10. Statistik, Zufall und Wahrscheinlichkeit

„Statistik ist: Wenn der Jäger am Hasen einmal links und einmal rechts vorbeischießt, dann ist der Hase im Durchschnitt tot.“

"Traue keiner Statistik, die du nicht selber gefälscht hast." [Winston Churchill]

10.1. Überblick

[Lit.: de.wikipedia.org, "Statistik"]

Historisch: Statistik = (vergleichende) Staatsbeschreibung (!), ital. *statista* = Staatsmann. Der Begriff wurde um 1749 von G. Achenwall geprägt.

Heute:

- **beschreibende (deskriptive) Statistik**: allgemeine Daten (nicht nur solche von Staaten!) verdichten zu Tabellen, graphischen Darstellungen oder Kennzahlen, Klasseneinteilung, Cluster finden
- **Wahrscheinlichkeitstheorie**: Kombinatorik, Wahrscheinlichkeitsräume, Ereignis, bedingte Wahrscheinlichkeit (Bayes), Zufallsvariable: diskret, stetig, Erwartungswert, Varianz, wichtige Verteilungen (binomial, normal, χ^2)
- **schließende (induktive) Statistik**: Schluss vom Besonderen auf das Allgemeine, von der Stichprobe auf die Gesamtheit: Parameterschätzung, Hypothesentests

STOCHASTIK

(Stochastik = „die Kunst des Vermutens“)

10.1.1. Warum InformatikerInnen Statistik brauchen

Statistik hat viel mit Daten und deren Verarbeitung zu tun, und damit ist der Bezug zur Informatik (= Datenverarbeitung) schon mehr als klar:

- Viele Aspekte der deskriptiven Statistik können wir hier nur anreißen, hier gibt es noch wesentlich mehr zu entdecken, wenn Sie später Vertiefungen in den Richtungen **Data Mining** und/oder **Visualisierung** von Daten studieren. Datenanalyse und Datenaufbereitung spielt eine wesentliche Rolle in vielen Informationsmanagementsystemen (=Anwendungsfeld für Informatiker in der betrieblichen Praxis, Stichwort Business Intelligence). Die Beschreibende Statistik (Kap. 10.2) legt hierfür die Grundlagen.
- Die **Kombinatorik** (Kap. 10.3.2) ist die "Kunst des Zählens". Sie bildet die Grundlage für viele Zufallsprozesse, und Informatiker brauchen sie, um sich einen Überblick über die Komplexität von Problemen zu verschaffen.
  Beispiele: Wie viele Möglichkeiten gibt es beim n-Städte-TSP (Kap. 9.4.2)?
  Wie viele Passwörter der Länge 5 enthalten "AA"?
  Wie viele Sudokus gibt es?
- Das Theorem von Bayes (bedingte Wahrscheinlichkeit) ist die Grundlage für **Klassifikation**. Beispielsweise können Sie damit einen **Spam-Filter** bauen, der anhand verschiedener Merkmale die Wahrscheinlichkeit für Spam bewertet.
- Bei jeder **Qualitätskontrolle** müssen Sie Stichproben bewerten und danach Entscheidungen fällen. Hier spielen **Zufallsvariablen** (Kap. 10.3.3) und **Normalverteilung** (Kap. 10.3.4) eine große Rolle.
- Bei den meisten Entscheidungen müssen Sie verschiedene Unwägbarkeiten ins Kalkül ziehen. Hier spielen **Zufallsvariablen** (Kap. 10.3.3) eine große Rolle >> **Risikominimierung**.

10.2. Beschreibende Statistik

[Teschl, Bd. 2, S. 199-224 + Stingl03, S. 581-598]

10.2.1. Merkmale und Merkmalstypen

-- dieses Kapitel im Selbststudium --

Die in der beschreibenden Statistik entwickelten, recht anschaulichen Begriffe spielen "Pate" für die abstrakteren Begrifflichkeiten der Wahrscheinlichkeitsrechnung.

Die beschreibende Statistik befasst sich mit der Darstellung von Daten.

Nehmen wir gleich ein konkretes Beispiel und betrachten wir Daten über die Fußballbundesliga. Die Rohdaten einer Spielzeit sehen z.B. wie folgt aus

Tabelle 10-1

Datum	Mannschaft		Tore		Zuschauer
	Heim	Gast	Heim	Gast	
01. März	Vfl Bochum	BVB	3	1	44.000
07. März	FC Bayern	FC Schalke	0	5	66.000
...	...	...	...	...	...

Im Laufe einer Spielzeit kommen hier eine ganze Menge Daten zusammen, und Aufgabe der beschreibenden Statistik ist es, durch geeignete Methoden einen guten Überblick herzustellen. Aussagekräftiger als die "nackte" Tabelle sind zum Beispiel: (a) Ranglisten, (b) (kumulierte) Tordifferenzen, (c) durchschnittliche Zuschauerzahlen usw.

Es ist zu unterscheiden zwischen den **Merkmalen** (z. B. Mannschaft, Spieltag, Tordifferenz) und den **Ausprägungen**, die diese Merkmale annehmen können (z.B. "VFL Bochum", "FC Bayern", ... für Merkmal Mannschaft)

		ist analog zu
Mathematische Analogie:	Merkmal	Funktion f
	Ausprägung	Funktionswert $f(x)$

Für die beschreibende Statistik sind verschiedene Merkmalstypen zu unterscheiden:

Def D 10-1 Merkmalstypen

Ein Merkmal heißt **qualitativ** oder **nominal**, wenn sich seine Ausprägungen durch Worte (*Nomen*) beschreiben lassen.

Bei einem **Rangmerkmal** lassen sich die Merkmale in eine lineare Ordnung bringen.

Ein Merkmal heißt (metrisch-)**quantitativ**, wenn sich die Ausprägungen durch Zahlen erfassen lassen, mit den für Zahlen üblichen Abstandsbegriffen ("liegt nahe bei", "ist größer als" usw.).

Ein quantitatives Merkmal heißt **diskret**, wenn die Ausprägungen deutlich voneinander abgrenzbar sind. Es heißt **stetig** (**kontinuierlich**), wenn innerhalb von bestimmten Intervallen prinzipiell alle Werte als Ausprägung auftreten können.

Anmerkungen:

- Der Begriff "diskret" wird oft mit "ganzzahlig" gleichgesetzt, was zwar in der Praxis häufig der Fall ist, aber keinesfalls notwendigerweise so sein muss.
- Ein quantitatives Merkmal, das nur abzählbar viele Werte annimmt, ist immer diskret. Auch wenn die Ausprägungen "krumme", z.B. irrationale Zahlen wie π, 2π, 3π, ... sind.

- Jedes quantitative Merkmal besitzt eine lineare Ordnung.
- Jedes quantitative Merkmal und jedes Rangmerkmal ist auch qualitativ.
- **Tabelle 10-2**

Typ	**qualitativ = nominal**	**Rangmerkmal**	**quantitatives Merkmal**	
Wertemenge	diskret	diskret	diskret	stetig
Skala	Nominalskala	Ordinalskala	metrische Skala	
Beispiel	Farbe	Tabellenplatz	RAM in kByte	Temperatur
	rot, grün, blau, ...	1., 2., 3., ...	44.512, 32.128, 16.000, 0, ...	0.51⁰C, 27.36⁰C, ...
Ordnung?	nein	ja	ja	
Summen- und ∅-Werte?	nein	fragwürdig (!!) [1]	ja	

[1] Wieso ist bei Rangmerkmalen die Summen- und Durchschnittsbildung zumindest fragwürdig? – Weil der Rang nichts über den tatsächlichen Abstand aussagt, auch nichts über die involvierten absoluten Summen. Eine Saison mit Kopf-an-Kopf-Rennen und eine "Michael-Schumacher-Deklassierung" sehen in der Rangstatistik u.U. völlig gleich aus. Die Rangfolge der Wochenumsätze einer Filialkette ist u.U. wenig aussagekräftig, wenn die Woche vor Weihnachten 10x so hohe Umsätze hat.

Beispiele:

- Der Name der Vereine der Fußballbundesliga ist ein qualitatives Merkmal
- Der Tabellenrang ist ein Rangmerkmal über den Vereinen der Liga
- Die Zuschauerzahl ist ein quantitaiv-diskretes Merkmal (ganzzahlige Werte), die Temperatur auf dem Rasen ein quantitativ-stetiges Merkmal des jeweiligen Spiels.
- Die Dateigröße in kByte auf der Festplatte meines Laptops ist auch ein diskretes Merkmal, auch wenn es in der Regel nicht ganzzahlig sein wird (!)

10.2.2. Relative Häufigkeiten und Klasseneinteilung

Eine Stichprobe ist die wiederholte (N-malige) Messung eines Merkmals.

Stichproben lassen sich übersichtlich zusammenfassen, wenn man absolute Häufigkeiten n_i und relative Häufigkeiten h_i bildet:

Def D 10-2 Häufigkeiten

Sei X ein quantitativ-diskretes Merkmal mit den Ausprägungen $X_1, X_2, \dots X_m$. Sei $S = \{x_1, x_2, \dots, x_N\}$ eine Stichprobe vom Umfang N. Dann ist:

1. **Absolute Häufigkeit** n_i = Anzahl der x_k mit Ausprägung X_i.
2. **Relative Häufigkeit** $h_i = \frac{n_i}{N}$.

Tabelle 10-3

Merkmal

Ausprägungen

	Gruppengröße Urlaubsreisende		
Gruppe	Anzahl n_i (absolute Häufigkeit)	relative Häufigkeit $h_i = \frac{n_i}{N}$	
Alleinreisende	2	2/15 = 0.1333	
Zweiergruppe	5	5/15 = 0.3333	
Dreiergruppe	5	5/15 = 0.3333	
4er-Gruppe	2	2/15 = 0.1333	
5 und mehr	1	1/15 = 0.0666	
Summe	15 = N	1.00000	

Für die Häufigkeiten gelten folgende, unmittelbar einsichtige Beziehungen:

Satz S 10-1

$$n_1 + n_2 + \ldots + n_m = \sum_{j=1}^{m} n_j = N \qquad \textit{(Summe der Datensätze)}$$

$$h_1 + h_2 + \ldots + h_m = \sum_{j=1}^{m} h_j = 1$$

Ü

<u>Übung:</u> Gegeben sei ein Merkmal X, das die Ausprägungen X_i = 1,...,8 annehmen kann. In einer Stichprobe sind diese Ausprägungen mit folgenden absoluten Häufigkeiten vertreten:

x_i	1	2	3	4	5	6	7	8
n_i	20	25	10	2	8	5	0	30

Berechnen Sie h_i. Mit welcher Häufigkeit gilt $4 \le x_i < 7$? Mit welcher Häufigkeit gilt $2 < x_i \le 6$?

Wenn bei einem quantitativen Merkmal zu viele Ausprägungen im Datensatz vorliegen (dies wird regelmäßig bei quantitativ-stetigen Merkmalen der Fall sein, jeder Wert tritt in der Regel nur einmal auf), dann bringt eine direkte Häufigkeitsdarstellung wenig. Deshalb gruppiert man die Daten in einer **Klasseneinteilung**:

Def D 10-3 Klasseneinteilung

Sei X ein quantitatives Merkmal. Eine **Klasseneinteilung** von X genügt folgenden Anforderungen:

1. Die Klassen sind paarweise disjunkt.
2. Die Klassen stoßen lückenlos aneinander.
3. Die Vereinigung aller Klassen überdeckt jeden Merkmalswert.

Beispiel: Sei X ein Merkmal mit Werten zwischen 0.4 und 5.0. Dann ist

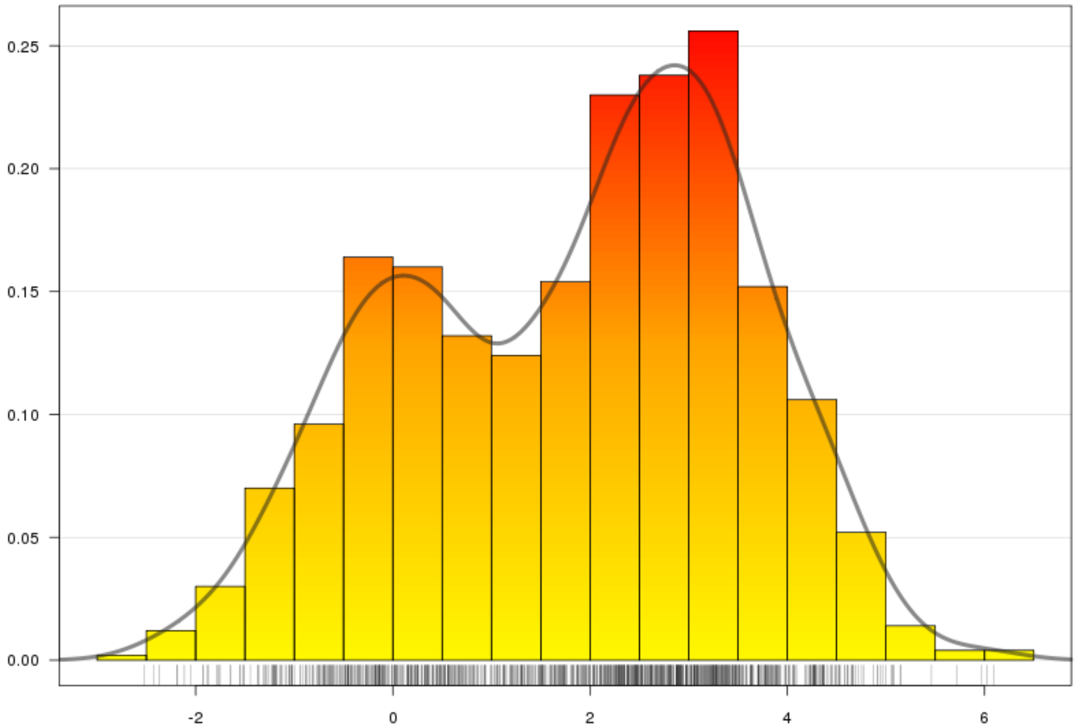
$$[0.0, 1.5[, \quad [1.5, 3.0[, \quad [3.0, 4.5[, \quad [4.5, 6.0[$$

eine gültige Klasseneinteilung. Im Allgemeinen können die Klassen unterschiedlich breit sein. können. (Wir behandeln hier aber nur gleichbreite Klassen. Bei unterschiedlich breiten Klassen müssen Histogramme etwas anders behandelt werden.)

Def D 10-4 Histogramm

Sei K eine Klasseneinteilung mit gleichbreiten Klassen. Die relative Häufigkeit „Wieviel Prozent der Daten fallen in Klasse K_i?" bezeichnet man mit h_i.

Ein **Histogramm** $f(x)$ besteht aus Rechtecken über den einzelnen Klassen, mit Breite Δx_i und Höhe h_i (oder auch n_i).



Mit dem Histogramm führt man ein quantitativ-kontinuierliches Merkmal zurück auf ein quantitativ-diskretes und gewinnt schnell einen Überblick, welche Klassen häufig / weniger häufig sind.

Wie viele Klassen? – Faustformel: Hat man N Werte in seinem Datensatz, so sollte man ca. $N^{1/2}$ Klassen wählen, dann kann im Mittel jede Klasse $N^{1/2}$ Daten enthalten.

Beispiel:

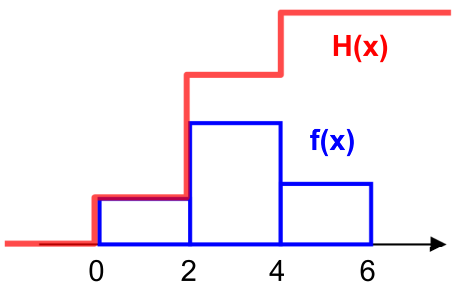
Gegeben seien die Daten

0.5	0.7	1.2	1.9	2.0	2.2	2.6	2.7
2.9	2.9	3.2	3.4	3.5	3.7	3.8	4.2
4.7	5.0	5.3	5.7				

Das Histogramm für die Klasseneinteilung [0,2[, [2,4[, [4,6[errechnet sich daraus wie folgt:

m=3 Klassen	n_i	h_i	x_i	

[0.0, 2.0[	4	0.20	0.0	
[2.0, 4.0[	11	0.55	2.0	
[4.0, 6.0[	5	0.25	4.0	
Summen (Kontrolle)	20	1.00		



(Zur Häufigkeitsverteilung H(x) siehe weiter unten Exkurs Häufigkeitsverteilung)

Ü

Übung: Gegeben sei eine Messreihe für Temperaturen T, die bei einem industriellen Prozess gemessen werden. Die Reihe liegt in geordneter Form vor:

T	-3	-1	-1	1	1	3	10	12
	12	14	18	19	19	20	25	27
	31	33	46	46	50	52	89	90
	90	101	110	110	124	134		

Berechnen und zeichnen Sie das Histogramm $f(x)$ nach Def D 10-4 für die Klasseneinteilung

[-10, 20[, [20, 50[, [50, 80[, [80, 110[, [110, 140[, [140, 170[

10.2.3. Kennzahlen einer Stichprobe

[Stingl04, S. 589-594]

Ein anderer Weg, eine Menge von Daten zu charakterisieren, besteht darin, (möglichst aussagekräftige) Kennzahlen zu ermitteln. Idealerweise spiegelt sich dann, wenn wir regelmäßig wiederkehrend bestimmte Daten erheben, eine interessierende Veränderung in der Datenzusammensetzung in einer "signifikanten" Veränderung der Kennzahl nieder.

Betrachten wir eine **aus Zahlen bestehende Stichprobe**.

Beispiel: Temperaturwerte werden an einer Messstation stündlich erhoben. Die 24 Messungen bilden eine Stichprobe. Die **mittlere Tagestemperatur** ist eine Kennzahl, die die Gesamtheit von jeweils 24 Messungen charakterisiert. Weitere Kennzahlen:

Def D 10-5 Mittelwert und Median

Der **arithmetische Mittelwert** $\overline{x}$ einer Stichprobe $X = \{x_1,\dots,x_n\}$ ist

$$\overline{x} = \frac{1}{n}\left(x_1 + \dots + x_n\right) = \frac{1}{n}\sum_{i=1}^{n} x_i \ .$$

Für den Median einer Stichprobe $X = \{x_1,\dots,x_n\}$ ordnet man die x_i zunächst der Größe nach: $x'_1 \le x'_2 \le \dots \le x'_n$. Der **Median** $m(X)$ ist der Wert, bei dem 50% der Werte "kleiner-gleich" sind und 50% "größer-gleich". Für ungerades n ist $m(X) = x'_{(n+1)/2}$.

Für gerades n ist $m(X) = \dfrac{x'_{n/2} + x'_{n/2+1}}{2}$.

Def D 10-6 p-Quantil

Das **p-Quantil q_p** einer Stichprobe ist der Wert, **unterhalb** dem genau der Anteil p aller Daten liegt.

Ist n·p nicht ganzzahlig, dann $q_p = x'_{\lceil n\cdot p \rceil}$.

Ist n·p ganzzahlig, dann $q_p = \dfrac{x'_{n\cdot p} + x'_{n\cdot p+1}}{2}$.

Anmerkungen:

- Der Median ist (etwas) aufwendiger zu berechnen als der Mittelwert, da zunächst die Werte sortiert werden müssen.
- Der Median ist aber auch "robuster": Ein einzelner Ausreißer verändert den Median kaum, den Mittelwert aber u.U. stark.
- Der Median ist der Spezialfall eines Quantils, nämlich das 0.5-Quantil.
- Das 0.25-Quantil nennt man auch (unteres) **Quartil**, da genau ein Viertel aller Daten unterhalb liegt.
- Das 0.1- und 0.9-Quantil nennt man auch **unteres** bzw. **oberes Dezil**.

Beispiel in Vorlesung.

Def D 10-7 Varianz und Standardabweichung

Die **(empirische) Varianz s^2** einer Stichprobe $X=\{x_1,...,x_n\}$ ist definiert als

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

mit dem Mittelwert $\bar{x}$ nach Def D 10-5. Die Größe $s = \sqrt{s^2}$ heißt **(empirische) Standardabweichung**. Je größer s oder s^2, desto mehr streuen die Werte der Stichprobe.

Def D 10-8 Interquartilsabstand IQR (inter-quartile range)

Ein alternatives Maß für die Streuung einer Stichprobe ist der **Interquartilsabstand**

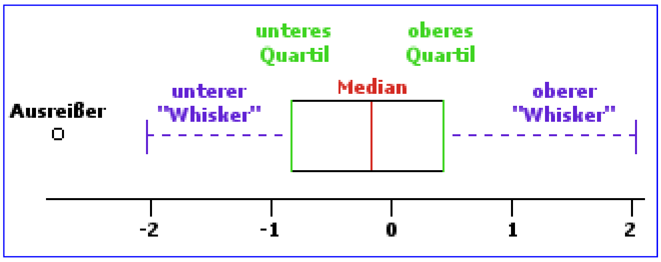
$$\mathbf{IQR = q_{0.75} - q_{0.25}}$$

(zu Quartil vgl. Def D 10-5). Im Intervall [$q_{0.25}$, $q_{0.75}$] liegen genau 50% aller Daten. Je größer der IQR, desto mehr streuen die Werte der Stichprobe.

Vorteil IQR: robuster gegen Ausreisser.

Warum Faktor $\frac{1}{n-1}$ bei Varianz? – Die Erklärung ist etwas komplexer, wer es genau wissen will, kann es in empirVari.pdf nachlesen (aber erst, wenn wir Zufallsvariablen gehabt haben).

10.2.4. Boxplot: Visualisierung einer Stichprobe

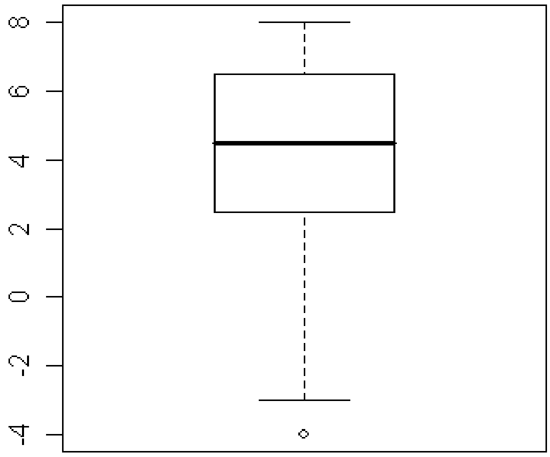


Der Boxplot ist eine kompakte Methode, die wesentlichen Parameter einer Datenreihe in einem Bild zu visualisieren:

- Das Rechteck wird durch das untere Quartil und obere Quartil begrenzt. (Das untere Quartil ist die Linie, unterhalb der 25% der Daten liegen, analog für oberes Quartil.)
- Das Rechteck wird durch den Median (s. Def D 10-5) geteilt
- Für die Whisker (engl. „Schnurrhaare“) gibt es verschiedene Konventionen. Eine Konvention ist: Die Länge jedes Whiskers beträgt maximal das 1.5-fache des Interquartilsabstandes IQR und wird immer durch einen Punkt aus den Daten bestimmt.
- Punkte, die außerhalb der Whisker liegen, werden einzeln als Ausreißer dargestellt.

Beispiel: Man zeichne den Boxplot für folgende Stichprobe:

-4	-3	2	3	3	4	5	5	6	7	7	8



Ü

Übung: Zeichnen Sie den Boxplot für folgende Stichprobe:

4	5	5	10	10	11	11	12
12	13	13	14	15	15	25	27

10.3. Wahrscheinlichkeitstheorie

10.3.1. Der Wahrscheinlichkeitsbegriff

[Teschl, Bd. 2, S. 225ff. + Stingl03, S. 606ff.]

Bei der Entwicklung der Wahrscheinlichkeitstheorie hat man sich von den Methoden der (empirischen) Statistik und Stichprobe leiten lassen und man hat diese Begriffe auf ein *theoretisches Modell* übertragen: Ein Versuch kann beliebig oft wiederholt werden (zumindest im Prinzip), aber wegen unkontrollierbarer Einflüsse kann man den Ausgang nicht präzise vorhersagen.

Begriffliche Gegenüberstellung:

Tabelle 10-4

Beschreibende Statistik	**Wahrscheinlichkeitstheorie**	**Kap.**
Menge d. Ausprägungen	Ergebnismenge	10.3.1
Ausprägung	Versuchsausgang	10.3.1
rel. Häufigkeit	Wahrscheinlichkeit	10.3.1
Häufigkeits-Verteilungsfkt.	Verteilungsfunktion	10.3.3
Histogramm	Dichtefunktion	10.3.3
Merkmal (quantitativ)	Zufallsvariable	10.3.3

HINWEIS: Es lohnt sich, im Laufe von Kap. 10 immer mal wieder zu dieser Tabelle zurückzukehren und sich selbst zu kontrollieren, ob man mit den Begriffen etwas anfangen kann und ob man den jeweiligen Bezug sieht.

Def D 10-9 Zufallsexperiment, Ergebnismenge, Ereignismenge

Unter einem Zufallsexperiment versteht man einen beliebig oft unter gleichen Bedingungen wiederholbaren Versuch, dessen Ausgang wegen unkontrollierbarer Einflüsse dem Zufall unterworfen ist. Man definiert:

Ergebnismenge Ω: die Menge der möglichen Versuchsausgänge,

Ereignismenge $\mathcal{A}$: die Menge aller Teilmengen von Ω.

Die leere Menge { } und Ω selbst sind ebenfalls Teilmengen von Ω und damit Elemente von $\mathcal{A}$. Es heißt { } das **unmögliche Ereignis** und Ω das **sichere Ereignis**.

Ist $A \in \mathcal{A}$ ein Ereignis, so heißt $\overline{A}$ = "A tritt nicht ein" das **zu A komplementäre Ereignis**.

Die Elemente von Ω sind ebenfalls Elemente von $\mathcal{A}$ und heißen **Elementarereignisse**.

Beispiele, Anmerkungen:

1. Beim Würfeln ist Ω = {1, 2, 3, 4, 5, 6}. Mögliche Ereignisse sind
   A = {6} "Würfeln einer Sechs" (Elementarereignis)
   B = {1,3,5} "ungerade Augenzahl"
   C = {1,2} "weniger als 3"
2. Das zu C komplementäre Ereignis ist $\overline{C} = \{3,4,5,6\}$.
3. { } und Ω sind zueinander komplementäre Ereignisse.

Def D 10-10 Wahrscheinlichkeitsmaß

Eine Funktion P: $\mathcal{A} \rightarrow \mathbf{R}$, die jedem Ereignis $A \in \mathcal{A}$ eine reelle Zahl zuordnet, heißt **Wahrscheinlichkeitsmaß** P(A) (engl. "probability"), wenn gilt:

1. $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. $P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$ falls sich A_1, A_2, ... paarweise ausschließen.
   (Solche Ereignisse A_1, A_2,... nennt man auch **unvereinbar**.)

Aus diesen sog. *Wahrscheinlichkeitsaxiomen* kann man weitere Eigenschaften das Wahrscheinlichkeitsmaßes ableiten:

Satz S 10-2 Konsequenzen

1. $P(\overline{A}) = 1 - P(A)$
2. $P(\{\,\}) = 0$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Beispiel: Beim Würfelexperiment ist die Ergebnismenge {1,2,3,4,5,6}. Welche Wahrscheinlichkeit müssen wir bei "fairem" Würfel den einzelnen Elementarereignissen zuschreiben?

Dies folgt direkt aus den Axiomen in Def D 10-10:

$$1 = P(\Omega) = P(\{1\} \cup \{2\} \cup ... \cup \{6\}) = P(\{1\}) + ... + P(\{6\})$$

weil die einzelnen Elementarereignisse paarweise unvereinbar sind. Bei einem "fairen" Würfel sind alle Elementarereignisse gleichwahrscheinlich, also $P(\{i\})=1/6$.

Die Wahrscheinlichkeit für Ereignis A = "Augenzahl < 3" ist $P(A)=2/6$.

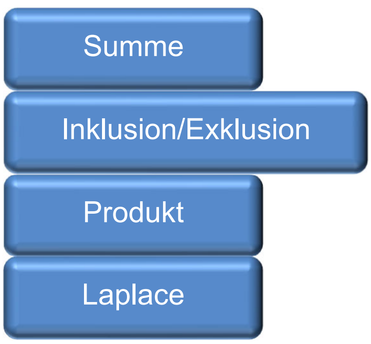
Die Wahrscheinlichkeit für $\overline{A}$="Augenzahl nicht < 3" ist $P(\overline{A}) = 1 - P(A) = 4/6$.

10.3.2. Kombinatorik

[Teschl, Bd. 1, S. 171-186]

Kombinatorik-Trainer: www.gm.fh-koeln.de/kombinatorik, NEU: Der Kombinatorik-Trainer wurde erweitert um Inklusions-Exklusions-Prinzip (s.u.), Binomial- und hypergeometrische Verteilung (s. Kap. 10.3.4)

Einfache Zählprinzipien



Summenregel und Produktregel

Summenregel: Hat man n Objekte mit Eigenschaft $a \in A$, m Objekte mit Eigenschaft $b \in B$, wobei sich a und b gegenseitig ausschließen (A und B sind also disjunkte Mengen), dann gibt es n+m Möglichkeiten, ein Objekt mit Eigenschaft „a oder b“ auszuwählen.

Beispiel: Mietwagenfirma, 3 Kleinwagen, 7 Mittelklassewagen → 10 Möglichkeiten.

Mengenschreibweise: Für disjunkte Mengen A und B gilt: $|A \cup B| = |A| + |B|$

NEU: Inklusions-Exklusions-Prinzip: Was ist, wenn A und B nicht disjunkt sind? – Dann ist die Summenregel zu erweitern, denn mit |A| + |B| würde man alle Elemente in der Schnittmenge $A \cap B$ doppelt zählen. Die Lösung ist, alle doppelt vorkommenden Elemente *einmal* abzuziehen, also

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Das ist die Inklusions-Exklusions-Formel.

Dies lässt sich auch auf mehr als zwei Mengen verallgemeinern mit der Regel: „Addiere die Anzahlen in den Mengen und ziehe die N-mal vorkommenden Objekte jeweils (N-1)mal ab.“

Das bedeutet für 3 Mengen A, B, C:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

Beispiele: siehe Übungen und www.gm.fh-koeln.de/kombinatorik.

Übung Inklusion/Exklusion: In einer Stadt gibt es 1000 Leute, die Skifahren oder Paragliding als Hobby haben. 900 sind Skifahrer. 100 betreiben sowohl Skifahren als auch Paragliding. Wie viele Paraglider wohnen in der Stadt?

Produktregel: Kann man eine Aufgabe in 2 Teilschritte zerlegen und hat man n Möglichkeiten für Schritt 1 und m Möglichkeiten für Schritt 2, dann gibt es $n \cdot m$ Möglichkeiten für die gesamte Aufgabe.

Beispiel: 3 Routen von GM nach K, 4 Routen von K nach AC → 12 Routen von GM nach AC

Mengenschreibweise: Die Produktmenge $A \otimes B$ hat $|A \otimes B| = |A| \cdot |B|$ Elemente.

Bsp: A = {a,b}, B = {1,2,3} → $A \otimes B$ = { (a,1), (a,2), (a,3),
(b,1), (b,2), (b,3) } → $|A \otimes B| = 2 \cdot 3 = 6$

Laplacescher Spezialfall

Zahlreiche Zufallsexperimente kann man auf den sogenannten *Laplaceschen Spezialfall* zurückführen:

> *Es gibt einen endlichen Ergebnisraum Ω, dessen Elementarereignisse (s.* Def D 10-9*) alle* ***gleichwahrscheinlich*** *sind.*

Satz S 10-3 Laplacesche Wahrscheinlichkeiten

Im *Laplaceschen Spezialfall* gilt:

$$P(A) = \frac{Anzahl\ der\ zu\ A\ gehörigen\ Versuchsausgänge}{Anzahl\ der\ überhaupt\ möglichen\ Versuchsausgänge} = \frac{günstige\ Fälle}{mögliche\ Fälle}$$

Das Berechnen von Wahrscheinlichkeiten ist also auf das Zählen von Ereignissen zurückgeführt. Die Kombinatorik als die "**Kunst des Zählens**" liefert hierzu die Grundlage.

Beispiel Würfelsumme in Übung

Zurück zum Laplaceschen Spezialfall:
Der Prototyp für solche Laplaceschen Spezialfälle ist das **Urnenexperiment**: Viele Dinge des realen Lebens lassen sich, wenn es nur auf die Wahrscheinlichkeiten ankommt, gedanklich auf eine Urne mit **verschieden bezeichneten** Kugeln zurückführen (denken Sie an die Ziehung der Lottozahlen), aus der mit oder ohne Zurücklegen (ungeordnete) Teilmengen oder (geordnete) Listen gezogen werden.

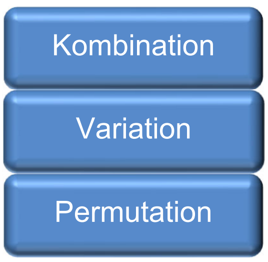
Binomialkoeffizienten (Wdh.)

Def D10-11: Für n, k $\in \mathbb{N}_0$ mit $k \le n$ definiert man:

- die **Fakultät** $n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$ für n>0 sowie 0! = 1
- den **Binomialkoeffizienten** $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1}$

Die letzte Umformung gilt nur für k>0.

Weitere Zählprinzipien



Kombination, Variation, Permutation

Kombination: Auswahl von k Elementen aus einer n-elementigen Menge, Reihenfolge egal. Implementierung: Menge, Set. {1,2,5} und {5,1,2} sind dieselben Mengen. Beispiel: Lotto.

Variation: Auswahl von k Elementen aus einer n-elementigen Menge, Reihenfolge wichtig. Implementierung: Liste. [1,2,5] und [5,1,2] sind verschiedene Listen. Beispiel: Pferderennen.

Permutation: Variation mit k=n. Beispiel: Mischung eines Skatspiels.

Satz S10-4 Stichproben aus Mengen
Zieht man aus einer n-elementigen Menge eine k-elementige Stichprobe (geordnet oder ungeordnet), so gibt es dafür, je nachdem ob dies mit/ohne Zurücklegen geschieht, folgende Anzahl von Möglichkeiten:

	Reihenfolge wichtig (geordnet) (**Variation**)	Reihenfolge egal (ungeordnet) (**Kombination**)
Ziehen mit Zurücklegen (**ZmZ**)	n^k	$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$
Ziehen ohne Zurücklegen (**ZoZ**)	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$

Beweis von Satz S10-4 in Vorlesung!

Aktivierung: Können Sie **Anwendungsfälle** für jeden der 4 Bereiche nennen?

Was sind **Anwendungsbeispiele** für jeden der 4 Fälle?

	geordnet (**Variation**)	ungeordnet (**Kombination**)
Ziehen mit Zurücklegen	Wörter aus Alphabet	geheime Wahlausgänge
Ziehen ohne Zurücklegen	Rangfolgen (Pferdewette)	Lotto, k-Teilmengen aus n-Menge, Positionierungen

Tablet: Systematik zur Kombinatorik im Flussdiagramm

Beispiele und Übungen:

Beispiel 1: In einer Urne sind 10 weiße und 20 schwarze Kugeln. Wie groß ist die Wahrscheinlichkeit, in einer 4er-Ziehung ohne Zurücklegen 3 weiße und 1 schwarze zu ziehen?

Lösung: Kombination, ZoZ. Wir nummerieren alle Kugeln gedanklich durch, dann haben wir wieder lauter unterscheidbare Objekte und können Satz S10-4 anwenden.

Laplacescher Spezialfall: Mögliche Fälle: Es gibt $\binom{30}{4}$ Ziehungen überhaupt. Günstige Fälle: Dazu bilden wir zwei **Hilfsurnen** H1 und H2: H1 enthält nur 10 weiße und H2 nur 20 schwarze Kugeln. Jeder günstige Fall besteht aus 3 Ziehungen aus H1 und 1 aus H2, also $\binom{10}{3}\binom{20}{1}$ Möglichkeiten (Produktregel). Setzt man beides ins Verhältnis, so erhält man

$$\frac{\binom{10}{3}\binom{20}{1}}{\binom{30}{4}} = \frac{\frac{10\cdot 9\cdot 8}{1\cdot 2\cdot 3}\cdot 20}{\frac{30\cdot 29\cdot 28\cdot 27}{1\cdot 2\cdot 3\cdot 4}} = 8.757\%$$

Ü1. *(a) Wie viele Wörter mit 4 Buchstaben kann man aus dem Alphabet {a,b,...,z} von 26 Buchstaben bilden? (b) Wie wahrscheinlich ist es, dass ein zufällig gezogenes Wort nur aus den ersten 5 Buchstaben besteht?*

Ü2. *Bei einer Pferdewette sind bei einem Lauf mit 8 Pferden die Pferde zu erraten, die als Erster, Zweiter und Dritter durchs Ziel gehen. (a) Wieviel mögliche Wettausgänge gibt es? (b) Wie groß ist die Wahrscheinlichkeit, durch zufälliges Tippen zumindest den Ersten richtig zu raten?*

Anwendungsbeispiel: Binomischer Satz

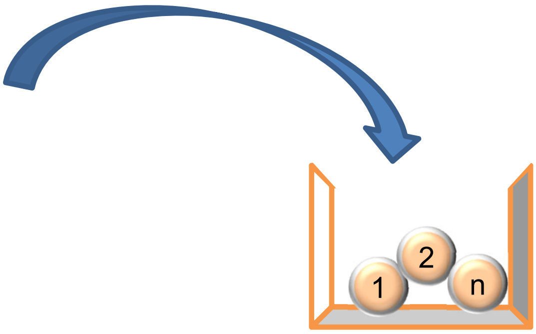
Satz S10-5 (Binomischer Satz): Für n $\in \mathbf{N}$ und a,b $\in \mathbf{R}$ gilt:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Beweis:

Wir schreiben das Produkt aus

$$(a+b)^n = \underbrace{(a+b)(a+b)\cdot\ldots\cdot(a+b)}_{n-\text{mal}}$$



Die Kugeln sind die **Positionsmenge** {1,2,...,n}. Für einen Term mit genau k a's: k-mal Ziehen aus Positionsmenge, ungeordnet → genau $\binom{n}{k}$ Möglichkeiten. Alle anderen Positionen werden mit b's besetzt → genau 1 Möglichkeit. ■

Die **Positionsmenge** ist ein wichtiges Hilfskonstrukt für viele Zählvorgänge.

Ü3. *Beim Lotto werden 6 aus 49 Zahlen gezogen. (a) Wie viele Möglichkeiten gibt es insgesamt? (b) Wie wahrscheinlich sind 4 Richtige?*

Ü4. *Wieviele Wörter der Länge 5 über dem Alphabet A={a,b,c} enthalten genau zwei a's? [Hinweis: Machen Sie's ähnlich wie beim Binomischen Satz!]*

Ü5. *Im Staate Mathelan wird der Präsident durch ein 60-köpfiges Gremium gewählt, 3 Präsidentschaftskandidaten stehen zur Auswahl. Die Wahl ist geheim, Enthaltungen sind nicht erlaubt, jeder hat genau eine Stimme. Wie viele verschiedene Wahlausgänge gibt es?*

Wichtige Anwendungen der Kombinatorik sind Qualitätsprüfungen durch Stichproben:

Beispiel: Bei einer Lieferung von 100 Chips dürfen nur weniger als 15% defekt sein. Zur Überprüfung wird eine Stichprobe von k=4 Chips entnommen und vermessen. Wie groß ist die Wahrscheinlichkeit, dass eine "schlechte" Lieferung mit 15% Ausschuss akzeptiert wird, obwohl in der Stichprobe kein fehlerhafter Chip war? Um wieviel sinkt diese Wahrscheinlichkeit, wenn man auf k=6 erhöht?

Lösung: Es handelt sich um Ziehen ohne Zurücklegen. Die Lieferung enthält 15 schlechte und 85 gute Chips. Die Wahrscheinlichkeit für Stichprobe mit nur guten Chips ist (Laplace):

$$P(k=4) = \frac{\binom{85}{4}}{\binom{100}{4}} = \frac{85 \cdot 84 \cdot 83 \cdot 82}{100 \cdot 99 \cdot 98 \cdot 97} = 51.6\%$$

$$P(k=6) = \frac{85 \cdot 84 \cdot 83 \cdot 82 \cdot 81 \cdot 80}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96 \cdot 95} = 36.6\%$$

Es gibt viele weitere Anwendungen, die wir z.T. in den Übungen besprechen:

1. Ab welcher Gruppengröße lohnt sich die Wette "Wetten, dass in dieser Gruppe von Personen mindestens zwei im gleichen Monat Geburtstag haben?"
2. bzw. "... am gleichen Tag ..."?
3. Dieses Problem hat eine sehr praktische Anwendung in der Informatik: Mit **Hashtabellen** ordnet man Objekten, die "von sich aus" keinen (kleinen) Index haben, einen solchen Index zu. Bsp.: Aus der 10-stelligen ISBN eines Buches bilden wir den Rest bei Division durch 101. Mögliche Hashwerte sind also 0,1,...,100. Wie wahrscheinlich ist eine *Kollision* in der Hashtabelle, d.h. das Ereignis, dass zwei Bücher auf denselben Hashwert abgebildet werden?

ZUM WEITEREN ÜBEN: Kombinatorik-Lernprogramm: www.gm.fh-koeln.de/kombinatorik

Lösung Ü5: Da die Wahl geheim ist, ist Reihenfolge irrelevant → Kombination. Weil jeder Wahlmann/jede Wahlfrau aus der gleichen 3er-Kandidatenliste wählen kann, ist es Ziehen mit Zurücklegen. Es gibt also nach Satz S10-4 mit n=3, k=60:

$$\binom{n+k-1}{n-1} = \binom{62}{2} = \frac{62 \cdot 61}{2 \cdot 1} = 1891 \text{ Wahlausgänge.}$$

10.3.3. Zufallsvariablen

[Teschl05, Bd. 2, S. 245-280 + Stingl, S. 619-624]

Motivation: In vielen praktischen Entscheidungssituationen hat man es mit Unwägbarkeiten zu tun: Eine Investition (z.B. in eine Startup-Firma) endet zu 20% in einem Desaster (alles Kapital verloren), zu 70% bei einer Rendite von +20% und zu 10% in einem märchenhaften Gewinn (Verdreifachung des eingesetzten Kapitals). Soll ich investieren oder nicht? Man hat es mit nur abzählbar vielen Kategorien zu tun.

Andere zufällige Dinge, wie die Verspätungszeit eines Zuges, die Füllmenge einer Flasche, die Schadenshöhe eines Versicherungsfalls können Werte aus einem ganzen Kontinuum annehmen (reelles Intervall).

Zufallsvariablen ermöglichen es, beide Bereiche **einheitlich**, d.h. mit einer Technik zu behandeln.

Zufallsvariablen sind ein wichtiges – eigentlich ***das*** wichtigste – Mittel der praktischen Statistik, denn mit Zufallsvariablen kann man solche Fragen ganz systematisch entscheiden!

Diskrete Zufallsvariablen

Kurz und knapp: Eine Zufallsvariable ist eine **Funktion**, die jeder Durchführung eines Versuchs eine reelle Zahl *und eine Wahrscheinlichkeit* zuordnet.

Genauer:

Def D10-12 Zufallsvariable

Unter einer **Zufallsvariablen** X versteht man eine Funktion $X\colon\ \Omega \to \mathbb{R}$, die jedem möglichen Ergebnis ω eines Zufallsexperimentes (s. **Def D 10-8**) eine reelle Zahl $X(\omega)$ zuordnet.

Wenn X nur abzählbar viele Werte x_m annehmen kann, spricht man von einer **diskreten Zufallsvariablen.**

Jedem Wert x_m der Zufallsvariable wird eine **Wahrscheinlichkeit** $p_m=P(X=x_m)$ zugeordnet.

Def D10-13 Verteilungsfunktion

Die Funktion $F\colon \mathbb{R} \to [0,1]$ mit $F(t) = P(X \le t)$ heißt **Verteilungsfunktion von** X.

F ist monoton wachsend.

Für diskrete Zufallsvariablen gilt einfach:

$$F(t) = \sum_{x_m \le t} P(X = x_m)$$

Zur Begriffsklärung:

- Wahrscheinlichkeitsverteilung $p_m\colon \Omega \to \mathbb{R}$, Summe ist 1 = 100%.
- (kumulative) Verteilungsfunktion $F\colon \mathbb{R} \to [0,1]$, strebt gegen 1 für $t \to \infty$.

Ohne den Begriff „kumulativ“ (= anhäufend, aufsummierend) verwechselt man die beiden Funktionen leicht, deshalb denken Sie am besten „kumulativ“ immer mit.
Im Englischen heißt es *cdf = cumulative distribution function*.

Beispiele und Anmerkungen:

- X = "Augensumme bei zwei Würfeln" ist eine diskrete Zufallsvariable. Das zugrundeliegende Zufallsexperiment: "Werfen zweier Würfel".

Tabelle 10-5 Augensumme zweier Würfel

Wert x_m von X	ω mit $X(\omega)=x_m$	$p_m = P(X=x_m)$	$F(x_m) = P(X \leq x_m)$
2	(1,1)	1/36	1/36
3	(1,2), (2,1)	2/36	3/36
4	(1,3), (2,2), (3,1)	3/36	6/36
...	...	...	...

<u>Übung:</u> Füllen Sie den Rest der Tabelle aus! Was ist dann $\sum_m x_m p_m$?

Beispiel „Urne“ in Vorlesung.

Def D10-14 Erwartungswert einer diskreten Zufallsvariablen

Für eine diskrete Zufallsvariable $X: \Omega \to \mathbb{R}$ gemäß Def D10-11 ist der Erwartungswert μ definiert durch:

$$\mu = E(X) = \sum_m x_m p_m$$

Der Erwartungswert gibt an, welcher Wert sich ergibt, wenn man X über sehr viele Zufallsexperimente mittelt.

<u>Beispiel:</u>

- Der Erwartungswert für die Augensumme bei zwei Würfeln ist (s. Tabelle 10-5):

$$\mu = E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \ldots + 12 \cdot \frac{1}{36}$$
$$= (2+12) \cdot \frac{1}{36} + (3+11) \cdot \frac{2}{36} + \ldots + (6+8) \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} = 7$$

Erwartungswerte spielen eine große Rolle bei der Bewertung von Situationen mit Unsicherheit und der <u>rationalen Entscheidung unter Unsicherheit</u>, wie nachfolgende Übungen zeigen:

<u>Übungen:</u>

Ü1. *Bewerten Sie, ob es sich lohnt, an folgendem Spiel teilzunehmen, indem Sie den Erwartungswert für X = "Gewinn – Einsatz" (oder für X‘ = "Gewinn") ausrechnen: Beim Würfeln mit zwei Würfeln erhält man einen Gewinn von 20€ für "Augensumme 12" und 5€ für "Augensumme 11", ansonsten geht man leer aus. Pro Spiel ist ein Einsatz von 1€*

zu zahlen.

Ü2. *Beim Würfeln mit 2 Würfeln sei d = Augendifferenz "groß – klein". Für einen Einsatz von 2€ kann man an folgendem Gewinnspiel teilnehmen*:

d	Gewinn
5	30 €
4	10 €

Spielen Sie?

Ü3. *Lösen Sie die Aufgabe aus der Motivationseinleitung (Invest in Startup-Firma): Eine Investition (z.B. in eine Startup-Firma) endet zu 20% in einem Desaster (alles Kapital verloren), zu 70% bei einer Rendite von +20% und zu 10% in einem märchenhaften Gewinn (Verdreifachung des eingesetzten Kapitals). Soll ich investieren, d.h. ist die Rendite besser als Sparbuch (3%), oder nicht?*

Folgende Gesetzmäßigkeiten und Definitionen können gleichermaßen für diskrete und stetige Zufallsvariablen (s.u) formuliert werden:

Satz S10-6 Linearität des Erwartungswertes

Für Zufallsvariablen X,Y und reelle Zahlen a,b∈**R** gilt der wichtige Satz

$$E(aX+b)=aE(X)+b \quad \text{und} \quad E(X+Y)=E(X)+E(Y)$$

Beispiel für diskret:

- Für einen einzelnen Würfel X gilt offensichtlich E(X)=3.5, wenn wir die Wahrscheinlichkeit $w_m = \frac{1}{6}$ für jedes $x_m \in \{1,2,3,4,5,6\}$ einsetzen. Nach Satz S10-6 gilt dann für zwei Würfel X und Y:

  $$E(X+Y) = E(X) + E(Y) = 3.5 + 3.5 = 7$$

  in Übereinstimmung mit dem Beispiel nach Def D10-14. Der Rechenweg hier ist viel einfacher (!) Und für 10 Würfel gilt entsprechend

  $$E(X_1 + \cdots + X_{10}) = E(X_1) + \cdots + E(X_{10}) = 10 \cdot 3.5 = 35$$

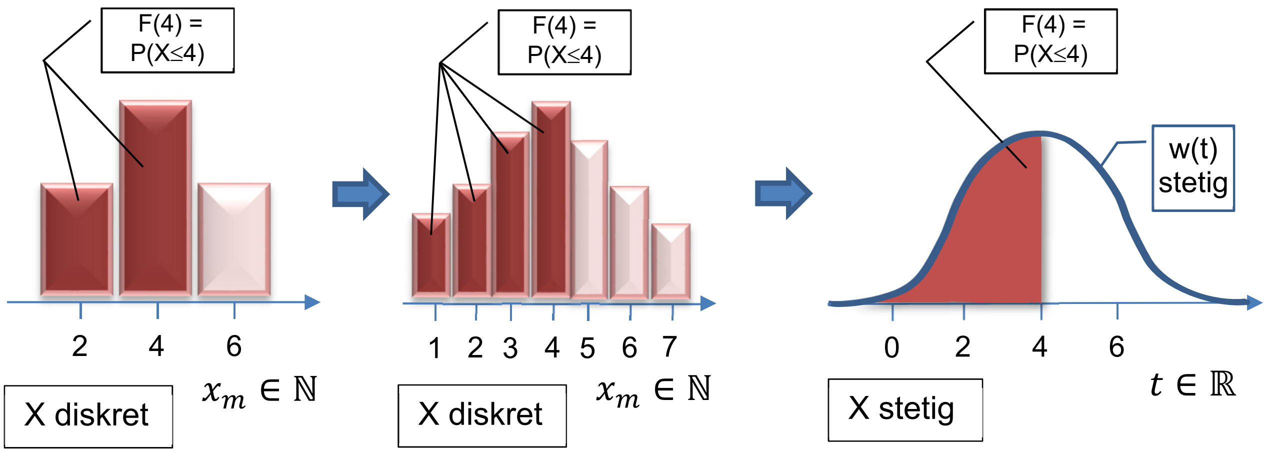


Abbildung 1: Übergang von diskreten zu stetigen Zufallsvariablen

Stetige Zufallsvariablen

Def D10-15 Stetige Zufallsvariable, Verteilungsfunktion

Unter einer stetigen **Zufallsvariablen** X versteht man eine Zufallsvariable (s. **Def D10-12**), bei der X beliebige Werte aus einem reellen Intervall annehmen kann.

Beispiele:

- X = "Lebensdauer einer Glühbirne in h" ist eine stetige Zufallsvariable.
- X = "Stellung des Stundenzeigers einer Uhr". Das Zufallsexperiment ist die zufällige Auswahl eines Zeitpunktes zum Uhr-Ablesen. Ereignismenge Ω ist die Menge der möglichen Zeigerstellungen und $X\colon \Omega \to]0,12]$ ist eine reelle Zufallsvariable.

Anmerkungen:

- Es macht keinen Sinn, bei einer stetigen Zufallsvariablen nach der Wahrscheinlichkeit $P(X=t)$ zu fragen, denn die ist 0. (Der Stundenzeiger steht praktisch nie auf "genau 3 Uhr").
- Es macht dagegen Sinn, nach der Wahrscheinlichkeit zu fragen, mit der X in ein **Intervall** fällt: Wie groß ist die Wahrscheinlichkeit, dass der Stundenzeiger zw. "12" und "1" steht? - Antwort $F(1) = P(X \leq 1)$ = 1/12.

Zur Verdeutlichung der Verteilungsfunktion F(t) gehen wir noch einmal kurz zurück in die Beschreibende Statistik:

Exkurs: Häufigkeitsverteilung (Beschreibende Statistik)

Bei quantitativen Merkmalen kann man die kumulierten relativen Häufigkeiten H_i bilden, diese bilden die Grundlage für die (kumulierte) **Häufigkeits-Verteilungsfunktion** $H(x)$.

Def D 10-16 Häufigkeits-Verteilungsfunktion

Sei X ein quantitativ-diskretes Merkmal mit den Ausprägungen $X_1 < X_2 < \ldots < X_m$. Sei h_i die relative Häufigkeit von X_i (Def D 10-1) Dann ist $H_i = \sum_{j=1}^{i} h_j$ die **kumulierte relative Häufigkeit** (Wieviel Prozent der Daten haben Ausprägung $x \le X_i$?) und

$$H : \mathbf{R} \to [0,1] \quad \text{mit} \quad H(x) = \begin{cases} 0 & \text{für} \quad x < x_1 \\ H_i & \text{für} \quad x_i \le x < x_{i+1} \\ 1 & \text{für} \quad x \ge x_m \end{cases}$$

ist die **Häufigkeits-Verteilungsfunktion**.

Beispiel: Ein Touristikkonzern will wissen, in welchen Gruppengrößen seine Kunden typischerweise buchen (Alleinreisende, Paare, 3er-, 4er-, 5er-Familien, ...)

Tabelle 10-6

	Buchungen mit Reisendenzahl i		
Anzahl Reisende i	Anzahl n_i (absolute Häufigkeit)	relative Häufigkeit h_i	kumulierte relative Häufigkeit H_i
1	5123	10.7%	10.7%
2	24510	51.3%	62.0%
3	13340	28.0%	90.0%
4	3270	6.8%	96.8%
5	1500	3.2%	100.0%
Summe	47743	100%	

(Wir nehmen hier vereinfachend an, dass 5 die maximale Gruppengröße ist.)

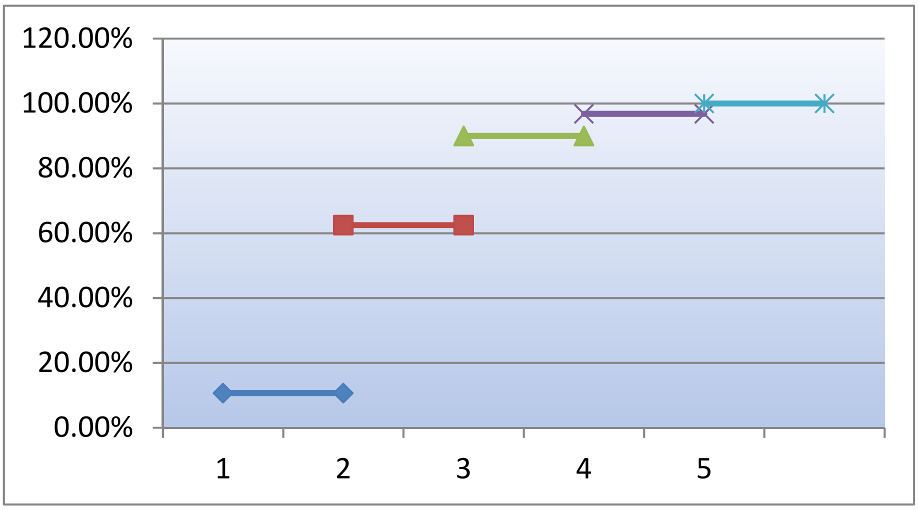


Abbildung 5: Häufigkeitsverteilungsfunktion H(x)

Damit lässt sich die Antwort auf eine Frage wie "Wieviel % meiner Buchungen haben eine Gruppengröße ≤ 3?", nämlich 90%, unmittelbar aus der kumulierten Häufigkeit H_3 ablesen.

Für die Häufigkeitsverteilung gelten folgende, unmittelbar einsichtige Beziehungen:

Satz S 10-7

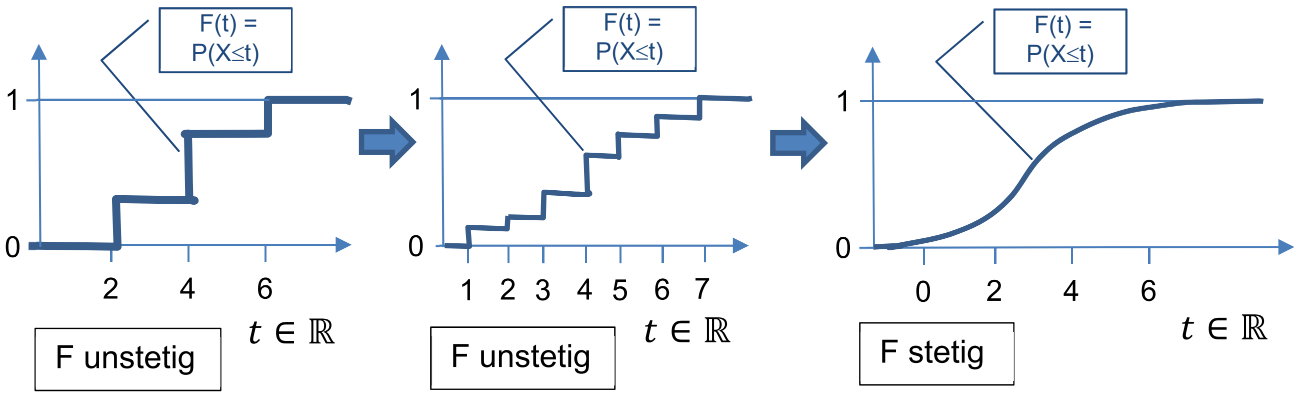
$$H_r = H_{r-1} + h_r \quad \text{für} \quad r \geq 2$$

und $H(x)$ ist monoton wachsend.

Ende Exkurs, weiter mit stetigen Zufallsvariablen:

Der Häufigkeitsverteilung H(x) bei (beschreibenden) Merkmalen entspricht bei Zufallsvariablen die (kumulativen) Verteilungsfunktion F(t).

Zur Veranschaulichung der (kumulierten) Verteilungsfunktion F(t) malen wir die Verteilungsfunktionen zur obigen **Abbildung 1** „diskret → stetig":



Def D10-17 Kumulierte Verteilungsfunktion

Die Funktion $F\colon \mathbb{R} \to [0,1]$ mit $F(t) = P(X \leq t)$ heißt **(kumulierte) Verteilungsfunktion von** X.
F ist monoton wachsend.

F(t) geht immer von 0 nach 1. Bild Schwimmer über Fluss: von Ufer „0" nach Ufer „1". Ferner ist F(t) monoton. Also langweilig? Keineswegs! ALLE relevanten Wahrscheinlichkeiten zu X stecken in F(t).

Satz S10-8 Eigenschaften der Verteilungsfunktion

1. Es gilt für die Verteilungsfunktion $F(t) = P(X \leq t)$ einer jeden Zufallsvariablen
$$\lim_{t\to-\infty} F(t) = 0 \quad \text{und} \quad \lim_{t\to+\infty} F(t) = 1$$
2. $P(a < X \leq b) = F(b) - F(a)$

Def D10-18 Wahrscheinlichkeitsdichte

Für eine stetige Zufallsvariable X: $\Omega \to \mathbb{R}$ heißt eine integrierbare, nichtnegative reelle Funktion $w: \mathbb{R} \to \mathbb{R}$, mit

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} w(t')dt'$$

die **Dichte** oder **Wahrscheinlichkeitsdichte** der Zufallsvariablen X.

Anmerkungen:

- Warum heißt es **Dichte**funktion? – Wenn hoch, liegen die Werte dort dicht, wenn niedrig, wenig dicht. Siehe Processing-Animation GaussianDensity.pde.
- Die Verteilungsfunktion $F(t)$ ist sowohl für diskrete als auch stetige X definiert.
- Die Verteilungsfunktion $F(t)$ für stetige X ist eine Stammfunktion zur Wahrscheinlichkeitsdichte $w(t)$.
- Es gilt $\int_{-\infty}^{\infty} w(u)du = 1$ wegen $\lim\limits_{t\to\infty} F(t) = 1$.

Die Wahrscheinlichkeit, dass X in ein Intervall $]a,b]$ fällt, ist gegeben durch

$$P(a < X \leq b) = \int_a^b w(t)dt = F(b) - F(a)$$

(Das Integral gilt nur bei stetigem X, die Identität mit $F(b) - F(a)$ gilt auch für diskretes X.)

In Vorlesung: Übergang erklären an Bild-Folge.

Beweis zu Satz S10-8

- Punkt 1. ist das Wahrscheinlichkeitsaxioms Def D 10-10, Nr. 2, verallgemeinert für Zufallsvariablen: Wenn wir die Grenze t gegen +∞ verschieben, haben wir das sichere Ereignis: $\lim\limits_{t\to+\infty} F(t) = P(X \leq \infty) = P(\Omega) = 1$. Wenn wir die Grenze t gegen -∞ verschieben, haben wir das unmögliche Ereignis:
$$\lim_{t\to-\infty} F(t) = P(X \leq -\infty) = P(\{\}) = 0$$
- Bew. zu 2.: $P(X \leq a) + P(a < X \leq b) = P((X \leq a) \vee (a < X \leq b)) = P(X \leq b)$. Die 1. Umformung gilt, weil $(X \leq a)$ und $(a < X \leq b)$ unvereinbare Ereignisse sind (s. Def D 10-10, 3. Wahrscheinlichkeitsaxiom)
- Punkt 2. besagt: Kennen wir die Verteilungsfunktion, so können wir die Wahrscheinlichkeit für jedes Intervall]a,b] bequem angeben.

Erwartungswert: In Verallgemeinerung der Summe aus Def D10-14 für diskrete Zufallsvariablen benutzen wir für stetige Zufallsvariablen das Integral:

Def D10-19 Erwartungswert einer stetigen Zufallsvariablen

Für eine stetige Zufallsvariable X mit Wahrscheinlichkeitsdichte $w(t)$ ist Erwartungswert μ:

$$\mu = E(X) = \int_{-\infty}^{\infty} t \cdot w(t)dt$$

Übergang von Formel „E(X) diskret“ erklären!

Die Regel aus Satz S10-6 (Linearität des Erwartungswertes) gilt genauso (!) für stetige Zufallsvariablen.

wenn Zeit: Beweis von E(X+Y)=E(X)+E(Y) über Doppelintegral zeigen (s. Beiblatt)

Über den Erwartungswert kann man auch die Varianz (Maß für die Streuung) berechnen:

Def D10-20 Varianz und Standardabweichung

Für eine Zufallsvariable $X: \ \Omega \to \mathbb{R}$, die den Erwartungswert μ besitzt, ist die Varianz $Var(X) = \sigma^2$ definiert durch:

$$\sigma^2 = Var(X) = E((X-\mu)^2)$$

Die Standardabweichung σ ist die Wurzel aus der Varianz: $\sigma = \sqrt{Var(x)}$

Diese Varianz-Formeln gelten gleichermaßen für **diskrete** und **stetige Zufallsvariablen** (!)

Die Varianz gibt an, wie sehr die Ergebnisse für X um den Wert $E(X)$ streuen: gar nicht (Varianz Null), wenig (Varianz klein) oder viel (Varianz groß).

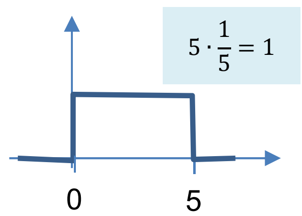
Übungen:

Ü1. *Eine im Intervall $[0,5]$ **gleichverteilte** Zufallsvariable X hat innerhalb des Intervalls die konstante Wahrscheinlichkeitsdichte $w(t) = \frac{1}{5}$ und ist außerhalb Null.*

(a) Zeichnen Sie w(t)! Wieso ist im Intervall die Höhe von $w(t)$ genau $\frac{1}{5}$? Welche Verteilungsfunktion hat X?

(b) Welchen Erwartungswert E(X) hat X?

(c) Welche Varianz Var(X) hat X?



Lösung:

(a) Damit Fläche unter w(t) gleich 1 ist: $\int_{-\infty}^{\infty} w(t)dt = \int_0^5 \frac{1}{5} dt = 1$

Verteilungsfunktion $F(t) = P(X \le t) = \int_{-\infty}^{t} w(t')dt' = \left[\frac{1}{5} t'\right]_0^t$,

also ist F(t)=0 für t<0, $F(t) = \frac{t}{5}$ *für 0≤t≤a und F(t)=1 für t>a*

(b) $\mu = E(X) = \int_0^5 t \cdot \frac{1}{5} dt = \frac{1}{5} \cdot \frac{1}{2} t^2 \Big|_0^5 = \frac{5}{2}$

(c) $\sigma^2 = Var(X) = \int_0^5 \left(t-\frac{5}{2}\right)^2 \cdot \frac{1}{5} dt = \frac{1}{5} \cdot \left[\frac{1}{3}\left(t-\frac{5}{2}\right)^3\right]_0^5 = \frac{1}{5} \cdot \frac{2}{3}\frac{5^3}{8} = \frac{25}{12}$

Ü2. *Sei X eine Zufallsvariable mit* ***dreiecksförmiger*** *Wahrscheinlichkeitsdichte*

$$w(t) = \begin{cases} \alpha t & \text{für} \quad 0 \le t \le T \\ 0 & \text{sonst} \end{cases}.$$

(a) Zeichnen Sie w(t)! Bestimmen Sie die Konstante α.
(b) Welchen Mittelwert E(X) hat die Zufallsvariable X?

10.3.4. Wichtige Verteilungen

[Teschl, Bd. 2, S. 281-308]

Warum sind Verteilungen wichtig? – Für bestimmte Verteilungen haben kluge Köpfe schon ausgerechnet, wie wahrscheinlich jedes $P(X \le b)$ ist. Wir brauchen also nicht immer wieder neu zu rechnen, sondern können in Tabellen o.ä. nachschauen, wenn wir wissen, dass X „Z-verteilt" ist, wobei Z als Platzhalter für den Namen einer Verteilung steht.

Wir haben die Wahrscheinlichkeiten für unendlich viele b direkt verfügbar.

Bei Verteilungsfunktionen von Zufallsvariablen unterscheidet man zwischen **diskreten** und **stetigen Verteilungsfunktionen**, je nachdem, ob die zugrundeliegende Zufallsvariable diskret oder stetig ist. Die nachfolgende Tabelle stellt die wichtigsten Verteilungen vor:

Tabelle 10-7 Wichtige Verteilungen

Typ	Name	Vorkommen	Bemerkung
diskrete Verteilung	**Binomialverteilung**	Ziehen mit Zurücklegen	nur 2 Versuchsausgänge
	hypergeometrische Verteilung	Ziehen ohne Zurücklegen, nur 2 Versuchsausgänge	geht für "große Urne" in Binomialverteilung über
	Poissonverteilung	atomarer Zerfall, Server-Requests	gilt für kleine p [Hartmann, S. 425-430]
stetige Verteilung	**Gleichverteilung**		
	Normalverteilung = Gaußverteilung	Vielfachausführung von Zufallsexperimenten	"Gaußglocke", Grenzverteilung für Binomialverteilung
	Chi-Quadrat-Vert.	statistische Tests	[Hartmann, S. 440ff]
	Exponential-Vert.	Lebensdauer	Dichte = const * e-Funktion [s. Kap. 6.6, Ü Glühbirnen]

Die *kursiven, grün unterlegten* Verteilungen behandeln wir im Rahmen dieser Einführung nicht.

Binomialverteilung

Def D10-21 Bernoulli-Experiment

Ein **Bernoulli-Experiment** ist ein Zufallsexperiment, bei dem es nur **zwei** Ausgänge gibt: das Ereignis A tritt ein (mit Wahrscheinlichkeit p) oder nicht, also tritt Ereignis $\bar{A}$ ein (mit Wahrscheinlichkeit 1-p). Wird ein Bernoulli-Experiment n-mal hintereinander ausgeführt, so spricht man von einer **Bernoulli-Kette** der Länge n.

Wie wahrscheinlich ist es, dass in einer Bernoulli-Kette der Länge n genau k-mal A eintritt? Sei X die Zufallsvariable, die das Eintreten von A zählt. Wie wahrscheinlich ist $P(X=k)$?

Ein typischer Experimentausgang für n=5 und k=3 ist $\underbrace{(\bar{A}, A, A, \bar{A}, A)}_{n-\text{mal}}$

p^k	Wahrsch., dass k bestimmte Positionen (hier: 2,3,5) mit A besetzt sind
$(1-p)^{n-k}$	Wahrsch., dass die (n-k) anderen (hier: 1,4) mit $\bar{A}$ besetzt sind
$\binom{n}{k}$	Anzahl der k-Teilmengen in Positionsmenge {1,2,…,n}

Satz S10-9 Binomialverteilung

Gegeben sei eine Bernoulli-Kette der Länge n, bei der Ereignis A mit $P(A)=p$ eintritt. Sei X eine diskrete Zufallsvariable, die die Anzahl der Versuche zählt, in denen Ereignis A eintritt. X heißt **binomialverteilt mit den Parametern n und p** oder kurz $b_{n,p}$-verteilt und es gilt:

$$b_{n,p}(k) = P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

(zu $\binom{n}{k}$ siehe Def D10-11)

Erwartungswert $E(X)$ und Varianz $Var(X)$ einer binomialverteilten Zufallsvariablen sind

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

Den Beweis der (überraschend einfachen!) Formeln für E(X) und Var(X) findet man in [Teschl, Bd. 2, S. 287 od. Hartmann04, S. 421]. Er ist nicht schwer.

Beispiele in Excel → s. Mathe-Reihen-V2.xlsx, Blatt Binomial.

Wann Binomial, wann nicht? - Antwort: „Bi-nomial“ heißt „2 Namen“ → Binomial nur anwendbar, wenn ein Versuch mit **genau zwei** Ausgängen gemacht wird.

Hypergeometrische Verteilung

Diese Verteilung hatten wir schon in Kapitel 10.3.2 "Kombinatorik", Übung Ü3 (4 Richtige bei 6-aus-49), berechnet. Es gilt

Satz S10-10 hypergeometrische Verteilung

Eine Urne enthalte N Kugeln, davon S schwarze. Eine diskrete Zufallsvariable Y, die bei n Zügen **ohne Zurücklegen** aus einer Urne die Anzahl der schwarzen Kugeln zählt, heißt **hypergeometrisch verteilt mit den Parametern N, S und n** oder kurz $h_{N,S,n}$ -verteilt. Es ist

$$h_{N,S,n}(k) = P(Y=k) = \frac{\binom{S}{k}\binom{N-S}{n-k}}{\binom{N}{n}}$$

(zu $\binom{S}{k}$ siehe Def D10-11)

Für $N \gg n$ gilt mit $p=S/N$ als gute Näherung:

$$h_{N,S,n}(k) \approx b_{n,p}(k)$$

Auch hier ist für große N, n und k die Berechnung mühsam. Es gibt wieder entsprechende Vereinfachungen (Wenn das Reservoir N groß ist, ist der Unterschied zwischen "Ziehen mit" und "Ziehen ohne Zurücklegen" gering >> Binomialverteilung)

Übung: Aus Urne mit N=60 Kugeln, davon 6 weiße, werden 2 Kugeln mit/ohne Zurücklegen gezogen. Wie wahrscheinlich ist "weiß-weiß"?

Gleichverteilung

Dies ist die einfachste stetige Verteilung. Wir hatten ihre wichtigsten Eigenschaften bereits in dem Beispiel nach **Def D10-20** notiert. Es gilt

Satz S10-11 Gleichverteilung

Eine in [a,b] $\subset$ R **gleichverteilte** stetige Zufallsvariable X, besitzt folgende Eigenschaften:

Wahrscheinlichkeitsdichte $w(t) = \begin{cases} \frac{1}{b-a} & \text{für} \quad a \le t \le b \\ 0 & \text{sonst} \end{cases}$,

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

Varianz $Var(X) = \frac{(b-a)^2}{12}$

Anmerkungen

- Für [0,1]-gleichverteilte Zufallsvariablen gilt also Erwartungswert $\mu = 0.5$ und $\sigma = \sqrt{\frac{1}{12}}$. D.h. im Intervall [μ-σ, μ+σ] liegen $[0.5 + \frac{1}{\sqrt{12}} - (0.5 - \frac{1}{\sqrt{12}})] = \frac{2}{\sqrt{12}} = 57.7\%$, also rund 60% der Daten. Diese Aussage „**Es liegen 57.7% der Daten in [μ-σ, μ+σ]**" gilt auch allgemein für in [a,b]-gleichverteilte Zufallszahlen.
- Ein Zufallsgenerator auf dem Computer muss notwendigerweise diese beiden Bedingungen erfüllen (darüber hinaus noch weitere Bedingungen wie "frei von Korrelation", die wir hier nicht behandeln)
- Die Gleichverteilung kommt in der Natur eher selten vor. Sie ist aber bei Computersimulationen oft der Ausgangspunkt, um diskret-zufällige Ereignisse zu würfeln.
  Beispiel: Erzeugt die Funktion `rnd()` [0,1[-verteilte Zufallszahlen, dann ist `int(37*rnd())` geeignet, um ein Roulette zu simulieren.

Normalverteilung = Gaußverteilung

Die Normalverteilung ist die wichtigste stetige Verteilung. Sie spielt in praktisch allen Anwendungen der Statistik eine große Rolle.

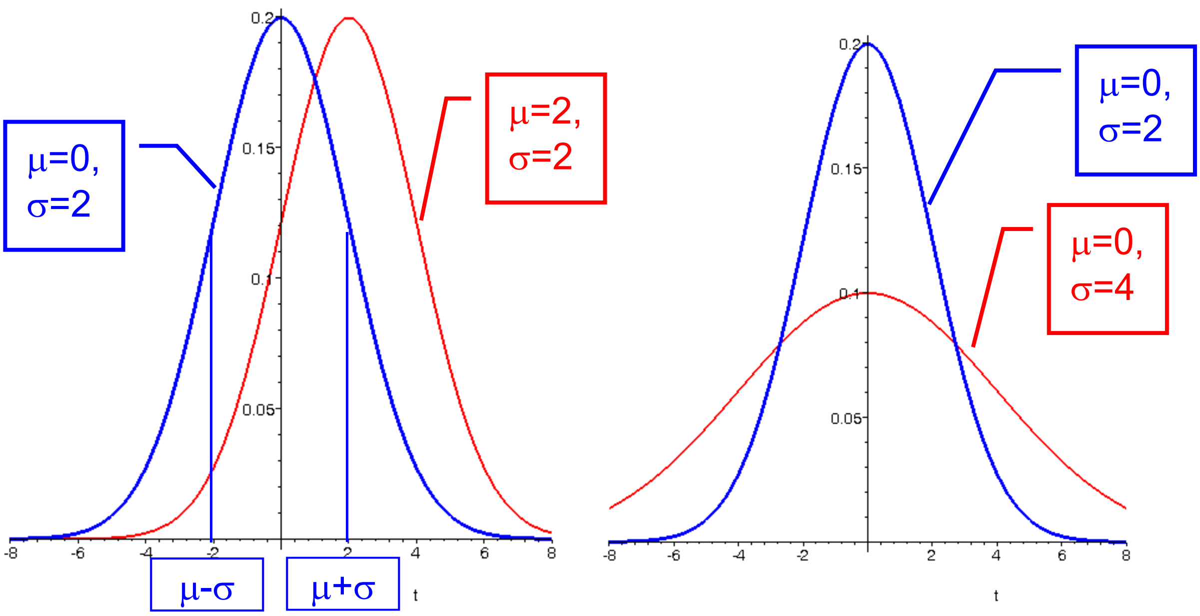
Def D10-22 Normalverteilung (Gaußverteilung)

Eine stetige Zufallsvariable **X: Ω → R** heißt **normalverteilt mit Mittelwert μ und Standardabweichung σ** oder kurz N(μ,σ²)-verteilt, wenn ihre Dichtefunktion

$$w(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2} \right)$$

lautet.

Die Normalverteilung hat die typische Form der Gauß'schen Glockenkurve:



Die Parameter μ und σ lassen sich auch unmittelbar aus der grafischen Darstellung der Dichtefunktion ablesen: Die Gauß'sche Glockenkurve hat ihr Maximum bei $t=\mu$ und ihre Wendepunkte bei $\mu-\sigma$ und $\mu+\sigma$.

Bei der Gaußverteilung liegen 68.2% der Daten in $[\mu-\sigma,\ \mu+\sigma]$. (Beweis s. Ü2)

Für praktische Anwendungen braucht man neben der Dichte auch die Verteilungsfunktion $F(t) = P(X\leq t)$ (s. Def D10-12). Diese ist leider für die Normalverteilung nicht mehr über elementare Funktionen darstellbar, sondern man muss Tabellen oder Näherungsverfahren benutzen. Das Problem lässt sich aber für alle μ und σ auf eine Tabelle zurückführen:

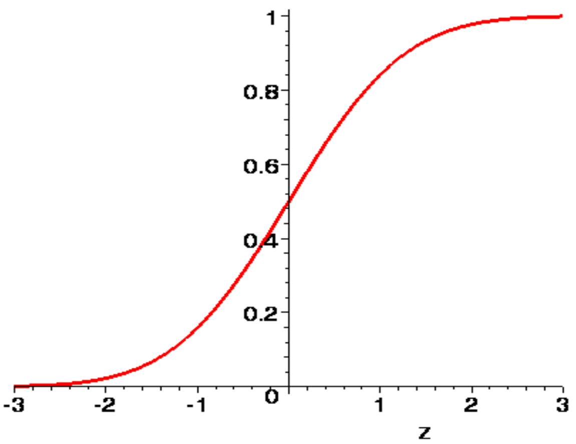
Def D10-23 Standardnormalverteilung, Verteilungsfunktion Φ(x)

Die Normalverteilung $N(0,1^2)$ mit Erwartungswert 0 und Standardabweichung 1 heißt **Standardnormalverteilung**. Ihre Verteilungsfunktion ist

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$$

lautet. $\Phi(z)$ gibt also die Wahrscheinlichkeit an, dass eine standardnormalverteilte Zufallsvariable Z nicht größer als z ist.

Die Verteilungsfunktion (engl. *cdf = cumulative density function*) hat die folgende Form

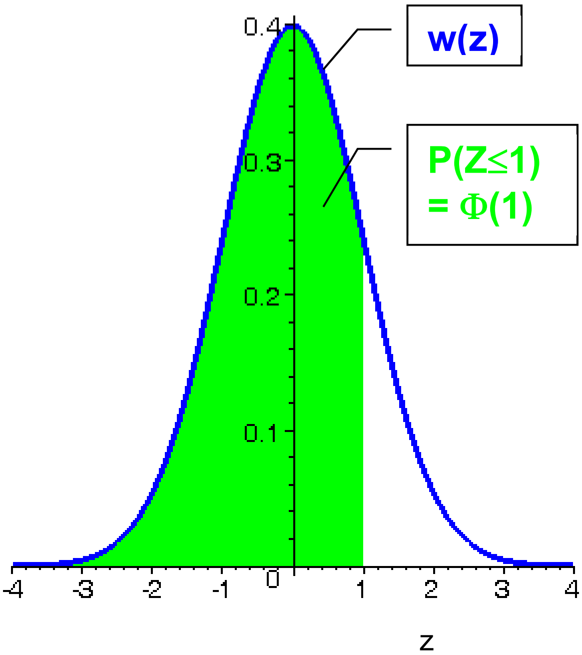


Maple:

```
with(stats):

plot(statevalf[cdf,
  normald[0,1]], -3..3,
colour=red,thickness=3);
```

Alternative Darstellung: Die Verteilungsfunktion ist die Fläche unter der Standard-Dichtefunktion bis zum Punkt z:



Maple:

```
w:= z->statevalf[pdf,
       normald[0,1]](z);
p1:=plot(w(z),z=-4..4,
thickness=3,color=blue):

p2:=plot(w(z),z=-4..1,
filled=true,color=green,
thickness=2):

display(p1,p2);
```

Tabelle 10-8 Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung (Ausschnitt)

z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	$\boxed{0.8413}$	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

$\Phi(1) = 0.8413$

[Nachkommastellen erläutern]

In vielen Fällen interessiert auch die **inverse Verteilungsfunktion der Standardnormalverteilung**. Man sucht bei vorgegebenem $q \in [0,1]$ diejenige Stelle z_q mit $\Phi(z_q)=q$. Anschaulich bedeutet z_q die Stelle, bis zu der unter der Dichtefunktion die Fläche q aufgelaufen ist. Man nennt z_q das **q-Quantil**.

[an Bild erklären!]

Beispiel: Man bestimme aus Tabelle 10-8 das q-Quantil für q=0.9.

Lösung mit "nächster Nachbar": Im Tabelleninnern den Wert suchen, der 0.9 am nächsten ist: $\Phi(1.28)=0.8997$ und damit $z_q=1.28$.

Lösung mit "linearer Interpolation": Aus der Tabelle entnimmt man $\Phi(1.28)=0.8997$ und $\Phi(1.29)=0.9015$. Zwischen 1.28 und 1.29 liegt also der Punkt z_q. Via Dreisatz bzw. lineare Interpolation erhalten wir

$$\frac{z_q - 1.28}{0.9 - 0.8997} = \frac{1.29 - 1.28}{0.9015 - 0.8997} \quad \Leftrightarrow \quad z_q - 1.28 = 0.0003 \cdot \frac{0.01}{0.0018}$$

und damit $z_q=1.2816$.

Für Berechnungen mit Normalverteilungen gelten folgende nützlichen Beziehungen:

Satz S10-12 Regeln für Normalverteilungen

1. $\Phi(-z) = 1 - \Phi(z)$.
2. Ist X eine $N(\mu,\sigma^2)$-verteilte Zufallsvariable, so ist $Z = \frac{X-\mu}{\sigma}$ $N(0,1^2)$-verteilt.
3. Für die Verteilungsfunktion $F(b)=P(X\le b)$ gilt: $F(b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$
4. $P(a < X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$

5. $q = \Phi(z_q) \quad \Leftrightarrow \quad 1 - q = \Phi(-z_q)$

6. Ist $\mathbf{z_q}$ das q-Quantil einer $N(0,1^2)$-Verteilung, so ist $x_q = \sigma \cdot z_q + \mu$ das q-Quantil einer $N(\mu,\sigma^2)$-Verteilung.

Beispiel 1: Die Körpergröße in Metern bei einer Gruppe von Menschen sei normalverteilt mit Mittelwert 1.75 und Standardabweichung 0.20. Man bestimme die Körpergröße, die Menschen nicht überschreiten, welche zum (unteren) 0.06-Quantil gehören.

Lösung: Zunächst bestimmt man das 0.06-Quantil der Standardnormalverteilung

$$0.06 = P(Z \le z_q) = \Phi(z_q) \quad \Leftrightarrow \quad 0.94 = \Phi(-z_q)$$

Diese Umformung gilt wg. Satz S10-12, Nr. 5. Der **Tabelle 10-8** entnehmen wir $-z_q=1.56$ (nächstgelegener Wert, ohne lineare Interpolation).

Nach Satz S10-12, Nr. 6 ist dann das Quantil x_q der $N(1.75,(0.2)^2)$-Normalverteilung gegeben durch $x_q = \sigma z_q + \mu = 0.2 \cdot (-1.56) + 1.75 = 1.438$.

Für die kleinsten 6% aus der Menschengruppe gilt also, dass sie eine Körpergröße von höchstens 1.438 m haben.

Ü

Ü1. *Wie groß ist bei obiger Verteilung die Wahrscheinlichkeit, dass ein Mensch größer als 2.00 ist?*

Ü2. *X sei eine $N(\mu,\sigma^2)$-verteilte Zufallsvariable. Wie groß ist die Wahrscheinlichkeit, dass X innerhalb des 1σ (bzw. 2σ, 3σ)-Intervalls um μ herum liegt?*

Ü3. *Sie sind Sys-Admin. Die durchschnittliche Wartezeit zwischen zwei Hacker-Attacken auf Ihrem zentralen Server sei $N(48h,(6h)^2)$-verteilt. Gerade ist eine Attacke passiert. In welchem Zeitintervall ist mit 82% mit der nächsten Attacke zu rechnen?*

[Ü3 wird in den Übungen gelöst!]

Ü4.

Eine Zufallsvariable X ist $N(2,1^2)$-verteilt. Wie groß ist die Wahrscheinlichkeit $P(1.5<X<2.5)$?

Ü5.

Wo liegt das 0.8-Quantil einer $N(3,2^2)$-verteilten Zufallsvariablen X?

10.3.5. Der zentrale Grenzwertsatz

[Teschl, Bd. 2, S. 309-313]

Motivation: Ein Versuch mit Ausgang A oder $\overline{A}$ mit P(A)=40% wird 1000-mal wiederholt. Wir zählen in X die Anzahl der A's. Wie wahrscheinlich ist P(X<450)?

Nach der Binomialverteilung müssten wir $b_{n,p}(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ für n=1000 und k=0,1,...,449 ausrechnen und alles aufaddieren. Nicht nur dass das eine Riesenarbeit ist, die Zahlen würden so groß und so klein, dass sie jeden Taschenrechner sprengen. Was also tun?

Kurz und knapp: **Wenn Sie dasselbe Experiment ganz oft wiederholen, dann sind die Summe der Ergebnisse und ihr Mittelwert normalverteilt** [Maas14, S. 164].

Genauer:

Satz S10-13 Der zentrale Grenzwertsatz

Ein Experiment wird n-mal unabhängig und gleich durchgeführt und die Ergebnisse in den Zufallsvariablen $X_1, X_2, \ldots, X_n$ notiert. Der Erwartungswert der X_i sei jeweils μ, die Varianz σ^2. Dann hat die Summe $\boldsymbol{S_n} = X_1 + X_2 + \ldots + X_n$ den Erwartungswert $n\mu$ und die Varianz $n\sigma^2$. Ihre Verteilung wird für große n einer **Normalverteilung** immer ähnlicher.

Die zugehörige standardisierte Zufallsvariable

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

ist für große n näherungsweise **standardnormalverteilt**. Es gilt:

$$\lim_{n\to\infty} P(Z_n \le z) = \Phi(z)$$

Folgerungen:

- Die Summe $\boldsymbol{S_n}$ hat also die Standardabweichung $\sqrt{n\sigma^2} = \sqrt{n}\sigma$.
- Wichtiger ist aber noch der Mittelwert $M_n = \frac{S_n}{n}$: Er hat klarerweise den Erwartungswert μ, (wie die einzelnen X_i), aber die Standardabweichung von M_n ist mit $\frac{\sqrt{n}\sigma}{n} = \frac{\sigma}{\sqrt{n}}$ **um den Faktor $\frac{1}{\sqrt{n}}$ kleiner**. Das ist der Grund, warum man in der Forschung oft Experimente wiederholt! Denn bei 10maliger Wiederholung ist die Streuung von M_n um den Faktor $\sqrt{10} \approx 3.16$ kleiner. Bei 100maliger Wiederholung um den Faktor 10.

Anmerkungen:

- Für alle Experimente und Messungen, die n-mal unabhängig wiederholt werden, trifft dieser Satz zu.
- Man beachte, dass der Satz völlig unabhängig von der Art der Verteilung gilt, die die X_i haben (!!). Jede Verteilung strebt bei n-facher Wiederholung und Summation gegen die Normalverteilung.
- Wann ist n groß? – Typischerweise ist für $n \ge 30$ die Näherung gut.

Ein wichtiger Spezialfall des zentralen Grenzwertsatzes ist der **Satz von Moivre-Laplace**, der die Binomialverteilung behandelt:

Für große n und k ist die Berechnung der Binomialkoeffizienten mühsam. Noch mühsamer ist für "k in der Mitte" die Berechnung von Wahrscheinlichkeiten $P(X \le k)$ wg. der Summen über Binomialkoeffizienten. Glücklicherweise gibt es, gerade für große n, eine Vereinfachung, nämlich die Gaußverteilung nach dem zentralen Grenzwertsatz:

Satz S10-14 Satz von Moivre-Laplace

Seien X eine $b_{n,p}$-verteilte Zufallsvariable. Dann ist, <u>falls np>5 und n(1-p)>5</u>, folgende Rechnung in guter Näherung möglich:

$$P(r \le X \le s) \approx \Phi\left(\frac{s-np+0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{r-np-0.5}{\sqrt{np(1-p)}}\right)$$

Spezialfall $r \to -\infty$:

$$P(X \le s) \approx \Phi\left(\frac{s-np+0.5}{\sqrt{np(1-p)}}\right)$$

Da X diskret ist, ist hier der Unterschied zwischen < und ≤ relevant.

Excel-Visualisierung Binomial – Gauss zeigen!

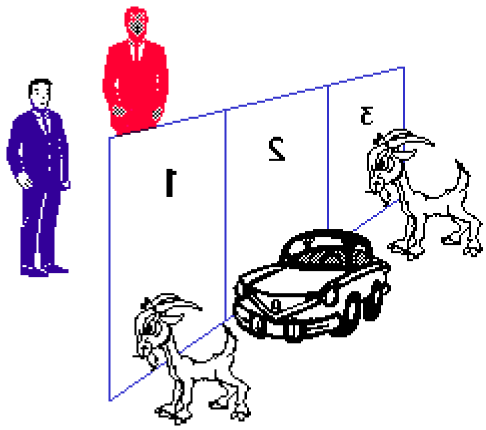
Ü

Ü1. *Die Wahrscheinlichkeit einer Jungengeburt sei 0.52. Wie groß ist die Wahrscheinlichkeit, dass unter 1000 Geburten mehr als 500 Mädchen sind?*

Ü2. *Lösen Sie die Aufgabe aus der Motivation: Ein Versuch mit Ausgang A oder $\overline{A}$ mit P(A)=40% wird 1000-mal wiederholt. Wir zählen in X die Anzahl der A's. Wie wahrscheinlich ist P(X<450)?*

10.3.6. Bedingte Wahrscheinlichkeiten

[Teschl, Bd. 2, S. 232-240]



<u>Motivation:</u> Das berühmte 3-Türen-Ziegenproblem wird in Vorlesung erläutert. Soll ich mich umentscheiden, wenn der Moderator mir eine Tür mit Ziege öffnet? Begründung?

Die richtige Lösung können wir erklären, wenn wir bedingte Wahrscheinlichkeiten verstehen.

$P(A|B)$ = Wahrscheinlichkeit, dass A (auch noch) eintritt, wenn B bereits eingetreten

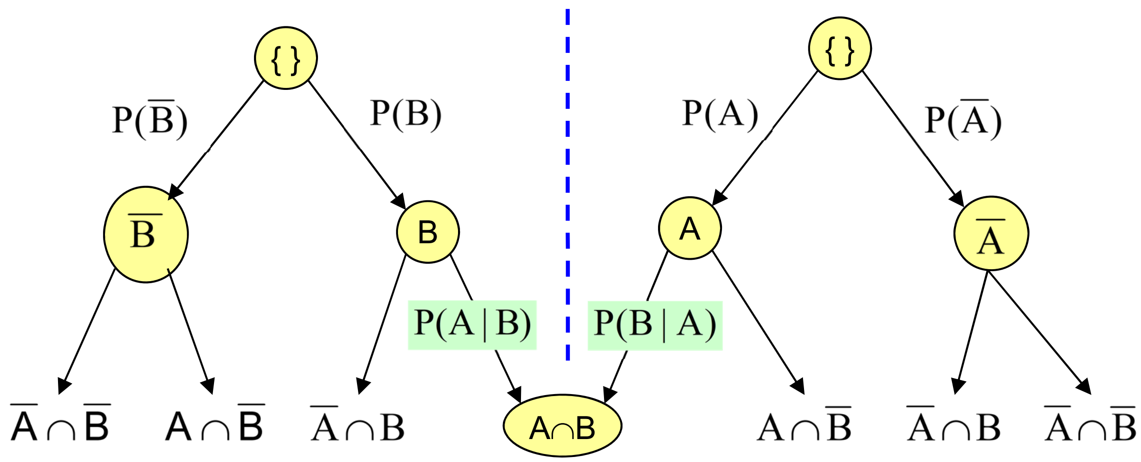
<u>Beispiel Urne:</u>

In einer Urne liegen 1 weiße und 7 schwarze Kugeln. Alex und ich ziehen ohne Zurücklegen. Sei

A = "Alex zieht eine weiße Kugel"

B = "Ich ziehe **keine** weiße Kugel"

Entscheidungsbäume (Knoten = Ereignisse, Kanten = Wahrscheinlichkeit, dass Kind-Ereignis eintritt, wenn Eltern-Ereignis bereits eingetreten):



Übung: Überlegen Sie, welche konkreten Zahlen beim Urnen-Beispiel zu den 4 Wahrscheinlichkeiten P(A), P(B), P(B|A), P(A|B) und P(A∩B) gehören!

Es gilt die Formel

> **Satz S10-15 Multiplikationssatz für Wahrscheinlichkeiten**
>
> (*) $$P(B\,|\,A)P(A) = P(A\cap B) = P(A\,|\,B)P(B)$$
>
> In Worten:
>
> *P(A|B) ist die Wahrscheinlichkeit, mit der man P(B) multiplizieren muss, um P(A ∩ B) zu erhalten. Dabei ist P(A|B) nur für P(B)≠0 definiert.*

Dies bestätigt sich im konkreten Beispiel: $1\cdot\frac{1}{8} = \frac{1}{8} = \frac{1}{7}\cdot\frac{7}{8}$

Übung: Überlegen Sie, welche konkreten Zahlen beim Urnen-Beispiel mit **nun 2 weißen und 7 schwarzen Kugeln** zu den 4 Wahrscheinlichkeiten P(A), P(B), P(B|A), P(A|B) und P(A∩B) gehören!

Falls P(B)≠0 ist, können wir Satz S10-15 zur Def. der bedingten Wahrscheinlichkeit nutzen:

> **Def D10-24 Bedingte Wahrscheinlichkeit**
>
> Seien A und B zwei Ereignisse mit P(B) ≠ 0. Dann heißt
>
> $$P(A\,|\,B) = \frac{P(A\cap B)}{P(B)}$$
>
> die **bedingte Wahrscheinlichkeit** von A unter der Bedingung B.

Satz S10-16 Bayes-Formel

Seien A und B zwei Ereignisse mit $P(B) \neq 0$. Dann gilt:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Bew.: folgt unmittelbar aus Formel in Satz S10-15.

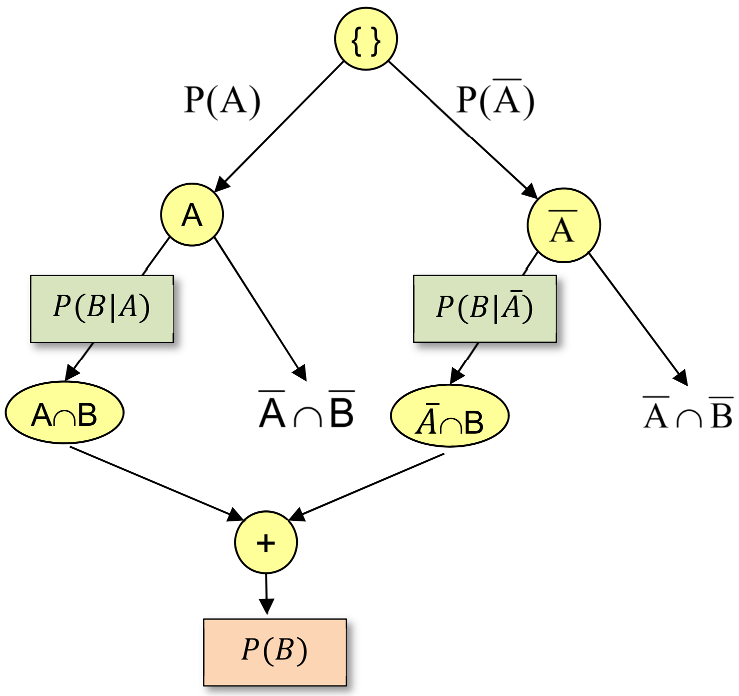
Manchmal ist P(B) nicht direkt gegeben. Dann hilft oft der folgende Satz:

Satz S10-17 Satz von der totalen Wahrscheinlichkeit

Seien A_1, A_2, … Ereignisse, die sich paarweise ausschließen und sei $\bigcup_i A_i = \Omega$. Dann gilt:

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

Beweis in Vorlesung [über Bild: Nehmen wir o.B.d.A. an: $A_1 = A,\ A_2 = \bar{A}$]



Wir laufen alle gelben Wege entlang, die auf ein Blatt mit B führen, multiplizieren die Wahrscheinlichkeiten und addieren sie:

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$$

Geht genauso für mehr als zwei A_i, wenn wir $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$ haben.

Beispiel Autohersteller [Hartmann, S. 395]

	Lieferant 1	Lieferant 2	Lieferant 3
Anteil	45%	35%	20%
Ausschuss	2%	3%	1%

Übung:

1. Wie groß ist die Wahrscheinlichkeit, dass ein geliefertes Teil fehlerhaft ist?
   [B = „Teil ist Ausschuss“]
2. Wie groß ist die Wahrscheinlichkeit, dass ein fehlerhaftes Teil von Lieferanten 1, 2 oder 3 stammt?

Def D10-25 Statistische Unabhängigkeit

Seien A und B zwei Ereignisse. A und B heißen **statistisch unabhängig** genau dann, wenn

$$P(A \cap B) = P(A) \cdot P(B).$$

Falls $P(A) \neq 0$, so gilt: A und B statistisch unabhängig $\Leftrightarrow$ $P(B|A) = P(B)$.

In Worten: Der Eintritt oder Nicht-Eintritt von Ereignis A ändert nichts an der Wahrscheinlichkeit P(B).

Übung: In einer Urne befinden sich 10 Kugeln, darunter 4 schwarze und 6 weiße. 2 Kugeln werden gezogen. Sei

A = "Die 1. gezogene Kugel ist schwarz"

B = "Die 2. gezogene Kugel ist schwarz"

Wie lauten die Wahrscheinlichkeiten P(A), P(B|A) und P(B), wenn man die Kugeln ohne Zurücklegen entnimmt? Sind A und B statistisch unabhängig? Berechnen Sie die Wahrscheinlichkeit, 2 schwarze Kugeln zu ziehen, $P(A \cap B)$.

Beantworten Sie die gleichen Fragen, wenn man die Kugeln mit Zurücklegen entnimmt.

In der Vorlesung klären wir mit unserem Wissen über bedingte Wahrscheinlichkeiten auch das Ziegen-Problem.

[mit eingebetter Übung]

[Selbststudium]

In der Vorlesung gehen wir kurz auf das Buch von Gigerenzer ein:

[**Gigerenzer09**] Gerd Gigerenzer: *Das Einmaleins der Skepsis: Über den richtigen Umgang mit Zahlen und Risiken, Berlin Verlag, 2009.* Wieso man Zusammenhänge mit bedingten Wahrscheinlichkeiten viel besser begreifen kann, wenn man sie als natürliche Häufigkeiten ausdrückt. Positiv rezensiert unter http://www.gavagai.de/philrezi/HHPRZ18.htm.

→ siehe Gigerenzer.pdf, Gigerenzer.pptx

10.4. Fazit Statistik

Viele wichtige Begriffe der Beschreibenden Statistik und Wahrscheinlichkeit wurden in diesem Kapitel eingeführt. Versuchen Sie die Lücken in nachfolgender Tabelle zu füllen:

Tabelle 10-9

Ü

Beschreibende Statistik	Wahrscheinlichkeitstheorie	Kap.
Merkmal (quantitativ)		
Ausprägung eines Merkmals		
Menge der Ausprägungen		
rel. Häufigkeit h_i		
kumulierte Häufigkeit H_i	---	
	Verteilungsfunktion F(x)	
Histogramm f(x)		
	Erwartungswert μ, E(X)	
(empirische) Varianz s^2		
(empirische) Std.-abweichung s		

Was hängenbleiben sollte:

- aus der beschreibenden Statistik:
  - wie man relative **Häufigkeiten** und **Histogramme** berechnet,
  - wie man Mittelwert, Median und Varianz einer Stichprobe ermittelt
  - wie man einen Boxplot erstellt bzw. was seine Elemente bedeuten
- aus der Wahrscheinlichkeitstheorie
  - die 4 Grundformeln der Kombinatorik (**Urnenexperimente**)
  - wie man bedingte Wahrscheinlichkeiten ausrechnet
  - wie man Erwartungswert und Varianz von (stetigen oder diskreten) **Zufallsvariablen** ausrechnet
  - der Zusammenhang zwischen **Wahrscheinlichkeitsdichte** und **Verteilungsfunktion**,
  - was die **Binomialverteilung** ist und wann sie durch Normalverteilung approximierbar ist
  - was die **Normalverteilung** (Gaußverteilung) ist und wie man Wahrscheinlichkeiten für normalverteilte Zufallsvariablen ausrechnet

10.4.1. Where to go from here

Vertiefungsmöglichkeiten: Wenn Sie mehr über Statistik lernen wollen und wissen wollen, was man noch mit Statistik machen kann:

Interessante Statistik-Apps:

http://www.graphpad.com/quickcalcs/

[**Gigerenzer09**] Gerd Gigerenzer: *Das Einmaleins der Skepsis: Über den richtigen Umgang mit Zahlen und Risiken, Berlin Verlag, 2009.* Wieso man Zusammenhänge mit bedingten Wahrscheinlichkeiten viel besser begreifen kann, wenn man sie als natürliche Häufigkeiten ausdrückt. Positiv rezensiert unter http://www.gavagai.de/philrezi/HHPRZ18.htm.

Das wichtige Gebiet der **Schließenden Statistik** (manche sagen: hier fängt die Statistik erst an) haben wir in dieser Einführung nicht behandelt. Themen der Schließenden Statistik sind:

- Wie kann ich aus Messungen an einer Stichprobe schließen, wie sich (wahrscheinlich) die Gesamtheit verhält? (Bsp. Wahl-Hochrechnungen, Konsumentenbefragungen). Wie sicher kann ich mir sein? Diese Grundfrage gliedert sich in folgende Teilgebiete:
  - **Parameterschätzung**: Welchen Wert nimmt eine Zufallsvariable an? In welchem Konfidenzintervall liegt sie mit welcher Sicherheit?
  - **Hypothesentests**: Sind die Mittelwerte zweier Stichproben signifikant verschieden? (t-Test). Sind zwei empirisch erhobene Verteilungen signifikant unterschiedlich? (Chi-Quadrat-Test)
  - **Anpassungstests**: Welche Dichtefunktion beschreibt meine zahlreichen Zufallsexperimente bestmöglich? (Anpassen einer Funktion an Daten)

Solche Themen werden im **Master Informatik** (MI + INF) behandelt.

Lit: [Maas14, S. 171ff], [Stingl02] oder [Teschl05, Bd. 2, S. 325-357]

10.4.2. Klausurvorbereitung

Hier noch ein Hinweis auf Online-Übungsaufgaben, die eine sehr gute Hilfe sind, wenn Sie sich auf die Klausur vorbereiten wollen:

Kombinatorik-Lernprogramm: www.gm.fh-koeln.de/kombinatorik. NEU 2020: Nun auch erweitert um Binomial- und hypergeometrische Verteilung.

MathWeb, Kurs Statistik: Eine kleine Sammlung von Aufgaben zu Normalverteilung und bedingter Wahrscheinlichkeit, die ich gemeinsam mit meinem Kollegen Klaus Giebermann von der HS Ruhr-West entwickelt habe: https://wmint.de/mathematik/chapter12.php.
Das Schöne daran: Sie können sich – nachdem Sie es selber versucht haben (!) – den Lösungsweg anschauen und sich auf Knopfdruck immer neue, ähnlich gelagerte Aufgaben mit anderen Zahlen ziehen.