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]

" Alles was lediglich wahrscheinlich ist, ist wahrscheinlich falsch." [René Descartes (1596 – 1650)]

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, Clusterung
- **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

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, Stichworte OLAP, SAP). Die (beschreibende) Statistik (Kap. 10.2) legt hierfür die Grundlagen.
  Wer solche und ähnliche Anwendungen interessant findet: **WPF Data Mining praktisch – Vorbereitung DMC** (W. Konen, T. Bartz-Beielstein)
- Die **Kombinatiorik** (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: Wieviele Möglichkeiten gibt es beim n-Städte-TSP (Kap. 9.4.2)? Wieviele Passwörter der Länge 5 enthalten "AA"?)
- 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. 10.3.5)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

[Stingl03, S. 581-598]

10.2.1. Merkmale und Merkmalstypen

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) durschnittliche 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 Nachbarschaftsprinzipien ("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 prinzipielle 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**	**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, ...	0.51^{0}C, 27.36^{0}C, ...
Ordnung?	nein	ja	ja	
Summen- und ∅-Werte?	nein	fragwürdig (!!)[5]	ja	

Beispiele:

- Der wöchentliche Spitzenreiter der Fussballbundesliga ist ein qualitatives Merkmal der Wochen der Saison
- Der Tabellenrang ist ein Rangmerkmal der Vereine 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 ihre graphische Darstellung

Für jedes Merkmal, ob qualitativ oder quantitativ, lassen sich große Tabellen oft übersichtlich zusammenfassen, wenn man absolute Häufigkeiten n_i und relative Häufigkeiten h_i bildet:

Tabelle 10-3

Merkmal

Aus-prä-gun gen

	Wochen mit Mannschaft i als Spitzenreiter		
Mannschaft i	Anzahl n_i (absolute Häufigkeit)	relative Häufigkeit $h_i = \frac{n_i}{N}$	
Werder Bremen	2	2/15 = 0.1333	
Schalke 04	5	5/15 = 0.3333	
FC Bayern	5	5/15 = 0.3333	
VFB Stuttgart	2	2/15 = 0.1333	
VFL Bochum	1	1/15 = 0.0666	

[5] 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.

Summe	15 = N	1.00000	

(Für Rangmerkmale kann man die relativen Häufigkeiten zwar auch bilden, dies macht aber in der Regel nicht viel Sinn: An wieviel % aller Wochen war eine Mannschaft auf dem 1. Tabellenplatz? – Klarerweise 100%!)

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

Def D 10-2 Häufigkeits-Verteilungsfunktion

Sei X ein quantitativ-diskretes Merkmal mit den Ausprägungen $x_1 < x_2 < \ldots < x_m$. Dann ist

$H_i = \sum_{j=1}^{i} h_j$ die **kumulierte relative Häufigkeit** (Für wieviel % der Datensätze gilt $x \le x_i$?) 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, Familien, ...)

Tabelle 10-4

	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%	

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 Häufigkeiten gelten folgende, unmittelbar einsichtige Beziehungen:

Satz S 10-1

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

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

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

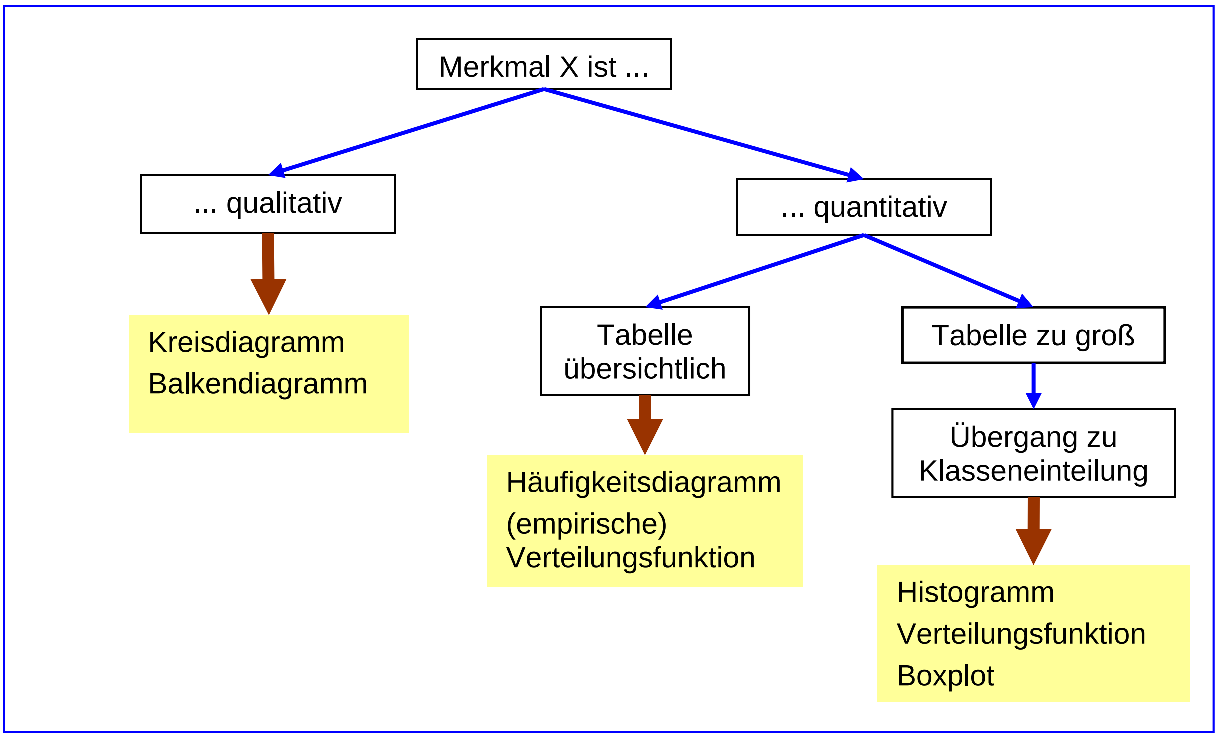
und $H(x)$ ist monoton wachsend.

<u>Übung:</u> Gegeben sei ein Merkmal x_i, 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 und H_i. Mit welcher Häufigkeit gilt $4 \leq x_i < 7$? Mit welcher Häufigkeit gilt $2 < x_i \leq 6$?

Grafische Darstellung von relativen Häufigkeiten:



Beispiele in Vorlesung! [s. Mathe-Reihen-V2.xls]

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 sind

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

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

gültige Klasseneinteilungen. Man beachte, dass die Klassen unterschiedlich breit sein können. Mit $K_i = [x_i,x_{i+1}[$ läßt sich eine Einteilung in m Klassen $K_1,\ldots, K_m$ durch $m+1$ Zahlen $x_1, \ldots , x_{m+1}$ charakterisieren. Die Klasse K_i hat die Breite $\Delta x_i = x_{i+1} - x_i$.

Offene Randklassen (z.B. $[4.5, \infty[$)sind zwar prinzipiell zulässig, bereiten aber bei der weiteren Auswertung (Histogramm, s.u.) Schwierigkeiten und sollten daher vermieden 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).

Die Häufigkeits-Verteilungsfunktion $H(x)$ (s. Def D 10-2) über der Klasseneinteilung nennt man auch **kumuliertes Histogramm**.

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.

Wieviele 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.

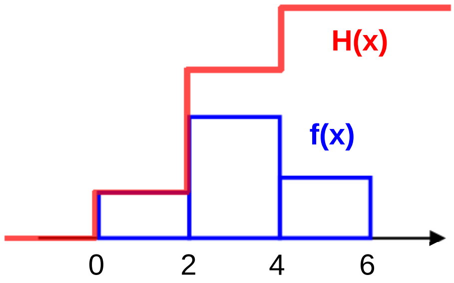
Beispiel:

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	$H(x)$ f. $x \geq x_i$
[0.0, 2.0[	4	0.20	0.0	0.2
[2.0, 4.0[	11	0.55	2.0	0.75
[4.0, 6.0[	5	0.25	4.0	1.00
Summen	20	1.00		



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 und die Häufigkeits-Verteilungsfunktion $H(x)$ nach Def D 10-2 für die Klasseneinteilung

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

m=6 Klassen	n_i	h_i	x_i	$H(x)$ f. $x \geq x_i$
[-10, 20[	13	13/30	-10	13/30
[20, 50[	7	7/30	20	20/30
[50, 80[	2	2/30	50	22/30
[80, 110[	4	4/30	80	26/30
[110, 140[	4	4/30	110	30/30
[140, 170[	0	0	140	30/30
Summen	30	1.00	210	

10.2.3. Parameter 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.

Beispiel: Temperaturwerte werden an einer Meßstation stündlich erhoben. Die mittlere Tagestemperatur ist eine Kennzahl, die die Gesamtheit von jeweils 24 Messungen charakterisiert. Die 24 Messungen bilden eine **Stichprobe** und man definiert folgende Parameter (Kennzahlen):

Def D 10-5 Mittelwert, Median und p-Quantil

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

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

Für den Median einer Stichprobe X={$x_1,\ldots,x_n$} ordnet man die x_i zunächst der Größe nach: $x'_1 \le x'_2 \le \ldots \le x'_n$. 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 $m(X) = \frac{x'_{n/2} + x'_{n/2+1}}{2}$.

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

Anmerkungen:

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

Beispiel in Vorlesung.

Def D 10-6 Varianz und Standardabweichung

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

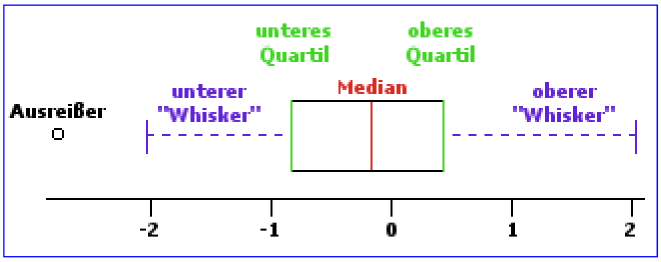
Def D 10-7 Interquartilsabstand IQR

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.

10.2.4. Boxplot: Visualisierung einer Stichprobe

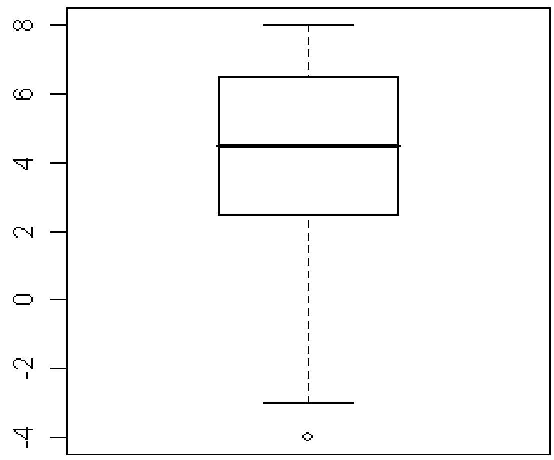


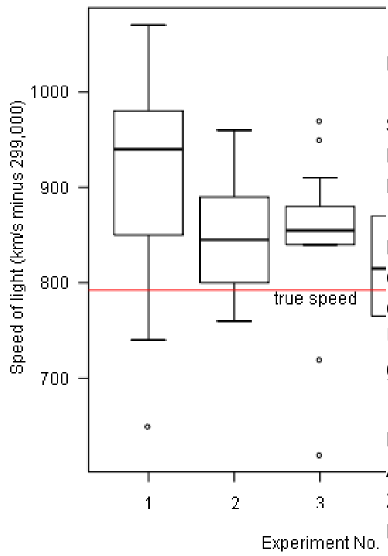
Der Boxplot ist eine kompakte Methode, die wesentlichen Parameter einer Datenreihe in einem Bild zu visualisieren:

- 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 Whisker beträgt maximal das 1.5-fache des Interquartilabstandes 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



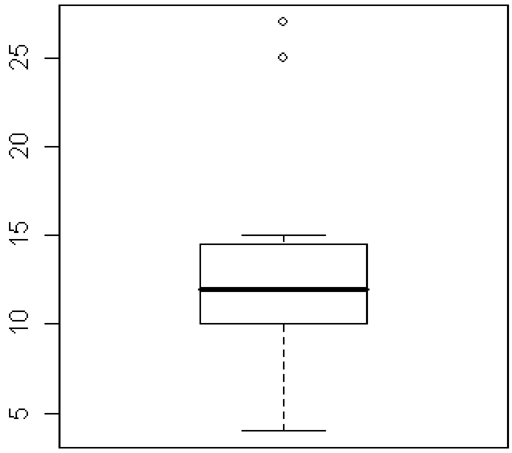


Hat man mehrere Messreihen, so kann man mit mehreren Boxplots nebeneinander schnell einen Überblick gewinnen. Nebenstehende Abbildung zeigt mehrere Experimente (jedes mit zahlreichen Einzelmessungen) zum Michelson-Morley-Versuch, der die Lichtgeschwindigkeit

299.792 km/s hochgenau bestimmt.

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

[Stingl03, S. 606ff. + Hartmann04, S. 387ff.]

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 10-5

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

Def D 10-8 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. Die Menge der möglichen Versuchsausgänge heißt **Ergebnismenge Ω** und die Menge aller Teilmengen von Ω heißt **Ereignismenge $\mathcal{A}$**.

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}. Zufällige Ereignisse sind
   A = {6} "Würfeln einer Sechs"
   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 10-9 Wahrscheinlichkeitsmaß

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

1. $0 \le P(A) \le 1$
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 10-9:

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

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

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

Es gibt einen endlichen Ergebnisraum Ω, dessen Elementarereignisse (s. Def D 10-8*) alle **gleichwahrscheinlich** sind.*

Satz S 10-3 Laplacesche Wahrscheinlichkeiten

Im *Laplaceschen Spezialfall* gilt:

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

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 Vorlesung

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 D10-10: Für n, k $\in \mathbf{N}_0$ mit k≤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.

Permutationen

Eine **geordnete Stichprobe** ist eine Liste, bei der es auf die Reihenfolge ankommt: [1,2,5] und [5,1,2] sind verschiedene Listen.

Eine **ungeordnete Stichprobe** ist eine Menge, bei der es auf die Reihenfolge der Elemente NICHT ankommt: {1,2,5} und {5,2,1} sind dieselben Mengen.

Die verschiedenen Listen, die man aus einer k-elementigen Menge bilden kann, nennt man **Permutationen**. Es gibt k! solcher Permutationen. (s.u., Ziehen einer k-elementigen Liste aus einer k-elementigen Menge ohne Zurücklegen)

Ob zwei Listen durch Permutation auseinander hervorgehen, kann man entscheiden, indem man ihre Elemente gemäß einer beliebigen Ordnungsrelation ordnet und prüft, ob die geordneten Listen gleich sind.

Beispiele:

1. [1,3,5,2,7] ist Permutation von [7,2,5,1,3], weil
   [1,2,3,5,7] = [1,2,3,5,7].
2. [1,3,5,2,7] ist KEINE Permutation von [1,3,5,7,5], weil
   [1,2,3,5,7] $\neq$ [1,3,5,5,7].

Satz S10-4 Stichproben
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:

	geordnet	ungeordnet
Ziehen mit Zurücklegen	n^k	$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$
Ziehen ohne Zurücklegen	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$

Spezialfall Ziehen ohne Zurücklegen und k=n:

1. Es gibt $\frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$ Möglichkeiten, eine n-elementige Liste aus einer n-elementigen Menge zusammenzustellen (Anzahl der Permutationen)
2. Es gibt $\binom{n}{n} = 1$ Möglichkeit, aus einer n-elementigen Menge eine n-elementige Menge zu ziehen (klar).

Beweis von Satz S10-4 in Vorlesung!

Anwendungsbeispiel: Binomischer Satz

Ein wichtiger Anwendungsfall der Kombinatorik ist der Binomische 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 in Vorlesung!

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: Wir nummerieren alle Kugeln gedanklich durch, dann haben wir wieder lauter unterscheidbare Objekte und können Satz S10-4 anwenden. Es gibt $\binom{30}{4}$ Ziehungen überhaupt. Wieviele Fälle von diesen sind für unseren Wunschergebnis günstig? 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. Setzt man beides ins Verhältnis, so erhält man

$$\left(\frac{10\cdot 9\cdot 8}{1\cdot 2\cdot 3}\cdot 20\right)/\left(\frac{30\cdot 29\cdot 28\cdot 28}{1\cdot 2\cdot 3\cdot 4}\right)=8.757\%$$

Ü1. *(a) Wieviele 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?*

Ü3. *Beim Lotto werden 6 aus 49 Zahlen gezogen. (a) Wieviele 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. Wieviele verschiedene Wahlausgänge gibt es?*

Da die Wahl geheim ist, ist die Stichprobe ungeordnet. Weil jeder Wahlmann/jede Wahlfrau aus der gleichen 3er-Kandidatenliste wählen kann, ist es Ziehen mit Zurücklegen. Es gibt also nach Satz S10-4, Nr. 4.

$$\binom{n+k-1}{k}=\binom{62}{60}=\frac{62\cdot 61}{2\cdot 1}=1891 \text{ Wahlausgänge.}$$

<u>Fazit Urnenexperimente:</u> Es gibt also folgende Systematik der Anwendungsfälle:

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

Eine weitere wichtige Anwendungen sind Qualitätsprüfungen durch Stichproben:

Beispiel: Bei einer Lieferung von 100 Rohren dürfen nur weniger als 15% vom Normdurchmesser um mehr als 1mm abweichen. Zur Überprüfung wird eine Stichprobe von N=4 Rohren entnommen und vermessen. Wie groß ist die Wahrscheinlichkeit, dass eine "schlechte" Lieferung mit 15% Ausschuss akzeptiert wird, obwohl in der Stichprobe kein fehlerhaftes Rohr war? Um wieviel sinkt diese Wahrscheinlichkeit, wenn man auf N=6 erhöht?

Lösung: Es handelt sich um Ziehen ohne Zurücklegen. Die Lieferung enthält 15 schlechte und 85 gute Rohre. Es gilt

$$P(N=4)=\frac{\binom{85}{4}}{\binom{100}{4}}=\frac{85\cdot 84\cdot 83\cdot 82}{100\cdot 99\cdot 98\cdot 97}=51.6\% \qquad P(N=6)=\frac{85\cdot 84\cdot 83\cdot 82\cdot 81\cdot 80}{100\cdot 99\cdot 98\cdot 97\cdot 96\cdot 95}=36.6\%$$

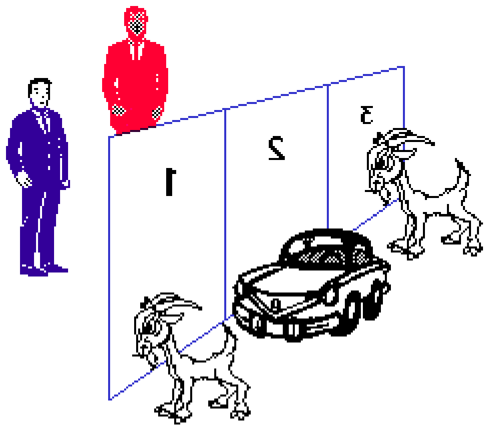
Es gibt viele weitere Anwendungen, die wir z.T. in den Übungen besprechen:

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?

10.3.3. Bedingte Wahrscheinlichkeiten

Motivation: 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

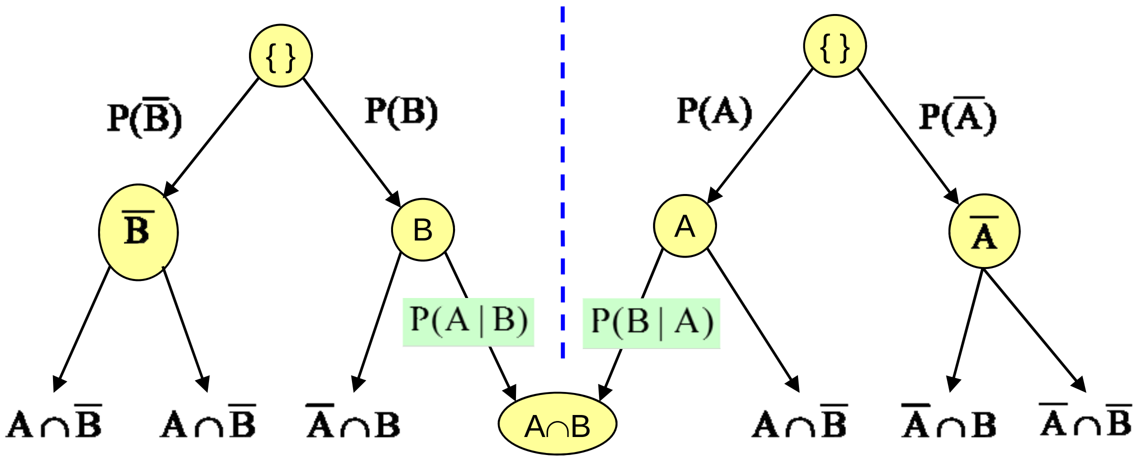
Beispiel Skatspiel: (wird in Vorlesung näher erläutert)

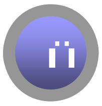
Beim Skatspiel bekommen von 32 Karten 3 Spieler je 10 Karten, 2 Karten wandern in den Stock. Sei

A = "Alex hat das Pik-As"

B = "Ich habe das Pik-As nicht"

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





<u>Übung:</u> Überlegen Sie, welche konkreten Zahlen beim Skatspiel zu den 4 Wahrscheinlichkeiten P(A), P(B), P(B|A), P(A|B) gehören!

Den für uns wichtigen Teil aus den obigen Entscheidungsbäumen können wir wie folgt zusammenfassen:

Es gilt die Formel

Satz S10-6 Multiplikationssatz für Wahrscheinlichkeiten

(*) $$P(B\mid A)P(A) = P(A\cap B) = P(A\mid B)P(B)$$

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: $1\cdot\frac{10}{32} = \frac{10}{32} = \frac{10}{22}\cdot\frac{22}{32}$

Gleichwertig zu Satz S10-6 ist die Definition:

Def D10-11 Bedingte Wahrscheinlichkeit

Seien A und B zwei Ereignisse mit P(B) ≠ 0. Dann heißt

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

die **bedingte Wahrscheinlichkeit** von A unter der Bedingung B.

Satz S10-7 Bayes-Formel

Seien A und B zwei Ereignisse mit P(B) ≠ 0. Dann gilt:

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

Bew.: folgt unmittelbar aus Formel in Satz S10-6.

Satz S10-8 Satz von der totalen Wahrscheinlichkeit

Seien B_1, B_2, ... Ereignisse, die sich paarweise ausschließen und sei $\bigcup_i B_i = \Omega$. Dann gilt:

$$P(A) = \sum_i P(A \mid B_i) P(B_i)$$

Beweis in Vorlesung [evtl. über Bild? verschiedene Pfade über die B_i, Summenzeichen, dann Ereignis B]

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?
2. Wie groß ist die Wahrscheinlichkeit, dass ein fehlerhaftes Teil von Lieferanten 1, 2 oder 3 stammt?

Def D10-12 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 $\Leftrightarrow$ $P(B|A) = P(B)$.

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) und P(B|A), 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.

[Lösung in den Übungen]

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

10.3.4. Zufallsvariablen

[Stingl, S. 619.624], [Hartmann, S. 404-418] oder [Teschl05, Bd. 2, S. 245-280]

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?

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

Def D10-13 Zufallsvariable, Verteilungsfunktion

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

Wenn X nur abzählbar viele Werte annehmen kann, spricht man von einer **diskreten Zufallsvariablen**. Wenn X beliebige Werte aus einem reellen Intervall annehmen kann, spricht man von einer **stetigen Zufallsvariablen**.

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

Beispiele und Anmerkungen:

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

Tabelle 10-6 Augensumme zweier Würfel

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

Übung: Füllen Sie den Rest der Tabelle aus!

- 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 $X\colon \Omega \rightarrow\ ]0,12]$ ist eine reelle Zufallsvariable.

- 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"). Dagegen ist die Wahrscheinlichkeit, dass der Stundenzeiger zw. "12" und "1" steht, gegeben durch $F(1) = P(X \le 1)$ = 1/12.

Satz S10-9 Eigenschaften der Verteilungsfunktion

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

Punkt 1. ist das Wahrscheinlichkeitsaxioms Def D 10-9, Nr. 2, verallgemeinert für Zufallsvariablen: Wenn wir die Grenze t gegen $+\infty$ verschieben, haben wir das sichere Ereignis: $\lim\limits_{t\to+\infty} F(t) = P(X \le \infty) = P(\Omega) = 1$. Wenn wir die Grenze t gegen $-\infty$ verschieben, haben wir das unmögliche Ereignis: $\lim\limits_{t\to-\infty} F(t) = P(X \le -\infty) = P(\{\}) = 0$

Bew. zu 2.: $P(X \le a) + P(a < X \le b) = P((X \le a) \vee (a < X \le b)) = P(X \le b)$. Die 1. Umformung gilt, weil $(X \le a)$ und $(a < X \le b)$ unvereinbare Ereignisse sind (s. Def D 10-9, 3. Wahrscheinlichkeitsaxiom)

Punkt 2. besagt: Kennen wir die Verteilungsfunktion, so können wir die Wahrscheinlichkeit für jedes Intervall]a,b] bequem angeben.

Die Formel in Satz S10-9, Punkt 2. sieht verdächtig nach einem bestimmten Integral aus, man kann sich fragen, ob es eine Funktion gibt, deren Stammfunktion die Verteilungsfunktion ist. Dies ist mit der <u>*Wahrscheinlichkeitsdichte*</u> in der Tat der Fall und so ist die Wahrscheinlichkeitstheorie ein wichtiger Anwendungsfall für die Integralrechnung:

Def D10-14 Wahrscheinlichkeitsdichte

Für eine stetige Zufallsvariable $X: \Omega \to \mathbf{R}$ heißt eine integrierbare, nichtnegative reelle Funktion $w: \mathbf{R} \to \mathbf{R}$ mit $F(x) = P(X \le x) = \int_{-\infty}^{x} w(t)dt$ die **Dichte** oder **Wahrscheinlichkeitsdichte** der Zufallsvariablen X.

Anmerkungen:

- Die Verteilungsfunktion $F(t)$ ist also eine Stammfunktion zur Wahrscheinlichkeitsdichte $w(t)$.
- Obwohl gilt $1 = \lim\limits_{t\to\infty} F(t) = \lim\limits_{t\to\infty} \int_{-\infty}^{t} w(u)du = \int_{-\infty}^{\infty} w(u)du$, kann $w(u)$ an einzelnen Stellen u sehr wohl größer als 1 werden.

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

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

Ähnlich wie in Kapitel 10.2.3 für Mittelwert Varianz einer Stichprobe, definieren wir hier Erwartungswert und Varianz einer Zufallsvariablen:

Def D10-15 Erwartungswert einer Zufallsvariablen

Für eine diskrete Zufallsvariable $\mathbf{X}:\ \Omega \to \mathbf{R}$, die Werte $x_m \in M$ annehmen kann, seien $w_m = P(X = x_m)$ die Wahrscheinlichkeiten. Der Erwartungswert μ ist definiert durch:

$$\mu = E(X) = \sum_{x_m \in M} x_m w_m$$

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

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

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

Satz S10-10 Linearität des Erwartungswertes

Für Zufallsvariablen X,Y und reelle Zahlen $a,b \in \mathbf{R}$ gilt der wichtige Satz

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

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

Def D10-16 Varianz einer Zufallsvariablen

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

$$\sigma^2 = Var(X) = E\left[(X - \mu)^2\right]$$

Dies gilt gleichermaßen für diskrete und stetige Zufallsvariablen.

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

Anmerkung und Beispiele:

- Der Erwartungswert für die Augensumme bei zwei Würfeln ist (s. Tabelle 10-6):

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

- Eine in $[0,a]$, $a>0$ gleichverteilte Zufallsvariable X hat innerhalb des Intervalls die konstante Wahrscheinlichkeitsdichte $w(t) = 1/a$ und ist ausserhalb gleich Null (klar? [zeichnen]). Der Erwartungswert und die Varianz sind:

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

$$\sigma^2 = V(X) = \int_0^a (t-\frac{a}{2})^2\cdot\frac{1}{a}dt = \frac{1}{a}\cdot\left[\tfrac{1}{3}\left(t-\frac{a}{2}\right)^3\right]_0^a = \frac{1}{a}\cdot\frac{2}{3}\frac{a^3}{8} = \frac{a^2}{12}$$

- 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> [Ü3 + Ü4: Lösung in den Übungen]

Ü1. *Bewerten Sie, ob es sich lohnt, an folgendem Spiel teilzunehmen, indem Sie den Erwartungswert für X = "Gewinn – Einsatz" ausrechnen: Beim Würfeln mit zwei Würfeln erhält man einen Gewinn von 20€ für "Augensumme 12" und 5€ für "Augesumme 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?

d	Wert x_m von X	Ereignisse ω	$P(X= x_m)$
5	28 €	(6,1), (1,6)	2/36
4	8 €	(6,2), (2,6), (5,1), (1,5)	4/36
<4	- 2 €	der Rest	30/36

Ü3. *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) und bestimmen Sie die Konstante α!
(b) Welchen Mittelwert E(X) hat die Zufallsvariable X?
(c) Überlegen Sie eine sinnvolle Definition des Median t_m für stetiges X und berechnen Sie t_m für die konkrete Dichte w(t).

Ü4. *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?*

Lösung in Vorlesung!

10.3.5. Wichtige Verteilungen

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 Binomialvert.
	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 D10-17 Bernoulli-Experiment

Ein **Bernoulli-Experiment** ist ein Zufallsexperiment, bei dem es nur zwei Ausgänge gibt: Ereignis das Ereignis A tritt ein (Wahrscheinlichkeit p) oder nicht, also tritt Ereignis $\overline{A}$ ein (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? Dies ist gleichwertig zu einem Urnenexperiment mit W+S=N weißen und schwarzen Kugeln, mit p=W/N und 1-p = S/N und A="Ziehen einer weißen Kugel, **mit Zurücklegen**".

Sei X die Zufallsvariable, die das Eintreten von A zählt. Wie wahrscheinlich ist $P(X=k)$?

Der Ausgang eines n-fachen Experimentes ist $\underbrace{\left(\overline{A},A,A,\overline{A},\ldots,A,\overline{A},A\right)}_{n-\text{mal}}$. Wieviele solcher Ereignisse gibt es mit k-mal A? – Nach Satz S10-4 gibt es dafür genau $\binom{n}{k}$ Möglichkeiten (klar? – Ziehen der A-Positionen aus der Positionsmenge {1,2,...,n}). Wie wahrscheinlich ist jedes dieser Ereignisse "k-mal A"? – Offensichtlich ist die Wahrscheinlichkeit $p^k(1-p)^{n-k}$.

Satz S10-11 Binomialverteilung

Gegenben 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 D10-10)

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 [Hartmann04, S. 421]. Er ist nicht schwer.

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 (Gaußverteilung, s.u., Satz S10-16, oder Poissonverteilung für große n und kleine p).

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 S10-12 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 D10-10)

Für $N>>n$ 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 D10-15 notiert. Es gilt

Satz S10-13 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 $E(X) = \frac{a+b}{2}$

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

Anmerkungen

- Für [0,1]-gleichverteilte Zufallsvariablen gilt also Erwartungswert 0.5 und Varianz = 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 diskrete 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 D10-18 Normalverteilung (Gaussverteilung)

Eine stetige Zufallsvariable **X: $\Omega \rightarrow$ R** heißt **normalverteilt mit Mittelwert μ und Standardabweichung σ** oder kurz N(μ,σ)-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 [μ-σ, μ+σ]. (Beweis s. Ü1)

Für praktische Anwendungen braucht man neben der Dichte auch die Verteilungsfunktion $F(t) = P(X \leq t)$ (s. Def D10-13). 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 <u>eine</u> Tabelle zurückführen:

Def D10-19 Standardnormalverteilung, Verteilungsfunktion $\Phi(x)$

Die Normalverteilung $N(0,1)$ 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

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

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

[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!]

<u>Beispiel:</u> Man bestimme aus Tabelle 10-8 das q-Quantil für $q=0.9$.

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

<u>Lösung mit "linearer Interpolation":</u> 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 S10-14 Regeln für Normalverteilungen

1. $\Phi(-z) = 1 - \Phi(z)$.

2. Ist X eine $N(\mu,\sigma)$-verteilte Zufallsvariable, so ist $Z = \dfrac{X-\mu}{\sigma}$ $N(0,1)$-verteilt.
3. Für die Verteilungsfunktion $F(b)=P(X\leq b)$ gilt: $F(b) = \Phi\left(\dfrac{b-\mu}{\sigma}\right)$
4. $P(a < X \leq b) = \Phi\left(\dfrac{b-\mu}{\sigma}\right) - \Phi\left(\dfrac{a-\mu}{\sigma}\right)$
5. $q = \Phi(z_q) \quad \Leftrightarrow \quad 1-q = \Phi(-z_q)$
6. Ist $\mathbf{z_q}$ das q-Quantil einer $N(0,1)$-Verteilung, so ist $x_q = \sigma \cdot z_q + \mu$ das q-Quantil einer $N(\mu,\sigma)$-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

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

Diese Umformung gilt wg. Satz S10-14, Nr. 5. Der Tabelle 10-8 entnehmen wir $-z_q = 1.56$ (nächstgelegener Wert, ohne lineare Interpolation).

Nach Satz S10-14, Nr. 6 ist dann das Quantil x_q der $N(1.75, 0.2)$-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.

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

Ü2. *X sei eine $N(\mu,\sigma)$-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 durchschnittle Wartezeit zwischen zwei Hacker-Attacken auf Ihrem zentralen Server sei N(48h,6h)-verteilt. Gerade ist eine Attacke passiert. In welchem Zeitintervall ist mit 82% mit der nächsten Attacke zu rechnen?*

10.3.6. Der zentrale Grenzwertsatz

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 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?

Die Rettung kommt in Form der Normalverteilung. Sie ist – und das ist ein ***wichtiges*** und tiefliegendes Resultat der Statistik – die Grenzverteilung für viele wichtige Zufallsexperimente ist, die sich entweder sonst nur schwer ausrechnen lassen und/oder die oft vorkommen:

Satz S10-15 Der zentrale Grenzwertsatz

Seien $X_1, X_2, \ldots, X_n$ unabhängige Zufallsvariablen, die alle die gleiche Verteilung mit Erwartungswert μ und Varianz σ^2 haben. Sei $S_n = X_1 + X_2 + \ldots + X_n$ die Summe, eine Zufallsvariable mit Erwartungswert $n\mu$ und Varianz $n\sigma^2$. Dann konvergiert diese Zufallsvariable gegen eine gemäß $N(n\mu, \sqrt{n\sigma^2})$ verteilte Zufallsvariable, das heißt

$$\lim_{n\to\infty} P(S_n \le x) = \lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le z\right) = \Phi(z) \quad \text{mit} \quad z = \frac{x - n\mu}{\sqrt{n\sigma^2}}$$

Die Konvergenz erfolgt recht schnell, schon für $n \geq 30$ können wir meist die Rechenregel

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le z\right) \approx \Phi(z)$$

(ohne Limes) mit guter Genauigkeit anwenden.

Anmerkungen:

- o Für <u>alle</u> Experimente und Messungen, die n-mal unabhängig wiederholt werden, trifft dieser Satz zu.
- o Man beachte, dass der Satz völlig unabhängig von der Art der Verteilung gilt, die die X_i haben (!!). <u>Jede</u> Verteilung strebt bei n-facher Wiederholung und Summation gegen die Normalverteilung.

Ein wichtiger Spezialfall ist der Satz von Moivre-Laplace, der die Binomialverteilung behandelt:

Satz S10-16 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:

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

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

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<550)?*

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

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
- 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 Gaussverteilung (Normalverteilung) approximierbar ist
  - was die **Normalverteilung** 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 Java-Applets zu Statistik-Grundlagen:

[1] Charles Stanton: *Java Demos for Probability and Statistics*, www.math.csusb.edu/faculty/stanton/m262/probstat.html

Das wichtige Gebiet der schließenden Statistik (manche sagen: hier fängt die Statistik erst an) haben wir in dieser Einführung nicht behandelt. Themen hier 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**: Chi-Quadrat-Schätzungen
  - **Anpassungstests**: Welche Dichtefunktion beschreibt meine zahlreichen Zufallsexperimente bestmöglich? (Anpassen einer Funktion an Daten)

Lit: [Hartmann04, S. 443-465], [Stingl02] oder [Teschl05, Bd. 2, S. 325-357]