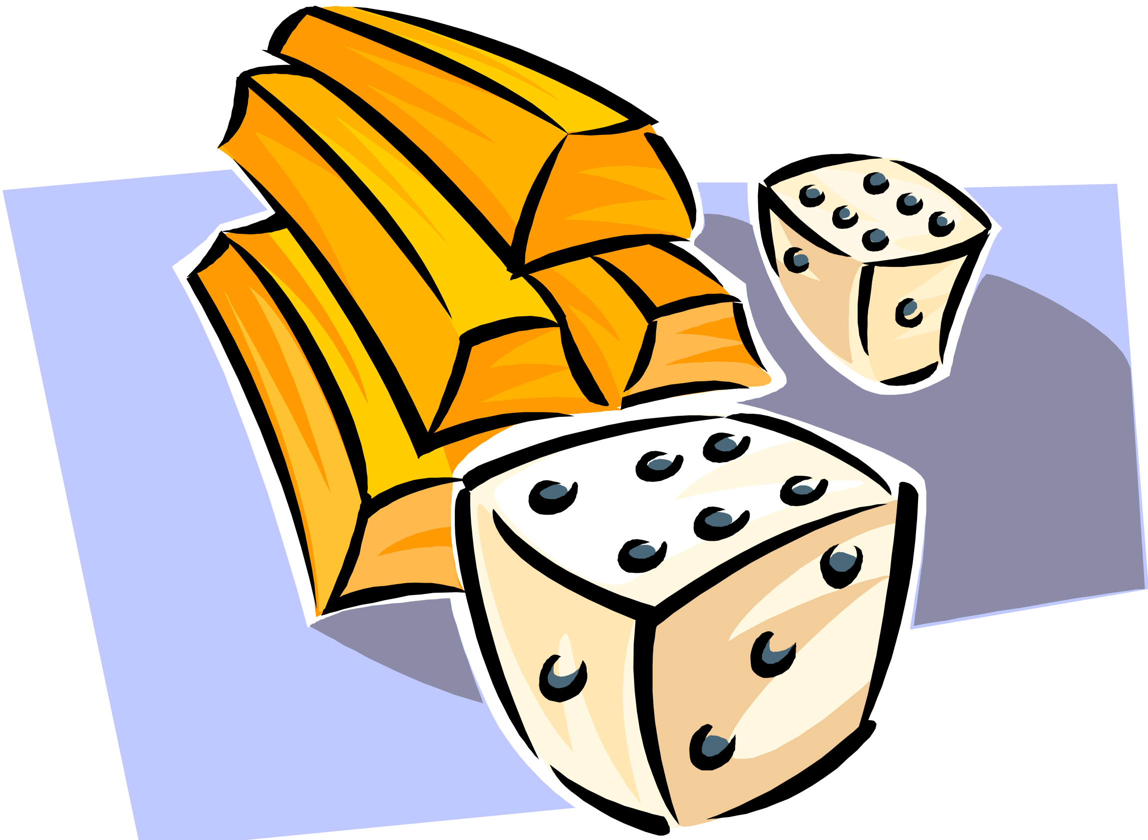
Prof. Dr. Hans-Wolfgang Henn

UNIVERSITÄT DORTMUND
Fachbereich Mathematik
Institut für Entwicklung und Erforschung
des Mathematikunterrichts

Einführung in die Stochastik



Skriptum zur Vorlesung im WS 2002/2003

Inhaltsverzeichnis

1. Wahrscheinlichkeitsrechnung

1.1 Einige stochastische Probleme

1. Lebenserwartungsproblem
Die mittlere Lebenserwartung von männlichen Lebendgeburten hat von 44,6 im Jahre 1900 auf 66,2 im Jahre 1970 zugenommen. Das heißt doch, dass die Menschen im Mittel um 21,6 Jahre älter werden, oder?

2. Geburtstagsproblem
Wie viele Personen müssen mindestens zusammenkommen, damit ich (mit fairen Chancen) darauf wetten kann, dass mindestens zwei Personen am selben Tag Geburtstag haben?

3. Lottoproblem
Welche Chancen haben ich, im Lotto „6 aus 49" genau 3 Richtige zu haben, und welche für 6 Richtige plus Zusatzzahl?

4. De-Méré-Problem
Die Erfahrung zeigt: Man kann aussichtsreich darauf wetten, dass beim mehrmaligen Werfen eines Würfels spätestens bis zum vierten Wurf eine „6" fällt. Eine „Doppel-6" beim Werfen zweier Würfel ist sechsmal seltener als eine „6" bei einem Würfel. Ist es also aussichtsreich, darauf zu wetten, dass beim mehrmaligen Werfen zweier Würfel spätestens bis zum vierundzwanzigsten Wurf (denn $6 \cdot 4 = 24$) eine „Doppel-6" fällt?

5. Teilungsproblem
Zwei Spieler führen einen Wettkampf durch. Wer zuerst 5 Einzelspiele gewonnen hat, ist Gesamtsieger und erhält einen Geldpreis von 10.000,- DM. Durch höhere Gewalt muss der Wettkampf beim Stande von 4 : 3 abgebrochen werden. Wie soll der Geldpreis unter den beiden Spielern aufgeteilt werden?

6. Dreitürenproblem
Am Ende einer Quizsendung darf der Kandidat eine von drei gleichaussehenden Türen wählen. Hinter einer der Türen verbirgt sich der Hauptgewinn, ein Auto, hinter den beiden anderen Türen befinden sich Ziegen. Nachdem der Kandidat eine Tür ausgewählt hat, öffnet der Quizmaster eine der beiden anderen Türen und zeigt eine Ziege. Dann bietet er dem Kandidaten die Möglichkeit, seine ursprüngliche Wahl zu revidieren und die andere noch geschlossene Tür zu nehmen. Was soll der Kandidat tun: Auf alle Fälle wechseln, auf alle Fälle die Erstwahl beibehalten, oder ist es sowieso egal?

7. Kugelproblem
Zuerst Armin und dann Beate ziehen abwechselnd blind eine Kugel (ohne sie wieder zurückzulegen) aus einem Sack mit 50 Kugeln, davon 49 schwarz und 1 weiß. Wer zuerst die weiße Kugel zieht, hat gewonnen. Hat Armin oder Beate die besseren Chancen?

8. Geschwisterproblem
Von einer Familie weiß ich, dass sie zwei Kinder hat, kenne aber nicht deren Geschlecht. Nun erfahre ich durch Zufall, dass eines der Kinder ein Mädchen ist. Wie groß ist die Chance, dass das andere Kind auch ein Mädchen ist?

9. Wasserbeutelproblem

100 Viertklässler werfen Wasserbeutel auf 100 Drittklässler. Jeder Viertklässler wirft einen Beutel auf einen zufällig ausgesuchten Drittklässler und trifft ihn auch sicher. Wie viele Drittklässler werden wohl trocken bleiben?

10. Zollhundproblem

Der Hund eines Zollbeamten bellt, wenn er Rauschgift erschnuppert. 98% aller Rauschgift-Schmuggelfälle entdeckt er. In 3% aller Fälle, in denen kein Rauschgift geschmuggelt hat, bellt er versehentlich trotzdem. Die Erfahrung zeigt, dass bei 1% sämtlicher Grenzübertritte Rauschgift geschmuggelt wird. Nun bellt der Hund bei einem gerade ankommenden Grenzgänger, beim nächsten nicht. Wie sicher kann der Zollbeamte sein, dass der erste tatsächlich Rauchgift schmuggelt, und wie sicher, dass der zweite es tatsächlich nicht tut?

11. Bertrand-Problem

Man zeichnet ganz zufällig eine Sehne in einen gegebenen Kreis. Wie groß ist die Chance, dass diese Sehne kürzer ist als eine Seite eines dem Kreis einbeschriebenen gleichseitigen Dreiecks?

Behandlung im Skriptum:

Nr.	Seite	Kapitel
1	8	1.3.1.2
2	43	1.6.2
3	24	1.4.4, b)
4	32	1.5.3.2, Beispiel g
5	33	1.5.3.2, Beispiel h
6	31	1.5.3.2, b)
7	53	1.6.4, Beispiel 6
8	54	1.6.4, Beispiel 8
9	54	1.6.4, Beispiel 7
10	60	1.6.5, Beispiel 1
11	34	1.5.3.4, Beispiel b

1.2 Einführende Bemerkungen

1.2.1 Meinungen zur Statistik

Benjamin Disraeli (1804 – 1881, britischer Schatzkanzler, Romancier)
There are there kinds of lies: lies, damned lies, and statistics

Andrew Lang (1844 – 1912, schottischer Gelehrter und Schriftsteller)
Wir benutzen die Statistik wie ein Betrunkener einen Laternenpfahl: vor allem zur
Stütze unseres Standpunktes und weniger zum Beleuchten eines Sachverhalts.

Otto Fürst von Bismarck (1815 – 1898)
Ich glaube nur den von mir selbst gefälschten Statistiken

Aus der Siemens-Broschüre „Argumente“: Ein Unfall mit Kernschmelzen ist für heutige Kernkraftwerke mit entsprechender Sicherheitstechnik aufgrund der extremen Unwahrscheinlichkeit praktisch ausgeschlossen.“ → *Glauben macht selig?*
(aus SPIEGEL 11/1994, Hohlspiegel, 14.3.94)

Der Gewinner der Spanischen Weihnachtslotterie 1993 wurde gefragt, woher er gewusst habe, welches Los er kaufen müsse:
„Ich habe solange gesucht, bis ich die Losnummer 48 gefunden habe.“
Wieso die 48?
„Ich habe sieben Nächte hintereinander von der Nummer 7 geträumt, also habe ich 7 mal 7 gerechnet, eben die 48!“

1.2.2 Historische Bemerkungen

Statistik:	„Ich würfle n mal und berechne....“ Verfahren, um empirische Daten zu gewinnen, darzustellen, zu verarbeiten, zu analysieren, ..
Wahrscheinlichkeitstheorie:	„Ich sage für die Zukunft voraus....“ Bestimmung eines Maßes für den Grad der Möglichkeit des Eintreffens noch unverwirklichter Ereignisse.
Stochastik:	„Mathematik des Zufalls“, Sammelbegriff für die Gebiete Statistik und Wahrscheinlichkeitstheorie

Experimente, die wir heute als Zufallsexperimente bezeichnen, hat man bereits in der Antike ausgeführt. Beispielsweise gab es Glücksspiele, deren Züge mit Tierknochen, den Astragalen, aus der Fußwurzel von Schafen oder Ziegen erwürfelt wurden (vgl. Bild in Engel, Stochastik, S. 37). Auch wurden für wichtige Entscheidungen Lose gezogen. Ein Beispiel hierfür findet sich im Alten Testament (Lev 16,8). Insbesondere versuchte man mit Astragali und Losen die Zukunft und den göttlichen Willen zu erforschen. Im Alten Rom gab es schon Versicherungen. Die Kaiser Nero und Augustus veranstalteten Lotterien mit kostbaren Gewinnen anlässlich der Saturnalien. Es bestand allerdings kein Bedürfnis zu einer „Theorie des Glücks und Ratens“.

Statistische Fragen haben ihren Ursprung im Staatswesen, z.B. Volkszählungen, Erhebung von Bevölkerungszahlen.
J. Graunt (1620 – 1674) analysierte umfangreiche Geburts- und Sterbelisten, sucht Gesetzmäßigkeiten beim Geschlechterverhältnis.
E. Halley (1656 – 1742, Astronom) untersucht Sterbetafeln als Grundlage für Leibrenten

Wahrscheinlichkeitstheorie als Wissenschaft ist sehr viel jünger, wenngleich sie ihre Ursprünge in Glücksspiel-Problemen schon vor dem 13. Jahrhundert hat. Eines dieser sehr alten Probleme ist das Teilungsproblem (vgl. 1.1 Problem 5). Bekannt ist auch das Problem des Chevalier de Méré, mit dem er sich 1654 an Blaise Pascal (1623 – 1652, Paris) wandte: „Was ist wahrscheinlicher, bei 4 Würfen mit einem Würfel mindestens eine 6 zu werfen, oder bei 24 Würfen mit 2 Würfen mindestens eine Doppel-6 zu werfen“ (vlg. 1.1 Problem 4). Dieses Jahr 1654 sieht man aufgrund des Briefwechsels zwischen Fermat und Pascal zum Teilungsproblem und zum Problem des Chevalier de Méré und ihren auch heute anerkannten Lösungen als Geburtsstunde der Wahrscheinlichkeitstheorie als Wissenschaft an. Pierre de Fermat (1601 – 1665) war übrigens Parlamentsrat in Toulouse.

Einer aus der großen Bernoulli-Familie, Jacob (1654 – 1705) erkannte das Gesetz der großen Zahl und wandte es schon auf statistische Fragen an. P.S. Laplace (1749 – 1829) verdanken wir die Definition der Wahrscheinlichkeit als Quotienten günstiger und möglicher Fälle: $p = \frac{g}{m}$; dabei müssen allerdings alle Fälle gleichwahrscheinlich sein. Wichtige Beiträge leistete C.F. Gauss (1777 – 1855), genannt seien nur Gausssche Normalverteilung, Streumaße und Fehlerrechnung. R.v. Mises (1883 – 1953) versuchte, den Begriff der Wahrscheinlichkeit allgemein als Grenzwert relativer Häufigkeiten zu *definieren*, ein Versuch, der scheitern musste. Erst A.v. Kolmogoroff (1903 – 1987) entwarf 1932 sein heute noch akzeptiertes Axiomensystem auf mengentheoretischer Grundlage. Wie Hilbert bei den Grundlagen der Geometrie verzichtet auch Kolmogoroff bei den Grundlagen der Wahrscheinlichkeit auf die ontologische Bindung der Grundbegriffe.

Zu historischen Fragen vgl. auch H. Kütting: Beschreibende Statistik – ein historischer Abriss, S. 117 – 126 in der Glatfeld-Festschrift „Beiträge zum Lernen und Lehren von Mathematik“ (Hrsg. Padberg).

1.3 Sammeln, Darstellen, Interpretieren statistischer Daten

Ziel dieses Abschnitts ist die Sensibilisierung für einige Probleme bei den vielen Graphiken und Statistiken, denen wir täglich konfrontiert sind. Die mathematischen Grundlagen sind sehr einfach.

1.3.1 Darstellungsarten und Kenngrößen

1.3.1.1 Darstellungsarten

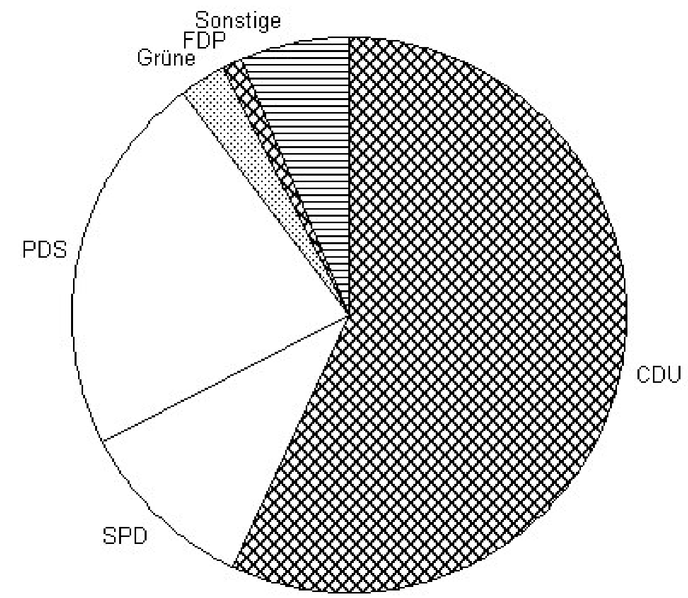
Daten eines **Merkmals** mit den Ausprägungen $x_1, ..., x_s$ (z.B. Stimmenzahl m_i für die Partei x_i bei einer Wahl) werden erhoben. Aus den **absoluten Häufigkeiten** m_i werden die **relativen Häufigkeiten** $h_i = \frac{m_i}{n}$ mit $n = m_1 + m_2 + ... + m_s$ berechnet.

Bekannte **Darstellungsarten** solcher Daten sind Strichlisten, Kreisdiagramme, Liniendiagramme, Stabdiagramme, Histogramme (Balkendiagramme), Stamm-und-Blatt-Darstellung , ...

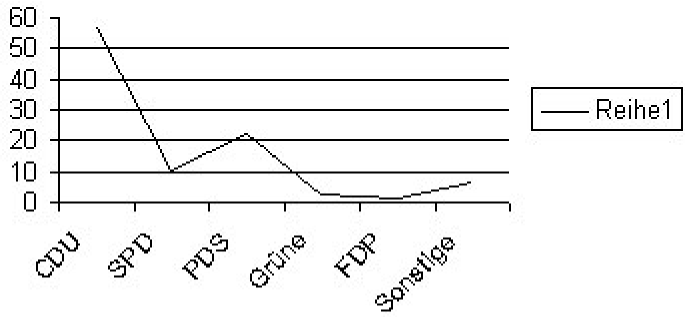
Als Beispiel werden im folgenden Daten der Landtagswahl in Sachen am 19.9.99 verschieden dargestellt (ob die jeweilige Darstellung für diese Daten Sinn macht, sei dahingestellt):

CDU	56,9 %	Grüne	2,6 %
SPD	10,7 %	FDP	1,1 %
PDS	22,2 %	Sonstige	6,5 %

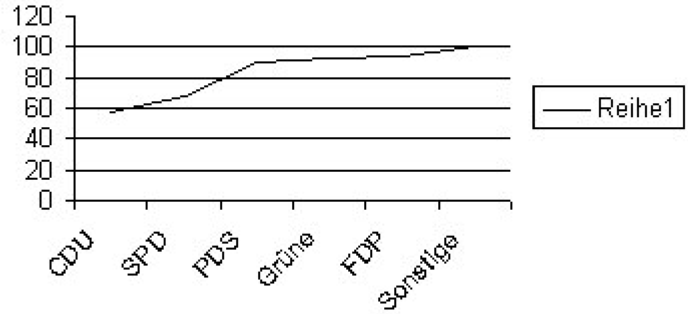
Kreisdiagramm (Winkel entspricht Anteil)



Liniendiagramm (Häufigkeitspolygon)

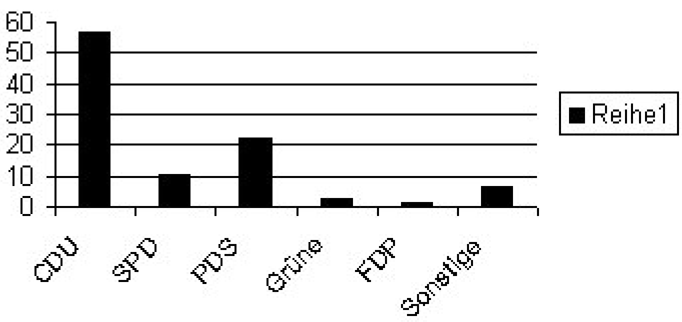


Summenlinienpolygon

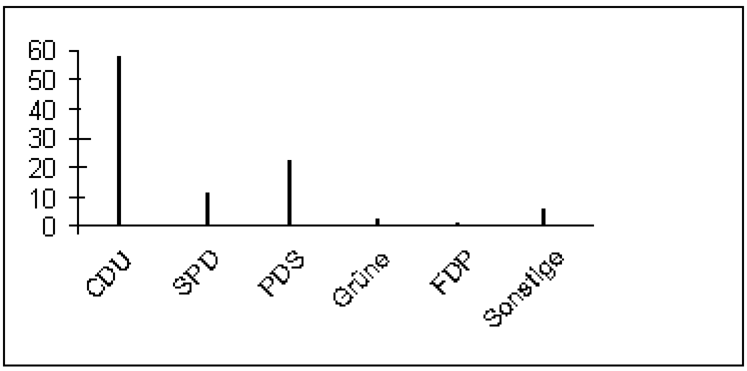


Stabdiagramm (Säulendiagramm) in drei verschiedenen Ausprägungen:

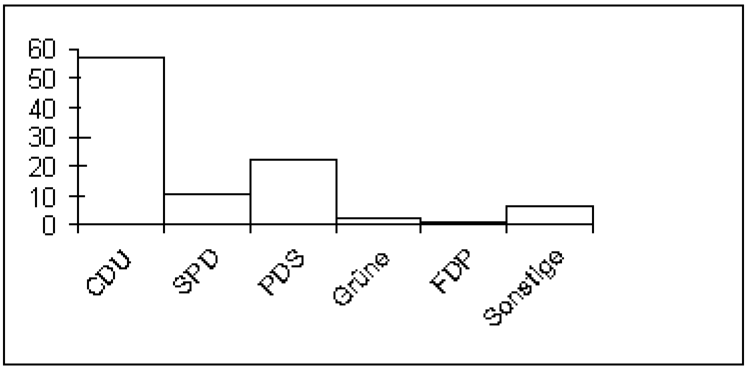
- **Computer-Darstellung**



- **„Ausprägung Stabdiagramm“**



- **„Ausprägung Histogramm (Balkendiagramm)“**



P(B / A)

Histogramme dienen der vergröberten, aber übersichtlicheren Darstellung von Merkmalen mit vielen Ausprägungen oder stetigen Merkmalen durch Klasseneinteilung: Die Stichprobenwerte $x_1, ..., x_n$ werden in s Klassen $[a_1, a_2[$, $[a_2, a_3[$, ..., $[a_s, a_{s1}[$ eingeteilt. Fallen m_j Werte in die Klasse $[a_j, a_{j+1}[$, so berechnet sich die Höhe d_j des Histogramms über dieser Klasse aus der Gleichung

$$d_j \cdot (a_{j+1} - a_j) = \frac{m_j}{n}.$$

Die Daten im folgenden Beispiel geben die jährliche Milchleistung von Kühen in dz an.

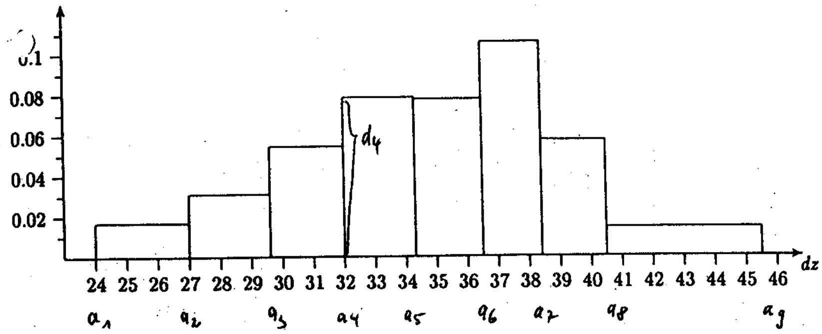
Histogramme

Vergröberung durch Klasseneinteilung bei Merkmalen mit vielen Ausprägungen oder bei (prinzipiell) stetigen Merkmalen, z.B. jährliche Milchleistung von Kühen in dz, Einteilung in 8 Klassen

37.4	37.8	29.0	35.1	30.9	28.5	38.4	34.7	36.3	30.4
39.1	37.3	45.3	32.2	27.4	37.0	25.1	30.7	37.1	37.7
26.4	39.7	33.0	32.5	24.7	35.1	33.2	42.4	37.4	37.2
37.5	44.2	39.2	39.4	43.6	28.0	30.6	38.5	31.4	29.9
34.5	34.3	35.0	35.5	32.6	33.7	37.7	35.3	37.0	37.8
32.5	32.9	38.0	36.0	35.3	31.3	39.3	34.4	37.2	39.0
41.8	32.7	33.6	43.4	30.4	25.8	28.7	31.1	33.0	39.0
37.1	36.2	28.4	37.1	37.4	30.8	41.6	33.8	35.0	37.4
33.7	33.8	30.4	37.4	39.3	30.7	30.6	35.1	33.7	32.9
35.7	32.9	39.2	37.5	26.1	29.2	34.8	33.3	28.8	38.9

Diese Daten werden in 8 Klassen eingeteilt und in folgendem Diagramm dargestellt. Die Rechtecksfläche über den Teilintervalle ist gleich der relativen Klassenhäufigkeit. Bei gleichen Klassenbreiten ist auch die Höhe proportional zur Klassenhäufigkeit.

Rechtecksfläche über den Teilintervallen ist üblicherweise gleich der relativen Klassenhäufigkeit. Bei gleichen Klassenbreiten ist auch die Höhe proportional zur Klassenhäufigkeit.



$$d_4 \cdot (a_5 - a_4) = \frac{\#_4}{\#_{alle}}$$

Dieselben Milchleistungsdaten sind in der folgenden Abbildung in der kompakten und überschtlichen **Stamm- und Blatt-Darstellung** zusammengefasst: Dabei ist z.B. die 6. Zeile so zu lesen, dass die Milchleistungen 29,0 dz, 29,2 dz und 29,9 dz vorkommen.

Stamm- und Blatt-Darstellung

Kompakte und übersichtliche Darstellung

24	7
25	8 1
26	4 1
27	4
28	4 5 0 7 8
29	0 2 9
30	4 9 4 8 7 6 6 7 4
31	3 1 4
32	5 9 7 9 2 5 6 9
33	7 8 0 6 7 2 8 3 0 7
34	5 3 8 7 4
35	7 0 1 5 3 1 3 1 0
36	2 0 3
37	4 5 1 8 3 1 4 5 4 0 7 1 4 0 2 7 2 8 4
38	0 4 5 9
39	1 7 2 2 4 3 3 0 0
40	
41	8 6
42	4
43	4 6
44	2
45	3
↓ *Stamm*	→ *Blatt*

Oft werden Daten nicht absolut sondern prozentual dargestellt. Missverständnisse können auftreten, wenn es um „% von %“ geht. Typische Beispiele sind Analysen von Wahlergebnissen:

> Eine der üblichen Gewinn-Verlust-Angaben ist
> *„Die Partei A hat 1 % gewonnen.“*
> Was bedeutet die Aussage?
> 40 % $\rightarrow$ 41 %: Zunahme um 2,5 %
>
> 4,5 % $\rightarrow$ 5,5 %: Zunahme um 22,2 %

1.3.1.2 Mittelwerte

Als übersichtliche Kenngröße gelten **Mittelwerte**. Damit ist fast immer das **arithmetische Mittel** gemeint:

$$\overline{x} = \frac{\text{Summe aller Werte}}{\text{Anzahl der Werte}} = \frac{m_1 x_1 + m_2 x_2 + ... + m_s x_s}{n} = \sum_{i=1}^{s} \frac{m_i x_i}{n} = \sum_{i=1}^{s} h_i x_i \; .$$

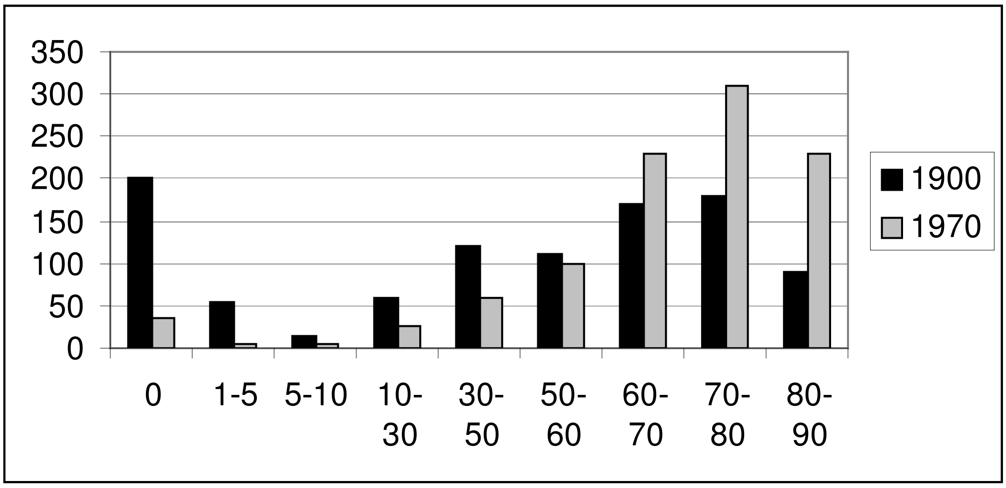
Ausrutscher bei den Daten können aber den Informationswert diese Mittelwerts sehr verfälschen:
Die folgenden Daten mögen das monatliche Einkommen von 9 Personen darstellen:
1.500, 1.100, 1.200, 1.300, 25.000, 1.000, 1.400, 1.600, 1.500.
Durch den „Ausrutscher“ 25.000 ist der Mittelwert die wenig aussagekräftige Zahl 3.956. Auf diese Art und Weise wird z.B. festgestellt, dass die Lehrerversorgung an Schulen ausreichend ist, dass das durchschnittliche Einkommen gut ist usw.

Bei diesen Daten wäre der **Median** (oder **Zentralwert**) die bessere Kenngröße. Die Daten werden als Rangreihe der Größe nach geordnet. Der Median ist die mittlere Zahl bei ungerader Anzahl von Werten, der Mittelwert der beiden mittleren Zahlen bei gerader Anzahl von Werten. Im obigen Beispiel ist die geordnete Datenreihe
1.000, 1.100, 1.200, 1.300, 1.400, 1.500, 1.500, 1.600, 25.000
mit dem Median 1.400.

Ein weiteres Beispiel, wo der Mittelwert falsche Informationen vermitteln kann, ist das **Lebenserwartungsproblem** (vgl. 1.1. Problem 1): Grundlage sind die folgenden Zahlen.[Borovnic 1992, S. 143] aus einer Absterbeordnung, wie sie in Statistischen Jahrbüchern zu finden sind:

	Von 1.000 männlichen Lebendgeborenen sterben im Alter von								
Jahr	**0**	**1 - 5**	**5 - 10**	**10 - 30**	**30 - 50**	**50 - 60**	**60 - 70**	**70 – 80**	**80 - 90**
1900	200	55	15	60	120	110	170	180	90
1970	35	5	5	25	60	100	230	310	230

In der Tat hat der Mittelwert von 44,6 im Jahre 1900 auf 66,2 im Jahre 1970 zugenommen. Diese Durchschnittszahlen kann man aber nicht als „typisches Sterbealter“ interpretieren. Die 1900-Verteilung mit Zentrum 44, 6 ist nicht einfach um 21,6 verschoben, die Menschen also alle um 21,6 Jahre älter geworden sind. Das folgende Histogramm zeigt, dass die starke Erhöhung des arithmetischen Mittels nur durch die radikale Abnahme der Säuglingssterblichkeit bewirkt wurde, während die Verteilung der über Fünfjährigen sich nur leicht zu höheren Werten verschoben hat.



Zwei weitere Mittelwerte der Daten $y_1, ..., y_n$ seien nur der Vollständigkeit halber erwähnt:

Das **geometrische Mittel** $\sqrt[n]{\prod_{i=1}^{n} y_i}$ und das **harmonische Mittel** $\dfrac{n}{\dfrac{1}{y_1}+...+\dfrac{1}{y_n}}$.

1.3.1.3 Streumaße

5 Mädchen und 5 Jungen haben jeweils mit einem Peilgerät die Höhe des Schulhauses gemessen.

Ergebnis:

	x_1	x_2	x_3	x_4	x_5	$\bar{x}$
Mädchen	10,5	9,5	10	11	9	10
Jungen	8	11	9	8	14	10

Wie kann man die qualitative Aussage, „die Mädchen haben besser gemessen“, quantifizieren?

- **Spannweite**: Differenz größter und kleinster Wert (empfindlich auf Ausreißer); im Beispiel 2 bei den Mädchen, 6 bei den Jungen.
- Maximaler Abstand vom Mittelwert max $|x_i - \bar{x}|$ (empfindlich auf Ausreißer); im Beispiel 1 bei den Mädchen, 4 bei den Jungen.
- Mittelwert der Differenzen $\sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n}$ (sinnlos, das stets = 0).
- **Mittlere lineare Abweichung** (Mittelwert der Beträge der Differenzen) $\sum_{i=1}^{n} \frac{|x_i - \bar{x}|}{n}$ (im Prinzip sinnvoll, aber Beträge schlecht manipulierbar); im Beispiel 0,6 bei den Mädchen, 2 bei den Jungen.

- **Mittlere quadratische Abweichung (empirische Varianz)** (Mittelwert der Quadrate der Differenzenquadrate, Idee von Gauss) $s^2 := \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n}$;
  im Beispiel 2,5 bei den Mädchen, 26 bei den Jungen.
  Die bestechende Idee dieses Ansatzes ist es, dass der Mittelwert $\overline{x}$ für eine Messreihe x_1, ..., x_n derjenige Schätzwert ist, der die quadratische Funktion Q mit Gleichung Q(x) = $\sum_{i=1}^{n} \frac{(x_i - x)^2}{n}$ minimiert (was man leicht durch quadratische Ergänzung nachrechnen kann.
- **Empirische Standardabweichung** $s := \sqrt{s^2}$ (begründet wegen der Einheit);
  im Beispiel 1,6 bei den Mädchen, 5,1 bei den Jungen.
- Erwähnt sei der für Messwertanalyse interessante **Standardfehler** $\Delta x = \frac{s}{\sqrt{n}}$

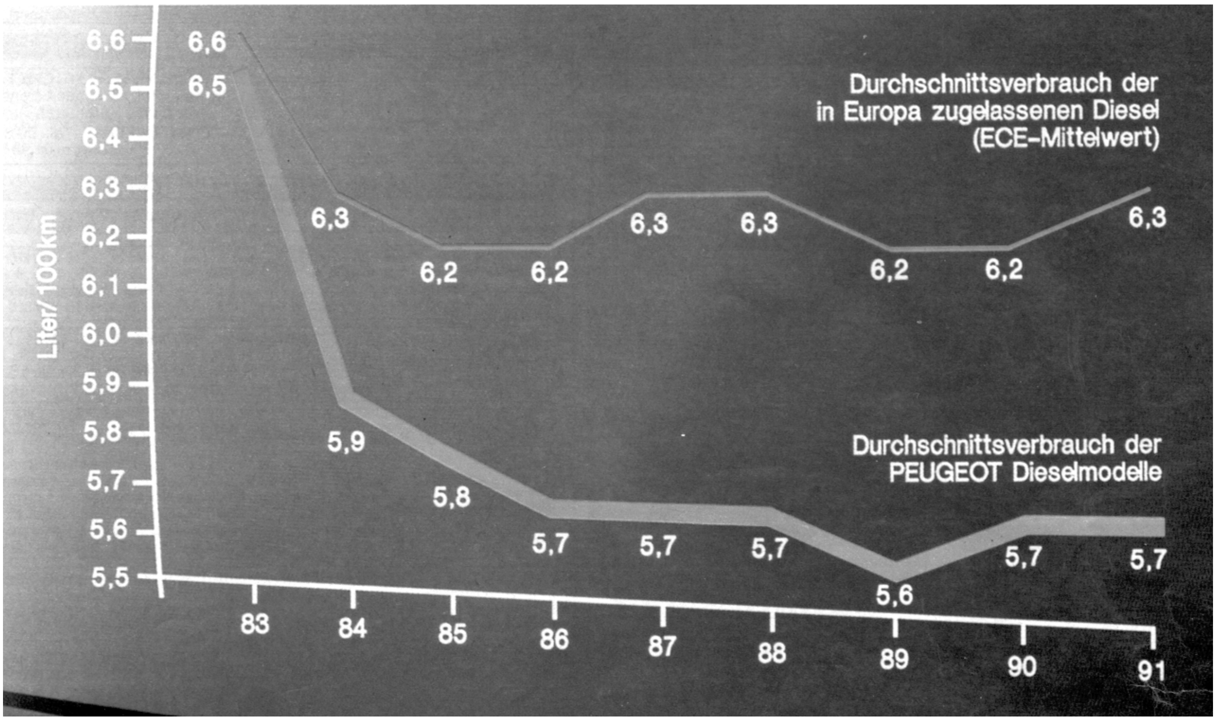
1.3.2 Bewusste und unbewusste Manipulation von Daten

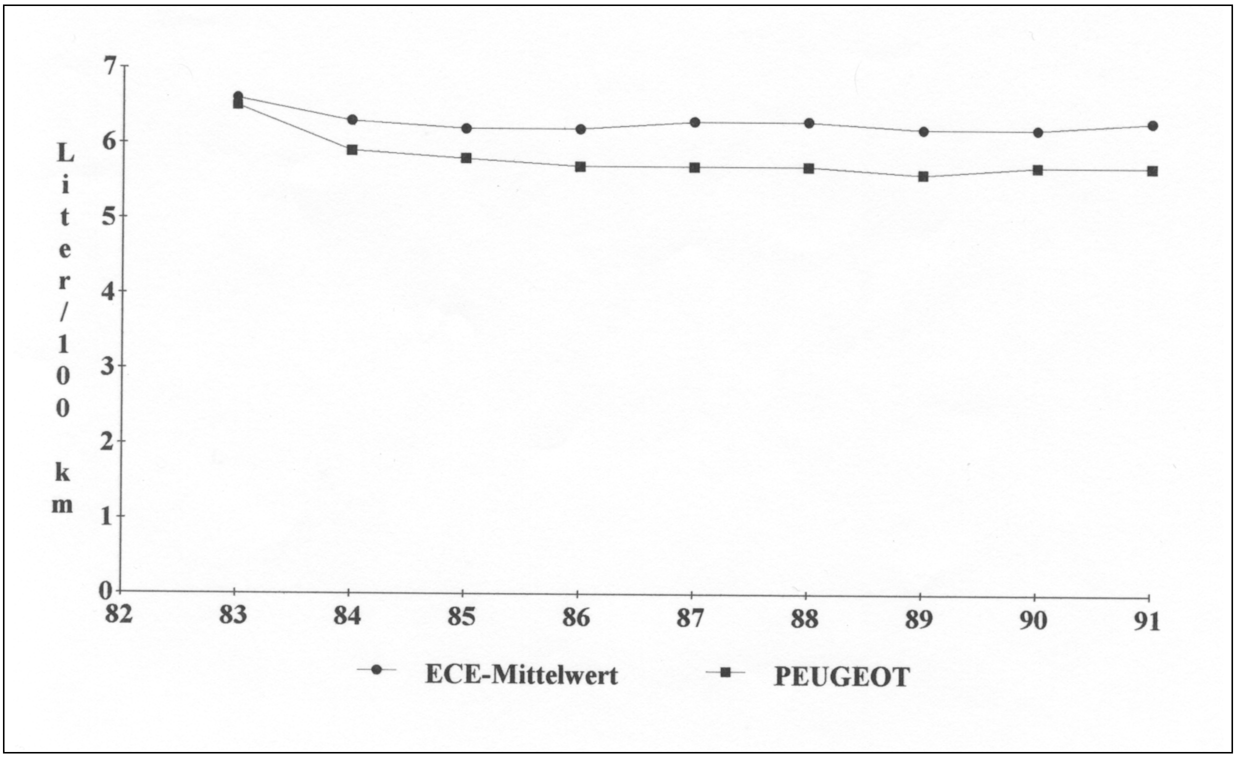
In den vielen graphischen Darstellungen in Zeitungen und Zeitschriften, die uns täglich begegnen, werden oft objektive Daten bewusst oder unbewusst falsch dargestellt oder manipuliert.

1.3.2.1 Bewusste oder unbewusste falsche Darstellung in Histogrammen und Graphiken

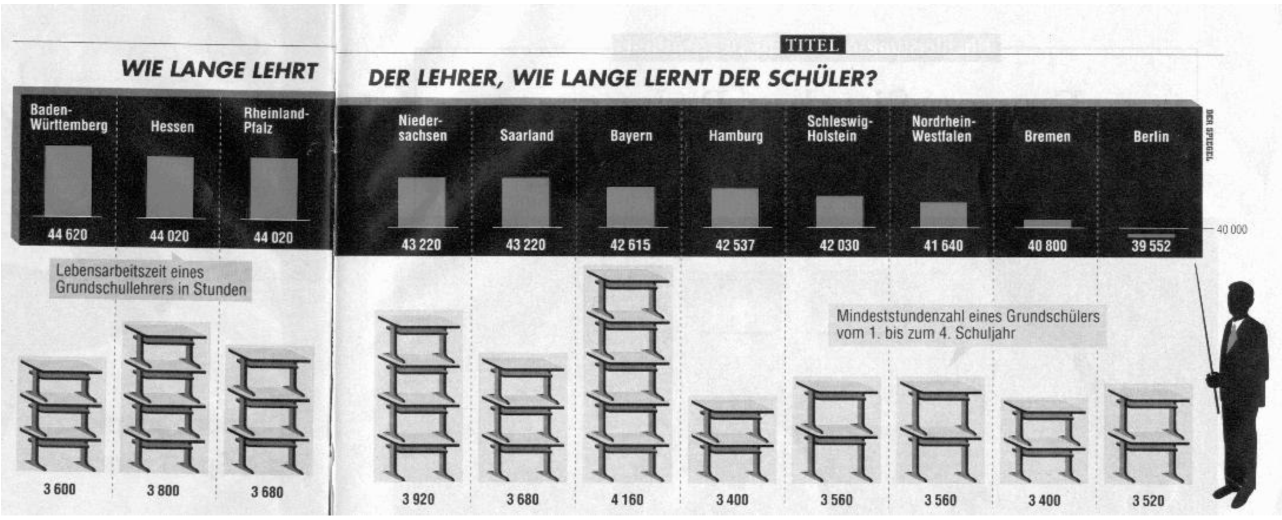
Die folgenden Beispiele stammen aus Zeitungen und Zeitschriften:

- Die Peugeot-Anzeige suggeriert aufgrund des manipulativ gewählten y-Achsenmaßstabs einen besonders niedrigen Verbrauch der Peugeot-Fahrzeuge. Die „normale Darstellung darunter zeigt, dass der Unterschied fast verschwindet:

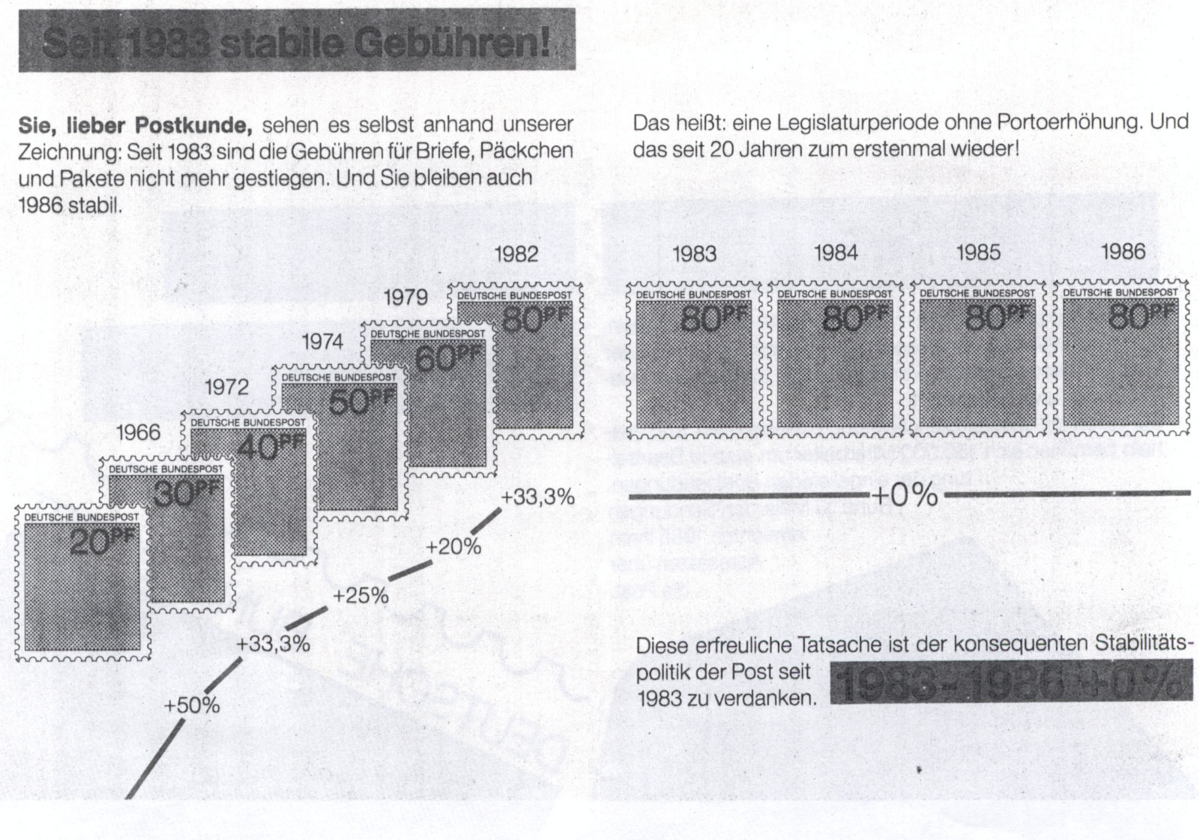


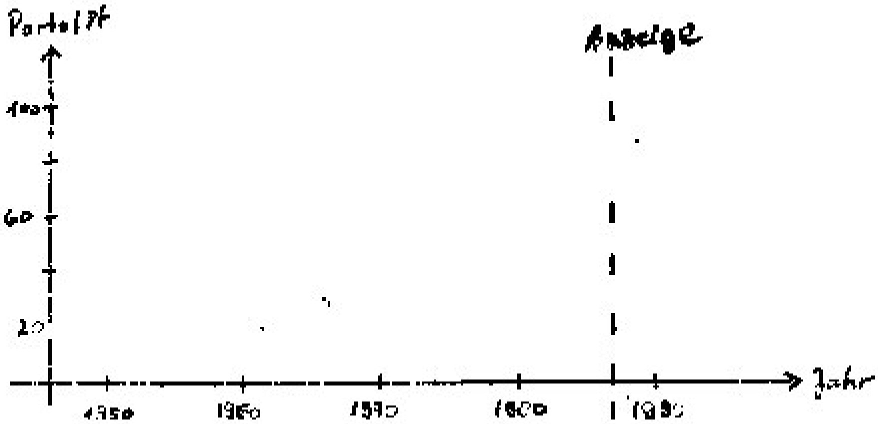


- Bei der Graphik „Wie lange lehrt der Lehrer, wie lange lernt der Schüler“ (SPIEGEL vom 7.10.91) ist auch der y-Achsenmaßstab manipulativ gewählt. Man fertige eine „normale“ Graphik an!

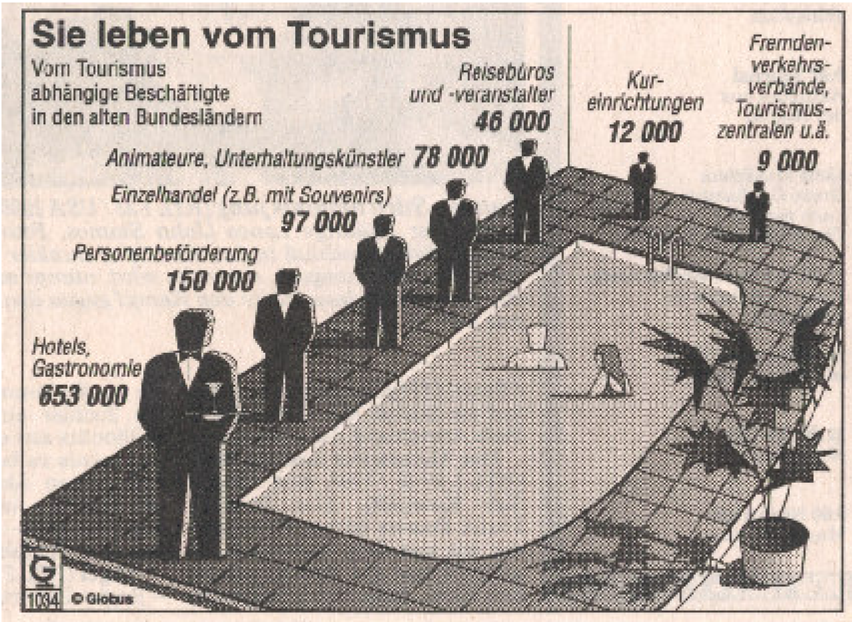


- Bei der Anzeige der damaligen Bundesregierung „Seit 1983 stabile Gebühren“ ist der x-Achsenmaßstab manipulativ gewählt (vgl. Anzeige und korrekte Darstellung):

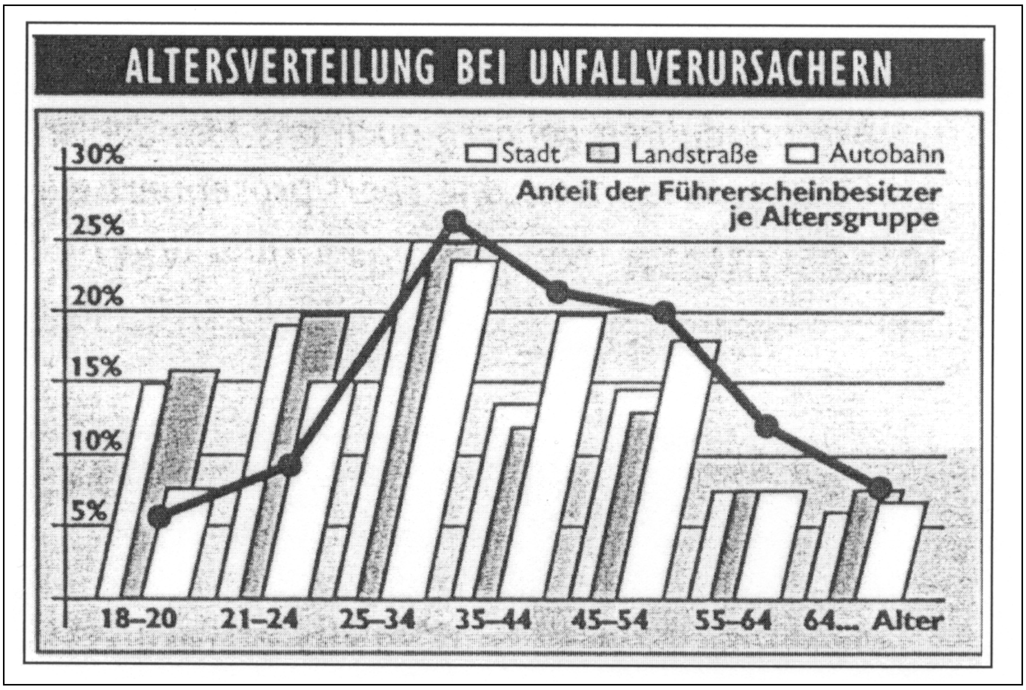




- Bei den Graphiken „Reiseziel Deutschland“ und „Sie leben vom Tourismus“ (Offenburger Tageblatt vom 8.7.95) ist nicht klar, was die Ballone bzw. die Männchen darstellen sollen? Sind die Daten proportional zu Länge, Flächeninhalt oder Volumen?



- Beim Histogramm „Altersverteilung bei Unfallursachen“ (*auto, motor und sport* 25/1994, S. 184 f) sind scheinbar gleichbreite Intervalle abgebildet, obwohl unterschiedlich breite Jahrgänge zusammengelegt wurden.



1.3.2.2 Manipulative Interpretation objektiver Daten

„Fliegen ist am sichersten.“

„Da gibt es keine Mirakel“

Airbus-Chef Hartmut Mehdorn über die Sicherheit des Airbus A 320

Mehdorn, 49, ist Geschäftsführer der Deutschen Airbus GmbH in Hamburg Finkenwerder, die zusammen mit dem französischen Partner den A 320 entwickelt hat und baut.

SPIEGEL: Viele deutsche Manager zeigten sich in den letzten Tagen von diesem Komfort unbeeindruckt. Sie versuchten, Flüge mit dem Airbus A 320 zu vermeiden.

MEHDORN: Pro Tag finden bestimmt Hunderte A-320-Flüge durch Europa statt. Klar, ängstliche Leute gibt's immer. Ich denke, daß diese Menschen, wenn sie konsequent wären, nie wieder in ihr Auto steigen dürften. Pro Milliarde Personenkilometer gibt es beim Straßenverkehr 19,6 Tote, bei der Eisenbahn 2,35 und beim Linienflugverkehr 1,3 Tote. Für eine Milliarde Kilometer müssen Sie 25 000mal die Erde umkreisen. Dazu brauchen Sie rund 150 Jahre. ◄

Airbus-Manager Mehdorn
„Ängstliche gibt's immer“

DER SPIEGEL 5/1992 193

In einem Bericht nach dem dritten Airbus-Absturz in Folge im Jahr 1992 interviewte der SPIEGEL den damaligen Airbus-Chef Mehdorn (vgl. Ausschnitt, SPIEGEL 5/1992, S. 192 f). Im letzten

Satz „beweist“ Mehdorn, dass Fliegen die sicherste Art der Fortbewegung ist. Wir setzten die Richtigkeit der Daten voraus:

$$\left.\begin{array}{lr} \text{Auto} & 19{,}6 \\ \text{Bahn} & 2{,}35 \\ \text{Flugzeug} & 1{,}3 \end{array}\right\} \text{ Tote pro } 10^9 \text{ Personen-km}$$

Bei einer fiktiven Jupiterreise argumentieren wir genauso: Es sind Hin- und Zurück $1260 \quad 10^6$ km. Je Reise sind 4 Raumfahrer dabei, also $5{,}04 \quad 10^9$ Personen-km pro Reise. Die 1. und 2. Reise verlaufen gut, die 3. Reise endet mit einem Unfall und 4 Toten, also haben wir 4 Tote auf $15{,}12 \quad 10^9$ Personen-km. Damit treten bei Jupiter-Reisen 0,25 Tote pro 10^9 Personen-km auf, womit sie – der Mehdornschen Argumentation folgend - wesentlich sicherer als Fliegen sind.

Ein anderes, vielleicht sinnvolleres Argument zur Beurteilung könnte ja die Anzahl der Tote pro Zeiteinheit im jeweiligen Fahrzeug sein („ich habe Angst, solange ich im Flugzeug sitze“). Nehmen wir als Durchschnittsgeschwindigkeiten

$$\left.\begin{array}{lr} \text{Auto} & 50 \\ \text{Bahn} & 80 \\ \text{Flugzeug} & 800 \end{array}\right\} \text{ km/h},$$

so erhalten wir eine total andere Aussage:

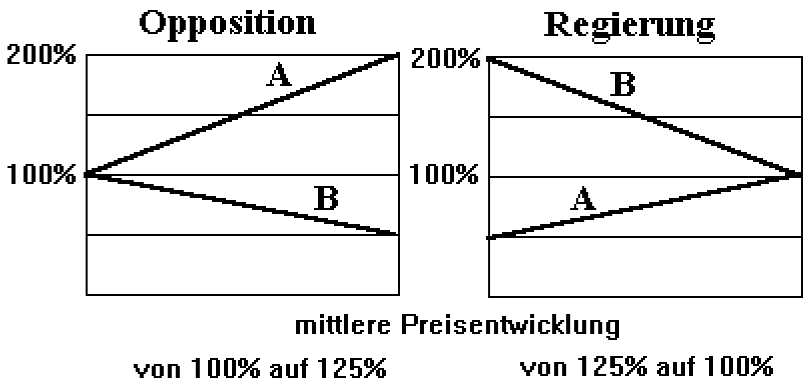
Auto:	19,6 Tote auf $\frac{10^9}{50}$ h	also 0,98 Tote auf 10^6 h
Bahn:	2,35 Tote auf $\frac{10^9}{80}$ h	also 0,188 Tote auf 10^6 h
Flugzeug:	1,3 Tote auf $\frac{10^9}{800}$ h	also 1.04 Tote auf 10^6 h

Ein weiteres Argument wäre das Abzählen von Starts und Landungen (wo die meisten Unfälle mit Flugzeugen passieren).

1.3.2.3 Subjektive Darstellung objektiver Daten

Die folgenden beiden Graphiken „Doppelte und halbe Preise“ stellen dieselben objektiven Daten graphisch unterschiedlich dar:

> Die Ware A verdoppelt ihren Preis von 1990 bis 1996.
> Die Ware B halbiert ihren Preis von 1990 bis 1996.



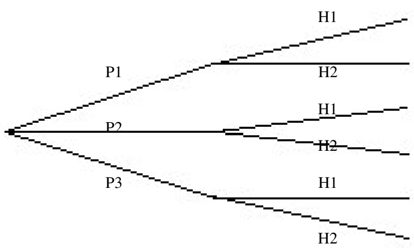
Die Opposition argumentiert, dass die Preise steigen, wogegen die Regierung von fallenden Preisen sprechen kann, beide aufgrund „objektiver Daten“.

1.4 Kombinatorik

1.4.1 Die Produktregel der Kombinatorik

Beispiel:
Torstens Barbiepuppe besitzt 3 Pullover und 2 Hosen. Wie viele Möglichkeiten gibt es, der Puppe einen Pullover und eine Hose anzuziehen?

Lösung: (Entscheidungs-)Baum:



Zu jedem der 3 Pullover kann Torsten eine der 2 Hosen auswählen (oder: Zu jeder der 2 Hosen ...). Insgesamt gibt es also $2 \cdot 3 = 6$ Möglichkeiten.

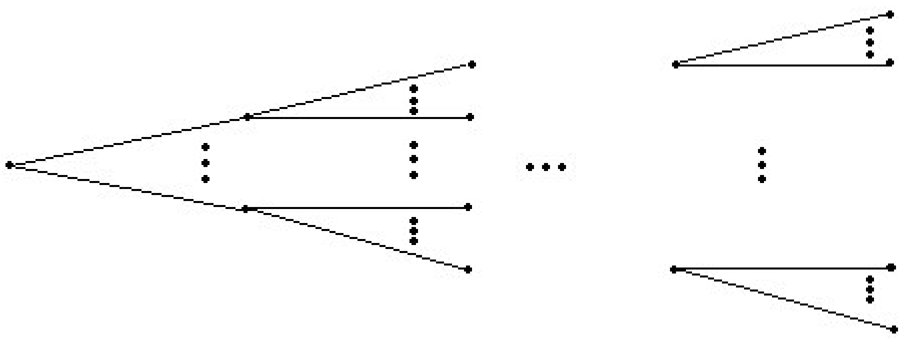
Entscheidung „Pullover“: 3 Möglichkeiten	Entscheidung „Hose“: 2 Möglichkeiten

Die Verallgemeinerung ist die

Produktregel der Kombinatorik

Es seien nacheinander und unabhängig voneinander n Einzel-Entscheidungen zu treffen, wobei es jeweils r_i (i = 1, ... , n) Entscheidungs-Möglichkeiten gebe. Dann gibt es insgesamt $r_1 \cdot r_2 \cdot \ldots \cdot r_n$ Entscheidungs-Möglichkeiten.

Begründung und Veranschaulichung am Baum:



1. Entscheidung r_1 Möglichkeiten	2. Entscheidung jeweils r_2 Möglichkeiten	n-te Entscheidung jeweils r_n Möglichkeiten

zusammen $r_1 \cdot r_2 \cdot \ldots \cdot r_n$ „Enden“ des Baumes.

Die Produktregel ist ein wichtiges Hilfsmittel schwer überschaubare Mengen abzuzählen. Dabei treten 4 typische Fälle auf.

1.4.2 Die kombinatorischen Grundaufgaben

- Ein Fahrradschloss hat ein 3-stelliges Zahlenschloss . Wie viele Einstellmöglichkeiten gibt es?

Lösung: je 10 Ziffern, also $10^3 = 1000$ Einstellmöglichkeiten.

Verallgemeinerung:

Es gibt n^k Möglichkeiten, k-Tupel einer n-elementigen Menge zu bilden (Anordnung von k Elementen einer n-elementigen Menge mit Wiederholung).

- Wie viele „Wörter" lassen sich aus den vier Buchstabensteinen $\boxed{A}$, $\boxed{B}$, $\boxed{E}$, $\boxed{R}$ bilden?

Lösung: 4 Möglichkeiten für den 1. Stein,
3 Möglichkeiten für den 2. Stein,
2 Möglichkeiten für den 3. Stein,
1 Möglichkeit für den 4. Stein,
also $4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24$ Möglichkeiten.

Verallgemeinerung:

Es gibt $n!$ Möglichkeiten, die Elemente einer n-elementigen Menge ohne Wiederholung anzuordnen (Permutationen (ohne Wiederholung)).

- Beim Rennquintett muss man die Reihenfolge der ersten 3 Pferde von 15 Pferden beim Zieleinlauf vorhersagen. Wie viele Möglichkeiten gibt es?

Lösung: 15 Möglichkeiten für Platz 1,
14 Möglichkeiten für Platz 2,
13 Möglichkeiten für Platz 3,
also $15 \cdot 14 \cdot 13 (= 2730)$ Möglichkeiten.

Zur Struktur:

$$15 \cdot 14 \cdot 13 = \frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot \ldots \cdot 1}{12 \cdot 11 \cdot 10 \cdot \ldots \cdot 1} = \frac{15!}{12!} = \frac{15!}{(15-3)!}$$

Verallgemeinerung:

Es gibt $n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \dfrac{n!}{(n-k)!}$ Möglichkeiten k Elemente aus einer n-elementigen Menge $(k \le n)$ ohne Wiederholung anzuordnen (k-Permutationen einer n-elementigen Menge).

• Wie viele verschiedene 6er-Tipps gibt es beim Zahlenlotto „6 aus 49"?

Lösung: Zunächst zieht man 6 Kugeln der Reihe nach: $49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44$ Möglichkeiten. Bei einem Tipp kommt es nicht auf die Reihenfolge an, das heißt „1, 3, 5, 20, 40, 49" und „49, 1, 20, 40, 3, 5" sind derselbe Tipp. Genauer führen alle 6er Permutationen mit denselben 6 Ziffern, also 6!, zum selben Tipp: Es gibt also insgesamt

$$\frac{49 \cdot 48 \cdot \ldots \cdot 44}{6!} = \frac{49!}{43! \cdot 6!} = \frac{49!}{(49-6)!} = \binom{49}{6} = 13.983.816$$

verschiedene Tips beim Lotto. Diese oft benötigten Zahlen vom Typ

$$\binom{49}{6} = \frac{49!}{(49-6)! \cdot 6!} \qquad \text{„49 über 6"}$$

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!} \qquad \text{„}n\text{ über }k\text{"}$$

heißen Binomialkoeffizienten.

Verallgemeinerung:

Es gibt $\binom{n}{k} = \dfrac{n!}{(n-k)! \cdot k!}$ *k*-elementige Teilmengen einer *n*-elementigen Menge.

Deutung der vier Grundaufgaben im Urnenmodell!

k mal ziehen mit Zurücklegen
k mal ziehen ohne Zurücklegen

1.4.3 Bemerkungen zu den Binomialkoeffizienten:

Der Fall „$k = n$" legt die Definition $0! = 1$ nahe.
Der Fall „$k = 0$" legt die Definition $\binom{n}{0} = 1$ nahe.

Binomische Formel $(a + b)^2 = a^2 + 2ab + b^2$

Ihre Verallgemeinerung klärt den Namen „Binomialkoeffizienten":

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \quad (a,b \in \mathbb{R},\ n \in \mathbb{N})$$

Man kann Binomialkoeffizienten im Pascalschen Dreieck anordnen:

allg.:

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

$$
\begin{array}{ccccccccc}
 & & & & \binom{0}{0} & & & & \\
 & & & \binom{1}{0} & & \binom{1}{1} & & & \\
 & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\
 & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\
\binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4}
\end{array}
$$

Jede Zahl ist die Summe der beiden darüber liegenden. Es gilt nämlich

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} \qquad (0 \le k \le n-1) \qquad \text{(nachrechnen!)}$$

Man kann offenbar auch schreiben

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

Klar ist die Symmetrie $\binom{n}{k} = \binom{n}{n-k}, k = 0,\ldots,n$

Anwendung:
Eine n-elementige Menge hat 2^n Teilmengen.

1.4.4 Anwendungsbeispiele

a.) Wie viele Möglichkeiten gibt es, 4 rote, 3 schwarze und 5 gelbe Spielsteine in eine Reihe zu legen?

Lösung: Man hat 4 + 3 + 5 = 12 Plätze zu belegen. $\binom{12}{4}$ Möglichkeiten für die roten, dann $\binom{8}{3}$ für die schwarzen; es bleiben $\binom{5}{5}$ Möglichkeiten für die gelben, also $\binom{12}{4} \cdot \binom{8}{3} \cdot \binom{5}{5} = 27720$ Möglichkeiten. Ein anderer Ansatz: $\frac{12!}{3! 4! 5!}$. Verallgemeinerung?

b.) Man hat 1 DM, 2 DM und 5 DM Münzen. Wie viele Möglichkeiten gibt es, vier Münzen in einen Automaten einzuwerfen?
Lösung: Münzen Nr. 1, 2, 3 (Geldwert irrelevant).

Tupel $T = a_1a_2a_3a_4$ mit $1 \le a_1 \le a_2 \le a_3 \le a_4 \le 3$

$$T \leftrightarrow R = \underbrace{a_1(a_2+1)(a_3+2)(a_4+3)}_{\text{Vierer-Teilmenge} \subseteq \{1,\, 2,\, ...,\, 6\}}$$

also gibt es $\binom{6}{4} = \binom{6}{2} = \frac{6 \cdot 5}{2} = 15$ verschiedene Tupel T

Verallgemeinerung: k Objekte von n Arten kombinieren (im Beispiel k = 4, n = 3). Es gibt $\binom{k+n-1}{k}$ Möglichkeiten.

Urnenmodell: je k Kugeln mit Nr. 1, 2, ..., n. Es werden auf einen Griff k Stück gezogen.

c.) Beim Zahlenlotto „6 aus 49“ gibt es folgende Ränge: (vgl. Lotto-Problem, Nr. 3, S.3)

1. Rang: 6 Richtige
2. Rang: 5 Richtige mit Zusatzzahl
3. Rang: 5 Richtige
4. Rang: 4 Richtige
5. Rang: 3 Richtige

Wie sind die Chancen für die 5 Ränge? Wie ist die Chance, keine einzige Zahl der 6 Richtigen angekreuzt zu haben?
Bemerkung: Genauer müsste man für den 3. Rang sagen „5 Richtige und Zusatzzahl falsch“.

Lösung: Für die Gewinnzahlen $\{z_1, \dots, z_6\}$ gibt es $\binom{49}{6}$= 13.983.816 Ziel-Möglichkeiten.

Die Chance für den ersten Rang ist also $\frac{1}{\binom{49}{6}} = \frac{1}{13.983.816} \approx 0{,}7 \cdot 10^{-7}$

Die Anzahl der Möglichkeiten für genau k Richtige ($k = 0, \dots, 6$): Dann müssen k Zahlen in der Gewinnmenge $\{z_1, \dots, z_6\}$ und $6 - k$ Zahlen in der Restmenge $\{z_7, \dots, z_{49}\}$ liegen.

Dafür gibt es $\binom{6}{k} \cdot \binom{43}{6-k}$ Möglichkeiten.

6 Richtige (1. Rang): $\binom{6}{6} \cdot \binom{43}{0} = 1$ Möglichkeit.

5 Richtige: $\binom{6}{5}\cdot\binom{43}{1} = 6\cdot 43 = 258$ Möglichkeiten.

- <u>2. Rang</u>: Zusätzlich Zusatzzahl z_7 richtig, also
  5 Zahlen in $\{z_1,...,z_6\}$,
  Zahl $=z_7$, also $\binom{6}{5} = 6$ Möglichkeiten.

Damit Chance für 2. Rang $\dfrac{6}{\binom{49}{6}} = \dfrac{1}{2.330.636} \approx 0{,}43\cdot 10^{-6}$

- <u>3.Rang</u>: es bleiben 258 - 6 = 252 Möglichkeiten,
  damit Chance für 3. Rang $\dfrac{252}{\binom{49}{6}} \approx 0{,}00018$.

- <u>4. Rang</u>: $\binom{6}{4}\binom{43}{2} = 13.545$ Möglichkeiten, Chance $\dfrac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} \approx 0{,}0067$.

- <u>5. Rang</u>: $\binom{6}{3}\binom{43}{3} = 246.820$ Möglichkeiten, Chance $\dfrac{\binom{6}{3}\binom{43}{3}}{\binom{49}{6}} \approx 0{,}177$.

Keine richtige Zahl: $k = 0$, $\binom{6}{0}\cdot\binom{43}{6} = 6096454$ Möglichkeiten, also eine Chance von $\dfrac{\binom{6}{0}\cdot\binom{43}{6}}{\binom{49}{6}} \approx 0{,}436$.

d.) Fußball-Toto: bei der 11er Wette ist für 11 Begegnungen vorherzusagen, ob die Heimmannschaft gewinnt (1), verliert (2) oder unentschieden spielt (0). Wie viele Tipps mit x Fehlern
(0 ≤ x ≤ 11) gibt es?

Lösung:
Ein Tipp ist ein 11-er Tupel über der Menge {0, 1, 2}, also insgesamt 3^{11} = 177.147 mögliche Tipps. Es seien x Ergebnisse falsch, 11-x richtig vorhergesagt. Es gibt $\binom{11}{x}$ Möglichkeiten, x der 11 Ergebnisse falsch vorherzusagen. Bei jedem falschen Ergebnis gibt es noch 2 Möglichkeiten für "falsch", also $\binom{11}{x}\cdot 2^{x}$ falsche Tipps mit x falschen Vorhersagen.

x	0	1	2	3	4	11
$\binom{11}{x}\cdot 2^x$	1	22	220	1.320	5.280	2.048

Beachte: Der Ansatz einer gleichen Chance bei allen Tipps ist nur bei einem absoluten Fußball-Laien sinnvoll (Modellansatz!).

e.) Rubbellose (in Baden-Württemberg):
Die bis Mai 1986 gültigen Teilnahmebedingungen der seit dem 19.02.86 laufenden staatlichen Losbrieflotterie Baden-Württemberg besagen, dass die Lotterie in Serien von jeweils 2 Millionen Losen zu je 1 DM aufgelegt wird. Weiter heißt es, dass die Gewinne einer Serie nach folgendem Gewinnplan ausgeschüttet werden:

Anzahl der Gewinne	Einzelgewinn	Gewinnsumme insgesamt
2	25 000 DM	50 000 DM
2	10 000 DM	20 000 DM
10	1 000 DM	10 000 DM
400	100 DM	40 000 DM
8 000	10 DM	80 000 DM
40 000	5 DM	200 000 DM
80 000	2 DM	160 000 DM
240 000	Freilos	240 000 DM

Das einschlägige Gesetz verlangt eine Auszahlquote von 40%. Entspricht dem der Spielplan?

Lösung: Das Finanzministerium berechnet gemäß dem Gesetz eine Gewinnsumme von 40%:

Einnahme: 2 000 000 DM
Gewinnsumme 800 000 DM (Summe der rechten Spalte)

also 40 %, wie das Gesetz es verlangt.

In Wirklichkeit werden die Freilose nicht bezahlt sondern sofort gegen ein anderes Los getauscht,
d. h. die Einnahmen sind nur 1 760 000 DM, die Auszahlungen nur 560 000 DM, also 31,82%.

Nach einem interessanten Briefwechsel, in den auch ein Leistungskurs Stochastik eines Heidelberger Gymnasiums eingeschaltet war, musste das Finanzministerium Baden-Württemberg den Gewinnplan abändern. Nach dem neuen Gewinnplan ist der Gewinner eines Freiloses nicht mehr gezwungen, dieses gegen ein Freilos einzutauschen. Statt dessen kann sich der Spieler auch 1 DM auszahlen lassen. Der wohl weitaus größte Teil der Spieler wird jedoch weiterhin einen Freilosgewinn gegen ein neues Los eintauschen, also auf die Auszahlung von 1 DM verzichten. Für solche Spieler bleibt letztlich alles beim Alten, nach wie vor haben sie eine Auszahlungserwartung von 31,82% des eingesetzten Betrages.
Spielerfreundlicher wäre die Regelung, an Stelle der 240 000 Freilosgewinne etwa 240 weitere Gewinne zu je 1 000 DM und 239 760 zusätzliche Nieten einzuführen. Die allermeisten Gewinner von 1 000 DM dürften nämlich ihren Gewinn nicht in weiteren Losen investieren.

Der Leser möge nun selbst darüber befinden, ob der neue Gewinnplan die Spieler manipuliert:

Spielbedingungen ab 09.11.88

§ 5 Gewinnausschüttung, Gewinnauszahlung

(1) Das Spielkapital einer Serie beträgt zwei Millionen DM. Davon werden 42,5 v. H. nach folgendem Gewinnplan ausgeschüttet:

Anzahl der Gewinne	Einzelgewinn	Gesamt-Gewinnsumme
1	50 000 DM	50 000 DM
1	10 000 DM	10 000 DM
1	5 000 DM	5 000 DM
10	1 000 DM	10 000 DM
50	500 DM	25 000 DM
300	100 DM	30 000 DM
4 000	20 DM	80 000 DM
8 000	10 DM	80 000 DM
2 000	5 DM	160 000 DM
80 000	2 DM	160 000 DM
240 000	1 DM	240 000 DM
364 363		850 000 DM

(2) Gewinne bis einschließlich 100 DM werden in der Annahmestelle, in der das Gewinnlos erworben wurde, gegen Rückgabe des Loses ausgezahlt.
(3) Gewinne ab 500 DM werden von der Staatlichen Sport-Toto GmbH, Postfach 10 43 52, 7000 Stuttgart 10, nach Rückgabe des Gewinnloses über die Annahmestelle, in der das Gewinnlos erworben wurde, oder nach Einsendung des Gewinnloses an die Gesellschaft zugestellt. Die Annahmestelle bestätigt den Erhalt des Gewinnloses, ohne damit zugleich den Gewinnanspruch anzuerkennen.
(4) Die Zustellung bzw. Auszahlung der Gewinne durch die Gesellschaft bzw. ihre Annahmestellen erfolgt mit befreiender Wirkung an den Besitzer des Gewinnloses.

1.5 Der Wahrscheinlichkeitsbegriff

1.5.1 Zufallsexperimente

Unterscheide

- Experimente der Physik (unter gleichen Bedingungen beliebig wiederholbar, (starker) Determinismus)
- Zufallsexperimente (kein kausaler Zusammenhang, Ausgang unvorhersehbar, nur von Zufall abhängig). Beschreibung durch Festlegung der zu beobachtenden Merkmalsausprägungen ((zunächst) endlich viele Ergebnisse, Ausgänge oder Ausfälle). Diese werden in der Menge Ω aller möglichen Ergebnisse, dem Ergebnisraum zusammengefasst.

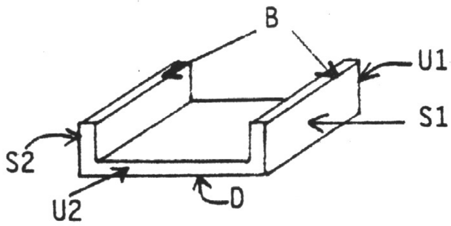
Wichtig: Die Wahl von Ω bei einem Zufallsexperiment hängt auch vom Interesse des „Experimentators" ab. Durch Festlegung von Ω werden stochastische Realsituationen mathematisch modelliert.

Beispiele für Zufallsexperimente:

a.) Riemer-U-Würfel werfen

$$\Omega_1 = \{S_1, S_2, U_1, U_2, B, D\}$$

$$\Omega_2 = \{S, U, B, D\}$$



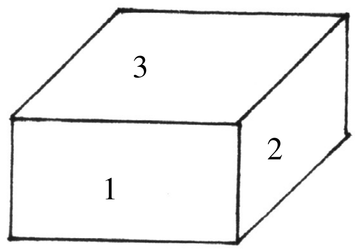
b.) Riemer-Quader werfen

$$\Omega_1 = \{1,2,3,4,5,6\}$$

$$\Omega_2 = \{\boxed{1}, \boxed{6}, \boxed{2}, \boxed{5}, \boxed{3}, \boxed{4}\}$$

$$\Omega_3 = \{\text{gerade, ungerade}\}$$

$$\Omega_4 = \{\text{Primzahl, keine Primzahl}\}$$



c.) Würfeln mit Farbwürfel

$$\Omega = \{\text{gelb, violett, grün, schwarz, blau, rot}\}$$

d.) 2 Münzen werfen

$\Omega_1 = \{WW, ZZ, WZ\}$

$\Omega_2 = \{W_1W_2, Z_1W_2, W_1Z_2, Z_1Z_2\}$

$\Omega_3 = \{0, 1, 2\}$ (Anzahl von „Wappen")

$\Omega_4 = \{\text{Ja, Nein}\}$ (Ausgang gleich?)

Die Mengenschreibweise für den Ergebnisraum ist sehr praktisch. Betrachtet man zum Beispiel beim Würfeln mit einem normalen Würfel das Ereignis „eine Primzahl fällt", so kann man dies als Teilmenge $E_1 = \{2, 3, 5\}$ des Ergebnisraums $\Omega = \{1, 2, 3, 4, 5, 6\}$ beschreiben. Ein Ergebnis, zum Beispiel „es fällt die 6", wird durch eine einelementige Menge, das Elementarereignis $E_2 = \{6\}$ beschrieben.
Wir identifizieren in der mathematischen Modellierung also Ereignisse mit den sie beschreibenden Ergebnismengen, so dass „Ereignis" ein Begriff im mathematischen Modell wird.

Definition der üblichen Begriffe:

a.) Ergebnisraum eines Zufallsexperiments: Menge Ω aller möglichen Ergebnisse des Zufallsexperiments.
b.) Ereignis: Teilmenge $E \subseteq \Omega$, Ereignisraum: (Ω) Potenzmenge von Ω, Elementarereignis: Einelementige Teilmenge $\{\omega\} \subset \Omega$
c.) Sicheres Ereignis Ω. Unmögliches Ereignis $\varnothing$.
d.) Unvereinbare Ereignisse: $E_1, E_2 \subseteq \Omega$ mit $E_1 \cap E_2 = \varnothing$.
e.) Gegenereignis eines Ereignisses: $E \subseteq \Omega$: Komplement $\overline{E} = \Omega \setminus E$.
f.) Und-Ereignis zweier Ereignisse: $E_1, E_2 \subseteq \Omega : E_1 \cap E_2$. Oder-Ereignis: $E_1 \cup E_2$.

Wichtige Sprechweise: Man sagt, dass bei einem Zufallsexperiment „das Ereignis *E* eingetreten" ist, wenn das aufgetretene Ergebnis ω zu *E* gehört: $\omega \in E$.

Beispiel: Würfel mit $\Omega = \{1,2,3,4,5,6\}$

$E_1 = \{2,4,6\}$" Augenzahl gerade"; $E_2 = \{3,6\}$ "3 teilt Augenzahl".
Und-Ereignis $E_1 \cap E_2 = \{6\}$ "Augenzahl gerade und durch 3 teilbar".
Oder-Ereignis $E_1 \cup E_2 = \{2,3,4,6\}$ "Augenzahl gerade oder durch 3 teilbar".
Gegenereignis $\overline{E_1} = \{1,3,5\}$ „Augenzahl ungerade"

1.5.2 Häufigkeiten und Wahrscheinlichkeiten

Beispiel: Ein Riemer-U-Würfel wurde 2000 Mal mit folgenden Ergebnissen geworfen:

Obenliegende Seite	D	S1	S2	U1	U2	B
Absolute Häufigkeit	526	177	195	68	92	942
Relative Häufigkeit	26.3%	8.85%	9.75%	3.4%	4.6%	47.1%

Stelle die Häufigkeiten graphisch dar (Kreisdiagramm, Säulendiagramm,...).

Allgemein: Ein Zufallsexperiment wird n-mal durchgeführt, dabei tritt das Ereignis E k-mal auf.

$H_n(E) = k$ absolute Häufigkeit von E,

$h_n(E) = \frac{k}{n}$ relative Häufigkeit von E.

Eigenschaften von relativen Häufigkeiten (klar!):

(1) $0 \le h_n(E) \le 1$ für alle $E \subseteq \Omega$.

(2) $h_n(\Omega) = 1$; $h_n(\varnothing) = 0$.

(3) $h_n(\overline{E}) = 1 - h_n(E)$ für alle $E \subseteq \Omega$.

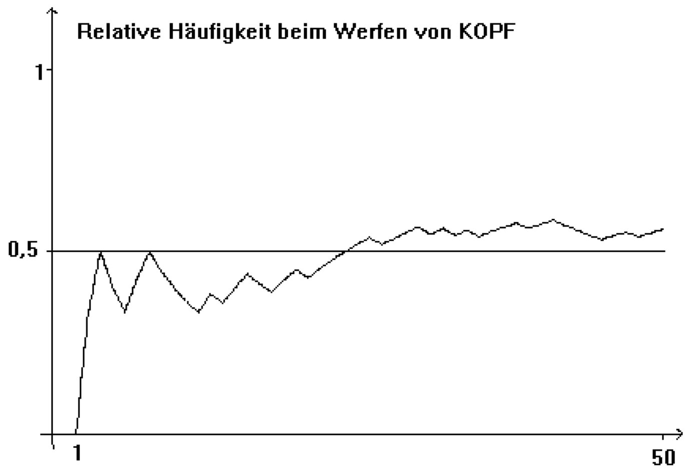
(4) $h_n(E_1 \cup E_2) = h_n(E_1) + h_n(E_2)$ für $E_1 \cap E_2 = \varnothing$; speziell:

$h_n(E) = \sum_{i=1}^{m} h_n(\{\omega_i\})$ für $E = \{\omega_1, ..., \omega_m\}$; insbesondere: $\sum_{\omega \in \Omega} h_n(\{\omega\}) = 1$.

Empirisches Gesetz der großen Zahlen:

Mit wachsender Versuchszahl stabilisiert sich die relative Häufigkeit eines gegebenen Ereignisses im Allgemeinen bei einer bestimmten Zahl p, die relativen Häufigkeiten unterscheiden sich immer weniger von dieser Zahl.

Beispiel: Münze werfen, E = „Kopf oben“



Diese Zahl p kann man sinnvollerweise als die Wahrscheinlichkeit $p = P(E)$ von E ansehen. Relative Häufigkeiten bieten demnach gute Näherungs- oder Schätzwerte für Wahrscheinlichkeiten. Man bezeichnet Wahrscheinlichkeiten, die man derart a posteriori aus relativen Häufigkeiten erhält, auch als empirische (oder statistische) Wahrscheinlichkeiten.

Mit Hilfe von Wahrscheinlichkeiten kann man dann auch vernünftige Prognosen abgeben: Wenn ich ein Zufallsexperiment n mal durchführe, so erwarte ich, dass ein bestimmtes Ereignis E mit Wahrscheinlichkeit $p = P(E)$ dabei ungefähr $n \cdot p$- mal eintritt ($\frac{k}{n} \approx p$, somit $k \approx n \cdot p$). Dies gilt natürlich nur für „große“ n.

Beispiel: Beim Riemer-U-Würfel kann man aufgrund der Daten in der Tabelle und unter Beachtung der Teilsymmetrien zum Beispiel ansetzen

$P(D) = 26{,}3\ \%$, $P(S_1) = P(S_2) = 9{,}3\ \%$, $P(U_1) = P(U_2) = 4{,}0\ \%$, $P(B) = 47{,}1\ \%$.

Beachte: Diese Zahlen sind subjektive Schätzungen der „innewohnenden“, unbekannten Wahrscheinlichkeiten! Der Wahrscheinlichkeitsbegriff ist noch nicht formal definiert worden.

Die Idee von Richard von Mises, die Wahrscheinlichkeit als Grenzwert der relativen Häufigkeiten zu definieren, wirft unlösbare mathematische Probleme auf, insbesondere:

- Nur endliche Anzahl von Versuchen möglich.
- Stabilisierung der relativen Häufigkeiten nur „i. a.“, da beliebig viele Ausreißer möglich, d. h. keine Konvergenz im analytischen Sinne (vgl. Folie).

1.5.3 Gleichwahrscheinlichkeit

1.5.3.1 Laplace-Experimente

Bei manchen Zufallsexperimenten kann man aus Symmetrie- (oder Indifferenz-)Gründen alle Ergebnisse a priori als „gleichberechtigt“ ansehen.

Beispiele: Werfen eines („guten“) Würfels, Lotto, Ziehen aus einer Urne,...

Experimente, denen diese „Laplace-Annahme“ zugesprochen wird, heißen Laplace-Experimente; der zugehörige Ergebnisraum wird dann auch Laplace-Raum genannt.

Die Laplace-Annahme läßt sich nicht beweisen, sondern nur aufgrund von Versuchen mehr oder weniger zuverlässig bestätigen (Stabilisieren sich die relativen Häufigkeiten sämtlicher Elementarereignisse bei derselben Zahl?). Es handelt sich also um eine Modell-Annahme.

Definition:

Sei Ω ein endlicher Laplace-Raum, und sei $E \subseteq \Omega$ ein Ereignis.
Die Zahl $P(E) = \frac{\lvert E \rvert}{\lvert \Omega \rvert}$ heißt die Laplace-Wahrscheinlichkeit von E.

Man nennt die Elemente von E auch die „für E günstigen“ (Anzahl $g = \lvert E \rvert$) und die Elemente von Ω die „für E möglichen“ Ergebnisse (Anzahl $m = \lvert \Omega \rvert$). Daher sagt man kurz

$$P(E) = \frac{g}{m} = \frac{\text{Anzahl der günstigen}}{\text{Anzahl der möglichen}}.$$

1.5.3.2 Beispiele für Laplace-Experimente

a.) Würfel:

$\Omega = \{1, ..., 6\}$, $E = \{2, 4, 6\}$ „Augenzahl gerade“

$$p = P(E) = \frac{|E|}{|\Omega|} = \frac{3}{6} = 0{,}5\,.$$

b.) 3-Türen-Problem (Problem 6, S. 3):

P(„Nicht-Wechsel“) = $\frac{1}{3}$, denn der Kandidat wählt gleichberechtigt zwischen 3 gleichberechtigten Möglichkeiten, eine davon ist günstig. Der Quizmaster ist irrelevant.

P(„Wechsel“) = $\frac{2}{3}$, denn der Kandidat gewinnt genau dann, wenn er ursprünglich eine Niete gewählt hat, d. h. in zwei von drei Fällen. Stochastisch gleichwertiges Experiment: Der Kandidat wählt für sich zwei Türen und nennt dem Quizmaster die dritte!

P(„Münze werfen, ob wechseln“) = $\frac{1}{2}$, denn der Kandidat wählt jetzt gleichberechtigt zwischen zwei Türen, die ursprüngliche Wahl ist irrelevant.

c.) Lotto-Problem (Problem 3, Seite 3, vgl. auch Beispiel c (1.4.4, Seite 25 und 26). Die dort berechneten „Chancen“ entsprechen genau dem Modell der Laplace-Wahrscheinlichkeiten.

d.) Werfen von 2 Münzen, $\Omega = \{WW, WZ, ZZ\}$

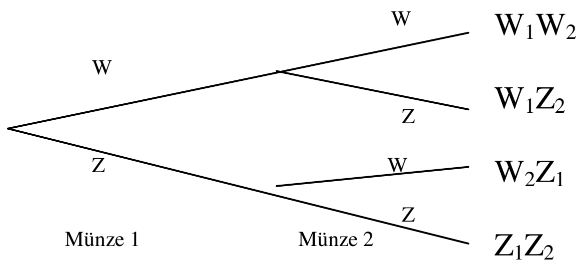
1. Idee: Laplace–Annahme

$P(\{WW\}) = P(\{WZ\}) = P(\{ZZ\}) = \frac{1}{3}$

Versuche zeigen: Die relativen Häufigkeiten stabilisieren sich nicht bei diesen Zahlen, die Laplace–Annahme ist nicht vernünftig. Die Menge Ω lässt sich jedoch zu einem Laplace–Raum verfeinern, indem WZ in W_1Z_2 und W_2Z_1 aufgeteilt wird. Der neue Raum ist, wie Versuche zeigen, sinnvoll als Laplace–Raum zu sehen.

$\Omega' = \{W_1W_2, W_1Z_2, W_2Z_1, Z_1Z_2\}$ mit $P(\{W_1W_2\}) = ... = \frac{1}{4}$.

Dies bestätigt auch ein Baum bzw. eine Tabelle aller Möglichkeiten:



Damit vernünftiger Ansatz

P({WW}) = P({ZZ}) = $\frac{1}{4}$, P({WZ}) = $\frac{1}{2}$.

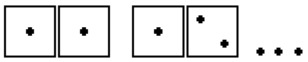
e.) Beim Werfen von 2 Würfeln sind nach folgendem Argument die Summe 11 und 12 gleich wahrscheinlich:

$\Omega = \{ab / a, b \in \{1, ..., 6\}, a \le b\}$,

Augensumme 11: $E_1 = \{56\}$,

Augensumme 12: $E_2 = \{66\}$.

Dies entspricht aber nicht der Spielwirklichkeit, wo 11 viel häufiger vorkommt. Folglich ist Ω kein Laplace–Raum. Verbesserung durch unterscheidbare Würfel und Fälle:



Also $\Omega' = \{ab / a, b \in \{1, ..., 6\}\}$, $|\Omega| = 36$

Augensumme 11: $E_1' = \{56, 65\}$,

Augensumme 12: $E_2' = \{66\}$,

also $P(E_1') = \frac{2}{36} \approx 0{,}056$, $P(E_2') = \frac{1}{36} \approx 0{,}028$

f.) 1. De-Méré-Problem (Analog zu e)

Nach folgender Überlegung von de Méré sind beim Werfen dreier Würfel die Augensummen 11 und 12 gleich wahrscheinlich:

$\Omega = \{abc / a, b, c \in \{1, ..., 6\}, a \le b \le c\}$

Abzählen von Ω:

$\binom{6}{3} = 20$ Möglichkeiten aus 3 verschiedenen Ziffern,

$2 \cdot \binom{6}{2} = 30$ Möglichkeiten aus 2 verschiedenen Ziffern,

6 Möglichkeiten aus 1 Ziffer, also $|\Omega| = 56$.

Augensumme 11: $E_1 = \{146, 155, 236, 245, 335, 344\}$,

Augensumme 12: $E_2 = \{156, 246, 255, 336, 345, 444\}$,

also $P(E_1) = P(E_2) = \frac{6}{56} \approx 0{,}107$.

In der Realität ist aber E_1 etwas häufiger als E_2! Das heißt, dass Ω kein Laplace-Raum ist. Werden die 3 Würfel als unterscheidbar angenommen, so erhält man den Laplace-Raum:

$\Omega' = \{abc / a, b, c \in \{1, ..., 6\}\}$ mit $|\Omega'| = 6^3 = 216$.

Jetzt erhält man Augensumme 11: $E_1' = \{146, 164, 461, 416, 614, 641, 155, ...\}$

Augensumme 12: $E_2' = \{156, 165, 516, 561, 615, 651, ...\}$

Mit $|E_1'| = 27$, $|E_2'| = 25$,

also $P(E_1') = \frac{27}{216} \approx 0{,}125$, $P(E_2') = \frac{25}{216} \approx 0{,}116$.

g.) 2. De Méré-Problem (Problem 4, Seite 3)

4 mal Werfen mit einem Würfel:

$\Omega_1 = \{1, ..., 6\}^4$, $|\Omega_1| = 6^4 = 1296$

E_1 „Bis zum 4. Wurf mindestens eine 6“

$\overline{E_1} = \{1, ..., 5\}^4$, $|\overline{E_1}| = 5^4 = 625$,

$P(E_1) = 1 - \frac{625}{1296} = \frac{671}{1296} \approx 0{,}518$.

24 mal mit Doppelwürfel werfen:

$\Omega_2 = \left[\{1, ..., 6\}^2\right]^{24}, |\Omega_2| = \left(6^2\right)^{24} = 36^{24}$

E_2: „bis zum 24. Wurf mindestens eine Doppelsechs“

$\overline{E_2} = \{\{1, ..., 6\}^2 \backslash \{6{,}6\}\}^{24} = 35^{24}$

$P(E_2) = 1 - \frac{35^{24}}{36^{24}} \approx 0{,}491$

Der einfache Proportionalschluss von de Méré ist also falsch!

h.) Teilungsproblem (Problem 5, Seite 3)

Aufteilung des Geldpreises z. B.

a) 1:1 aufgrund von Unwissenheit;
b) 4:3 entsprechend Spielstand;
c) 2:1 entsprechend fehlenden Siegen.

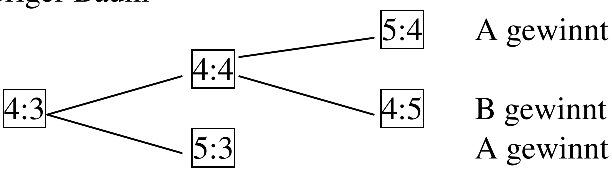
Bei gleicher Spielstärke der beiden Spieler erscheint folgende Möglichkeit angemessen:

d) Spieler A führe 4:3 gegen Spieler B. Zweimal Münzwerfen simuliere Fortsetzung des Wettkampfs. „*W*“ bedeute „Spieler A gewinnt“, „*Z*“ bedeute „Spieler B gewinnt“. Klar: Genau dann wird Spieler B Gesamtsieger, wenn „*ZZ*“ fällt. Also

   P(„Spieler B Gesamtsieger“) = $\frac{1}{4}$ und P („Spieler A Gesamtsieger“) = $\frac{3}{4}$.

   Daher Aufteilung 3:1 , d.h. 75 000.- DM für Spieler A und 25 000.- DM für Spieler B.

   Zugehöriger Baum



1.5.3.3 Eigenschaften von Laplace-Wahrscheinlichkeiten

Aufgrund der Definition der Laplace-Wahrscheinlichkeiten sind diese Eigenschaften trivial. Sie sind später Grundlage für die Definition des allgemeinen Wahrscheinlichkeitsbegriffs.

(1) $0 \le P(E) \le 1$ für alle $E \subseteq \Omega$.

(2) $P(\Omega) = 1;\ P(\emptyset) = 0$.

(3) $P\left(\overline{E}\right) = 1 - P(E)$ für alle $E \subseteq \Omega$.

(4) $P\left(E_1 \cup E_2\right) = P\left(E_1\right) + P\left(E_2\right)$ für $E_1 \cap E_2 = \emptyset$; speziell:

$$P(E) = \sum_{i=1}^{m} P(\{\omega_i\}) \text{ für alle } E = \{\omega_1, ..., \omega_m\};\ \text{insbesondere: } \sum_{\omega \in \Omega} P(\{\omega\}) = 1.$$

1.5.3.4 Geometrische Wahrscheinlichkeiten

Sie sind eine Verallgemeinerung der Laplace-Wahrscheinlichkeit.
Beispiele:

a.) Drehen eines Glücksrades:

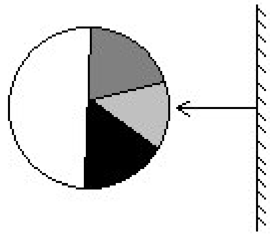
Man kann z. B. sinnvoll festlegen:

P(„Pfeil zeigt auf schwarz“)

$$= \frac{\text{Flächeninhalt schwarzer Sektor}}{\text{Flächeninhalt Kreis}}$$

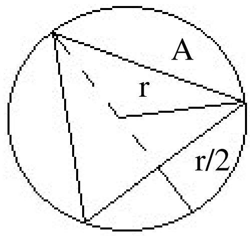
$$= \frac{\text{Winkel im schwarzen Sektor}}{360^\circ}$$

$$= \frac{\text{Bogenlänge zu schwarzem Sektor}}{\text{Kreisumfang}}.$$



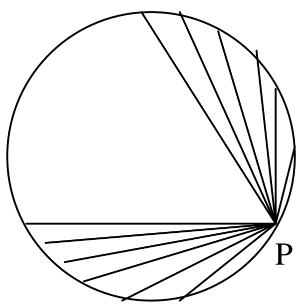
b.) Bertrand-Problem (Problem 11, Seite 3)

Kreisradius sei r. Man überlegt elementargeometrisch sofort: Seite des gleichseitigen Dreiecks halbiert Radius, Seitenlänge ist
$a = \sqrt{3}r$ (nach Pythagoras).



Die gesuchte Wahrscheinlichkeit sei p.
Wir werden 3 verschiedene Modellierungen machen:

1)

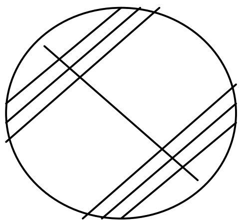


a) Wähle P auf Kreis beliebig. Klar: Alle Sehnen kürzer als a, die von P ausgehen, enden auf Kreisbogen, der $\frac{2}{3}$ des Umfangs ausmacht (entsprechende Argumentation für Winkel). Also: $p = \frac{2}{3}$

b) Anders: Alle Sehnen kürzer a, die von P ausgehen, liegen in 2 Kreisabschnitten über den Dreiecksseiten. Also:

$$p = \frac{\text{Fl Kreisabschnitte}}{\text{Fl Kreis}} = \frac{\frac{2}{3}(\pi r^2 - \overbrace{\frac{1}{2}a \cdot \frac{3}{2}r}^{A_\Delta})}{\pi r^2} = \frac{2}{3} - \frac{\sqrt{3}}{2\pi} \approx 0.391.$$

2)



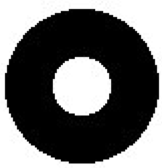
Wähle beliebigen Durchmesser. Klar: Alle Sehnen kürzer als a und senkrecht zum Durchmesser liegen in dessen 1. Und 4. Viertel.

Also: $p = \frac{1}{2}$.

Anders: Entsprechend über Flächen (wie eben!).

Also: $p = \frac{2}{3} - \frac{\sqrt{3}}{2\pi}$.

3)



Klar: Alle Sehnen kürzer als a liegen im Kreisring mit innerem Radius $\frac{r}{2}$ (Inkreis!).

Also: $p = \frac{\text{Fl Kreisring}}{\text{Fl Kreis}} = \frac{\pi r^2 - \frac{1}{4}\pi r^2}{\pi r^2} = \frac{3}{4}$.

Nicht-Eindeutigkeit der Lösung resultiert aus Nicht-Eindeutigkeit der Fragestellung: Was heißt „ganz zufällig“?

1.5.4 Wetten und subjektive Wahrscheinlichkeiten

Teilungsproblem, Spielstand 4 : 3, Spiel soll fortgesetzt werden. Ich wette 10 DM auf den Gesamtsieg von Spieler B. Was muss mein Freund auf den Sieg von Spieler A setzen, damit die Wette fair ist? Vorausgesetzt sind gleiche Spielstärken der Spieler A und B).

Chancenverhältnis: P („B gewinnt“) : P („A gewinnt“) = $\frac{1}{4} : \frac{3}{4} = 1 : 3$.

Also muss der Freund $3 \cdot 10$ DM = 30 DM setzen, damit sich Gewinn und Verlust langfristig ausgleichen (im Mittel verliere ich bei 4 Spielen 3 mal 10 DM und gewinne einmal 30 DM). Meine „Wettquote“, d.h. das Verhältnis von Auszahlung (ohne Einsatz) und Einsatz im Gewinnfalle, beträgt 3 : 1, also das Reziproke des Chancenverhältnisses.

Verallgemeinerung:

Sei $P(E) = \frac{a}{c}$. Ich wette auf E. Es gilt $P(\overline{E}) = 1 - P(E) = 1 - \frac{a}{c} = \frac{b}{c}$ ($b = c - a$).

Somit beträgt mein Chancenverhältnis

$P(E) : P(\overline{E}) = \frac{a}{c} : \frac{b}{c} = a : b$ (wobei $a + b = c$).

Eine faire Wettquote ist dann offensichtlich $b : a$, d.h. pro eingesetzte DM erhalte ich im Gewinnfall $\frac{b}{a}$ DM ausbezahlt (und zudem meinen Einsatz zurück).

Umgekehrt: Beträgt die Wettquote $b : a$, so ist das Chancenverhältnis $P(E) : P(\overline{E}) = a : b$, und dies liefert wegen $P(\overline{E}) = 1 - P(E)$ nach leichter Rechnung $P(E) = \frac{a}{a+b}$ (und $P(\overline{E}) = \frac{b}{a+b}$).

Sei $E \subseteq \Omega$ ein Ereignis. Folgende Aussagen sind gleichwertig: (1) Die Wettquote beim Wetten auf E beträgt $b : a$. (2) Das Chancenverhältnis ist $P(E) : P(\overline{E}) = a : b$. (3) Es ist $P(E) = \frac{a}{a+b}$.

Man kann Wahrscheinlichkeiten also auch über Chancenverhältnisse oder (noch besser) über Wettquoten quantifizieren (das eine reziprok zum anderen). Das tut man insbesondere bei Ereignissen, die nicht zu wiederholbaren Zufallsexperimenten gehören, sondern einmalig sind. Diese Wahrscheinlichkeiten drücken dann einen individuellen, subjektiven Vertrauensgrad in das Eintreten solcher Ereignisse aus. Man nennt solche Wahrscheinlichkeiten <u>subjektive Wahrscheinlichkeiten</u>.

Natürlich darf man Chancenverhältnisse bzw. Wettquoten (und damit subjektive Wahrscheinlichkeiten) nicht völlig beliebig festsetzen. Z.B. muß sinnvollerweise stets $P(\overline{E}) = 1 - P(E)$ gelten; ist meine Wettquote für E dann $b : a$, so muss ich für $\overline{E}$ eine Wettquote $a : b$ nehmen.

Beispiele:

a.) „Ich wette 3 : 2, dass Pete Sampras 1998 Wimbledon gewinnt."
D.h. ich quantifiziere meine subjektive Wahrscheinlichkeit für E: „P. S. gewinnt" zu $P(E) = \frac{2}{2+3} = \frac{2}{5} = 40\%$

b.) Wettervorhersage: „Die Regenwahrscheinlichkeit für morgen beträgt 30 %."
E: „es regnet", $P(E) : P(\overline{E}) = 3 : 7$.

<u>1.5.5 Ein Axiomensystem für Wahrscheinlichkeiten</u>

Der bisherige Wahrscheinlichkeitsbegriff ist bisher inhaltlich mit der Realität verbunden: „objektivistisch" mit Häufigkeiten und Laplaceannahmen und „subjektivistisch" als individuelle Prognose. Um eine heutigen mathematischen Ansprüchen genügende <u>Definition</u> der Wahrscheinlichkeit zu erhalten, muss diese ontologische Bindung aufgegeben werden. Dies geschieht durch den Ansatz von Kolmogorov (vgl. Axiomensystem von Hilbert für die Geometrie).

Definition: Axiomensystem von Kolmogorov

Gegeben sei eine (höchstens abzählbare) Menge Ω (der Ergebnisraum eines Zufallsexperiments). Eine Funktion P: $\mathcal{P}(\Omega) \to \mathbb{R}$, die jeder Teilmenge $E \subseteq \Omega$ (d.h. jedem Ereignis) eine reelle Zahl $P(E)$ zuordnet, heißt ein <u>Wahrscheinlichkeitsmaß</u> (Wahrscheinlichkeitsverteilung) auf Ω, wenn gilt:
(1) $0 \le P(E) \le 1$ für alle $E \subseteq \Omega$.
(2) $P(\Omega) = 1$.
(3) $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ falls $E_1 \cap E_2 = \varnothing$.
Die Zahl $P(E)$ heißt dann die Wahrscheinlichkeit von $E \subseteq \Omega$.

Bemerkungen:

a) Analog zu den Hilbertschen Geometrie-Axiomen sind Kolmogorovs Axiome nicht willkürlich sondern aus den Häufigkeits-Eigenschaften abgeleitet.
b) Ω ist bei uns im Allgemeinen endlich. Bei überabzählbarem Ω (z. B. $\Omega = \mathbb{R}$) wird die Definition komplizierter.
c) Ein Wahrscheinlichkeitsmaß P ist durch seine Werte für die Elementarereignisse eindeutig festgelegt. Wird jedem $\omega \in \Omega$ eine Zahl $P(\{\omega\}) \in [0;1]$ zugeordnet, wobei gilt $\sum_{\omega\in\Omega} P(\{\omega\}) = 1$, und wird festgesetzt $P(E) := \sum_{\omega \in E} P(\{\omega\})$, so ist P ein Wahrscheinlichkeitsmaß auf Ω, und umgekehrt.
d) Einige Folgerungen aus den Axiomen:

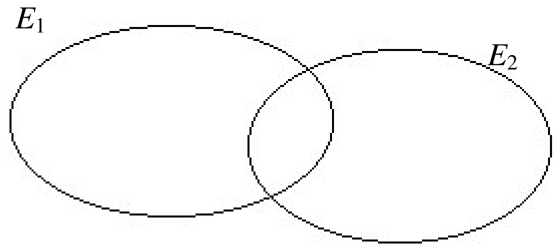
Sei P ein Wahrscheinlichkeitsmaß auf Ω. Dann gilt
(4) $P(\varnothing) = 0$.
(5) $P(\overline{E}) = 1 - P(E)$ für alle $E \subseteq \Omega$.
(6) $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$ für alle $E_1, E_2 \subseteq \Omega$.

Beweis:

Ad (4): $P(E) = P(E \cup \varnothing) \underset{(3)}{=} P(E) + P(\varnothing)$, also $P(\varnothing) = 0$.

Ad (5): $1 \underset{(2)}{=} P(\Omega) = P(E \cup \overline{E}) \underset{(3)}{=} P(E) + P(\overline{E})$, da $E \cap \overline{E} = \varnothing$.

Ad (6):



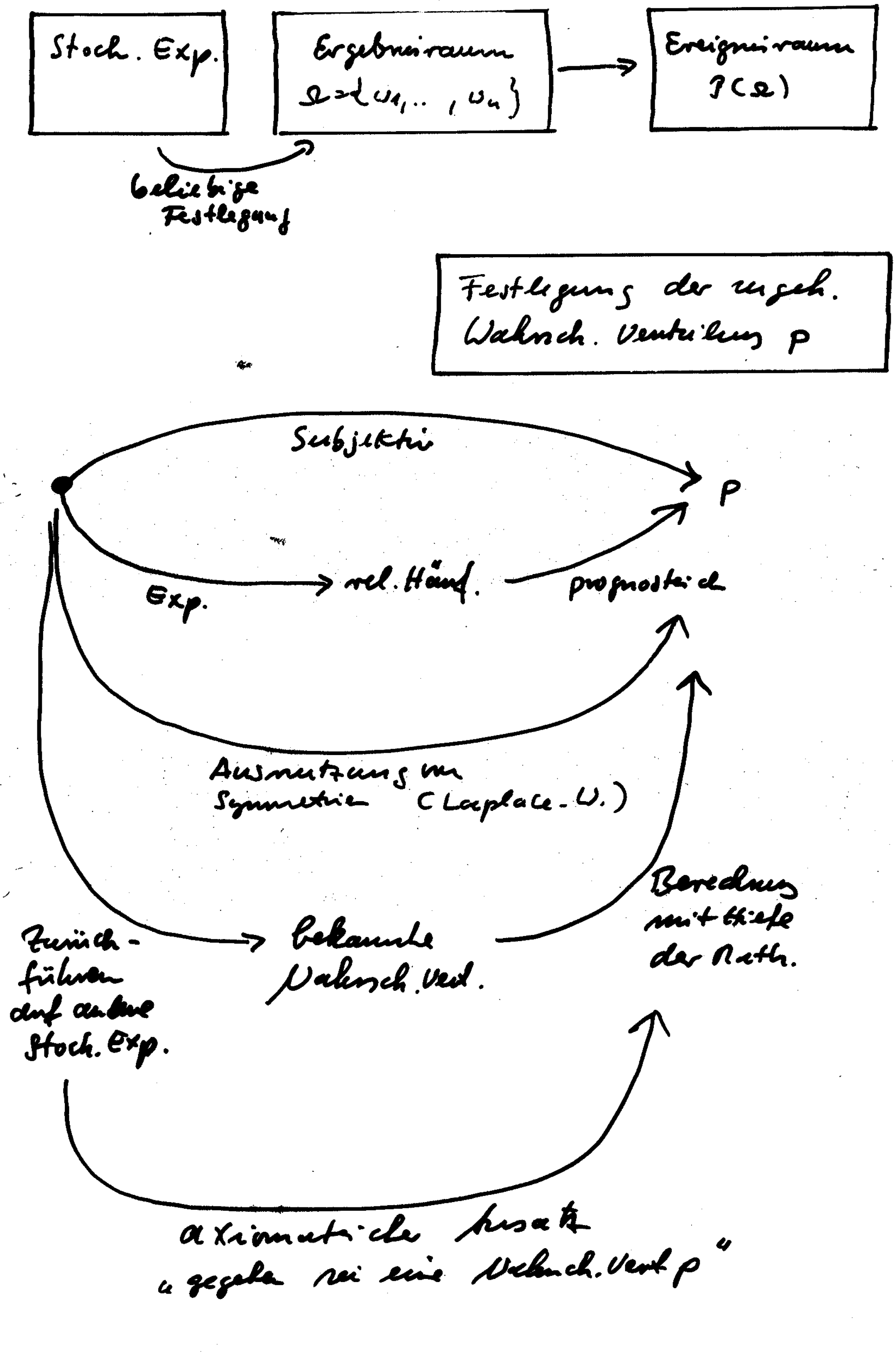
$P(E_1 \cup E_2) = P(E_1 \cup (E_2 \setminus E_1)) \underset{(3)}{=} P(E_1) + P(E_2 \setminus E_1)$ (*)

$P(E_2) = P((E_2 \setminus E_1) \cup (E_1 \cap E_2)) \underset{(3)}{=} P(E_2 \setminus E_1) + P(E_1 \cap E_2)$, also

$P(E_2 \setminus E_1) = P(E_2) - P(E_1 \cap E_2)$; Einsetzen in (*) liefert (6).

Die Eigenschaft (6) heißt $\left\{\begin{matrix}\text{allgemeiner}\\ \text{spezieller}\end{matrix}\right\}$ Additionssatz für $\left\{\begin{matrix}E_1 \cap E_2 \ne \varnothing\\ E_1 \cap E_2 = \varnothing\end{matrix}\right\}$.

a) Die folgende Folie gibt einen Überblick über die Möglichkeiten, eine Wahrscheinlichkeitsverteilung P festzulegen.



1.6 Rechnen mit Wahrscheinlichkeiten

1.6.1 Bedingte Wahrscheinlichkeit, Unabhängigkeit

Zwei- Beispiele sollen die Problematik klarmachen:

> Neulich erfuhren wir durch die Wettervorhersage, dass es am Samstag mit fünfzigprozentiger Wahrscheinlichkeit und am Sonntag ebenfalls mit fünfzigprozentiger Wahrscheinlichkeit regnen werde. Mein Freund, ein promovierter Jurist meinte daraufhin, dass es dann ja mit hundertprozentiger Wahrscheinlichkeit regnen werde.

Es bereitete mir viel Mühe, dem recht eingebildeten, aber unlogischen Juristen sein mathematisches Missverständnis klar zu machen und ihn mit Hilfe einer Münze zu überzeugen, dass es am Wochenende nur mit der Wahrscheinlichkeit $(1 - 1/2 \cdot 1/2) = 0{,}75$ regnen werde. Nachts wachte ich mit einem schlechten Gewissen auf und überlegte, wer der Dummkopf gewesen sei. Ist wirklich der doppelte Münzwurf das adäquate Modell, d.h. sind die Ereignisse "Sa mit Regen", "So mit Regen" unabhängig? Kommt nicht vielleicht ein Tiefdruckgebiet, das uns wenn nicht Samstags, so doch Sonntags mit Regen überschütten wird? Obwohl am Freitag für jeden der beiden Tage die Wahrscheinlichkeit je nur 50% ist, kann sie am Samstag, an dem es nicht regnet, für den Sonntag 100% sein. Es ist nicht so ohne weiteres klar, ob beide Ereignisse unabhängig oder abhängig sind. (Beispiel nach John A. Paulos: Zahlenblind - Mathematisches Analphabetentum und seine Konsequenzen. Heyne 1990 S. 44 f)

> Der Arzt nach der Untersuchung zu seinem Patienten: "Also, die Lage ist ziemlich ernst. Sie sind sehr krank. Statistisch gesehen überleben 9 von 10 Menschen diese Krankheit nicht." Der Patient erbleicht. "Sie haben aber Glück", beruhigt ihn der Arzt. "Ich hatte schon neun Patienten mit den gleichen Symptomen, und die sind alle tot". (Beispiel nach George Polya).

Hier wird vorgegaukelt, die Ereignisse seinen voneinander abhängig. Tausende von Lottospielern, die auf lange nicht gezogene Zahlen setzen, unterliegen demselben Irrtum, genauso wie die vielen Roulettespieler, die nach 10maligem Rouge hoffnungsvoll auf Noir setzen.

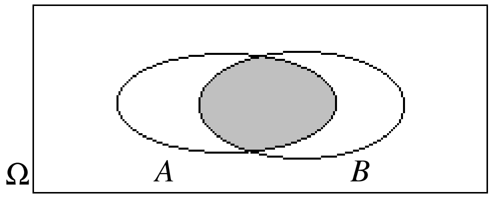
- Die umgangssprachlichen Begriffe „das eine bedingt das andere“, „das eine ist von anderen (un-)abhängig“ muss zuerst präzisiert werden (was dann zu Verständnisproblemen führen kann).
- Beide Definitionen, „bedingte Wahrscheinlichkeit“ und „Unabhängigkeit“, lehnen sich an die Verhältnisse bei Laplace-Experimenten an, wo sich der Umgangssprachliche Sinn am leichtesten präzisieren lässt.

<u>Bedingte Wahrscheinlichkeit</u>: Wie groß ist die Wahrscheinlichkeit für das Eintreffen von A, wenn ich <u>schon weiß</u>, dass B eingetreten ist? Bezeichnung: $P(A/B)$ oder $P_B(A)$.

Der Begriff „bedingte Wahrscheinlichkeit“ ist recht subtil:
- So, wie wir ihn gerade definiert haben, scheint er inhaltlich festzuliegen. Leider haben wir zunächst keine Möglichkeit, solche Wahrscheinlichkeiten zu berechnen. Ziel ist also, aus

der (vorher auch immer) gegebenen Wahrscheinlichkeitsverteilung P: $\mathcal{P}(\Omega) \to [0, 1]$ die Wahrscheinlichkeiten P(A/B) für alle A, B $\in \mathcal{P}(\Omega)$ zu berechnen.

- Beschränken wir uns auf (endliche) Laplace-Experimente, so können wir die bedingten Wahrscheinlichkeiten leicht berechnen, was man am einfachsten anhand eines Venndiagramms einsieht:



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

- Jetzt erhebt man die deduzierte Formel für Laplace-Experimente zur Definition bedingter Wahrscheinlichkeiten bei beliebigen Wahrscheinlichkeitsverteilungen;

<u>Definition:</u> Für $P(B) \neq 0$ heißt $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$ <u>bedingte Wahrscheinlichkeit</u> (für das Eintreten von A unter Bedingung B).

Unsere Hoffnung ist, dass diese Definition das trifft, was man gefühlsmäßig meint!
Beispiele:

A: „ist US-Staatsbürger“,
B: „hat Muttersprache Englisch“
$P(A/B)$, $P(B/A)$?

A: „ist Vater“
B: „ist männlich“
$P(A/B)$, $P(B/A)$?

3) 50 % aller Frauen sind verheiratet
50 % aller Verheirateten sind Frauen

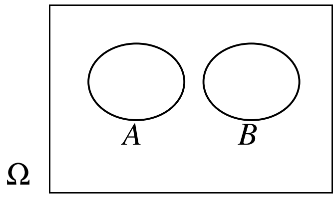
Man beachte, dass durch P(*/B) eine neue Wahrscheinlichkeitsverteilung auf Ω definiert wird (was man auch leicht anhand der Kolmogorov–Axiome nachrechnen kann).

Die Unabhängigkeit von Ereignissen lässt sich jetzt leicht fassen: Zwei Ereignisse (die sinnvollerweise $\neq 0$ und Ω sein sollen) sind unabhängig voneinander, wenn die Wahrscheinlichkeit des Eintretens des ersten unabhängig davon ist, ob das zweite eingetreten ist oder nicht, d.h. $P(A) = P(A/B) = P(A \cap B) / P(B)$ und analog für B. Äquivalent dazu ist die übliche

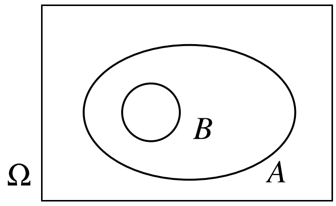
<u>Definition:</u> A, B heißen stochastisch <u>unabhängig</u>, falls $\emptyset \neq A, B \neq \Omega$ und $P(A) \cdot P(B) = P(A \cap B)$ gilt.

Leicht folgt:
- mit A; B sind auch A, $\overline{B}$; $\overline{A}$, B und $\overline{A}$, $\overline{B}$ unabhängig.
- sind A, B unvereinbar, so sind sie abhängig!

Veranschaulichung an Venn-Diagrammen (bei Laplace-Experimenten)



A, B unvereinbar, also A, B abhängig



B zieht *A* nach sich, also *A*, *B* abhängig

<u>1.6.2 Ergänzungen und Beispiele</u>

Mit <u>Vierfeldertafeln</u> lassen sich Fragen der bedingten Wahrscheinlichkeit $P(A/B)$ besonders übersichtlich darstellen: Die folgende Tabelle gibt die Schülerdaten des Lessing-Gymnasiums an:

Lessing-G.	A Mädchen	$\overline{A}$ Jungen	Summe
B Oberstufe	33	43	76
$\overline{B}$ Unterstufe	112	103	215
Summe	145	146	291

Übertragen auf Wahrscheinlichkeiten stehen in der Vierfeldertafel:

	A	$\overline{A}$	
B	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{B}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	$P(\overline{B})$
	$P(A)$	$P(\overline{A})$	

	A	$\overline{A}$	
B	0,113	0,148	0,261
$\overline{B}$	0,385	0,354	0,739
	0,498	0,502	

Die Definition der bedingten Wahrscheinlichkeit ergibt nun z. B. sofort $P(A/B) = P(A \cap B)/P(B) = 0{,}113/0{,}261 = 0{,}433$.
Auch die Unabhängigkeit lässt sich in der Vierfeldertafel überprüfen: Falls A und B unabhängig sind, gilt $P(A \cap B) = P(A) \cdot P(B)$ usw., d.h. die Vierfeldertafel zeigt eine Multiplikationstafel der Wahrscheinlichkeiten $P(A)$ usw.:

	A	$\overline{A}$	
B	$P(A)\cdot P(B)$	$P(\overline{A})\cdot P(B)$	$P(B)$
$\overline{B}$	$P(A)\cdot P(\overline{B})$	$P(\overline{A})\cdot P(\overline{B})$	$P(\overline{B})$
	$P(A)$	$P(\overline{A})$	

Ist bekannt (oder kann vorausgesetzt werden), dass A, B unabhängig sind, so können neue Wahrscheinlichkeiten berechnet werden (Definition als Rechenregel!).

- Warnung: Für die Unabhängigkeit von mehr als 2 Ereignissen reicht es nicht, $P(A_1\cap \dots \cap A_n) = P(A_1)\cdot \dots \cdot P(A_n)$ zu definieren, sondern die Formel muss für alle Schnitte gelten.

Versuch: A, B, C unabhängig, wenn $P(A\cap B\cap C) = P(A)\cdot P(B)\cdot P(C)$.
Beispiel: Urnen $U_1 = \{a, b, c\}$, $U_2 = \{1, 2, 3, 4\}$,

$\Omega = \{a1, a2, \dots ,c4\}$, $|\Omega| = 12$
$A = \{b1, b2, c1, c2\}$, $|A| = 4$, $P(A) = 1/3$
$B = \{a1, a2, b1, c1, c2, c3\}$, $P(B) = 1/2$
$C = \{a1, a2, a3, b1, b2, b3\}$, $P(C) = 1/2$
$A\cap B = \{b1, c1, c2\}$, $P(A\cap B) = 1/4$
$A\cap C = \{b1, b2\}$, $P(A\cap C) = 1/6$
$B\cap C = \{a1, a2, b1\}$, $P(B\cap C) = 1/4$
$A\cap B\cap C = \{b1\}$, $P(A\cap B\cap C) = 1/12$

Es gilt zwar $P(A\cap B\cap C) = P(A)\cdot P(B)\cdot P(C)$, aber $P(A)\cdot P(B) = \frac{1}{6} \neq \frac{1}{4} = P(A\cap B)$.

Definition: $A_1, \dots , A_n$ unabhängig: <=> für alle möglichen Schnitte gilt $P(A_{i1}\cap \dots \cap A_{is}) = P(A_{i1})\cdot \dots \cdot P(A_{is})$.

- Das folgende Beispiel zeigt, dass die Definition der Abhängigkeit nicht einmal bei Laplace-Experimenten unbedingt anschaulich ist.

  Werfe n-mal eine Münze und betrachte die beiden folgenden Ereignisse:
  A = „es kommt höchstens einmal Zahl“
  B = „jede Seite kommt mindestens einmal vor“.

Lösung: $\Omega = \{W, Z\}^n$, $|\Omega| = 2^n$, $P(E) = \left(\frac{1}{2}\right)^n$ für jedes Elementarereignis E.

$A = \{WWW...W, ZW...W, WZW...W, \dots , W...WZ\}$, $|A| = n+1$, $P(A) = \frac{n+1}{2^n}$

$B = \Omega \setminus \{WW...W, ZZ...Z\}$, $|B| = 2^n - 2$, $P(B) = \frac{2^{n-2}}{2^n}$

Damit gilt:

$$P(A\cap B)=P(A)\cdot P(B) \quad\Leftrightarrow\quad \frac{n}{2^n}=\frac{n+1}{2^n}\cdot\frac{2^n-2}{2^n} \quad\Leftrightarrow\quad n=(n+1)\cdot\left(1-\frac{1}{2^{n-1}}\right)=n+1-\frac{n+1}{2^{n-1}}$$

$$\Leftrightarrow\ 1=\frac{n+1}{2^{n-1}}\ \Leftrightarrow\ 2^{n-1}=n+1\ \Leftrightarrow\ n=3$$

Dies führt zu dem merkwürdigen Ergebnis, dass A und B für $n = 3$ unabhängig, sonst aber immer abhängig sind.

- Roulettespieler warten oft, bis n-mal Rot kam (n = individuelle Zahl) und setzen dann auf Schwarz.

- Fallschirme öffnen sich in 999 von 1000 Fällen. Deshalb haben Fallschirmspringer 2 Fallschirme, deren Versagen selbstverständlich als unabhängige Ereignisse angesehen werden können.

  Also: P(„beide versagen“) = $\frac{1}{1000}\cdot\frac{1}{1000}=\frac{1}{1.000.000}$.

  Wirklich: Fallschirm versagt, weil eine Naht aus schlechtem Material besteht, das nach kurzem Lagern brüchig wird. Beide Fallschirme stammen aus derselben Charge. !?!

Analog: Hardware- und Software-Redundanz

Die Luftfahrtindustrie verweist mit Hilfe der Multiplikations-Regel auf ihre Sicherheit durch mehrfach ausgelegte Systeme. Parallel geschaltete Sensoren und Aktoren können von derselben Fehlerquelle zum selben Zeitpunkt gestört werden. In verschiedenen Systemen (von verschiedenen Teams geschrieben) kann derselbe Programmierfehler stecken.

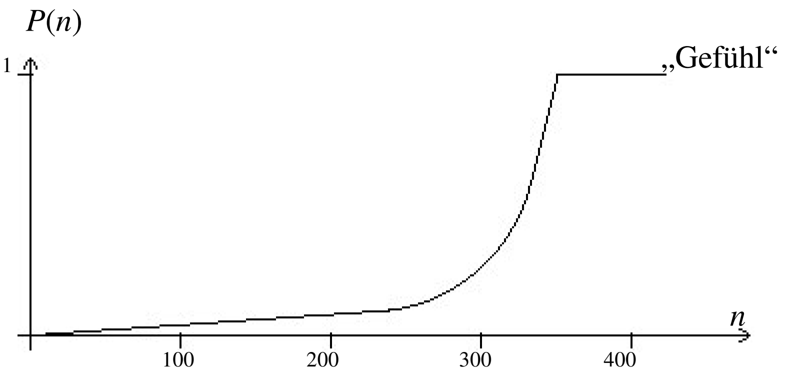
- Flugreisende, die Angst vor versteckten Bomben haben, sollten stets eine Bombe mit ins Flugzeug nehmen.

- **Geburtstagsproblem:** (Problem 2, Seite 3)

  Bei einer Talkshow (ca. 80 Teilnehmer) behauptet der befragte Gast, es seien sicher zwei Personen mit gleichem Geburtstag anwesend. Der Showmaster glaubt das nicht und fragt spontan: „Ich habe am 26.10. Geburtstag, wer noch?“ Keiner meldet sich.

  - Was denken Sie?
  - Wie schätzen Sie die Wahrscheinlichkeit ein?

Das naive Gefühl für die Wahrscheinlichkeit, dass zwei Personen am gleichen Tag Geburtstag haben, wird durch die folgende Graphik beschrieben.



Beim Geburtstagsproblem sind die ähnlich klingenden Ereignisse zu unterscheiden:

a) P(„Zwei Personen haben am gleichen Tag Geburtstag“)
b) P(„Zwei Personen haben am 26.10. Geburtstag“)
c) P(„Eine weitere Person hat wie ich am 26.10. Geburtstag“)
(eigentlich müsste es jeweils „mindestens zwei Personen...“ heißen.)

<u>Lösung</u> über Gegenereignis:

a) $|\Omega| = 365^n$, $\overline{E}_a =$ “Keine 2 haben am gleichen Tag Geburtstag“

$|\overline{E}_a| = 365 \cdot 364 \cdot ... \cdot (365-(n-1))$

$$P\left(\overline{E_a}\right) = \frac{365 \cdot 364 \cdot ... \cdot (365-(n-1))}{365^n},$$

$P(E_a) = 1 - P(\overline{E_a})$, z.B. $n = 30$: $P(E_a) \approx 71\%$; $n = 80$: $P(E_a) \approx 100\%$.

b) $|\Omega| = 365^n$,

$\overline{E}_b =$ „alle haben nicht am 26.10. Geburtstag“ oder „genau einer (der 1. oder 2. oder ... der n-te) hat am 26.10. Geburtstag“

$$\left|\overline{E}_b\right| = 364^n + n \cdot 364^{n-1}$$

$$p(\overline{E}_b) = \frac{364^n + n \cdot 364^{n-1}}{365^n}$$

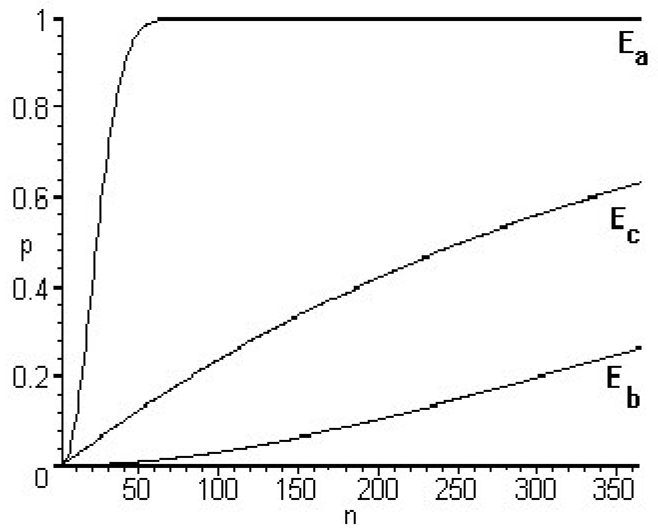
$p(E_b) = 1 - p(\overline{E}_b)$, z.B. $n = 30, P(E_b) \approx 0{,}3\%; n = 80,$
$P(E_b) \approx 2{,}1\%$

c) Jetzt geht es nur noch um n-1 Personen, die alle <u>nicht</u> am 26.10. Geburtstag haben, also

$$\left|\Omega\right| = 365^{n-1}, \left|\overline{E_c}\right| = 364^{n-1}, \quad P\left(\overline{E_c}\right) = \left(\frac{364}{365}\right)^{n-1}$$

$$P(E_c) = 1 - p(\overline{E_c}), \ z.B. \ n = 30: \ P(E_c) \approx 7{,}5\%; n = 80: \ P(E_c) \approx 19{,}7\%$$

Die wirkliche Wahrscheinlichkeit k für die 3 Fälle wird durch die folgende, mit MAPLE erstellte Graphik beschrieben:

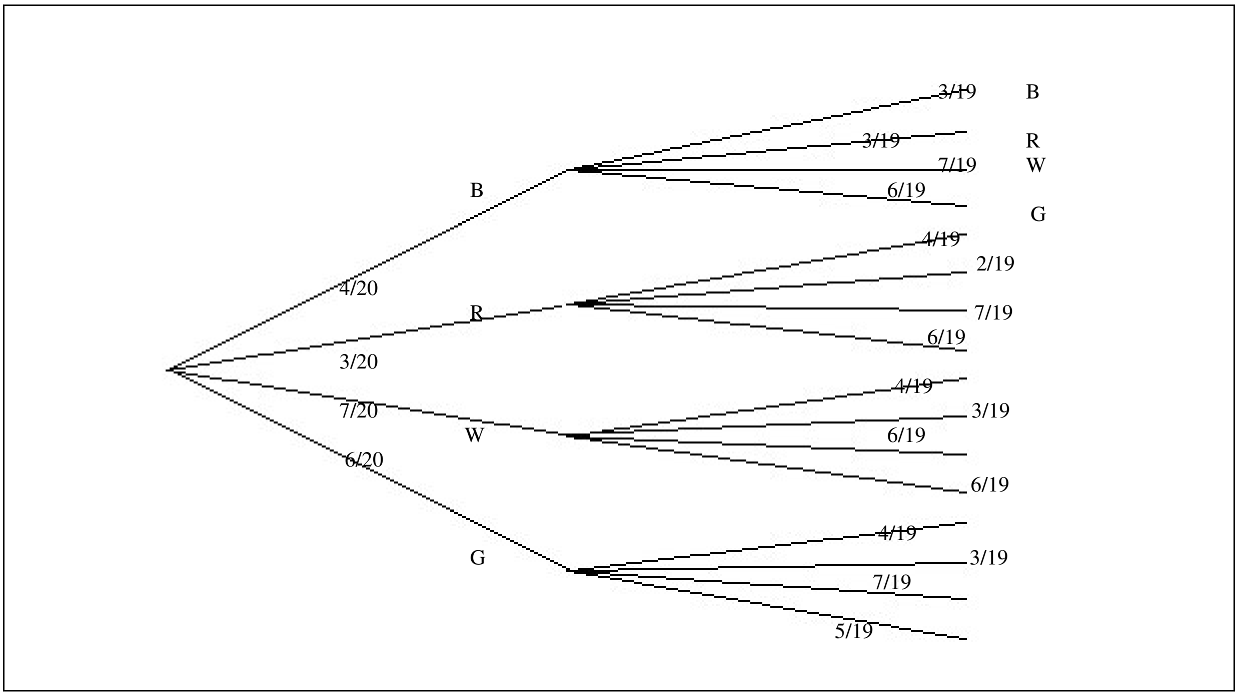


1.6.3 Mehrstufige Zufallsexperimente und Pfadregeln

Beispiel 1:

In einer Urne liegen 20 Kugeln: 4 blaue, 3 rote, 7 weiße, 6 grüne. Ich ziehe nacheinander zwei Kugeln ohne Zurücklegen. Wie groß sind die Laplace-Wahrscheinlichkeiten folgender Ereignisse?

a) Die erste Kugel ist blau, die zweite rot
b) Beide Kugeln sind gleichfarbig
c) Die zweite Kugel ist blau oder rot
d) Die erste Kugel ist nicht weiß und die zweite Kugel ist grün
e) Die zweite Kugel ist grün, wobei wir bereits wissen, dass die erste Kugel nicht weiß ist.



Das Experiment wird besonders übersichtlich, wenn man es als Nacheinanderausführung zweier eigenständiger Zufallsexperimente, d.h. als zweistufiges Zufallsexperiment betrachtet, wobei jeweils eine Kugel gezogen wird. Dabei ist das 2. Experiment vom Ausgang des ersten abhängig. Veranschaulicht werden solche Experimente durch Bäume, an deren „Ästen" die einzelnen (Zweig-)Wahrscheinlichkeiten geschrieben werden. Jedes Ereignis und seine Wahrscheinlichkeit kann durch das obenstehende Baumdiagramm beschrieben werden. Die einzelnen Wahrscheinlichkeiten sind durch die Anzahl der Kugeln bestimmte Laplace-Wahrscheinlichkeiten. Ein Ergebnis, z. B.
1. Kugel blau, 2. Kugel rot, wird durch einen Pfad in Teilaufgabe a) beschrieben, seine Wahrscheinlichkeit durch das Produkt der Wahrscheinlichkeiten entlang des Pfades, im Beispiel P(rot / blau) = $\frac{4}{20} \cdot \frac{3}{19} = \frac{12}{380}$. Jedes Ereignis wird durch die zugehörigen Pfade beschrieben, z.B. in Teilaufgabe b)

E_b = „beide Kugeln gleichfarbig" = {(b/b), (r/r), (w/w), (g/g)} mit

$$P(E_b) = \frac{4}{20}\cdot\frac{3}{19}+\frac{3}{20}\cdot\frac{2}{19}+\frac{7}{20}\cdot\frac{6}{19}+\frac{6}{20}\cdot\frac{5}{19} = \frac{12+6+42+30}{380} = \frac{90}{380} = \frac{9}{38}.$$

Entsprechend lässt sich jedes Ereignis c) – e) beschreiben:

Teilaufgabe c)

Bei E_c = „zweite Kugel blau oder rot" kann man alle zugehörigen Pfade berücksichtigen:

$$P(E_c) = \frac{4}{20}\cdot\left(\frac{3}{19}+\frac{3}{19}\right)+\frac{3}{20}\cdot\left(\frac{4}{19}+\frac{2}{19}\right)+\frac{7}{20}\cdot\left(\frac{4}{19}+\frac{3}{19}\right)+\frac{6}{20}\cdot\left(\frac{4}{19}+\frac{3}{19}\right) = \frac{133}{380} = \frac{7}{20}$$

Man kann natürlich auch einfacher argumentieren, dass der 1. Zug irrelevant ist, also 7 von 20 günstige Kugeln da sind.

<u>Teilaufgabe d):</u> Die zu E_d = „1. Kugel nicht weiß und 2. Kugel grün" gehörigen Pfade ergeben

$$P(E_d) = \frac{4}{20}\cdot\frac{6}{19} + \frac{3}{20}\cdot\frac{6}{19} + \frac{6}{20}\cdot\frac{5}{19} = \frac{72}{380} = \frac{18}{95} \approx 19\%$$

<u>Teilaufgabe e):</u> Hier ist die bedingte Wahrscheinlichkeit $P(E_d)$ = P(„2. Kugel grün"/"1. Kugel nicht weiß")

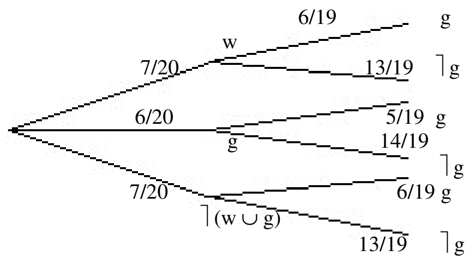
$$= \frac{P(2.\mathit{Kugel\ grün}"\ \mathit{und}\ "1.\mathit{Kugel\ nicht\ weiß}")}{P("1.\mathit{Kugel\ nicht\ weiß}")}$$

$$= \frac{P(E_d)}{P("1.\mathit{Kugel\ nicht\ weiß}")}$$

$$\frac{\frac{18}{95}}{\frac{13}{20}} = \frac{72}{247} \approx 29\%$$

Natürlich kann man auch jeweils speziell zugeschnittene Bäume verwenden:

Das Ereignis E_d = „1. Kugel nicht weiß, 2. Kugel grün" wird durch den folgenden vereinfachten Baum beschrieben:

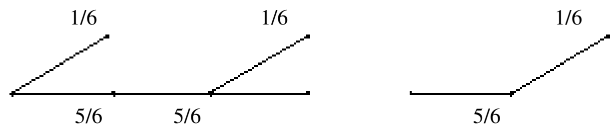


$$\leftarrow\ P(E_d) = \frac{6}{20}\cdot\frac{5}{19} + \frac{7}{20}\cdot\frac{6}{19} = \frac{72}{380} = \frac{18}{95}$$

$\leftarrow$

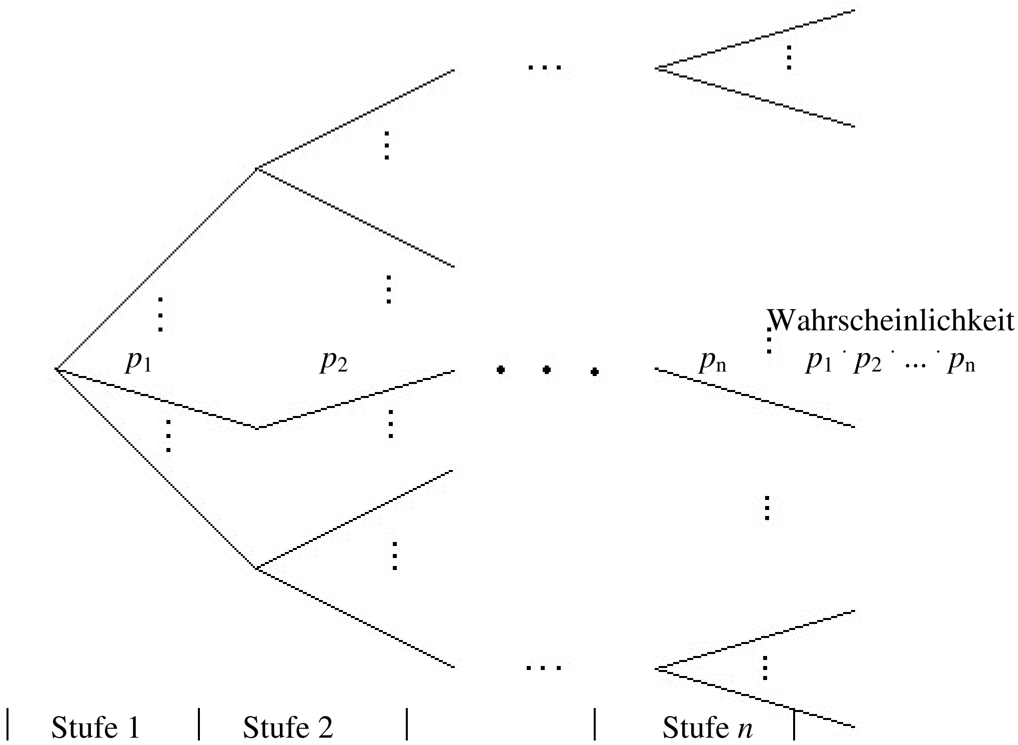
Beispiel 2: Gesucht ist die Wahrscheinlichkeit, genau beim n-ten Würfelwurf eine „6“ zu werfen.

Lösung: n-stufiges Zufallsexperiment, zugehöriger Baum:



$$P = \left(\frac{5}{6}\right)^{n-1} \cdot \frac{1}{6}$$

Allgemein versteht man unter einem mehrstufigen Zufallsexperiment die Nacheinanderausführung mehrerer Zufallsexperimente. Man veranschaulicht ein n-stufiges Zufallsexperiment durch einen n-stufigen Baum, wobei in jeder Stufe die Äste den interessierenden Einzel-Ergebnissen entsprechen. Die Wahrscheinlichkeiten der Einzelergebnisse notiert man an den zugehörigen Ästen:



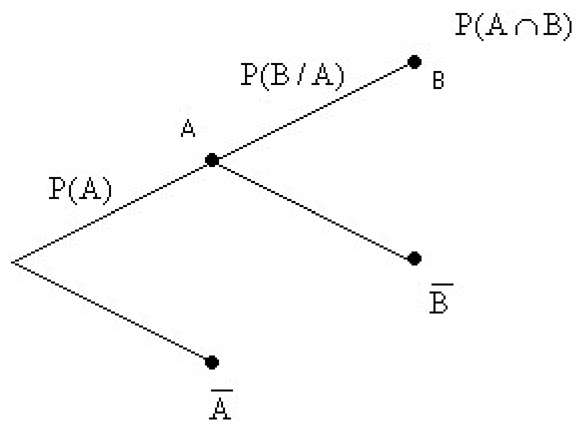
Für mehrstufige Zufallsexperimente gelten die Pfadregeln:

Für ein mehrstufiges Zufallsexperiment – veranschaulicht durch einen Baum – gilt: (1) Die Wahrscheinlichkeit eines Ergebnisses erhält man durch Multiplikation aller Einzel-Wahrscheinlichkeiten längs des zugehörigen Pfades im Baum (Pfad-Multiplikationsregel). (2) Die Wahrscheinlichkeit eines Ereignisses erhält man durch Addition aller zugehörigen Ergebnis-Wahrscheinlichkeiten (Pfad-Additionsregel).

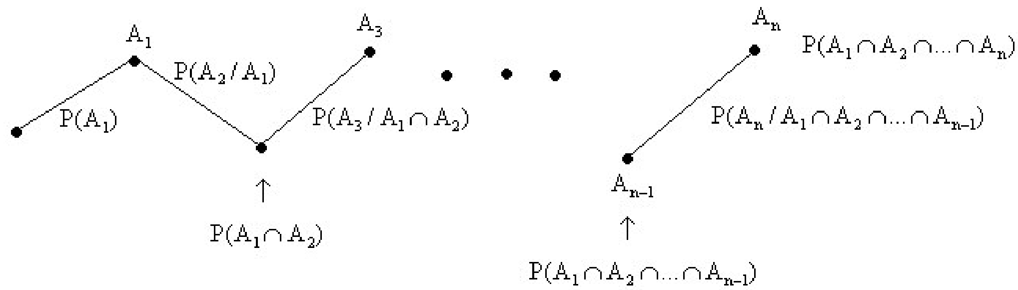
Diese Regeln ermöglichen im konkreten Fall eine rasche und einfache Berechnung von Wahrscheinlichkeiten, ohne einen Ergebnisraum und entsprechende Ereignisse explizit als Mengen darstellen und ohne kombinatorische Berechnungen anstellen zu müssen.
Es ist klar, dass im Baum nach jeder Stufe die Summe aller Wahrscheinlichkeiten an den Ast-Enden (erhalten durch Multiplizieren längs der Pfade) gleich 1 sein muss. Dies ermöglicht eine Kontrolle der Rechnungen.

Die Pfadregeln folgen aus der Definition der bedingten Wahrscheinlichkeit (bzw. direkt bei mehrstufigen Laplace-Experimenten):

Die Pfad-Multiplikationsregel für n = 2 ist die Umformung der Definition der bedingten Wahrscheinlichkeit: $P(A \cap B) = P(A) \cdot P(B / A)$



Die Verallgemeinerung auf beliebigen n ist die Pfad-Multiplikationsregel:



$$P(A_1 \cap A_2 \cap A_3 \cap ... \cap A_n) = P(A_n / A_1 \cap A_2 \cap ... \cap A_{n-1}) = ... =$$
$$= P(A_1) \cdot P(A_2 / A_1) \cdot P(A_3 / A_1 \cap A_2) \cdot ... \cdot P(A_n / A_1 \cap A_2 \cap ... \cap A_{n-1})$$

Die Pfad-Additionsregel folgt auf den Kolmogorov-Axiomen.

1.6.4 Beispiele zu den Pfadregeln

Beispiel 1: Der Dieb von Bagdad

Bevor der berühmte Dieb endlich in den Kerker kommt, darf er blind in eine von drei Kisten greifen. Sie enthalten 4 weiße und 2 schwarze, je 3 weiße und schwarze, und die dritte Kiste 5 weiße und eine schwarze Kugel. Der Dieb wird noch einmal laufen gelassen, wenn er eine weiße Kugel zieht.

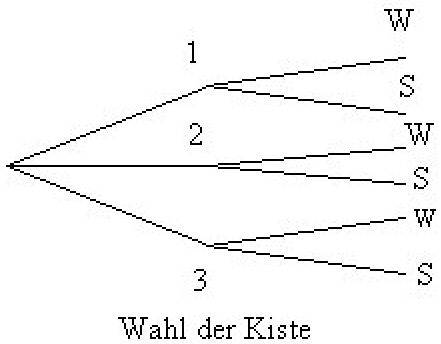
a) Wie groß ist seine Chance?

b) Der Dieb fragt, ob er vor dem Ziehen die Kugeln umordnen darf. Bringt das etwas?

Lösung:

Beschreibe das Ziehen als 2-stufiges Zufallsexperiment:

a)



$$P(\text{„weiß“}) = \frac{1}{3}\cdot\frac{4}{6}+\frac{1}{3}\cdot\frac{3}{6}+\frac{1}{3}\cdot\frac{5}{6} = \frac{2}{3}$$

b) Falls man beim Umordnen jeweils 6 Kugeln pro Kiste beibehält, so bleibt $P(\text{„weiß“}) = \frac{2}{3}$!

Lösen von dieser Beschränkung ergibt das Optimum:

1 W 0	1 W 0	10 W 6 S

$$P(\text{„weiß“}) = \frac{1}{3}+\frac{1}{3}+\frac{1}{3}\cdot\frac{10}{16} = \frac{7}{8}.$$

Beispiel 2:

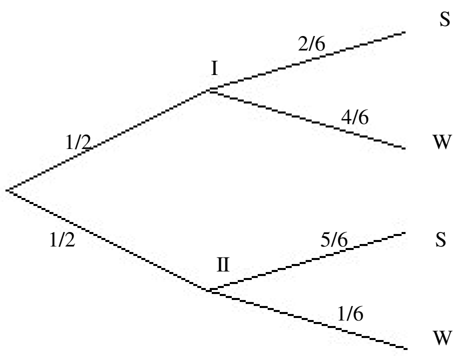
Urne I: 4 weiße und 2 schwarze Kugeln, Urne II: 1 weiße und 5 schwarze Kugeln.
Zuerst wird die Urne gewählt, dann eine Kugel gezogen.

a) Wie hoch ist die Wahrscheinlichkeit, eine weiße Kugel zu ziehen?

b) Wie hoch ist die Wahrscheinlichkeit, dass eine weiße Kugel aus Urne I stammt?

Lösung:

a)



2 Lösungsideen:
$P(W) = \frac{1}{2} \cdot \frac{4}{6} + \frac{1}{2} \cdot \frac{1}{6} = \frac{5}{12}$ (Pfadregeln)

oder: schütte alle Kugeln zusammen!
5 von 12 sind weiß!

b) Man weiß, dass die Kugel weiß ist, d.h. es ist gesucht *P*(„Urne I/ Farbe weiß") = *P*(*I*/ *W*):

$$P(I/W) = \frac{P(I \cap W)}{P(W)} = \frac{\frac{1}{2} \cdot \frac{4}{6}}{\frac{1}{2} \cdot \frac{4}{6} + \frac{1}{2} \cdot \frac{1}{6}} = \frac{4}{5} \quad \text{(Pfadregeln)}$$

oder: $P(I/W) = \frac{4}{5}$ ← #weiße in I
← # alle weißen

Achtung: Hier geht die sprachliche Deutung mit ein (vgl. „Der brave Mann denkt an sich, selbst zuletzt!" versus „Der brave Mann denkt an sich selbst zuletzt.")

Wir haben den Satz gelesen als „..., dass eine weiße Kugel aus Urne I stammt" (d. h. „weiß" ist schon bekannt). Man könnte auch lesen „..., dass eine weiße Kugel aus Urne I stammt (d. h. „Urne I" ist schon bekannt). Dann wäre das Ergebnis

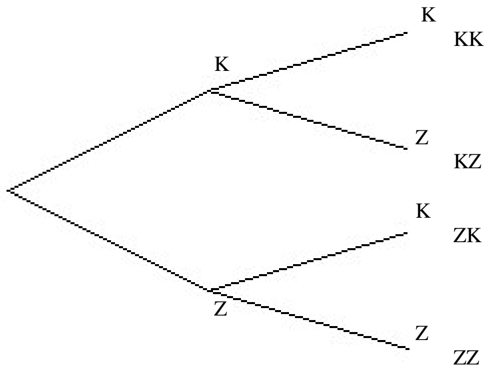
$$P(W/I) = \frac{4}{6} \quad \begin{array}{l} \leftarrow \#\, weiße\, in\, I \\ \#\, alle\, Kuge\ln\, in\, I \end{array}$$

Wohl kaum denkbar ist die Deutung „und" $P(I \cap W) = \frac{1}{2} \cdot \frac{4}{6} = \frac{4}{12}$.

Beispiel 3:
Eine Münze wird nacheinander geworfen und das Ergebnis notiert; Vor Beginn des Spieles wählen Spieler A und B eine der Kombinationen KK, KZ, ZK, ZZ. Gewonnen hat der, dessen Kombination zuerst auftaucht. Sind alle Wahlen gleich gut?

Lösung:

Nein, z.B. A wählt KK, B wählt ZK:



KK: A gewinnt

KZ → B muss irgendwann gewinnen

ZK: B gewinnt

ZZ → B muss irgendwann gewinnen

Der Spielbaum ist zwar unendlich, aber es ist klar:

$$P(\text{„A gewinnt“}) = \frac{1}{4},\ P(\text{„B gewinnt“}) = \frac{3}{4}.$$

Bei einer anderen Regel kann das Ergebnis anders aussehen, z. B. ausgeglichen bei der Wahl KK und KZ.

Beispiel 4:

Im Sack sind r rote und s schwarze Kugeln, 2 mal Ziehen ohne Zurücklegen:

a) Ich ziehe und zeige, dass ich rot gezogen habe.

b) Ich ziehe und zeige Kugel nicht.

Berechne P(„2. Kugel ist rot“)

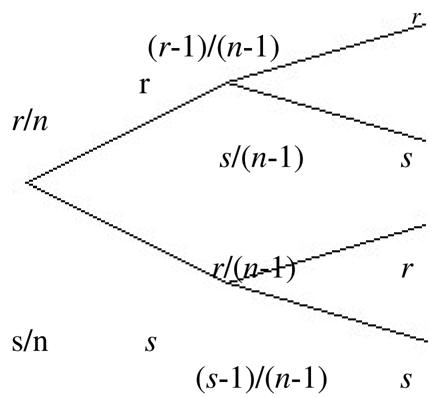
Lösung:

a) $P(\text{„2. Kugel rot“}) = \dfrac{r-1}{r+s-1}$;

b) $P(\text{„2. Kugel rot“}) = \dfrac{r}{r+s}$;

anschauliche Deutung: Da ich nichts weiß, ist wieder alles möglich.

Formale Deutung: Es sei n = r + s



Günstig für b): ((*r*/*r*),(*s*/*r*)), also

$$p = \frac{r}{n} \cdot \frac{r-1}{n-1} + \frac{s}{n} \cdot \frac{r}{n-1}$$

$$= \frac{r(r-1+s)}{n(n-1)} = \frac{r}{n}$$

<u>Beispiel 5</u>:

Wie oft muss man einen Würfel werfen, um mit 98% Sicherheit mindestens eine 6 zu werfen?

<u>Lösung</u>:

Berechne $P(A_n)$ mit A_n = „Keine 6 bei n Würfen"

$P(A_n) = \left(\frac{5}{6}\right)^n$, also $p = 1 - P(A_n) = 1 - \left(\frac{5}{6}\right)^n \geq 0,98$, oder $\left(\frac{5}{6}\right)^n \leq 0,02$.

- Taschenrechner:

n = 2	$P(A_n)$ =	0,69...
n = 4		0,48...
n = 8		0,23...
n = 16		0,05...
n = 32		0,0029...

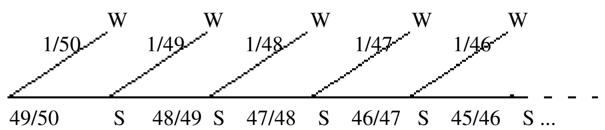
$$\left.\begin{array}{ll} n = 20 & 0,0026... \\ n = 21 & 0,0021... \\ n = 22 & 0,0018... \end{array}\right\} \text{Also mindestens 22 mal}$$

- Logarithmieren

$$\left(\frac{5}{6}\right)^n \leq 0,02 \Leftrightarrow n \cdot \log\left(\frac{5}{6}\right) \geq \log(0,02) \;\Rightarrow\; n \geq 21,46.$$

(Umkehrung des Ungleichheitszeichens, da log negativ ist)

<u>Beispiel 6</u>: (Problem 7 von Seite 3, Kugelproblem)



$P(\text{W beim 1. Griff}) = \frac{1}{50}$

$P(\text{W genau beim 2. Griff}) = \frac{49}{50} \cdot \frac{1}{49} = \frac{1}{50}$

$P(\text{W genau beim 3. Griff}) = \frac{49}{50} \cdot \frac{48}{49} \cdot \frac{1}{48} = \frac{1}{50}$

•

•

•

$P(\text{W genau beim 49. Griff}) = \frac{49}{50} \cdot \frac{48}{49} \cdot \frac{47}{48} \cdot ... \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{50}$

$P(\text{W genau beim 50. Griff}) = \frac{49}{50} \cdot \frac{48}{49} \cdot \frac{47}{48} \cdot ... \cdot \frac{1}{2} \cdot 1 = \frac{1}{50}$

Für Armin zählen der 1., 3., ..., 49. Griff, für Beate der 2., 4., 6., ..., 50. Griff, also

$P(\text{Armin}) = P(\text{Beate}) = 25 \cdot \frac{1}{50} = \frac{1}{2}$.

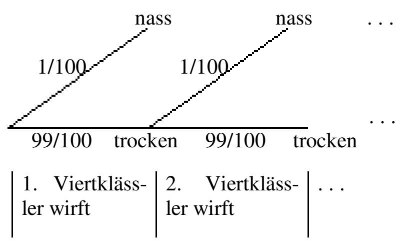
Wären es 49 Kugeln gewesen, so wären die Einzelwahrscheinlichkeiten je $\frac{1}{49}$, für Armin würden der 1., 3., ..., 49., für Beate der 2., 4., ...,48. Zug zählen, also

$P(\text{Armin}) = \frac{25}{49} > P(\text{Beate}) = \frac{24}{49}$.

Beispiel 7: (Problem 9 von Seite 3, Wasserbeutelproblem)

Lösung:

Wir betrachten das Werfen als 100-stufiges Zufallsexperiment und achten auf einen bestimmten, beliebig herausgegriffenen Drittklässler DK; zugehöriger Baum:



Klar $p = P(\text{„DK bleibt trocken“}) = \left(\frac{99}{100}\right)^{100} \approx 0{,}366$.

Nun interpretieren wir die Fragestellung ein wenig um und stellen uns vor, dass das Zufallsexperiment „DK wird beworfen“ 100 mal wiederholt wird. Klar: Ich erwarte dabei, dass das Ereignis „DK bleibt trocken“ ungefähr $100 \cdot p$ mal, d.h. etwa 37 mal eintritt. Also erwarte ich, dass 37 Drittklässler bei der Wurfaktion trocken bleiben. Damit ist Problemaufgabe 9 gelöst!

Verallgemeinerung auf n Dritt- und Viertklässler: Der erwartete relative Anteil der trocken gebliebenen Drittklässler ist $\left(1-\frac{1}{n}\right)^n$. Für wachsendes n wächst diese Zahl nicht unbeschränkt, sondern strebt gegen $\frac{1}{e} \approx 36{,}8\%$.

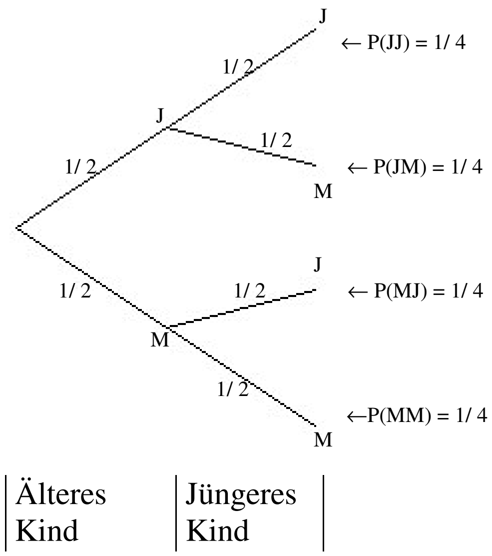
Beispiel 8:
a) Geschwisterproblem (Problem 8, Seite 3)
b) Analog, nur erfahre ich, dass das ältere Kind ein Mädchen ist.

Lösung:
Gefragt sind bei

a) p_1 = P(„Beide Kinder Mädchen“| „(Mindestens) ein Kind ist Mädchen“)
b) p_2 = P(„Jüngeres Kind Mädchen“| „Älteres Kind Mädchen“).

Es seien

Ω = {JJ, JM, MJ, MM} (Codierung „älteres/jüngeres Kind“)
A = {MM}, B = {JM, MJ, MM}, C = {MJ, MM}



Gefragt sind also $p_1 = P(A|B)$ und $p_2 = P(A|C)$. Wegen A∩B = A = A∩C gilt

$$p_1 = \frac{|A \cap B|}{|B|} = \frac{1}{3} \left(= \frac{\frac{1}{4}}{\frac{3}{4}} \right) \text{ und } p_2 = \frac{|A \cap C|}{|C|} = \frac{1}{2} \left(= \frac{\frac{1}{4}}{\frac{1}{2}} \right).$$

1.6.5 Bedingte Wahrscheinlichkeiten und der Satz von Bayes

Einführendes Beispiel „Wahlumfrage“

In einer repräsentativen Stichprobe aus der Bevölkerung wurde eine Umfrage veranstaltet. Dabei ergaben sich folgende Präferenzen für die drei Parteien Q, R, S, aufgeschlüsselt nach der Zugehörigkeit zu drei sozialen Schichten U, V, W:

	Q	R	S
U	60%	30%	10%
V	50%	40%	10%
W	30%	20%	50%

Es gilt also zum Beispiel $P(\mathrm{R}/\mathrm{V}) = 40\%$.

Typische Fragen sind:

a) Mit welcher Wahrscheinlichkeit wählt eine zufällig angesprochene Person X die Partei Q (bzw. R, bzw. S)?
b) Eine Person X wählt die Partei Q (bzw. R, bzw. S). Mit welcher Wahrscheinlichkeit gehört sie zur Schicht U (bzw. V, bzw. W)?

Gegeben sind also die bedingten Wahrscheinlichkeiten $P(\mathrm{Q}/\mathrm{U})$, $P(\mathrm{R}/\mathrm{U})$, $P(\mathrm{S}/\mathrm{U})$, ..., $P(\mathrm{S}/\mathrm{W})$, gesucht sind die Wahrscheinlichkeiten $P(\mathrm{Q})$, $P(\mathrm{R})$, $P(\mathrm{S})$ bei a) und die bedingten Wahrscheinlichkeiten $P(\mathrm{U}/\mathrm{Q})$, $P(\mathrm{V}/\mathrm{Q})$, ..., $P(\mathrm{W}/\mathrm{S})$ bei b) (dabei haben wir abkürzend Q für „X wählt Partei Q“ und U für „X gehört zur Schicht U“ geschrieben).

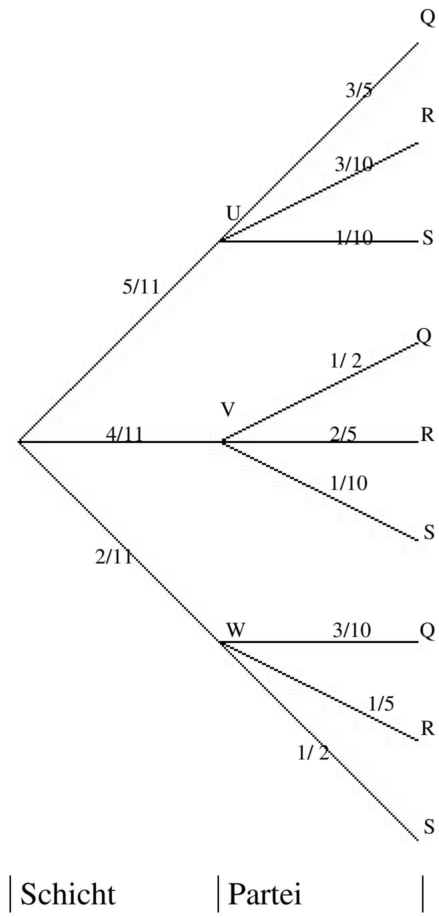
Es sei bekannt, dass sich die Bevölkerung in die Schichten zu 5 : 4 : 2 aufteilt.

Damit ist gegeben:

$$P(\mathrm{Q}/\mathrm{U}) = \frac{3}{5},\ P(\mathrm{R}/\mathrm{U}) = \frac{3}{10},\ P(\mathrm{S}/\mathrm{U}) = \frac{1}{10},\ \text{etc.}$$
$$P(\mathrm{U}) = \frac{5}{11},\ P(\mathrm{V}) = \frac{4}{11},\ P(\mathrm{W}) = \frac{2}{11}.$$

Antwort zu a) Gesucht ist $P(\mathrm{Q})$, ...

Zugehöriger Baum:



$\leftarrow P(\text{Q}\cap\text{U}) = \frac{5}{11}\cdot\frac{3}{5} = \frac{3}{11}$

$\leftarrow P(\text{Q}\cap\text{V}) = \frac{4}{11}\cdot\frac{1}{2} = \frac{2}{11}$

$\leftarrow P(\text{Q}\cap\text{W}) = \frac{2}{11}\cdot\frac{3}{10} = \frac{3}{55}$

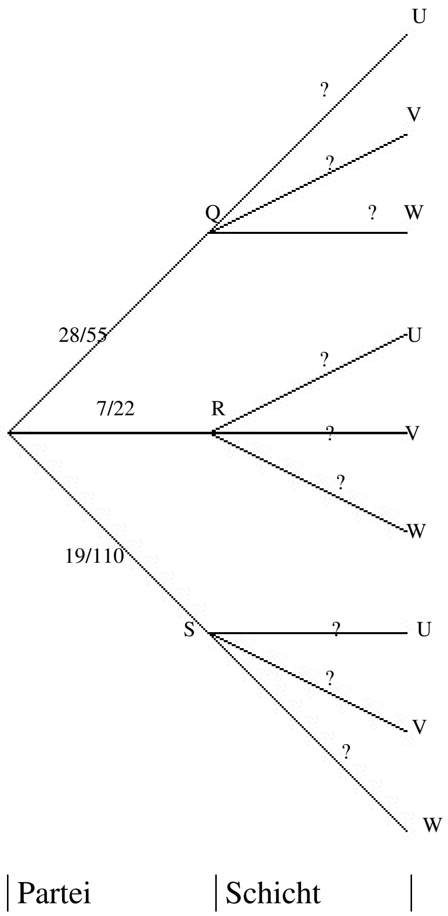
Klar:

$$P(\text{Q}) = P(\text{Q}\cap\text{U}) + P(\text{Q}\cap\text{V}) + P(\text{Q}\cap\text{W}) = \frac{3}{11}+\frac{2}{11}+\frac{3}{55} = \frac{28}{55} \approx 50{,}9\%$$

Analog: $P(\text{R}) = \frac{7}{22} \approx 31{,}8\%$, $P(\text{S}) = \frac{19}{110} \approx 17{,}3\%$

Antwort zu b): Gesucht $P(\text{U}/\text{Q})$, ...

Der zugehörige Baum ist umgekehrt wie der bei a), die bedingten Wahrscheinlichkeiten der 2. Stufe sind gesucht!



$\leftarrow P(\mathrm{Q}\cap\mathrm{U}) = \dfrac{3}{11}$ nach a)

Pfadmultiplikationstregel:

$P(\mathrm{Q}) \cdot P(\mathrm{U/Q}) = P(\mathrm{U}\cap\mathrm{Q})$, also

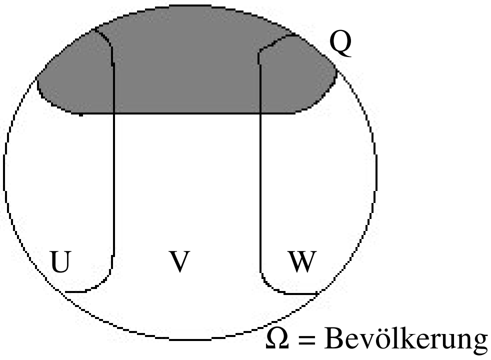
$$P(\mathrm{U/Q}) = \frac{P(\mathrm{U}\cap\mathrm{Q})}{P(\mathrm{Q})} = \frac{\frac{3}{11}}{\frac{28}{55}} = \frac{15}{28} \approx 53{,}6\%$$

(und analog $P(\mathrm{V/Q}) = \dfrac{5}{14}$, $P(\mathrm{W/Q}) = \dfrac{3}{28}$, $P(\mathrm{U/R}) = \dfrac{3}{4}$, ...)

Analysieren wir noch einmal genauer, wie wir ausgehend von den gegebenen bedingten Wahrscheinlichkeiten *P*(Q/U), P(R/U), ... und den gegebenen Wahrscheinlichkeiten *P*(U), *P*(V), *P*(W) die gesuchten Wahrscheinlichkeiten *P*(Q), *P*(R), *P*(S) und *P*(U/Q), *P*(R/Q), ... berechnet haben.

Unser Beispiel:

$\Omega = U \cup V \cup W$ disjunkt,
$Q \subseteq \Omega$



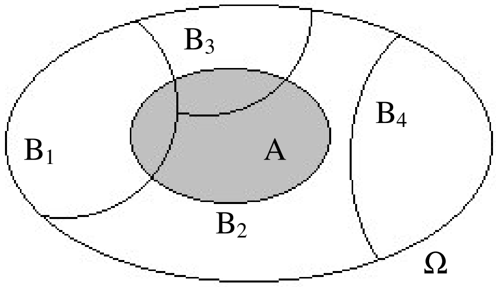
$$P(Q) = P((Q \cap U) \cup (Q \cap V) \cup (Q \cap W))$$

$$\overset{(1)}{=} P(Q \cap U) + P(Q \cap V) + P(Q \cap W)$$

$$\overset{(2)}{=} P(U) \cdot P(Q/U) + P(V) \cdot P(Q/V) + P(W) \cdot P(Q/W)$$

Allgemeine Situation:

$\Omega = \bigcup_{i=1}^{n} B_i$ disjunkt, $A \subseteq \Omega$



$$P(A) = P\left(\bigcup_{i=1}^{n} A \cap B_i\right)$$

$$\overset{(1)}{=} \sum_{i=1}^{n} P(A \cap B_i)$$

$$\overset{(2)}{=} \sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)$$

Dabei gilt das Gleichheitszeichen (1) wegen Axiom 3, der Kolmogrorov-Axiome (vgl. S. 38) das Gleichheitszeichen (2) nach Definition der bedingten Wahrscheinlichkeit. Diese Formel heißt auch der

Satz von der totalen Wahrscheinlichkeit:

> Sei P eine Wahrscheinlichkeitsverteilung auf Ω. Es gelte $\Omega = \bigcup_{i=1}^{n} B_i$ mit $B_i \cap B_j = \varnothing$ für $i \neq j$. Dann gilt für jedes $A \subseteq \Omega$:
>
> $$P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)$$

Betrachten wir weiter:

Unser Beispiel

$$P(U/Q) = \frac{P(Q \cap U)}{P(Q)}$$

es gilt auch genauso
$P(Q \cap U) = P(U) \cdot P(Q/U)$

also gilt zusammen
$P(U/Q) =$

$$\frac{P(U) \cdot P(Q/U)}{P(U) \cdot P(Q/U) + P(V) \cdot P(Q/V) + P(W) \cdot P(Q/W)}$$

Allgemeine Situation

$$P(B_j / A) = \frac{P(B_j \cap A)}{P(A)}$$

$$P(B_j \cap A) = P(B_j) \cdot P(A/B_j)$$

$$P(B_j / A) = \frac{P(B_j) \cdot P(A/B_j)}{\sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)}$$

Die letzte Formel heißt der

Satz von Bayes:

> Sei P eine Wahrscheinlichkeitsverteilung auf Ω. Es gelte $\Omega = \bigcup_{i=1}^{n} \mathrm{B}_i$ mit $\mathrm{B}_i \cap \mathrm{B}_j = \varnothing$ für $i \neq j$.
>
> Dann gilt für jedes j und jedes $\mathrm{A} \subseteq \Omega$ mit $P(\mathrm{A}) \neq 0$
>
> $$P(\mathrm{B}_j / \mathrm{A}) = \frac{P(\mathrm{B}_j) \cdot P(\mathrm{A}/\mathrm{B}_j)}{\sum_{i=1}^{n} P(\mathrm{B}_i) \cdot P(\mathrm{A}/\mathrm{B}_i)}$$

Diese „Bayes´sche Umkehrformel" wird in Situationen angewendet, in denen zu einem gegebenen Ereignis A die bedingten Wahrscheinlichkeiten $P(\mathrm{A}/\mathrm{B}_i)$ sowie die Wahrscheinlichkeiten $P(\mathrm{B}_i)$ der Bedingungen (mit $\Omega = \bigcup_{i=1}^{n} \mathrm{B}_i$) gegeben sind und umgekehrt die bedingte Wahrscheinlichkeit $P(\mathrm{B}_j/\mathrm{A})$ gesucht ist.

Eine wichtige Deutung des Bayes´schen Satzes ist die <u>subjektivistische Deutung</u>: Ich habe a-priori-Wahrscheinlichkeiten (B_j). Nun tritt das Ereignis A ein. Dadurch verändern sich meine a-posteriori-Wahrscheinlichkeiten $P(\mathrm{B}_j/\mathrm{A})$ („Lernen durch Erfahrung").

<u>Beispiel 1</u>: Zollhundproblem, Problem 10, Seite 3

B = „Zollhund bellt"
S = „Grenzgänger schmuggelt Rauschgift"

Gegeben:	$P(\mathrm{B}/\mathrm{S}) = 0{,}98$,	$P(\mathrm{B}/\overline{\mathrm{S}}) = 0{,}03$,	$P(\mathrm{S}) = 0{,}01$
Es folgt:	$P(\overline{\mathrm{B}}/\mathrm{S}) = 0{,}02$,	$P(\overline{\mathrm{B}}/\overline{\mathrm{S}}) = 0{,}97$,	$P(\overline{\mathrm{S}}) = 0{,}99$
Gesucht:	$P(\mathrm{S}/\mathrm{B})$ und $(\overline{\mathrm{S}}/\overline{\mathrm{B}})$		

Formale Lösung mit Bayes´scher Regel:

$$P(\mathrm{S}/\mathrm{B}) = \frac{P(\mathrm{S}) \cdot P(\mathrm{B}/\mathrm{S})}{P(S) \cdot P(\mathrm{B}/\mathrm{S}) + P(\overline{\mathrm{S}}) \cdot P(\mathrm{B}/\overline{\mathrm{S}})} = \ldots \approx 0{,}248 \approx 25\%$$

$$\mathrm{P}(\overline{\mathrm{S}}/\overline{\mathrm{B}}) = \frac{\mathrm{P}(\overline{\mathrm{S}}) \cdot \mathrm{P}(\overline{\mathrm{B}}/\overline{\mathrm{S}})}{\mathrm{P}(\overline{\mathrm{S}}) \cdot \mathrm{P}(\overline{\mathrm{B}}/\overline{\mathrm{S}}) + \mathrm{P}(\mathrm{S}) \cdot \mathrm{P}(\overline{\mathrm{B}}/\mathrm{S})} = \ldots \approx 1{,}000 = 100\%$$

Verblüffend ist, dass $P(\mathrm{S}/\mathrm{B})$ so klein ist! Man kann sich dies an der Vierfeldertafel klar machen (berechnet mit Hilfe von $P(\mathrm{S}) = 0{,}01$, $P(\overline{\mathrm{S}}) = 0{,}99$, $P(\mathrm{B} \cap \mathrm{S}) = P(\mathrm{B}/\mathrm{S}) \cdot P(\mathrm{S})$ und $P(\mathrm{B} \cap \overline{\mathrm{S}}) = P(\mathrm{B}/\overline{\mathrm{S}}) \cdot P(\overline{\mathrm{S}})$).

Da nur ganz wenige Grenzgänger schmuggeln, bellt der Hund viel öfter versehentlich ($P(\mathrm{B} \cap \overline{\mathrm{S}})$ = 0,0297 versus ($P(\mathrm{B} \cap \mathrm{S})$ = 0,0098)! Dagegen ist fast stets , wenn er nicht bellt,

	B	$\overline{\mathrm{B}}$	
S	0,0098	0,0002	0,01
$\overline{\mathrm{S}}$	0,0297	0,9603	0,99
	0,0395	0,9605	

auch tatsächlich kein Grund vorhanden ($P(\overline{\mathrm{B}} \cap \overline{\mathrm{S}}) = 0{,}9603$ versus $P(\overline{\mathrm{B}} \cap \mathrm{S}) = 0{,}0002$)!

Beispiel 2: AIDS-Test

Ihr AIDS-Test ist positiv. Ist Panik angesagt?

Analyse:

A = „Jemand ist HIV-positiv“
T = „Test ist positiv“

Wir gehen von folgenden Daten aus:

$P(\mathrm{A}) = 0{,}001$	Prävalenz von HIV-Infektion in der Bevölkerung
$P(\mathrm{T/A}) = 0{,}99$	Sensitivität des Test
$P(\overline{\mathrm{T}}/\overline{\mathrm{A}}) = 0{,}98$	Spezifizität des Test

Interessant hier ist P(A/T)!

Bayes´sche Regel

$$P(\mathrm{A/T}) = \frac{P(\mathrm{A}) \cdot P(\mathrm{T/A})}{P(\mathrm{A}) \cdot P(\mathrm{T/A}) + P(\overline{\mathrm{A}}) \cdot P(T/\overline{\mathrm{A}})} = \frac{0{,}001 \cdot 0{,}99}{0{,}001 \cdot 0{,}99 + 0{,}999 \cdot 0{,}02} = 0{,}05$$

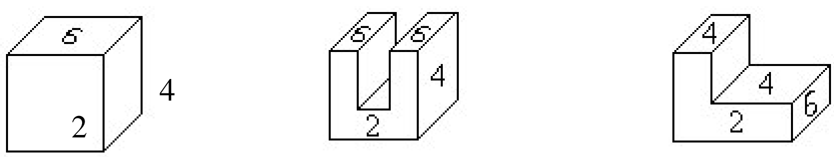
Die Chance, wirklich AIDS zu haben, ist also nur 5%. Entsprechend gering ist die Chance, dass ein AIDS-Fall nicht entdeckt wird.

$$P(\mathrm{A}/\overline{\mathrm{T}}) = \frac{P(\mathrm{A}) \cdot P(\overline{\mathrm{T}}/\mathrm{A})}{P(\mathrm{A}) \cdot P(\overline{\mathrm{T}}/\mathrm{A}) + P(\overline{\mathrm{A}}) \cdot P(\overline{\mathrm{T}}/\overline{\mathrm{A}})} = \frac{0{,}001 \cdot 0{,}01}{0{,}001 \cdot 0{,}01 + 0{,}999 \cdot 0{,}98} \approx 0{,}00001$$

Praktische Anwendung: billiger Massentest, um die potentiell Kranken heraus zu sieben, dann teure Spezialtests (vgl. Röntgen-Reihenuntersuchung).

Beispiel 3: „Beweiswürdigung“

Überprüfen und Revidieren von Hypothesen aufgrund neuer Informationen nennen die Juristen „Beweiswürdigung“. Armin hat 3 „Würfel“, einen Laplace-Würfel W, und je einen Riemer-U- und -L-Würfel:



W U L

Armin würfelt und nennt Beate seine Zahlen, zeigt aber nicht den verwendeten Würfel. Den soll Beate raten. Er würfelt dreimal, zuerst 3, dann 1 und schließlich 5. Welche Hypothesen kann Beate entwickeln?
Dies ist ein typisches Beispiel für die Bayes´sche Regel. Aufgrund von Experimenten mit U- und L-Würfeln hat Beate folgende Wahrscheinlichkeitsverteilungen für die 3 Würfel aufgestellt:

		1	2	3	4	5	6
W	Laplace-Würfel	0,17	0,17	0,17	0,17	0,17	0,17
L	L-„Würfel“	0,01	0,14	0,21	0,40	0,14	0,10
U	U-„Würfel“	0,24	0,06	0,10	0,10	0,06	0,44

<u>1. Schritt</u>: „3 fällt“

a-priori-Wahrscheinlichkeit
$P(\mathrm{W}) = P(\mathrm{L}) = P(\mathrm{U}) = 1/3$

A = „3 fällt“ tritt ein; es ist bekannt: $P(3/\mathrm{W}) = 0{,}17$, $P(3/\mathrm{L}) = 0{,}24$ und $P(3/\mathrm{V}) = 0{,}06$

a-posteriori-Wahrscheinlichkeiten (mit der Beyes´schen Regel)

$$P(\mathrm{W}/3) = \frac{P(\mathrm{W})\cdot P(3/\mathrm{W})}{P(\mathrm{W})\cdot P(3/\mathrm{W}) + P(\mathrm{L})\cdot P(3/\mathrm{L}) + P(\mathrm{U})\cdot P(3/\mathrm{U})}$$

$$= \frac{0{,}33\cdot 0{,}17}{0{,}33\cdot 0{,}17 + 0{,}33\cdot 0{,}21 + 0{,}33\cdot 0{,}10} \approx 0{,}35$$

$P(\mathrm{L}/3) = \ldots \approx 0{,}44$

$P(\mathrm{U}/3) = \ldots \approx 0{,}21$

<u>2. Schritt</u>: „1 fällt“

Neue a-priori-Wahrscheinlichkeiten sind die alten a-posteriori-Wahrscheinlichkeiten:
$P(\mathrm{W}) = 0{,}35$, $P(\mathrm{L}) = 0{,}44$, $P(\mathrm{U}) = 0{,}21$

A = „1 fällt“ tritt ein,
es ist bekannt $P(1/\mathrm{W}) = 0{,}17$, $P(1/\mathrm{L}) = 0{,}01$, $P(1/\mathrm{U}) = 0{,}24$

a-posteriori-Wahrscheinlichkeiten

$$P(\mathrm{W}/1) = \frac{P(\mathrm{W})\cdot P(1/\mathrm{W})}{P(\mathrm{W})\cdot P(1/\mathrm{W}) + P(\mathrm{L})\cdot P(1/\mathrm{L}) + P(\mathrm{U})\cdot P(1/\mathrm{U})}$$

$$= \frac{0{,}35\cdot 0{,}17}{0{,}35\cdot 0{,}17 + 0{,}44\cdot 0{,}01 + 0{,}21\cdot 0{,}24} \approx 0{,}52$$

$P(\mathrm{L}/1) = \ldots \approx 0{,}04$

$P(\mathrm{U}/1) = \ldots \approx 0{,}44$

<u>3. Schritt</u>: “5 fällt” analog

a-priori: $P(\mathrm{W}) = 0{,}52$, $P(\mathrm{L}) = 0{,}04$, $P(\mathrm{U}) = 0{,}44$

a-posteriori: $P(5/\mathrm{W}) \approx 0{,}73$, $P(5/\mathrm{L}) \approx 0{,}05$, $P(5/\mathrm{U}) \approx 0{,}22$.

Jetzt kann Beate mit großer Sicherheit behaupten, dass Armin den Laplace-Würfel verwendet hatte.

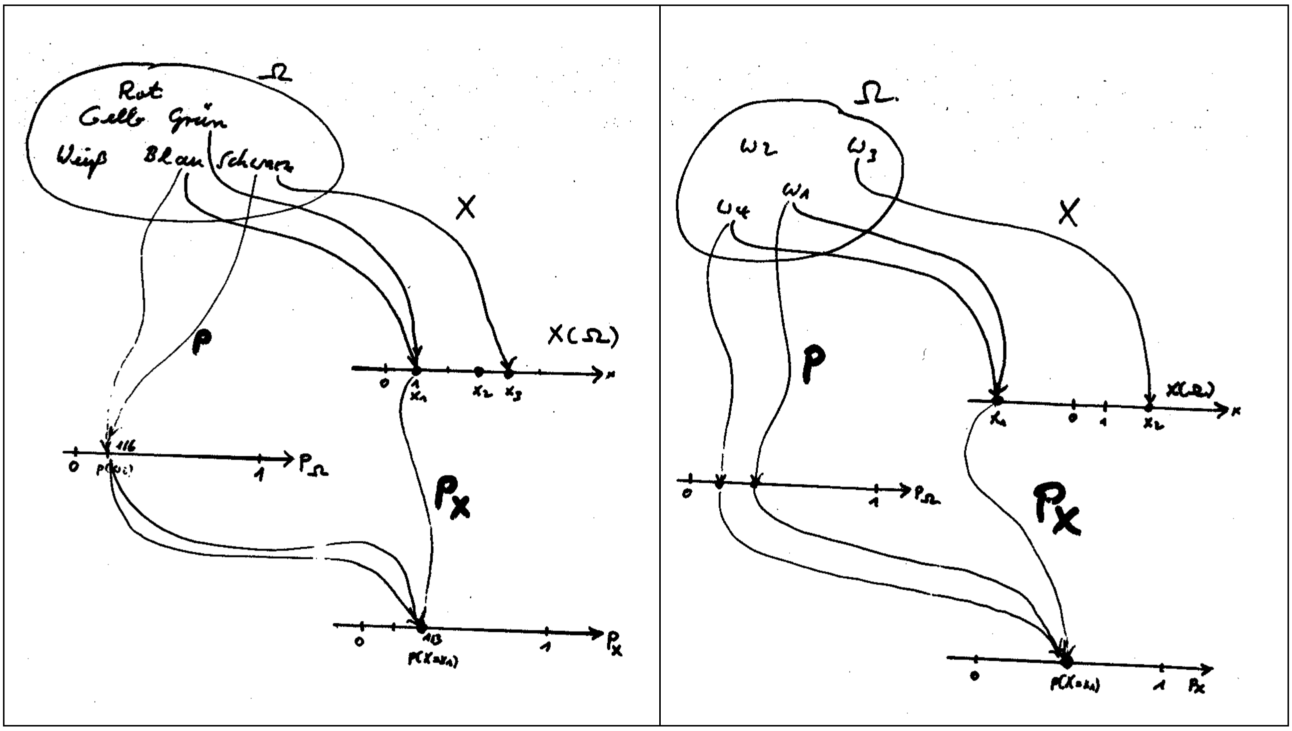
2. Wahrscheinlichkeitsverteilungen

2.1 Zufallsvariable

2.1.1 Zufallsvariable und ihre Wahrscheinlichkeitsverteilungen

Problemstellung
Wirft man einen Würfel oder zählt man die Autoinsassen der Autos, die morgens an einer Zählstelle vorbeikommen, so kann man z. B. Mittelwerte ausrechnen und sagen, im Mittel hat man 3,2 gewürfelt oder im Mittel saßen 1,6 Personen in den Autos. Aus den Häufigkeiten kann man Wahrscheinlichkeiten prognostizieren und z. B. auch sagen, beim weiteren Würfeln ist mit der Augenzahl 3,5 zu rechnen und in den weiteren Autos werden im Mittel 1,6 Personen sitzen.

Verwenden wir dagegen einen Farbwürfel mit den Farben Rot, Grün, Gelb, Weiß, Schwarz, Blau, beobachten das Fallen eines Reißnagels mit seinen beiden möglichen Lagen oder bestimmen beim Pkw-Beispiel das Geschlecht des Fahrers, so sind Mittelwerte sinnlos. Rechenoperationen lassen sich nur auf Zahlen anwenden, die Elementarereignisse, die wir beobachten, müssen also Zahlen sein. Andernfalls muss man jedem Elementarereignis eine Zahl zuordnen. Dies ist nichts anderes als eine (zunächst willkürliche) Funktion, die in der Stochastik Zufallsvariable (oder Zufallsgröße) heißt. Die Wahrscheinlichkeiten „vererben sich".



Definition
Ω sei ein (endlicher) Ergebnisraum mit Wahrscheinlichkeitsverteilung P. Eine Zufallsgröße (Zufallsvariable) ist eine Funktion X: $\Omega \to \mathbb{R}$, $\omega \mapsto a = X(\omega)$.

Leicht sieht man, dass X(Ω) ein neuer Ergebnisraum ($X(\Omega) \subseteq \mathbb{R}$) ist und dass P eine neue Wahrscheinlichkeitsverteilung P_X induziert durch:

$$\begin{tikzcd}
\Omega \arrow[rr, "P"] \arrow[dr, "X"'] & & {[0;1]} \\
& X(\Omega) \arrow[ur, "P_X"'] &
\end{tikzcd}$$

Genauer gilt $P_X(x) = P(X^{-1}(x)) = \sum\limits_{\substack{\omega \\ X(\omega)=x}} P(\omega)$.

<u>Bemerkung:</u> Es reicht, P_X für die Elementarereignisse zu definieren. Da bei uns Ω endlich (oder abzählbar unendlich ist), kann P_X als Funktion von $\mathbb{R} \to [0,1]$ betrachtet werden.

Jetzt sind Fragen der beurteilenden Statistik beantwortbar. X(Ω) fasst gewisse der bisherigen Ergebnisse so zusammen, wie es für die jeweilige Problemstellung zweckmäßig erscheint.

Übliche Sprechweise: „Zufallsgröße X hat Wert a angenommen“ statt „Ereignis ‚X = a’ ist eingetreten“, d. h. „für das aufgetretene Ergebnis ω gilt X(ω) = a“. Die entsprechende Wahrscheinlichkeit $P_X(a)$ wird auch als P(X=a) bezeichnet.

<u>Beispiel 1</u>: Skatspiel

$$\Omega = \{\text{KreuzAs, PikAs, ... , KaroSieben}\} = \{\omega_1, \omega_2, \dots, \omega_{32}\}$$

Unsere Zufallsgröße X ordne jeder Karte ihren Wert zu:

$$X: \Omega \to \mathbb{R}, \ \omega \to \text{Wert von } \omega.$$

Damit erhalten wir den neuen Ergebnisraum X(Ω) = {0, 2, 3, 4, 10, 11} mit der induzierten Wahrscheinlichkeitsverteilung

a	0	2	3	4	10	11
P(X = a)	3/8	1/8	1/8	1/8	1/8	1/8

<u>Beispiel 2</u>: Chuck-a-luck

Werfen dreier Würfel, $\Omega = \{1, 2, ..., 6\}^3$, $|\Omega| = 216$

Vorher Setzen auf eine Ziffer u, Einsatz e; bei i-maligem Auftreten von u Auszahlung von (i + 1)·e (für i ≥ 1), sonst Einsatz futsch.

ω	Ergebnisse	111	112	...	221	222	...	666
X(ω)	Reingewinn	-1	1	...	2	3	...	-1

Beispiel: u = 2, e = 1 (in DM); dann Zuordnung X: $\Omega \to \mathbb{R}$, mit Wertemenge {-1, 1, 2, 3}.

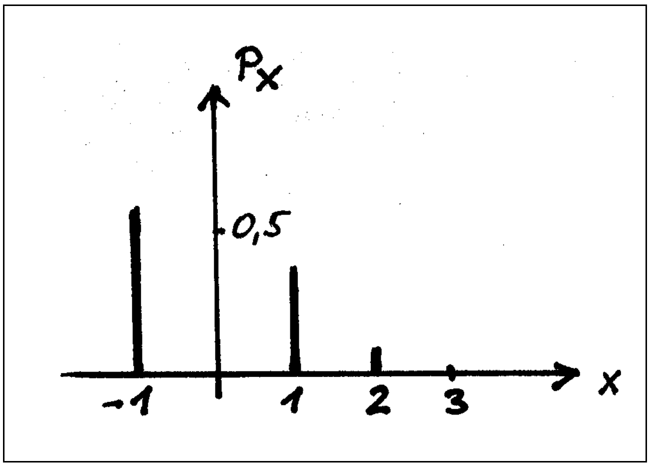
Wertetabelle:

$$P_X(3) = \frac{1}{216},\; P_X(2) = 3\cdot\left(\frac{1}{6}\right)^2\cdot\frac{5}{6} = \frac{15}{216},\; P_X(1) = 3\cdot\frac{1}{6}\cdot\left(\frac{5}{6}\right)^2 = \frac{75}{216}$$

a	-1	1	2	3	sonst.
P(X = a)	$\frac{125}{216}\approx 0{,}58$	$\frac{75}{216}\approx 0{,}35$	$\frac{15}{216}\approx 0{,}07$	$\frac{1}{216}\approx 0{,}00$	0

Wenn wir P_X als Funktion von $\mathbb{R} \to [0,1]$ betrachten, so ist $P_X(a) = 0$ für $a \in \mathbb{R} \setminus \{-1, 1, 2, 3\}$ zu setzen.

Veranschaulichung der „Verteilung“ P_x durch ein Stabdiagramm:



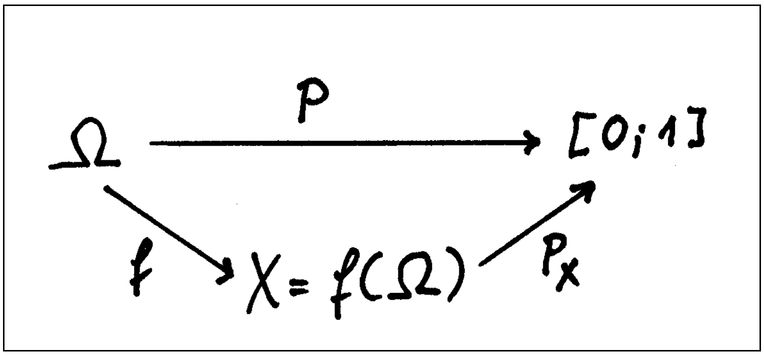
Noch offene Frage: ist dieses Spiel gerecht?

<u>Didaktischer Kommentar</u>

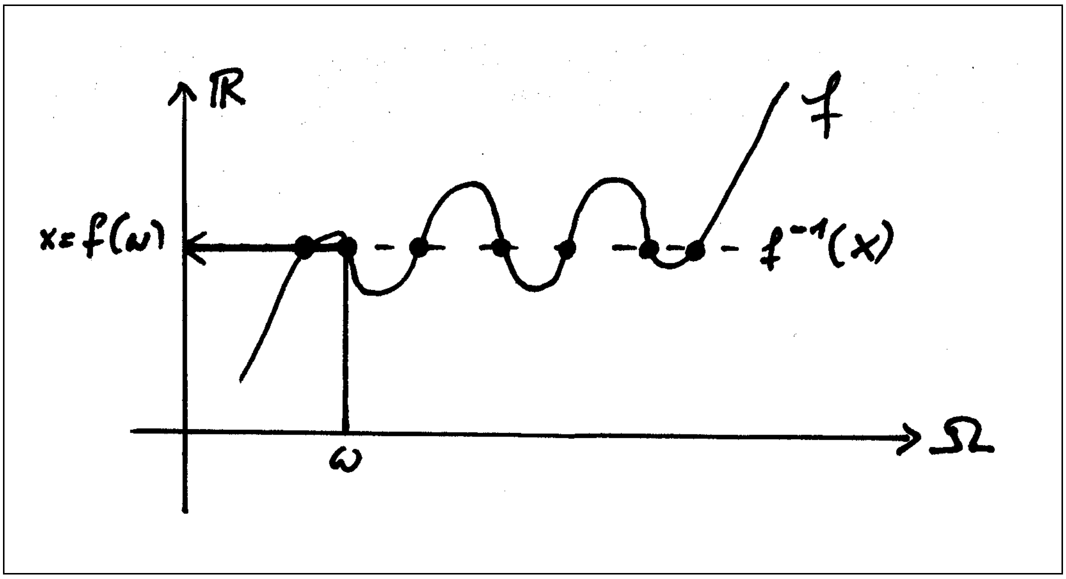
a) Die Berechnungen sind sehr gewöhnungsbedürftig und weichen von den üblichen Funktionsschreibweisen ab. Allein hieraus erwachsen Verständnisprobleme.

$$P(X=a)=P_X(a)$$
$$P(a \le X \le b)=\sum_{\substack{c\ mit\\ a\le c\le b}} P(X=c)$$
$$P(X \ge a)=\sum_{\substack{c\ mit\\ c\ge a}} P(X=c)$$

b) Die übliche Funktionsschreibweise zeigt, dass hier eine andere Sichtweise als üblich im Vordergrund steht.



mit $P_X(x)=P\left(f^{-1}(x)\right)$



Zur Bestimmung von P_X ist „gegeben: x, gesucht: $f^{-1}(x)$“ das Problem, d. h. „waagerechte Schnitte“, während sonst „senkrechte Schnitte“ ($x \mapsto f(x)$) gefragt sind.

Jede zu einer Zufallsgröße X gehörige Wahrscheinlichkeitsverteilung induziert eine Verteilungsfunktion:

Definition: $F:\mathbb{R}\to\mathbb{R},\ x\mapsto F(x)=P(X\le x)=\sum_{\substack{a\in X(\Omega)\\ a\le x}} P(X=a)$

heißt Verteilungsfunktion von X.

Beachte: F ist auf $\mathbb{R}$, nicht nur auf X(Ω) definiert! F wächst monoton von 0 auf 1.

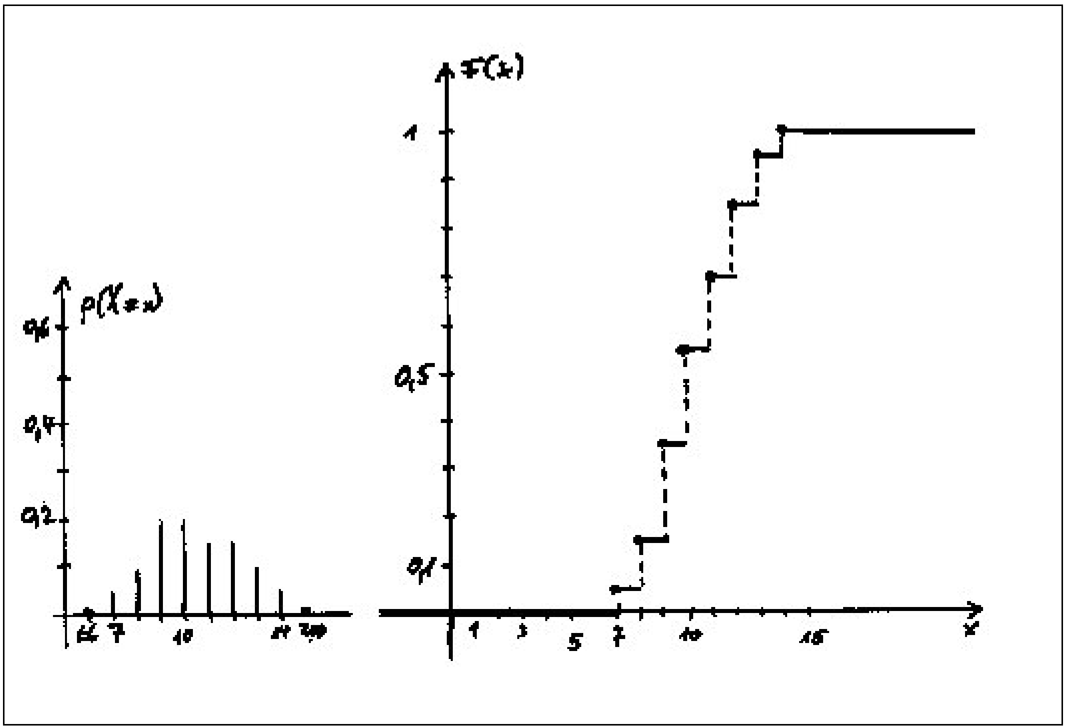
Beispiel: Erfahrungsgemäße Heilungsdauer bei Knieverletzungen

x/Wochen	≤ 6	7	8	9	10	11	12	13	14	> 14
p(X=x) in %	0	5	10	20	20	15	15	10	5	0

Der Torwart ist verletzt. Wie groß ist die Wahrscheinlichkeit, dass er
- 10 – 13 Wochen ausfällt
- höchstens 10 Wochen ausfällt
- mehr als 13 Wochen ausfällt

Aus der Tabelle ergibt sich:

$p_1 = P(10 \le X \le 13) = 0{,}2 + 0{,}15 + 0{,}15 + 0{,}1 = 0{,}6 = F(13) - F(9)$
$p_2 = F(10) = 0{,}55$
$p_3 = 1 - F(13) = 0{,}05$



Eigenschaften von Verteilungsfunktionen F
- F monoton steigend in $\mathbb{R}$
- $F(b) - F(a) = P(a < X \leq b)$ für alle $a, b \in \mathbb{R}$ mit $a \leq b$
- $P(X > a) = 1 - F(a)$
- $P(X = x_i) = F(x_i) - F(x_{i-1})$, falls $x_1 < x_2 < \ldots < x_n$ und $X(\Omega) = \{x_1, x_2, ..., x_n\}$

2.1.2 Stochastische Kenngrößen von Zufallsvariablen

Wir knüpfen an die Begriffsbildung der beschreibenden Statistik (vgl. 1.1) an.

Statistik Ich würfle n mal und berechne:	Wahrscheinlichkeitstheorie Ich sage für die Zukunft voraus:
die relative Häufigkeit h für 6	mit welcher Wahrscheinlichkeit p werfe ich eine 6,
den Mittelwert $\overline{x}$	welche Zahl $\mu = E(X)$ werde ich im Mittel werfen,
die (empirische) Standardabweichung s als Streumaß der Messreihe	mit welcher Standardabweichung σ die gewürfelten Zahlen streuen werden

Die Definitionen lassen sich direkt übertragen (dabei sei $X(\Omega) = \{x_1, x_2, ..., x_n\}$:

$h = \frac{a}{m}$ (a günstige bei insgesamt m Ergebnissen)	P
$\overline{x} = \frac{a_1 x_1 + a_2 x_2 + \ldots + a_n x_n}{m}$ $= \frac{a_1}{m} x_1 + \frac{a_2}{m} x_2 + \ldots + \frac{a_n}{m} x_n$ $= h_1 x_1 + h_2 x_2 + \ldots + h_n x_n$	$\mu = E(X) = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n$
$s = \sqrt{\sum_{i=1}^{n} h_i (x_i - \overline{x})^2}$	$\sigma = \sqrt{\sum_{i=1}^{n} p_i (x_i - \mu)^2}$
(empirische) Varianz s^2	Varianz σ^2

Definition: Seien P eine Wahrscheinlichkeitsverteilung und X eine Zufallsgröße auf Ω. Dann heißt die Zahl

$$\mu = E(X) := \sum_{a \in \mathbb{R}} a \cdot P_X(a)$$

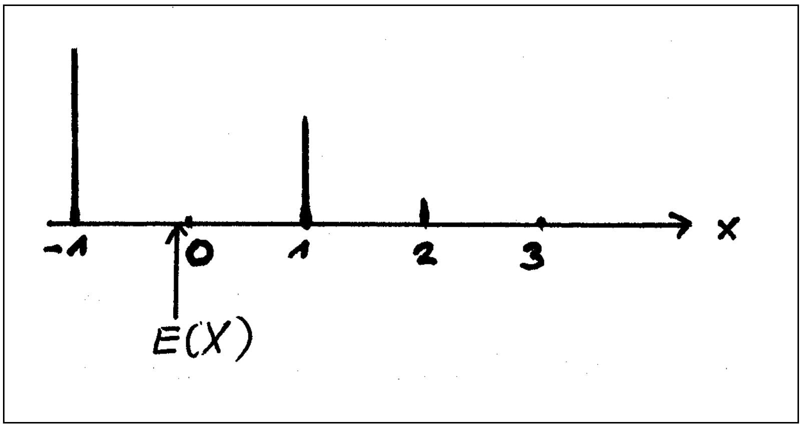
(sofern sie existiert) der Erwartungswert von X (wir könnten natürlich auch nur über $a \in X(\Omega)$ summieren).

Beispiel 1: „Chuck-a-luck"

X: Gewinn (vgl. Beispiel 2 auf S. 68)

$$E(X) = (-1) \cdot \frac{125}{216} + 1 \cdot \frac{75}{216} + 2 \cdot \frac{15}{216} + 3 \cdot \frac{1}{216} = -\frac{17}{216} \approx -0,07$$

Das Spiel ist also nicht fair, man verliert im Mittel $\frac{17}{216}$ des Einsatzes („fair" wäre E(X) = 0).



Der Erwartungswert ist zu unterscheiden vom „Modalwert", d. h. vom Wert mit höchster Wahrscheinlichkeit; hier: -1. Weiter ist der – dem praktischen „Median" entsprechende „Zentralwert" zu unterscheiden, d. h. der Wert „in der Mitte"; hier 1.5 (zwischen 1 und 2).

Beispiel 2:
Ich würfle bis zur ersten „6". Wie viele Würfe brauche ich wohl durchschnittlich?
Allgemein: Ich warte auf das Eintreten des Ereignisses E mit P(E) = p. Wie lange muss ich wohl im Mittel warten?

Lösung
Konkrete Experimente liefern einen Mittelwert von $\bar{x} \approx 6$.

3 Argumente: X: Anzahl der Würfe bis zur ersten „6".

- Im stochastischen (Laplace-) Modell sind alle sechs Zahlen gleichberechtigt. Also tritt beim Würfeln aus Symmetriegründen jede Zahl im Mittel jedes sechste Mal auf. Somit gilt E(X) = 6.

- Andere (verallgemeinerbare) Überlegung: Bei n Würfen erwarten wir etwa k mal „6", wobei $\frac{k}{n} \approx \frac{1}{6}$. Die mittlere Anzahl der Versuche bis zu einer „6" ist somit (klar!) $\frac{n}{k} \approx 6.$

- formale Bestätigung:

  p_n = P („genau beim n-ten Wurf erste „6") = $\left(\frac{5}{6}\right)^{n-1} \cdot \frac{1}{6}$,

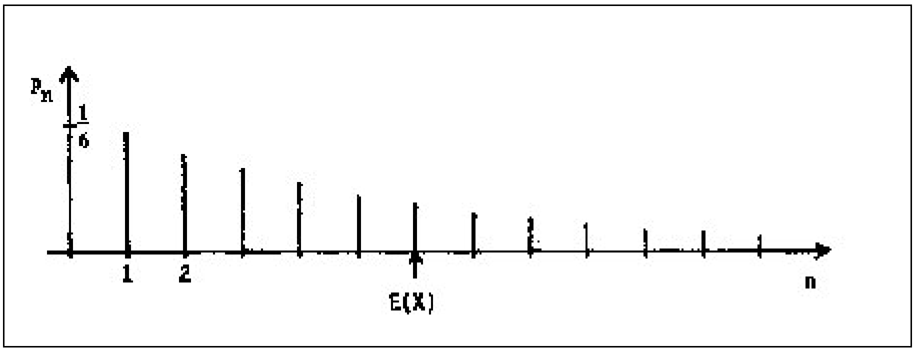
  also gilt

  $$E(X) = \sum_{n=1}^{\infty} n \cdot p_n = \sum_{n=1}^{\infty} n \cdot \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} = \frac{1}{6} \cdot \frac{6}{5} \cdot \sum_{n=1}^{\infty} n \left(\frac{5}{6}\right)^{n}$$

  Hinweis (ohne Beweis): $\sum_{n=1}^{\infty} na^n = \frac{a}{(1-a)^2} \quad \textit{für } |a| < 1$,

  also $E(X) = \frac{1}{5} \cdot \frac{\frac{5}{6}}{\left(1 - \frac{5}{6}\right)^2} = 6,$ wie zu erwarten war!

Graphische Veranschaulichung:



Im allgemeinen Fall sei X: Anzahl der Versuche bis zum Eintreten von E. Genau wie eben überlegt man: Bei n Versuchen k-maliges Eintreten von E, wobei $\frac{k}{n} \approx p.$ Somit mittlere Anzahl der Versuche bis zu E gleich $\frac{n}{k} \approx \frac{1}{p},\ \textit{also}\ E(X) = \frac{1}{p}.$

Definition der Streumaße:

Seien P ein Wahrscheinlichkeitsmaß und X eine Zufallsgröße auf Ω, mit Erwartungswert μ = E(X). Dann heißt die Zahl

$$V(X) = \sum_{a \in \mathbb{R}} (a - \mu)^2 \cdot P_X(a)$$

die Varianz und die Zahl $\sigma = \sqrt{V(X)}$ die Standardabweichung von X (falls ex.).

σ-Umgebungen bzw. nσ-Umgebungen sind die Intervalle [μ-σ, μ+σ] bzw.[μ-n·σ, μ+n·σ] für n = 1, 2, ...

Beispiel 1

A und B spielen Roulette. A setzt auf „17 plein“, B auf „rouge“. Beurteilen Sie die Risiken!

x	-1	35
$P(X=x)$	$36/37$	$1/37$

$$\mu = -1 \cdot \frac{36}{37} + 35 \cdot \frac{1}{37} = -\frac{1}{37}$$

$$\sigma = \sqrt{\left(-1 + \frac{1}{37}\right)^2 \cdot \frac{36}{37} + \left(35 + \frac{1}{37}\right)^2 \cdot \frac{1}{37}} = \sqrt{3408} = 58{,}4$$

x	-1	1
$P(X=x)$	$19/37$	$18/37$

$$\mu = -1 \cdot \frac{19}{37} + 1 \cdot \frac{18}{37} = -\frac{1}{37}$$

$$\sigma = \sqrt{\left(-1 + \frac{1}{37}\right)^2 \cdot \frac{19}{37} + \left(1 + \frac{1}{37}\right)^2 \cdot \frac{18}{37}} = \sqrt{99{,}9} = 10{,}0$$

Trotz gleichen Erwartungswertes geht A ein viel größeres Risiko ein.

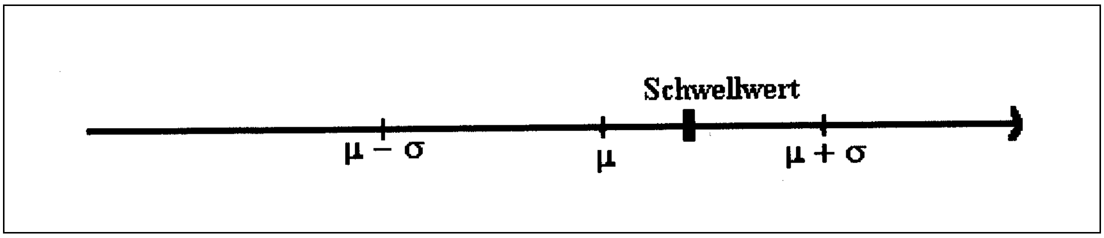
Beispiel 2

Für die Praxis interessant sind die σ-Umgebungen: Für annähernd symmetrische Verteilungen sind

$$\text{ca.} \begin{Bmatrix} 68\% \\ 95\% \\ 99\% \end{Bmatrix} \text{der Ergebnisse im Intervall} \begin{Bmatrix} [\mu\text{-}\sigma;\mu+\sigma] \\ [\mu\text{-}2\sigma;\mu+2\sigma] \\ [\mu\text{-}3\sigma;\mu+3\sigma] \end{Bmatrix}$$

zu erwarten (genaueres auf S. 104).

Leider wird bei Wahlprognosen immer nur der zu erwartende Mittelwert μ angegeben. Noch schlimmer ist es z. B. bei folgendem Beispiel: Der Firma BASF leitet jeden Tag eine mehr oder weniger große Menge Chemikalien in den Rhein. Gefährlich für Fische ist erst eine Einleitung oberhalb eines gewissen Schwellwertes. Angegeben wird von der Firma der zu erwartende Mittelwert μ der Gifteinleitung, der weit unterhalb des Schwellenwertes liegt. Das schließt aber nicht aus, dass schon μ+σ größer als der Schwellwert ist.



2.1.3 Verknüpfungen von Zufallsgrößen

2.1.3.1 Definition

Wie bei Abbildungen (= Funktionen) üblich, lassen sich auch Zufallsgrößen verknüpfen.
X, Y seien Zufallsgrößen auf demselben Ergebnisraum Ω. Dann gilt

$$Z := X + Y : \Omega \to \mathbb{R},\ \omega \mapsto Z(\omega) := X(\omega) + Y(\omega).$$

Für die induzierte Wahrscheinlichkeitsverteilung P_z gilt

$$P_z(a) = \sum_{\substack{\omega \\ Z(\omega)=a}} P(\omega) = \sum_{\substack{\omega \\ X(\omega)+Y(\omega)=a}} P(\omega)$$

Diese Darstellung vereinfachen wir für endliches Ω. (analog geht es für abzählbar unendliches Ω):

Es seien $a = x_1 + y_1 = x_2 + y_2 = \ldots = x_n + y_n$ die n verschiedenen Darstellungen von a als Summe mit $x_i \in X(\Omega)$ und $y_i \in Y(\Omega)$. Damit gilt

$$P_z(a) = \sum_{i=1}^{n} \sum_{\substack{\omega \\ X(\omega)=x_i \wedge \\ Y(\omega)=y_i}} P(\omega) = \sum_{i=1}^{n} P(X = x_i \wedge Y = y_i) = \sum_{\substack{x+y=a \\ x \in X(\Omega), y \in Y(\Omega)}} P(X = x \wedge Y = y)$$

Die nötigen Wahrscheinlichkeiten $P(X = x \wedge Y = y)$ gehören zur sogenannten gemeinsamen Verteilung von X und Y:

Definition: X und Y seien Zufallsgrößen auf demselben endlichen Ereignisraum Ω. Dann heißt

$$f_{X,Y} :\ \mathbb{R} \times \mathbb{R} \to [0,1], (a,b) \mapsto P(X = a \wedge Y = b)$$

die gemeinsame Verteilung von X und Y.

Analog sind definiert

$$Z = X \cdot Y : \Omega \to \mathbb{R},\ \omega \mapsto Z(\omega) := X(\omega) \cdot Y(\omega)$$

$$P(Z = a) = P_z(a) = \sum_{\substack{\omega \\ X(\omega)\cdot Y(\omega)=a}} P(\omega) = \sum_{\substack{x\cdot y=a \\ x\in X(\Omega),\, y\in Y(\omega)}} P(X = x \wedge Y = y)$$

$$Z = k \cdot X : \Omega \to \mathbb{R},\ \omega \mapsto Z(\omega) = k \cdot X(\omega) \ \ \textit{für}\ k \in \mathbb{R}$$

$$P(Z = a) = P_z(a) = \sum_{\substack{\omega \\ h\cdot X(\omega)=a}} P(\omega) = \sum_{\substack{k\cdot x=a \\ x\in X(\Omega)}} P(X = x)$$

und entsprechend für die anderen möglichen Verknüpfungen.

2.1.3.2 Beispiele

Beispiel 1: Würfeln mit 2 Würfeln
X: Pasch dabei? X(Pasch) = 1, X($\overline{Pasch}$) = 0
Y: Augensumme

Verteilung von X:

X	0	1
P(X=x)	30/36	6/36

Verteilung von Y:

Y	2	3	4	5	6	7	8	9	10	11	12
P(Y=y)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Gemeinsame Verteilung von X und Y :

Y X	2	3	4	5	6	7	8	9	10	11	12
0	0/36	2/36	2/36	4/36	4/36	6/36	4/36	4/36	2/36	2/36	0/36
1	1/36	0/36	1/36	0/36	1/36	0/36	1/36	0/36	1/36	0/36	1/36

U : = X+Y
U(22) = X(22) + Y(22) = 1 + 4 = 5

$$P(U=5)=\sum_{\substack{x+y=5\\ x\in X(\Omega),\,y\in Y(\Omega)}} P(X=x \wedge Y=y)$$

$$=P(X=0 \wedge Y=5)+P(X=1 \wedge Y=4)=\frac{4}{36}+\frac{1}{36}=\frac{5}{36}$$

$$V:=\frac{X\cdot Y}{2}$$

$$V(22)=\frac{1\cdot 4}{2}=2$$

$$P(V=2)=\sum_{\substack{\frac{X\cdot Y}{2}=2\\ x\in X(\Omega),\,y\in Y(\Omega)}} P(X=x \wedge Y=y)=P(X=1 \wedge Y=4)=\frac{1}{36}$$

Beispiel 2 : Würfeln mit 3 Würfeln

X: „Gewinn beim chuck-a-luck (1 DM Einsatz, Ziffer 2)"
Y: „Zweite Ziffer"
Z: „Summe der Ziffern"

Hieraus neue Zufallsgrößen, z. B.

U = X + Y + Z, V = $\frac{Y\cdot Z}{2}$

U(123) = 1 + 2 + 6 = 9

$$P(U=9)=\sum_{x+y+z=9} P(X=x \wedge Y=y \wedge Z=z)=\ldots \qquad A)$$

$$V(123)=\frac{2\cdot 6}{2}=6$$

$$P(V=6)=\sum_{\frac{y\cdot z}{2}=6} P(Y=y \wedge Z=z)=\ldots \qquad B)$$

Übungsaufgabe ! Lösung für die obigen Beispiele : Dabei bedeutet „./.", dass es keine Lösung gibt.

A) Fallunterscheidung nach Y = 1, …, 6, dann jeweils X = -1, 1, 2, 3.
Hieraus ergibt sich dann Z:

Y = 1:	X = -1, also Z = 9	zugehörige Tripel	(3/1/5), (4/1/4), (5/1/3)
	X = 1, also Z = 7		(2/1/4), (4/1/2)
	X = 2, also Z = 6		./.
	X = 3, also Z = 5		./.

Y = 2 : X = -1, also Z = 8	./.
X = 1, also Z = 6	(1/2/3), (3/2/1)
X = 2, also Z = 5	(2/2/1), (1/2/2)
X = 3, also Z = 4	./.
Y = 3 : X = -1, also Z = 7	(1/3/3), (3/3/1)
X = 1, also Z = 5	(2/3/4), (4/3/2)
X = 2, also Z = 4	./.
X = 3, also Z = 3	./.
Y = 4 : X = -1, also Z = 6	(1/4/1)
X = 1, also Z = 4	./.
X ≥ 2, also Z ≤ 3	./.
Y = 5 : X = -1, also Z = 5	./.
X ≥ 1, also Z ≤ 3	./.
Y = 6 : X ≥ -1, also Z ≤ 4	./.

Damit gilt $P(U = 9) = \frac{3}{216}+\frac{2}{216}+\frac{2}{216}+\frac{2}{216}+\frac{2}{216}+\frac{2}{216}+\frac{1}{216}=\frac{14}{216}=\frac{7}{108}$

B) $\frac{Y \cdot Z}{2} = 6$

Y = 1, also Z = 12	zugehöriger Tripel	(5/1/6), (6/1/5)
Y = 2, also Z = 6	´´	(1/2/3), (2/2/2), (3/2/1)
Y = 3, also Z = 4		./.
Y ≥ 4, also $Z \leq \frac{12}{4} = 3$		./.

Damit folgt $P(V = 6) = P(Y = 1 \wedge Z = 12) + P(Y = 2 \wedge Z = 6) = \frac{2}{216}+\frac{3}{216}=\frac{5}{216}$

<u>Beispiel 3 :</u> Roulette (europäische Variante ; in den USA gibt es noch die 00)
$\Omega = \{0, 1, 2, \ldots, 36\}$

Beim Roulette gibt es 16 Einsatzmöglichkeiten mit folgenden Reingewinnen in Vielfachen des Einsatzes :

„Eine volle Zahl“ (pleine)	35facher Einsatz
„Zwei verbundene Zahlen“(cheval)	17facher Einsatz
„Eine Querreihe von 3 Zahlen“ (transversale pleine)	11facher Einsatz
„Vier Zahlen im Viereck“ (carré)	8facher Einsatz
„Die ersten 4 Zahlen“ (0, 1, 2, 3)	8facher Einsatz
„Zwei Querreihen mit 6 Zahlen“ (transversale pleine)	5facher Einsatz
„Eine Längsreihe von 12 Zahlen“ (colonne)	2facher Einsatz
„Das erste Dutzend 1-12“ (12p)	2facher Einsatz
„Das zweite Dutzend 13-24“ (12 m)	2facher Einsatz
„Das dritte Dutzend 25-36“ (12 d)	2facher Einsatz
„Sämtliche gerade Zahlen“ (pair)	1facher Einsatz
„Sämtliche ungerade Zahlen“ (impair)	1facher Einsatz
„Sämtliche rote Zahlen“ (rouge)	1facher Einsatz
„Sämtliche schwarze Zahlen“ (noir)	1facher Einsatz
„Die erste Hälfte aller Zahlen von 1-18“ (manque)	1facher Einsatz
„Die zweite Hälfte aller Zahlen von 19-36“ (passe)	1facher Einsatz

Die Einsatzmöglichkeiten werden bezüglich ihrer Gewinnerwartung durch je eine Zufallsgröße beschrieben,

z. B. „1. Dutzend“ $A = \{1, ..., 12\}$, $\overline{A} = \{0, 13 ... 36\}$

$$X : \omega \mapsto 2 \quad \textit{für } \omega \in A$$
$$\omega \mapsto -1 \quad \textit{für } \omega \in \overline{A}$$

„1. Querreihe von 3 Zahlen“, $B = \{1, 2, 3\}$, $\overline{B} = \Omega \setminus B$

$$Y : \omega \mapsto 11 \quad \textit{für } \omega \in B$$
$$\omega \mapsto -1 \quad \textit{für } \omega \in \overline{B}$$

$$P(X = 2) = \frac{12}{37}, \; P(X = -1) = \frac{25}{37}$$

$$P(Y = 11) = \frac{3}{37}, \; P(Y = -1) = \frac{34}{37}.$$

Nun setzt der Ehemann die erste, die Frau die zweite Möglichkeit. Was ist die Wahrscheinlichkeit, dass beide gewinnen?

$$P(X = 2 \wedge Y = 11) = \frac{3}{37} \qquad \text{(zugehörige } \omega \in \{1, 2, 3\}\text{)}$$

Wahrscheinlichkeit, dass Frau verliert und Mann gewinnt?

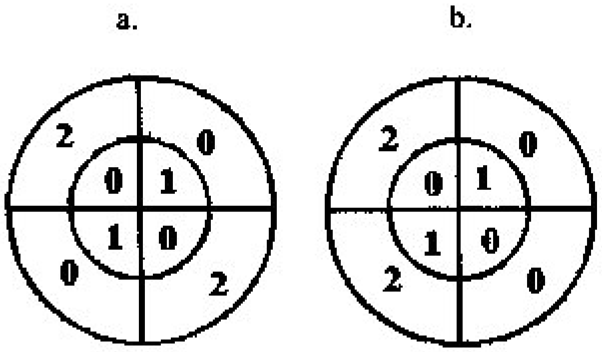
$$P(X = 2 \wedge Y = -1) = \frac{9}{37} \qquad \text{(zugehörige } \omega \in \{4, ..., 12\}\text{)}$$

In beiden Fällen gilt

$$P(X = a \ \wedge \ Y = b) \neq P(X = a) \cdot P(Y = b),$$

also Abhängigkeit der Ereignisse.

<u>Beispiel 4:</u>



Wir definieren jeweils zwei Zufallsgrößen:
X: Zahl außen
Y: Zahl innen

Bestehen zwischen X und Y „Abhängigkeiten"?
Dazu berechnen wir bedingte Wahrscheinlichkeiten P(X=a | Y=b).
Übersicht durch Vierfeldertafeln der gemeinsamen Verteilung P(X=a ∧ Y=b):

a.

Y \ X	0	2	
0	0	0,5	0,5
1	0,5	0	0,5
	0,5	0,5	1

b.

Y \ X	0	2	
0	0,25	0,25	0,5
1	0,25	0,25	0,5
	0,5	0,5	1

Glücksrad a: Es gilt z.B. P(X=0|Y=1) = 1 ≠ P(X=0) = 0,5, d.h. X und und Y sind abhängig.
Glücksrad b: Es gilt P(X=0|Y=1) = 0,5 = P(X=0) = 0,5
P(X=2|Y=1) = 0,5 = P(X=2) usw., d.h. X und Y sind unabhängig

Bei den letzten beiden Beispielen haben wir den Begriff Unabhängigkeit auf Zufallsgrößen übertragen. Allgemein gilt sinngemäß die

Definition:

Sei P ein Wahrscheinlichkeitsmaß auf Ω und seien X, Y Zufallsgrößen auf Ω. X und Y heißen (stochastisch) unabhängig, wenn gilt
$P(X = a \mid Y = b) = P(X = a) \quad \text{für alle } a, b \in \mathbb{R}$

Unabhängigkeit lässt sich auch charakterisieren durch

$$P(X = a \wedge Y = b) = P(X = a) \cdot P(Y = b) \quad \text{für alle } a, b \in \mathbb{R},$$

d. h. die folgende „Mehrfeldertafel“ der gemeinsamen Verteilung muss eine Multiplikationstafel sein. (vgl. Vierfeldertafel)

X Y	…	a	…	
. . .		. . .		
B	…	P(X=a ∧ Y=b)	…	$P_Y(b)$
. . .		. . .		
		$P_X(a)$		1

Ausblick: Der Korrelationskoeffizient misst quantitativ, wie stark X, Y voneinander abhängen.

2.1.3.3 Erwartungswert und Varianz für verknüpfte Zufallsgrößen

Für eine einzelne Zufallsgröße folgt jetzt ganz einfach

Satz 1:
$a)\ Var(X) = E(X^2) - E(X)^2$
$b)\ Var(X) = E\left[\left(X - E(X)\right)^2\right]$

Beweis:

$$a)\; Var(X) = \sum (x_i - \mu)^2 \; P(X = x_i) =$$
$$= \sum x_i^2 P(X = x_i) - 2\mu \underbrace{\sum x_i \, P(X = x_i)}_{=\mu} + \mu^2 \underbrace{\sum P(X = x_i)}_{=1}$$
$$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - E(X)^2$$

$$b)\; Wegen \;\; V(X) = \sum (x_i - \mu)^2 \; P(X = x_i) \;\; gilt \; V(X) = E(Y) \;\; für \; Y = (X - \mu)^2$$

Aufgrund der inhaltlichen Bedeutung des Erwartungswertes als theoretischer Mittelwert und der Varianz als theoretisches quadratisches Streuungsmaß sind die ersten drei Eigenschaften in folgendem Satz klar:

Satz 2:

Seien X, Y Zufallsgrößen auf Ω und seien a, b $\in \mathbb{R}$. Dann gilt (sofern die jeweiligen Werte existieren):
(1) E(aX + b) = aE(X) + b
(2) E(X +Y) = E(X) + E(Y)
(3) V(aX + b) = a²V(X)
(4) V(X + Y) = V(X) + V(Y) + 2(E(X·Y) – E(X)·E(Y))

Sind X, Y unabhängig, so gilt überdies :
(5) E(X·Y) = E(X)·E(Y)
(6) V(X + Y) = V(X) + V(Y)

Beweis :

Vorbemerkung : Wegen $\mathrm{P_X(c)} = \sum\limits_{\substack{\omega\in\Omega \\ X(\omega)=c}} \mathrm{P}(\omega)$ für $c \in X(\Omega)$ können wir

$$\mathrm{E(X)} = \sum_{c\in X(\Omega)} c \cdot \mathrm{P_X(c)}$$

auch schreiben als

$$\mathrm{E(X)} = \sum_{\omega\in\Omega} X(\omega) P(\omega)$$

(und analog für V(X)).

Ad (1) :

$$E(aX + b) = \sum_{\omega\in\Omega} (a \cdot X(\omega) + b) \cdot P(\omega) = a \cdot \sum_{\omega\in\Omega} X(\omega) \cdot P(\omega) + b \cdot \sum_{\omega\in\Omega} P(\omega) = a \cdot E(X) + b \cdot 1.$$

Ad (2) :

$$E(X+Y)=\sum_{\omega\in\Omega}\big(X(\omega)+Y(\omega)\big)\cdot P(\omega)=\sum_{\omega\in\Omega}X(\omega)\cdot P(\omega)+\sum_{\omega\in\Omega}Y(\omega)\cdot P(\omega)=E(X)+E(Y).$$

Ad (3) :

$$V(aX+b)\overset{Satz1b}{=}E\left((aX+b-E(aX+b))^2\right)\underset{(1)}{=}E\left((aX+b-aE(X)-b)^2\right)=E\left(a^2(X-E(X))^2\right)=$$

$$\underset{(1)}{=}a^2E\left((X-E(X))^2\right)\underset{Satz1b}{=}a^2V(X).$$

Ad (4) :

$$V(X+Y)\underset{Satz1a}{=}E\left((X+Y)^2\right)-\left(E(X+Y)\right)^2\underset{(2)}{=}E\left(X^2+2XY+Y^2\right)-\left(E(X)+E(Y)\right)^2=$$

$$\underset{(2)}{=}E(X^2)+2E(XY)+E(Y^2)-E(X))^2-2E(X)E(Y)-(E(Y))^2=$$

$$\underset{Satz1a}{=}V(X)+V(Y)+2(E(XY)-E(X)E(Y)).$$

Ad (5) :

$$E(XY)=\sum_{c\in\mathbb{R}}c\cdot P_{XY}(c)=\sum_{a,b\in\mathbb{R}}a\cdot b\ P(X=a\wedge Y=b)\underset{da\ unabh.}{=}\sum_{a,b\in\mathbb{R}}a\cdot b\ P(X=a)\cdot P(Y=b)$$

$$=\sum_{a\in\mathbb{R}}a\ P(X=a)\sum_{b\in\mathbb{R}}b\ P(Y=b)=E(X)\,E(Y)$$

Ad (6) :

Klar nach (4) und (5)

Bem.: Wie üblich ist Ω endlich (oder höchstens abzählbar unendlich).

2.1.3.4 Standardisierte Zufallsgrößen

Definition: Ist X Zufallsgröße mit μ = E(X), σ² = Var(X) so heißt $T:=\frac{X-\mu}{\sigma}$ die zu X gehörige standardisierte Zufallsgröße (normierte Zufallsgröße)

Satz: E(T) = 0, Var(T) = 1

Beweis: Nachrechnen!

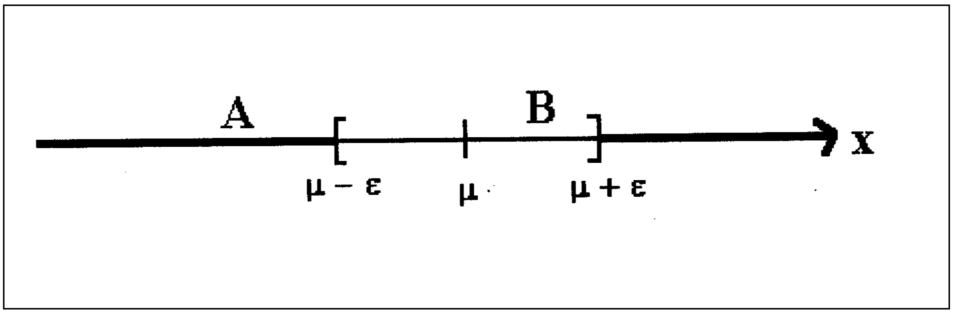
$$T = \frac{X - \mu}{\sigma} = \frac{1}{\sigma} X - \frac{\mu}{\sigma} \Rightarrow E(T) = \frac{1}{\sigma} E(X) - \frac{\mu}{\sigma} = 0$$

$$Var(T) = \left(\frac{1}{\sigma}\right)^2 Var(X) = 1$$

2.1.3.5 Die Ungleichung von Tschebyscheff und σ-Umgebungen

Industriell gefertigte Produkte müssen gewisse Normbedingungen erfüllen, z. B. Metallbohrer mit Querschnitt 4,8 mm² und maximalen Abweichungen von 0,05 mm². Interessant ist die Kenntnis der Wahrscheinlichkeit, dass die Bohrer im Toleranzintervall liegen. Bei Kenntnis der zugrundeliegenden Zufallsgröße X (genauer reichen E(X) und Var(X)) lässt sich hierfür eine Abschätzung angeben, nämlich die Ungleichung von Tscheyscheff.

Es seien also X, μ, σ^2 bekannt (ideale Annahme!)
Gesucht ist $P(|X - \mu| > \varepsilon)$ *für* $\varepsilon \in \mathbb{R}^+$.



$$\sigma^2 = \sum_{x_i \in A} (x_i - \mu)^2 P(x_i) + \sum_{x_i \in B} (x_i - \mu)^2 P(x_i) \geq \sum_{x_i \in A} (x_i - \mu)^2 P(x_i) > \varepsilon^2 \sum_{x_i \in A} P(x_i)$$

$$\Leftrightarrow \sum_{x_i \in A} P(x_i) < \frac{\sigma^2}{\varepsilon^2}, \text{ also}$$

Satz: Ungleichung von Tschebyscheff
P sei ein Wahrscheinlichkeitsmaß und X eine Zufallsgröße auf Ω.

Dann gilt

$$a)\ P(|X - \mu| > \varepsilon) < \frac{\sigma^2}{\varepsilon^2} \quad \text{für alle } \varepsilon \in \mathbb{R}^+$$

$$b)\ P(|X - \mu| > k \cdot \sigma) < \frac{1}{k^2} \quad \text{für alle } k \in \mathbb{N}$$

Beweis für b): setze $\varepsilon = k\sigma$!

Die so gelieferten Vertrauensintervalle sind <u>nicht</u> besonders scharf (z. B. k = 1 !!).
Für <u>jede</u> Zufallsgröße X gilt

im Intervall	[μ-2σ; μ+2σ] liegen im Mittel mindestens	75 %
	[μ-3σ; μ+3σ]	89 %
	[μ-4σ, μ+4σ]	94%

aller Werte. Weiß man mehr über X, so lassen sich schärfere Schranken angeben (vgl. S. 104).

<u>Beispiel:</u> Konservendosen sollen 1000g Fleisch enthalten, die maximal erlaubte Abweichung ist 30g. Bei Prüfung von 200 Dosen ergaben sich $\bar{x}$ = 1000g und s² = 100.

Deutung: Setze

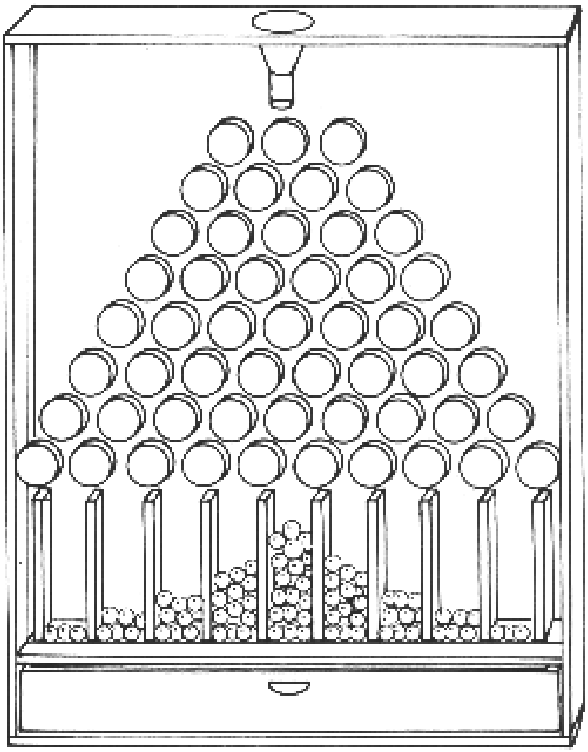
$$\mu = \bar{x},\ Var(x) = \sigma^2 = s^2.$$

$$P(|X - \mu) > 30| < \frac{10^2}{30^2} = 11{,}1\%$$

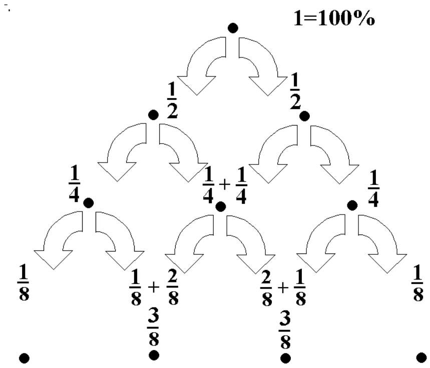
2.2 Binomialverteilung

2.2.1 Galtonbrett und binomialverteilte Zufallsgrößen

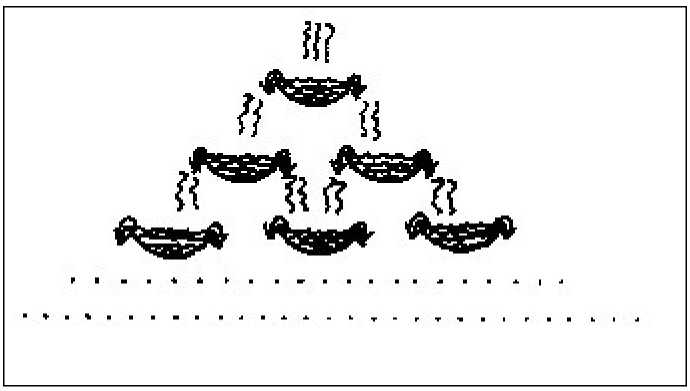
Auf den englischen Naturforscher Francis Galton (1822-1911) geht ein schönes Stochastisches Experiment, das Galton-Brett zurück (vgl. Abb.). Galton interessierte sich für Merkmalsprägungen und ihre Vererbung und war einer der Väter der Daktyloskopie.



Von oben fallen Kugeln in das Galtonbrett und entscheiden sich bei jedem Plättchen zufällig, ob sie nach rechts oder links fallen. Lässt man viele Kugeln oben in das Brett fallen, so entsteht unten in den Auffangkästen eine glockenförmige Verteilung. Die Wahrscheinlichkeiten, mit der die Kugeln in die unteren Fächer fallen, lassen sich leicht berechnen, wenn man bei jeder Verzweigung eine Chance von 1:1 für rechts oder links annimmt (vgl. auch Folie).



Ein anderes, analoges Beispiel ist ein „römischer Brunnen“, bei dem ähnlich wie beim Galtonbrett, Überlaufgefäße übereinander angeordnet sind. Der Überlauf kann so eingerichtet werden, dass links und rechts jeweils ein unterschiedlicher Anteil des Wassers überläuft (dies entspricht einem schief aufgestellten Galtonbrett).



Schließlich kann auch das 30-malige Werfen eines Würfels und Achten auf die 6 ähnlich gesehen werden: von links nach rechts entsprechen die „Fächer“ dem Auftreten von 0 bis 30 mal einer 6:

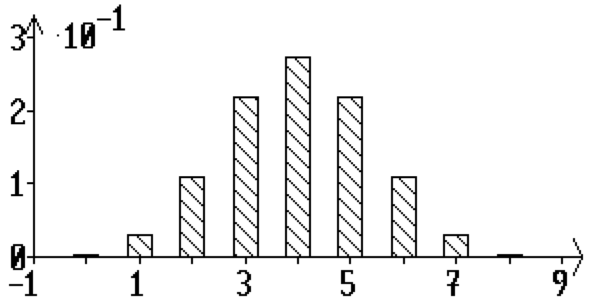
In jedem Fall haben wir ein mehrstufiges Zufallsexperiment bestehend aus n gleichartigen Bernoulli-Experimenten, d. h. Experimenten mit genau 2 Ausgängen $\Omega = \{T, N\}$ oder besser gleich Zufallsgrößen $X = \Omega = \{0,1\}$ mit $P(X=1) = p$, $P(X=0) = q = 1-p$. Eine Bernoulli-Kette der Länge n mit Parameter p ist das n-fache Durchführen eines Bernoulli-Experiments (die Teil-Experimente sind unabhängig und gleichartig).

Am Baum des Würfelbeispiels macht man sich leicht den folgenden Satz klar:

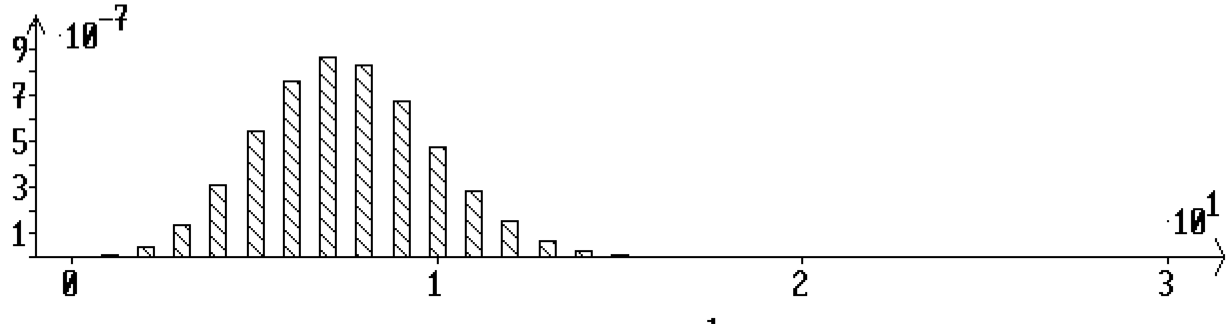
Satz: Bei einer n-stufigen Bernoulli-Kette mit Parameter p ist die Wahrscheinlichkeit für das Ereignis „genau k mal 1“ die Zahl $B_{n,p}(k) = B(n,p,k) = \binom{n}{k} p^k (1-p)^{n-k}$

Der Beweis ist die Anwendung der Pfadregeln: ein Pfad mit genau k Treffern hat nach der Pfad-Multiplikationsregel die Wahrscheinlichkeit $p^k (1-p)^{n-k}$. Zum gewünschten Ergebnis „genau k mal 1“ führen alle Pfade mit genau k mal einer 1 und (n-k) mal einer 0; dies entspricht der kombinatorischen Grundaufgabe, k Kugeln auf n Plätze zu legen, was genau $\binom{n}{k}$ mal geht.

Die entsprechende Wahrscheinlichkeitsverteilungen beim Galton-Brett und beim „6-er-Würfeln“ zeigt die folgende Graphik:



Binomialverteilung beim Galtonbrett (n = 8, p = 0.5)



Binomialverteilung beim "6-er Würfeln" (n = 30, p = $\frac{1}{6}$)

Der Name „Binomialverteilung“ erklärt sich aus den auftretenden Binomialkoeffizienten. Die Verallgemeinerung der 3 Beispiele führt zur

Definition der Binomialverteilung

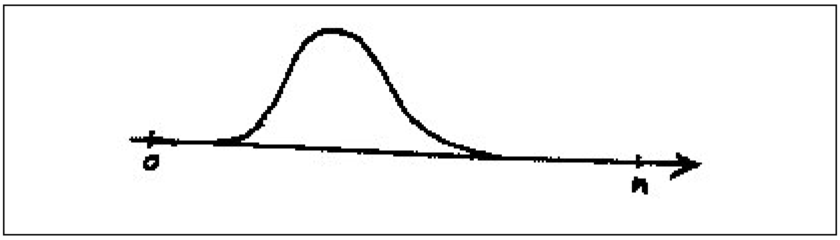
Sei P ein Wahrscheinlichkeitsmaß auf Ω und sei X eine Zufallsgröße auf Ω mit Werten 0,1,...,n $n \in \mathbb{N}$). Wenn gilt

$$P_x(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \text{ für } k = 0,1,...,n$$

mit einem gewissen $p \in (0;1)$, so heißt X <u>binomialverteilt</u> mit den Parametern n und p (kurz: B(n,p)-verteilt), und die Verteilung P_x heißt eine <u>Binomialverteilung.</u>

Die Wahrscheinlichkeiten werden meistens mit B(n,p,k) oder $B_{n,p}(k)$ bezeichnet.

Man kann leicht beweisen: Die Werte $B_{n,p}(k)$ steigen von k = 0 ausgehend monoton, bis sie ein Maximum bei $k \approx np$ erreichen. Dann fallen sie monoton bis k = n. Deshalb haben die Diagramme aller Binomialverteilungen einen charakteristischen, „glockenförmigen“ Verlauf. Sehr schön kommt dies beim Riesen-Galtonbrett im Science-Museum in Seattle heraus (Folie!).



Viele reale Experimente lassen sich gut mit Binomialverteilungen modellieren. Beispielsweise seien in einer Urne schwarze und weiße Kugeln. Ist die Wahrscheinlichkeit p(S) für das Ziehen einer schwarzen Kugel bekannt, so ist es eine Aufgabe der Wahrscheinlichkeitsrechnung, die Wahrscheinlichkeit für das k-malige Ziehen einer schwarzen Kugel bei n-maligem Ziehen mit Zurücklegen zu berechnen. Ist p(S) dagegen nicht bekannt, so ist es eine Aufgabe der Statistik nach Ziehen einer Stichprobe mit Zurücklegen auf diese Wahrscheinlichkeit zu schließen. Im ersten Fall könnte es sich um Schrauben handeln, die mit Wahrscheinlichkeit p den Normmaßen entsprechen, und es wird vorhergesagt, wie viele defekte Schrauben im Mittel in einer 1000er Packung sind. Im zweiten Fall kann aufgrund eines Tests an n Patienten auf die Wirksamkeit eines neuen Medikaments geschlossen werden.

Durch die fast ausschließliche Behandlung der Binomialverteilung in der Schule vergisst man gerne, dass es sehr viele verschiedene Verteilungen als stochastisches Modell gibt. Der Käufer der Schrauben wird bei der Wareneingangskontrolle Stichproben ziehen, um die behauptete Fehlerfreiheitsquote zu überprüfen. Dieses Stichprobenziehen wird ein Ziehen ohne Zurücklegen sein, das durch eine hypergeometrische Verteilung (vgl. 2.3.1) beschrieben wird. Allerdings weicht diese Verteilung für große n kaum von der Binomialverteilung ab.

Die Berechnung der Koeffizienten der Binomialverteilung ist schon bei kleinem n mühevoll. Auch beim Schreiben eines kleinen Computerprogramms muß man durchaus trickreich rechnen, um die Binomialkoeffizienten zu erhalten. Man berechnet die $B_{n,p}(k)$ am besten rekursiv.

$$B_{n,p}(k+1) = \frac{n-k}{k+1} \cdot \frac{p}{1-p} \cdot B_{n,p}(k) \text{ für } k = 0,\dots,n\text{-}1.$$

Die in Formelsammlungen tabellierten Werte haben den Nachteil, dass nur ganz bestimmte n- und p-Werte vorhanden sind, reale Probleme also kaum behandelbar sind.

2.2.2 Erwartungswert und Varianz binomialverteilter Zufallsgrößen

Inhaltlich ist klar, dass E(x) = np gilt. Dies lässt sich auch nachrechnen:

$$E(X) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = n \cdot p \cdot \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \cdot \underbrace{\sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i}}_{=(p+(1-p))^{n-1}=1} = np$$

$$\boxed{\text{denn } k \cdot \binom{n}{k} = n \cdot \binom{n-1}{k-1}}$$

Diese Rechnerei ist unschön. Einfachere Berechnung über Zerlegung $X = \sum_{i=1}^{n} X_i$, wobei X_i Anzahl der Treffer beim i^{ten} Versuch; klar $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$.

Wegen $E(X_i) = 0 \cdot (1 - p) + 1 \cdot p = p$ folgt aus 2.1.3.3, Satz 2(2), sofort

$$E(X) = \sum_{i=1}^{n} E(X_i) = n \cdot p$$

da die Zufallsgrößen X_i unabhängig voneinander sind.
Klar: $p(X_i = a_1\ X_j = b) = p(X_i = a) \cdot p(X_j = b)$, folgt aus 2.1.3.3, Satz 2(6)

$$Var(X) = \sum_{i=1}^{n} Var(X_i).$$

$$Var(X_i) = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p(1-p),$$

also zusammen

$$Var(X) = np(1\text{-}p)$$

Dies könnte man auch mühsam direkt aus der Definition nachrechnen:

$$\operatorname{Var}(X)=\sum_{k=1}^{n}(k-np)^2\binom{n}{k}p^k(1-p)^{n-k}$$

Satz: Für eine B(n,p) – verteilte Zufallsgröße X gilt E(X) = np und Var (X) = np(1-p)

2.2.3 Beispiele

1. Lehrer Schlau hat je einen gelben, silbernen, weißen und roten Würfel. Ein Schüler wählt einen Würfel. Schlau nimmt einen anderen. Schlau setzt 1 DM. Dann wird 11 mal gewürfelt. Beim Einzelspiel gewinnt die höhere Augenzahl, insgesamt gewinnt der mit mehr gewonnenen Einzelspielen. Was riskiert Lehrer Schlau?

Bem.: Die Schüler merken schnell, dass die Ziffern nicht 1,...,6, sondern anders verteilt sind:

Gelb Silber Weiß Rot

Lösung

Analyse der möglichen Paarungen zeigt schnell, dass zu jeder Farbe eine andere gewählt werden kann mit Gewinnchance $p=\frac{2}{3}$:

$$\to \text{Gelb} \to \text{Rot} \to \text{Weiß} \to \text{Silber} \to \text{Gelb} \to \ldots$$

Einzelspiel: Bernoulli-Experiment

$$p(\text{Schlau gewinnt}) = \frac{2}{3}$$

Gesamtspiel: Bernoulli-Kette mit $p = \frac{2}{3}$, $n = 11$

$$P(\text{Schlau gewinnt}) = \sum_{k=6}^{11} B(11,\tfrac{2}{3},k) = \sum_{k=6}^{11}\binom{11}{k}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{11-k} \approx 0{,}8779 \approx 88\%$$

2. Eine Fabrik gibt den Ausschussanteil bei der Produktion elektrischer Sicherungen mit 1% an. Der Käufer einer größeren Stückzahl entnimmt eine Stichprobe von 100 Stück und entscheidet nach folgendem Stichprobenplan:
   Sind unter den 100 Prüfstücken mehr als zwei defekt, wird die Lieferung zurückgewiesen.
   a) Wie groß ist die Wahrscheinlichkeit, dass der Kaufmann die Lieferung zurückweist, obwohl der Ausschussanteil der Angabe entspricht?
   b) Wie groß ist die Wahrscheinlichkeit, dass der Kaufmann die Lieferung annimmt, wenn der Ausschussanteil in Wirklichkeit sogar 5% ist?

Lösung

a) Aufgrund der großen Gesamtzahl kann bei einer Entnahme von 100 Stück eine Bernoulli-Kette mit n=100, p=0,01 unterstellt werden, d. h. der Ausschussanteil möge korrekt sein. Zurückgewiesen wird bei i defekten Sicherungen mit $3 \le i \le 100$, d. h.

$$\begin{aligned} P(\text{Zurückweisung}) &= \sum_{i=3}^{100} B(100, 0{,}01, i) = 1 - \sum_{i=0}^{2} B(100, 0{,}01, i) \\ &= 1 - 0{,}99^{100} - 100 \cdot 0{,}01 \cdot 0{,}99^{99} - \frac{100 \cdot 99}{2} \cdot 0{,}01^2 \cdot 0{,}99^{98} \\ &\approx 0{,}0794 = 7{,}94\% \end{aligned}$$

a) Jetzt ist in Wirklichkeit p = 0,05, q = 0,95. Annahme bei i Fehlern mit $0 \le i \le 2$, also

$$\begin{aligned} P(\text{Annahme}) &= \sum_{i=0}^{2} B(100, 0{,}05, i) \\ &= 0{,}95^{100} + 100 \cdot 0{,}05 \cdot 0{,}95^{99} + \frac{100 \cdot 99}{2} \cdot 0{,}05^2 \cdot 0{,}95^{98} \\ &\approx 0{,}1183 = 11{,}83\% \end{aligned}$$

3. Bei der Prüfung wird ein sogenannter „Multiple-choice-Test“ mit Fragen angewendet. Es werden fünf Fragen gestellt. Zu jeder der Fragen sind in zufälliger Anordnung eine richtige und zwei falsche Antworten angegeben. Die Prüfung gilt als bestanden, wenn bei mindestens vier Fragen die richtige Antwort angekreuzt ist. Ein unvorbereiteter Prüfling wählt seine Antworten rein zufällig aus.
   a) Mit welcher Wahrscheinlichkeit besteht er die Prüfung?
   b) Wie groß sind Erwartungswert und Standardabweichung der Anzahl X seiner richtigen Antworten?

Lösung

Bernoulli-Kette mit n = 5, p = $\frac{1}{3}$ Wahrscheinlichkeit für „richtige Antwort“

a) Bestanden, falls 4 oder 5 richtige Kreuze,

$$p(\text{Bestanden}) = p(X \ge 4) = B(5, \tfrac{1}{3}, 4) + B(5, \tfrac{1}{3}, 5)$$

$$= \binom{5}{4} \cdot \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right) + \binom{5}{5} \cdot \left(\frac{1}{3}\right)^5 \approx 0{,}0452675 \approx 4{,}5\%$$

b) E(X) = np = 5/3; Var(X) = npq = $\frac{10}{9}, \sigma = \sqrt{\mathrm{Var}X} = \frac{\sqrt{10}}{3} \approx 1{,}05$

2.2.4 Das Bernouillesche Gesetz der großen Zahl

Es geht hierbei um die Präzisierung des empirischen Gesetzes der großen Zahl: $h_n \approx p(E)$. Wir betrachten das n-malige Durchführen eines Experiments als Bernouilli-Experiment mit Eins = „E ist eingetreten" mit p(E) = p. Die zugehörige Zufallsgröße X hat die Wertmenge {0,1,2,...,n}. Erwartungswert und Varianz ergeben sich zu

$$E(X) = n \cdot p\ ,\ \mathrm{Var}(X) = n \cdot p \cdot (1-p).$$

Die neue Zufallsgröße $\overline{X} = \frac{1}{n} X$ beschreibt die relative Häufigkeit des Eintretens von E bei n Versuchen mit der Wertmenge

$$W = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\right\}$$

Die entsprechenden Größen von $\overline{X}$ ergeben sich zu

$$E\left(\overline{X}\right) = \frac{1}{n} E\left(X\right) = p, \quad \sigma^2 = \mathrm{Var}\left(\overline{X}\right) = \frac{1}{n^2} \mathrm{Var}\left(X\right) = \frac{p(1-p)}{n}.$$

Der Erwartungswert der relativen Häufigkeit ist also p, die Abweichung davon ist umso geringer, je größer n ist. Setzen wir diese Ergebnisse in die Ungleichung von Tschebyschew

$$p(\left|X - E(X)\right| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

ein, so folgt

$$p(|\,\overline{X} - p\,|) \geq \varepsilon \leq \frac{p(1-p)}{n\varepsilon^2}$$

oder gleichbedeutend

$$p(|\,\overline{X} - p\,| < \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2} \geq 1 - \frac{1}{4n\varepsilon^2}\ ,$$

wobei wir bei der letzten Ungleichung verwendet haben, dass der Scheitel der Parabel y = x(1-x) der Punkt $\left(\frac{1}{4}\middle|\frac{1}{2}\right)$ ist.

Damit gilt der

Satz (Bernouillisches Gesetz der großen Zahl)

Sei P ein Wahrscheinlichkeitsmaß auf Ω. X sei B(n,p)-verteilt, und $H_n := \frac{X}{n}$.
Dann gilt für jedes $\varepsilon > 0$.

$$\lim_{n\to\infty} P(|H_n - p| < \varepsilon) = 1$$

In Worten: Welche positive Schranke ε man auch vorgibt, stets kommen die Wahrscheinlichkeiten, dass die relative Trefferhäufigkeiten in einer Bernoulli-Kette der Länge n von der Trefferwahrscheinlichkeit p um höchstens ε abweicht, für genügend große n beliebig dicht an 1 heran. Diese Tatsache heißt das Bernoullische Gesetz der großen Zahlen.

Dieses Gesetz stellt offenbar das Modell-Analogon zum empirischen Gesetz der großen Zahlen dar (vgl. S. 31). Das Bernoullische Gesetz ist nun aber tatsächlich eine Grenzwertaussage im Sinne der Analysis.

Die Schweiz gab übrigens zum ICM in Zürich im Jahr 1954 eine Sonderbriefmarke zu diesem Thema heraus:

Beispiel 1
Eine angebliche Laplace-Münze zeigt bei 1000-maligem Werfen 600 mal Kopf. Ist das noch zu akzeptieren?

Lösung
Es gilt $h = \frac{600}{1000} = 0{,}6$, $p = 0{,}5$, $n = 1000$

$$p(|h - p| < 0{,}1) \geq 1 - \frac{1}{4000 \cdot 0{,}01} = 97{,}5\%,$$

also handelt es sich vermutlich um keine Laplace-Münze.

Beachte: Die direkte Berechnung

$$p(400 < X < 600) = \sum_{k=401}^{599} \binom{1000}{k} \frac{1}{2}^{k} \cdot \frac{1}{2}^{1000-k}$$

wäre nicht so grob, aber viel komplizierter!

Beispiel 2
Eine Partei lässt 3 Wochen vor der Wahl eine Umfrage machen, um auf 2% genau ihren derzeitigen Stimmenanteil zu ermitteln. Das beauftragte Meinungsforschungsinstitut befragt 5000 Personen.

Lösung
Gefordert ist $|h-p| \leq 0{,}02$, wobei das Ereignis E = „ein Bürger wählt die Partei" betrachtet wird. Die Sicherheit der Prognose beträgt

$$P(|h-p| \leq 0{,}02) > 1 - \frac{1}{4 \cdot 5000 \cdot 0{,}02^2} = 87{,}5\%$$

Beispiel 3
Während des Auszählens der Wahlergebnisse werden nach Zählen von n Zetteln 11% für Partei A festgestellt. Die Hochrechnung behauptet dann „11% ± 1%" mit 99,9%-iger Sicherheit. Wie groß muß n sein?

Lösung
Behauptet wird

$$P(|11\%-p| \leq 0{,}01) \geq 0{,}999 \text{ , also}$$

$$1 - \frac{1}{4 \cdot n \cdot 0{,}01^2} \geq 0{,}999 \Leftrightarrow 4n \cdot 0{,}01^2 \geq \frac{1}{0{,}001} = 10^3 \Leftrightarrow 4n \geq 10^7 \Leftrightarrow n \geq 2500000$$

2.3 Andere Verteilungen

2.3.1 Hypergeometrische Verteilung

Die Binomialverteilung lässt sich an einem Urnenmodell mit M weißen und N-M schwarzen Kugeln durch Ziehen mit Zurücklegen erklären. Die Wahrscheinlichkeit, eine weiße Kugel zu ziehen, ist jeweils $p = \frac{M}{N}$. Jede Versuchswiederholung ist unabhängig von dem vorherigen Versuch.

Bei Stichproben von Bauteilen auf Fehler werden die gezogenen Bauteile nicht zurückgelegt, bei Befragungen, die befragten Personen nicht ein zweites Mal befragt. Hier ist das Urnenmodell mit Ziehen ohne Zurücklegen angemessen.

Wir haben also im Modell wieder M weiße und N-M schwarze Kugeln. Es wird eine Stichprobe von n (≤ N) Kugeln entnommen. Die Zufallsgröße X beschreibt die Anzahl k der gezogenen weißen Kugeln. Es gibt bekanntlich

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

Definition: Eine Zufallsgröße X heißt hypergeometrisch verteilt mit den Parametern n, M und N ($n, M, N \in \mathbb{N}$, $1 \le M < N$), wenn gilt

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

für $0 \le k \le \min(n,M)$. Man sagt dann auch, X sei H(N,M,n)-verteilt.

Satz: Für eine H(N,M,n)-verteilte Zufallsgröße X gilt mit $p=\frac{M}{N}$

$$E(X) = n \cdot p$$

$$V(X) = n \cdot p(1-p) \cdot \frac{N-n}{N-1}$$

Beweis

a) Die weißen Kugeln seien nummeriert von 1 bis M. Die Zufallsvariable X_i beschreibt, ob die i-te gezogene Kugel der Stichprobe eine weiße ist.

$$X_i=\begin{cases}1, \text{falls } i-\text{te Kugel in Stichprobe weiß}\\ 0, \text{ falls nicht}\end{cases}$$

Es gilt: $P(X_i=1)=\frac{M}{N}$, da jede der N Kugeln die gleiche Chance hat, als i-te gezogen werden (dies ist ein qualitatives Argument; die Zufallsvariablen sind ja voneinander abhängig. Dies lässt sich jedoch auch formal nachrechnen, vgl. für den einfachsten Fall n=2, Beispiel 4 in 1.6.4, S. 51).

Wegen $E(X_i)=1\cdot\frac{M}{N}+0\cdot\frac{N-M}{N}=p$ und $X=X_1+X_2+...+X_n$

folgt wie behauptet aus Satz 2(2) in 2.1.3.3 für den Erwartungswert E(X) = np.

Die Varianz berechnen wir mit Hilfe des Verschiebungssatzes Satz 1 (a) in 2.1.3.3:

$$V(X) = E(X^2) - E(X)^2.$$

Es gilt

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

$$= n\frac{M}{N} \sum_{k=1}^{n} k$$

Dabei haben wir verwendet

$$k \cdot \binom{M}{k} = k \cdot \frac{M(M-1)\cdot \ldots \cdot (M-k+1)}{k!} = M\frac{(M-1)\cdot \ldots \cdot (M-k+1)}{(k-1)!} = M\binom{M-1}{k-1}$$

$$\binom{N}{n} = \frac{N(n-1)\cdot \ldots \cdot (N-n+1)}{n!} = \frac{N}{n}\,\frac{(N-1)\ldots(N-n+1)}{(n-1)!} = \frac{N}{n}\binom{N-1}{n-1}$$

Mit k = (k-1) + 1 gilt weiter nach einigen einfachen Umformungen

$$E(X^2) = n\frac{M}{N}\sum_{k=1}^{n}(k-1)\frac{\binom{M-1}{k-1}\binom{(N-1)-(M-1)}{(n-1)-(k-1)}}{\binom{N-1}{n-1}} + n\frac{M}{N}\sum_{k=1}^{n}\frac{\binom{M-1}{k-1}\binom{(N-1)-(M-1)}{(n-1)-(k-1)}}{\binom{N-1}{n-1}}$$

Die erste Summe ist der Erwartungswert für n-1, die zweite Summe ist 1, da über alle Wahrscheinlichkeiten für n-1 summiert wird. Zusammen folgt

$$E(X^2) = n\frac{M}{N}\cdot(n-1)\frac{M-1}{N-1} + n\frac{M}{N} = n \cdot p\left((n-1)\frac{M-1}{N-1} + 1\right)$$

Damit folgt für die Varianz

$$V(X) = np\left((n-1)\frac{M-1}{N-1} + 1\right) - (np)^2 = np\left((n-1)\frac{M-1}{N-1} + 1 - n\frac{M}{N}\right)$$

$$= np\frac{(n-1)(M-1)N + (N-1)N - nM(N-1)}{(N-1)N}$$

$$= np\frac{(N-M)(N-n)}{N(N-1)} = np(1-p)\frac{N-n}{N-1} \ ,$$

was zu beweisen war.

Bem.: Bei umfangreicher Grundgesamtheit und kleiner Stichprobe, also n << N, kann man die hypergeometrische Verteilung durch die Binomialverteilung approximieren.

2.3.2 Wartezeitprobleme und Geometrische Verteilung

Typische Situation: Warte auf die erste 6 beim Mensch-ärgere-Dich-nicht, auf den ersten Pasch beim Monopoly ...

Wir hatten schon das Warten auf die erste 6 beim Würfeln behandelt (Beispiel 5 in 1.6.4, Beispiel 2 in 2.1.2). Beschreibt die Zufallsgröße X die Anzahl der nötigen Würfe bis zur ersten 6, so gilt

$$P(X=k) = \left(\frac{5}{6}\right)^{k-1}\frac{1}{6} \qquad (k = 1, 2, ...)$$

Definition: Eine Zufallsgröße X heißt geometrisch verteilt mit Parameter q, wenn für ihre Wahrscheinlichkeitsverteilung gilt

$$P(X=k) = \begin{cases} q^{k-1}(1-q) & \text{für } k \in \mathbb{N} \\ 0 & \text{sonst} \end{cases}$$

Satz: Für eine geometrisch verteilte Zufallsgröße gilt

$$E(X) = \frac{1}{1-q}$$

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

Beweis

$$E(X) = \sum_{k=1}^{\infty} k\, q^{k-1}(1-q) = \frac{1-q}{q}\sum_{k=1}^{\infty} k\, q^k \ = \frac{1-q}{q}\frac{q}{(1-q)^2} = \frac{1}{1-q}$$

V(X) analog mit Verschiebungssatz (vgl. Kütting, Elementare Stochastik, S. 188)

2.3.3 Poisson-Verteilung

Um die mühevolle Berechnung der Binomialkoeffizienten zu erleichtern, gab Poisson (1781-1840) im Jahr 1837 eine Näherungsformel an, die für große n die Binomialkoeffizienten durch zu seiner Zeit sehr gut tabellierte Funktionswerte ersetzte:

Satz 1:

$$\boxed{\lim_{\substack{n\to\infty \\ np=\mu}} B(n,p,k) = \frac{\mu^k}{k!} e^{-\mu}}$$

Beweis:

Unter Verwendung von E(X) = μ = np, also $p = \frac{\mu}{n}$ bei der Binomialverteilung gilt

$$B(n,p,k) = \frac{n(n-1)\cdot \ldots \cdot (n-k+1)}{k!} \cdot \frac{\mu^k}{n^k} \cdot \frac{\left(1-\frac{\mu}{n}\right)^n}{\left(1-\frac{\mu}{n}\right)^k}$$

$$= \frac{1\cdot\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right)}{\left(1-\frac{\mu}{n}\right)^k} \cdot \frac{\mu^k}{k!} \cdot \left(1-\frac{\mu}{n}\right)^n$$

$$= \underbrace{\frac{1}{1-\frac{\mu}{n}} \cdot \frac{1-\frac{1}{n}}{1-\frac{\mu}{n}} \ldots \frac{1-\frac{k-1}{n}}{1-\frac{\mu}{n}}}_{\substack{\to 1\cdot 1\cdot 1 \\ n\to\infty}} \cdot \frac{\mu^k}{k!} \cdot \underbrace{\left(1-\frac{\mu}{n}\right)^n}_{\substack{\to e^{-\mu} \\ n\to\infty}}$$

Satz: Dies beweist die Poisson-Näherung der Binomialverteilung

Für große n und μ = np gilt $B(n,p,k) \approx \frac{\mu^k}{k!} \cdot e^{-\mu}$

Faustregel: Die Näherung ist gut für $\mu << n$ und $|k-\mu| << n$; i. A. sinnvoll für $p \le 0{,}1$ und $n \ge 100$.

Definition: Eine Zufallsgröße X heißt Poisson-verteilt mit dem Parameter μ (≥ 0), wenn gilt

$$P(X=k)=\begin{cases}\dfrac{\mu^k}{k!}e^{-\mu} & \text{für } k\in\mathbb{N}_0\\ 0 & \text{sonst}\end{cases}$$

> Satz 2: X sei Poisson-verteilt mit Parameter μ.
> Dann gilt $E(X)=\mu$
> $V(X)=\mu$

Beweis:

Falls die unendliche Summe existiert, so gilt:

$$E(X)=\sum_{k=0}^{\infty}k\frac{\mu^k}{k!}e^{-\mu}=e^{-\mu}\cdot\sum_{k=0}^{\infty}k\frac{\mu^k}{k!}=$$

$$=e^{-\mu}\mu\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!}=\mu e^{-\mu}\sum_{k=0}^{\infty}\frac{\mu^k}{k!}=e^{-\mu}\cdot\mu\cdot e^{\mu}=\mu$$

$$V(X)=E(X^2)-E(X)^2$$

$$E(X^2)=\sum_{k=0}^{\infty}k^2\frac{\mu^k}{k!}e^{-\mu}=\sum_{k=1}^{\infty}k\frac{\mu^k}{(k-1)!}e^{-\mu}=\mu\left[\underbrace{\sum_{k=1}^{\infty}(k-1)\frac{\mu^{k-1}}{(k-1)!}e^{-\mu}}_{\mu}+\underbrace{\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!}}_{e^{\mu}}e^{-\mu}\right]$$

$$=\mu\left[\mu+e^{-\mu}\cdot e^{\mu}\right]=\mu^2+\mu$$

$V(X)=\mu^2+\mu-\mu^2=\mu$, was zu beweisen war.

Beispiel 1

Lotto: P(„Sechser“) ≈ $0{,}7\cdot 10^{-7}$ (vgl. 1.4.4 c). Pro Woche werden ca. 56 Millionen Tippreihen abgegeben. Wie groß ist die Wahrscheinlichkeit für weniger als 3 Sechser?

Lösung

Binomialverteilung mit $p = 0{,}7 \cdot 10^{-7}$, $n = 56 \cdot 10^{6}$, X = Anzahl der „Sechser“.

Es gilt $\mu = np \approx 4$. Damit gilt für die gesuchte Wahrscheinlichkeit

$$P(X \le 2) = P(X=0) + P(X=1) + P(X=2), \qquad \approx e^{-4} + \frac{4}{1!}e^{-4} + \frac{4^2}{2!}e^{-4} \approx 0{,}26$$

2.3.4 Normal-Verteilung

2.3.4.1 Lokaler Grenzwertsatz von De Moivre

Für $p = \frac{1}{2}$ entwickelte De Moivre (1667-1754) eine Näherungsformel, die Laplace (1740-1827) auf beliebiges $p \in (0;1)$ verallgemeinern konnte:

Satz: (lokale Näherungsformel oder lokaler Grenzwertsatz von De Moivre-Laplace)

$$B(n,p,k) \approx \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2}$$

mit $\mu = np$ und $\sigma = \sqrt{np(1-p)}$

Definition:

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{Gauss-Funktion}$$

$$\phi(x) := \int_{-\infty}^{x} \varphi(t)\,dt \qquad \text{Gaussche Integralfunktion}$$

Damit kann man schreiben

a) $B(n,p,k) \approx \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right)$

b) $P(k_1 \le X \le k_2) \approx \phi\left(\frac{k_2-\mu}{\sigma}\right) - \phi\left(\frac{k_1-\mu}{\sigma}\right)$

c) $F(x) \approx \phi\left(\frac{x-\mu}{\sigma}\right)$

Insbesondere gilt $\phi\left(\frac{x-\mu}{\sigma}\right) \to 1$ für $x \to \infty$.

Faustregel: Die Näherung ergibt sinnvolle Werte für $\sigma^2 = npq > 9$

<u>Bemerkung:</u> φ - und ϕ - Werte sind seit langem tabelliert. Deshalb wurden auf ihnen beruhende stochastische Modelle bevorzugt. Da φ symmetrisch zur y-Achse ist, lassen sich alle interessierenden Flächen auf ϕ(x)-Werte mit x > 0 zurückführen, was die Tabellierung erleichtert.

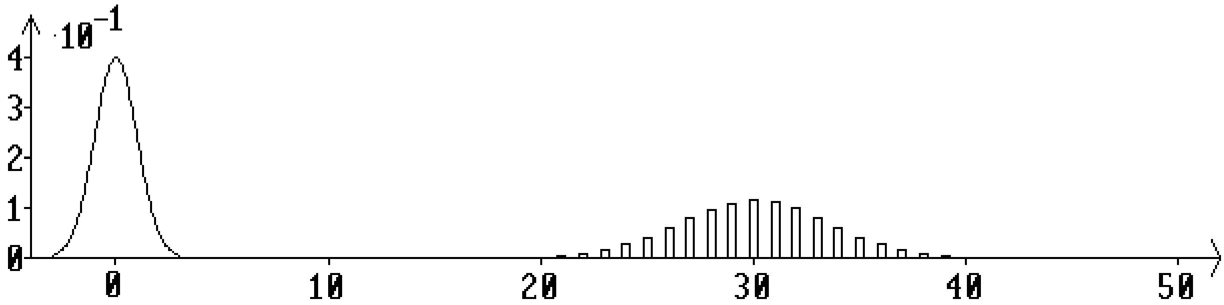
<u>Begründung des lokalen Grenzwertsatzes</u>

Wir können diesen Satz nicht beweisen, aber anschaulich begründen (Visualisierung durch ein Computerprogramm, z. B. folgende Abbildungen mit LAPLACE, einem Stochastik-Programm):
Die Histogramme der B(n,p)-Verteilung nähern sich mit großem n einer glockenförmigen Kurve. Diese soll durch einen Funktionsgraphen angenähert werden. *Eine* mögliche glockenförmige Kurve ist die Gaussche Glockenkurve, der Graph der Gausschen Phi-Funktion φ mit

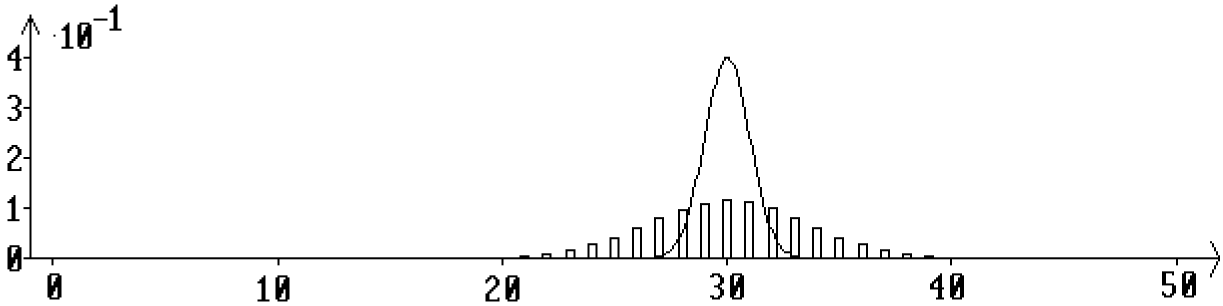
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x^2}$$

Eigenschaften von φ: Symmetrie zur y-Achse, $\int_{-\infty}^{\infty} \varphi(x)dx = 1$

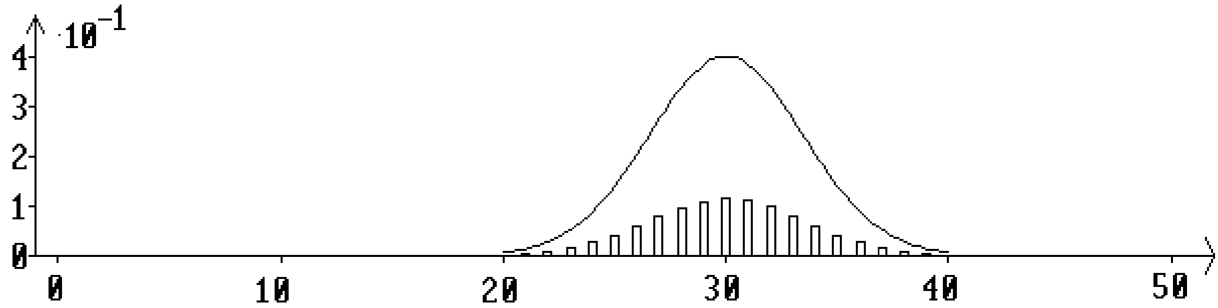
Eine gegebene Binomialverteilung B(n,p) wird in drei Schritten angenähert. Im Beispiel ist n = 50 und p = 0,6 (mit μ = 30 und σ^2 = 12).



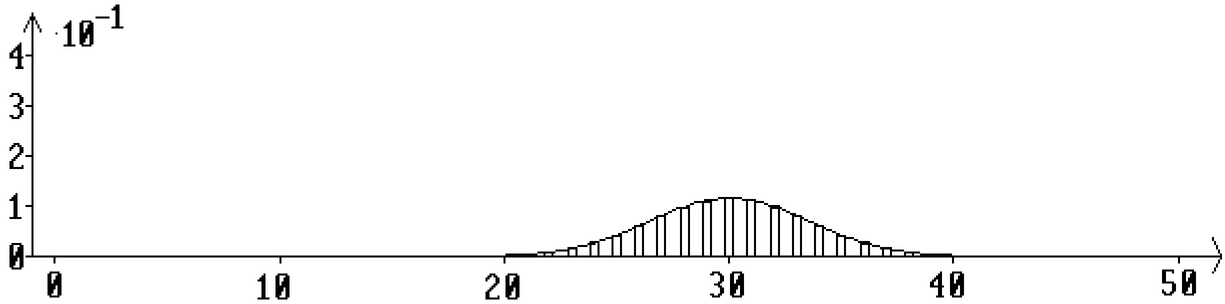
Ausgangssituation φ(x)-Graph und B(n,p)-Verteilung.



1. Schritt: φ(x) → φ(x-A), „Zentrieren“, optimal für A = μ



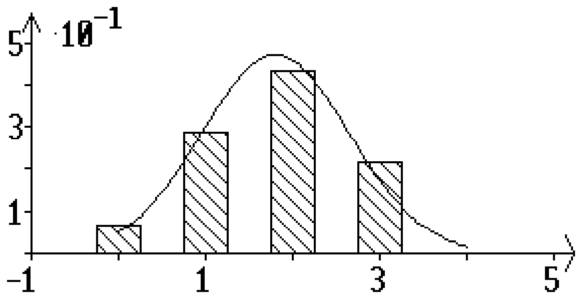
2. Schritt: $\varphi(x) \to \varphi(x\text{-}A)$, „Stauchen längs x-Achse“, optimal für $B = \dfrac{1}{\sigma}$



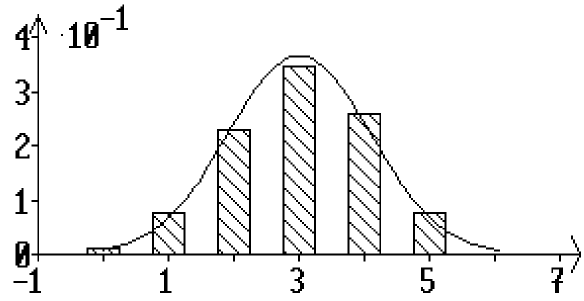
3. Schritt: $\varphi(x) \to C\cdot\varphi(x\text{-}A)$, „Stauchen längs y-Achse“, optimal für $C = \dfrac{1}{\sigma}$

Damit ist $f(x) = \dfrac{1}{\sigma}\varphi\left(\dfrac{x-\mu}{\sigma}\right)$ die optimale Anpassungskurve für die B(n,p)-Verteilung.

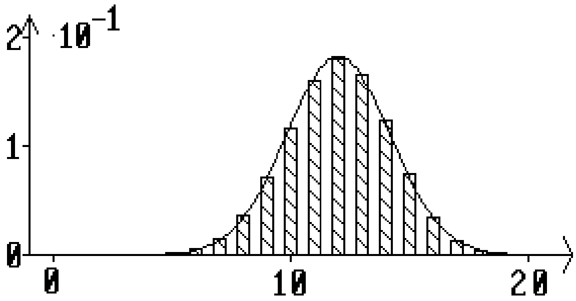
Die folgende Abbildung zeigt diese Anpassungskurve B(n,0.6) und n = 3, 5, 20 und 50:



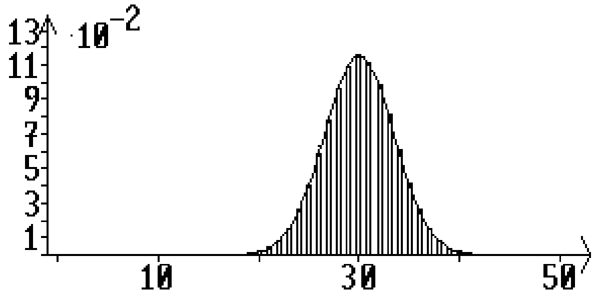
$n = 3$, $p = 0{,}6$, $\sigma^2 = npq = 0{,}72$



$n = 5$, $p = 0{,}6$, $\sigma^2 = npq = 1{,}2$



$n = 20$, $p = 0{,}6$, $\sigma^2 = npq = 4{,}8$



$n = 50$, $p = 0{,}6$, $\sigma^2 = npq = 12$

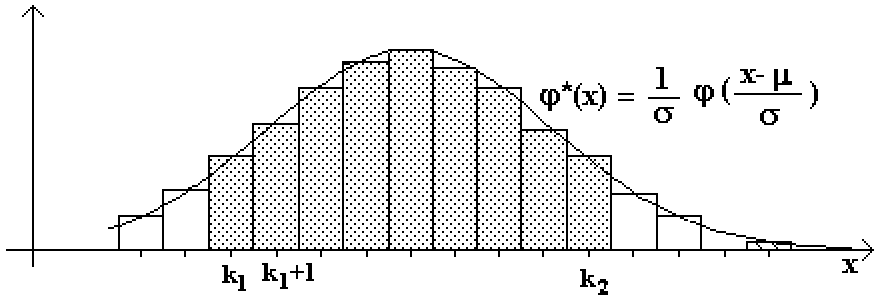
Beweise des lokalen Grenzwertsatzes übersteigen die Möglichkeiten dieser Vorlesung.

Eine bessere Abschätzung der Werte der Verteilungsfunktion zu B(n,p) liefert die
Integrale Näherungsformel

$$P(X \le k) \approx \phi\left(\frac{k + 0{,}5 - \mu}{\sigma}\right)$$

$$P(k_1 \le X \le k_2) \approx \phi\left(\frac{k_2 + 0{,}5 - \mu}{\sigma}\right) - \phi\left(\frac{k_1 - 0{,}5 - \mu}{\sigma}\right)$$

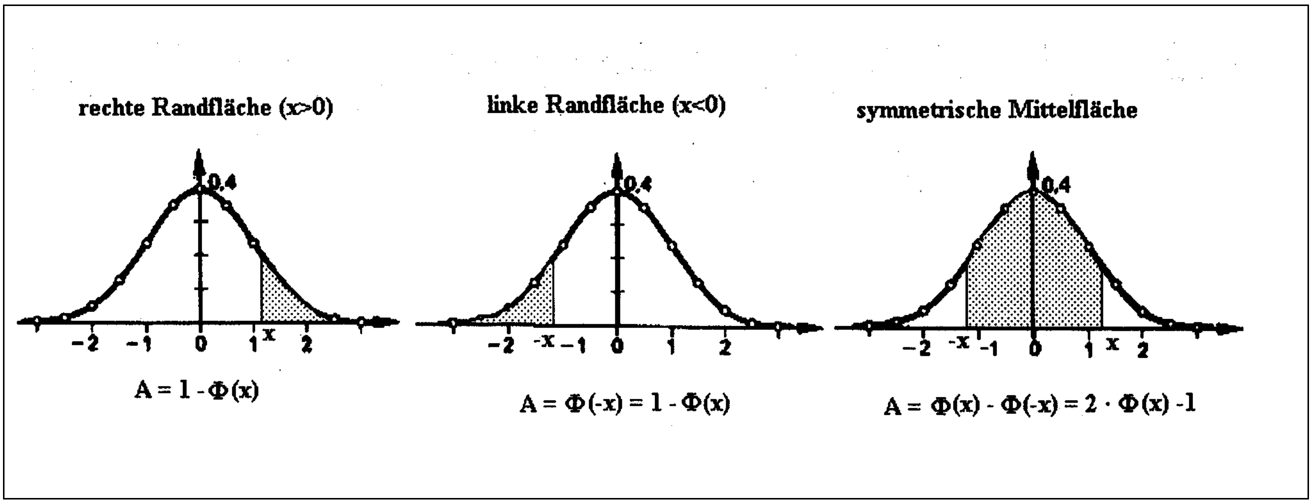
Die Begründung dieser „Stetigkeitskorrektur"-Terme folgt aus der folgenden Abbildung:



$P(k_1 \leq X \leq k_2) \,\hat{=}\,$ Flächeninhalt der Balken von k_1 bis k_2. Diese Fläche soll durch die φ^*-Funktion (mit $\varphi^x(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$ approximiert werden. Also ist von k_1-0,5 bis k_2+0,5 zu integrieren.

<u>Beispiel</u>

Zunächst überlegt man leicht, dass sich alle interessierende Flächen auf $\phi(x)$-Werte mit x>0 zurückführen lassen:



Daher reichte es, die $\phi(x)$-Werte (die sich nicht direkt berechnen lassen) nur für X > 0 zu tabellieren. Anwendung im folgenden Beispiel:

<u>Beispiel</u>
Eine Firma liefert Glühlämpchen in Kartons zu je 1000 Stück. Der Ausschußteil wird von der Firma mit 1% angegeben. Wie groß ist die Wahrscheinlichkeit, dass ein Karton nicht mehr als 10 defekte Lämpchen enthält?

<u>Lösung</u>: Binomialverteilung mit n = 1000, p = 0,01, μ = 10, $\sigma = \sqrt{9,9}$

$$p = \sum_{x=0}^{10} \binom{1000}{x} 0,01^x \, 0,99^{1000-x} = P(0 \le X \le 10) =$$

$$\approx \phi\left(\frac{10-10+0,5}{\sqrt{9,9}}\right) - \phi\left(\frac{0-10-0,5}{\sqrt{9,9}}\right) \approx \phi(0,1589) - \phi(-3,3371)$$
$$= \phi(0,1589) - (1 - \phi(3,3371))$$
$$= \text{(Tabelle!) } 0,5632 - (1 - 0,9996)) = 0,5628$$

Damit haben wir die Summe der Wahrscheinlichkeiten zur Berechnung der Verteilungsfunktion als Flächen gedeutet, die wir durch die Fläche unter der φ^*-Funktion approximiert haben. Diese erlaubt beim Übergang zur stetigen Grenzkurve eine Berechnung als Integral (obwohl die einzelne Ordinate der φ-Funktion keine Deutung als Wahrscheinlichkeit mehr hat).

Mit Hilfe der Gauss-Funktion können wir auch die Tschebycheff-Ungleichung verschärfen und die Bemerkung auf Seite 73 begründen, falls sich eine Zufallsgröße X durch eine Gaussfunktion beschreiben lässt (normalverteilte Zufallsgrößen, vgl. Seite 110) oder wenigstens approximieren lässt (z. B. oft binomialverteilte Zufallsgrößen nach dem lokalen Grenzwertsatz von Moivre-Laplace). Dann lässt sich die Wahrscheinlichkeit des Abweichens von X vom Erwartungswert μ = E(X) um höchstens den Wert $c \in \mathbb{R}^+$ leicht ausdrücken:

$$P(|X-\mu| \le c) = P(\mu - c \le X \le \mu + c)$$
$$= P(X \le \mu + c) - P(X < \mu - c)$$
$$= \phi\left(\frac{\mu + c - \mu}{\sigma}\right) - \phi\left(\frac{\mu - c - \mu}{\sigma}\right)$$
$$= \phi\left(\frac{c}{\sigma}\right) - \phi\left(\frac{-c}{\sigma}\right) = 2\phi\left(\frac{c}{\sigma}\right) - 1.$$

Hierbei haben wir die Symmetrieeigenschaften der φ-Funktion benutzt.

Sinnvollerweise drückt man die Abweichung vom Erwartungswert durch Vielfache der Standardabweichung σ aus, also $c = k\cdot\sigma$, $k \in \mathbb{R}^+$.

Für Zufallsgrößen X, die unseren Voraussetzungen genügen, haben wir damit

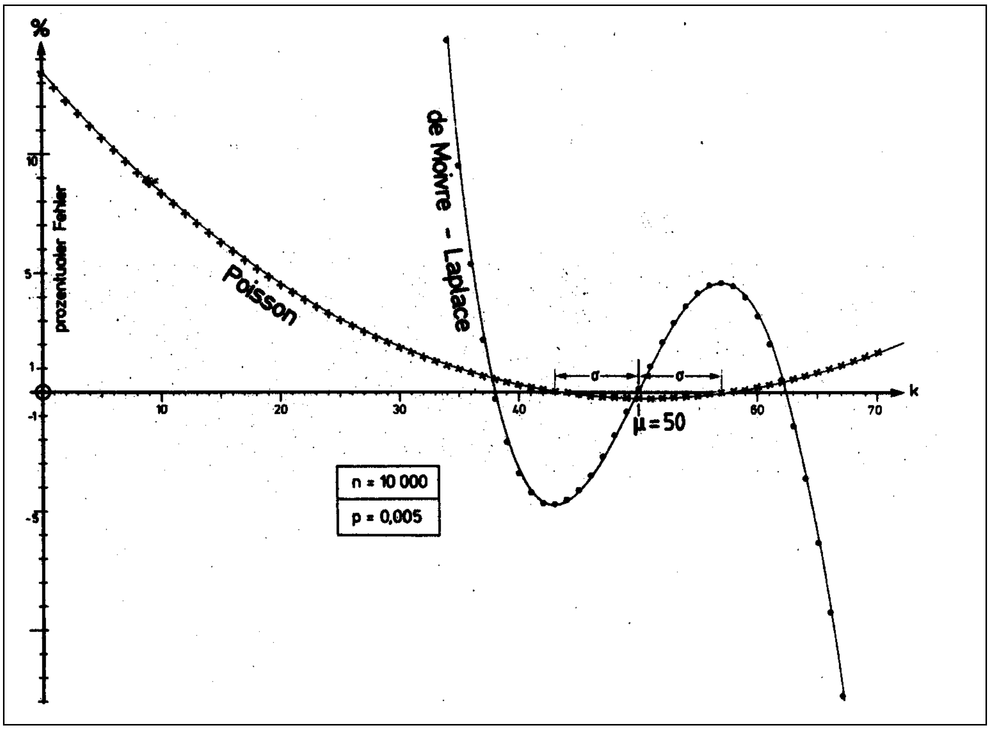
$$P(|X-\mu|) \le r \cdot \sigma = 2 \cdot \phi(r) - 1.$$

Da die φ-Funktionen numerisch berechnet werden, erhält man für die am häufigsten benötigten Werte die σ-Regeln

$$P(|X-\mu| \leq \sigma = 0,6827$$
$$P(|X-\mu| \leq 2\sigma) = 0,9545$$
$$P(|X-\mu| \leq 3\sigma) = 0,9973$$

Bemerkung zur Güte der Approximation

Das folgende Bild zeigt den prozentualen Fehler bei der Approximation der Binomialverteilung B(10.000, 0,005) mit De Moivre – Laplace und mit Poisson: μ = 50, σ = 7,05.



2.3.4.2 Der zentrale Grenzwertsatz

Die Bedeutung der φ und ϕ-Funktion reichen weit über die Approximation der Binomialverteilung hinaus. Am wichtigsten ist der zentrale Grenzwertsatz, dessen Beweis schulische Methoden bei weitem übersteigt. Er besagt, dass die durch φ bzw. ϕ ausgedrückte Näherungsformel praktisch für alle Verteilungen gilt, die als Summe von hinreichend vielen unabhängigen Zufallsgrößen aufgefaßt werden können.

$X_1, ..., X_n$ seien unabhängige Zufallsgrößen, und $X = X_1 + ... + X_n$. T sei die „standardisierte" Zufallsgröße zu X, also

$$T = \frac{X - \mu(X)}{\sqrt{\mathrm{Var}(X)}} = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu(X_i)}{\sqrt{\sum_{i=1}^{n} \mathrm{Var}(X_i)}}$$

mit $\mu(T) = 0$, $\sigma(T) = 1$.

Dann gilt für genügend großes n

$$P(T \leq k) \approx \phi(k)$$
$$P(k_1 \leq T \leq k_2) \approx \phi(k_2) - \phi(k_1) \quad \text{bzw.}$$
$$P(X \leq a) \approx \phi\left(\frac{a - \mu(X)}{\sigma(X)}\right)$$

<u>Beispiel 1</u>: Messfehler

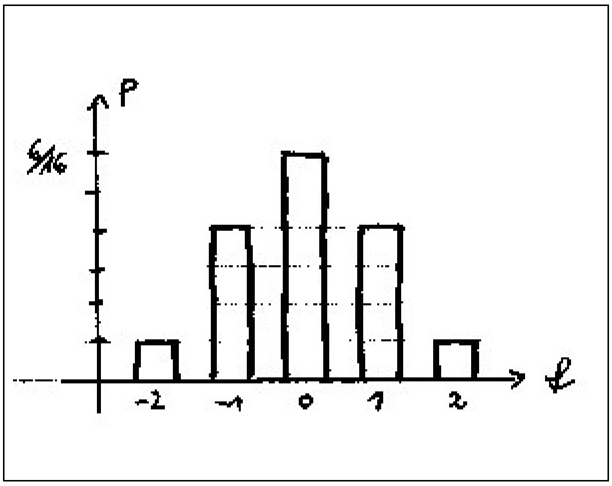
Ein typisches Beispiel sind Messungen, z. B. die Messung der Länge eines Stabes: Wir gehen von 4 voneinander unabhängigen Fehlerquellen aus

- Veränderung des Stabes X_1
- Veränderung des Maßstabes X_2
- ungenauer Beobachtung X_3
- sonstige Einflüsse X_4

Die an sich stetig eintretenden Fehlerquellen seien der Einfachheit halber mit möglichen Ergebnissen ± 0,5 angenommen, die jeweils mit diskreter Wahrscheinlichkeit p = ½ auftreten. Der Gesamtfehler f kann dann als Wert der Zufallsgröße $F = X_1+X_2+X_3+X_4$ angenommen werden:

f_i	-2	-1	0	1	2
$p(F = f_i)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

Das zu diesem einfachen Mess-Modell gehörige Diagramm zeigt schon deutliche Glockenform:

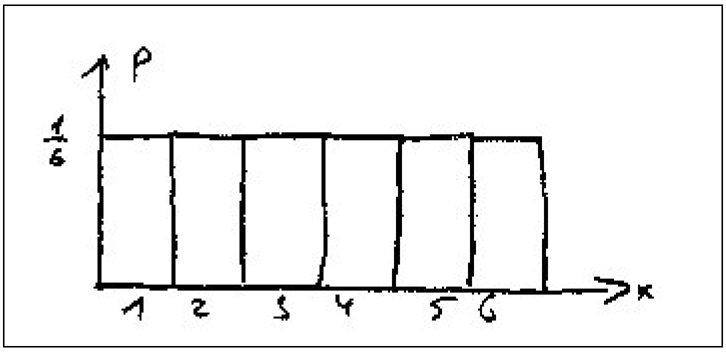


Beispiel 2 : „Zensuren-Würfeln“
X = Augenzahl bei Würfeln mit Laplace-Würfel

Würfeln mit n-Würfeln: $Y = X_1 + X_2 + ... + X_n$

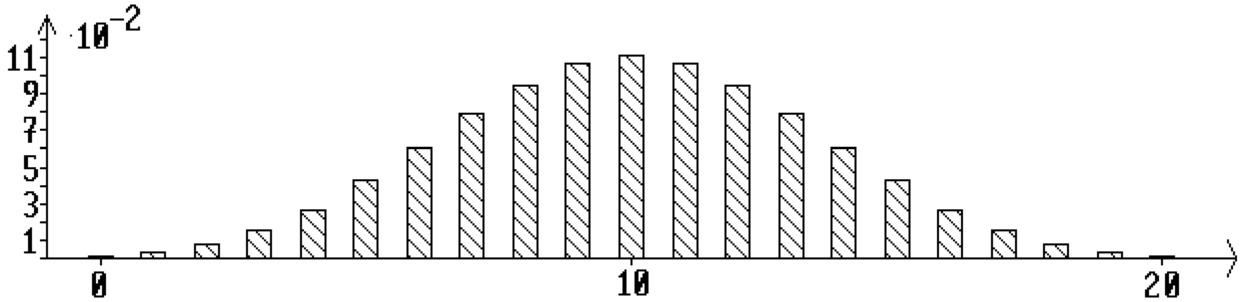
Klar ist, dass die X_i unabhängig sind.

n = 1: $Y = X_1$ Gleichverteilung



n = 4: $Y = X_1 + X_2 + X_3 + X_4$

(für eine Note ordne man z. B. dem Wert 24 die Note 1^+, 23 die Note 1, 22 die Note 1^- usw. zu). Die Verteilung von Y zeigt schon deutlich die Glockenform!



2.3.4.3 Stetige Zufallsgrößen

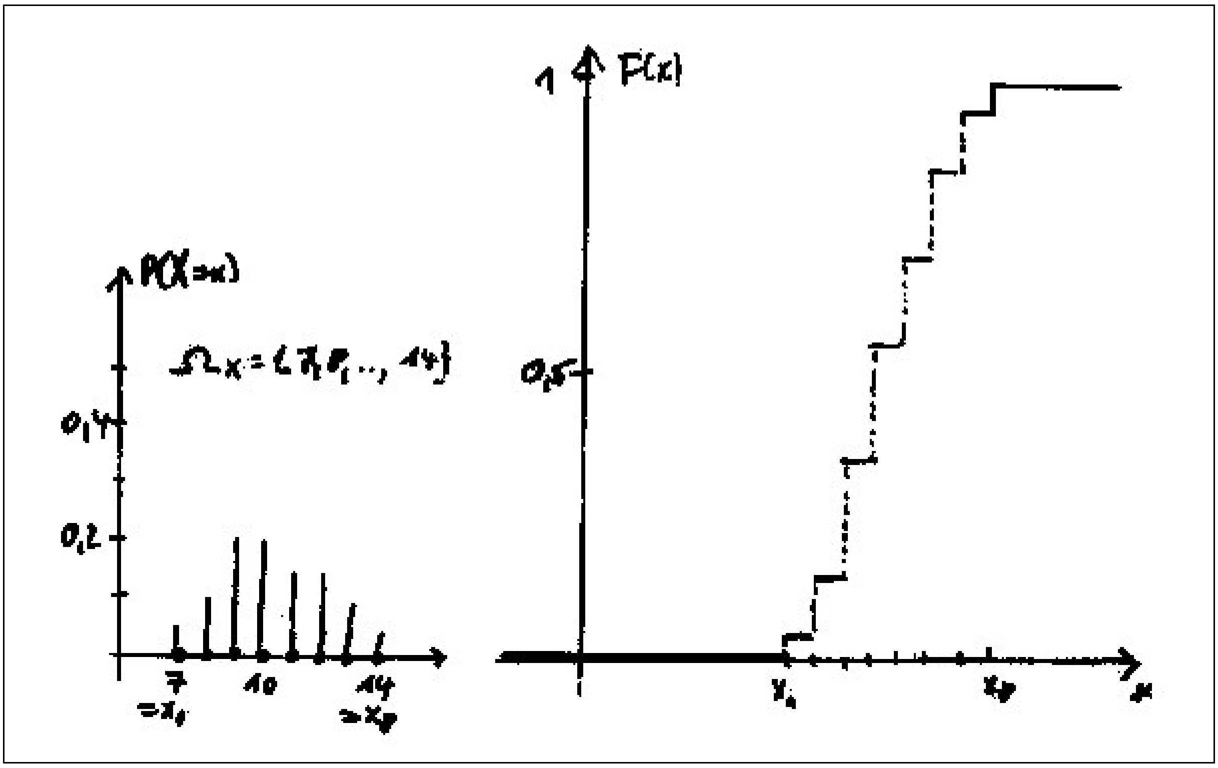
Bisher hatten wir nur endliche Zufallsgrößen mit diskreten Wahrscheinlichkeitsverteilungen betrachtet. Die stetigen Funktionen φ und ϕ dienten nur der Approximation der diskreten Binomialverteilung. Nun gibt es sicherlich Zufallsexperimente, zu deren Modellierung stetig verteilte Merkmale sinnvoll wären, etwa die oben betrachteten Messfehler.

Bisher hatten wir eine Zufallsgröße X mit Wertemenge $\{x_1,...,x_n\}$ und Wahrscheinlichkeitsverteilung P. Die zugehörige Verteilungsfunktion

$$F:\mathbb{R} \to \mathbb{R}, x \to F(x) = P(X \le x) = \sum_{x_i \le x} P(X = x_i)$$

ist zwar für ganz $\mathbb{R}$ definiert, ihre Funktionswerte sind aber nur endliche Summen, deren Werte klar definierte Wahrscheinlichkeiten sind, ihr Graph ist monoton wachsend von 0 auf 1 und besteht aus zur x-Achse parallelen Geradenstückchen. Die folgende Abbildung zeigt ein Beispiel.

Zufallsgröße X mit $\Omega_x = \{x_1, ..., x_n\}$
Verteilungsfunktion $F:\mathbb{R} \to \mathbb{R},\ x \mapsto F(x) = P(X \le x) = \sum_{x_i \le x} P(X = x_i)$



In dieser Verteilungsfunktion liegt der Ansatz, stetige Wahrscheinlichkeitsverteilungen zu betrachten:

Definition:

a) Eine nicht negative Funktion f heißt Dichtefunktion, wenn gilt $\int_{-\infty}^{\infty} f(t)dt = 1$.

b) Eine Zufallsgröße X heißt stetig, wenn sich ihre Verteilungsfunktion F darstellen läßt als

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

mit einer Dichtefunktion f.

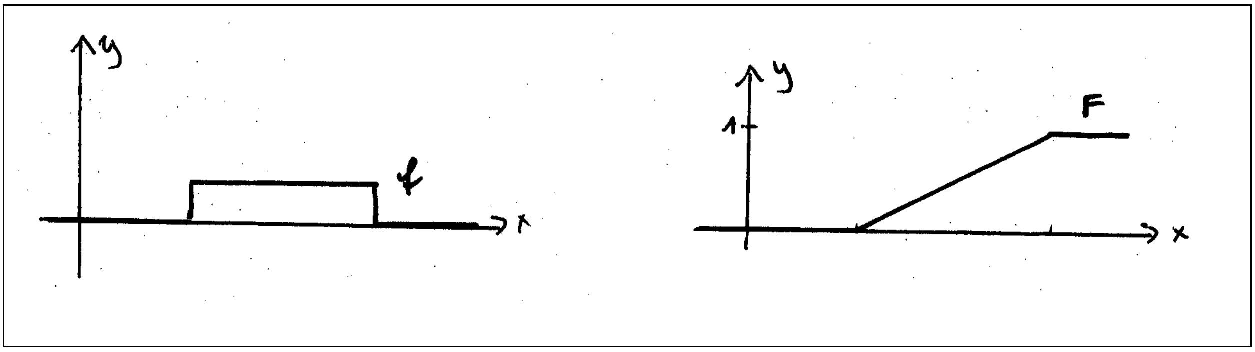
Im diskreten Fall entspricht die Wahrscheinlichkeitsverteilung P der Dichtefunktion f. Im stetigen Fall haben aber die Abszissen von f keine Deutung als Wahrscheinlichkeiten, erst die Flächeninhalte unter der f-Kurve lassen sich so deuten:

$P(X \le x) = F(x)$
$P(x_1 \le X \le x_2) = F(x_2) - F(x_1)$.

Einige Beispiele für stetige Verteilungen sind:

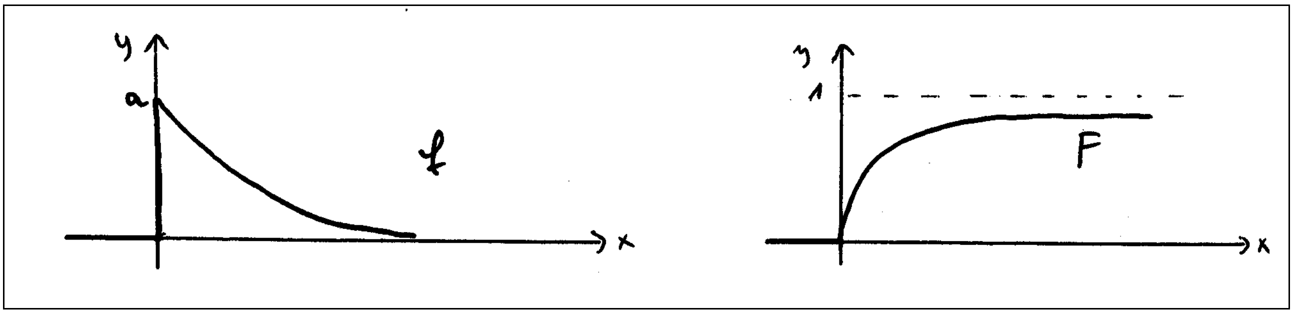
a. **Gleichverteilung auf einem Intervall**,
z. B. ein Zufallsgenerator erzeugt eine reelle Zahl zwischen a und b.

$$f(x) = \begin{cases} 1/(b-a) & \text{für } a \le x \le b \\ 0 & \text{sonst} \end{cases} \qquad F(x) = \begin{cases} 0 & \text{für } x < a \\ \dfrac{x-a}{b-a} & \text{für } a \le x \le b \\ 1 & \text{für } x > b \end{cases}$$



b. **Exponentialverteilung,**
z. B. Wartezeiten (Eintreffen zum Zeitpunkt 0, Wartezeit X)

$$f(x) = \begin{cases} a \cdot e^{-ax} & \text{für } x \ge 0 \\ 0 & \text{für } x < 0 \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-ax} & \text{für } x \ge 0 \\ 0 & \text{für } x < 0 \end{cases}$$



c. **<u>Normalverteilung,</u>**
z. B. Verteilung zufälliger Messfehler

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) \qquad = \phi\left(\frac{x-\mu}{\sigma}\right)$$

Diese Verteilung nimmt eine Sonderstellung in der Wahrscheinlichkeitsrechnung ein:
- die mathematische und numerische Behandelbarkeit im Vor-Computerzeitalter;
- nach dem zentralen Grenzwertsatz und vielen anderen Grenzwertsätzen ist die Verteilungsfunktion einer Zufallsgröße zumindest approximativ durch die φ-Funktion berechenbar, wenn die Zufallsgröße ihre Werte unter dem Einfluss mehrerer unabhängig additiv wirkender Faktoren annimmt.

Auf dem derzeitigen deutschen 10 DM-Schein ist die Gauss untersuchte φ-Kurve abgebildet. (vgl. Abb.)

Beispiele für solche Verteilungen:
- Verteilung von Körpergröße und Gewicht bei Personen eines Geschlechts und einer Altersgruppe in Deutschland;
- Lebensdauer technischer Produkte;
- Treffgenauigkeit bei Schießwettbewerben.

Gegenbeispiele:
- Häufigkeitsverteilung der Dauer des Schulbesuchs;
- Anzahl der Kinder bei deutschen Familien;
- Abweichungen vom Fahrplan;
- Alterspyramide der deutschen Bevölkerung.

Analog zu den diskreten Zufallsgrößen definiert man bei stetigen Zufallsgrößen X:
<u>Definition:</u>

$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$ <u>Erwartungswert von X</u>

$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} \left((x-\mu)^2 \cdot f(x)\right) dx$ <u>Varianz von X</u>

Damit gilt für normalverteilte Zufallsgrößen X mit $f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$

<u>Satz:</u> $E(X) = \mu, Var(X) = \sigma^2$

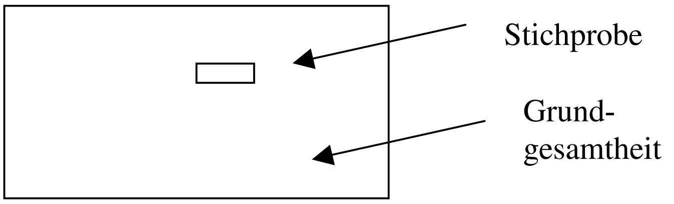
(ohne Beweis)

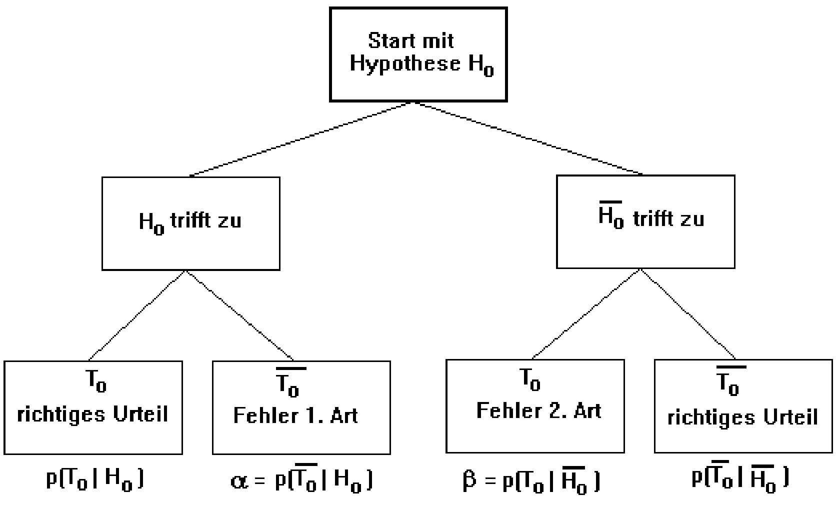
3. Grundlagen der „beurteilenden Statistik“

3.1 Testen

3.1.1 Das Testproblem

Die folgende vereinfachte Situation soll das Testproblem erläutern: Wir haben von vornherein gewisse Annahmen über die Grundgesamtheit, die Nullhypothese H_0. Die Gegenhypothese $H_1 = \overline{H_0}$ ist die Verneinung der Nullhypothese. Aufgrund einer zufälligen Stichprobe entscheidet man sich für T_0, d. h. „Annahme“ von H_0, oder T_1, d. h. Zurückweisung von H_0 (also Annahme von H_1). Hierbei sind vier Fälle möglich:





Für eine sinnvolle Entscheidung müssen das Signifikanzniveau $\alpha = p\ (T_1 \mid H_0)$ („Lieferantenrisiko“) und $\beta = p\ (T_0 \mid H_1)$ („Käuferrisiko“) möglichst klein sein. Die Namen werden klar, wenn man an die Stichprobe eines Käufers bei einer Sendung mit vom Lieferanten zugesicherten Eigenschaften denkt.

Die Berechnung des Unsicherheitsgrades der Entscheidung ist eine Hauptaufgabe der beurteilenden Statistik.

3.1.2 Einseitige Tests

Beispiel: Qualitätskontrolle
Ein Betrieb stellt Keramikisolatoren her. Aus langjähriger Erfahrung wird mit 20 % Ausschuss kalkuliert. In gewissen Abständen wird anhand von Stichproben überprüft, ob dies weiterhin der Fall ist, wobei nur interessiert, ob die Ausschussquote gestiegen ist. Falls auffällige Abweichungen nach oben auftreten, erfolgt Meldung an die Betriebsleitung. Konkret: Es werden 100 Isolatoren zufällig entnommen und geprüft. Vorausgesetzt wird die „Null-Hypothese“.

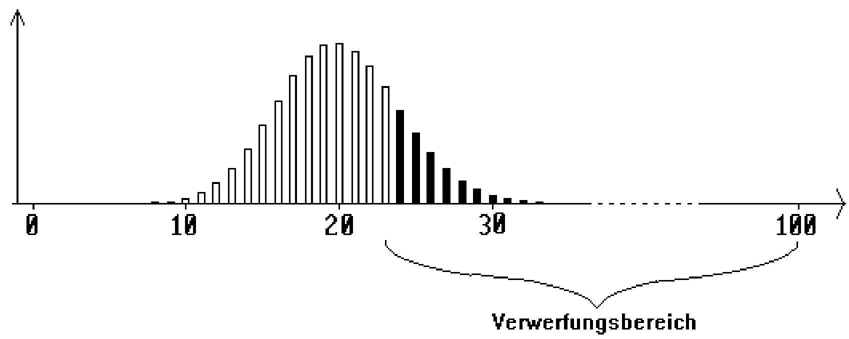
H_0: „Wahrscheinlichkeit für defekte Isolatoren ist $p \leq 20$ %
$H_1 = \overline{H_0}$ logisches Gegenteil, also hier H_1 = „$p > 0{,}2$“

Man rechnet mit p = 20 %. Wenn man dies ablehnt, so auch alle anderen Fälle mit $p < 0{,}2$. Bei Nichtablehnen hat man ohnedies keine Erkenntnis.

Die „Prüfgröße“ sei

X: „Anzahl der defekten unter den 100 geprüften Isolatoren“

X ist B(100, 0.2)-verteilt; man erwartet also $\mu = E(X) = 100 \cdot 0{,}2 = 20$ defekte Isolatoren mit einer Standardabweichung von $\sigma = \sqrt{100 \cdot 0{,}2 \cdot 0{,}8} = 4$. Es ist Meldung zu machen, wenn mindestens $\mu + \sigma$ defekte Isolatoren dabei sind. Der „Verwerfungsbereich“ für H_0 ist also $V = \{24, 25, ..., 100\}$ (vgl. Abb.) (Natürlich kann man dann auch alle Hypothesen mit $p < 0{,}2$ verwerfen). Interessant sind nur die Abweichungen nach oben; der Verwerfungsbereich ist zusammenhängend rechts von μ konstruiert worden. Solche Tests nennt man einseitige Tests.



Auch bei korrektem p = 0,2 liegt rein zufällig die Anzahl defekter Isolatoren mit der Wahrscheinlichkeit

$$\alpha = P_{0,2}(V) = \sum_{k=24}^{100} B(100,\ 0.2,\ k) \approx 0{,}189\ ,$$

im Verwerfungsbereich, was dann zu einem Fehler 1. Art mit der Irrtumswahrscheinlichkeit $\alpha \approx 19\%$ führt.

Aber auch für fehlerhafte Produktion, also p > 0,2, kann zufälligerweise die Anzahl defekter Isolatoren im „Annahmebereich" $\overline{V} = \{0,1,...,23\}$ liegen, was dann zu einem Fehler 2. Art führt. Das große Problem ist natürlich, dass man p nicht kennt! Es gelte z. B. in Wirklichkeit

H': „Wahrscheinlichkeit für defekte Isolatoren ist p = 0,25"

Dann ist die Wahrscheinlichkeit β für einen Fehler 2. Art.

$$\beta = P_{0,25}\left(\overline{V}\right) = \sum_{k=0}^{23} B(100,\ 0.25,\ k) \approx 0,371.$$

Die Wahrscheinlichkeit β = β(p) für p > 0,2 eines Fehlers 2. Art hängt also von der wahren Ausschusswahrscheinlichkeit ab: Die für alle p ∈ [0,1] definierte Funktion

$$\chi : p \mapsto \chi(p) = \sum_{k=0}^{23} B(100, p, k)$$

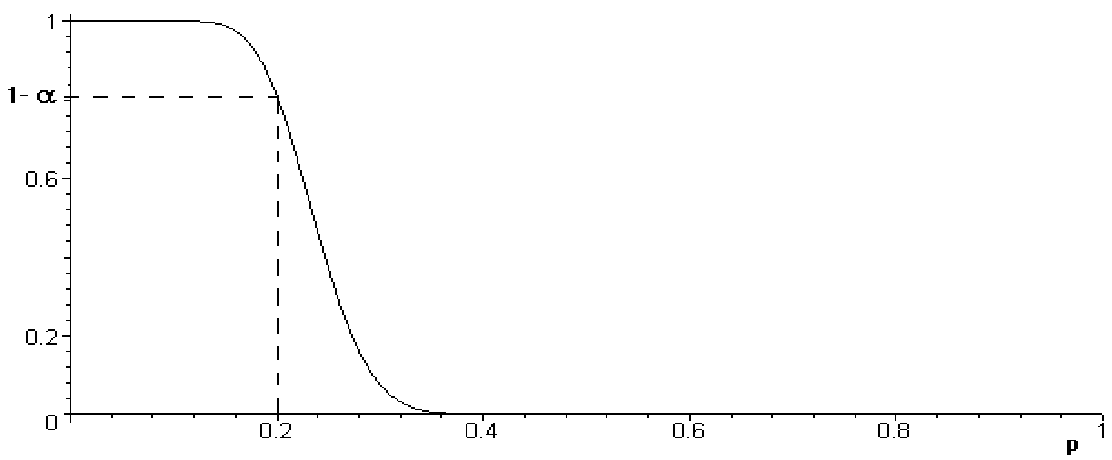
heißt <u>Operationscharakteristik</u> des Tests, einige Werte in unserem Beispiel sind

P	0	0,1	0,2	0,25	0,3	0,4	1
χ (p)	1	1,000	0,811	0,371	0,076	0,000	0

Wenn p > 0,2 ist, so ist χ(p) = β(p) der β-Fehler.

Ideal wäre $\chi(p) = \begin{cases} 1 \text{ für } p \leq 0,2 \\ 0 \text{ für } p > 0,2 \end{cases},$

was leider bei realen Tests keinesfalls möglich ist (und eventuell auch gar nicht wünschenswert ist). Im Gegenteil ist für p nur wenig größer als 0,2 sogar β ≈ 1 - α = 0,811 in unserem Beispiel! Der Graph der χ-Funktion zeigt dies deutlich:



χ(p)

Tatsächlich\Entscheidung	H_o verwerfen	H_o beibehalten
H_o richtig (d. h. p ≤ 0.2)	Fehler 1. Art Wahrscheinlichkeit $\alpha = P_{0.2}(V)$	Okay! Wahrscheinlichkeit $P_{0.2}(\overline{V})=1-\alpha$
H_o falsch (d. h. p > 0.2)	Okay! Wahrscheinlichkeit $P_p(V) = 1 - \beta_p$	Fehler 2. Art Wahrscheinlichkeit $\beta_p = P_p(\overline{V})$

In der Praxis gibt man die Irrtumswahrscheinlichkeiten, d. h. das Signifikanzniveau α vor, z. B. Industrie 5 %, Medizin 1% oder 0,1%. Etwa scheint in unserem Beispiel eine Irrtumswahrscheinlichkeit von 18,9 % nicht tragbar zu sein, man gibt $\alpha_0 = 5$ % vor. Gesucht wird jetzt ein Verwerfungsbereich V‘ derart, dass $\alpha' \le \alpha_0$, d. h. dass $P_{0,2}(V') \le \alpha_0 = 0{,}05$ gilt. Man findet z. B. durch eine Tabelle oder Computereinsatz:

$\alpha' = P_{0,2}\,(27,28,...,100) \approx 0{,}056$ noch zu groß,
$\alpha' = P_{0,2}\,(28,29,...,100) \approx 0{,}034$ also okay.

Diese Irrtumswahrscheinlichkeit α' ist maximal unter allen möglichen, die $\le \alpha_0$ sind, damit die Fehlermöglichkeiten 2. Art möglich klein sind. Allerdings bedeutet eine Verkleinerung von α und V stets eine Vergrößerung von $\overline{V}$ ‘ und damit von β. Hier gilt für p = 0,25

$$\beta(0,25) = P_{0,25}\left(\overline{V}'\right) = \sum_{k=0}^{27} B(100,\ 0.25,\ k) \approx 0{,}722,$$

was den Test fragwürdig macht. Wie der Test tatsächlich ablaufen wird, hängt von vielerlei Faktoren ab (z. B. Abnehmererwartungen, Firmenimage, Kosten für Kontrollen, Kosten für Änderungen usw.). Die Mathematik kann nur Hilfe geben, aber nicht entscheiden.

Allgemeines Vorgehen bei Hypothesentests:

1) Es gibt Anlass, die Richtigkeit einer Hypothese H_o, zu überprüfen.
2) Man nimmt an, dass H_o richtig ist. (Null-Hypothese).
3) Man legt ein Signifikanzniveau α_o fest.
4) Man definiert ein Zufallsexperiment zum Überprüfen von H_o und bestimmt – entsprechend dem Testinteresse – einen geeigneten, größtmöglichen Verwerfungsbereich V (d. h. eine Menge von Ergebnissen, die zu der Entscheidung führen sollen, H_o als falsch anzusehen), so dass $P\,(V|H_o) \le \alpha_o$.
5) Man führt das Experiment durch. Liegt das aufgetretene Ergebnis in V, so sagt man „Der Test hat ein signifikantes Ergebnis erbracht.“
6) Man entscheidet aufgrund des aufgetretenen Ergebnisses. Dabei können Fehler unterlaufen:
   a) Fehler 1. Art: H_o wird irrtümlich verworfen, obwohl H_o richtig ist. Die Irrtumswahrscheinlichkeit $\alpha = P(V \mid H_o)$ ist die Wahrscheinlichkeit für Fehler 1. Art.
   b) Fehler 2. Art: H_o wird irrtümlich nicht verworfen, obwohl H_o falsch ist.
7) Die Wahrscheinlichkeit β für einen Fehler 2. Art lässt sich im allgemeinen nicht bestimmen.

Bem. zu 4) Dies muss vor Ziehen der Stichprobe geschehen! Es ist Selbstbetrug, erst nach Vorliegen der Stichprobe (z. B. medizinische Ergebnisse) einen Test zu konstruieren, der das gewünschte Resultat „beweist“.

Die Funktion

$\chi : [0,1] \to \mathbb{R}, p \mapsto \chi(p) = P_p(\overline{V}) = P(\overline{V} \mid \text{"p liegt vor"})$ heißt Operationscharakteristik eines Tests, der überprüft, ob die Wahrscheinlichkeit p die richtige ist. Ist p_0 die (Grenz-) Wahrscheinlichkeit, die in der Nullhypothese angenommen wird, so gilt $\chi(p) \approx 1-\alpha$ für p in der

Nähe von p_0. Die analog gebildete Funktion $\gamma = 1-\chi : p \mapsto P_p(V) = 1-\chi(p)$ heißt <u>Gütefunktion</u> des Tests.

<u>Bemerkungen</u>

(A) <u>Analogie zu Gerichtsverfahren</u>

ad 1)	Es gibt Anlass, einen Menschen in Bezug auf eine begangene Tat zu überprüfen (evtl. anzuklagen).
ad 2)	Es gilt die Unschuldsvermutung (als Grundprinzip unserer Rechtsordnung).
ad 3)/4)	Es werden Kriterien festgelegt, welche über Schuld oder Unschuld entscheiden sollen, und es werden (i. a. hohe) Standards für diese Kriterien gesetzt.
ad 5)	Es werden Indizien entsprechend diesen Kriterien zusammengetragen.
ad 6)	Es wird entsprechend diesen Indizien ein Urteil gefällt. Dabei können Fehler unterlaufen: a) 1. Art: Ein Unschuldiger wird irrtümlich verurteilt. Das sollte in unserer Rechtsordnung möglichst vermieden werden (d. h. es sollte nur mit sehr geringer Wahrscheinlichkeit passieren; daher die hohen Standards). b) 2. Art: Ein Schuldiger wird irrtümlich „mangels Beweisen" freigesprochen. Das sollte zwar ebenfalls vermieden werden, ist aber eher in Kauf zu nehmen, um Ersteres zu vermeiden (d. h. kann ggfs. mit nicht geringer Wahrscheinlichkeit passieren).

(B) <u>Analogie zu indirekten Beweisen</u>

Bei diesen schließt man: „Falls A, dann B. Da aber nicht B, gilt folglich nicht A". Bei der statistischen Schlussweise beim Testen argumentiert man: „Falls H_0 richtig, so ist mit hoher Wahrscheinlichkeit das Ergebnis nicht in V. Da aber das Ergebnis in V ist, ist folglich vermutlich H_o falsch".

(C) <u>Irreführende Deutungen</u>

Wir haben die beiden Wahrscheinlichkeiten

$$\alpha = P(V/H_0),\quad \beta = P\left(\overline{V}/\overline{H_0}\right).$$

Diese werden oft mit den folgenden verwechselt:

$$\tilde{\alpha} = P(H_0/V),\ \tilde{\beta} = P\left(\overline{H_0}/\overline{V}\right).$$

$\tilde{\alpha}$ ist die „Wahrscheinlichkeit", dass die Ausschussrate 20 % beträgt, obwohl mindestens 28 Isolatoren der Stichprobe defekt sind. Wer beim Auftreten von 28 defekten Isolatoren das Ergebnis als „mit 95%-iger Sicherheit ist die Ausschussrate höher als 20 %" beschreibt, missdeutet α im Sinne von $\tilde{\alpha}$. Die übliche Bezeichnung, 1 - α als „statistische Sicherheit" zu beschreiben, ist etwas irreführend. $\overline{V}$ heißt üblicherweise „Annahmebereich" des Tests, was irreführend ist, da H_0 höchstens „beibehalten" wird, d. h. „nicht verworfen" wird, aber eben nicht „akzeptiert" wird. (vgl. „Freispruch mangels Beweisen")

Man beachte, dass man eigentlich <u>nicht</u> von den Wahrscheinlichkeiten „$p(H_0)$" und „$p(H_0/V)$ sprechen kann, obwohl dies ja gerade das Wünschenswerte wäre. Die Hypothese Ho gilt oder gilt nicht. Im frequentistischen Sinn ist „$p(H_0)$" nicht definiert.

Man beachte die Analogie zu Gerichtsverfahren: Der Angeklagte ist der Mörder oder ist nicht der Mörder, es gibt keine Wahrscheinlichkeit *per se*, dass er der Mörder ist. Im Indizienbeweis wird aufgrund von Indizien die Hypothese H_0 „Er ist unschuldig" beibehalten („Freispruch mangels Beweisen", entspricht dem Bereich $\overline{V}$) oder verworfen („Verurteilung aufgrund von Indizien", entspricht dem Verwerfungsbereich V).

Die subjektivistische Bayessche Sichtweise würde es dagegen schon erlauben, von subjektiven Wahrscheinlichkeiten $p(H_0)$ und $p(H_0/V)$ zu sprechen. Wir müssen uns aber im klassischen frequentistischen Sinn mit „$p(V/H_0)$" begnügen.

(D) <u>Alternativtests</u>

Genaue Aussagen über β lassen sich bei Alternativtests machen, wenn wirklich nur 2 Fälle für die Wahrscheinlichkeit möglich sind, z. B. echte Laplace-Münze oder gefälschte Münze mit $p = 0{,}6$ für „Kopf". Schauen wir dieses Beispiel genauer in 2 Fällen an: Für eine Münze ist bekannt

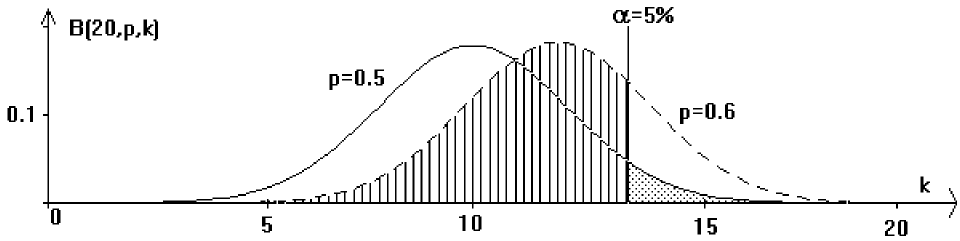
a) entweder $p = 0{,}5$ oder $p = 0{,}6$ für Kopf
b) entweder $p = 0{,}5$ oder $p = 0{,}8$ für Kopf

H_o: Die Münze ist gut.

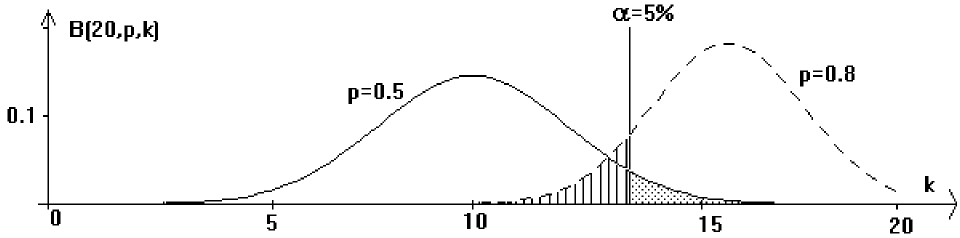
Wir werfen 20 mal und entscheiden uns für die Annahme der Hypothese auf dem $\alpha = 5\% - \text{Niveau}$, was zu einem Verwerfungsbereich V = {14,15,...,20} führt. In den folgenden beiden Abbildungen haben wir die Binomialverteilung durch die Normalverteilung

$$\varphi^*(x) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x-\mu}{\sigma}\right)$$

approximiert (vgl. 2.3.4).



0

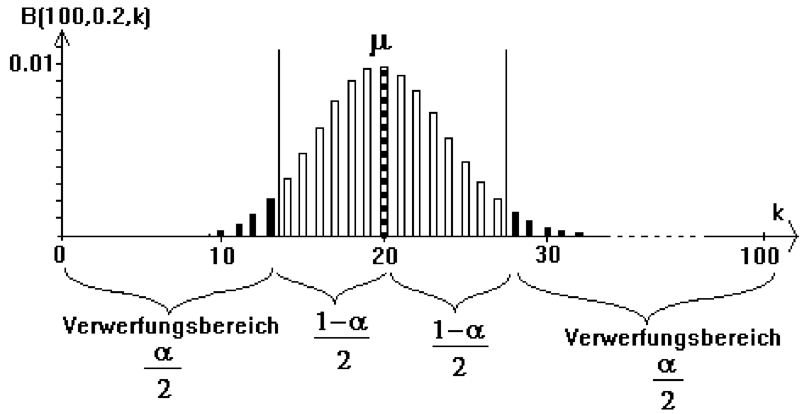


Die rechts des α = 5%-Balkens liegende punktierte Fläche unter der Kurve für p = 0,5 entspricht dem Fehler 1. Art auf 5%-Niveau. Die jeweils links dieses Balkens gelegene schraffierte Fläche unter der p = 0,6- bzw. der p = 0,8-Kurve entspricht dem Fehler 2. Art mit β = 77,5° im ersten Fall (also Test unbrauchbar) und β = 9,1 % im 2. Fall.

3.1.3 Zweiseitige Tests

In Erweiterung des einseitigen Tests beim Isolatoren-Beispiel sollen jetzt auch Abweichungen nach unten interessieren (z. B. weil bei deutlich geringerem Ausschussanteil der Preis reduziert werden kann). Auch beim Münzenbeispiel sind natürlich Abweichungen nach oben und nach unten von μ relevant, wenn man schlechte Münzen ausschließen will.

Die Isolatoren mögen wie eben mit n = 100 und entsprechendem H_0 (d. h. H_0: Wahrscheinlichkeit für defekte Isolatoren ist 0,2) getestet werden. Das Signifikanzniveau sei α_o = 10 %; diese 10 % sollen möglichst gleichmäßig auf die beiden Teile des Verwerfungsbereichs V verteilt werden; d. h. genauer, dass der Bereich „1- α" möglichst gleichmäßig um den Erwartungswert μ herumgelegt werden soll. (Bei stetigen Zufallsgrößen lässt sich das mit Hilfe des Integral exakt beschreiben.)



Mit einer Tabelle oder mit Computerhilfe erhält man

$$\sum_{k=0}^{13} B(100,\ 0.2,\ k) \approx 0{,}047 \quad \text{und} \quad \sum_{k=28}^{100} B(100,\ 0.2,\ k) \approx 0{,}034\,,$$

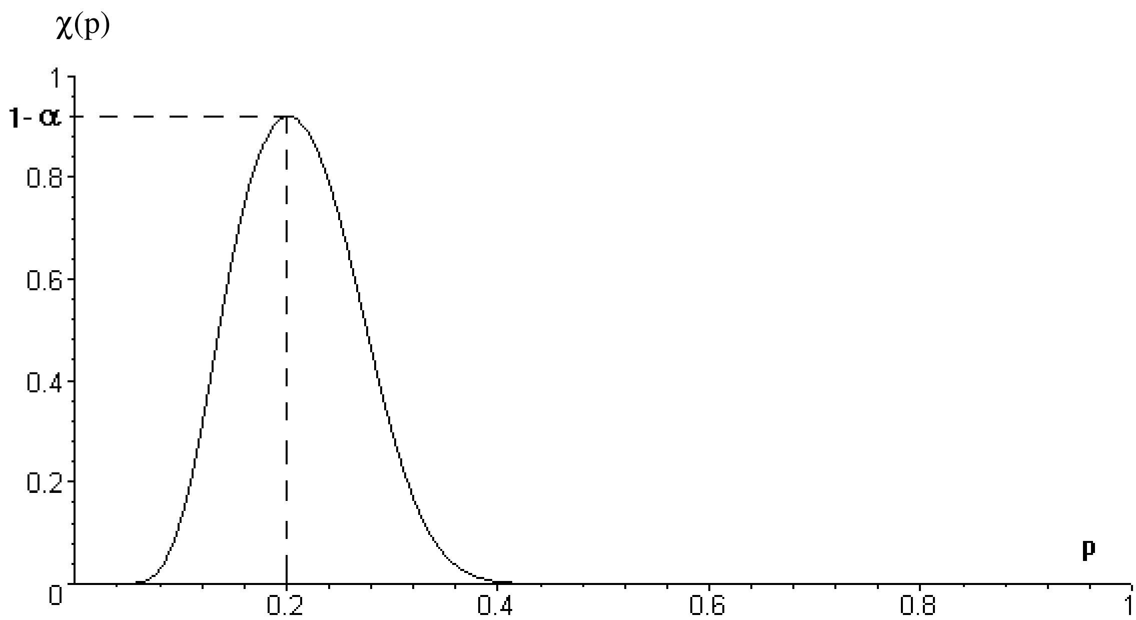
und dies ist optimal. Für $V = \{0,1,...,13,28,29,...,100\}$ gilt also $\alpha = P_{0,2}(V) \approx 0{,}081 < 0{,}1$ wie verlangt.

Die Testvorschrift ist somit: Wähle zufällig 100 Isolatoren und prüfe sie. Falls die Anzahl der „defekten" zwischen 14 und 27, so tue nichts. Falls aber diese Anzahl kleiner 14 oder größer 27 ist, so erstatte Meldung.

Die Operationscharakteristik hat nun einen anderen, für zweiseitige Tests typischen Verlauf:

$$\chi : p \mapsto \chi(p) = \sum_{k=14}^{27} B(100, p, k)$$

P	0	0.1	0.2	0.25	0.3	0.4	0.5	1
$\chi(p)$	0	0,124	0,919	0,720	0,296	0,005	0,000	0



In diesem Fall ist $\chi(p) = \beta(p)$ für $p \neq 0{,}2$.

Wieder ideal, aber unerreichbar, wäre hier

$$\beta(p) = \begin{cases} 1 & \text{für } p = p_0 = 0{,}2 \\ 0 & \text{sonst} \end{cases}$$

Man überlege, wie sich die Operationscharakteristik ändert, wenn bei gleichem Signifikanzniveau (hier $\alpha_o = 10\%$) die Größe der Stichprobe verändert wird (n = 200, 500, 1000,...). Liefert das eine bessere Entscheidungshilfe?

3.2 Schätzen

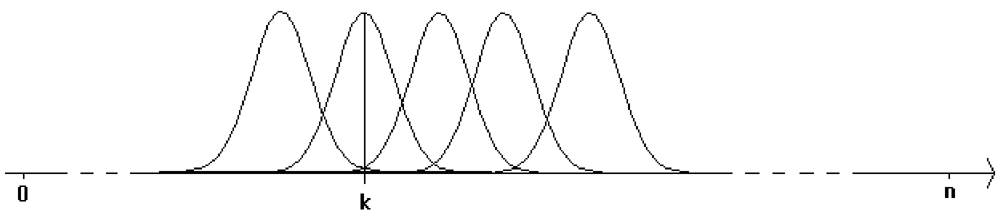
3.2.1 Punktschätzen

Oft ist die Wahrscheinlichkeit p für das Auftreten eines Ergebnisses E unbekannt. Wir haben uns dann auf das Gesetz der großen Zahl berufen, n mal das zugehörige Experiment durchgeführt, wobei k mal E eingetreten sein möge, und dann $p \approx \frac{k}{n}$ geschätzt. Die Frage ist, inwieweit diese Schätzung optimal ist. Es sei

X: Anzahl des Eintretens von E bei n Versuchen.

Klar ist, dass X $B(n,p_*)$-verteilt ist mit einer unbekannten Wahrscheinlichkeit p_*. Nun sei bei der konkreten Durchführung k von n mal das Ereignis E eingetreten.

Dann schätzen wir als Zahlwert für p_* dasjenige $p_o \in [0;1]$, das unter allen p den größten Wert für B(n,p,k) liefert (sofern dieser eindeutig existiert). Geometrisch bedeutet das: Wir suchen unter allen Glockenkurven diejenige, bei welcher der Wahrscheinlichkeitswert für k am größten ist (vgl. Abb.)



Formal ausgedrückt: Wir suchen das Maximum der Funktion $L: p \mapsto B(n,p,k)$. Es zeigt sich tatsächlich, dass dies bei $p_0 = \frac{k}{n}$ angenommen wird.

Zum Beweis bestimmen wir den maximalen Wert von L(p) = B(n,p,k) für p ∈ [0;1] mittels Differentialrechnung:

$$L'(p) = \frac{d}{dp}\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \binom{n}{k} \cdot \left(kp^{k-1} \cdot (1-p)^{n-k} + p^k \cdot (n-k)(1-p)^{n-k-1}(-1)\right) =$$

$$= \binom{n}{k} p^{k-1}(1-p)^{n-k-1}\left(k(1-p) + p(n-k)(-1)\right) = \binom{n}{k} p^{k-1}(1-p)^{n-k-1}(k-pn).$$

Damit folgt wie behauptet

$$L'(p) = 0 \Leftrightarrow p = \frac{k}{n}.$$

Bei $p_0 = \frac{k}{n}$ hat L‘ einen Vorzeichenwechsel von + nach -, also liegt ein Maximum vor!

Das unserem Vorgehen zugrundeliegende allgemeine Prinzip beim Schätzen des (unbekannten) wahren Werts eines Parameters λ in einer (vom Typ her bekannten) Verteilung P_x lautet also („Maximum-Likelihood-Prinzip“): Bei der Durchführung des zugehörigen Zufallsexperiments sei das Ergebnis ω aufgetreten.

Die Wahrscheinlichkeit $P_x(\omega)$ hängt von λ ab: $P_x(\omega) = L(\lambda)$ (L heißt („Likelihood-Funktion“). Dann schätzen wir den Wert λ_0, für den $L(\lambda)$ maximal wird, als den wahren Wert (sofern dieser eindeutig existiert).

Diesem Prinzip liegt offenbar die naheliegende Annahme zugrunde, dass unter allen Möglichkeiten diejenige die beste ist, die am plausibelsten ist, d. h. die das aufgetretene Ergebnis am besten erklärt.

3.2.2 Intervallschätzen

Beim Intervallschätzen fragt man nach einem Konfidenzintervall für die unbekannte Wahrscheinlichkeit eines Zufallsexperiments, d. h. Werte für p, die mit einem vorliegenden Ergebnis „verträglich“ sind. Wir betrachten hierzu ein Beispiel:

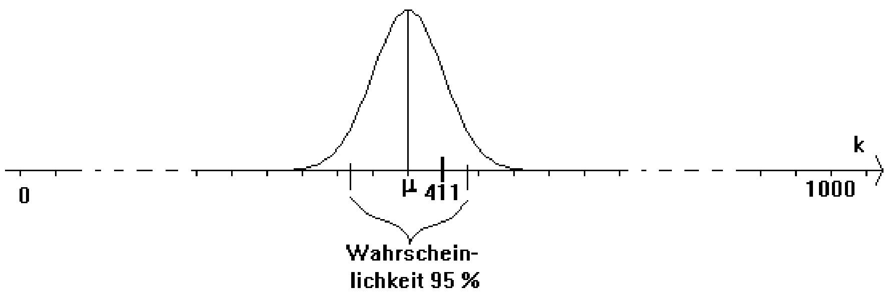
Wahlprognose:
Durch eine Umfrage unter 1000 zufällig ausgewählten Wahlberechtigten soll der Anteil von Partei Q ermittelt werden. Bei dieser Umfrage sprechen sich 411 Personen für Q aus. Wie groß ist wohl der tatsächliche Wähleranteil von Q?

Sinnvolle Modellannahme (falls die Gesamtzahl aller Wahlberechtigten sehr groß gegenüber 1000 ist): Diese Befragungsaktion stellt eine Bernoulli-Kette der Länge 1000 mit unbekanntem Parameter $p_* = P(Q)$ dar.

Wie in 3.2.1 können wir

$$p_* = \frac{411}{1000} = 41{,}1\%$$

schätzen. Nun will man aber genauer wissen, mit welchen Werten p für die unbekannte Wahrscheinlichkeit dieses Ergebnis „411 von 1000" verträglich ist und mit welchen nicht. Wir quantifizieren das, indem wir eine „Sicherheitsmarge" $\gamma_0 = 95\ \%$ festlegen und alle Zahlen p $\in [0; 1]$ suchen derart, dass die Zahl 411 in der 95 %-Umgebung des Erwartungswertes $\mu = np = 1000p$ liegt.



Wir wissen, σ-Regeln auf S.105, dass für die Binomialverteilung für genügend großes n gilt

$$P(\,|X-\mu| \le 2\sigma) \approx 95\%$$

Gesucht sind somit alle p mit

$$|411-\mu| \le 2\sigma\ ,$$

wobei $\mu = 1000$ p und $\sigma = \sqrt{1000 \cdot p(1-p)}$ gelten.

Die Berechnung dieser p ergibt

$$\begin{aligned} & |411-1000\cdot p| \le 2\sqrt{1000\cdot p(1-p)} \\ \Leftrightarrow\quad & (411-1000\cdot p)^2 \le 2^2 \cdot 1000 \cdot p(1-p) \end{aligned}$$

Hieraus folgt (leicht gerundet)

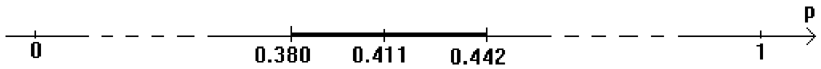
$$p^2 - 0.8227 \cdot p + 0{,}1682 \le 0\ .$$

Die Lösungen der zugehörigen quadratischen Gleichung sind

$$p_1 \approx 0.380 \; , \; p_2 \approx 0.442 \; ,$$

so dass wir folgendes Ergebnis erhalten:

$p \in [0.381; 0.442]$ (und genau diese p) sind auf dem 95 %-Niveau mit dem Ergebnis 411 verträglich.



Dabei liegt, wie intuitiv erwartet, 0.411 etwa in der Mitte dieses Intervalls. Man nennt es das „95%-Konfidenzintervall" für $p_* = P(Q)$ zum Wert 411. Dieses Intervall hat die Breite 0.062, d. h. wir haben etwa 6 % Streubreite.

Man sagt kurz – und äußerst missverständlich - : „Der tatsächliche Wert $p_* = P(Q)$ liegt mit 95%iger Wahrscheinlichkeit in diesem Intervall". Richtig muss man sagen (s. o.): Für die Zahlen p aus diesem Intervall (und genau für diese p) gilt: Ist p der tatsächliche Wert der Wahrscheinlichkeit P(Q), so liegt das aufgetretene Ergebnis 411 in einem (symmetrischen) Intervall um den erwarteten Wert 1000p herum, in das solche Ergebnisse mit Wahrscheinlichkeit 95 % fallen; anders ausgedrückt: Ist p der tatsächliche Wähleranteil von Q, so fallen 95 % sämtlicher solcher Umfrage-Ergebnisse in ein Intervall um 1000p herum, das 411 enthält.

<u>Definition:</u>
Sei γ_0 mit $0 < \gamma_0 < 1$ gegeben (im Beispiel oben war $\gamma_o = 95\%$). Es sei der (unbekannte) wahre Wert λ_0 eines Parameters λ in einer (vom Typ her bekannten) Wahrscheinlichkeitsverteilung P_x – mit jeweiligem Erwartungswert $\mu_\lambda = E(X)$ – zu schätzen. Bei der Durchführung des zugehörigen Zufallsexperiments sei das Ergebnis ω aufgetreten.

(a) Das wie folgt gebildete Intervall I heißt γ_0 –<u>Konfidenzintervall</u> für λ_0 und ω : λ liegt in I genau dann, wenn ω zu der symmetrisch zu μ_λ gelegenen Menge M mit $P_x(M) = \gamma_0$ gehört.
(b) γ_0 heißt das <u>Konfidenzniveau</u> dieser <u>Intervallschätzung</u>.

Es gibt, wie man sofort überlegt, einen einfachen und erhellenden Zusammenhang zwischen Schätzen und Testen:

Im γ_0-Konfidenzintervall zum aufgetretenen Ergebnis ω liegen genau die λ, für die beim Testen von

H_0: „Der wahre Wert des Parameters ist λ" (kurz: „$\lambda_0 = \lambda$")

(mit demselben Zufallsexperiment) das Ergebnis ω nicht in den (symmetrisch zu μ konstruierten) Verwerfungsbereich V zum Signifikanzniveau $\alpha_0 = 1 - \gamma_0$ fällt (d. h. in Definition der obigen ist $M = \overline{V}$).
Im obigen Beispiel gilt: Für genau die $p \in [0.380, 0.442]$ liegt 411 nicht im Verwerfungsbereich des Tests zu H_0: „$P(Q) = p$" mit Signifikanzniveau 5%.