$ so, dass

$$\frac{rq}{p} = \lambda \tag{6.92}$$

so strebt die negative Binomialverteilung gegen die Poissonverteilung.

Beispiel
Von einem sehr grossen Warenposten sei bekannt, dass er 10% defekte Elemente enthält. Es werden solange Elemente entnommen, bis 5 defekte gezogen sind. Mit welcher Wahrscheinlichkeit werden diese 5 defekten Elemente in weniger als 11 Zügen gezogen? Wieviele Züge braucht man im Mittel, bis erstmals 5 defekte Elemente gezogen sind?

X = Anzahl Entnahmen nicht defekter Elemente vor dem 5. defekten Element.

1.) $P(X=0) = \binom{4}{4} 0.1^5 \cdot 0.9^0 \quad = 0.000010$

$P(X=1) = \binom{5}{4} 0.1^5 \cdot 0.9 \quad = 0.000045$

$P(X=2) = \binom{6}{4} 0.1^5 \cdot 0.9^2 \quad = 0.000121$

$P(X=3) = \quad = 0.000255$

$P(X=4) = \quad = 0.000459$

$P(X=5) = \binom{9}{4} 0.1^5 \cdot 0.9^5 \quad = 0.000744$

0.001635

2.) $$E(X) = \frac{rq}{p} = \frac{5 \cdot 0.9}{0.1} = 45 \tag{6.93}$$

6.2.11 Zusammenhang zwischen der Binomialverteilung und der negativen Binomialverteilung

X sei binomialverteilt mit den Parametern n und p.
Y sei negativ binomialverteilt mit den Parametern r und p.

X = Anzahl Erfolge in n Versuchen
Y = Anzahl Fehl-Versuche zur Realisierung von r Erfolgen

Satz 6.10

$$\begin{aligned} P(Y \leq n-r) &= P(X \geq r) \\ P(Y > n-r) &= P(X < r) \end{aligned} \tag{6.94}$$

Beweis

Falls in den ersten n Versuchen r oder mehr Erfolge vorliegen, so wurden höchstens n-r Fehl-Versuche zur Realisierung der ersten r Erfolge benötigt.

Gibt es weniger als r Erfolge in den n ersten Versuchen, so benötigt man mehr als n-r Fehl-Versuche, um r Erfolge zu erhalten.

Für das Beispiel von S. 110 erhält man

$$\begin{aligned} P\left(Y_5^{0.1} \leq 5\right) = P\left(X_{10}^{0.1} \geq r\right) &= 1 - P\left(X_{10}^{0.1} \leq 4\right) \\ &= 1 - 0.998365 = 0.001635 \end{aligned} \tag{6.95}$$

7. SPEZIELLE STETIGE VERTEILUNGEN

7.1 Stetige Gleichverteilung

Def. 7.1

Eine Zufallsvariable X heisst über dem Intervall $[a,b]$ stetig gleichverteilt, falls deren Dichte gegeben ist durch

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{sonst} \end{cases} \tag{7.1}$$

$$\text{mit} \quad -\infty < a < b < \infty$$

Für die Verteilungsfunktion gleichverteilter Zufallsvariablen gilt

$$F_X(x) = P\left(X \le x\right) = \begin{cases} 0 & x < a \\ \displaystyle\int_a^x \frac{1}{b-a}\,dx = \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases} \tag{7.2}$$

Satz 7.1

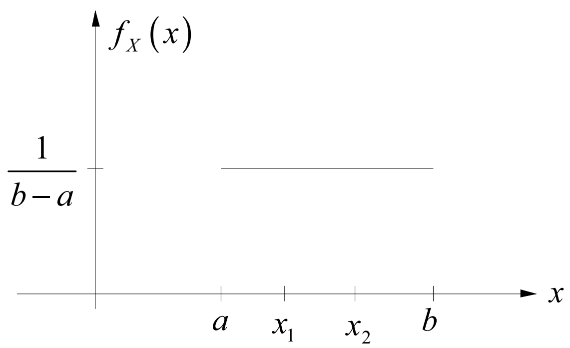
Mittelwert, Varianz, momenterzeugende Funktion

$$\begin{aligned} E\left(X\right) &= \frac{a+b}{2} \\ V\left(X\right) &= \frac{\left(b-a\right)^2}{12} \\ m_X\left(t\right) &= \frac{e^{bt}-e^{at}}{t\left(b-a\right)} \end{aligned} \tag{7.3}$$

Beweis

$$
\begin{aligned}
m_X(t) = E\left(e^{tX}\right) = \int_a^b e^{tx} \, \frac{1}{b-a} \, dx = \frac{e^{tx}}{t(b-a)} \Big|_a^b \\
= \frac{e^{bt} - e^{at}}{t(b-a)}
\end{aligned}
\tag{7.4}
$$

Die Gleichverteilung wird auch als Verteilung bei *vollständiger Unsicherheit* bezeichnet. Man weiss nur, dass X Werte im Intervall $[a,b]$ annehmen kann und dass Wahrscheinlichkeiten über beliebigen Teilintervallen gleicher Länge aus $[a,b]$ gleich gross sind.

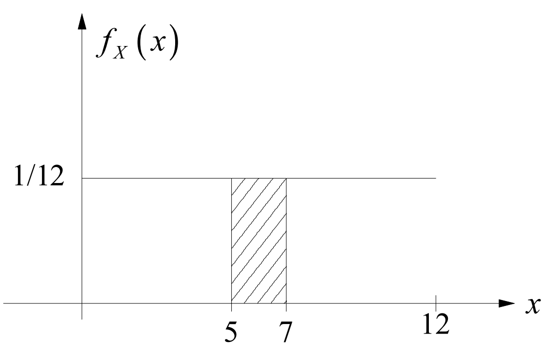


$$P(x_1 \le X \le x_2) = \frac{x_2 - x_1}{b-a} \tag{7.5}$$

Die Wahrscheinlichkeit hängt nur von der Intervalllänge $[x_1, x_2]$ ab.

Beispiel

Eine Uhr wird nicht mehr aufgezogen. Mit welcher Wahrscheinlichkeit fällt die Stillstandzeit zwischen 5 und 7 Uhr?



$$P(5 \le X \le 7) = \frac{7-5}{12} = \frac{1}{6} \tag{7.6}$$

7.2 Exponentialverteilung

Eine Zufallsvariable X heisst exponentialverteilt, falls ihre Dichte folgende Form besitzt.

Def. 7.2

$$f_X(x) = \begin{cases} \alpha\, e^{-\alpha x} & x \geq 0, \alpha > 0 \\ 0 & \text{sonst} \end{cases} \tag{7.7}$$

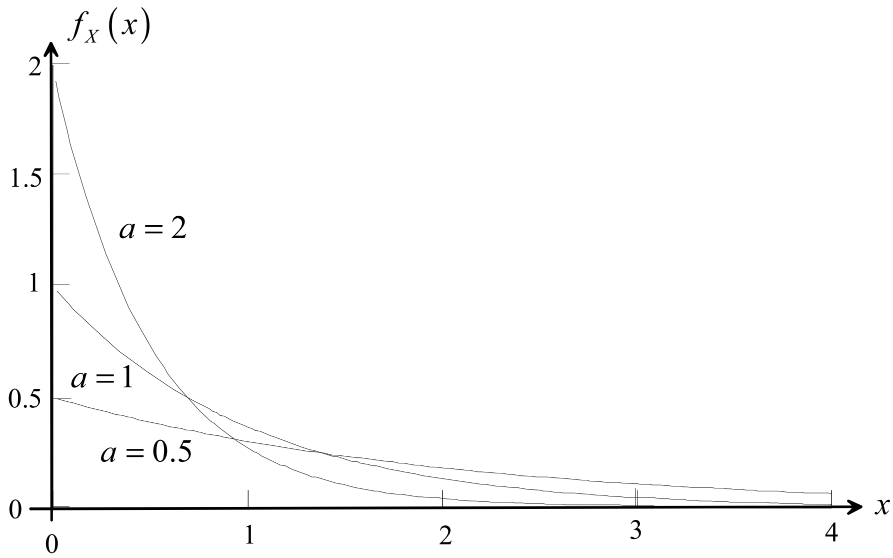
Verteilungsfunktion

$$F_X(x) = P(X \leq x) = \int_0^x \alpha\, e^{-\alpha t} dt = \begin{cases} 0 & x < 0 \\ 1 - e^{-\alpha x} & x \geq 0 \end{cases} \tag{7.8}$$

Satz 7.2

Mittelwert, Varianz, momenterzeugende Funktion (sh. S. 74)

$$E(X) = \frac{1}{\alpha}$$
$$V(X) = \frac{1}{\alpha^2} \tag{7.9}$$
$$m_X(t) = \frac{\alpha}{\alpha - t}$$



Die Exponentialverteilung wird u.a. als Modell zur Beschreibung von Zeitspannen verwendet. Ein wichtiges Anwendungsgebiet ist die Warteschlangentheorie, bei der Zufallsvariablen wie

- Zeit zwischen dem Eintreffen von Kunden an einem Schalter
- Bedienungszeit an einem Schalter
- Lebensdauer von Verschleissteilen

zur Diskussion stehen.

Der bevorzugte Einsatz der Exponentialverteilung beruht nicht nur auf ihrer einfachen mathematischen Handhabbarkeit; sie besitzt auch andere wünschbare Eigenschaften.

Satz 7.3

Ist X exponentialverteilt, so gilt

$$P(X > t + s \mid X > t) = P(X > s) \tag{7.10}$$

Der Beweis erfolgt unmittelbar aus der Verteilungsfunktion. Bezeichnet X z.B. die Lebensdauer eines Verschleisselementes, so besagt der obige Satz, dass die bedingte Verteilung der Lebensdauer unabhängig ist von der bereits erreichten Lebensdauer t. Diese Eigenschaft haben wir beim diskreten Analogon der geometrischen Verteilung als *Prozess ohne Gedächtnis* bezeichnet.

Satz 7.4

Sind die Voraussetzungen eines Poissonprozesses mit der Rate α erfüllt, so genügt die Zeit zwischen 2 aufeinanderfolgenden Ereignissen einer Exponentialverteilung mit dem Parameter α .

Es seien die Voraussetzungen eines Poissonprozesses erfüllt und

$X(t)$ = Anzahl Ereignisse in $(0,t)$

Dann ist $X(t)$ bekanntlich poissonverteilt mit dem Parameter αt . Wie gross ist die Wartezeit W bis zum ersten Eintreten von E?

Es gilt

$$\begin{aligned} P(W \le t) &= 1 - P(W > t) = 1 - P(X(t) = 0) \\ &= 1 - e^{-\alpha t} = F_X(t) \end{aligned} \tag{7.11}$$

Beachte: $P(X(t) = 0)$ bezeichnet die Wahrscheinlichkeit, dass im Intervall $[0,t]$ kein Ereignis eintritt. Dies ist aber gerade die Verteilungsfunktion einer exponential verteilten Zufallsvariablen.

Beispiel

Es sei bekannt, dass das Auftreten eines bestimmten Defektes die Voraussetzungen eines Poissonprozesses erfülle. Im Mittel rechnet man mit 3 Defekten pro Stunde. Mit welcher Wahrscheinlichkeit vergeht zwischen 2 Defekten mindestens 1/2 Std.?

X = Zeit zwischen 2 Defekten

X exponentialverteilt mit dem Parameter $\alpha t\,(\alpha = 3)$

$$P\left(X \ge \frac{1}{2}\right) = 1 - P\left(X \le \frac{1}{2}\right) = 1 - \left(1 - e^{-3\cdot\frac{1}{2}}\right) = e^{-1.5} = 0.2231 \tag{7.12}$$

$E(X) = \dfrac{1}{\alpha} = \dfrac{1}{3}$, d.h. im Mittel treten die Defekte im Abstand von 20 Minuten auf. (Rate = 3 Defekte/Stunde)

In statistischen Softwarepaketen (z.B. SYSTAT) sind oft nur Spezialfälle von Wahrscheinlichkeits- und Dichtefunktionen verfügbar. Für die Exponentialverteilung ist beispielsweise nur die Variante $\alpha = 1$ verfügbar. Die dazugehörige Zufallsvariable Y besitzt die Dichte

$$f_Y(y;1) = e^{-y} \qquad y \ge 0 \tag{7.13}$$

Die aus Y abgeleitete Zufallsvariable

$$X = g(Y) = \frac{1}{\alpha} Y \tag{7.14}$$

besitzt dann die (allgemeine) Dichte

$$f_X(x,\alpha) = \alpha\, e^{-\alpha x} \tag{7.15}$$

Der Beweis folgt sofort aus der allgemeinen Regel über Funktionen stetiger Variablen.

$$\begin{aligned} &X = g(Y) = \frac{1}{\alpha} Y; \qquad Y = g^{-1}(X) = \alpha X \\ &\frac{dg^{-1}(x)}{dx} = \alpha \\ &f_X(x) = e^{-\alpha x} \cdot \alpha = f_X(x;\alpha) \end{aligned} \tag{7.16}$$

Allgemein gilt

$$\begin{aligned} f_X(x;\alpha) &= \alpha\, f_Y(\alpha x;1) \\ f_Y(y;1) &= \frac{1}{\alpha} f_X\left(\frac{y}{\alpha};\alpha\right) \end{aligned} \tag{7.17}$$

7.3 Normalverteilung

Die Normalverteilung gehört zu den wichtigsten Verteilungen. Diese Tatsache wird einerseits durch ihre *mathematischen Eigenschaften* begründet und anderseits unterstützt durch das breite Feld von *Anwendungsmöglichkeiten* in fast allen Wissenschaftsbereichen.

Die Anfänge der Normalverteilung gehen auf de Moivre (1733) mit seinen Untersuchungen über das asymptotische Verhalten einer Binomialverteilung zurück. Laplace (1774) und Gauss (1809) haben die Untersuchungen vor allem im Zusammenhang mit der *Theorie über Messfehler* vorangetrieben.

Im Zuge weiterer Forschung wurde gezeigt, dass eine Vielzahl von stetigen und diskreten Verteilungen unter bestimmten Voraussetzungen durch eine Normalverteilung *approximiert* werden können.

Schliesslich sind nach dem *zentralen Grenzwertsatz* auch Summen von beliebig verteilten Zufallsvariablen unter recht allgemeinen Bedingungen approximativ normalverteilt.

Def. 7.3

Eine Zufallsvariable X mit der Dichte

$$f_X(x) = \frac{1}{a\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-b}{a}\right)^2} \qquad \begin{array}{c} -\infty < x < +\infty \\ -\infty < b < +\infty \\ a > 0 \end{array} \tag{7.18}$$

heisst normalverteilt.

Satz 7.5

Mittelwert, Varianz, momenterzeugende Funktion

$$\begin{aligned} E(X) &= b = \mu \\ V(X) &= a^2 = \sigma^2 \\ m_X(t) &= e^{\mu t + \sigma^2 \frac{t^2}{2}} \end{aligned} \tag{7.19}$$

Die Dichte $f_X(\cdot)$ wird bestimmt durch die beiden Parameter Erwartungswert und Varianz. Um dies deutlicher zum Ausdruck zu bringen, verwendet man für f_X die Darstellung

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{7.20}$$

<u>Bezeichnung</u>

Ist X normalverteilt mit den Parametern μ und σ, so schreiben wir

$$X \sim N(\mu,\sigma) \tag{7.21}$$

Eigenschaften der Dichtefunktionen

- f_X erfüllt die Voraussetzungen einer Dichte, nämlich

$$\begin{aligned} &f_X(x) \geq 0 \\ &\int_{-\infty}^{+\infty} f_X(x)\, dx = 1 \qquad \text{(ohne Beweis)} \end{aligned} \tag{7.22}$$

- f_X ist axialsymmetrisch bezüglich einer Ordinate in $x = \mu$, d.h

$$f_X(\mu + x) = f_X(\mu - x) \qquad \forall\ x \tag{7.23}$$

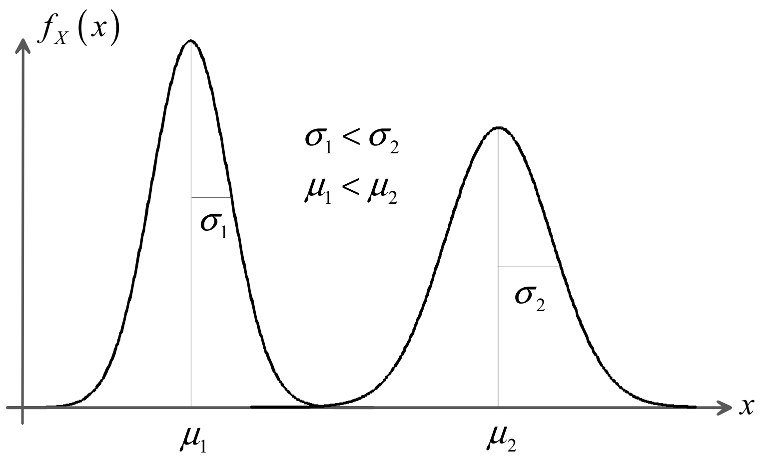
- f_X besitzt an der Stelle $x = \mu$ ein Maximum; d.h.

$$f_X(\mu) = \frac{1}{\sigma\sqrt{2\pi}} \tag{7.24}$$

- f_X besitzt Wendepunkte an den Stellen $x = \mu \pm \sigma$

- Ist σ klein (gross), so besitzt f_X ein hohes (niedriges) Maximum und die Wendepunkte liegen nahe (entfernt) von $x = \mu$.

- μ ist ein Lage- und σ ein Formparameter. Der speziellen Gestalt des Graphen wegen nennt man die durch f_X erzeugte Kurve auch *Glockenkurve*.

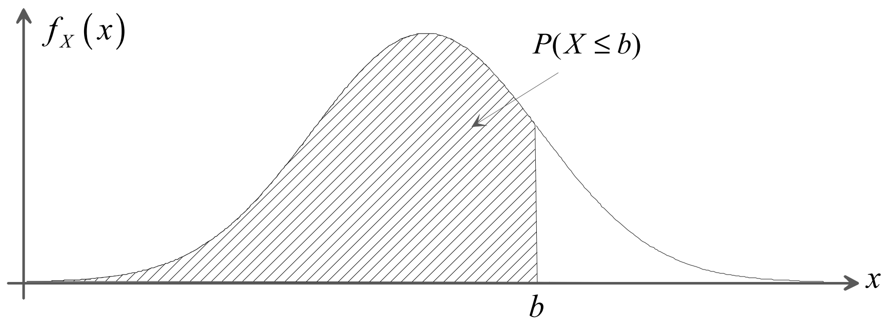


Die Funktion

$$F_X(x) = P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \tag{7.25}$$

heisst *Verteilungsfunktion* einer normalverteilten Zufallsvariablen mit den Parametern μ und σ.

Es lässt sich zeigen, dass das obige Integral geschlossen nicht lösbar ist. Man kann es jedoch durch numerische Integration beliebig genau approximieren.



Satz 7.6

Ist $X \sim N(\mu, \sigma)$ und $Y = c_1 X + c_2 \quad (c_1 \ne 0)$, so gilt

$$Y \sim N(c_1\mu + c_2, |c_1|\sigma) \tag{7.26}$$

Lineare Transformationen normalverteilter Zufallsvariablen sind wiederum normalverteilt.

<u>Beweis</u>

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P\left(X \le \frac{y-c_2}{c_1}\right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\frac{y-c_2}{c_1}} e^{-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2} du \end{aligned} \tag{7.27}$$

Mit

$$u = \frac{t - c_2}{c_1} \qquad du = \frac{1}{|c_1|}\,dt \tag{7.28}$$

folgt

$$\begin{aligned} F_Y(y) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left(-\frac{1}{2}\left(\frac{\frac{t-c_2}{c_1} - \mu}{\sigma} \right)^2 \right) \frac{1}{|c_1|}\,dt \\ &= \frac{1}{|c_1|\,\sigma\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left(-\frac{1}{2}\left(\frac{t-(c_1\mu + c_2)}{c_1\,\sigma} \right)^2 \right) dt \end{aligned} \tag{7.29}$$

Dies ist jedoch die Verteilungsfunktion einer Normalverteilung mit den Parametern $\mu^* = c_1\mu + c_2$ und $\sigma^{*2} = c_1^2\sigma^2$.

<u>Beispiele</u>
$X \sim N(20, 2)$

$Y = 2X$	$\sim$	$N(40, 4)$
$Y = -X$	$\sim$	$N(-20, 2)$
$Y = 2X - 10$	$\sim$	$N(30, 4)$

<u>Standardisierung</u>

$$X \sim N(\mu,\sigma) \;\Rightarrow\; Z = \frac{X-\mu}{\sigma} \sim N(0,1) \tag{7.30}$$

Dichte und Verteilungsfunktion standardisierter normalverteilter Variablen werden speziell bezeichnet.

$$\begin{aligned} \varphi(z) &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \\ \Phi(z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\,dt \end{aligned} \tag{7.31}$$

Zur Wahrscheinlichkeitsberechnung normalverteilter Zufallsvariablen ist es nicht notwendig, in jedem Einzelfall eine spezielle numerische Integration durchzuführen.

Nach Satz 7.6 stehen alle Normalverteilungen so zueinander in Beziehung, dass die Kenntnis einer einzigen Verteilungsfunktion genügt, um Wahrscheinlichkeiten beliebiger Normalverteilungen zu bestimmen.

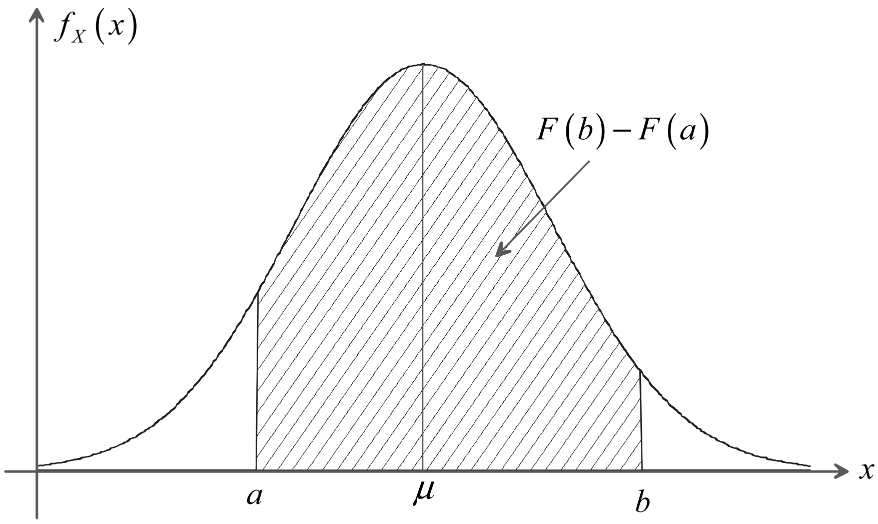
Für die standardisierte Verteilungsfunktion $\Phi(z)$ existieren umfangreiche Tabellenwerke. Diese können benützt werden zur Berechnung normaler Wahrscheinlichkeiten. Nach Satz 7.6 gilt nämlich

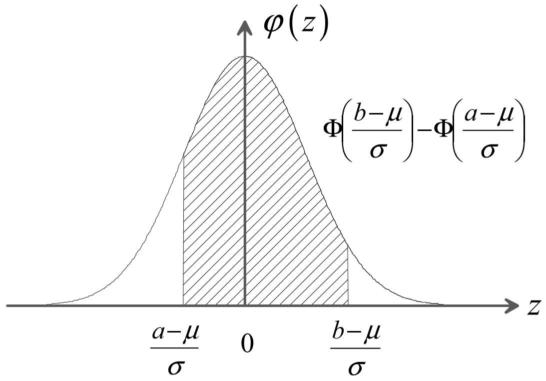
$$X \sim N(\mu,\sigma)$$
$$F_X(x) = P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{7.32}$$

und es gilt

Satz 7.7

$$X \sim N(\mu,\sigma)$$
$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \tag{7.33}$$

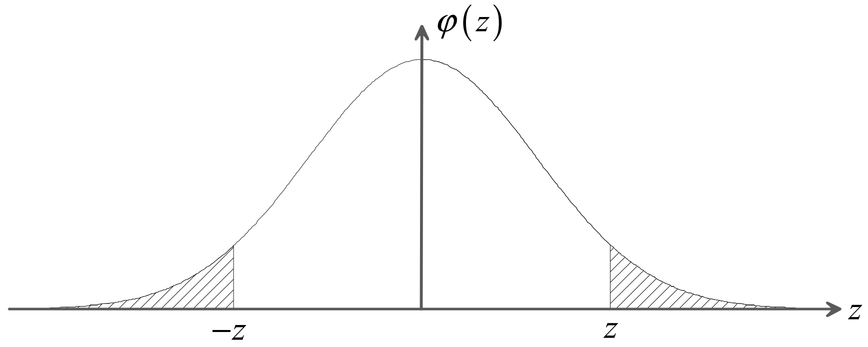




Wegen der Symmetrie der Normalverteilung gilt ferner

$$\begin{aligned}\Phi(-z) &= P(Z \le -z) = P(Z \ge z) = 1 - P(Z \le z) \\ &= 1 - \Phi(z)\end{aligned} \tag{7.34}$$

$\Phi(-z) = 1 - \Phi(z)$



Beispiel

Das Abfüllgewicht eines bestimmten Produkts sei normalverteilt mit dem Mittelwert 1000 und der Varianz 225.

$$X \sim N(1000{,}15) \tag{7.35}$$

Man berechne für eine zufällig ausgewählte Packung die Wahrscheinlichkeit, dass sie

1.) höchstens 980 gr. schwer ist

$$P(X \le 980) = \Phi\left(\frac{980-1000}{15}\right) = \Phi\left(-\frac{4}{3}\right) = 0.0918 \tag{7.36}$$

2.) mindestens 1010 gr. schwer ist

$$P(X \geq 1010) = 1 - P(X \leq 1010) = 1 - \Phi\left(\frac{1010-1000}{15}\right) = 0.2514 \qquad (7.37)$$

3.) zwischen 985 und 1010 gr. schwer ist

$$\begin{aligned} P(985 \leq X \leq 1010) &= \Phi\left(\frac{1010-1000}{15}\right) - \Phi\left(\frac{985-1000}{15}\right) \\ &= 0.7486 - 0.1587 = 0.5899 \end{aligned} \qquad (7.38)$$

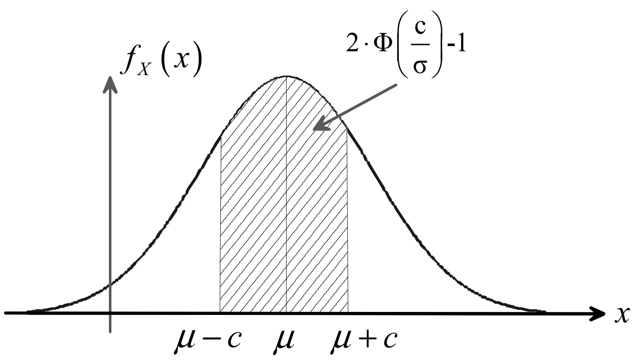
4.) um höchstens 10 gr. vom Mittelwert 1000 abweicht.

$$\begin{aligned} &P(1000-10 \leq X \leq 1000+10) = \\ &\Phi\left(\frac{1000+10-1000}{15}\right) - \Phi\left(\frac{1000-10-1000}{15}\right) = 0.4972 \end{aligned} \qquad (7.39)$$

Als Verallgemeinerung von Aufgabe 4.) bestätigt man sofort den

Satz 7. 8

$$P(\mu - c \leq X \leq \mu + c) = 2\Phi\left(\frac{c}{\sigma}\right) - 1 \qquad (7.40)$$



Aus dem obigen Satz folgen die

σ - Regeln:

$$P\left(\mu-k\sigma \le X \le \mu+k\sigma\right)=\begin{cases} 0.6827 & \text{falls} \quad k=1 \\ 0.9545 & \text{falls} \quad k=2 \\ 0.9973 & \text{falls} \quad k=3 \end{cases} \tag{7.41}$$

Die Standardabweichung besitzt demnach im Falle normalverteilter Zufallsvariablen eine intuitive Bedeutung, indem sie jenen zu μ symmetrischen Bereich absteckt, dem ca. die Wahrscheinlichkeit 2/3 zukommt.

<u>Beispiel</u>

Sei $X \sim N(10,\sigma)$ $\qquad$ σ unbekannt.

Wie gross darf σ höchstens werden, damit $P(X \ge 15)$ 5% nicht übersteigt?

$$P\left(X \ge 15\right)=1-P\left(X \le 15\right)=1-\Phi\left(\frac{15-10}{\sigma}\right)\le 0.05 \;\Rightarrow\; \Phi\left(\frac{5}{\sigma}\right)\ge 0.95 \tag{7.42}$$

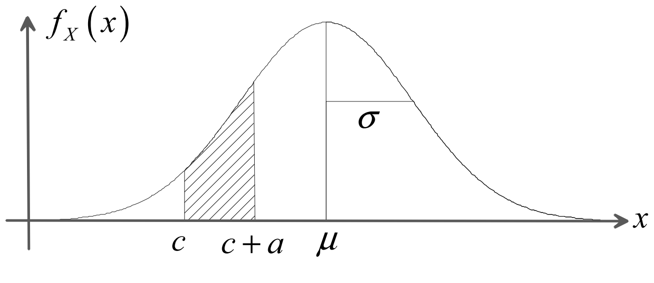
Gesucht ist jenes Argument von Φ, sodass gilt $\Phi(z)\ge 0.95$. Auch für diese inverse Fragestellung existieren Tabellen. Man findet

$$\frac{5}{\sigma}\ge 1.645 \qquad\qquad \sigma \le \frac{5}{1.645}\approx 3.04 \tag{7.43}$$

<u>Beispiel</u>

Sei $X \sim N(\mu,\sigma)$. Für welchen Wert von c gilt

$P(c \le X \le c+a) \rightarrow$ max ? $\quad$ ($a > 0$, konst.)



$$P(c \le X \le c+a) = \Phi\left(\frac{c+a-\mu}{\sigma}\right) - \Phi\left(\frac{c-\mu}{\sigma}\right) = g(c) \tag{7.44}$$

$$\begin{aligned}
&g'(c) = \varphi\left(\frac{c+a-\mu}{\sigma}\right)\cdot\frac{1}{\sigma} - \varphi\left(\frac{c-\mu}{\sigma}\right)\cdot\frac{1}{\sigma} = 0 \\
&\varphi\left(\frac{c+a-\mu}{\sigma}\right) = \varphi\left(\frac{c-\mu}{\sigma}\right) \qquad \text{falls} \\
&\frac{c+a-\mu}{\sigma} = -\frac{c-\mu}{\sigma} \qquad (\text{Symmetrie von } \varphi) \\
&c = \mu - \frac{a}{2}
\end{aligned} \tag{7.45}$$

7.4 Die Lognormalverteilung

Eine Zufallsvariable X heisst lognormalverteilt, falls der natürliche Logarithmus von X normalverteilt ist.

Sei $Y \sim N(\mu_Y, \sigma_Y)$ und $Y = \ln X$ resp. $X = e^Y$, dann genügt X einer Lognormalverteilung. X ist beschränkt auf positive Werte und im Gegensatz zur (zugeordneten) Normalverteilung nicht mehr symmetrisch. Die Hauptanwendungsgebiete liegen in der Ökonomie (Finance), Biologie und Physik.

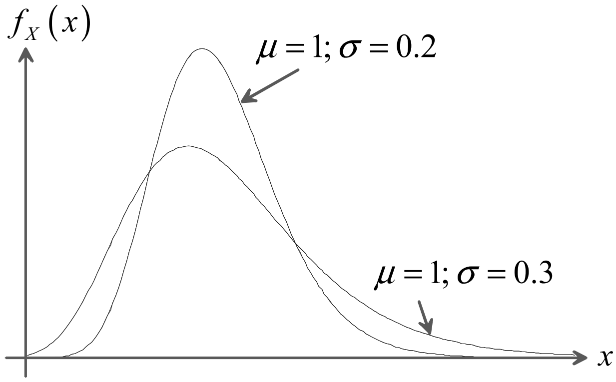
Def. 7.4

X heisst lognormalverteilt, falls ihre Dichte folgende Form besitzt

$$\begin{gathered}
f_X(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln x - \mu_Y}{\sigma_Y}\right)^2} \\
x > 0 \\
\sigma_Y > 0 \\
-\infty < \mu_Y < +\infty
\end{gathered} \tag{7.46}$$

Die Lognormalverteilung ist eine zweiparametrige Verteilung. μ_Y und σ_Y sind relative Lage- und Dispersionsmasse. Die Form der Verteilung (Schiefe) kann durch die Wahl von σ beeinflusst werden.

Notation $\qquad X \sim \ln(\mu_Y, \sigma_Y)$



Für die Momente von X gilt

$$\begin{aligned} \mu_X &= E(X) = e^{\left(\mu_Y + \frac{\sigma_Y^2}{2}\right)} \\ \sigma_X^2 &= V(X) = e^{2\left(\mu_Y + \sigma_Y^2\right)} - e^{2\mu_Y + \sigma_Y^2} \end{aligned} \tag{7.47}$$

Median = e^{μ_Y}

Modus = $e^{\mu_Y - \sigma_Y^2}$

$$\begin{aligned} \mu_Y &= \ln\left(\sqrt{\frac{\mu_X^4}{\mu_X^2 + \sigma_X^2}}\right) \\ \sigma_Y &= \sqrt{\ln\left(\frac{\mu_X^2 + \sigma_X^2}{\mu_X^2}\right)} \end{aligned} \tag{7.48}$$

Der Zusammenhang zwischen der Lognormal- und der damit verbundenen Normalverteilung folgt aus

$$F_X(x) = P(X \le x) = P(\ln X \le \ln x) = P(Y \le \ln x)$$
$$= F_Y(\ln x) \qquad \text{mit } Y \sim N(\mu_Y, \sigma_Y) \tag{7.49}$$

$$f_X(x) = \frac{dF_X(x)}{dx} = \frac{dF_Y(y)}{dy} \cdot \frac{dy}{dx} = \frac{1}{x\sigma_Y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln x - \mu_Y}{\sigma_Y}\right)^2} \tag{7.50}$$

mit $\quad y = \ln x$

Die Berechnung lognormalverteilter Wahrscheinlichkeiten erfolgt mit Hilfe der zugeordneten Normalverteilung.

$$P(a \le X \le b) = P(a \le e^Y \le b) = P(\ln a \le Y \le \ln b)$$
$$= \Phi\left(\frac{\ln b - \mu_Y}{\sigma_Y}\right) - \Phi\left(\frac{\ln a - \mu_Y}{\sigma_Y}\right) \tag{7.51}$$

<u>Beispiel</u>

X sei lognormalverteilt mit $\mu_X = 5$ und $\sigma_X^2 = 4$. Gesucht ist $P(X \le 4.5)$.

$$\mu_Y = \ln\left(\sqrt{\frac{5^4}{5^2+4}}\right) = 1.535$$
$$\sigma_Y = \sqrt{\ln\left(\frac{5^2+4}{5^2}\right)} = 0.385 \tag{7.52}$$

$$P(X \le 4.5) = \Phi\left(\frac{\ln 4.5 - 1.535}{0.385}\right) = \Phi(-0.080) = 0.4681 \tag{7.53}$$

Die Lognormalverteilung spielt eine wichtige Rolle im Zusammenhang mit der Verteilung von (stetigen) Wertpapier-Renditen. Die Notation erfolgte in Anlehnung an Hull (1993).

Bezeichnet S_t den Preis eines Wertpapiers zum Zeitpunkt t, so folgt

$$G = \ln S_t \tag{7.54}$$

einem allgemeinen Wienerprozess, welcher besagt, dass Änderungen von G zwischen den Zeitpunkten t und T einer Normalverteilung mit dem Mittelwert $\left(\mu - \frac{\sigma^2}{2}\right)(T-t)$ (drift rate) und der zeitproportionalen Varianz $\sigma^2\,(T-t)$ genügen.

$$\ln S_T - \ln S_t \sim \mathrm{N}\left(\left(\mu - \frac{\sigma^2}{2}\right)(T-t),\, \sigma\sqrt{T-t}\right) \tag{7.55}$$

resp.

$$\ln S_T \sim N\left(\ln S_t + \left(\mu - \frac{\sigma^2}{2}\right)(T-t),\, \sigma\sqrt{T-t}\right) \tag{7.56}$$

Aus der letzten Beziehung folgt sofort, dass der Preis des Wertpapiers S_T einer Lognormalverteilung genügt.

Unterstellt man eine stetige Rendite R (Rate of Return), so kann deren Verteilung im Falle lognormalverteilter Preise abgeleitet werden. Mit der Bezeichnung

$$S_T = S_t\, e^{R(T-t)} \tag{7.57}$$

gilt

$$\frac{S_T}{S_t} = e^{R(T-t)} \tag{7.58}$$

resp.

$$\ln \frac{S_T}{S_t} = R(T-t) \sim N\left(\left(\mu - \frac{\sigma^2}{2}\right)(T-t),\, \sigma\sqrt{(T-t)}\right) \tag{7.59}$$

Für die Verteilung von R folgt daraus

$$R \sim N\left(\mu - \frac{\sigma^2}{2}, \frac{\sigma}{\sqrt{T-t}}\right) \tag{7.60}$$

Stetige Renditen sind unter den gemachten Annahmen normalverteilt.

Beispiel
Eine Aktie besitze eine Jahresrendite von 8.5% bei einer Standardabweichung (Volatilität) von 6%. Die Verteilung der stetigen Rendite nach 2 Jahren ist normal mit dem Mittelwert

$$0.085-\frac{0.0036}{2}=0.0832 \tag{7.61}$$

und der Standardabweichung

$$\frac{0.06}{\sqrt{2}}=0.0424 \tag{7.62}$$

$$R \sim N(0.0832, 0.0424)$$

Das bekannte Verteilungsgesetz von R gestattet wahrscheinlichkeitstheoretische Aussagen, etwa dass R mit einer Wahrscheinlichkeit von 95% um höchstens 2 Standardabweichungen (0.0848) vom Mittelwert (0.0832) abweicht.

Zur Bestimmung des Erwartungswertes betrachten wir zunächst den praktisch wichtigen Spezialfall eines sog. gestutzten Erwartungswertes, nämlich

$$E(X|X>c) \tag{7.63}$$

Fragen dieser Art treten etwa bei der Herleitung der Black-Scholes Formel über die risikoneutrale Bewertung auf.

Aus der Theorie der Funktionen von Zufallsvariablen leitet sich die Beziehung zwischen der Lognormalverteilung von X und der korrespondierenden Normalverteilung von

$$Y=g(X)=\ln X \quad \text{resp.} \quad X=g^{-1}(Y)=e^{Y} \tag{7.64}$$

ab, nämlich

$$f_Y(y)=f_X(e^y)e^y=\frac{1}{\sigma_Y\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2} \tag{7.65}$$

Für den gestutzten Mittelwert folgt dann

$$
\begin{aligned}
E(X \mid X > c) &= E\left(e^Y \mid e^Y > c\right) = E\left(e^Y \mid Y > \ln c\right) \\
&= \frac{1}{\sigma_Y \sqrt{2\pi}} \int_{\ln c}^{\infty} e^y \, e^{-\frac{1}{2\sigma_Y^2}(y-\mu_Y)^2} dy
\end{aligned}
\tag{7.66}
$$

und nach erfolgter quadratischer Ergänzung des Exponenten

$$
\begin{aligned}
E(X|X > c) &= e^{\mu_Y + \frac{\sigma_Y^2}{2}} \left(1 - \Phi\left(\frac{\ln c - \left(\mu_Y + \sigma_Y^2\right)}{\sigma_Y}\right)\right) \\
&= e^{\mu_Y + \frac{\sigma_Y^2}{2}} \, \Phi\left(\frac{\mu_Y - \ln c}{\sigma_Y} + \sigma_Y\right)
\end{aligned}
\tag{7.67}
$$

Für den ungestutzten Erwartungswert folgt sofort

$$
E(X) = \lim_{c \to 0} E(X|X > c) = e^{\mu_Y + \frac{\sigma_Y^2}{2}}
\tag{7.68}
$$

Wie später noch darzustellen sein wird, liegt das Hauptanwendungsgebiet lognormaler Verteilungen bei Situationen, bei denen eine bestimmte Grösse unabhängigen zufälligen Einflüssen unterliegt, deren Wirkungen jeweils proportional zum aktuellen Bestand sind.

7.5 Pareto-Verteilung

Def. 7.5

Eine stetige Zufallsvariable X heisst paretoverteilt, falls ihre Dichte die Form

$$
f_X(x) = \begin{cases} \dfrac{\alpha}{x_0}\left(\dfrac{x_0}{x}\right)^{\alpha+1} & \text{falls } x > x_0 \\ 0 & \text{falls } x \le x_0 \end{cases}
\tag{7.69}
$$

$$
x_0 > 0\,, \quad \alpha > 0
$$

besitzt.

Für die Verteilungsfunktion erhält man

$$F_X(x) = P\left(X \le x\right) = \int_{x_0}^{x} \frac{\alpha}{x_0}\left(\frac{x_0}{t}\right)^{\alpha+1} dt = -\left(\frac{x_0}{t}\right)^{\alpha} \Big|_{x_0}^{x}$$
$$= 1-\left(\frac{x_0}{x}\right)^{\alpha} \qquad x > x_0 \tag{7.70}$$

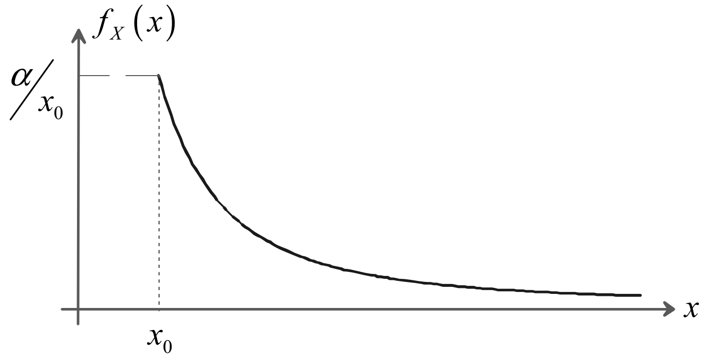
Satz 7.9

Mittelwert, Varianz, Median

$$E\left(X\right) = \frac{\alpha}{x_0}\int_{x_0}^{\infty} x\left(\frac{x_0}{x}\right)^{\alpha+1} dx = -\alpha\, x_0^{\alpha}\, \frac{x^{-\alpha+1}}{\alpha-1} \Big|_{x_0}^{\infty}$$
$$= \frac{\alpha}{\alpha-1}\, x_0\,, \quad \text{falls } \alpha > 1 \tag{7.71}$$

$$V\left(X\right) = \sigma^2 = \frac{\alpha}{\left(\alpha-1\right)^2\left(\alpha-2\right)}\, x_0^2\,, \quad \text{falls } \alpha > 2 \tag{7.72}$$

$$Me = x_0 \cdot 2^{\frac{1}{\alpha}}$$



Beachte: Die momenterzeugende Funktion existiert nicht.

Die Paretoverteilung beschreibt die Verteilung von Zufallsvariablen, die nur Werte über einem (positiven) Schwellenwert x_0 annehmen können. Anwendungsgebiete finden sich u.a. in der Soziologie und in der Ökonomie; z.B. bei der Verteilung von Einkommen und Vermögen, der Grösse von Städten, usw.

7.6 Gammafunktion

Def. 7.6

Die Gammafunktion $\Gamma(x)$ wird definiert durch

$$\Gamma(x) = \int_0^\infty t^{x-1}\, e^{-t}\, dt\,,\ \ x > 0 \tag{7.73}$$

und es gelten folgende Beziehungen

$$\begin{aligned} &1.)\ \ \Gamma(1) = 1 \\ &2.)\ \ \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \\ &3.)\ \ \Gamma(x+1) = x\,\Gamma(x) \end{aligned} \tag{7.74}$$

1.) folgt unmittelbar aus der Definition und 3.) ergibt sich durch partielle Integration. Aus 3.) folgt für ganzzahlige Werte von $x > 0$

$$\Gamma(x) = (x-1)! \tag{7.75}$$

<u>Beweis von 2.)</u>

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty t^{-\frac{1}{2}}\, e^{-t}\, dt \tag{7.76}$$

Mit $t = \dfrac{u^2}{2}$ und $dt = u\, du$ erhält man

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{\sqrt{2}}{u} e^{-\frac{1}{2}u^2} u\, du = \sqrt{2}\int_0^\infty e^{-\frac{1}{2}u^2}\, du = \sqrt{\pi} \tag{7.77}$$

mit

$$\int_0^\infty e^{-\frac{1}{2}u^2}\, du = \frac{\sqrt{2\pi}}{2} \tag{7.78}$$

Für $0 < x < 1$ ist die Gammafunktion ausführlich tabelliert. Für Computerapplikationen existieren Approximationspolynome (vgl. Hastings, Approximations for digital computers, Princeton N.J., 1955).

7.7 Betafunktion

Def. 7.7

Die Betafunktion $B(x,y)$ wird definiert als

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt \qquad x, y > 0 \tag{7.79}$$

Die Beta- und Gammafunktion sind miteinander verknüpft. Es gilt der

Satz 7.10

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{7.80}$$

7.8 Die Chi-quadratverteilung

Def. 7.8

Es seien $Z_1, Z_2 \ldots, Z_n$ unabhängige standardnormalverteilte Zufallsvariablen. Die Summe ihrer Quadrate

$$X = Z_1^2 + Z_2^2 + \ldots. + Z_n^2 \tag{7.81}$$

bezeichnet man als chi-quadratverteilte Zufallsvariable mit *n* Freiheitsgraden. Notation: $X \sim \chi^2(n)$

Satz 7.11

X sei eine chi-quadratverteilte Zufallsvariable mit *n* Freiheitsgraden. Ihre Dichte ist dann

$$f_X(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x \geq 0 \\ 0 & \text{sonst} \end{cases} \tag{7.82}$$

$$\begin{aligned} E(X) &= n \\ V(X) &= 2n \\ m_X(t) &= \left(\frac{1}{1-2t}\right)^{\frac{n}{2}}, \quad \text{für} \quad t < \frac{1}{2} \end{aligned} \tag{7.83}$$

Wir beweisen, dass $f_X(x)$ die Dichte einer Chi-quadratverteilung ist für den Fall $n = 1$.

Gegeben: $Z \sim N(0,1)$

Gesucht: Dichte von $X = Z^2$

Es gilt

$$
\begin{aligned}
F_X(x) &= P(X \le x) = P\left(Z^2 \le x\right) = P\left(-\sqrt{x} \le Z \le +\sqrt{x}\right) \\
&= 2\,\frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{1}{2}t^2}\, dt
\end{aligned}
\tag{7.84}
$$

Wir setzen $t = \sqrt{u}$ resp. $dt = \dfrac{1}{2\sqrt{u}}$ und erhalten

$$
F_X(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{u}{2}}\, \frac{1}{2\sqrt{u}}\, du \tag{7.85}
$$

woraus unter Benützung von $f_X(x) = F_X'(x)$ folgt

$$
f_X(x) = F_X'(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x}{2}}\, x^{-\frac{1}{2}} \qquad \text{q.e.d.} \tag{7.86}
$$

<u>Beweis für die momenterzeugende Funktion</u>

Wir benutzen die Tatsache, dass die momenterzeugende Funktion einer Summe unabhängiger Zufallsvariablen gleich dem Produkt der momenterzeugenden Funktionen der Summanden ist und ermitteln deshalb zuerst $m_{Z^2}(t)$ wobei $Z \sim N(0,1)$

$$
\begin{aligned}
m_{Z^2}(t) = E\left(e^{tz^2}\right) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tz^2}\, e^{-\frac{1}{2}z^2}\, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2(1-2t)}\, dz \\
&= \frac{1}{\sqrt{1-2t}}\, \frac{\sqrt{1-2t}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2(1-2t)}\, dz \\
&= \frac{1}{\sqrt{1-2t}}, \quad t < \frac{1}{2} \qquad\qquad q.e.d.
\end{aligned}
\tag{7.87}
$$

Daraus erhält man

$$m_X(t) = m_{Z_1^2+\ldots+Z_n^2}(t) = \left(\frac{1}{\sqrt{1-2t}}\right)^n = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}} \quad q.e.d.$$

$$m'_X(t) = n(1-2t)^{-\frac{n}{2}-1} \tag{7.88}$$

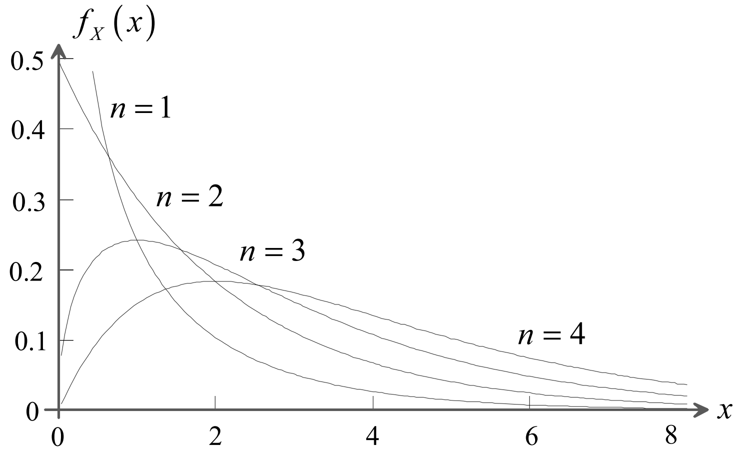
$$m''_X(t) = 2n\left(\frac{n}{2}+1\right)(1-2t)^{-\frac{n}{2}-2}$$

Für den Mittelwert und die Varianz von X folgt daraus

$$E(X) = m'(0) = n$$

$$V(X) = m''(0) - (m'(0))^2 = 2n\left(\frac{n}{2}+1\right) - n^2 = 2n \tag{7.89}$$

<u>Dichten chi-quadratverteilter Zufallsvariablen</u>



Für grosse Werte von n lässt sich die Chi-quadratverteilung durch die Normalverteilung *approximieren*. Für die Zufallsvariable $X \sim \chi^2(n)$

$$P(X \le x) = F_X(x) \approx \Phi\left(\sqrt{2x} - \sqrt{2n-1}\right) \tag{7.90}$$

Die Chi-quadratverteilung spielt vor allem im Zusammenhang mit Stichprobenfunktionen eine wichtige Rolle. Es gilt der

Satz 7.12

$X_1, X_2, \ldots, X_n$ seien *n* unabhängige identisch normalverteilte Zufallsvariablen mit

$$\begin{aligned} E(X_i) &= \mu \\ & \qquad\qquad i = 1, 2, \ldots., n \\ V(X_i) &= \sigma^2 \end{aligned} \tag{7.91}$$

Dann ist

1.)

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right) = \frac{\bar{X}-\mu}{\sigma} \sim N\left(0, \frac{1}{\sqrt{n}}\right) \tag{7.92}$$

2.) (ohne Beweis)

$$\begin{aligned} \bar{Z} = \frac{\bar{X}-\mu}{\sigma} \quad &\text{und} \quad \sum\left(Z_i - \bar{Z}\right)^2 \quad \text{stochastisch unabhängig} \\ \bar{X} \quad &\text{und} \quad \sum\left(X_i - \bar{X}\right)^2 \quad \text{stochastisch unabhängig} \end{aligned} \tag{7.93}$$

3.)

$$\sum_{i=1}^{n}\left(Z_i - \bar{Z}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1) \tag{7.94}$$

Definiert man z.B. die sog. Stichprobenvarianz

$$S^2 = \frac{1}{n-1}\sum\left(X_i - \bar{X}\right)^2 \tag{7.95}$$

so folgt aus 3.) unmittelbar, dass

$$V = \frac{(n-1)S^2}{\sigma^2} \tag{7.96}$$

chi-quadratverteilt ist mit (*n*-1) Freiheitsgraden.

7.9 Die Gammaverteilung

Def. 7.9

Eine stetige Zufallsvariable mit der Dichte

$$f_X(x) = \begin{cases} \frac{\lambda}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & \text{sonst} \end{cases} \tag{7.97}$$

$(\lambda > 0, \alpha > 0)$ heisst gammaverteilt mit den Parametern λ und α .

Satz 7.13

Mittelwert, Varianz, momenterzeugende Funktion

$$\begin{aligned} E(X) &= \frac{\alpha}{\lambda} \\ V(X) &= \frac{\alpha}{\lambda^2} \\ m_X(t) &= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}, \quad t < \lambda \end{aligned} \tag{7.98}$$

Wir führen den Beweis nur für die momenterzeugende Funktion. Mittelwert und Varianz folgen nach den bekannten Sätzen aus $m_X(t)$.

$$\begin{aligned} m_X(t) = E\left(e^{tX}\right) &= \int_0^{\infty} e^{tx} \frac{\lambda}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} e^{-\lambda x} \, dx \\ &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-x(\lambda-t)} dx \end{aligned} \tag{7.99}$$

mit $x(\lambda - t) = u$ und $dx = \frac{1}{\lambda - t} du$ folgt

$$
\begin{aligned}
m_X(t) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{u}{\lambda - t} \right)^{\alpha-1} e^{-u} \frac{1}{\lambda - t} du \\
&= \left(\frac{\lambda}{\lambda - t} \right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^\infty u^{\alpha-1} e^{-u} du = \left(\frac{\lambda}{\lambda - t} \right)^\alpha
\end{aligned}
\tag{7.100}
$$

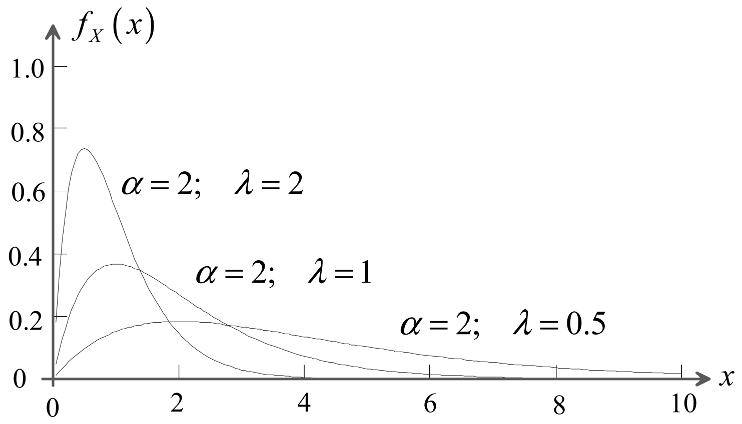
Spezialfälle der Gammaverteilung

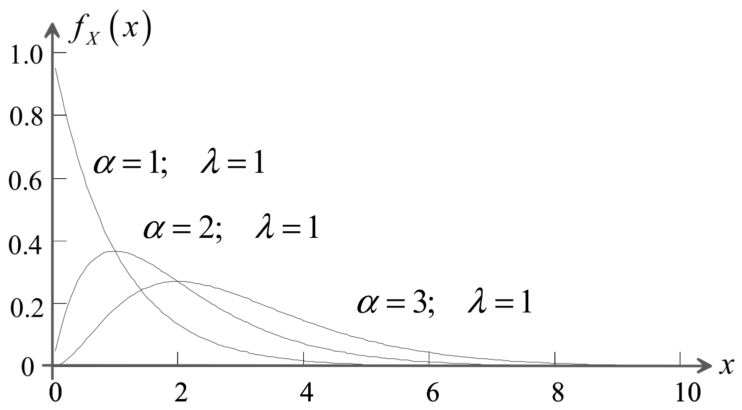
$\alpha = \frac{n}{2}$; $\lambda = \frac{1}{2}$: Gammaverteilung → $\chi^2(n)$

$\alpha = 1$: Gammaverteilung → Exponentialverteilung

Besitzt X eine Gammaverteilung und ist $\alpha > 0$ ganzzahlig, so findet man durch fortgesetzte partielle Integration

$$
F_X(x) = P(X \le x) = 1 - \sum_{j=0}^{\alpha-1} \frac{(\lambda x)^j}{j!} e^{-\lambda x} \tag{7.101}
$$





7.10 Die t -Verteilung (Student-Verteilung)

Def. 7.10

Sind $Z \sim N(0,1), Y \sim \chi^2(n)$ sowie Z und Y unabhängig voneinander, so ist der Quotient

$$X = \frac{Z}{\sqrt{\frac{Y}{n}}} \tag{7.102}$$

eine stetige Zufallsvariable mit der Dichte

$$f_X(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \tag{7.103}$$

mit

$$\begin{aligned} \mu &= E(X) = 0 && n > 1 \\ \sigma^2 &= V(X) = \frac{n}{n-2} && n > 2 \end{aligned} \tag{7.104}$$

Eine Zufallsvariable mit der obigen Dichte wird als *t-verteilte Zufallsvariable mit n Freiheitsgraden* bezeichnet.

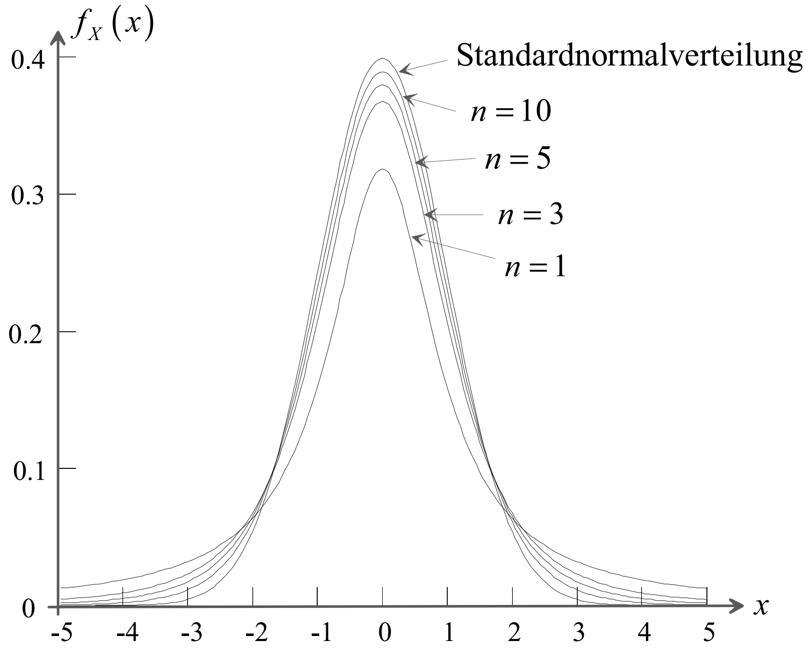
Beachte: Die momenterzeugende Funktion existiert nicht.

Spezialfall

Falls $n = 1$, nennt man die t-Verteilung *Cauchy-Verteilung*, für deren Dichte gilt

$$f_X(x) = \frac{1}{\pi}\left(1 + x^2\right)^{-1} \qquad -\infty < x < +\infty \tag{7.105}$$

Mittelwert und Varianz einer Cauchy-Verteilung existieren nicht.



Wie die obige Graphik nahelegt, konvergiert die *t*-Verteilung mit wachsendem *n* gegen die standardisierte Normalverteilung. Der wesentliche Unterschied besteht darin, dass die *t*-Verteilung den Rändern ein grösseres Wahrscheinlichkeitsmass zuordnet als die Normalverteilung.

Satz 7.14

$$\lim_{n\to\infty} \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{7.106}$$

<u>Beweis</u>

Wir untersuchen zuerst

$$\frac{1}{\sqrt{\pi}}\lim_{n\to\infty}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} = \frac{1}{\sqrt{\pi}}\lim_{m\to\infty}\left(\left(1+\frac{\frac{x^2}{2}}{m-\frac{1}{2}}\right)^m\right)^{-1} \quad \text{mit} \quad \frac{n+1}{2}=m$$

$$= \frac{1}{\sqrt{\pi}}\, e^{-\frac{x^2}{2}} \tag{7.107}$$

Es verbleibt der Faktor

$$\frac{1}{\sqrt{n}}\,\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = g(n) \tag{7.108}$$

und es gilt

$$g(n+1) = \frac{1}{\sqrt{n+1}}\,\frac{\Gamma\left(\frac{n+1+1}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} = \frac{1}{\sqrt{n+1}}\,\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n+1}{2}\right)} \tag{7.109}$$

Ferner gilt

$$\Gamma\left(\frac{n}{2}+1\right) = \frac{n}{2}\,\Gamma\left(\frac{n}{2}\right) \tag{7.110}$$

und

$$\Gamma\left(\frac{n+1}{2}\right) = \sqrt{n}\;\Gamma\left(\frac{n}{2}\right) g(n) \tag{7.111}$$

woraus folgt

$$g(n+1) = \frac{\frac{n}{2}\,\Gamma\left(\frac{n}{2}\right)}{\sqrt{n+1}\,\sqrt{n}\;\Gamma\left(\frac{n}{2}\right) g(n)} = \frac{1}{2\sqrt{1+\frac{1}{n}}\; g(n)} \tag{7.112}$$

Existiert der Grenzwert

$$\lim_{n\to\infty} g(n) = g \tag{7.113}$$

$[g(n) \to g; g(n+1) \to g]$, so erhält man

$$\lim_{n\to\infty} g(n+1) = g = \lim_{n\to\infty} \frac{1}{2\sqrt{1+\frac{1}{n}}\, g(n)} = \frac{1}{2g} \tag{7.114}$$

woraus für g folgt

$$g = \frac{1}{\sqrt{2}} \tag{7.115}$$

Eingesetzt erhält man

$$\begin{aligned}
&\lim_{n\to\infty}\left(\frac{1}{\sqrt{n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{\pi}} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}\right) \\
&= \lim_{n\to\infty} \frac{1}{\sqrt{n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{\pi}} \lim_{n\to\infty}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \\
&= \frac{1}{\sqrt{2}} \frac{1}{\sqrt{\pi}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \text{q.e.d.}
\end{aligned} \tag{7.116}$$

Beispiel

$X_1, X_2, \ldots, X_n$ seien unabhängige identisch normalverteilte Zufallsvariablen mit

$$X_i \sim N(\mu, \sigma) \qquad i = 1,2,\ldots.,n \tag{7.117}$$

Ferner seien

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \tag{7.118}$$

Dann sind

$$Z = \frac{\bar{X} - \mu}{\sigma}\sqrt{n} \sim N(0,1)$$
$$Y \sim \chi^2 (n-1) \tag{7.119}$$
$$Z, Y \text{ stochastisch unabhängig}$$

Also ist die Zufallsvariable

$$T = \frac{Z}{\sqrt{\frac{Y}{n-1}}} = \frac{\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma}}{\sqrt{\frac{1}{n-1}\sum\left(\frac{X_i - \bar{X}}{\sigma}\right)^2}}$$
$$= \frac{\sqrt{n}\left(\bar{X}-\mu\right)}{\sqrt{\frac{1}{n-1}\sum\left(X_i - \bar{X}\right)^2}} = \sqrt{n}\,\frac{\bar{X}-\mu}{S} \tag{7.120}$$

t-verteilt mit (*n*-1) Freiheitsgraden.

Beachte: $T = \sqrt{n}\,\frac{\bar{X}-\mu}{S}$ ist nicht mehr abhängig von σ.

7.11 Die F - Verteilung

Def. 7.11

X_m^2 und X_n^2 seien zwei unabhängige chi-quadratverteilte Zufallsvariablen mit m resp. n Freiheitsgraden. Man nennt

$$X = \frac{\frac{X_m^2}{m}}{\frac{X_n^2}{n}} \tag{7.121}$$

eine sog. F-verteilte Zufallsvariable mit den Parametern m und n. Wir schreiben

$$X \sim F_{(m,n)} \tag{7.122}$$

Satz 7.15

X besitzt die Dichte

$$f_X(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} \left(\dfrac{m}{n}\right)^{\frac{m}{2}} x^{\frac{(m-2)}{2}} \left[1 + \dfrac{m}{n}x\right]^{-\frac{(m+n)}{2}} & x > 0 \\ 0 & sonst \end{cases} \tag{7.123}$$

$$\begin{aligned} E(X) &= \frac{n}{n-2} & n > 2 \\ V(X) &= \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} & n > 4 \end{aligned} \tag{7.124}$$

Besitzt X eine F-Verteilung mit den Parametern m und n, so besitzt $\frac{1}{X}$ ebenfalls eine F-Verteilung, jedoch mit den Parametern n und m. Dank dieser Eigenschaft kann die Tabellierung wesentlich vereinfacht werden. Es gilt nämlich

$$P\left(X_{(m,n)} \le x\right) = P\left(\frac{1}{X_{(m,n)}} \ge \frac{1}{x}\right) = P\left(X_{(n,m)} \ge \frac{1}{x}\right) = 1 - P\left(X_{(n,m)} \le \frac{1}{x}\right) \qquad (7.125)$$

Beispiel

Seien $X_1, X_2, \ldots, X_m$ und $Y_1, Y_2, \ldots, X_n$ $m + n$ unabhängige normalverteilte Zufallsvariablen

$$\begin{aligned} X_i &\sim N(\mu_X, \sigma) \qquad i = 1, 2, \ldots., m \\ Y_j &\sim N(\mu_Y, \sigma) \qquad j = 1, 2, \ldots., n \end{aligned} \qquad (7.126)$$

Dann sind

$$\begin{aligned} V_m &= \sum_{i=1}^{m} \left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \\ V_n &= \sum_{j=1}^{n} \left(\frac{Y_i - \overline{Y}}{\sigma}\right)^2 \end{aligned} \qquad (7.127)$$

unabhängige chi-quadratverteilte Zufallsvariablen mit (m-1) resp. (n-1) Freiheitsgraden. Also ist der Quotient

$$U = \frac{\sum_{i=1}^{m} \frac{\left(X_i - \overline{X}\right)^2}{(m-1)}}{\sum_{j=1}^{n} \frac{\left(Y_i - \overline{Y}\right)^2}{(n-1)}} \qquad (7.128)$$

F-verteilt mit den Parametern (m-1) und (n-1). Wie das Beispiel zeigt, treten F-verteilte Zufallsvariablen speziell im Zusammenhang mit *Quotienten von Varianzen* auf.

Querverbindung zwischen der F- und Binomialverteilung

Ist X F-verteilt mit den Parametern df_1 und df_2 und Y binomialverteilt mit den Parametern n und p, so gilt der

Satz 7.16

$$\begin{aligned} & P(Y \le y) = 1 - P\left(X < \frac{n-y}{y+1} \frac{p}{1-p} \right) \\ & \text{wobei} \quad Y \sim B(n, p) \\ & \qquad\quad X \sim F_{(df_1, df_2)} \\ & \text{mit} \qquad df_1 = 2(y+1) \\ & \qquad\quad df_2 = 2(n-y) \end{aligned} \tag{7.129}$$

Beispiel

$$\begin{aligned} & Y \sim B(10, 0.2) \\ & P(Y \le 2) = 1 - P\left(X < \frac{10-2}{2+1} \frac{0.2}{0.8} \right) \\ & = 1 - P\left(X < \frac{2}{3} \right) \qquad \text{wobei} \quad X \sim F_{(6,16)} \end{aligned} \tag{7.130}$$

Aus der Tabelle für die F-Verteilung erhält man

$$P(Y \le 2) = 1 - P\left(X < \frac{2}{3} \right) = 1 - 0.3222 = 0.6778 \tag{7.131}$$

7.12 Betaverteilung

Eine stetige Zufallsvariable X heisst über dem (offenen) Intervall (a,b) betaverteilt mit den Parametern $\alpha > 0$ und $\beta > 0$, falls sie die Dichte

$$f_X(x;\alpha,\beta) = \frac{(b-a)^{1-\alpha-\beta}}{B(\alpha,\beta)} (x-a)^{\alpha-1} (b-x)^{\beta-1} \, I_{(a,b)}(x) \tag{7.132}$$

besitzt.

Wir verwenden die Notation

$$X \sim B(\alpha, \beta) \tag{7.133}$$

Für die Parameter der Betaverteilung findet man

Modus:

$$M_0 = \frac{b(\alpha-1)+a(\beta-1)}{\alpha+\beta-2} \tag{7.134}$$

Mittelwert:

$$\mu = E(X) = a + (b-a)\,\frac{\alpha}{\alpha+\beta} \tag{7.135}$$

Varianz:

$$\sigma^2 = V(X) = \frac{(b-a)^2\ \alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{7.136}$$

Wählt man speziell $a = 0$ und $b = 1$, so gilt für die Dichte der dazugehörigen Zufallsvariablen Y

$$f_Y(y;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\, y^{\alpha-1}(1-y)^{\beta-1}\, I_{(0,1)}(y) \tag{7.137}$$

Zwischen X und Y gilt die Beziehung

$$\begin{aligned} X &= a+(b-a)Y \\ E(X) &= a+(b-a)\,E(Y) \\ V(X) &= (b-a)^2\, V(Y) \end{aligned} \tag{7.138}$$

Die Beweise für die Parameter von X können somit aus jenen für Y abgeleitet werden. Dabei gilt insbesondere

$$
\begin{aligned}
E\left(Y^k\right) &= \frac{1}{B(\alpha,\beta)} \int_0^1 y^k \; y^{\alpha-1} \, (1-y)^{\beta-1} \, dy \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{\alpha+k-1} \, (1-y)^{\beta-1} \, dy \\
&= \frac{B(\alpha+k,\beta)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha+k)\,\Gamma(\beta)}{\Gamma(\alpha+\beta+k)} \, \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \\
&= \frac{\Gamma(\alpha+k)\,\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\alpha+\beta+k)}
\end{aligned}
\tag{7.139}
$$

Für $k = 1$ folgt

$$
E(Y) = \frac{\Gamma(\alpha+1)\,\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\alpha+\beta+1)} = \frac{\alpha\,\Gamma(\alpha)\,\Gamma(\alpha+\beta)}{\Gamma(\alpha)(\alpha+\beta)\,\Gamma(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta} \tag{7.140}
$$

Analog verwendet man für die Varianz die Beziehung

$$
V(Y) = E\left(Y^2\right) - \left[E(Y)\right]^2 \tag{7.141}
$$

<u>Bemerkungen</u>

- Falls $\alpha = \beta = 1$, geht die Betaverteilung über in eine Gleichverteilung über dem Intervall (a,b).

- Für die Verteilungsfunktion

$$
F_Y(y;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \int_0^y t^{\alpha-1} (1-t)^{\beta-1} \, dt \tag{7.142}
$$

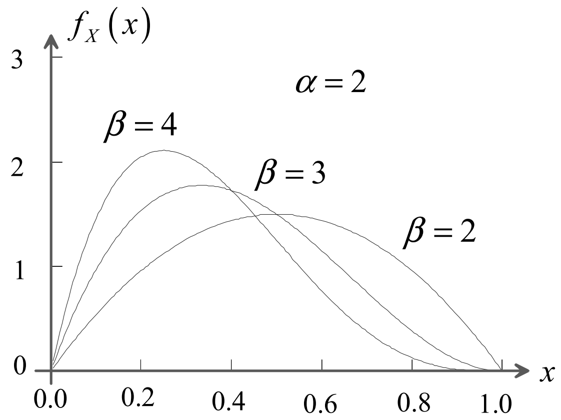
  existieren umfangreiche Tabellen und Computerprogramme. Für die Verteilung von X gilt,

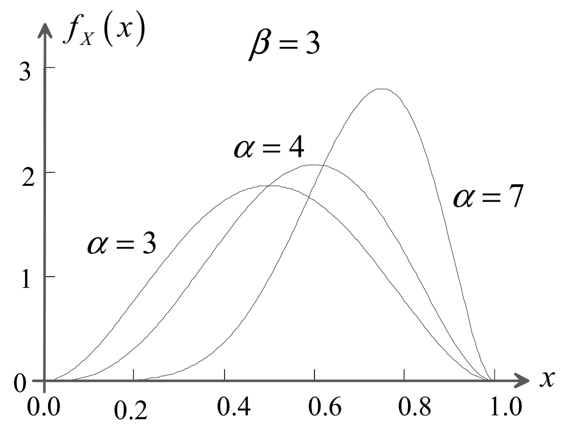
$$F_X(x;\alpha,\beta) = P(X \le x) = P(a + (b-a)Y \le x) = P\left(Y \le \frac{x-a}{b-a}\right)$$
$$= F_Y\left(\frac{x-a}{b-a};\alpha,\beta\right) \tag{7.143}$$

- Die Betaverteilung ist symmetrisch, falls $\alpha = \beta$; sie ist linksschief, falls $\beta < \alpha$ und rechtsschief, falls $\alpha < \beta$.

- In der Netzplantechnik verwendet man oft den Spezialfall $\alpha = 3 + \sqrt{2}$ und $\beta = 3 - \sqrt{2}$. Dann gilt

$$E(X) = \frac{1}{6}(a + 4\,Mo + b)$$
$$V(X) = \left(\frac{b-a}{6}\right)^2 \tag{7.144}$$



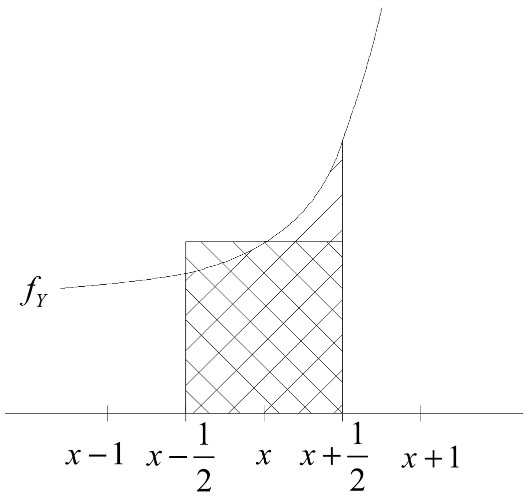


8. APPROXIMATION DISKRETER VERTEILUNGEN DURCH STETIGE VERTEILUNGEN

8.1 Kontinuitätskorrektur

Wir nehmen an, eine diskrete Verteilung sei auf der Menge der nichtnegativen ganzen Zahlen definiert (Binomial-, Poisson-, Hypergeometrische Verteilung). Wir haben schon darauf hingewiesen, dass man unter bestimmten Voraussetzungen diskrete Verteilungen durch stetige Verteilungen approximieren kann. Dabei ist aber zu berücksichtigen, dass diskrete Wahrscheinlichkeiten einzelnen Punkten, stetige Wahrscheinlichkeiten hingegen Intervallen zugeordnet werden.

Man führt deshalb folgende Kontinuitätskorrektur durch. X sei die diskrete und Y die stetige Zufallsvariable. $P(X = x)$ kann als Rechtecksfläche über dem Intervall $\left[x - \frac{1}{2}, x + \frac{1}{2}\right]$ aufgefasst werden. Diese Fläche wird dann approximiert durch die Fläche, welche die Dichtefunktion über demselben Intervall abgrenzt.



$$P(X = x) \approx P\left(x - \frac{1}{2} \le Y \le x + \frac{1}{2}\right) \tag{8.1}$$

8.2 Approximation der Binomialverteilung durch die Normalverteilung

X sei eine binomialverteilte Zufallsvariable mit den Parametern n und p

$$X \sim B(n,p)$$

$$P(X=x)=\binom{n}{x} p^x \cdot (1-p)^{n-x} \qquad x=0,1,2,....,n \tag{8.2}$$

$$\begin{aligned} E(X) &= np \\ V(X) &= npq \end{aligned} \tag{8.3}$$

Satz 8.1

Für grosse Werte von n ist $X \sim B(n,p)$ approximativ normalverteilt mit den Parametern

$$\mu = np \quad \text{und} \quad \sigma = \sqrt{npq} \tag{8.4}$$

$$\begin{aligned} P(a \le X \le b) &= \sum_{x=a}^{b} \binom{n}{x} p^x \, q^{n-x} \\ &\approx \Phi\left(\frac{b+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-\frac{1}{2}-np}{\sqrt{npq}}\right) \end{aligned} \tag{8.5}$$

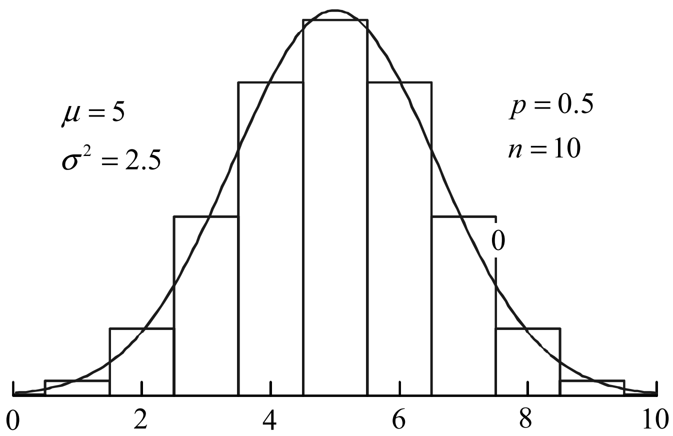
$$X \sim B(n,p) \quad \xrightarrow{n \text{ gross}} \quad N\left(np, \sqrt{npq}\right) \tag{8.6}$$

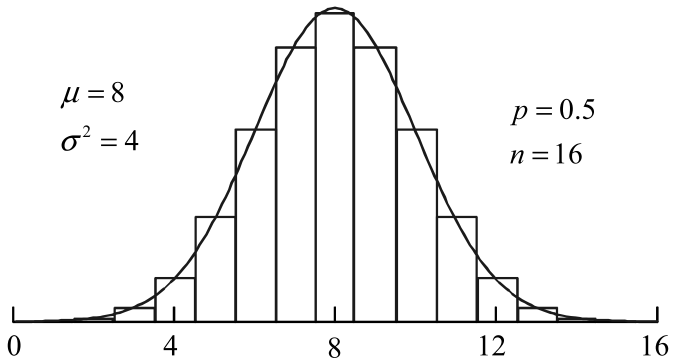
Als *Faustregel* für die Zulässigkeit der Approximation ist bekannt

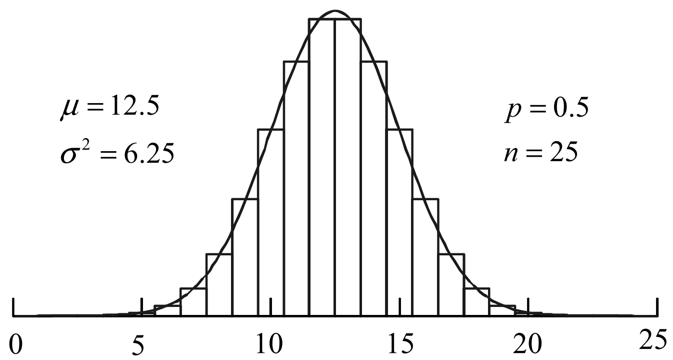
$$n > \frac{9}{pq} \tag{8.7}$$

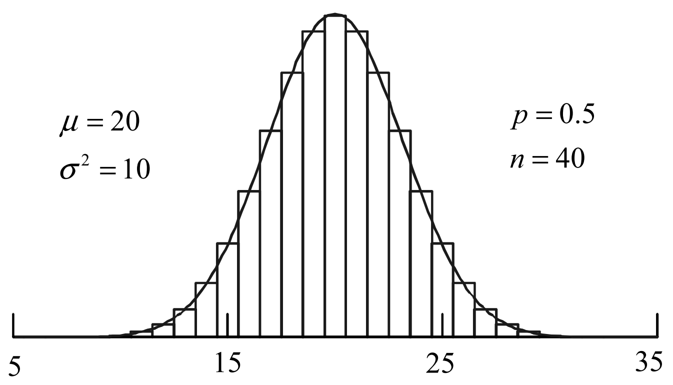
Binomialverteilung - Normalverteilung

$p = 0.50$



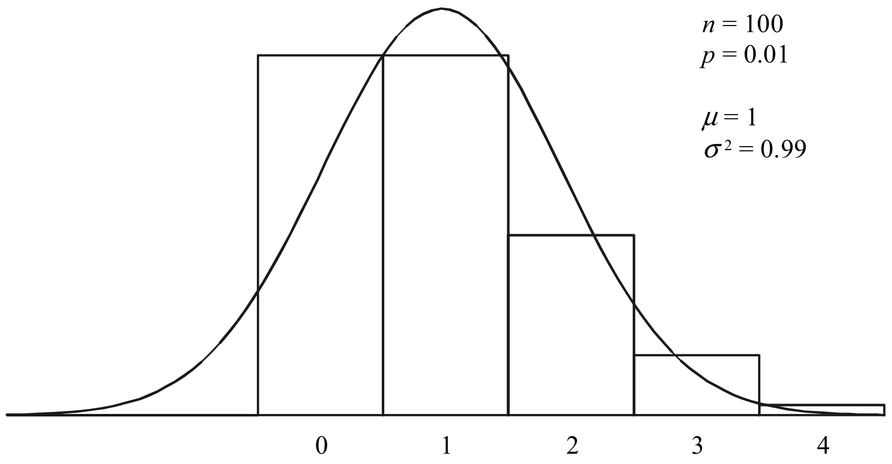


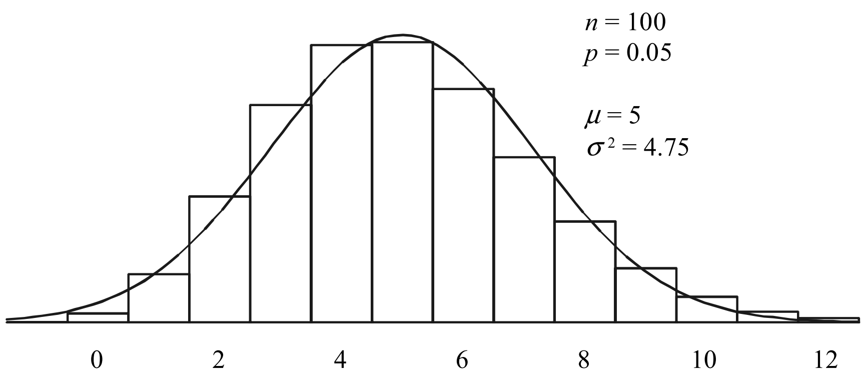


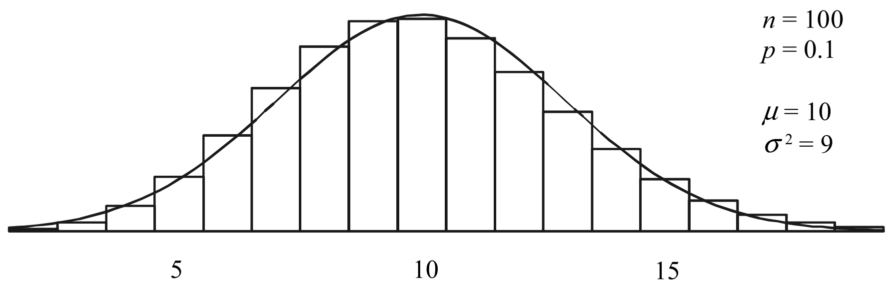


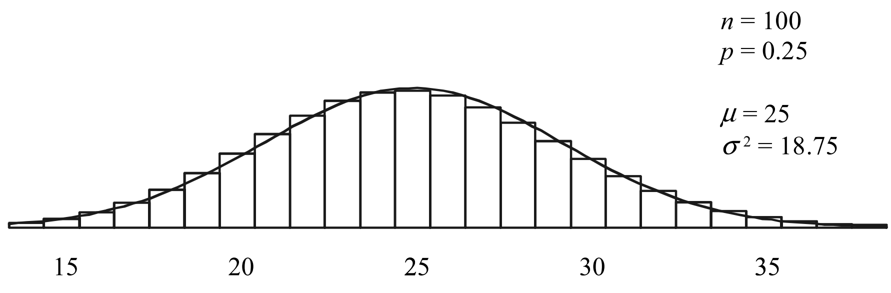
Binomialverteilung - Normalverteilung

$n = 100$









Beispiel

Eine Prüfung wird nach dem "multiple choice"-System durchgeführt. Es sind 100 Fragen zu beantworten. Bei jeder Frage stehen 3 Alternativen zur Verfügung, wobei eine davon richtig ist. Ein Student erscheint völlig unvorbereitet zur Prüfung und kreuzt rein zufällig je eine der Möglichkeiten an.

1.) Mit wieviel richtigen Antworten kann er im Mittel rechnen?
2.) Mit welcher Wahrscheinlichkeit realisiert er mehr als 25 richtige Antworten?
3.) Die Prüfung gilt als bestanden, wenn mindestens die Hälfte aller Fragen richtig beantwortet wurden. Mit welcher Wahrscheinlichkeit wird die Prüfung bestanden?

X = Anzahl richtige Antworten

$$\begin{aligned} &X \sim B\left(100, \frac{1}{3}\right) \\ &E(X) = 33\frac{1}{3} \\ &V(X) = 22.22 \\ &\sigma(X) = \sqrt{22.22} = 4.71 \end{aligned} \tag{8.8}$$

Die Verteilung von X kann approximiert werden durch eine Normalverteilung
$X \sim N(33.33, 4.71)$

1.)

$$E(X) = np = 33\frac{1}{3} \tag{8.9}$$

2.)

$$\begin{aligned} P(X > 25) &= 1 - P(X \leq 25) \approx 1 - \Phi\left(\frac{25 + 0.5 - 33.33}{4.71}\right) \\ &= 1 - \Phi(-1.66) = 1 - 0.485 = 0.9515 \end{aligned} \tag{8.10}$$

3.)

$$\begin{aligned} P(X \geq 50) &= 1 - P(X \leq 49) \approx 1 - \Phi\left(\frac{49 + 0.5 - 33.33}{4.71}\right) \\ &= 1 - \Phi(3.43) = 0 \end{aligned} \tag{8.11}$$

Beispiel
Im Kapitel über die Binomialverteilung haben wir die einen Random-Walk beschreibende Zufallsvariable X_n eingeführt.

$$P(X_n = x) = \binom{n}{\frac{x+n}{2}} p^{\frac{x+n}{2}} q^{\frac{n-x}{2}} \qquad x = -n, -n+2, \ldots, n-2, n \tag{8.12}$$

$$\begin{aligned} E(X_n) &= n(p-q) \\ V(X_n) &= 4\,npq \end{aligned} \tag{8.13}$$

Zur Bildung des Grenzüberganges nehmen wir an, dass die ursprüngliche Zeiteinheit verkürzt werde, wir bezeichnen sie mit γ. Während der Zeit t erfolgen mit der neuen Zeiteinheit $\frac{t}{\gamma}$ Bewegungen, jedoch immer noch nach dem Bernoullimuster. Wenn die Zeitintervalle zwischen zwei Bewegungen verkürzt werden, so ist auch die Einheitsstrecke zu verkürzen. Als neue Einheitsstrecke wählen wir δ. Bezeichnet Y_i wieder die im i-ten Schritt zurückgelegte Strecke, so gilt mit den neuen Einheiten

$$Y_i = \begin{cases} \delta & \text{mit Wahrscheinlichkeit } p \\ -\delta & \text{mit Wahrscheinlichkeit } q \end{cases} \tag{8.14}$$

und für die Positionen zum Zeitpunkt t

$$X_t = \sum_{i=1}^{\frac{t}{\gamma}} Y_i \tag{8.15}$$

Mit

$$\begin{aligned} E(Y_i) &= \delta(p-q) \\ V(Y_i) &= 4\delta^2\, pq \end{aligned} \tag{8.16}$$

folgt sofort

$$E(X_t) = \frac{t}{\gamma}\,\delta(p-q)$$
$$V(X_t) = \frac{t}{\gamma}\,4\delta^2\,pq \qquad (8.17)$$

Durch die Anpassung der Einheitsstrecke wurde erreicht, dass sich die Varianz proportional zu t verhält.

Sowohl beim Mittelwert als auch bei der Varianz tritt die arbiträre Schrittlänge δ explizit in Erscheinung. Durch eine geeignete Wahl von δ können wünschbare Eigenschaften insbesondere bei der Varianz erzeugt werden.

In wichtigen Anwendungsgebieten möchte man eine zur Zeit t proportionale Varianz von X_t. Dies erreicht man durch die Wahl

$$\delta = \sqrt{\gamma} \qquad (8.18)$$

und es folgt

$$E(X_t) = \frac{t}{\sqrt{\gamma}}(p-q)$$
$$V(X_t) = 4tpq \qquad (8.19)$$

Wird in dieser Notation γ verglichen mit t klein so, dass $\frac{t}{\gamma}$ genügend gross wird, so folgt für die (approximative) standardisierte Normalverteilung von X_t

$$Z_t = \frac{X_t - \frac{t}{\sqrt{\gamma}}(p-q)}{4tpq} \sim N(0,1) \qquad (8.20)$$

Für den wichtigen Spezialfall $p = q = \frac{1}{2}$ folgt schliesslich

$$
\begin{aligned}
&E(X_t) = 0 \\
&V(X_t) = t \qquad (8.21) \\
&X_t \sim N\left(0, \sqrt{t}\right)
\end{aligned}
$$

Die Position des Partikels zum Zeitpunkt t (vom Ausgangspunkt aus betrachtet) ist normalverteilt mit der Standardabweichung $\sqrt{t}$.

Der oben beschriebene Prozess über X_t heisst Brown'sche Bewegung oder Wiener Prozess. Praktische Anwendungen finden sich insbesondere in der Finanzmarkttheorie etwa bei der Modellierung von Kursreihen für Aktien.

8.3 Approximation der hypergeometrischen Verteilung durch die Normalverteilung

Ist X hypergeometrisch verteilt, so haben wir festgestellt (vgl. S. 97f.), dass unter den Bedingungen

$$n \ll N, K, N-K \qquad (8.22)$$

X approximativ binomialverteilt ist (Stichprobe ohne Zurücklegen - Stichprobe mit Zurücklegen). Wir erwarten deshalb, dass auch hypergeometrische Verteilungen unter den obigen Bedingungen für relativ grosse Werte von n gegen eine Normalverteilung streben. Die Bestätigung liefert der

Satz 8.2

X sei hypergeometrisch verteilt mit der Wahrscheinlichkeitsfunktion

$$f_X(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad (8.23)$$

und den Parametern

$$
\begin{aligned}
E(X) &= n\,\frac{K}{N} \\
V(X) &= n\,\frac{K}{N}\left(1-\frac{K}{N}\right)\frac{N-n}{N-1}
\end{aligned}
\tag{8.24}
$$

Dann ist X approximativ normalverteilt

$$
X \sim N\left(np, \sqrt{npq\,\frac{N-n}{N-1}}\right)
\tag{8.25}
$$

wobei $p = \dfrac{K}{N}$, $q = 1 - p$.

Es gilt

$$
\begin{aligned}
P(a \le X \le b) &= \sum_{x=a}^{b} \frac{\begin{pmatrix} K \\ x \end{pmatrix}\begin{pmatrix} N-K \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \\
&\approx \Phi\left(\frac{b+0.5-\mu_X}{\sigma_X}\right) - \Phi\left(\frac{a-0.5-\mu_X}{\sigma_X}\right)
\end{aligned}
\tag{8.26}
$$

wobei

$$
\begin{aligned}
\mu_X &= n\,\frac{K}{N} \\
\sigma_X &= \sqrt{n\,\frac{K}{N}\left(1-\frac{K}{N}\right)\frac{N-n}{N-1}}
\end{aligned}
\tag{8.27}
$$

Als Faustregel für die Zuverlässigkeit der Approximation erhält man (vgl. Approximation der Binomialverteilung)

$$\begin{aligned} n &> \frac{9}{p(1-p)}, \quad \frac{n}{N} < 0.05 \\ \frac{9}{p(1-p)} &< n < 0.05N \end{aligned} \tag{8.28}$$

<u>Beispiel</u>

Eine Packung enthalte 1000 Elemente, wobei bekannt ist, dass 20% davon defekt sind. Die Überprüfung der Elemente führt zu deren Zerstörung.

Der Packung werden 20 Elemente entnommen, wobei die Zufallsvariable X die Anzahl defekter Elemente in der Stichprobe beschreibt.

Für die hypergeometrische Verteilung von X gilt

$N = 1000$ $K = 200$ $n = 20$

$$\begin{aligned} E(X) &= 20 \cdot 0.2 = 4 \\ V(X) &= 20 \cdot 0.2 \cdot 0.8 \cdot \frac{980}{999} = 3.139 \end{aligned} \tag{8.29}$$

x	F-Hypergeo	F-Binomial	F-Normal
0	0.0110	0.0115	0.0241
1	0.0673	0.0692	0.0791
2	0.2033	0.2061	0.1986
3	0.4097	0.4114	0.3889
4	0.6301	0.6296	0.6111
5	0.8062	0.8042	0.8014
6	0.9154	0.9133	0.9209
7	0.9693	0.9679	0.9759
8	0.9907	0.9900	0.9945
9	0.9977	0.9974	0.9990
10	0.9995	0.9994	0.9999
11	0.9999	0.9999	1.0000
12	1.0000	1.0000	1.0000
	$N = 1000$ $K = 200$ $n = 20$	$n = 20$ $p = 0.20$	$\mu = 4$ $\sigma = 1.772$

8.4 Approximation der Poissonverteilung durch die Normalverteilung

Eine poissonverteilte Zufallsvariable X mit der Wahrscheinlichkeitsfunktion

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \tag{8.30}$$

kann für genügend grosse Werte von λ ebenfalls durch eine Normalverteilung approximiert werden. Es gilt der

Satz 8.3

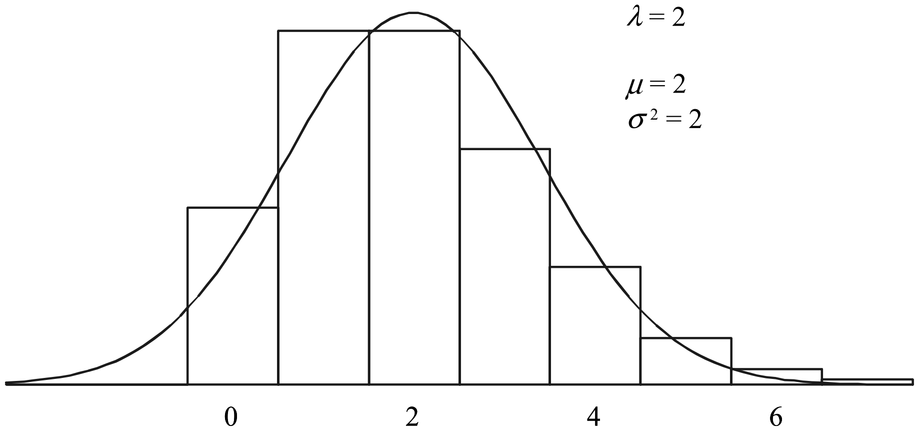
Ist X poissonverteilt mit dem Parameter λ, so ist X approximativ $N\left(\lambda, \sqrt{\lambda}\right)$ -verteilt.

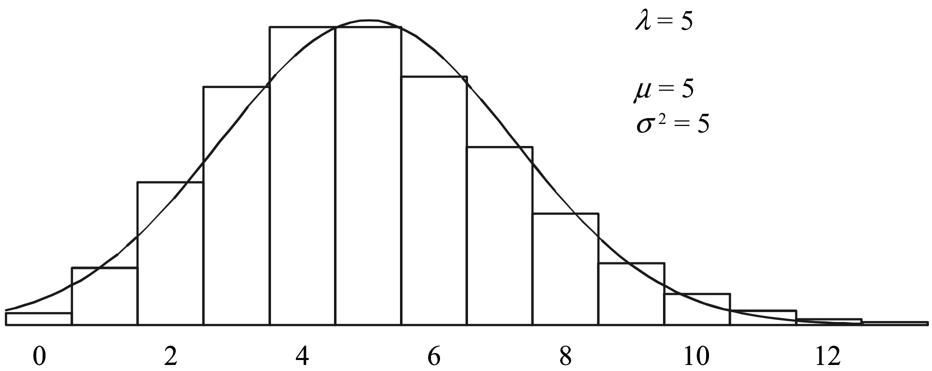
$$P(a \le X \le b) \approx \Phi\left(\frac{b + 0.5 - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - 0.5 - \lambda}{\sqrt{\lambda}}\right) \tag{8.31}$$

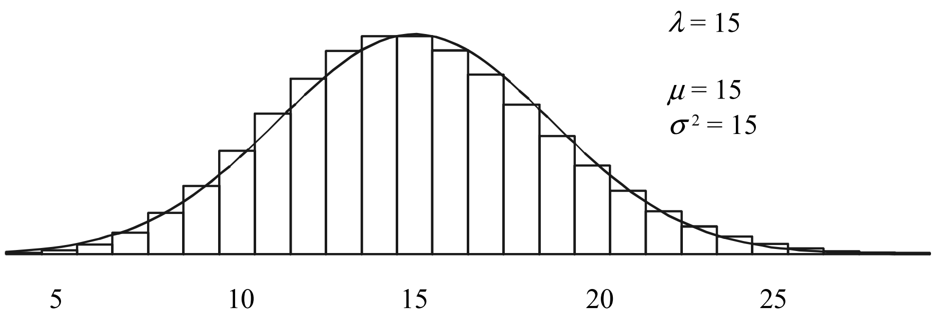
Als *Faustregel* für die Anwendbarkeit des obigen Satzes ist bekannt

$$\lambda \ge 12$$

Graphische Veranschaulichung







Beispiel
Innert Jahresfrist ereigneten sich insgesamt 10547 Strassenverkehrsunfälle auf dem Arbeitsweg. Wir nehmen an, dass die einzelnen Unfälle unabhängig (?) voneinander erfolgen und die Voraussetzungen eines Poissonprozesses erfüllt seien. Aus den obigen

Daten leiten wir eine Rate von 44 Unfällen pro Arbeitstag (240 Arbeitstage pro Jahr) ab.

Mit welcher Wahrscheinlichkeit ereignen sich pro Tag

1.) mehr als 56 Unfälle
2.) zwischen 40 und 50 Unfälle
3.) weniger als 35 oder mehr als 55 Unfälle?

X = Anzahl Unfälle pro Tag. Approximativ gilt

$$X \sim N\left(44, \sqrt{44}\right) \tag{8.32}$$

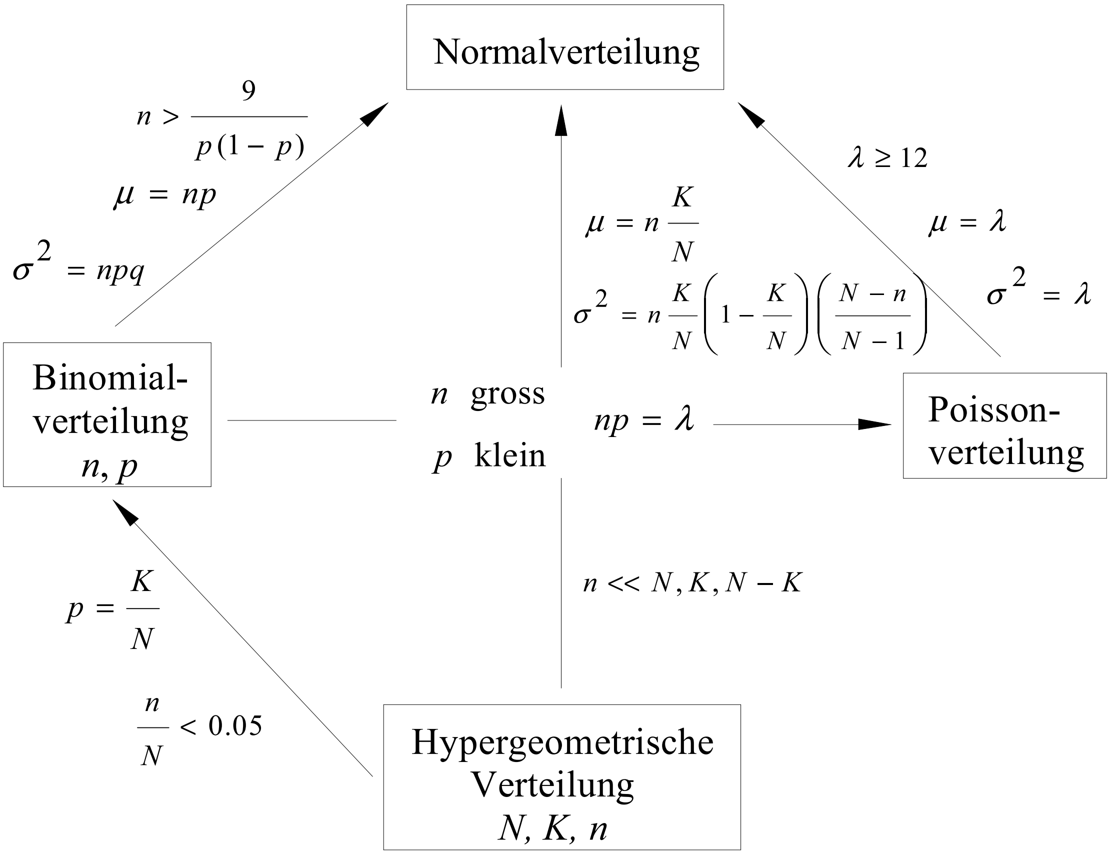
1.)
$$\begin{aligned} P(X > 56) &= 1 - P(X \le 56) \approx 1 - \Phi\left(\frac{56 + 0.5 - 44}{\sqrt{44}}\right) \\ &= 1 - \Phi(1.88) = 1 - 0.9699 = 0.0301 \;\; (0.0337) \end{aligned} \tag{8.33}$$

2.)
$$\begin{aligned} P(40 \le X \le 50) &\approx \Phi\left(\frac{50 + 0.5 - 44}{\sqrt{44}}\right) - \Phi\left(\frac{40 - 0.5 - 44}{\sqrt{44}}\right) \\ &= \Phi(0.98) - \Phi(-0.68) = 0.8365 - 0.2483 = 0.5882 \;\; (0.5838) \end{aligned} \tag{8.34}$$

3.)
$$\begin{aligned} P(X < 35 \text{ oder } X > 55) &= P(X < 35) + P(X > 55) \\ &\approx \Phi\left(\frac{35 - 0.5 - 44}{\sqrt{44}}\right) + 1 - \Phi\left(\frac{55 + 0.5 - 44}{\sqrt{44}}\right) \\ &= \Phi(-1.43) + 1 - \Phi(1.73) = 0.1182 \;\; (0.1174) \end{aligned} \tag{8.35}$$

In Klammern sind die “exakten” Poissonwahrscheinlichkeiten angegeben.

8.5 Schematische Zusammenfassung der Approximationsmöglichkeiten

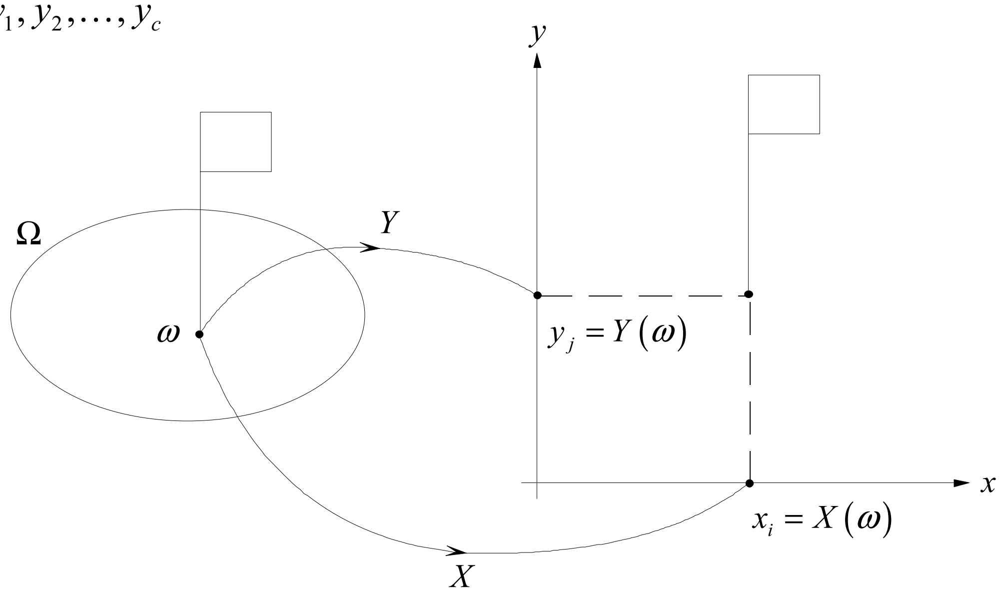


9. MEHRDIMENSIONALE ZUFALLSVARIABLEN

9.1 Diskrete zweidimensionale Zufallsvariablen

Wir betrachten mit dem Zufallsexperiment ε dessen Wahrscheinlichkeitsraum (Ω, P) und definieren darüber zwei Zufallsvariablen (Funktionen) X und Y. Die daraus entstehenden Bilder seien

$$X : x_1, x_2, \ldots, x_r$$
$$Y : y_1, y_2, \ldots, y_c$$



$$x_i = X(\omega); y_j = Y(\omega)$$

Analog zum eindimensionalen Fall wird auch jetzt mit den beiden Abbildungen das Wahrscheinlichkeitsmass von (Ω, P) übertragen, allerdings nicht mehr auf die reelle Achse sondern auf die (x,y)-Ebene. $P(\omega)$ wird dem Gitterpunkt (x_i, y_j) zugeordnet.

Das Paar (X, Y) heisst zweidimensionale Zufallsvariable mit der gemeinsamen Wahrscheinlichkeitsfunktion

$$P\left(X = x_i, Y = y_j\right) = f_{XY}(x_i, y_j) \tag{9.1}$$

wobei jedem Gitterpunkt (x_i, y_j) eindeutig das entsprechende Wahrscheinlichkeitsmass zugeordnet wird. Ausserhalb der Gitterpunkte gilt

$$f_{XY}(x, y) = 0 \quad \text{für} \quad (x, y) \neq (x_i, y_j) \tag{9.2}$$

Beispiel

Eine Urne enthält 20 rote, 30 weisse und 50 schwarze Kugeln. Es wird eine Zufallsstichprobe mit Zurücklegen vom Umfang $n = 3$ entnommen. Wir definieren die Zufallsvariablen

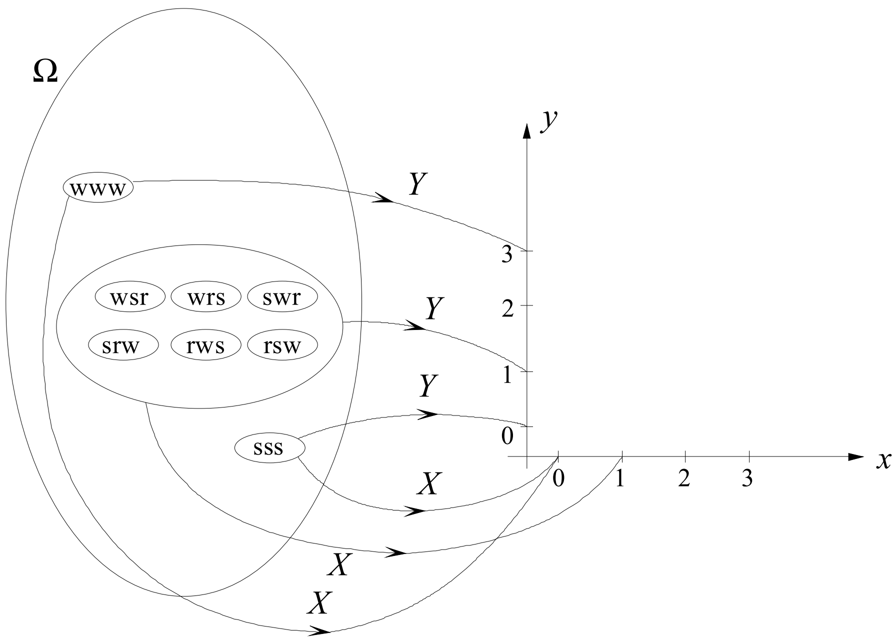
X = Anzahl rote (r) Kugeln in der Stichprobe

Y = Anzahl weisse (w) Kugeln in der Stichprobe

Wertebereiche

X: 0,1,2,3

Y: 0,1,2,3



Die gemeinsamen Wahrscheinlichkeiten werden auf die Gitterpunkte (x_i, y_j) in der Ebene übertragen. Sie lassen sich in einer zweidimensionalen Matrix darstellen. Für das Wertepaar X=1 und Y=2 erhält man z.B.

$$
\begin{aligned}
P(X=1, Y=2) &= P\big((rww)\cup(wrw)\cup(wwr)\big) \\
&= P(rww)+P(wrw)+P(wwr) \\
&= \begin{pmatrix} & 3 & \\ 1 & 2 & 0 \end{pmatrix}\cdot 0.2^{1}\cdot 0.3^{2}\cdot 0.5^{0} \\
&= \frac{3!}{1!2!0!}\cdot 0.2^{1}\cdot 0.3^{2}\cdot 0.5^{0} \\
&= 0.054
\end{aligned}
\tag{9.3}
$$

X \ Y	0	1	2	3
0	0.125	0.225	0.135	0.027
1	0.150	0.180	0.054	0
2	0.060	0.036	0	0
3	0.008	0	0	0

Für die *Verteilungsfunktion* F_{XY} der zweidimensionalen Zufallsvariablen (X,Y) gilt die

Def. 9.1

$$F_{XY}(x,y) = P(X \le x, Y \le y) = \sum_{x_i \le x} \sum_{y_j \le y} f_{XY}(x_i, y_j) \tag{9.4}$$

Beide Summationskriterien müssen gleichzeitig erfüllt sein.

Im obigen Beispiel gilt

$$
\begin{aligned}
F_{XY}(1,2.5) &= f_{XY}(0,0)+f_{XY}(0,1)+f_{XY}(0,2)+f_{XY}(1,0)+f_{XY}(1,1)+f_{XY}(1,2) \\
&= 0.869
\end{aligned}
\tag{9.5}
$$

Die Summe über alle gemeinsamen Wahrscheinlichkeiten ergibt wieder 1.

$$\sum_{i} \sum_{j} f_{XY}(x_i, y_j) = 1 \tag{9.6}$$

Offenbar sind mit jeder zweidimensionalen Verteilung zwei eindimensionale Verteilungen, nämlich jene von X und jene von Y verbunden. Man nennt diese eindimensionalen Verteilungen die sog. Randverteilungen von X resp. Y.

9.1.1 Randverteilungen

Wir gehen vom obigen Urnenbeispiel aus. Die Matrix der gemeinsamen Wahrscheinlichkeiten enthält die Wahrscheinlichkeiten $P\left(X = x_i \cap Y = y_j\right)$. Interessiert man sich für die Wahrscheinlichkeit des Ereignisses

$$A = \{X = x_i\,,\ Y \text{ beliebig}\}$$

so stellt man A als Vereinigung paarweise disjunkter Ereignisse dar.

$$A = \left\{\left(X = x_i \cap Y = y_1\right) \cup \left(X = x_i \cap Y = y_2\right) \cup ...\right\} \tag{9.7}$$

z.B. für ($X = 1$, Y beliebig)

$$\begin{aligned} A = \{&\left(X = 1, Y = 0\right) \cup \left(X = 1, Y = 1\right) \\ &\cup \left(X = 1, Y = 2\right) \cup \left(X = 1, Y = 3\right)\} \end{aligned} \tag{9.8}$$

Daraus erhält man

$$\begin{aligned} P\left(X = 1, Y \text{ beliebig}\right) &= f_{XY}\left(1,0\right) + f_{XY}\left(1,1\right) + f_{XY}\left(1,2\right) + f_{XY}\left(1,3\right) \\ &= 0.15 + 0.18 + 0.054 + 0 = 0.384 \end{aligned} \tag{9.9}$$

Die Randverteilungen von X resp. Y erhält man durch *Summation der gemeinsamen Wahrscheinlichkeiten* über die Spalten resp. Zeilen der gemeinsamen Wahrscheinlichkeitsmatrix.

$$\begin{aligned} f_X(x_i) &= \sum_j f_{XY}(x_i, y_j) \qquad \text{Randverteilung von } X \\ f_Y(y_j) &= \sum_i f_{XY}(x_i, y_j) \qquad \text{Randverteilung von } Y \end{aligned} \tag{9.10}$$

X \ Y	0	1	2	3	f_X
0	0.125	0.225	0.135	0.027	0.512
1	0.150	0.180	0.054	0	0.384
2	0.060	0.036	0	0	0.096
3	0.008	0	0	0	0.008
f_Y	0.343	0.441	0.189	0.027	1

In diesem Beispiel sind die Randvariablen binomialverteilt (warum?)

$$\begin{aligned} X &\sim B(3, 0.2) \\ Y &\sim B(3, 0.3) \end{aligned} \tag{9.11}$$

Beachte: Aus den gemeinsamen Verteilungen lassen sich stets die Randverteilungen ableiten. Die Umkehrung ist aber i.a. nicht möglich.

9.1.2 Bedingte Verteilungen

Praktisch bedeutsam sind Situationen, bei denen bekannt ist, in welcher Realisation z.B. Y vorliegt. Man interessiert sich dann für die Wahrscheinlichkeitsverteilungen der Zufallsvariablen X unter der Voraussetzung (Bedingung), dass die Ausprägung y_j von Y bekannt ist. In Analogie zur bedingten Wahrscheinlichkeit von Ereignissen nennt man den Ausdruck

$$P\left(X = x_i | Y = y_j\right) = \frac{P\left(X = x_i,\, Y = y_j\right)}{P\left(Y = y_j\right)} \tag{9.12}$$

bedingte Wahrscheinlichkeit, dass X den Wert x_i annimmt.

Im Urnenbeispiel gilt

$$P(X = 1 | Y = 0) = \frac{P(X = 1,\, Y = 0)}{P(Y = 0)} = \frac{0.15}{0.343} = 0.4373 \tag{9.13}$$

Def. 9.2

$$f_{X|Y}(x_i\,|\,y_j)=P\left(X=x_i\,|\,Y=y_j\right)=\frac{f_{XY}(x_i,y_j)}{f_Y(y_j)} \tag{9.14}$$

heisst bedingte Wahrscheinlichkeitsfunktion von X, gegeben $Y=y_j$.

Beispiel (Urnenmodell)
Bedingte Verteilung von X, gegeben $Y=2$.

x	0	1	2	3
$f_X\left(x\,\middle\|\,y=2\right)$	0.7143	0.2857	0	0

Bedingte Verteilung von Y, gegeben $X=1$.

x	0	1	2	3
$f_Y\left(y\,\middle\|\,x=1\right)$	0.3906	0.4688	0.1406	0

Die bedingten Wahrscheinlichkeiten erfüllen alle Bedingungen, die wir an Wahrscheinlichkeiten stellen. So gilt z.B. für festes y_j

$$\sum_i f_{X|Y}(x_i\,|\,y_j)=\sum_i\frac{f_{XY}(x_i,y_j)}{f_Y(y_j)}=\frac{1}{f_Y(y_j)}\sum_i f_{XY}(x_i,y_j)=\frac{1}{f_Y(y_j)}\,f_Y(y_j)=1 \tag{9.15}$$

9.1.3 Unabhängige Zufallsvariablen

Vergleicht man im Urnenbeispiel die Randverteilungen und die bedingten Verteilungen, so stellt man fest, dass sie voneinander abweichen. Für $X=1$ gilt z.B.

$$P(X=1)=0.384 \quad\neq\quad P(X=1\,|\,Y=0)=0.437 \tag{9.16}$$

Die Wahrscheinlichkeit für die Ausprägung von X hängt offenbar davon ab, welcher Wert von Y realisiert wurde. X und Y heissen in einem solchen Fall *abhängige Zufallsvariablen*.

Def. 9.3

Die Zufallsvariablen X und Y heissen *voneinander unabhängig*, falls für alle x_i, y_j gilt

$$f_{XY}(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j)$$
$$\text{resp.} \quad P\left(X = x_i, Y = y_j\right) = P\left(X = x_i\right) \cdot P\left(Y = y_j\right) \tag{9.17}$$

Beachte:

- Bei unabhängigen Zufallsvariablen stimmen bedingte und unbedingte Wahrscheinlichkeiten überein.

$$f_{X|Y}(x_i \mid y_j) = \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)} = \frac{f_X(x_i) f_Y(y_j)}{f_Y(y_j)} = f_X(x_i) \tag{9.18}$$

- Bei unabhängigen Zufallsvariablen lassen sich die gemeinsamen Wahrscheinlichkeiten aus den Randverteilungen ableiten.

- Bei unabhängigen Zufallsvariablen stimmen die bedingten Verteilungen und die entsprechenden Randverteilungen überein.

Beispiel (Goldberg)

Eine symmetrische Münze wird 3 mal geworfen. Wir definieren die Zufallsvariablen

$$X = \begin{cases} 0 & \text{falls beim ersten Wurf "Zahl"} \\ 1 & \text{falls beim ersten Wurf "Kopf"} \end{cases}$$

$$\begin{array}{ll} Y = & \text{absoluter Betrag der Differenz zwischen der Anzahl} \\ & \text{„Kopf" und „Zahl"} \\ Z = & \text{Anzahl „Kopf"} \end{array} \tag{9.19}$$

Aus dem Ereignisraum

Ω = {KKK, KKZ, KZK, KZZ, ZKK, ZKZ, ZZK, ZZZ}

mit den 8 gleichwahrscheinlichen Ausgängen verifiziert man die nachfolgenden gemeinsamen Wahrscheinlichkeiten

X \ Y	1	3	f_X
0	3/8	1/8	1/2
1	3/8	1/8	1/2
f_Y	3/4	1/4	1

X und *Y* sind unabhängig, da sich die gemeinsamen Wahrscheinlichkeiten multipliziert aus den Randwahrscheinlichkeiten ergeben.

Man zeige, dass *X* und *Z* im obigen Beispiel abhängige Zufallsvariablen sind!

9.2 Stetige Zweidimensionale Zufallsvariablen

Def. 9.4

(X,Y) heisst stetige zweidimensionale Zufallsvariable, falls sie in einem Bereich B der (x,y)-Ebene jeden beliebigen reellen Wert annehmen kann.

Beispiele

- (X,Y) kann alle Werte im Rechteck $\{(x,y) | a \le x \le b, c \le y \le d\}$ annehmen.
- (X,Y) kann alle Werte der Kreisfläche $\{(x,y) | x^2 + y^2 \le 1\}$ annehmen.

Def. 9.5

(X,Y) sei eine stetige zweidimensionale Zufallsvariable. Die Funktion $f_{XY}(x,y)$ heisst *gemeinsame Dichtefunktion* auf der Menge B, falls sie folgende Bedingungen erfüllt

$$\begin{aligned} &1.) \quad f_{XY}(x,y) \ge 0 \qquad (x,y) \in B \\ &2.) \quad \iint_B f_{XY}(x,y)\, dx\, dy = 1 \end{aligned} \tag{9.20}$$

Beachte:

- Durch Bedingung 2 wird das Volumen unter dem "Dichtegebirge" auf 1 normiert.
- $f_{XY}(x, y)$ sind *keine* Wahrscheinlichkeiten. Für genügend kleine Δx und Δy gilt hingegen

$$P(x \leq X \leq x + \Delta x,\ y \leq Y \leq y + \Delta y) \approx f_{XY}(x, y)\, \Delta x\, \Delta y \tag{9.21}$$

- $f_{XY}(x, y) = 0$ falls $(x,y) \prec B$, womit Bedingung 2 geschrieben werden kann als

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dx\, dy = 1 \tag{9.22}$$

<u>Beispiel</u>

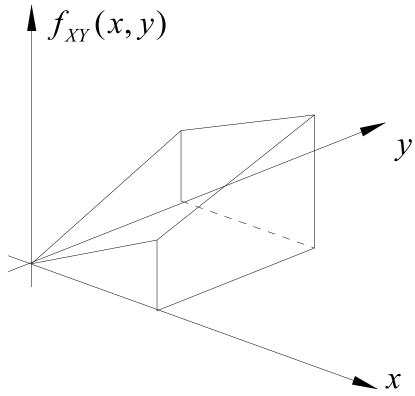
(X,Y) sei eine stetige zweidimensionale Zufallsvariable über dem Rechteck $B = \{(x, y) | 0 \leq x \leq 1, 0 \leq y \leq 1\}$ mit der Dichte

$$f_{XY}(x, y) = \begin{cases} x + y & (x, y) \in B \\ 0 & \text{sonst} \end{cases} \tag{9.23}$$

Es gilt

1.) $f_{XY}(x, y) \geq 0$ für alle (x,y)

2.)
$$\int_0^1 \int_0^1 (x + y)\, dx\, dy = \int_0^1 x \left(\int_0^1 dy \right) dx + \int_0^1 y \left(\int_0^1 dx \right) dy \tag{9.24}$$
$$= \int_0^1 x\, dx + \int_0^1 y\, dy = \frac{1}{2} + \frac{1}{2} = 1$$



Def. 9.6

$$\begin{aligned} F_{XY}(x,y) &= P(X \le x, Y \le y) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)\, du\, dv \end{aligned} \tag{9.25}$$

heisst gemeinsame Verteilungsfunktion der stetigen Zufallsvariablen (X,Y).

Im obigen Beispiel gilt

$$\begin{aligned} F(0.5,1) &= \int_{0}^{0.5}\int_{0}^{1} = (x+y)\, dx\, dy = \int_{0}^{0.5} x \int_{0}^{1} dy\, dx + \int_{0}^{1} y \int_{0}^{0.5} dx\, dy \\ &= \int_{0}^{0.5} x\, dx + 0.5 \int_{0}^{1} y\, dy = 0.375 \end{aligned} \tag{9.26}$$

<u>Beziehungen zwischen Dichte- und Verteilungsfunktion</u>

- für eindimensionale Zufallsvariablen X gilt:

$$f_X(x) = F'_X(x) \tag{9.27}$$

- für zweidimensionale Zufallsvariablen (X,Y) gilt:

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x\, \partial y} \tag{9.28}$$

9.2.1 Randverteilungen

Wie im diskreten Fall sind auch im stetigen Fall mit der gemeinsamen Dichte $f_{XY}(x,y)$ der Zufallsvariablen (X,Y) die eindimensionalen Dichten von X resp. Y verbunden. Man nennt letztere wieder die *Randverteilungen*, für die gilt

Def. 9.7

Sei $f_{XY}(x,y)$ die gemeinsame Dichte der zweidimensionalen Zufallsvariablen (X,Y). Dann heissen

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx \end{aligned} \tag{9.29}$$

Randverteilungen von X resp. Y.

<u>Beispiel</u>

Die gemeinsame Dichte der Zufallsvariablen (X,Y) sei

$$f_{XY}(x,y) = \begin{cases} 2(x+y-2xy) & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{sonst} \end{cases} \tag{9.30}$$

Für die Randverteilungen gilt dann

$$\begin{aligned} f_X(x) &= \int_0^1 2(x+y-2xy)\,dy = 2\left(xy + \frac{y^2}{2} - 2x\frac{y^2}{2}\right)\Big|_0^1 \\ &= 2\left(x + \frac{1}{2} - 2x\frac{1}{2}\right) = 1 \end{aligned} \tag{9.31}$$

Aus Symmetriegründen folgt

$$f_Y(y) = \int_0^1 2(x + y - 2xy)\, dx = 1 \qquad 0 \le y \le 1 \tag{9.32}$$

Beide Randverteilungen sind stetige Gleichverteilungen über dem Intervall [0,1].

9.2.2 Bedingte Verteilungen

Def. 9.8

(X,Y) sei eine stetige zweidimensionale Zufallsvariable mit der gemeinsamen Dichte $f_{XY}(x,y)$.

$f_X(x)$ und $f_Y(y)$ seien die jeweiligen Randverteilungen. Dann heissen

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} \qquad f_Y(y) > 0 \tag{9.33}$$

bedingte Dichte von X für gegebenes y und

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} \qquad f_X(x) > 0 \tag{9.34}$$

bedingte Dichte von Y für gegebenes x.

Beispiel

Die Zufallsvariable (X,Y) besitze die gemeinsame Dichte

$$f_{XY}(x,y) = \begin{cases} x + y & 0 \le x \le 1 \quad 0 \le y \le 1 \\ 0 & \text{sonst} \end{cases} \tag{9.35}$$

Wir bestimmen die Randverteilungen und die bedingten Verteilungen.

1.) Randverteilungen

$$f_X(x) = \int_0^1 (x+y)\,dy = \left(xy + \frac{y^2}{2} \right) \Big|_0^1 = x + \frac{1}{2} \tag{9.36}$$

$$f_Y(y) = \int_0^1 (x+y)\,dx = \left(\frac{x^2}{2} + xy \right) \Big|_0^1 = y + \frac{1}{2} \tag{9.37}$$

2.) bedingte Verteilungen

$$f_{X|Y}(x|y) = \frac{x+y}{y+\frac{1}{2}}, \quad y \text{ fest} \tag{9.38}$$

$$f_{Y|X}(y|x) = \frac{x+y}{x+\frac{1}{2}}, \quad x \text{ fest} \tag{9.39}$$

Dichten bedingter Verteilungen entstehen (abgesehen von einer Normierung), indem die Fläche im Raum, aufgespannt durch die gemeinsame Dichte $f_{XY}(x,y)$, mit Parallelebenen zu den Koordinatenebenen im Abstand y resp. x geschnitten wird.

9.2.3 Unabhängige Zufallsvariablen

Def. 9.9

(X,Y) sei eine stetige zweidimensionale Zufallsvariable mit der gemeinsamen Dichte $f_{XY}(x,y)$. Die Zufallsvariablen X und Y heissen genau dann unabhängig, wenn

$$f_{XY}(x,y) = f_X(x)\, f_Y(y) \tag{9.40}$$

für alle (x,y). f_X und f_Y sind die Randverteilungen von X und Y.

Lässt sich die gemeinsame Dichte $f_{XY}(x,y)$ der zweidimensionalen Zufallsvariablen (X,Y) so in 2 Faktoren aufspalten, dass jeder nur x- resp. y-Terme enthält, so sind X und Y unabhängig.

<u>Beispiel</u>
Ist

$$f_{XY}(x,y) = e^{-x-y} = e^{-(x+y)} \quad x \geq 0 \quad y \geq 0 \tag{9.41}$$

die gemeinsame Dichte der beiden Zufallsvariablen X und Y, so sind X und Y unabhängig, denn $f_{XY}(x,y)$ lässt sich als Produkt darstellen:

$$f_{XY}(x,y) = e^{-x-y} = e^{-x}e^{-y} = f_X(x)\, f_Y(y) \tag{9.42}$$

Aus der obigen Definition folgt unmittelbar, dass die Unabhängigkeit von X und Y auch die Beziehungen

$$\begin{aligned} f_{X|Y}(x\,|\,y) &= f_X(x) \\ & \qquad\qquad \text{für alle } (x,y) \\ f_{Y|X}(y\,|\,x) &= f_Y(y) \end{aligned} \tag{9.43}$$

impliziert.

10. FUNKTIONEN VON ZUFALLSVARIABLEN

10.1 Gemeinsame Verteilung

$X_1, X_2, \ldots, X_n$ seien auf demselben Wahrscheinlichkeitsraum (Ω,*P*) als Zufallsvariablen definiert. $g\left(X_1, \ldots, X_n\right)$ sei eine reelle Funktion von *n* Variablen. Dann ist auch $Z = g\left(X_1, X_2, \ldots, X_n\right)$ eine Zufallsvariable mit einer allerdings erst noch zu bestimmenden Wahrscheinlichkeits- resp. Dichtefunktion.

Verteilungsfunktion von *Z*

$$F_Z(z) = P\left(g\left(X_1, X_2, \ldots, X_n\right) \leq z\right) \tag{10.1}$$

bezeichnet die allgemeine Verteilungsfunktion. Der Einfachheit halber beschränken wir uns auf den Fall *n*=2 und bezeichnen die Zufallsvariablen wieder mit *X* und *Y*.

(*X*,*Y*) sei eine zweidimensionale Zufallsvariable, von der wir die Existenz einer gemeinsamen Wahrscheinlichkeits- resp. Dichtefunktion $f_{XY}(x, y)$ voraussetzen.

1.) (*X*,*Y*) sei eine diskrete zweidimensionale Zufallsvariable mit

$$\begin{aligned} P\left(X = x_i, Y = y_j\right) &= f_{XY}\left(x_i, y_j\right) \\ F_{XY}(x, y) &= \sum_{x_i \leq x} \sum_{y_j \leq y} f_{XY}\left(x_i, y_j\right) \end{aligned} \tag{10.2}$$

Ist $Z = g(X,Y)$, dann gilt

$$\begin{aligned} P(Z = z) &= f_Z(z) = \sum\sum_{g\left(x_i, y_j\right) = z} f_{XY}\left(x_i, y_j\right) \\ P(Z \leq z) &= F_Z(z) = \sum\sum_{g\left(x_i, y_j\right) \leq z} f_{XY}\left(x_i, y_j\right) \end{aligned} \tag{10.3}$$

Beispiel
Gegeben sei eine Zufallsvariable (X,Y) mit nachfolgenden gemeinsamen Wahrscheinlichkeiten

$X \backslash Y$	0	1	2	f_X
1	0.1	0.15	0	0.25
2	0.25	0.10	0.20	0.55
3	0.05	0.05	0.10	0.20
f_Y	0.40	0.30	0.30	1

Wir bestimmen die Wahrscheinlichkeitsfunktionen von

$$Z_1 = g_1(X,Y) = X + Y$$
$$\text{und} \tag{10.4}$$
$$Z_2 = g_2(X,Y) = X \cdot Y$$

z_1	1	2	3	4	5
$f_{Z_1}(z_1)$	0.1	0.4	0.15	0.25	0.10

z_2	0	1	2	3	4	6
$f_{Z_2}(z_2)$	0.40	0.15	0.10	0.05	0.20	0.10

$P(Z = z)$ ergibt sich durch Summation aller gemeinsamen Wahrscheinlichkeiten $f_{XY}(x_i, y_j)$, deren Argumente x_i und y_j die Bedingung $g(x_i, y_j) = z$ erfüllen.

2.) (X,Y) sei eine stetige zweidimensionale Zufallsvariable mit der gemeinsamen

Dichte $f_{XY}(x,y)$. Damit erhält man die Verteilungsfunktion

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv \tag{10.5}$$

Ist

$$Z = g(X,Y) \tag{10.6}$$

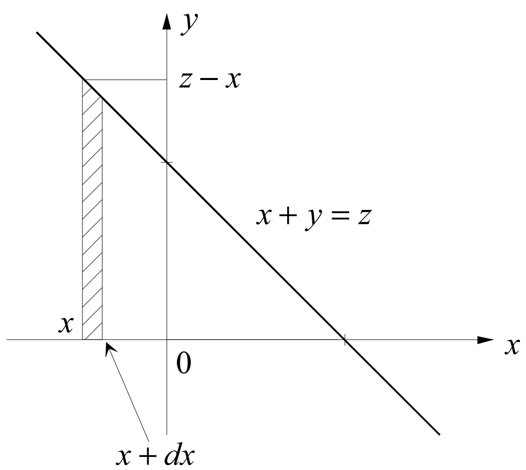
so erhält man die Verteilungsfunktion von Z gemäss

$$P(Z \le z) = F_Z(z) = \iint_{g(x,y)\le z} f_{XY}(x,y)\,dx\,dy \tag{10.7}$$

und

$$f_Z(z) = \frac{dF_Z(z)}{dz} \tag{10.8}$$

<u>Beispiel</u>
Verteilung der Summe von zwei Zufallsvariablen $Z = X + Y$



$$
\begin{aligned}
F_Z(z) = P(Z \le z) = P(X + Y \le z) &= \iint\limits_{x+y\le z} f_{XY}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} f_{XY}(x,y)\,dy\,dx \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{z} f_{XY}(x,u-x)\,du\,dx \qquad (y = u - x) \\
&= \int_{-\infty}^{z}\left(\int_{-\infty}^{+\infty} f_{XY}(x,u-x)\,dx\right)du
\end{aligned}
\tag{10.9}
$$

Die Ableitung von F_Z nach z ergibt die Dichte

$$
f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{-\infty}^{+\infty} f_{XY}(x, z-x)\,dx \tag{10.10}
$$

Sind die Zufallsvariablen X und Y unabhängig, so modifiziert sich die Dichte

$$
f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)\, f_Y(z-x)\,dx \tag{10.11}
$$

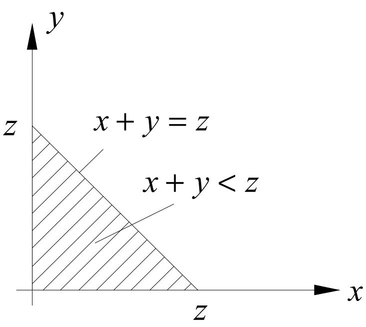
f_Z heisst Faltung der Funktionen f_X und f_Y.

Ist speziell

$$
f_{XY}(x,y) = e^{-(x+y)} \qquad x \ge 0, \quad y \ge 0 \tag{10.12}
$$

so folgt für $Z = g(X,Y) = X + Y$

$$
F_Z(z) = P(Z \le z) = P(X + Y \le z) \tag{10.13}
$$



$$f_Z(z) = \int_0^z f_{XY}(x, z-x)\, dx = \int_0^z e^{-(x+z-x)} dx$$

$$= e^{-z} \int_0^z dx = z e^{-z} \tag{10.14}$$

$$F_Z(z) = \int_0^z u e^{-u}\, du = 1 - (z+1) e^{-z}$$

10.2 Erwartungswert einer Funktion von Zufallsvariablen

$X_1, X_2, \ldots, X_n$ seien *n* Zufallsvariablen mit der gemeinsamen Wahrscheinlichkeits- resp. Dichtefunktion $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$. Dann ist auch $Z = g(X_1, X_2, \ldots, X_n)$ eine Zufallsvariable mit einer Wahrscheinlichkeits- resp. Dichtefunktion $f_Z(z)$. Für den Erwartungswert von *Z* erhält man

$$E(Z) = \sum_i z_i\, f_Z(z_i) \qquad \left(\text{bzw. } \int_{-\infty}^{+\infty} z\, f_Z(z)\, dz \right) \tag{10.15}$$

Der nachfolgende Satz erlaubt jedoch die Berechnung des Erwartungswertes von *Z* über die gemeinsame Verteilung der $X_1, \ldots, X_n$.

Satz 10.1

$X_1, X_2, \ldots, X_n$ seien Zufallsvariablen, definiert über demselben Ereignisraum mit der gemeinsamen Wahrscheinlichkeits- resp. Dichtefunktion

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \tag{10.16}$$

Ist ferner $Z = g(X_1,\ldots,X_n)$, so gilt

$$\begin{aligned} &E\big(g(X_1,\ldots,X_n)\big) \\ &= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} g(x_1,x_2,\ldots,x_n)\, f_{X_1,\ldots,X_n}(x_1,x_2,\ldots,x_n) \end{aligned} \tag{10.17}$$

im diskreten Fall und

$$\begin{aligned} &E\big(g(X_1,\ldots,X_n)\big) \\ &= \iint\cdots\int g(x_1,\ldots,x_n)\, f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\, dx_1 \ldots dx_n \end{aligned} \tag{10.18}$$

im stetigen Fall.

10.2.1 Bedingte Erwartung einer zweidimensionalen Zufallsvariablen

Sei (X,Y) eine zweidimensionale Zufallsvariable mit der gemeinsamen Verteilungsfunktion $F_{XY}(\cdot,\cdot)$. $E(Y|X=x)$ heisst bedingter Erwartungswert von Y bei vorgegebener Realisation von X.

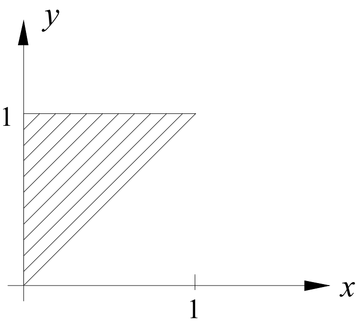
$$E(Y|X=x) = \sum_j y_j\, f_{Y|X}(y_j \mid x) = \sum_j y_j \frac{f_{XY}(x,y_j)}{f_X(x)} \tag{10.19}$$

bei gemeinsamer diskreter Verteilung von X,Y resp.

$$E(Y|X=x) = \int_{-\infty}^{+\infty} y\, f_{Y|X}(y \mid x)\, dy = \int_{-\infty}^{+\infty} y \frac{f_{XY}(x,y)}{f_X(x)}\, dy \tag{10.20}$$

bei gemeinsamer stetiger Verteilung von X,Y.

Beispiel
Die zweidimensionale Zufallsvariable (X,Y) besitze die Dichte
$f_{XY}(x,y)=2,\ 0<x<y<1$.



Mit der Randverteilung von X

$$f_X(x)=\int_x^1 2dy=2(1-x) \qquad 0<x<1 \tag{10.21}$$

folgt für die bedingte Erwartung von Y (bei gegebener Realisation von X)

$$E(Y|X=x)=\int_x^1 y\frac{2}{2(1-x)}dy=\frac{1+x}{2} \tag{10.22}$$

Beachte: Die bedingte Erwartung von Y ist eine Funktion von X.

Ein wichtiges Anwendungsfeld bedingter Erwartungen ist die Regressionsrechnung. Bezeichnet (X,Y) wieder eine zweidimensionale Zufallsvariable mit der gemeinsamen Dichte f_{XY}, so nennt man die Kurve mit den Koordinaten $(x, E(Y|X=x))$ Regressionskurve von X auf Y.

Im obigen Beispiel folgt für die Regression von X auf Y die lineare Funktion

$$y=\frac{1+x}{2} \tag{10.23}$$

Ist $g(X,Y)$ eine Funktion der beiden Zufallsvariablen, so heisst $E(g(X,Y)|X=x)$ bedingter Erwartungswert von $g(X,Y)$ bei gegebener Realisation von X

$$E\left(g\left(X,Y\right)\mid X=x\right)=\begin{cases}\sum_j g\left(x,y_j\right) f_{Y|X}\left(y_j\mid x\right)\\ \int_{-\infty}^{+\infty} g\left(x,y\right) f_{Y|X}\left(y\mid x\right)dy\end{cases} \tag{10.24}$$

10.3 Summe von Zufallsvariablen

Sowohl im Rahmen der Theorie über Stichprobenfunktionen als auch in der statistischen Praxis schlechthin besitzen Summen von Zufallsvariablen eine zentrale Bedeutung. Wir betrachten die spezielle Funktion der *n* Zufallsvariablen $X_1, X_2, \ldots, X_n$

$$S = g\left(X_1, X_2, \ldots, X_n\right) = X_1 + X_2 + \ldots + X_n \tag{10.25}$$

Die nachfolgenden Beweise werden für den Fall $n = 2$ geführt. Die Verallgemeinerung auf *n* Variablen verläuft analog.

10.3.1 Mittelwert und Varianz einer Summe von Zufallsvariablen

$X_1, X_2, \ldots, X_n$ seien *n* Zufallsvariablen mit der gemeinsamen Wahrscheinlichkeits- resp. Dichtefunktion $f_{X_1,\ldots,X_n}\left(x_1, \ldots, x_n\right)$.

Bezeichnet

$$S = X_1 + X_2 \ldots + X_n = \sum_{i=1}^{n} X_i \tag{10.26}$$

so gilt

Satz 10.2

$$E\left(X_1 + X_2 + \ldots + X_n\right) = E\left(X_1\right) + E\left(X_2\right) + \ldots + E\left(X_n\right) \tag{10.27}$$

“Erwartungswert einer Summe = Summe der Erwartungswerte”

Beweis für $n = 2$

$$\begin{aligned} E(X+Y) &= \sum_i \sum_j (x_i + y_j) f_{XY}(x_i, y_j) \\ &= \sum_i x_i \sum_j f_{XY}(x_i, y_j) + \sum_j y_j \sum_i f_{XY}(x_i, y_j) \\ &= \sum_i x_i \; f_X(x_i) + \sum_j y_j \; f_Y(y_j) \\ &= E(X) + E(Y) \end{aligned} \tag{10.28}$$

Beachte: X und Y brauchen nicht unabhängig zu sein.

Allgemein

$$E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i \, E(X_i) \tag{10.29}$$

<u>Beispiel 1</u>

$X_1, X_2, \ldots, X_n$ seien n Zufallsvariablen, die alle denselben Mittelwert haben

$$E(X_i) = \mu \qquad i = 1, \ldots, n \tag{10.30}$$

Für $\overline{X} = \frac{1}{n}(X_1 + X_2 \ldots + X_n)$ gilt dann

$$\begin{aligned} E(\overline{X}) &= E\left[\frac{1}{n}(X_1 + X_2 \ldots + X_n)\right] \\ &= \frac{1}{n} E(X_1 + X_2 \ldots + X_n) \\ &= \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right] \\ &= \frac{1}{n}[\mu + \mu + \ldots + \mu] \\ &= \frac{1}{n} n\mu \\ &= \mu \end{aligned} \tag{10.31}$$

Beispiel 2

Ein Investor plaziert 30% in die Anlagekategorie *A*, 50% in *B* und 20% in *C*. Die erwarteten Renditen in den einzelnen Kategorien betragen 5%, 4% und 4.5%.
Für die erwartete Rendite des Portefeuilles

$$P = 0.3A + 0.5B + 0.2C \tag{10.32}$$

gilt

$$\begin{aligned} E(P) = \mu_P &= E(0.3A + 0.5B + 0.2C) \\ &= 0.3E(A) + 0.5E(B) + 0.2E(C) = 4.4\% \end{aligned} \tag{10.33}$$

10.3.2 Kovarianz und Korrelationskoeffizient

Wir betrachten nun die Varianz von $Z = X + Y$

$$\begin{aligned} V(Z) = V(X+Y) &= E\left\{\left(X + Y - E(X+Y)\right)^2\right\} \\ &= E\left\{\left(X - \mu_X + Y - \mu_Y\right)^2\right\} \\ &= E\left\{\left(X-\mu_X\right)^2 + \left(Y-\mu_Y\right)^2 + 2\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right\} \\ &= E\left\{\left(X-\mu_X\right)^2\right\} + E\left\{\left(Y-\mu_Y\right)^2\right\} + 2E\left\{\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right\} \\ &= V(X) + V(Y) + 2E\left\{\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right\} \end{aligned} \tag{10.34}$$

Die Varianz einer Summe entspricht i. a. nicht der Summe der Varianzen. Dazu kommt noch ein Term, den wir als sog. Kovarianz zwischen *X* und *Y* definieren.

Def. 10.1

$$Cov(X,Y) = E\left\{\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right\} \tag{10.35}$$

heisst Kovarianz zwischen den Zufallsvariablen *X* und *Y*, wobei $\mu_X = E(X)$ und $\mu_Y = E(Y)$.

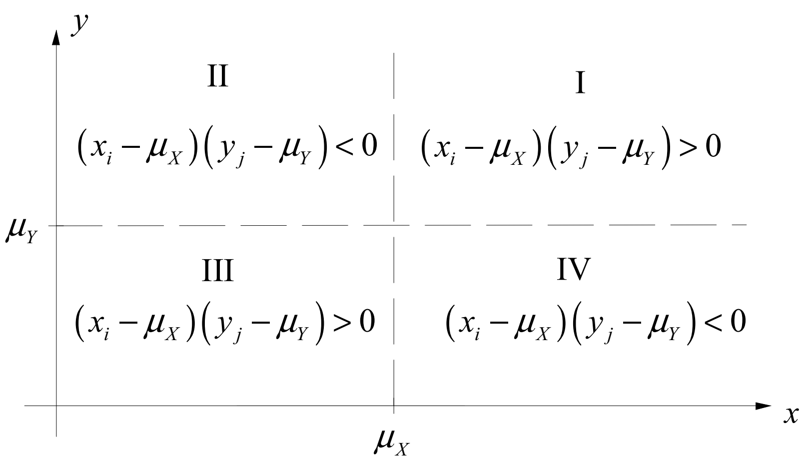
$$
\begin{aligned}
Cov(X,Y) &= E\{(X-\mu_X)(Y-\mu_Y)\} \\
&= \sum_i \sum_j (x_i-\mu_X)(y_j-\mu_Y)\, f_{XY}(x_i,y_j)
\end{aligned}
\tag{10.36}
$$

falls X,Y diskret und

$$
\begin{aligned}
Cov(X,Y) &= E\{(X-\mu_X)(Y-\mu_Y)\} \\
&= \iint (x-\mu_X)(y-\mu_Y)\, f_{XY}(x,y)\, dx\, dy
\end{aligned}
\tag{10.37}
$$

falls X,Y stetig.

Anschaulich lässt sich die Kovarianz durch folgende Graphik motivieren.



Fallen die Realisationen (x_i,y_j) mit grosser Wahrscheinlichkeit in die Gebiete I und III, so deutet dies auf eine positive Kovarianz hin. Umgekehrt ist eine negative Kovarianz mit einer Konzentration der (x_i,y_j) im II. und IV. ‘Quadranten’ verbunden.

Für die Kovarianz gilt

$$
Cov(X,Y) = E(XY) - \mu_X \mu_Y \tag{10.38}
$$

Beweis
Ausmultiplikation und Anwendung des Satzes über den Erwartungswert einer Summe von Zufallsvariablen.

Ebenso findet man

$$Cov(aX+b, cY+d) = ac\, Cov(X,Y) \tag{10.39}$$

Def. 10.2
Zwei Zufallsvariablen X und Y heissen genau dann *unkorreliert*, wenn

$$Cov(X,Y) = 0 \tag{10.40}$$

Satz 10.3
Unabhängige Zufallsvariablen sind stets unkorreliert.

Beweis

$$\begin{aligned} Cov(X,Y) &= E(XY) - \mu_X \mu_Y \\ &= \sum_i \sum_j x_i y_j \; f_{XY}(x_i, y_j) - \mu_X \mu_Y \\ &= \sum_i \sum_j x_i y_j \; f_X(x_i) f_Y(y_j) - \mu_X \mu_Y \quad (\text{Unabhängigkeit}) \\ &= \sum_i x_i \; f_X(x_i) \sum_j y_j \; f_Y(y_j) - \mu_X \mu_Y \\ &= \mu_X \mu_Y - \mu_X \mu_Y = 0 \end{aligned} \tag{10.41}$$

Beachte:
Die Umkehrung des obigen Satzes gilt i.a. nicht. (Ausnahme: (X,Y) zweidimensionale normalverteilte Zufallsvariable)

Beispiel für die Ungültigkeit der Umkehrung:
Wir betrachten die Zufallsvariablen

X	-1	0	1
$P(X=x)$	0.3	0.4	0.3

und

$Y=X^2$	0	1
$P(Y=y)$	0.4	0.6

wobei $E(X)=0$ und $E(Y)=0.6$

X und Y sind *nicht unabhängig*, denn

$$P\left(Y=0 \,\middle|\, X=0\right)=1 \neq P\left(Y=0\right)=0.4 \tag{10.42}$$

Aus der gemeinsamen Wahrscheinlichkeitsfunktion

X \ Y	0	1
-1	0	0.3
0	0.4	0
1	0	0.3

erhält man

$$E(XY)=(-1)\cdot 1\cdot 0.3+1\cdot 1\cdot 0.3=0 \tag{10.43}$$

woraus für die Kovarianz folgt

$$Cov(X,Y)=E(XY)-E(X)\cdot E(Y)=0-0=0 \tag{10.44}$$

(obschon X und Y abhängig!)

Hinweis

$\rho=\dfrac{Cov(X,Y)}{\sigma_X\sigma_Y}$ heisst Korrelationskoeffizient zwischen X und Y.

Die Umformung

$$\rho = \frac{E(X-\mu_X)(Y-\mu_Y)}{\sigma_X\sigma_Y} = E\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right) \tag{10.45}$$

zeigt, dass der Korrelationskoeffizient der Kovarianz der standardisierten Variablen von *X* und *Y* entspricht.

Mit Hilfe des Kovarianzbegriffes erhält man für die Varianz einer Summe von Zufallsvariablen $X_1, X_2, \ldots, X_n$

$$\begin{aligned} &V(X_1 + X_2 \ldots + X_n) \\ &= V(X_1) + V(X_2) + \ldots + V(X_n) + \sum_{\substack{i=1 \\ i\neq j}}^{n} \sum_{j=1}^{n} Cov(X_i, X_j) \end{aligned} \tag{10.46}$$

Sind die Zufallsvariablen $X_1, \ldots, X_n$ *paarweise unkorreliert*, so gilt

$$V(X_1 + X_2 \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n) \tag{10.47}$$

“Varianz einer Summe = Summe der Varianzen, falls die Zufallsvariablen unkorreliert sind.”

Beachte:
Die Varianz einer Summe unabhängiger Zufallsvariablen ist stets gleich der Summe der Varianzen der einzelnen Zufallsvariablen.

Für die Linearkombination

$$S = c_1X_1 + c_2X_2 + \ldots + c_nX_n = \sum_{i=1}^{n} c_iX_i \tag{10.48}$$

findet man analog

$$V(S) = V\left(\sum_{i=1}^{n} c_iX_i\right) = \sum_{i=1}^{n} c_i^2 \, V(X_i) + \sum_{\substack{i=1 \\ i\neq j}}^{n} \sum_{j=1}^{n} c_ic_j \, Cov(X_i, X_j) \tag{10.49}$$

Beispiel
Ein Portefeuille bestehe wieder aus 3 Anlagekategorien A, B und C mit den Gewichten $g_A = 0.3$, $g_B = 0.5$ und $g_C = 0.2$ und den erwarteten Renditen $E(A) = 0.05$, $E(B) = 0.04$ und $E(C) = 0.045$. Die Varianz-Kovarianzbeziehungen sind in der folgenden Matrix zusammengefasst. Die Hauptdiagonale enthält die Varianzen und den übrigen Feldern entsprechen die Kovarianzen.

	I	II	III
I	0.002	-0.003	0.004
II	-0.003	0.004	-0.001
III	0.004	-0.001	0.005

Die erwartete Rendite des Portefeuilles beträgt

$$\mu_P = 0.3 \cdot 0.05 + 0.5 \cdot 0.04 + 0.2 \cdot 0.045 = 0.044$$

Für die Varianz erhält man

$$\begin{aligned}
V(P) = \sigma_P^2 &= V(0.3 \cdot A + 0.5 \cdot B + 0.2 \cdot C) = 0.3^2 V(A) + 0.5^2 V(B) + 0.2^2 V(C) \\
&\quad + 2\,Cov(0.3A, 0.5B) + 2\,Cov(0.3A, 0.2C) + 2\,Cov(0.5B, 0.2C) \\
&= 0.3^2 V(A) + 0.5^2 V(B) + 0.2^2 V(C) \\
&\quad + 2 \cdot 0.3 \cdot 0.5\,Cov(A,B) + 2 \cdot 0.3 \cdot 0.2\,Cov(A,C) + 2 \cdot 0.5 \cdot 0.2\,Cov(B,C) \\
&= 0.00076
\end{aligned}$$

Man beachte, dass unter der obigen Kovarianzstruktur die Varianz (Risiko) des Portefeuilles kleiner ist als die Varianz (Risiko) der einzelnen Titel.

10.3.3 Gesetz der grossen Zahlen

$X_1, X_2, \ldots, X_n$ seien n gemeinsam stochastisch unabhängige Zufallsvariablen, die alle dieselbe Verteilungsfunktion besitzen. Für diese Voraussetzung verwendet man die Abkürzung iid (independent, identically distributed)

$$\left.\begin{aligned} E(X_i) &= \mu \\ V(X_i) &= \sigma^2 \end{aligned}\right. \qquad i = 1, 2, \ldots, n$$

Die Varianz des arithmetischen Mittels $\overline{X}=\frac{1}{n}\left(X_1+X_2+\ldots+X_n\right)$ ergibt

$$\begin{aligned}
V\left(\overline{X}\right)&=V\left(\frac{1}{n}\left(X_1+\ldots+X_n\right)\right)\\
&=\frac{1}{n^2}\left[V\left(X_1\right)+V\left(X_2\right)+\ldots+V\left(X_n\right)\right]\\
&=\frac{1}{n^2}\left(\sigma^2+\sigma^2+\ldots+\sigma^2\right)=\frac{1}{n^2}n\sigma^2\\
&=\frac{\sigma^2}{n}
\end{aligned}\tag{10.50}$$

Die Bezeichnung "Gesetz der grossen Zahlen" hängt damit zusammen, dass die Varianz des arithmetischen Mittels mit wachsendem *n* gegen 0 strebt.

10.3.4 Momenterzeugende Funktion einer Summe von unabhängigen Zufallsvariablen

$X_1, X_2, \ldots, X_n$ seien *n* unabhängige Zufallsvariablen mit der gemeinsamen Wahrscheinlichkeits- resp. Dichtefunktion $f_{X_1,X_2,\ldots,X_n}\left(x_1,x_2,\ldots,x_n\right)$. Mit $m_{X_i}\left(t\right)$ bezeichnen wir die momenterzeugende Funktion der Zufallsvariablen X_i und mit $m_{X_1+X_2+\ldots+X_n}\left(t\right)$ jene von $X_1+X_2+\ldots+X_n$.

Satz 10.4

Sind $X_1, X_2, \ldots, X_n$ unabhängige Zufallsvariablen mit der momenterzeugenden Funktion $m_{X_i}\left(t\right)$, so gilt für die momenterzeugende Funktion der Summe $X_1+X_2+\ldots+X_n$.

$$m_{X_1+X_2+\ldots+X_n}\left(t\right)=m_{X_1}\left(t\right)m_{X_2}\left(t\right)\ldots m_{X_n}\left(t\right)\tag{10.51}$$

Die momenterzeugende Funktion einer Summe unabhängiger Zufallsvariablen ist gleich dem Produkt der momenterzeugenden Funktionen der Summanden.

Beweis für $n = 2$ X,Y diskret

$$
\begin{aligned}
m_{X+Y}(t) &= E\left(e^{t(X+Y)}\right) = \sum_i \sum_j e^{t\left(x_i + y_j\right)} f_{XY}\left(x_i, y_j\right) \\
&= \sum_i \sum_j e^{tx_i} e^{ty_j} f_X(x_i) f_Y\left(y_j\right) \\
&= \sum_i e^{tx_i} f_X(x_i) \sum_j e^{ty_j} f_Y\left(y_j\right) = m_X(t)\, m_Y(t)
\end{aligned}
\tag{10.52}
$$

Die Bedeutung dieses Satzes wird klar, wenn man sich in Erinnerung ruft, dass die momenterzeugende Funktion eindeutig die Verteilung einer Zufallsvariablen beschreibt. Besitzen die Zufallsvariablen X und Y dieselbe momenterzeugende Funktion, so stimmen auch deren Verteilungsfunktionen überein.

<u>Beispiel</u>

$X_1, X_2, \ldots, X_n$ seien n unabhängige Bernoullivariablen, die alle dieselbe Wahrscheinlichkeitsfunktion und somit auch dieselbe momenterzeugende Funktion besitzen.

$$
\begin{aligned}
& f_X(x) = p^x (1-p)^{1-x} \qquad x = 0,1 \\
& m_X(t) = q + pe^t
\end{aligned}
\tag{10.53}
$$

Bilden wir

$$S = X_1 + X_2 + \ldots + X_n$$

so erhält man für die momenterzeugende Funktion einer Binomialverteilung mit den Parametern n und p

$$m_S(t) = \left(q + pe^t\right)^n$$

Die Summe von n unabhängigen Bernoullivariablen, die alle dieselbe Wahrscheinlichkeitsfunktion besitzen, ist binomialverteilt.

Wir haben die Technik der momenterzeugenden Funktion als sehr effizientes Mittel zur Bestimmung der Wahrscheinlichkeitsverteilung unabhängiger Zufallsvariablen kennengelernt. Wir werden die Technik an einigen weiteren Beispielen anwenden.

10.3.5 Summe unabhängiger Binomialvariablen

Seien

$$X_1 \sim B(n_1, p),\ X_2 \sim B(n_2, p), \ldots.,\ X_k \sim B(n_k, p) \tag{10.54}$$

unabhängige binomialverteilte Zufallsvariablen, die alle dieselbe Erfolgswahrscheinlichkeit besitzen. Dann gilt

$$\begin{aligned} m_{X_1+X_2+\ldots+X_n}(t) &= \left(q+pe^t\right)^{n_1}\left(q+pe^t\right)^{n_2} \ldots.\left(q+pe^t\right)^{n_k} \\ &= \left(q+pe^t\right)^{n_1+n_2+\ldots+n_k} \end{aligned} \tag{10.55}$$

Die momenterzeugende Funktion der Summe entspricht wieder derjenigen einer Binomialverteilung und es gilt

$$S = X_1 + X_2 + \ldots + X_k \sim B(n_1 + n_2 + \ldots + n_k, p) \tag{10.56}$$

Beispiel

Gegeben sind 2 Urnen mit $N_1 = 100$ und $N_2 = 10$ Kugeln. Der Anteil der "Erfolgskugeln" sei in beiden Urnen konstant, nämlich 0.3. Es werden unabhängig voneinander Stichproben mit Zurücklegen vom Umfang $n_1 = 5$ aus Urne 1 und $n_2 = 3$ aus Urne 2 gezogen.

Man bestimme die Wahrscheinlichkeitsverteilung der insgesamt gezogenen Erfolgskugeln

X_1: Anzahl Erfolge aus Urne 1
X_2: Anzahl Erfolge aus Urne 2
$S = X_1 + X_2$: Anzahl Erfolge insgesamt

$X_1 \sim B(5, 0.3)$ $X_2 \sim B(3,0.3)$ $S \sim B(8,0.3)$

s	0	1	2	3	4	5	6	7	8
$P(S=s)$	0.0576	0.1977	0.2965	0.2541	0.1361	0.0467	0.0100	0.0012	0.0001

10.3.6 Summe unabhängiger Poissonvariablen

$X_1, X_2, \ldots, X_n$ seien *n* unabhängige poissonverteilte Zufallsvariablen mit den Parametern $\lambda_1, \lambda_2, \ldots, \lambda_n$. Wahrscheinlichkeits- und momenterzeugende Funktionen sind dann

$$\begin{aligned} P\left(X_i = x\right) &= \frac{\lambda_i^x}{x!} e^{\lambda_i} & x &= 0,1,\ldots \\ m_{X_i}\left(t\right) &= e^{\lambda_i\left(e^t - 1\right)} & i &= 1,2,\ldots,n \end{aligned} \tag{10.57}$$

und es gilt unter diesen Bedingungen

Satz 10.5

Die Summe unabhängiger poissonverteilter Zufallsvariablen mit den Parametern λ_i ist wieder poissonverteilt mit dem Parameter $\sum \lambda_i$.

Beweis

$$\begin{aligned} m_{X_1+X_2+\ldots+X_n}\left(t\right) &= e^{\lambda_1\left(e^t-1\right)}\, e^{\lambda_2\left(e^t-1\right)} \ldots e^{\lambda_n\left(e^t-1\right)} \\ &= e^{\left(e^t-1\right)\left(\lambda_1+\lambda_2+\ldots+\lambda_n\right)} \end{aligned} \tag{10.58}$$

Dies ist gerade die momenterzeugende Funktion einer Poissonverteilung mit dem Parameter $\lambda_1 + \lambda_2 + \ldots + \lambda_n$.

10.3.7 Summe normalverteilter Zufallsvariablen

$X_1, X_2, \ldots, X_n$ seien unabhängige normalverteilte Zufallsvariablen

$$X_i \sim N\left(\mu_i, \sigma_i\right) \tag{10.59}$$

resp.

$$a_i X_i \sim N\left(a_i \mu_i, \left|a_i\right| \sigma_i\right) \tag{10.60}$$

Die momenterzeugende Funktion von $a_i X_i$ ist dann

$$m_{a_i X_i}(t) = e^{\left(a_i\,\mu_i\,t + \frac{1}{2} a_i^2\,\sigma_i^2\,t^2\right)} \tag{10.61}$$

woraus sich die momenterzeugende Funktion von $\sum_{i=1}^{n} a_i X_i$ ergibt

$$m_{\sum a_i X_i}(t) = e^{\left(t \sum_i a_i\,\mu_i + \frac{1}{2} t^2 \sum_i a_i^2 \sigma_i^2\right)} \tag{10.62}$$

also wiederum eine momenterzeugende Funktion einer Normalverteilung mit den Parametern

$$\begin{aligned} \mu &= \sum_i a_i\,\mu_i \\ \sigma^2 &= \sum_i a_i^2\,\sigma_i^2 \end{aligned} \tag{10.63}$$

Es gilt der Satz

Satz 10.6

Sind $X_1, X_2, \ldots, X_n$ unabhängige normalverteilte Zufallsvariablen, $X_i \sim N(\mu_i, \sigma_i)$, so ist jede Linearkombination

$$\sum_{i=1}^{n} a_i X_i \tag{10.64}$$

wiederum exakt normalverteilt mit den Parametern

$$\mu = \sum_i a_i \mu_i \quad \text{und} \quad \sigma^2 = \sum_i a_i^2\,\sigma_i^2 \tag{10.65}$$

Bei der Summe normalverteilter Zufallsvariablen ist die Voraussetzung der Unabhängigkeit nicht zwingend notwendig. Tatsächlich kann man zeigen, dass die Summe beliebig normalverteilter Zufallsvariablen ebenfalls normalverteilt ist.

Beispiele
Seien X und Y stochastisch unabhängig mit

$$\begin{aligned} X &\sim N(\mu_X, \sigma_X) \\ Y &\sim N(\mu_Y, \sigma_Y) \end{aligned} \tag{10.66}$$

dann

$$\begin{aligned} X+Y &\sim N\left(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right) \\ X-Y &\sim N\left(\mu_X - \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right) \end{aligned} \tag{10.67}$$

Sind $X_1, X_2, \ldots, X_n$ unabhängige normalverteilte Zufallsvariablen mit derselben Verteilungsfunktion

$$X_i \sim N(\mu, \sigma) \qquad i = 1, 2, \ldots, n \tag{10.68}$$

dann gilt für das arithmetische Mittel $\overline{X}$

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \tag{10.69}$$

Beispiel:
Seien X, Y, Z unabhängig standardnormalverteilte Zufallsvariablen.

Dann sind

$$\begin{aligned} U &= X + cY \\ V &= X + cZ \end{aligned}$$

jeweils wieder normalverteilt

$$\begin{aligned} U &\sim N\left(0, \sqrt{1+c^2}\right) \\ V &\sim N\left(0, \sqrt{1+c^2}\right) \end{aligned} \tag{10.70}$$

jedoch korreliert mit dem Korrelationskoeffizienten

$$\rho_{U,V} = \frac{Cov(U,V)}{\sigma_U \sigma_V} = \frac{1}{1+c^2} \tag{10.71}$$

da

$$\begin{aligned} Cov(U,V) &= E(U,V) = E(X+cY, X+cZ) \\ &= E(X^2) + cE(XZ) + cE(YX) + c^2 E(YZ) \\ &= E(X^2) + 0 \\ &= 1 \end{aligned} \tag{10.72}$$

10.3.8 Summe unabhängiger Chi-quadratvariablen

Sind U_1 und U_2 zwei unabhängige chi-quadratverteilte Zufallsvariablen mit n und m Freiheitsgraden, so ist

$$U = U_1 + U_2$$

wieder chi-quadratverteilt mit $n + m$ Freiheitsgraden.

Beweis

Für die momenterzeugenden Funktionen von U_1 und U_2 gilt

$$\begin{gathered} m_{U_1}(t) = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}} \qquad m_{U_2}(t) = \left(\frac{1}{1-2t}\right)^{\frac{m}{2}} \\ m_U(t) = m_{U_1+U_2}(t) = \left(\frac{1}{1-2t}\right)^{\frac{n+m}{2}} \end{gathered} \tag{10.73}$$

m_U ist die momenterzeugende Funktion einer Chi-quadratverteilung mit ($n + m$) Freiheitsgraden.

10.4 Produkt von Zufallsvariablen

10.4.1 Erwartungswert eines Produktes von zwei Zufallsvariablen

Wir betrachten die Zufallsvariable X,Y mit der gemeinsamen Dichte $f_{XY}(\cdot)$ und bestimmen den Mittelwert des Produktes XY.

Aus der Darstellung

$$XY = (X-\mu_X)(Y-\mu_Y) + \mu_Y X + \mu_X Y - \mu_X\mu_Y \tag{10.74}$$

folgt der Erwartungswert

$$\begin{aligned} E(XY) &= E(X-\mu_X)(Y-\mu_Y) + \mu_Y E(X) + \mu_X E(Y) - \mu_X\mu_Y \\ &= Cov(X,Y) + \mu_X\mu_Y \end{aligned} \tag{10.75}$$

Sind die Zufallsvariablen X und Y stochastisch unabhängig, so gilt

$$E(XY) = E(X)E(Y)$$

10.4.2 Produkt lognormalverteilter Zufallsvariablen

Das Bedürfnis nach einem Verteilungsgesetz eines Produktes zeigt sich etwa beim Zinseszinsmodell. Nach n Zuschlagsperioden beträgt das Endkapital

$$K_n = K_0 R_1 R_2 \dots R_n$$

Werden stochastische Renditen unterstellt, so wird naturgemäss auch das Endkapital zu einer Zufallsgrösse, dessen Verteilung von Interesse ist. Durch Logarithmieren kann das obige Problem sofort auf ein Summenproblem reduziert werden, nämlich

$$\ln K_n - \ln K_0 = \ln\left(\frac{K_n}{K_0}\right) = \sum_{i=1}^{n} \ln R_i \tag{10.76}$$

Setzt man für die (unabhängigen) Renditen R_i (i = 1, ..., n) eine Lognormalverteilung voraus, so sind deren Logarithmen ebenso wie die Summe der Logarithmen normalverteilt, womit die Aussagen über die Verteilung eines Produktes möglich werden.

Allgemein gehen wir aus von n unabhängig lognormalverteilten Zufallsvariablen $X_i\ (i = 1,\dots,n)$ mit der Dichte

$$f_{X_i}(x) = \frac{1}{x\,\sigma_i\,\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln x - \mu_i}{\sigma_i}\right)^2} \tag{10.77}$$

und betrachten ihr Produkt. Dann gilt

Satz 10.7

Das Produkt von n unabhängig lognormalverteilten Zufallsvariablen ist wieder lognormalverteilt.

Beweis

Für lognormalverteilte Zufallsvariablen mit der obigen Dichte verwenden wir die Darstellung

$$X_i \sim Ln(\mu_i, \sigma_i) \qquad i = 1, \dots, n \tag{10.78}$$

Dann gilt

$$\ln X_i \sim N(\mu_i, \sigma_i) \qquad i = 1, \dots, n \tag{10.79}$$

Durch Logarithmieren des Produktes

$$X = X_1^{c_1}\, X_2^{c_2} \dots X_n^{c_n} = \prod_{i=1}^{n} X_i^{c_i} \tag{10.80}$$

folgt

$$\ln X = \sum_{i=1}^{n} c_i \ln X_i \tag{10.81}$$

Als Summe unabhängig normalverteilter Zufallsvariablen ist $Ln\ X$ wieder normalverteilt

$$\ln X \sim N\left(\sum_{i=1}^{n} c_i\, \mu_i, \sqrt{\sum_{i=1}^{n} c_i^2\, \sigma_i^2} \right) \tag{10.82}$$

Konsequenterweise ist X selber lognormalverteilt mit

$$X \sim Ln\left(\sum_{i=1}^{n} c_i\, \mu_i, \sqrt{\sum_{i=1}^{n} c_i^2\, \sigma_i^2} \right) \tag{10.83}$$

Den Erwartungswert und die Varianz von X erhält man durch Einsetzen der entsprechenden Parameter in die Formel auf S. 130.

Für den Erwartungswert eines Produktes unabhängiger Variablen gilt nach 10.4.1, dass er dem Produkt der Erwartungswerte entspricht.

Für den Quotienten von zwei unabhängig lognormalverteilten Zufallsvariablen X_1 und X_2 gilt analog

$$X = \frac{X_1}{X_2}$$
$$\ln X = \ln X_1 - \ln X_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right) \tag{10.84}$$

womit der Quotient ebenfalls einer Lognormalverteilung genügt mit

$$X \sim Ln\left(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right) \tag{10.85}$$

10.5 Der zentrale Grenzwertsatz

Die dominante Stellung der Normalverteilung liegt im zentralen Grenzwertsatz begründet. Dieser gestattet nämlich Aussagen über die Verteilung einer Summe von n unabhängigen, identisch verteilten Zufallsvariablen

$$S_n = X_1 + X_2 + \ldots + X_n \tag{10.86}$$

Wir wollen nur voraussetzen, dass die einzelnen X_i demselben - allerdings beliebigen - Verteilungsgesetz unterliegen. Der zentrale Grenzwertsatz besagt, dass sich S_n für grosses n mit Hilfe der Normalverteilung hinreichend genau approximieren lässt.

Zentraler Grenzwertsatz 10.8 (ohne Beweis)

$X_1, X_2, \ldots, X_n$ seien unabhängige Zufallsvariablen, die alle die gleiche Verteilung, und damit auch denselben Mittelwert μ und dieselbe Varianz σ^2 besitzen. Ferner sei

$$S_n = X_1 + X_2 + \ldots + X_n$$
$$E(S_n) = n\mu \tag{10.87}$$
$$V(S_n) = n\sigma^2$$

Dann ist die Zufallsvariable

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma} \tag{10.88}$$

asymptotisch normalverteilt mit Mittelwert 0 und Standardabweichung 1.

Asymptotisch normalverteilt bedeutet

$$P(Z_n \le z) \xrightarrow{n\to\infty} \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt \tag{10.89}$$

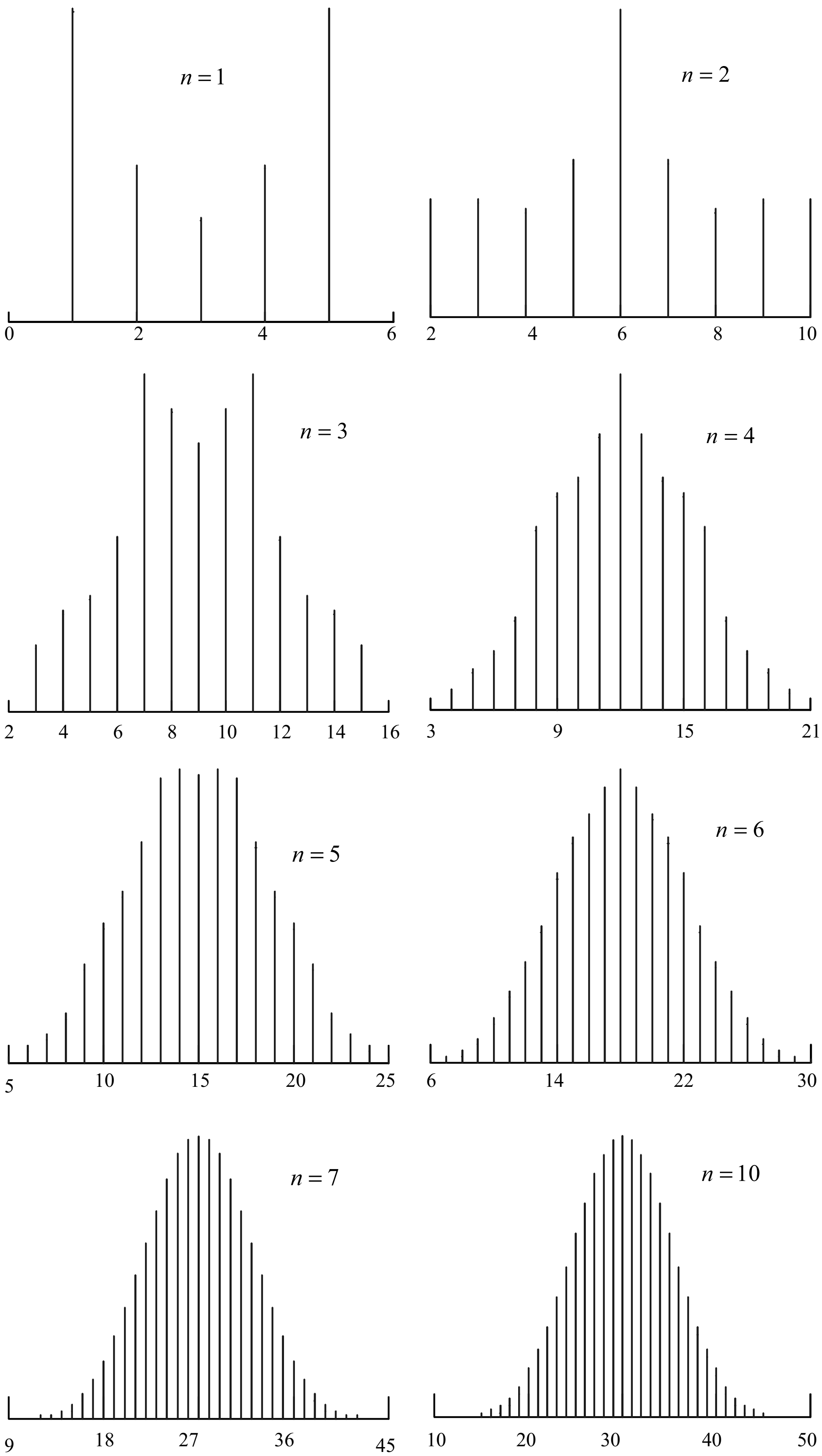
Für die praktische Anwendung existieren leider keine allgemein gültigen Regeln. Wünschenswert wären vor allem Angaben über die Grössen von n, um mit Hilfe der Approximation brauchbare Resultate zu erhalten.

Sind die Verteilungen nicht stark asymmetrisch, so wird man für $n > 30$ i.a. genügend genaue Approximationen erhalten. Für grössere Werte von n kann man sogar auf die Voraussetzungen der Unabhängigkeit und desselben Verteilungsgesetzes verzichten. Die fundamentale Bedeutung des zentralen Grenzwertsatzes tritt dadurch noch stärker in den Vordergrund. Zufallsvariablen, die sich aus additiven Komponenten einer Vielzahl von Einzeleinflüssen zuammensetzen (z.B. Messfehler, irreguläre Restschwankungen in Zeitreihen) können aufgrund des zentralen Grenzwertsatzes approximativ als normalverteilt betrachtet werden.

Graphische Motivation

Wir betrachten die Summe von n unabhängigen Zufallsvariablen, die alle dieselbe Wahrscheinlichkeitsfunktion besitzen. Für diese wählen wir

x	1	2	3	4	5
$f_X(x)$	0.30	0.15	0.10	0.15	0.30


n = 1
0
2
4
6
n = 2
2
4
6
8
10
n = 3
2
4
6
8
10
12
14
16
n = 4
3
9
15
21
n = 5
5
10
15
20
25
n = 6
6
14
22
30
n = 7
9
18
27
36
45
n = 10
10
20
30
40
50

Beispiel

$X_1, X_2, \ldots, X_n$ seien n unabhängige, identisch verteilte Zufallsvariablen mit den Parametern

$$\begin{aligned} E(X_i) &= \mu \\ & \qquad\qquad i = 1, 2, ..., n \\ V(X_i) &= \sigma^2 \end{aligned} \tag{10.90}$$

Bezeichnet

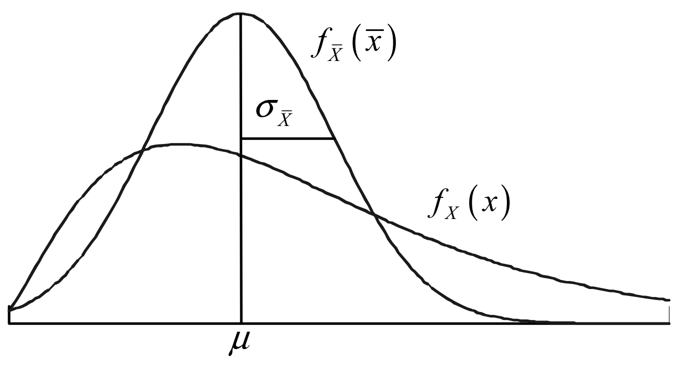
$$\bar{X} = \frac{1}{n}\left(X_1 + X_2 \ldots + X_n\right) \tag{10.91}$$

mit

$$\begin{aligned} E\left(\bar{X}\right) &= \mu \\ V\left(\bar{X}\right) &= \frac{\sigma^2}{n} \end{aligned} \tag{10.92}$$

das arithmetische Mittel der Zufallsvariablen X_i, so sagt der zentrale Grenzwertsatz, dass ungeachtet der Verteilung der X_i das arithmetische Mittel approximativ normalverteilt ist. Wir schreiben

$$\bar{X} \xrightarrow{n \text{ gross}} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \tag{10.93}$$



Beispiel

In Simulationsstudien benötigt man oft normalverteilte Zufallsvariablen. Der zentrale Grenzwertsatz gestattet, dass letztere aus stetig gleichverteilten Zufallsvariablen über [0,1] erzeugt werden können. Ist X über [0,1] stetig gleichverteilt, so gilt

$$
\begin{aligned}
f_X(x) &= 1\, I_{[0,1]}(x) \\
E(X) &= \frac{1}{2} \\
V(X) &= \frac{1}{12}
\end{aligned}
\tag{10.94}
$$

Sind $X_1, \ldots, X_{12}$ unabhängige Zufallsvariablen mit der obigen Dichte, so gilt für

$$Z = X_1 + X_2 + \ldots + X_{12} - 6 \tag{10.95}$$

$$
\begin{aligned}
E(Z) &= \sum_{i=1}^{12} E(X_i) - 6 = 6 - 6 = 0 \\
V(Z) &= \sum_{i=1}^{12} V(X_i) = 12 \cdot \frac{1}{12} = 1
\end{aligned}
\tag{10.96}
$$

und schliesslich approximativ

$$Z \sim N(0,1)$$

Beispiel

Wir betrachten den Kursverlauf einer Aktie in einer Phase ohne Dividendenausschüttung und nehmen an, dass die Kursentwicklung das Resultat einer Vielzahl multiplikativ wirkender Einzelschocks E_i sei. Jeder Einzelschock bewirke nur kleine Änderungen. Der Gesamteinfluss ist dann

$$Z_n = E_1 \cdot E_2 \cdot \ldots \cdot E_n \qquad \text{wobei} \quad E_i \approx 1 \tag{10.97}$$

Wir betrachten

$$\ln Z_n = \sum_{i=1}^{n} \ln E_i \qquad \text{mit} \;\; \ln E_i \approx 0 \tag{10.98}$$

und unterstellen, dass die Schocks unabhängig voneinander erfolgen und demselben Verteilungsgesetz genügen.

Nach dem zentralen Grenzwertsatz ist $\ln Z_n$ dann normal- und Z_n selber lognormal-verteilt.

10.6 Spezielle mehrdimensionale Verteilungen

10.6.1 Multinomialverteilung

Die Multinomialverteilung ist eine Verallgemeinerung der Binomialverteilung. Ihr liegt folgendes Urnenmodell zugrunde.

Eine Urne enthält N Kugeln, von denen N_1 der Sorte A_1, N_2 der Sorte A_2 und schliesslich N_k der Sorte A_k angehören. Innerhalb der einzelnen Sorten sind die Kugeln nicht unterscheidbar.

Es wird eine Stichprobe mit Zurücklegen vom Umfang n entnommen und nach der Wahrscheinlichkeitsverteilung für die Anzahl Elemente der einzelnen Sorten gefragt. Die Zufallsvariablen

X_i : Anzahl Kugeln der i-ten Sorte $\qquad\qquad i = 1,2,...,k$

beschreiben die Anzahl Kugeln in der Stichprobe. Gesucht ist die Wahrscheinlichkeit des Ereignisses

$$X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k$$

unter der Bedingung

$$x_1 + x_2 + \ldots + x_k = n$$

Die k (abhängigen) Zufallsvariablen $X_1, \ldots, X_k$ unterliegen der Bedingung $\sum_{i=1}^{k} X_i = n$, womit sich die Anzahl "freier" Variablen auf k - 1 reduziert.

Der Anteil der Kugeln in der Urne sei

p_i für Sorte i $\qquad$ $i = 1,...,k$

mit $\sum_{i=1}^{k} p_i = 1$.

Die Wahrscheinlichkeit für eine bestimmte Sequenz mit x_1 Kugeln der 1. Sorte, x_2 Kugeln der 2. Sorte usw., beträgt

$$p_1^{x_1}\ p_2^{x_2} \,....\, p_k^{x_k} \tag{10.99}$$

Nun gibt es aber insgesamt

$$\binom{n}{n_1\ n_2\ ...\ n_k} = \frac{n!}{n_1!\, n_2! ...\, n_k!} \tag{10.100}$$

verschiedene Anordnungen von x_1 Kugeln der 1. Sorte, x_2 Kugeln der 2. Sorte usw., die sich nur durch unterschiedliche Reihenfolge der Kugeln (nicht aber in der anteilsmässigen Zusammensetzung) voneinander unterscheiden. Wir erhalten deshalb die folgende gemeinsame Wahrscheinlichkeitsfunktion

$$P\left(X_1 = x_1, X_2 = x_2, ..., X_k = x_k\right) = \binom{n}{n_1\ n_2\ ...\ n_k} p_1^{x_1}\ p_2^{x_2} ... p_k^{x_k}$$

$$x_i \in \{1...n\}, \quad \text{so dass} \quad \sum_{i=1}^{k} x_i = n \tag{10.101}$$

$$\text{und} \qquad \sum_{i=1}^{k} p_i = 1$$

<u>Beispiel</u>
Gegeben ist ein Kartenspiel mit 36 Karten, wovon 4 Asse, 4 Könige, 4 Zehner und 24 andere. Es werden nacheinander 3 Karten mit Zurücklegen gezogen.

Mit welcher Wahrscheinlichkeit werden x_1 Asse, x_2 Könige oder Zehner und x_3 andere Karten gezogen?

X_1 : Anzahl Asse
X_2 : Anzahl Könige oder Zehner
X_3 : Anzahl übrige Karten

$$P\left(X_1=x_1,\, X_2=x_2,\, X_3=x_3\right)=\binom{3}{x_1\; x_2\; x_3}\left(\frac{1}{9}\right)^{x_1}\left(\frac{2}{9}\right)^{x_2}\left(\frac{6}{9}\right)^{x_3} \tag{10.102}$$

mit $\quad x_1+x_2+x_3=3$

<u>Beispiel</u>

$$P\left(X_1=2,\, X_2=0,\, X_3=1\right)=\frac{3!}{2!0!1!}\left(\frac{1}{9}\right)^{2}\left(\frac{2}{9}\right)^{0}\left(\frac{6}{9}\right)^{1}=0.0247 \tag{10.103}$$

<u>Randverteilungen</u>
Aus der Versuchsanordnung des Urnenmodells erkennt man sofort, dass alle Randverteilungen Binomialverteilungen sind.

$$\begin{aligned} &X_i \sim B\left(n, p_i\right) \\ &E\left(X_i\right)=np_i \qquad\qquad i=1,2,\ldots,k \\ &V\left(X_i\right)=np_i\left(1-p_i\right) \end{aligned} \tag{10.104}$$

<u>Kovarianzen</u>
Man kann zeigen

$$Cov\left(X_i, X_j\right)=-\,np_i\;p_j \qquad i\neq j \tag{10.105}$$

Das negative Vorzeichen lässt sich insofern motivieren, als bei festem Stichprobenumfang n mit grossen Werten von X_i tendenziell kleine Werte von X_j und umgekehrt verbunden sind.

<u>Beispiel</u>
Die Länge X eines bestimmten Werkstückes sei über dem Intervall [10, 12] gleichverteilt. Das Werkstück gehört zum Ausschuss, falls die Länge kleiner ist als 10.5 cm, wird akzeptiert, falls die Länge zwischen 10.5 und 11.3 cm ist und zur Überarbeitung zurückgewiesen, falls die Länge 11.3 cm übersteigt.

Der Produktion werden zufällig 10 Werkstücke entnommen. Man bestimme die Wahrscheinlichkeit, dass

1.) 2 Werkstücke defekt, 5 gut und 3 zur Überarbeitung zurückgewiesen werden müssen,

2.) keine Ausschusselemente enthalten sind.

Mit wievielen Ausschusselementen muss man im Mittel rechnen?

Wie der nachfolgende Satz zeigt, bestehen zwischen der Poissonverteilung und der Multinomialverteilung interessante Beziehungen.

Satz 10.9

$X_1, X_2, \ldots, X_n$ seien unabhängig poissonverteilte Zufallsvariablen mit den Parametern $\lambda_i \left(i = 1, 2, \ldots, n\right)$ und

$$X = \sum_{i=1}^{n} X_i \tag{10.106}$$

Dann ist die gemeinsame Verteilung der $X_1, X_2, \ldots, X_n$ unter der Bedingung $X = x$ eine Multinomialverteilung.

$$\begin{aligned} &P\left(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \middle| \sum_{i=1}^{n} X_i = x\right) \\ &= \frac{x!}{x_1!x_2!\ldots x_n!}\left(\frac{\lambda_1}{\sum \lambda_i}\right)^{x_1} \cdots \left(\frac{\lambda_n}{\sum \lambda_i}\right)^{x_n} \end{aligned} \tag{10.107}$$

Beweis

$$P\left(X_1 = x_1, \ldots, X_n = x_n \middle| \sum X_i = x\right) = \frac{P\left(X_1 = x_1, \ldots, X_n = x_n, \sum X_i = x\right)}{P\left(\sum X_i = x\right)} \tag{10.108}$$

Dann gilt

$$P\left(X_1 + X_2 + \ldots + X_n = x\right) = \frac{\left(\lambda_1 + \lambda_2 + \ldots + \lambda_n\right)^x}{x!} e^{-\left(\lambda_1 + \lambda_2 + \ldots + \lambda_n\right)} \tag{10.109}$$

Die Summe unabhängiger poissonverteilter Zufallsvariablen ist wieder poissonverteilt.

$$\begin{aligned} P\left(X_1 = x_1, \ldots, X_n = x_n, \sum X_i = x\right) &= \frac{\lambda_1^{x_1}}{x_1\,!}\, e^{-\lambda_1} \ldots \frac{\lambda_n^{x_n}}{x_n\,!}\, e^{-\lambda_n} \\ &= \frac{e^{-(\lambda_1+\lambda_2+\ldots+\lambda_n)}}{x_1\,!\,x_2\,!\ldots x_n\,!}\, \lambda_1^{x_1}\, \lambda_2^{x_2} \ldots \lambda_n^{x_n} \end{aligned} \tag{10.110}$$

sodass $x_1 + x_2 + \ldots + x_n = x$.

Bildet man den Quotienten der beiden Wahrscheinlichkeiten, so erhält man gerade die Behauptung des zu beweisenden Satzes.

10.6.2 Die zweidimensionale Normalverteilung

Wie im eindimensionalen Fall gehört auch die zweidimensionale Normalverteilung zu den zentralen Exponenten zweidimensionaler Verteilungen. Sie spielt eine fundamentale Rolle vor allem im Zusammenhang mit Problemen der Regressions- und Korrelationsanalyse (Inferenzstatistik).

Def. 10.3

Sei (X,Y) eine stetige zweidimensionale Zufallsvariable. (X,Y) besitzt eine sog. zweidimensionale Normalverteilung, falls deren gemeinsame Dichte die folgende Form besitzt

$$f_{XY}(x,y) = \frac{1}{2\pi\, \sigma_X \sigma_Y \sqrt{1-\rho^2}} \cdot \exp\left\{ \frac{-1}{2\left(1-\rho^2\right)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}$$

$$-\infty < x < +\infty \qquad -\infty < y < +\infty \tag{10.111}$$

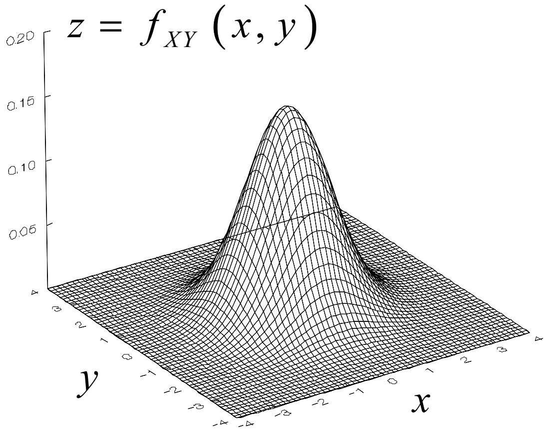
Damit $f_{XY}(x,y)$ die Bedingungen an eine Dichte

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dx\, dy = 1 \tag{10.112}$$

erfüllt, müssen an die 5 Parameter folgende Bedingungen gestellt werden

$$
\begin{aligned}
-\infty < \mu_X < +\infty \\
-\infty < \mu_Y < +\infty \\
\sigma_X > 0 \\
\sigma_Y > 0 \\
-1 < \rho < 1
\end{aligned}
\tag{10.113}
$$

Graphische Veranschaulichung der Dichte einer zweidimensionalen Normalverteilung



Der Graph von $f_{XY}(x,y)$ besitzt die Form einer Glocke. Ebenen parallel zur *x*-*y*-Ebene schneiden die Fläche im Raum in Form von Ellipsen.

Satz 10.10

Ist (X,Y) zweidimensional normalverteilt, so gilt für die Randverteilungen von X und Y

$$
\begin{aligned}
X \sim N(\mu_X, \sigma_X) \\
Y \sim N(\mu_Y, \sigma_Y)
\end{aligned}
\tag{10.114}
$$

Wir beweisen den obigen Satz für die Randverteilung von X.

Wir müssen zeigen, dass

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy = \frac{1}{\sigma_X\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \tag{10.115}$$

Wir setzen in $f_{XY}(x,y)$

$$v = \frac{y-\mu_Y}{\sigma_Y} \qquad \sigma_Y\,dv = dy \tag{10.116}$$

und erhalten

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{1}{\sigma_X\,2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{1}{2\left(1-\rho^2\right)}\left(v-\rho\frac{x-\mu_X}{\sigma_X}\right)^2\right\}dv \tag{10.117}$$

Substituiert man nochmals

$$w = \frac{1}{\sqrt{1-\rho^2}}\left(v-\rho\frac{x-\mu_X}{\sigma_X}\right) \text{ und } dv = \sqrt{1-\rho^2}\,dw \tag{10.118}$$

so erhält man schliesslich

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} \frac{1}{\sigma_X\,2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{1}{2}w^2\right\}\sqrt{1-\rho^2}\,dw \\ &= \frac{1}{\sigma_X\,2\pi}\,e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}w^2}\,dw \\ &= \frac{1}{\sigma_X\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \qquad \text{q.e.d} \end{aligned} \tag{10.119}$$

Satz 10.11

Ist (X,Y) eine zweidimensional normalverteilte Zufallsvariable, dann sind die bedingten Verteilungen von X (gegeben $Y = y$) und von Y (gegeben $X = x$) ebenfalls normalverteilt gemäss

$$\begin{aligned}(X|Y=y) &\sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y), \sigma_X\sqrt{1-\rho^2}\right)\\ (Y|X=x) &\sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y\sqrt{1-\rho^2}\right)\end{aligned} \tag{10.120}$$

Wir verzichten auf den ausführlichen Beweis dieses Satzes. Er besteht im wesentlichen aus algebraischen Umformungen bei der Bildung des Quotienten

$$f_{XY}(x\,|\,y) = \frac{f_{XY}(x,y)}{f_Y(y)} \tag{10.121}$$

Kovarianz zwischen X und Y:

Satz 10.12 (ohne Beweis)

Ist $f_{XY}(x,y)$ gemeinsame Dichte einer zweidimensionalen normalverteilten Zufallsvariablen (X,Y), so gilt

$$Cov(X,Y) = \rho\,\sigma_X\sigma_Y \tag{10.122}$$

ρ heisst *Korrelationskoeffizient* zwischen X und Y.

Satz 10.13

Ist (X,Y) zweidimensional normalverteilt, so sind X und Y genau dann unabhängig, wenn X und Y unkorreliert sind ($\rho = 0$).

Beweis

1.) Seien X und Y unabhängig. Dann gilt

$$f_{XY}(x,y) = f_X(x)\, f_Y(y) \tag{10.123}$$

wobei

$$\begin{aligned} X &\sim N(\mu_X, \sigma_X) \\ Y &\sim N(\mu_Y, \sigma_Y) \end{aligned} \tag{10.124}$$

also

$$f_{XY}(x,y) = \frac{1}{2\pi\, \sigma_X \sigma_Y} \exp\left\{ -\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\} \tag{10.125}$$

d.h. $\rho = 0$

2.) Sei $\rho = 0$. Dann gilt

$$\begin{aligned} f_{XY}(x,y) &= \frac{1}{2\pi\, \sigma_X \sigma_Y} \exp\left\{ -\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\} \\ &= \frac{1}{\sigma_X \sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \frac{1}{\sigma_Y \sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2} \end{aligned} \tag{10.126}$$

$f_{XY}(x,y)$ lässt sich so in zwei Faktoren aufspalten, dass jeder nur x resp. y enthält. Nach dem Faktorisierungstheorem sind X und Y unabhängig.

Zusammenfassung der wichtigsten Eigenschaften zweidimensional normalverteilter Zufallsvariablen (X,Y) mit der Dichte

$$f_{XY}(x,y) = \frac{1}{\sigma_X \sigma_Y\, 2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{ \frac{-1}{2\left(1-\rho^2\right)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left(\frac{x-\mu_X}{\sigma_X} \right)\left(\frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\} \tag{10.127}$$

Parameter

$$
\begin{aligned}
E(X) &= \mu_X \\
E(Y) &= \mu_Y \\
V(X) &= \sigma_X^2 \\
V(Y) &= \sigma_Y^2 \\
Cov(X,Y) &= \rho\, \sigma_X \sigma_Y \\
\rho &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \qquad (\text{Korrelationskoeffizient})
\end{aligned}
\tag{10.128}
$$

Graph

Glocke im $\forall^3$

Höhenlinien

- Ellipsen mit dem Mittelpunkt (μ_X, μ_Y).
- Länge der Halbachsen abhängig von σ_X, σ_Y, ρ und von $c = f_{XY}(x, y)$

  falls $\rho \to 1$ strebt die Länge der Halbachsen gegen 0

  falls $\sigma_X = \sigma_Y$ und $\rho = 0$ sind die Höhenlinien Kreise
- Richtung der Hauptachsen abhängig von ρ

  falls $\rho = 0$ sind die Hauptachsen parallel zu den *x*- und *y*- Achsen.

Randverteilungen

von *X*: $X \sim N(\mu_X, \sigma_X)$

von *Y*: $Y \sim N(\mu_Y, \sigma_Y)$

Bedingte Verteilungen

$$
\begin{aligned}
&\text{von X}\,|\,Y: \quad X|y \;\sim\; N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y), \sigma_X\sqrt{1-\rho^2}\right) \\
&\text{von } Y\,|\,X: \quad Y|x \;\sim\; N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y\sqrt{1-\rho^2}\right)
\end{aligned}
\tag{10.129}
$$

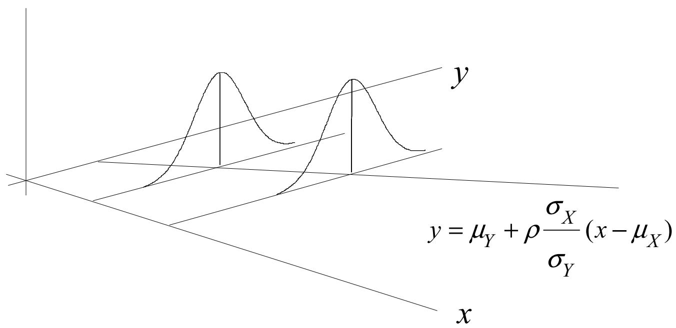
Die Standardabweichungen der bedingten Verteilungen sind proportional zu $\sqrt{1-\rho^2}$

falls $|\rho| \to 0$: bedingte Varianz ≈ unbedingte Varianz

falls $|\rho| \to 1$: bedingte Varianz → 0

Die (bedingten) Mittelwerte von $X|y$ und von $Y|x$ sind Funktionen von y resp. x. Solche Funktionen nennt man speziell *Regressionsfunktionen*. Der Mittelwert von Y in der bedingten Dichte von $Y|X = x$ ist im obigen Spezialfall eine lineare Funktion g von x, nämlich

$$y = g(x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X) \tag{10.130}$$



<u>Kovarianz-Unabhängigkeit</u>

$$Cov(X,Y) = \rho\, \sigma_X \sigma_Y \tag{10.131}$$

X und Y sind genau dann unabhängig, wenn $Cov(X,Y) = 0$ (resp. $\rho = 0$).

In der obigen Graphik verläuft z.B. die Regressionsgerade parallel zur x-Achse im Abstand μ_Y, falls $\rho = 0$, d.h. falls X und Y unabhängig.

LITERATURVERZEICHNIS

Basler, H.
Grundbegriffe der Wahrscheinlichkeitsrechnung und statistischen Methodenlehre. 11. A., Physica-Verlag, 1994

Callender, J.T., Jackson, R.
Exploring Probability and Statistics with Spreadsheets. Prentice Hall, 1995

Hoffmann-Jorgensen, J.
Probability with a View towards Statistics. Chapman and Hall, 1994

Hogg, R.V., Craig, A.T.
Introduction to Mathematical Statistics. 6. A., Prentice Hall, 2004

Hull, J.
Options, Futures, and other derivative Securities. 4. A., Prentice Hall, 2000

Kreyszig, E.
Statistische Methoden und ihre Anwendungen. 7. A., Vandenhoeck & Ruprecht, 1991 (Nachdruck)

Mood, A.M., Graybill, F.A., Boes, D.C.
Introduction to the Theory of Statistics. 3. A., McGraw Hill, 1974

Pfanzagl, J.
Elementare Wahrscheinlichkeitsrechnung. Verlag de Gruyter, 1991

Spiegel, M.R.
Statistics. McGraw Hill, 1992

INDEX

Additionssatz 18
Approximation der Binomialverteilung 101, 104, 156, 163
Approximationsmöglichkeiten 168
äquivalente Ereignisse 56, 57
Axiome der Wahrscheinlichkeitsrechnung 15
Bayes'sche Formel 27, 28
bedingte Erwartung 188, 189
bedingte Verteilungen 173, 180, 181, 221
bedingte Wahrscheinlichkeit 21, 22, 110, 173
Bernoulliexperiment 82
Bernoulliverteilung 83, 87
Betafunktion 137
Betaverteilung 151-154
binomialer Random walk 91
Binomialverteilung . 84, 86, 87, 89-91, 95, 97, 98, 101, 102, 104, 109, 111-113, 120, 150, 156, 157, 160, 163, 199, 200, 212
Boole'sche Ungleichung 19
Chi-quadratverteilung 138, 140, 204
Dichte 46, 49, 51, 52, 57, 58, 67, 71, 73, 115, 117, 119-122, 124, 129, 134, 138, 142, 144, 149, 151, 152, 177-182, 185, 186, 189, 204-206, 211, 216, 217, 219, 220, 222
diskrete Gleichverteilung 81
diskrete Zufallsvariable 37, 39, 50, 52, 55, 57
Eintretenswahrscheinlichkeit 39, 105
Ein-Ausschaltformel 19
endliche Additivität 17
Ereignis . . 1, 3, 6, 7, 9, 10, 12-15, 17, 21, 25, 27, 28, 30, 31, 39, 44, 46, 56, 85, 104, 108, 119
Ereignisraum 3, 5, 6, 15-17, 175, 188
Erwartungswert 50, 61, 62, 67-69, 73, 74, 77, 121, 134, 187-189, 194, 204-207
Erwartungswert einer Funktion von Zufallsvariablen 187
Exponentialverteilung 117-119, 143
Faustregel 98, 156, 163, 165
Funktionen von Zufallsvariablen 55, 61, 133, 183
F-Verteilung 149, 151
Gammafunktion 136, 137
Gammaverteilung 142, 143
gemeinsame Dichtefunktion 176
geometrische Verteilung 108
Häufigkeitswahrscheinlichkeit 13
hypergeometrische Verteilung 94, 97, 155, 164
Jensen'sche Ungleichung 69

Kontinuitätskorrektur . . . 155
Korrelationskoeffizient . . . 192, 195, 196, 219
Kovarianz . . . 192, 193, 195, 196, 219, 222
lineare Transformation . . . 66
Lognormalverteilung . . . 129-133, 205, 207
mathematische Erwartung . . . 61
Median . . . 78, 79, 130, 135
Mehrdimensionale Zufallsvariablen . . . 169
Mittelwert einer Zufallsvariablen . . . 50
Momente von Zufallsvariablen . . . 71
momenterzeugende Funktion . . . 73, 75, 76, 81, 83, 87, 99, 109, 112, 115, 117, 121, 135, 139, 142, 144, 198-202, 204
Multinomialverteilung . . . 212, 215
Multiplikationssatz . . . 24
negative Binomialverteilung . . . 111, 112
Normalverteilung . . . 120, 124, 126, 129-133, 140, 145, 156-159, 161, 162, 165, 202, 207, 216, 217
Pareto-Verteilung . . . 134
Poissonverteilung . . . 99-102, 104, 105, 107, 108, 112, 165, 201, 215
Poisson-Annahmen . . . 104
Poisson-Prozess . . . 104
Prozess ohne Gedächtnis . . . 110, 118
Randverteilungen . . . 172-175, 179-181, 214, 217, 221
Satz von der totalen Wahrscheinlichkeit . . . 25, 28
Schiefe der Verteilung . . . 72
Standardabweichung der Zufallsvariablen X . . . 52
standardisierte Variable . . . 66
Standardisierung . . . 124
stetige Gleichverteilung . . . 115
stetige Zufallsvariable . . . 44, 51, 52, 57, 67, 71, 134, 142, 144, 151, 155
Stichprobe ohne Zurücklegen . . . 94, 97, 162
Summe normalverteilter Zufallsvariablen . . . 201, 202
Summe unabhängiger Binomialvariablen . . . 200
Summe unabhängiger Chi-quadratvariablen . . . 204
Summe unabhängiger Poissonvariablen . . . 201
Summe von Zufallsvariablen . . . 190, 194, 196
Tschebyscheff . . . 67
unabhängige Ereignisse . . . 30, 32
unabhängige Zufallsvariablen . . . 174, 181, 194, 197, 198, 207, 211
unkorreliert . . . 194, 196, 219
unmögliches Ereignis . . . 7, 44, 46
unvereinbare Ereignisse . . . 30, 32
Varianz einer Zufallsvariablen . . . 51, 64

Verteilung von Anteilen . . . 92
Verteilungsfunktion der stetigen Zufallsvariablen X . . . 45
Verteilungsfunktion einer diskreten Zufallsvariablen . . . 41
vollständiges System (Partition) . . . 8
Wahrscheinlichkeit . . . 1, 10-13, 15, 16, 21-31, 33, 36, 39-41, 44, 46, 68, 81, 82, 84, 85, 89, 93, 94, 96, 101-105, 107-110, 113, 116, 119, 126, 128, 133, 159, 167, 172-174, 193, 212-214
Wahrscheinlichkeitsdichtefunktion . . . 45
Wahrscheinlichkeitsfunktion . . . 15, 17, 40, 43, 52, 57, 73, 75, 79, 81, 83, 86, 95, 99, 108, 111, 112, 162, 165, 169, 174, 195, 199, 208, 213
Wahrscheinlichkeitsmass . . . 20, 22, 37, 40, 145, 169
Wahrscheinlichkeitsmatrix . . . 172
Wahrscheinlichkeitsraum . . . 17, 22, 37-39, 55, 169, 183
Wahrscheinlichkeitsverteilung . . . 39, 45, 55, 57, 62, 65, 85, 94, 98, 199, 200, 212
Wartezeiten . . . 112
zentraler Grenzwertsatz . . . 207
Zufallsexperiment . . . 3, 4, 6, 10, 13, 82, 169
Zufallsvariable . . . 37, 39, 41, 43-46, 48, 50-52, 54-57, 60-62, 66-69, 71, 73-76, 80-82, 84, 85, 87, 89, 91, 92, 95, 98, 99, 105, 108, 111, 115, 117, 119, 121, 129, 134, 138, 140, 142, 144, 148, 149, 151, 155, 156, 160, 164, 165, 169, 176, 177, 180, 181, 183, 184, 187-189, 194, 204, 208, 216, 218
zweidimensionale Normalverteilung . . . 216

Verlag Wilhelm Surbir

Betten 10 · CH-9303 Wittenbach / SG
Tel. und Fax +41 (0)71 298 36 16
E-Mail verlag@surbir.ch · Internet www.surbir.ch

Lieferbare Titel

Allgoewer, Elisabeth, Dr.
Ökonomische Theoriebildung und Zeit. Eine methodenkritische Analyse anhand ausgewählter Arbeiten J.R. Hicks', 1992 (St. Galler Dissertation), Fr. 42.00

Bartmann, Hermann, Prof. Dr.
Allokationstheorie. Vorlesung, 2. Auflage 1993, Fr. 25.00

Bartmann, Hermann, Prof. Dr. und ***Borchers***, Henning, Dr.
Preistheorie. Vorlesung, 5. Auflage 1992 (unveränderter Nachdruck 1994), Fr. 30.00

Bartmann, Hermann, Prof. Dr. und ***John***, Klaus-Dieter, Prof. Dr.
Grundkonzeptionen der Konjunktur- und Wachstumsanalyse. Beiträge zur Wirtschaftstheorie
Band 1, Klassik, Neoklassik, Keynes und Keynesianismus, 4. Auflage 1994, Fr. 20.00
Band 2, Monetaristisch-neoklassische Position und Supply-Side-Ökonomien, 4. Auflage 1994, Fr. 20.00
Band 3, Postkeynesianismus, 4. Auflage 1994, Fr. 20.00

Bartmann, Hermann, Prof. Dr., ***Busch***, Andreas A., Diplom-Volkswirt und ***Schwaab***, Jan A., Diplom-Volkswirt, Preis- und Wettbewerbstheorie. Vorlesung, 6. Auflage 1999, Fr. 45.00

Beljean, Tobias, Dr. u.a.
Mikroökonomik II. Übung zur Mikroökonomik, 10. Auflage 2001, Fr. 20.00
[aktualisierte Online-Version unter http://www.surbir.ch/index.html]

Borchers, Henning, Dr.
Regulierte Strommärkte. Ein Beitrag zur (De-)Regulierungsdebatte in der Elektrizitätswirtschaft, 1994 (Mainzer Dissertation), Fr. 39.00

Brauchlin, Emil, Prof. Dr., ***Schips***, Bernd, Prof. Dr., ***Stier***, Winfried, Prof. Dr. und ***Studer***, Hans-Peter, Dr.
Statistische Methoden. Ihr sachgerechter Einsatz in der empirischen Wirtschafts- und Sozialforschung. Ein Kompendium, 3. Auflage 1987, Fr. 33.00

Einführung in die Wissenschaftstheorie für Nationalökonomen. Verfaßt von der Volkswirtschaftlichen Abteilung des Doktorandenseminars für Wissenschaftstheorie an der Hochschule St. Gallen
Band 1, hrsg. v. Prof. Dr. Walter Adolf ***Jöhr*** in Zusammenarbeit mit Dr. Gerhard ***Schwarz***, 1979, Fr. 33.00
Band 2, hrsg. v. Prof. Dr. Walter Adolf ***Jöhr*** und Prof. Dr. Bernd ***Schips*** in Zusammenarbeit mit Dr. Gerhard ***Schwarz***, 1980, Fr. 16.00

Filitti, Constantin A., Dr.
Portfolio Selection in Continuous Time, 2004 (St. Galler Dissertation), Fr. 40.00

Föller, Alex, Dr.
Umwelthaftungsrecht und Schadensprävention. Eine ökonomische Analyse der Haftung für Umweltschäden unter Einbeziehung juristischer, ökologischer und versicherungstheoretischer Aspekte, 1994 (Mainzer Dissertation), Fr. 45.00

Frenkel, Michael, Prof. Dr.
Einführung in die Makroökonomik offener Volkswirtschaften, 2. Auflage 1993 (unveränderter Nachdruck 1995), Fr. 39.50

Gauglhofer, Margrit, Prof. Dr. und ***Müller***, Heinz, Prof. Dr.
Mathematik für Oekonomen
Band 1, 14. Auflage 2004, Fr. 36.00
Band 2, 13. Auflage 2004, Fr. 20.00

Guyer, Philipp, Dr.
Der „Non-Market-Clearing“-Ansatz der Ungleichgewichtstheorie und seine Anwendung auf das keynesianische makroökonomische Standardmodell, 1981 (St. Galler Dissertation), Fr. 37.00

John, Klaus-Dieter, Prof. Dr.
Verteilungskonflikte, Inflation und Beschäftigung. Ungleichgewichtsökonomische Ansätze und sozialwissenschaftliche Erweiterungen, 1982 (Mainzer Dissertation), Fr. 44.00

KANTIge Worte und Sprueche aus berufenem Munde, von den Traegern eben derselben autorisiert und zurecht gerueckt. Zu Nutzen und Frommen nachfolgender Schuelergenerationen gesammelt an der hochwohlloeblichen und ehrbaren mathematischen Abteilung der Kantonsschule St. Gallen von deren ehemaligen Zoeglingen Carola und Matthias ***Reetz***, 1986, Fr. 12.00

Keel, Alex, Prof. Dr.
Statistik
Band 1, Beschreibende Statistik, 15. Auflage 2000, Fr. 21.00
Band 2, Wahrscheinlichkeit, 14. Auflage 2000, Fr. 23.00
Band 3, Induktive Statistik, 15. Auflage 2000, Fr. 23.00

Kippel-Chronik 1991-2001, hrsg. v. Christian ***Reetz*** und Christian ***Strehlau***, 2001, Fr. 18.00

Knecht, René, Dr.
Die Humankapitaltheorie als Ansatz zur Erklärung der personellen Arbeitseinkommensverteilung, 1988 (St. Galler Dissertation), Fr. 42.00

Koch, Christine, Dr.
Wachstum und Einkommensverteilung in postkeynesianischen Ansätzen, 1999 (Mainzer Dissertation), Fr. 48.00

Matthes, Rainer, Dr.
Zur ökonometrischen Spezifikation von Beschäftigungsfunktionen. Eine empirische Untersuchung für die BR Deutschland, 1991 (Mainzer Dissertation), Fr. 42.00

Von Musen, Müttern und der Mathematik: Frauen(an)sichten, hrsg. von Annabeth ***Naef-Hinderling*** und Johanna ***Schönenberger-Deuel***, 1998, Fr. 20.00

Räth, Norbert, Dr.
Die Zwangsanleihe als finanzpolitisches Instrument, 1980 (Mainzer Dissertation), Fr. 39.50

Reetz, Axel, Dr.
Die Entwicklung der Parteiensysteme in den baltischen Staaten. Vom Beginn des Mehrparteiensystems 1988 bis zu den dritten Wahlen, 2004 (Berliner Dissertation), Fr. 54.00

Reetz, Gesine, Sozialarbeiterin (grad.)
Rückfallprognose in der Bewährungshilfe. Eine Untersuchung anhand von Erfahrungen mit Probanden der Reutlinger Bewährungshilfe 1960-1971, 1979, Fr. 10.00

Reetz, Norbert, Prof. Dr.

> Symbole. Das griechische Alphabet und mathematische Symbole für WordPerfect und einen grafikfähigen Drucker, Version 6, 1986
> *vergriffen*
> [Online-Version unter http://www.surbir.ch/index.html]

Konjunktur und Wachstum. Eine Einführung in die reale Theorie. Vorlesung, 5. Auflage 1987, Fr. 20.00
[aktualisierte Online-Version unter http://www.surbir.ch/index.html]

Produktionstheorie. Vorlesung, 2. Auflage 1989, Fr. 20.00
[aktualisierte Online-Version unter http://www.surbir.ch/index.html]

Grundzüge der makroökonomischen Theorie. Vorlesung, 5. Auflage 1990, Fr. 30.00
[aktualisierte Online-Version unter http://www.surbir.ch/index.html]

Grundzüge der mikroökonomischen Theorie. Vorlesung, 5. Auflage 1991, Fr. 30.00

Einführung in die mikroökonomische Theorie. Vorlesung, 10. Auflage 2001, Fr. 28.00
[aktualisierte Online-Version unter http://www.surbir.ch/index.html]

Anhang zu „Einführung in die mikroökonomische Theorie. Vorlesung (10. Auflage 2001)“. Klausuren. Aufgaben und Lösungen, 2. Auflage 2001, 20.00 Fr.
[aktualisierte Online-Version unter http://www.surbir.ch/index.html]

Grundlagen der mikroökonomischen Theorie. Vorlesungen. Online-Publikation Version 12, September 2004
[http://www.surbir.ch/index.html]

Reine Theorie der Außenwirtschaft. Vorlesung, 1995, Fr. 40.00
[aktualisierte Online-Version unter http://www.surbir.ch/index.html]

Schierjott, Alexander, Dr.
Mengenrationierung und Arbeitsmarkt. Theoretische Untersuchungen und empirische Ergebnisse für die Bundesrepublik Deutschland, 1984 (Mainzer Dissertation), Fr. 29.00

Schilling, Günter, Dr.
Rationale Erwartungen in makroökonomischen Modellen, 1987 (Mainzer Dissertation), Fr. 35.00

Schindler, Rosemarie, Dr.
Die Marktpolitik des Roheisenverbandes während der Weimarer Republik, 1978 (Tübinger Dissertation), Fr. 39.20

Schlotjunker, Stefan, Dr.
The Constructed Evolution of Technology. A Constructivist-Evolutionary Approach to Technological Change and its Empirical Evidence, 1994 (St. Galler Dissertation), Fr. 42.00

Schmidt, Joachim, Dr.
Regionales Konsumverhalten. Theoretische Überlegungen und empirische Ergebnisse für ausgewählte Bundesländer der Bundesrepublik Deutschland, 1987 (Mainzer Dissertation), Fr. 42.00

Schmidt, Norbert, Dr.
Investorenverhalten und konjunkturelle Stabilität, 1987 (Mainzer Dissertation), Fr. 42.00