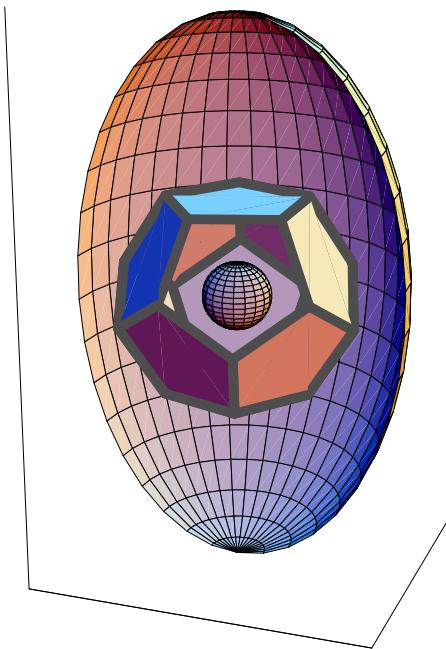


Script ◊ Math ◊ Ing
◊ Wahrscheinlichkeit & Statistik ◊
◊ Anhang ◊



Scripta bilingua

von

Rolf Wirz

Berner Fachhochschule BFH ◊ TI und AHB

V.1.12.23a d / 7. August 2012 **Deutsche Version**

Produziert mit LaTeX auf Win XP.

Einige Graphiken sind auch mit *Mathematica* entstanden.

Glück hilft manchmal, Arbeit immer ...

Brahmanenweisheit

Aktuelle Adresse des Autors (2007):

Rolf W. Wirz-Depierre
Prof. für Math.
Berner Fachhochschule (BFH), Dep. AHB und TI
Pestalozzistrasse 20
Büro B112 CH-3400 Burgdorf/BE
Tel. ++41 (0)34 426 42 30 / intern 230
Mail: Siehe <http://rowicus.ch/Wir/indexTotalF.html> unter „Koordinaten von R.W.“
Alt: Ingenieurschule Biel (HTL), Ing'schule des Kt. Bern, Fachhochschule ab 1997) // BFH HTA Biel // BFH TI //

©1996/2001/02/03/04/05/06/07/08/09/10/11/12

Vor allem die handgefertigten Abbildungen sind früheren öffentlichen Darstellungen des Autors entnommen. Die Urheberrechte dafür gehören dem Autor privat.

Die hier noch verwendete LaTeX-Version erlaubt keine Änderung der Überschrift des Inhaltsverzeichnisses ...

Inhaltsverzeichnis

A Aus dem DIYMU und einem Crashkurs	3
B Zum Begriff der Wahrscheinlichkeit	5
B.1 Über „Wahrscheinlichkeit“ und „Zufall“	5
B.1.1 Wahrscheinlichkeit und Zufall	5
B.2 Zur Entwicklungsgeschichte des Wahrscheinlichkeitsbegriffs	8
B.2.1 Die Frage nach der Gewinnchance des Chevaliers de Méré	8
B.2.2 Klassischer contra statistischer Wahrscheinlichkeitsbegriff	8
C Zum Begriff der Kombinatorik	15
C.1 Übersicht über die elementaren kombinatorischen Fälle	15
C.2 Formeln für die elementaren kombinatorischen Fälle	15
C.2.1 Anordnungen oder Permutationen ohne Wiederholung	15
C.2.2 Permutationen mit Wiederholung	17
C.2.3 Variationen ohne Wiederholung	18
C.2.4 Kombinationen ohne Wiederholung	19
C.2.5 Variationen mit Wiederholung	20
C.2.6 Kombinationen mit Wiederholung	21
C.3 Eine Aufgabe	21
C.4 Links zur Fortsetzung	22
D Fehler von statistischen Kenngrößen und Standardfehler	23
D.1 Fehler von statistischen Kenngrößen	23
D.1.1 Die Fragestellung	23
D.1.2 Fehler des Mittelwerts	24
D.1.3 Fehler der Standardabweichung	24
D.1.4 Beispiel	25
D.2 Der Standardfehler von statistischen Kenngrößen	25
D.2.1 Die Fragestellung	25
D.2.2 Einige Begriffe	26
D.2.3 Die Berechnung des Erwartungswerts	27
D.2.4 Formeln für den Erwartungswert	28
D.2.5 Formeln für die Varianz	29
D.2.6 Formel für den Erwartungswert des Mittelwerts	30
D.2.7 Nochmals Formeln für die Varianz	30
D.2.8 Formel für die Varianz des Mittelwerts	30
D.3 Messfehler und Standardfehler	31
D.4 Einige Links	31
D.5 Ein Beispiel	31
D.6 Der Erwartungswert eines Funktionswerts	32
D.7 Standardabweichung und Tschebyschow–Ungleichung	33

D.7.1	Zur Standardabweichung	33
D.7.2	Die Tschebyschow–Ungleichung	34
D.8	Messfehler (funktionale Abhangigkeit)	36
D.8.1	Das Problem der Verpflanzung — Le probleme de la dependance	36
D.8.2	Verwendung des totalen Differentials — Appliquer la differentielle totale	36
E	Monte-Carlo, Resampling und anderes	39
E.1	Monte-Carlo-Simulationen	39
E.1.1	Berechnung von Pi, Messeexperiment	39
E.1.2	Das Konzept der Monte-Carlo-Simulation resp. –Studie	40
E.1.3	Berechnung von Pi, Monte-Carlo–Flachensimulation	40
E.1.4	Berechnung von Pi, Buffon, Nadelexperiment	41
E.2	Resampling-Methoden (Einführung)	44
E.2.1	Die Idee von Bootstrap und Jackknife	44
E.2.2	Resampling, Beispiele	46
F	Eine Bootstrap–Anwendung Schritt fur Schritt	53
F.1	Aufgabenstellung und Konzept	53
F.1.1	Das praktische Problem der Verteilungsfunktion	53
F.1.2	Eine mogliche Aufgabenstellung	53
F.2	Beispiel 1: Generierung einer Zahlenmenge als Messwerte	53
F.3	Beispiel 1: Bearbeitung der Menge von Messwerten	55
F.4	Beispiel 2: Einlesen und bearbeiten von Messwerten, neue Messreihe	60
G	Anhang: Datensatznderung	63
G.1	Die Problemstellung	63
G.1.1	Beurteilung einer Korrekturmglichkeit	63
G.2	Die Minimaleigenschaft des Mittelwerts	64
G.2.1	Varianz, Standardabweichung und Wahrscheinlichkeitsfunktion	64
G.2.2	Minimaleigenschaft des Mittelwerts	65
G.3	Verkleinerung der StD bei Erweiterung eines Datensatzes	65
G.3.1	Verkleinerung	65
G.3.2	Konsequenzen fur den Mittelwert	66
G.4	Vergrosserung der StD bei Erweiterung eines Datensatzes	66
G.5	Umrechnung des Mittelwerts und der empirischen StD	68
G.5.1	Eine Formel fur den Mittelwert	68
G.5.2	Eine Formel fur die empirische Varianz	68
G.5.3	Anwendung: Berechnung einer veranderten StD	69
G.6	Ein Beispiel	70
G.7	Übung	71
H	Anhang: Spezielle Wahrscheinlichkeitssituationen	73
H.1	Kreuztabellen, Beispiel	73
H.2	Ein Beispiel mit einem Taxi, das Fragen aufwirft	75
H.3	„Wie alt ist der Kapitan?“	79
I	Hinweise zur Datenanalyse	81
I.1	Grundfragen — Modellierungen	81
I.1.1	Wichtige Ablufe	81
I.1.2	Vorgehen bei Modellierungen	82
I.1.3	Hypothesenmodellierung, Hypothesentest	83
I.1.4	Das Problem der Wechselwirkungen zwischen Variablen	84
I.2	Wirklichkeit, mathematische Modelle, Wesentlichkeit der Entscheide	84

I.2.1	Das Problem der sinnvollen Frage	84
I.2.2	Galilei und Archimedes	85
I.2.3	Extrapolation und mathematisches Modell	87
I.2.4	Wozu ein Modelle?	88
I.2.5	Wirklichkeit, Wahrscheinlichkeitsmodell und Entscheidung	90
I.3	Mathematische Modelle und Parameter	90
I.3.1	Zur Grundgesamtheit	90
I.3.2	Zur Stichprobe	91
I.3.3	Wahrscheinlichkeit, Verteilungsfunktion, Wahrscheinlichkeitsmodell	91
I.3.4	Prozesskontrolle, Prozesstauglichkeit	93
I.3.5	Zu den Box–Whisker–Plots	93
I.3.6	Tschebyschow–Ungleichung	93
I.4	Zu den Datenerhebungskonzepten	94
I.4.1	Wo und wieso Daten erheben?	94
I.4.2	Zur Datenqualitätssteuerung	95
I.4.3	Placebo–Effekt, Doppelblindstudien, Monitoring	97
I.4.4	Grundfragen und Datenerhebungsprinzipien	97
I.4.5	Statistiken	100
I.4.6	Konstruktion von Wahrscheinlichkeitsmodellen mit Statistiken	101
I.4.7	Zu den Fehlern bei Schätzungen	102
I.4.8	Verteilungsfreiheit und Vertrauensintervalle	104
I.4.9	Nochmals zu statistischen Tests	107
I.4.10	Bemerkung zu Darstellungsmethoden der explorativen Statistik	110
I.5	Nochmals zu den Bootstrap–Methoden	112
I.5.1	Zum Percentil–Lemma	112
J	Zusammenfassungen und Ausblicke	115
J.1	Parametertests — Signifikanztests	115
J.1.1	Das Problem des Lernenden	115
J.1.2	Das Denkgerüst hinter dem Test	115
J.2	Parameterschätzung und Maximum–Likelihood–Methode	116
J.2.1	Die Idee	116
J.2.2	Ein Beispiel zur Veranschaulichung	117
J.2.3	Erwartungstreue, Konsistenz, Effizienz	117
J.2.4	Wie weiter?	118
J.3	Ausblick: Weitere Links zu Stochastik und Statistik	118

Vorwort

Der Sinn dieses Anhangs:

Liebe Leserin, lieber Leser,

Dieser Anhang wurde erzeugt, weil im Skript „Wahrscheinlichkeit & Statistik“ Probleme mit der Kapazität von PC-TeX aufgetreten sind, welche mit den momentan zur Verfügung stehenden technischen Mitteln nicht befriedigend lösbar sind.

Im Herbst 2008

Der Autor

Anhang A

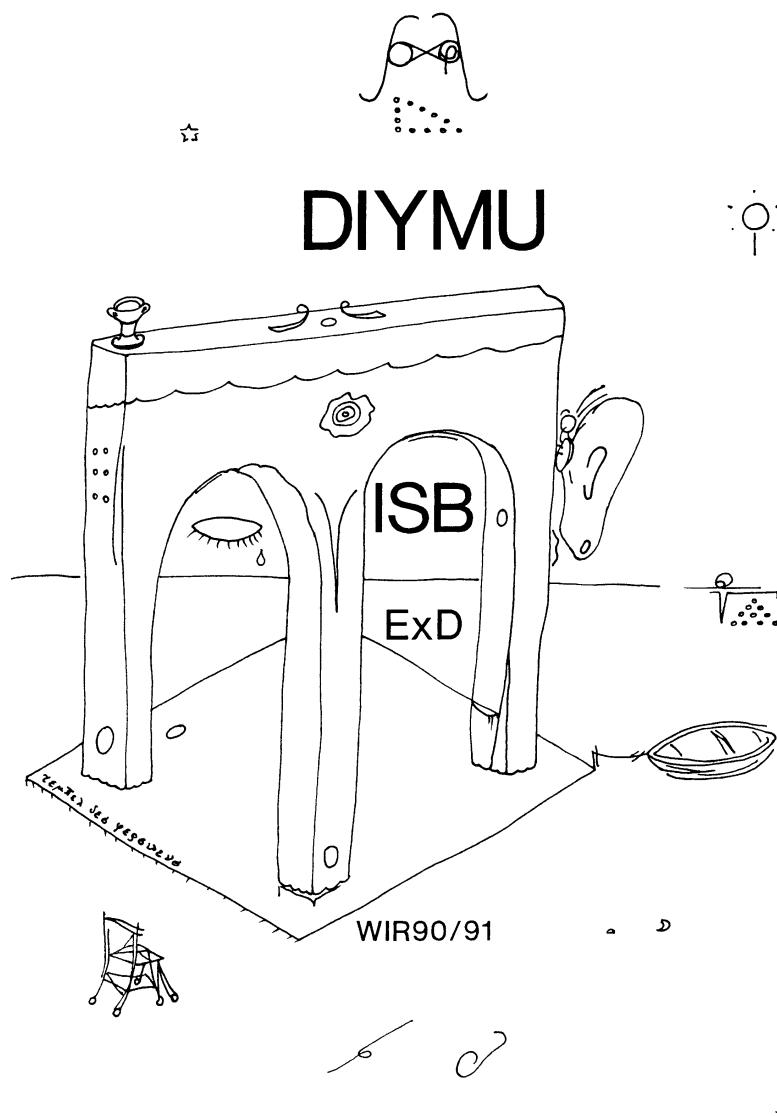
Aus dem DIYMU und einem Crashkurs

Übungen

<http://rowicus.ch/Wir/TheProblems/Problems.html>

<http://rowicus.ch/Wir/Scripts/TEIL6dCrashKursWahrschKomb.pdf>

Der Inhalt des Kurses „TEIL6dCrashKursWahrschKomb.pdf“ wird in den beiden nachfolgenden Kapiteln wiedergegeben.



Anhang B

Zum Begriff der Wahrscheinlichkeit

B.1 Über „Wahrscheinlichkeit“ und „Zufall“

B.1.1 Wahrscheinlichkeit und Zufall

Was ist mit „Wahrscheinlichkeit“ gemeint?

„Wahrscheinlichkeit“ ist das Dingwort (Substantiv) zum Eigenschaftswort (Adjektiv) „wahrscheinlich“. Dieses erkennt man als Zusammenzug von „wahr scheinlich“ in der Bedeutung von „wahr scheinen“. „Wahr scheinen“ aber steht im Gegensatz zu „wahr seiend“. Mit „Wahrscheinlichkeit“ meint man daher, dass die betreffende Sache nicht unbedingt wahr sei, sondern nur so scheine. Die Sache könnte daher auch durchaus falsch sein. Die dabei angegebene Wahrscheinlichkeitszahl steht als Gradmesser der Erwartung von Wahrheit. Bei einer hohen Wahrscheinlichkeitszahl erwartet man das Eintreffen der angesprochenen Sache sehr stark. Bei einer tiefen dagegen ist die Erwartung sehr schwach. Über die Sicherheit wird dabei im Einzelfalle nichts ausgesagt, denn es könnte alles auch anders eintreffen, als man es erwartet.

Was ist mit „Zufall“ gemeint?

Bei sprachlichen Konstruktionen wie etwa bei „Wahrscheinlichkeit“ oder bei „Zufall“ taucht ein neues Problem auf: Das Problem der Sinnhaftigkeit abstrakter Begriffe. Solche Begriffe bezeichnen nicht etwa Dinge der sinnlichen Erfahrungswelt, so konkrete Begriffe etwa wie z.B. „Tisch“ oder „Teller“. Abstrakte Begriffe lernt man als Kind nicht aus der eigenen Erfahrung kennen. Sie gehören zum Kulturschatz, den man sich erst mühsam erarbeiten oder erlernen muss wie z.B. die meisten Begriffe der Mathematik. Abstrakte Begriffe haben eine Begriffsgeschichte, eine Konventionsgeschichte und können ihre Bedeutung mit der Zeit auch ändern.

Man kann sagen, dass folgender Sachverhalt wohl feststeht: Mit dem Eigenschaftswort „zufällig“ drückt der Betrachter einer Sache eine Beziehung eines Beobachters zur besagten Sache aus. „Das Unwetter kam für mich zufällig“ meint, dass der Sprecher das Eintreffen des Unwetters nicht persönlich vorausgesehen hat. Es ist ihm einfach „zugefallen“. Was andere in dieser Sache vorausgesehen haben, wird hier nicht angesprochen. Für den amtierenden Wetterfrosch z.B. muss das Unwetter keineswegs Zufall gewesen sein, denn er hatte im Voraus aus vorhandenen Messdaten sein Resultat rechnerisch ermittelt und sogar als Prognose angeboten. Unser Freund, der jetzt verregnet worden ist, konnte das Angebot deshalb nicht annehmen, weil es ihn nicht erreicht hatte. Schade.

Ähnlich verhält es sich, wenn Hans Hansen am Morgen nichts ahnend zur Haustür heraustritt und gleich von einem neben dem Türpfosten versteckt lauernden Mann eine Ohrfeige kassiert. Für Hans Hansen ist das ein unglaublicher, schlimmer, schmerzhafter Zufall, denn das hätte dieser friedfertige Mensch nie erwartet. Für den Spender der Ohrfeige war es aber gewiss Absicht, denn jener hat dem Hans Hansen schon lange aufgelauert und sein Werk mit Hochgenuss ausgeführt. Vermutlich aus Rache, weil ihn Hans

Hansen einmal auf der Strasse vor zwei Jahren öffentlich zurechtgewiesen hatte, als er seinen Fischabfall einfach auf ein fremdes Autodach geschleudert hatte. Hans musste das längst vergessen haben. Der Spender der Ohrfeige übrigens auch. Nur seine Wut auf Hans ist geblieben, denn er war in seinem tiefsten Kern schon immer herabgewürdigt wie auch seiner Freiheit beraubt worden. Die ständigen Zurechtweisungen mochte er nicht mehr. Die Zufälligkeit verhält sich daher hier relativ zum Betrachter. Für Hans war es Zufall. Für den Schläger der Ohrfeige nicht.

Zufällig ist eine Sache für jemanden daher sicher dann, wenn der Betroffene für das Eintreffen der Sache keine Regeln ausmachen kann. Das bedeutet jedoch keineswegs, dass es keine Regeln dafür gibt. Denn es gibt ja immer auch Regeln, welche man noch nicht kennt, welche man vielleicht erst in den Wissenschaften neu entdecken wird und welche man daher vor ihrer Entdeckung nicht exakt kennen konnte. Es ist geradezu die Aufgabe der Wissenschaft, immer wieder solche neuen Regeln zu entdecken. Gäbe es keine unbekannten Regeln mehr, so wäre die Entwicklung der Wissenschaft an ein Ende gekommen. Zudem kann man schon aus Komplexitätsgründen nie alles wissen. Denn es sind ja Regeln denkbar, welche sich aufgrund ihrer Kompliziertheit und Länge nie von einem menschlichen Wesen oder einer Maschine aufschreiben lassen werden, auch mit der grössten Geschwindigkeit nicht. Denn unsere Welt ist endlich, so wie auch jede in ihr vorkommende Geschwindigkeit. Nicht jeder sehr lange Prozess kann daher innerhalb unseres Universum zu unserer Zeit in unserer erreichbaren Umgebung auch ausgeführt werden, bis ein brauchbares Resultat vorliegt. Es gibt Ausdrücke, welche länger sind, als alles bei der gegebenen Endlichkeit Mögliche es je sein kann.

Wir unterscheiden daher zwischen der Menge der wissbaren und in Zeichen aufschreibbaren Fakten und der Menge der noch nicht von uns entdeckten Fakten sowie der Menge der nicht entdeckbaren Fakten, weil es dazu prinzipiell, vielleicht aus Komplexitätsgründen, keine praktische Möglichkeit geben kann. Dazu gesellen sich noch die Irrtümmer, die man ebenfalls in einer Menge von wissbaren und in Zeichen aufschreibbaren Fakten zusammengefasst sich denken kann. Nie angeben könnte ein Mensch (oder auch ein anderes Wesen nicht) die prinzipiell nicht in Sprache oder Denkinhalte fassbaren Irrtümmer. Darüber sich auslassen zu wollen macht an dieser Stelle keinen Sinn. Denn solches übersteigt die uns zustehenden Mittel in jeder Weise. Die Frage nach solchem entpuppt sich als widersprüchliche Konstruktion wie etwa die Frage nach einem ebenen viereckigen Kreis mit vier Ecken auf dem Rande, an denen spitze Winkel zu finden sind. Oder die Frage nach einem ebenen Dreieck mit einer Winkelsumme von π^2 im Bogenmass.

Eine der bekannten „Frage nach dem absoluten Zufall“ vergleichbare Frage

Frage: „Wo ist das Nichts im Innern einer leeren Büchse?“

Im Innern einer leeren Büchse hier auf Erden befindet sich normalerweise Luft. Schicken wir die Büchse offen ins Weltall, so existiert an jedem Ort im Innern dieser Büchse noch immer der Ort eines geometrischen Punktes. Dort findet man nach der Theorie auch immer die Anwesenheit materieller, physikalischer Felder. Ebenfalls verhält es sich so, wenn wir in der Büchse hier auf Erden Vakuum erzeugen.

An jedem Ort irgendwo im Bereich der Orte existiert so im Voraus immer schon ein geometrischer Ort, also nicht ein Nichts. Daher kann im Innern einer Büchse kein Nichts anzutreffen sein. Und überhaupt, könnte man ein Nichts denn je irgendwo oder irgendwie antreffen? Ist etwas, das es nicht gibt, je zu treffen, anzutreffen, womit man ja eine Begegnung von Angesicht zu Angesicht meint? Und weiter am Rande des Innern unserer Büchse: Im Material der Büchse oder gar aussen, wie verhält es sich dort mit dem Nichts?

Wir gehen davon aus, dass man die Büchse nicht aus dem Universum herausbringen kann. Man kann sie wohl zerstören, jedoch bleibt sie als Geschichte, in der Realität über die Zeit, als Gewesenes dann gleichwohl immer hier in ihrer Zeit. Gewesenes kann nicht ungewesen werden.

Abgesehen vom vermessbaren Ort haben wir im Material der Büchse wohl kein Nichts. Auch zwischen den Atomen nicht, denn wir reden heute ja von Feldern sowie auch von geometrischen Orten. Denn auch zwischen den Atomen und innerhalb dieser befinden sich Felder sowie geometrische Orte. In allfälligen Zwischenräumen existiert also immer mindestens der geometrische Ort. Auch aussen, soweit das Weltall reicht, verhält es sich so. Der Begriff des physikalischen Ortes lässt ist ferner ans Weltall gebunden, nicht

aber derjenige des geometrischen Ortes. Ein Ausserhalb des Weltalls können wir materiell nicht denken, wohl aber geometrisch, zum Beispiel schon in der euklidischen Geometrie. Ein Nichts im Örtlichen kann so nie ein Es, also ein Objekt sein. Unser Denken und unsere Sprache handelt aber immer von Objekten. Ein Nichts kann daher nie in unserer Sprache ausgedrückt oder durch jemanden gedacht werden. Die Sprache, und auch schon der Gedanke, lokalisiert das Nichts in sich als Gedankenkonstruktion oder als sprachliche Konstruktion. Formal ist das schon ein Etwas, ein gedankliches Etwas, also kein Nichts. Und wie wäre es mit einem nicht formalen Nichts? — Was nicht sprachlich ist, kann nicht angesprochen werden. Wissen über Existenz setzt Information voraus. Diese muss von irgendwo kommen. Das Nichts müsste daher Spuren hinterlassen haben. Damit wird es zum Spuren erzeugenden Etwas, bleibt also keineswegs ein Nichts.

Somit handelt es sich beim absoluten Nichts um eine sprachlich unkorrekte Konstruktion. Das sieht man auch, wenn man die Menge aller existierenden Dinge gedanklich nimmt, welche physisch real existieren oder welche einmal gedacht worden sind — und noch gedacht sein werden — um diese sich in einer Menge zusammengefasst vorzustellen. Ausserhalb der physikalisch existierenden und der gedachten Dinge befindet sich dann auch die Menge der nicht gedachten Dinge, welche in einer Menge noch möglich wären. Das ist nicht das Nichts. Eine Menge ist kein Nichts. Sie ist ein Etwas, ein Exemplar aus den Mengen.

„Nichts“ bezogen auf Mengen kann man als „nichts von“, „nichts von einer Sache“ sehen. Die Sache wäre vermutlich, sofern sie fassbar ist, als Menge B denkbar, als Teilmenge ($B \subseteq M$) einer anderen Menge M , einer Grundmenge somit, bezüglich der „nichts von jener Sache“ das Komplement, also die Mengendifferenz der „Sache“ A bezüglich der Grundmenge erscheint: $B = \bar{A} = M \setminus A$. Als Komplement ist immer minimal die leere Menge $\{\}$ denkbar, eine Menge ohne Elemente, aber immer noch eine Menge mit der Mächtigkeit $|\{\}| = 0$ und daher nicht ein Nichts. Sie enthält nur „nichts von den unerwünschten Elementen“, also keines von den Elementen $a_k \in A$.

Auch die leere Menge ist nicht das Nichts, denn sie hat schon eine Mächtigkeit wie eben festgestellt, nämlich die Zahl null. Und selbst in der leeren Menge ist kein Nichts drin, denn sie ist leer von allem, daher also auch vom Nichts. Man landet hier bei einer Konstruktion wie derjenigen von Bertrand Russel, welche von einem Barbier handelt, der sich selber rasiert. Russel zeigt auf, dass bei dieser Konstruktion Regeln zur Anwendung kommen, welche man für die Benennungen von konkreten physikalischen Dingen wohl gebrauchen darf, die aber nicht automatisch auch für abstrakte Dinge Gültigkeit haben. Der entlarvt damit eine naive sprachliche Konstruktion im Abstrakten bezüglich Sinnhaftigkeit als Unsinn.

Das Nichts ist damit kein denkbarer Begriff, keine existierende Sache, sondern eine sprachliche, unüberlegte widersprüchliche Konstruktion einer abstrakten Sache, welche das einfache Denken in seiner Konsequenz weit übersteigt.

Fazit: Konstruktionen wie „das absolute Nichts“ kann man erzeugen, wenn man mit abstrakten sprachlichen Gebilden gedankenlos regelhaft so verfährt wie mit konkreten Dingen. Erst muss man ihre Sinnhaftigkeit prüfen, um sich von deren Existenzberechtigung zu überzeugen. Man kann sich vielleicht davon überzeugen, dass sich im Innern einer Büchse kein Goldstück befindet. Dann ist „nicht ein Goldstück“ drin, jedoch damit nicht automatisch das Nichts mit seiner Absolutheit an der Stelle eines Goldstucks.

Korrekt denken bedeutet daher sprachlich, begrifflich korrekt denken.

„Absoluter Zufall“ oder „absolutes Nichtwissen“?

Ebenso ist es unsinnig darüber nachzudenken, ob es den „absoluten Zufall“ oder etwa nur den „relativen“ (aus der Unwissenheit stammenden) Zufall gibt. Denn soweit reicht unser Denken nicht. Sind wir einmal mit einem unerklärbaren Phänomen konfrontiert, so finden wir nicht automatisch eine Handhabe um zu entscheiden, ob zu diesem Phänomen einmal eine Regel gefunden werden wird oder ob es vielleicht eine Regel gibt, welche so komplex ist, dass man sie schon der Komplexität wegen praktisch nicht zwingend finden kann. Ob zum Phänomen aber eine Regel in der Menge aller möglichen Regeln (nicht nur der heute oder auch später einmal zu unserem Wissen gehörenden Regeln) gehört, bleibt unserer Entscheidungsfähigkeit unnahbar. Denn über für den Menschen nicht fassbares, jedoch potentiell faktisch mögliches Wissen in Opposition zu nicht existierendem Wissen, also auf falschen sprachlichen Konstruktionen beruhendem vermeintlichen Wissen, lässt sich nicht streiten. Denn dazu sind uns keine Regeln

gegeben.

Die Frage nach dem „absoluten Zufall“ im Vergleich zum besprochenen „relativen Zufall“ in der Angelegenheit Hans Hansen gehört somit zu den nicht sinnvoll gestellten Fragen.

B.2 Zur Entwicklungsgeschichte des Wahrscheinlichkeitsbegriffs

B.2.1 Die Frage nach der Gewinnchance des Chevaliers de Méré

Am Anfang der Wahrscheinlichkeitsrechnung steht in der Neuzeit die unten gestellte Fragen des Chevaliers de Méré, Philosoph, Literat und begeisterter Spieler. Die Frage war an den Mathematiker, Physiker, Literaten und Philosophen Blaise Pascal gerichtet (1623 – 1662, Frankreich). Bemerkung: Die Fachkompetenzbezeichnungen Mathematiker, Physiker, Literat und Philosoph sind dabei nach dem heutigen Begriffsverständnis genommen zu verstehen.

Aus früheren Zeiten sind dem Autor keine vergleichbaren Fragen überliefert. Damit kann jedoch nicht behauptet werden, dass früher keine solchen Fragen gestellt worden sind.

Eine erste Frage (übersetzt in die heutige Sprechweise) des Chevaliers de Méré lautet:

Welche der beiden Varianten ist wahrscheinlicher: Beim Würfeln mit vier Würfen mindestens mit einem Würfel eine Sechs zu werfen oder beim Würfeln mit 24 Würfen mit zwei Würfeln mindestens eine Doppelsechs zu erhalten? — Wo ist die Gewinnchance größer? (Die Antwort wird das „Méré-Paradoxon“ genannt.) Wie man sieht impliziert die Frage, dass vorgängig man schon eine Vorstellung von einer „Gewinnchance“ haben muss, um diese Frage überhaupt stellen zu können.

Pierre de Fermat (ca. 1607 – 1665) soll dann geholfen haben, die Gewinnchancen genau zu ermitteln.

Eine zweite Frage ist angeblich etwa wie folgt überliefert: Wir betrachten die Situation, in der eine Münze wiederholt einmal geworfen wird. Für immer nur „Zahl“ erhält A einen Punkt und für immer nur „Kopf“ erhält B einen Punkt. A und B spielen. Wer zuerst 5 Punkte erzielt, gewinnt den Einsatz. Nach 7 Würfeln hat die Person A 4 Punkte und die Person B 3 Punkte. Das Spiel wird abgebrochen, obwohl noch keiner 5 Punkte gesammelt hat. Welches ist hier die gerechte Aufteilung des Einsatzes: Nach Maßgabe der gewonnenen Spiele (das heißt 4:3) oder nach Maßgabe der noch fehlenden Spiele (also 2:1, denn B braucht noch 2 Punkte und A nur noch einen) oder überhaupt nach anderen Gesichtspunkten?

Man konsultiere dazu auch: <http://de.wikipedia.org/wiki/De-M%C3%A9r%C3%A9-Paradoxon> oder <http://de.wikipedia.org/wiki/De-M%C3%A9r%C3%A9-Paradoxon> (falls der Link auf der momentan benutzten Maschine nicht klickbar ist, so kopiere man das URL in die Kopfzeile des Internet-Browsers.) Zu den Personen Méré, Pascal, Fermat, Huygens, Bernoulli Jakob I. und Laplace konsultiere man das Wikipedia

B.2.2 Klassischer contra statistischer Wahrscheinlichkeitsbegriff

Die Urheber des klassischen Wahrscheinlichkeitsbegriffs

Christiaan Huygens (1629 – 1665) und Jakob(I.) Bernoulli (1655 – 1705) haben die klassische Wahrscheinlichkeitsrechnung weiter entwickelt. Marquis Pierre Simon de Laplace (1749 – 1827) hat diese Rechnungsart dann ausformuliert. Daher sprechen wir von der „Laplace’schen“ oder der „klassischen“ Wahrscheinlichkeit.

Die klassische Wahrscheinlichkeit

Der Einfachheit halber studieren wir hier mit Vorteil das Modell eines idealen Würfels, bei dem gegenüberliegende Seiten immer exakt parallel und gleich weit voneinander entfernt sowie alle Kantenwinkel rechte Winkel sind. Das Material sei exakt homogen, ohne magnetische oder andere vergleichbare Anziehungseigenschaften und die angebrachten Bezeichnungen für die Augen oder Zahlen seien gewichtlos. Ebenso finde beim Würfeln keinerlei weitere Beeinflussung statt, etwa durch Wind, unebene,

inhomogene Landeoberfläche u.s.w.

Unter diesen Voraussetzungen kann man postulierend feststellen, dass bei „zufälligem Abwurf“ kein Ergebnis vor allen andern ausgezeichnet sein kann. Die „Chancen“ für jede Zahl sind somit gleich sowie untereinander ohne Einfluss auf die Grösse des Resultats vertauschbar. Diese Symmetrieeigenschaft nennen wir „Gleichwahrscheinlichkeit“. Von ihr wollen wir bei den folgenden Betrachtungen ausgehen.

Sei m die Anzahl der möglichen Fälle und g die Anzahl der günstigen Fälle für ein betrachtetes Ereignis. Zum Beispiel betrachten wir beim Würfeln mit einem idealen Würfel das Ereignis „würfeln einer Zahl grösser als 4“. Günstig sind hier zwei Fälle, nämlich die 5 und die 6. Daher ist $g = 2$. Möglich sind hingegen 6 Fälle. Bekanntlich finden wir bei einem Würfel die Zahlen 1, 2, 3, 4, 5, 6. Also ist $m = 6$.

Die Gewinnchance drücken wir bei diesem Ereignis durch das folgende Verhältnis aus:

$$g : m = 2 : 6 = \frac{2}{6} = \frac{1}{3}.$$

Dieses Vorgehen wollen wir bei der klassischen Wahrscheinlichkeit allgemein so halten. Wir definieren daher für diese Wahrscheinlichkeit:

$$P = \frac{g}{m}, \quad g, m \in \mathbb{N}$$

P steht hier für franz. „probabilité“ oder engl. „probability“.

Die klassische Wahrscheinlichkeit wird somit hier durch ein Verhältnis von natürlichen Zahlen ausgedrückt. Da immer $0 \leq g \leq m$ gilt, finden wir:

$$0 \leq P = \frac{g}{m} \leq 1$$

Denn m ist ja die Anzahl aller Ereigisse, eine grössere Zahl gibt es bei einem betrachteten Experiment nicht. Ebenso kann $g = |G|$ nicht negativ sein.

Wenn $g = m$ gilt, so wird $P = \frac{g}{m} = \frac{m}{m} = 1$. In diesem Fall sprechen wir von einem **sicheren Ereignis**.

Wenn hingegen $g = 0$ gilt, so wird $P = \frac{g}{m} = \frac{0}{m} = 0$. Wir sprechen von einem **unmöglichem Ereignis**.

Weitere Beispiele zur klassischen Wahrscheinlichkeit

Wir wollen nun noch einen Fall studieren, in dem mit zwei Würfeln gewürfelt wird. Gefragt ist die Wahrscheinlichkeit, dass die erzielte Summe der gewürfelten Zahlen kleiner gleich 3 ist.

Dazu bestimmen wir zuerst die Anzahl der günstigen Fälle g . Wenn wir einen roten und einen blauen Würfel benutzen, so kann beim roten Würfel die 1 kommen und beim blauen Würfel auch. Dann ist die Summe 2. Das erweist sich als die einzige Möglichkeit, um die Summe 2 zu erhalten. Für die Summen 3 gibt es jedoch 2 Möglichkeiten: 1 mit dem roten Würfel und 2 mit dem blauen. Oder auch 1 mit dem blauen Würfel und 2 mit dem roten. Weitere Möglichkeiten kann man nicht ausmachen, denn die Möglichkeiten $1+1=2$, $1+2=3$ und $2+1=3$ zur Bildung der Summe 3 sind damit ausgeschöpft. Die nächste Möglichkeit zur Bildung einer kleinen Summe wäre $1+3$ oder umgekehrt $3+1$ sowie $2+2$. Damit hätte man aber schon die Summe 4. Diese ist grösser als 3. Es gibt somit nur 3 günstige Möglichkeiten:
 $\Rightarrow g = 3$.

Um die Anzahl der möglichen Resultate m zu berechnen, argumentieren wir analog: Wenn man mit dem roten Würfel eine 1 würfelt, so kann dazu mit dem blauen eine 1, eine 2, eine 3, eine 4, eine 5 oder eine 6 erzielt werden. Das sind 6 Möglichkeiten. Wenn man mit dem roten Würfel eine 2 würfelt, so kann dazu mit dem blauen ebenfalls eine 1, eine 2, eine 3, eine 4, eine 5 oder eine 6 erzielt werden. Das sind wieder 6 Möglichkeiten. Ebenso gibt es 6 Möglichkeiten, wenn wir mit dem roten Würfel eine 3 erzielen. Genau gleich gibt es je 6 Möglichkeiten, wenn wir mit dem roten Würfel eine 4, eine 5 oder eine 6 erzielen. Somit hat man $6 \cdot 6$ Möglichkeiten, mit zwei Würfeln eine Summe zu bilden. Das macht total 36 Möglichkeiten. Damit ist $m = 36$.

Daraus ergibt sich, dass $P(\text{Summe } \leq 3) = \frac{3}{36} = \frac{1}{12} = 0.083\bar{3}\dots$ wird.

An diesem Beispiel wird klar, dass man es bei der Berechnung von g und m mit Abzählproblemen zu tun bekommt. Als Beispiel mag dazu noch das Lotto-Spiel dienen: Wieviele Möglichkeit m gibt es, aus 45 Zahlen 6 auszuwählen. Für g gibt es hier nur eine Möglichkeit, denn der Sechser ist immer eindeutig. Um solche Abzählprobleme zu lösen, ist die klassische Kombinatorik eine Hilfe, welche wir im nächsten Kapitel studieren wollen.

Da g und m Mächtigkeiten von Mengen sind, kommt bei der Behandlung der klassischen Wahrscheinlichkeit auch die Mengenlehre zum Zuge. Mengenoperationen sind hier wichtig.

Statistische Wahrscheinlichkeitsmodelle

Stellen wir uns einen Händler vor, der rote und gelbe Socken verkaufen will. Um vor Saisonbeginn eine Bestellung aufgeben zu können, muss er irgendwie vorauswissen, mit welcher Chance ein Durchschnittskunde rote und mit welcher Chance ein solcher Kunde gelbe Socken kaufen wird. Ebenso muss er etwa den Umfang der Nachfrage in der Saison abschätzen können. Für diese Abschätzungen hilft ihm die klassische Wahrscheinlichkeitsrechnung kaum, denn er kann nicht davon ausgehen, dass der Verkauf eines Paares roter Socken gleich wahrscheinlich ist wie der Verkauf eines Paares gelber Socken. Was also soll er tun?

Hier hilft ihm nur eine statistische Abschätzung weiter. Glücklicherweise hat unser Sockenhändler noch die Verkaufszahlen vom letzten Jahr zur Hand. Er hat mittels elektronischer Registrierung eines jeden Verkaufes durch seine Kasse eine Statistik erstellt. Daher kennt er nun die Häufigkeit des letztjährigen Verkaufs bei den roten Socken $H(\text{rot})$ und auch diejenige bei den gelben Socken $H(\text{gelb})$. Die totale Verkaufsmenge ist $H(\text{rot}) + H(\text{gelb})$. Damit erhält er die relativen Häufigkeiten:

$$h(\text{rot}) = \frac{H(\text{rot})}{H(\text{rot}) + H(\text{gelb})} \text{ und } h(\text{gelb}) = \frac{H(\text{gelb})}{H(\text{rot}) + H(\text{gelb})}. \text{ Dabei gilt:}$$

$$0 \leq h(\text{rot}) = \frac{H(\text{rot})}{H(\text{rot}) + H(\text{gelb})} \leq 1, \quad 0 \leq h(\text{gelb}) = \frac{H(\text{gelb})}{H(\text{rot}) + H(\text{gelb})} \leq 1$$

Man kann leicht sehen, dass die damit definierbare statistische Wahrscheinlichkeit sich nach den selben Regeln verhält wie die klassische, Laplace'sche Wahrscheinlichkeit. In der Praxis bleibt einem ja keine andere Wahl als mit dieser statistischen Wahrscheinlichkeit zu arbeiten. Das ist Pragmatismus wie ihn schon Descartes für die Modelle der Physik vorschlägt, denn weitere Informationen besitzt man ja nicht. Wenn trotzdem weiteren Informationen vorhanden wären, müsste man die relativen Häufigkeiten entsprechend gewichten, womit man sich das verwendete Modell ein wenig verkompliziert. Das dabei verwendete theoretische Gebäude stammt von Kolmogorow.

Das Realitätsdilemma

Wir wollen hier eine Geschichte diskutieren, welche sich auf einer Insel mit dem Namen Boayaky südlich von Samoa zugetragen hat.

Dort hat ein Eingeborener mit dem Namen Yuk im Beisein des Mathematikers Memo ein Würfelexperiment durchgeführt. Vor seiner Hütte sitzend und in Richtung Ost blickend hat er 5692 mal mit zwei nach menschlichem Ermessen als ideal akzeptierbaren Würfeln gewürfelt und dabei 5691 mal die Summe 2 erhalten und 1 mal die Summe 3. Das hat Memo dem Yuk dann nicht glauben wollen und scharf protestiert. Memo hatte während dem Experiment sehr aufmerksam den Vögeln zugeschaut. Er vermutete also gleich Betrug. Das sei Unfug, hat Memo dem Yuk ins Gesicht gesagt, worauf dieser sehr betroffen gewesen sei, denn das schien sehr beleidigend. Darauf sind die beiden übereingekommen, dass Memo jeden Wurf genau beobachten soll und dass das Experiment, trotz seiner langen Dauer, wiederholt werden soll. Da hat Memo nicht schlecht gestaunt. Denn diesmal ist, nachdem

sich Yuk wieder hingesetzt hatte, nun nach Westen blickend, 5690 mal die Summe 12 gekommen und 2 mal die Summe 11, entgegen jeder Erwartung von Memo. Denn Memo hatte nach der Theorie der klassischen Wahrscheinlichkeit je ein Verhältnis von $1 : 36 = \frac{1}{36}$ für die Summe 2 oder die Summe 12 erwartet. Yuk hat Memo darauf geantwortet, dass er sich jetzt überzeugt zeigen müsse, denn er müsse doch seinen eigenen Beobachtungen mehr Glauben schenken als seiner obskuren Theorie. Er habe ja jetzt selbst gesehen, dass die relative Häufigkeit und damit die statistisch abgestützte Wahrscheinlichkeit für die Summen 2 oder 12 praktisch 1 sei. Im ersten Fall sei sie etwa 0.999824 und im zweiten Fall etwa 0.999649. Und so ein Resultat sei ja nicht unmöglich. Es sie jetzt sogar statistisch erwiesen, denn es sei ja eingetroffen, wie man hätte sehen können. Die relativen Häufigkeiten und somit die statistischen Wahrscheinlichkeiten seinen zudem praktisch 1. Es sei also jeweils fast sicher, dass es so sein müsse: In den Fällen wo er mit dem Gesicht nach dem Osten würfle oder eben nach dem Westen.

Leider endet die Geschichte der Vorkommnisse auf Boayaky hier. Denn Memo sei am folgenden Tag mit seiner Jacht nach Samoa abgereist. Bevor er Gelegenheit hatte, die neu entdeckte Insel Boayaky genau zu erforschen, hatte sich in der Gegend ein Erdbeben ereignet. Ein Sunami hätte, so wird berichtet, die kleine und flache Sandinsel, samt allem was darauf lebte und wuchs oder in der Art der Eingeborenen gebaut war, einfach weggespült. Nun erzählt Memo überall, wo er vorbeikommt, seine Geschichte. Und niemand glaubt ihm diese. Die Leute sagen nur: „Armer Memo“, derweil Memo den Leuten sagt: „Armer Yuk“.

Nun wird das Geschehen zur Kriminalgeschichte. Und Memo, der jetzt fast immer ohne Freunde blieb, reiste viel in der Welt herum, um über sein Erlebnis Vorträge vor interessanten Leuten zu halten. So geschah es einmal, als er gerade vor der Handelskammer einer Stadt im Gebirge geredet hatte, dass ihn danach ein sehr böser und übel aussehender Kritiker sehr laut ansprach. Dieser Kritiker rechnete Memo die klassische Wahrscheinlichkeit der von Yuk erzielten Resultate vor. Danach lachte er öffentlich vor der Kamera über Memo, ja lachte ihn geradewegs durchs Mikrofon vor allen Leuten aus. Das war selbst für den sonst immer ruhigen Memo zu viel. Dieser sah den Kritiker verwirrt an und stellte murmelnd fest, dass von den etwa 1000 anwesenden Personen jener üble kritische Mensch der einzige war, der einem exakten Ein-Millimeter-Bürstenschnitt trug. Die Wahrscheinlichkeit für dieses Auftreten war also höchstens 10^{-3} . Ebenso war der Kritiker der einzige, welcher eine Tätowierung am Ohr trug. Die Wahrscheinlichkeit für dieses Auftreten war also wiederum höchstens 10^{-3} . Tätowierung und Bürstenschnitt sind unabhängige Dinge. Somit war die Wahrscheinlichkeit für das gleichzeitige Auftreten beider Phänomene zusammen höchstens $10^{-3} \cdot 10^{-3} = 10^{-6}$. Weiter trug der Kritiker unabhängig von diesen beiden Dingen kniehohe schwarze Lederstiefel. Also wieder mit einer Wahrscheinlichkeit für das Auftreten in solchen Stiefeln von höchstens 10^{-3} , zusammen damit höchstens von 10^{-9} . Weiter stank der Kritiker wie kein zweiter nach einer Mischung von Wein und Knoblauch. Wieder mit einer Wahrscheinlichkeit von höchstens 10^{-3} , zusammen höchstens 10^{-12} . Weiter trug der Kritiker als einziger eine rote Lederhose. Wieder mit einer Wahrscheinlichkeit von höchstens 10^{-3} , zusammen höchstens von 10^{-15} . Dann hatte er als einziger noch auf seinen Haaren einen violetten Kreis angefärbt, wieder mit einer Wahrscheinlichkeit von höchstens 10^{-3} , zusammen nun mit einer Wahrscheinlichkeit von höchstens 10^{-18} . Auf diese Weise entdeckte Memo noch viele andere seltsame und einzigartige Dinge am Kritiker, sodass Memo schliesslich für das Auftreten dieses einzigartigen Wesens eine Wahrscheinlichkeit von 10^{-153} berechnete. Nun folgerte er hieraus streng logisch, dass dieser überaus kritische Erdenbewohner trotz seiner üblen Anwesenheit praktisch nicht mehr zum Universum gehören könne, da die Wahrscheinlichkeit dafür praktisch beinahe null sei. Er stellte daher lauthals fest, dass es dieses kritische Wesen gar nicht geben könne, ja nicht geben dürfe! Denn seine Existenz widerspreche der dafür notwendigen Wahrscheinlichkeit. So zog Memo unverfroren schnell sein Messer aus der Tasche — und stach den Kritiker mit einem Hieb nieder. Denn es gibt ihn ja nicht, so schrie er! Und trotzdem stand er ihm zuvor im Wege. Nur das Messer konnte Memos Widerspruch beseitigen. Was nicht sein darf, verstösst gegen die Bedingungen der Existenz. Was nicht sein darf, das kann auch nicht sein. Damit ist es auch nie gewesen. Es hat keinerlei Existenz.

Darauf soll im grossen Saal ein schrecklicher Tumult entstanden sein. Menschen liefen durcheinander,

schrien, weinten, stiessen sich an, rannten sich gegenseitig um. Dann ging auf einmal das Licht aus. Denn im Tumult muss jemand auf den Nachbar eingeschlagen haben. Und statt diesem hatte dieser Jemand den Lichtschalter getroffen. Irgendwie sind dann trotzdem einige noch unversehrt ins Freie gelangt. Diese fragten dann später in der Gegend herum, ob da wer sei der wisse, was weiter passiert sein könnte. Die Antworten nahmen ihre Richtungen weit auseinander. Einige Befragten waren fest der Meinung, dass Memo jetzt im Zentralgefängnis einsitze. Andere behaupteten, er habe fliehen können. Er sei längst über alle Berge, ja sogar ausser Landes abgehauen. Man hielt ihn für unauffindbar. Gewiss habe er sein Äusseres verändert, mit Hilfe von Gesichtschirurgie natürlich. Und überhaupt, den Memo könne man so umgeformt jetzt nie mehr ergreifen, da er nicht mehr zu erkennen sei. Wieder andere waren der Ansicht, dass das alles sehr unwahrscheinlich sein müsse. Wahrscheinlicher sei vielmehr, dass es den Memo nie wirklich gegeben habe. Es sei sogar fast sicher, dass diese Geschichte frei erfunden sei, von einem Journalisten nämlich, natürlich, denn man müsse ja Zeitschriften und Zeitungen verkaufen. Man brauche also gute Drehbücher für das Hirn der Leser, hart an der Grenze des Erträglichen, denn nur so liesse sich die Auflage des eigenen Druckerzeugnisses noch steigern. Und wenn man Memo inzwischen nicht gefasst hat, so sucht man ihn heute noch. Wegen seiner Tat. Aber vielmehr auch wegen seiner Geschichte, die ihm trotzdem niemand glaubt — die ihm nie jemand glauben wird.

Beispiel aus der Literatur: Zum Münzwurf

Buffon und Pearson haben je wiederholt eine Münze geworfen, wie man der einschlägigen Literatur entnehmen kann. Damit liegt ein beglaubigtes Datenmaterial vor zum Vergleich der klassischen, idealen und der statistischen Wahrscheinlichkeit:

	Anzahl Würfe	Anzahl „Kopf“	relative Häufigkeit $h(Kopf)$
Buffon	4040	2048	0.5069
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

Wie man vermutet, hat die oben beschriebene Boayaky-Situation eher Seltenheitswert. Die Resultate zu den hier aufgelisteten Münzwürfen überraschen dagegen nicht. Sie entsprechen wohl etwa den Erwartungen des „Durchschnittsbetrachters“. Der Leser möge die Resultate selbst interpretieren und kommentieren.

Zusammenfassung

Klassisch versteht man unter Chance oder Wahrscheinlichkeit eine Zahl, welche den Grad des „wahr Scheinens“ eines betrachteten Sachverhalts charakterisiert. „Wahr scheinen“ muss hier von „wahr seiend“, also von „wahr sein“ unterschieden werden. Diese genannte Zahl ist das Verhältnis von günstigen zu möglichen genau definierten Fällen, ein Quotient also von Anzahlen, welche man als Mächtigkeit je einer Menge von Fällen interpretieren kann. Damit hat man es hier mit Mengenlehre und mit Abzählproblemen zu tun. Als Beispiel seien die Würfelprobleme etwa erwähnt.

Beim klassischen Wahrscheinlichkeitsbegriff handelt es sich also um eine theoretische Zahl, die vorerst keine wesenhafte Verbindung zur praktischen Realität besitzt, denn der erwähnte Quotient ist eine gedachte Sache und keine praktisch erfassbare relative Häufigkeit. In der praktischen Realität stellt man jedoch fest, dass diese theoretische Zahl mit gemessenen Häufigkeiten überraschend genau übereinstimmt. Dies ist ein induktiv oder heuristisch gewonnenes Ergebnis, also eine von vielen Leuten für evident gehaltene Erfahrung, welche so als eine „Erfahrungstatsache“ oder von andern auch als eine Interpretation der Erfahrung gewertet wird. Die klassische Wahrscheinlichkeit kann man daher als zulässiges Modell ansehen, das das Verhaltens des damit verbundenen Experiments in der Realität in guter Annäherung beschreibt. Die grosse Übereinstimmung von relativer Häufigkeit und theoretischer Wahrscheinlichkeit bedeutet, dass sich das Modell und die Realität analog verhalten. Man findet also hier eine Analogie vor und nicht etwa eine Kausalität.

Es gibt jedoch viele Fälle in der Praxis, wo der klassische Wahrscheinlichkeitsbegriff versagt. Auf der Grundlage der Theorie von Kolmogorow verwendet man hier dann eine statistische Interpretation der Wahrscheinlichkeit: Man interpretiert die durch Datenerfassung gefundene relative Häufigkeit als Wahrscheinlichkeit. Damit hat man einen Versuch zur Modellierung einer diesmal nicht genau bekannten Wahrscheinlichkeitsfunktion unternommen. Dieses pragmatische Vorgehen wählt man als Reaktion auf das Fehlen weiterer Informationen. Es bleibt einem nichts anderes übrig, denn man kann eben in solchen Fällen keine Möglichkeiten oder günstige Fälle exakt unterscheiden und abzählen. Falls die modellierte Wahrscheinlichkeitsfunktion dann eine gute Übereinstimmung mit den realen mittleren Häufigkeiten liefert, kann man wiederum vom einem guten Modell reden, das man hier aus der Nachbildung des realen Verhaltens im Sinne einer Analogie der Äusserungen als Zahlenwerte und nicht auf der Grundlage etwa eines Ansatzes aus einem inneren Verständnis des Geschehens heraus gefunden hat, wie etwa durch geometrischer Überlegungen. Für die darauf erhaltenen Resultate ist es einerlei, wieviel man über die Grundlagen weiß. Ohne solches Wissen wird es allerdings schwierig sein, die Modellanwendungen auszudehnen. Wichtig ist jedoch, dass man in jedem der beiden Fälle bei Berechnungen gute Resultate erzieht, was ja meistens auch den wichtigsten Beweggrund für solche Arbeiten befriedigt.

Anhang C

Zum Begriff der Kombinatorik

C.1 Übersicht über die elementaren kombinatorischen Fälle

In der klassischen Kombinatorik unterscheidet man 3 Fälle mit je zwei Ausprägungen. Es geht um die drei Fälle „Auswahl“, „Anordnung“ und „Auswahl mit anschliessender Anordnung“. Die Ausprägungen sind „mit Wiederholung“ und „ohne Wiederholung“. Bei den Auswahl- oder Anordnungsmöglichkeiten geht es um die Anzahl der verschiedenen möglichen Elementauswahlen oder die Reihenfolgen von ausgewählten Elementen. Eine Übersicht dazu bietet die nachfolgende Tabelle.

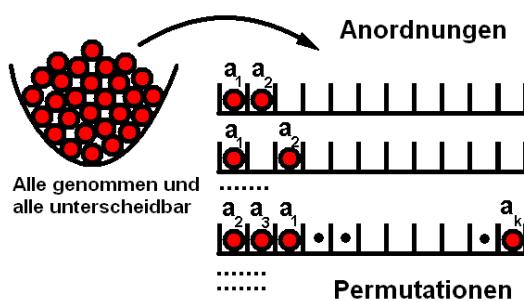
Anzahl Elemente $ M = m$ Daraus n_1, n_2, \dots, n_k auswählen	Ohne Wiederholung	Mit Wiederholung
Alle wählen Nur Anordnung (Reihenfolge) Beispiel: Anzahl Wörter mit m gegebenen Buchstaben	Permutation o.W. $P_m = P(m)$ Alle Elemente sind unterscheidbar	Permutation m.W. $P_m(n_1, \dots, n_k)$ Je n_j Elemente sind nicht unterscheidbar
k Elemente aus m Elementen wählen Nur Auswahl Beispiel: Mannschaftsauswahl bei einem FC.	Kombination o.W. $C(k, m)$	Kombination m.W. $\overline{C}(k, m)$
k Elemente aus m Elementen auswählen und anordnen Beispiel: Anzahl Wörter mit k beliebigen Buchstaben	Variation o.W. $V(k, m)$	Variation m.W. $\overline{V}(k, m)$

C.2 Formeln für die elementaren kombinatorischen Fälle

C.2.1 Anordnungen oder Permutationen ohne Wiederholung

Die Formel

Hier wollen wir untersuchen, auf wieviele mögliche Arten man m unterscheidbare Elemente in einer Reihenfolge anordnen kann. Anders gesprochen: Auf wieviele mögliche Arten kann man m unterscheidbare Elemente, z.B. verschiedene Bücher, in eine Rangfolge bringen, d.h. in unserem Beispiel auf einem Bibliotheksregal anordnen oder einreihen.



Ein konkretes Beispiel erhalten wir in der Frage, auf wie viele Arten man m nummerierte Studenten auf m nummerierte Stühle setzen kann. Das heisst: Wie viele Sitzordnungen gibt es zu einer Klasse von m Studenten.

Um dieses Problem zu lösen nehmen wir an, dass die Schulklasse vor der Tür wartet. Nun wird ein Student nach dem andern hinein gelassen. Der Ablauf präsentiert sich dann wie folgt:

1. Der Student mit der Nummer 1 wird hereingelassen. Ihm stehen m freie Plätze zur Auswahl. Er hat m Möglichkeiten sich zu setzen.
2. Der Student mit der Nummer 2 wird hereingelassen. Da der Student mit der Nummer 1 schon sitzt, stehen dem Studenten mit der Nummer 2 nur noch $m - 1$ freie Plätze zur Auswahl. Zu jeder der m Platzierungen des Studenten mit der Nummer 1 hat der Student mit der Nummer 2 noch $m - 1$ Möglichkeiten. Zusammen haben die ersten beiden Studenten also $m \cdot (m - 1)$ Möglichkeiten, sich zu setzen.
3. Der Student mit der Nummer 3 wird hereingelassen. Da die Studenten mit der Nummer 1 und 2 schon sitzen, stehen dem Studenten mit der Nummer 3 nur noch $m - 2$ freie Plätze zur freien Auswahl. Zu jeder der $m \cdot (m - 1)$ Platzierungen der Studenten mit der Nummer 1 und 2 hat der Student mit der Nummer 3 noch $m - 2$ Möglichkeiten. Zusammen haben die ersten drei Studenten also $m \cdot (m - 1) \cdot (m - 2)$ Möglichkeiten, sich zu setzen.
4. Der Student mit der Nummer 4 wird hereingelassen. Da die Studenten mit der Nummer 1, 2 und 3 schon sitzen, stehen dem Studenten mit der Nummer 4 nur noch $m - 3$ freie Plätze zur freien Auswahl. Zu jeder der $m \cdot (m - 1) \cdot (m - 2)$ Platzierungen der Studenten mit der Nummer 1, 2 und 3 hat der Student mit der Nummer 4 noch $m - 3$ Möglichkeiten. Zusammen haben die ersten vier Studenten also $m \cdot (m - 1) \cdot (m - 2) \cdot (m - 3)$ Möglichkeiten, sich zu setzen.
5. Und so fort. Immer wenn wieder ein Student hereinkommt, erhöht sich das Produkt wieder um einen um 1 kleineren Faktor im Vergleich zum vorher geschriebenen oder erhaltenen Faktor.
6. Schliesslich kommt der letzte Student herein. Er hat noch eine Möglichkeit zum Sitzen, denn nur noch einer der m Plätze ist frei. Damit wird das Produkt $= m \cdot (m - 1) \cdot (m - 2) \cdot (m - 3) \cdots 2 \cdot 1 := m!$. Das Zeichen $m!$ liest man als „ m Fakultät“. Hier ist die Betrachtung zu einem Ende gekommen. $P_m = m!$ ist die totale Anzahl Möglichkeiten. Diese Formel gilt auch allgemein, ob es sich nun um Studenten oder um andere Objekte handelt, die auf Plätzen angeordnet werden. Daher finden wir den unten angegebene Satz:

Satz:

$$P_m = m!$$

Das Problem der auftretenden grossen Zahlen bei den Permutationen

Wir wollen zu einer Klasse wie der betrachteten einmal alle Sitzordnungen in Gedanken durchspielen. Dabei gehen wir von einer Klasse mit 30 Studenten aus. Wie lange würde es dauern, wenn diese im Sekundentakt alle Sitzordnungen durchprobieren würden? Das ist natürlich viel zu schnell, denn vermutlich würde nicht einmal eine halbe Minute reichen. Wir nehmen zudem zur Vereinfachung weiter an, dass

die Studenten nie Pause machen, auch nicht zum Essen und Schlafen. Dass sie also Tag und Nacht immer durcharbeiten, fleissiger als fleissig und schneller als schnell. Nie sollen sie müde werden, nie krank. Gegen alle Regeln der Natur sind sie ständig an der Arbeit. Wie lange würde das wohl dauern? Die Antwort ist sehr einfach: Wir rechnen:

$$s = 30! \text{ sec} = 265252859812191058636308480000000 \text{ sec} \approx 2.6525285981219107 \cdot 10^{32} \text{ sec}$$

Das vergleichen wir mit dem Alter des Universums a in Sekunden:

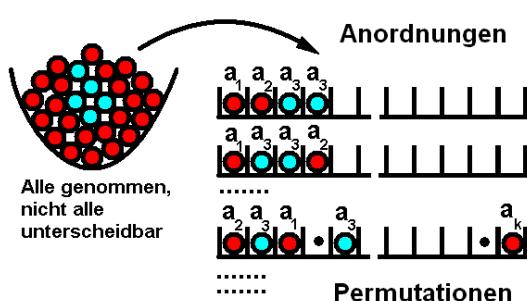
$$a \approx 13.6 \cdot 10^9 \text{ Jahre} = 13.6 \cdot 10^9 \cdot 365 \cdot 24 \cdot 60 \cdot 60 \text{ sec} \approx 4.288896 \cdot 10^{17} \text{ sec} (\text{Universenalter})$$

$$\frac{s}{a} \approx 6.18 \cdot 10^{14} \text{ sec}$$

Das werden unsere Studenten trotz ihrem Fleiss nie erleben. Denn nach einem weiteren Universumsalter wird die Sonne längst zu einem roten Riesenstern angewachsen sein, wie man aus Modellrechnungen weiss, und die Erde in sich aufgenommen „aufgefressen“ haben. Dann bleiben immer noch etwa $6.18 \cdot 10^{14}$ Universenalter, bis alle Sitzordnungen durchgespielt sind!

C.2.2 Permutationen mit Wiederholung

Die Formel



Wir nehmen an, dass sich bei einer Permutation von m Elementen ohne Wiederholung neu k_1 Elemente nicht unterscheiden. Bei jeder Anordnung der $m - k_1$ restlichen Elemente lassen sich daher die k_1 nicht unterscheidbaren Elemente unter sich austauschen oder permutieren. Das kann man nach der oben hergeleiteten Formel für Permutationen ohne Wiederholung auf $k_1!$ mögliche Arten tun.

Sei x_1 die Anzahl Anordnungen der besagten $m - k_1$ restlichen Elemente. Dann gilt: $x_1 \cdot k_1! = P_m = m!$. Daher wird die Permutation von m Elementen bei Ununterscheidbarkeit oder Wiederholung von k_1 Elementen zu $x_1 = P_m(k_1) = \frac{m!}{k_1!}$ berechnet.

Wenn man nun weitere k_2 Elemente unter sich nicht unterscheiden kann, so könnte man gleich wie eben gehabt argumentieren. Sei x_2 die Anzahl Anordnungen der besagten $m - k_1 - k_2$ restlichen Elemente. Dann gilt: $x_2 \cdot x_1 = P_m = m!$. Daher wird $x_2 = P_m(k_1, k_2) = \frac{x_1}{k_2!} = \frac{m!}{k_1! \cdot k_2!}$. Und so fort.

Schliesslich erhalten wir bei Fortsetzung dieser Argumentationsart bei $k_1, k_2, k_3, \dots, k_n$ Elementen, welche je unter sich nicht unterscheidbar sind:

Satz:

$$P_m(k_1, k_2, k_3, \dots, k_n) = \frac{m!}{k_1! \cdot k_2! \cdot \dots \cdot k_n!}$$

Beispiel

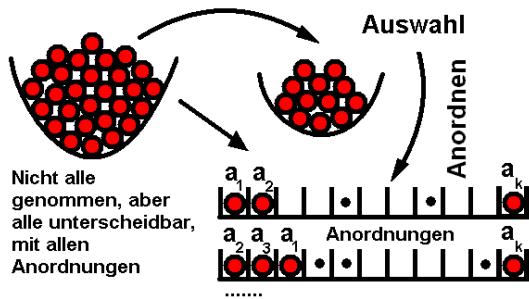
Auf wieviele Arten kann man mit den Buchstaben AABBCDDDEFGH Wörter bilden, auch wenn diese nicht sinnvoll sind?

Zur Lösung bemerken wir zuerst, dass ein neues Wort durch eine Umstellung, also durch eine Permutation der Zeichenkette gebildet wird. Daher reduziert sich die Frage auf das Problem: Auf wieviele Arten kann man 14 Buchstaben permutieren, wenn sich eine erste Gruppe von 2 Buchstaben (A), eine zweite Gruppe von 3 Buchstaben (B) und eine dritte Gruppe von 4 Buchstaben (D) nicht unterscheiden? Lösung:

$$P_{14}(2, 3, 4) = \frac{14!}{2! \cdot 3! \cdot 4!} = 302702400 \approx 3.027024 \cdot 10^8$$

C.2.3 Variationen ohne Wiederholung

Die Formel



Hier geht es darum zu berechnen, auf wie viele Arten man aus einem Vorrat von m Objekten k solche auswählen und diese danach anordnen kann. Dabei argumentieren wir ähnlich wie bei den Permutationen, betrachten aber als erstes die k freien Plätze. Fokussieren wir Nummer 1 dieser Plätze, so können wir aus der Urne jedes der m Objekte auswählen und auf diesen Platz setzen. Für den ersten Platz gibt es also m Möglichkeiten.

Ist dann dieser Platz besetzt, so kann man zu jeder dieser m Besetzungsmöglichkeiten den zweiten Platz auf $m - 1$ mögliche Arten besetzen. Denn in der Urne verbleiben immer noch $m - 1$ Objekte, von denen jedes ausgewählt werden kann. Total hat man für den ersten und zweiten Platz damit $m \cdot (m - 1)$ Möglichkeiten.

Zu jeder dieser $m \cdot (m - 1)$ Möglichkeiten für die ersten beiden Plätze kann man den dritten Platz auf $m - 2$ mögliche Arten besetzen, denn in der Urne bleiben jetzt noch $m - 2$ Objekte. Das gibt $m \cdot (m - 1) \cdot (m - 2)$ Möglichkeiten für die ersten drei Plätze. Entsprechend sind es für die ersten vier Plätze dann

$m \cdot (m - 1) \cdot (m - 2) \cdot (m - 3)$ Möglichkeiten und für die „ersten“ k Plätze darauf

$m \cdot (m - 1) \cdot (m - 2) \cdot (m - 3) \dots (m - k + 1)$ Möglichkeiten. Da es nur k Plätze gibt, hat man also hier für die Anzahl Möglichkeiten ein Produkt mit k Faktoren, in einer Formel also:

$m \cdot (m - 1) \cdot (m - 2) \cdot (m - 3) \dots (m - k + 1)$. Nun ist aber

$$m \cdot (m - 1) \cdot (m - 2) \cdot (m - 3) \dots (m - k + 1) =$$

$$\frac{m \cdot (m - 1) \cdot (m - 2) \cdot (m - 3) \dots (m - k + 1) \cdot (m - k) \cdot \dots \cdot 2 \cdot 1}{(m - k) \cdot \dots \cdot 2 \cdot 1} = \frac{m!}{k!}$$

Daher gilt:

Satz: $V(k, m) = \frac{m!}{(m - k)!} = m \cdot (m - 1) \cdot (m - 2) \cdot (m - 3) \dots (m - k + 1)$

Beispiele

- Wie viele verschiedene Worte zu 6 Buchstaben kann man mit 26 verschiedenen Buchstaben ohne Wiederholung bilden?

Lösung: Hier geht es um eine Auswahl mit Anordnung, also um eine Variation ohne Wiederholung. Denn Worte besitzen meist eine eindeutige Leserichtung. Damit wird

$$V(6, 26) = \frac{26!}{(26 - 6)!} = 165765600 \approx 1.657656 \cdot 10^8$$

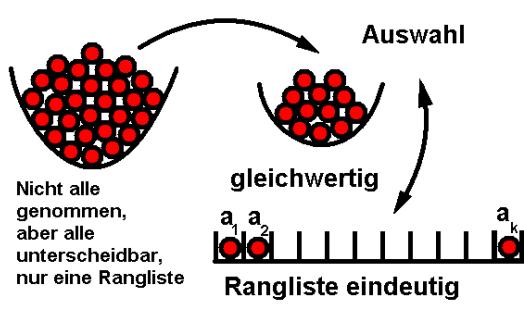
2. An einem Rennen nehmen 100 Pferde teil. Es gibt 10 Preise für die ersten 10 Plätze nach Rangliste. Bei Doppelbesetzung eines Ranges entscheidet das Los vor der Preisverleihung. Wie viele Möglichkeiten gibt es, die Preise auf die 100 Pferde zu verteilen?

Das Problem ist ähnlich dem Problem oben: Eine Auswahl von 10 aus 100 mit Anordnung. Dann wird:

$$V(10, 100) = \frac{100!}{(100 - 10)!} = 62815650955529472000 \approx 6.281565095552947 \cdot 10^{19}$$

C.2.4 Kombinationen ohne Wiederholung

Die Formel



Die Formel für die Kombination ohne Wiederholung $C(k, m)$ können wir aus der Variation ohne Wiederholung $V(k, m)$ herleiten. Dabei bemerken wir, dass bei der Kombination die ausgewählten k Objekte im Unterschied zur Variation nicht auch noch angeordnet werden. Für diese Anordnungen gibt es $k!$ Möglichkeiten, welche hier wegfallen. Daher gilt die Gleichung: $V(k, m) = C(k, m) \cdot k! \Rightarrow C(k, m) = \frac{V(k, m)}{k!} = \frac{m!}{(m - k)! \cdot k!} := \binom{m}{k}$

Satz:

$$C(k, m) = \frac{m!}{(m - k)! \cdot k!} = \binom{m}{k}$$

Bekanntlich nennt man den Ausdruck $\binom{m}{k} := \frac{m!}{(m - k)! \cdot k!}$ **Binomialkoeffizient**. Diese Binomialkoeffizienten treffen wir in der binomischen Formel:

$$\begin{aligned} (a + b)^m &= \binom{m}{0} a^m b^0 + \binom{m}{1} a^{m-1} b^1 + \binom{m}{2} a^{m-2} b^2 + \binom{m}{3} a^{m-3} b^3 + \dots + \\ &\quad + \dots + \binom{m}{m-1} a^{m-(m-1)} b^{m-1} + \binom{m}{m} a^{m-m} b^m \\ \Rightarrow (a + b)^m &= a^m + m a^{m-1} b + \binom{m}{2} a^{m-2} b^2 + \binom{m}{3} a^{m-3} b^3 + \dots + m a b^{m-1} + b^m \end{aligned}$$

Beispiele

1. Im pascalschen Dreieck finden wir die eben besprochenen Binomialkoeffizienten. Den Koeffizienten $\binom{m}{k}$ z.B. erhält man aus dem Produkt mit m Faktoren

$$(a + b)^m = (a + b) \cdot (a + b) \cdot (a + b) \cdot (a + b) \cdot \dots \cdot (a + b) \cdot (a + b),$$

wenn man aus k Faktoren das b und aus $m - k$ Faktoren das a auswählt. So wählen wir damit in der Reihenfolge der Platznummern diejenigen aus mit dem b . Zum Beispiel wird so $\binom{12}{7} = 792$.

2. Problem: In einem Vorrat befinden sich 34 rote und 28 blaue Steine. Wie viele mögliche Ketten mit unterschiedlichem Anfangs- und Endpunkt zu $34 + 28 = 62$ Steinen können wir damit bilden?

Lösung: Um die Kette zu bilden, müssen wir alle Steine auswählen. Zur Vereinfachung denken wir uns die Kette als Rangliste mit nummerierten Plätzen. Wählen wir zuerst nur die blauen Steine und platzieren wir diese auf die dafür ausgewählten 28 Plätze, so sind die roten Steine gesetzt. Die Platzierung dieser Steine ändert darauf die Anzahl Möglichkeiten nicht mehr. Daher reduziert sich das Problem auf die folgende Frage: Wie viele Möglichkeiten gibt es, die 28 aus den 62 Plätzen auszuwählen, auf die dann die blauen Steine zu liegen kommen? Man folgert sofort: Es gibt dafür

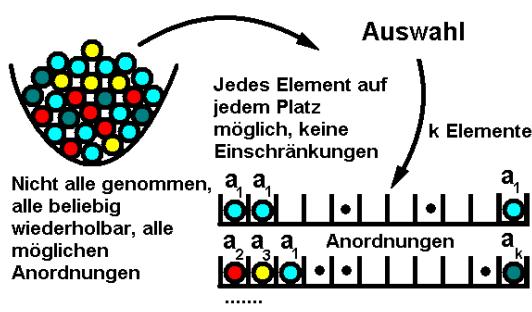
$$\binom{62}{28} = 349615716557887465 \approx 3.496157165578875 \cdot 10^{17} \text{ Möglichkeiten.}$$

3. Lotto: Was ist die Chance, einen 6-er zu haben, wenn mit 45 Kugeln gespielt wird?

Lösung: Die Chance ist $g : m = 1 : \binom{45}{6} = 1 : 8145060$.

C.2.5 Variationen mit Wiederholung

Die Formel



Bei der Variation mit Wiederholung ist die Sache sehr einfach. Man hat k Plätze und wählt aus m Objekten k Stück für diese aus. Jedes Objekt kann bis k mal wiederholt, also mehrmals ausgewählt werden. Auf den ersten Platz kann man so m mögliche Objekte setzen. Ebenso auf den 2. Platz. Damit hat man für die ersten beiden Plätze $m \cdot m = m^2$ Möglichkeiten. Auch auf den 3. Platz kann man m mögliche Objekte setzen. Damit hat man für die ersten drei Plätze $m^2 \cdot m = m^3$ Möglichkeiten u.s.w..

Bei k Plätzen hat man schliesslich m^k verschiedene Möglichkeiten. Damit sind alle Auswahl- und Anordnungsmöglichkeiten ausgeschöpft.

Satz:

$$\overline{V}(k, m) = m^k$$

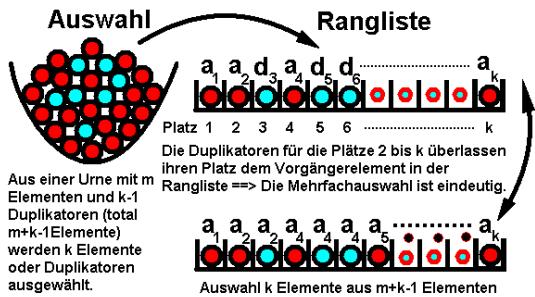
Beispiel

Wie viele Möglichkeiten gibt es, Worte mit 10 Buchstaben zu schreiben, wenn 26 verschiedene Buchstaben zur Verfügung stehen und jeder Buchstabe beliebig oft ausgewählt werden kann?

Lösung: $\overline{V}(k, m) = 26^{10} = 141167095653376 \approx 1.41167095653376 \cdot 10^{14}$.

C.2.6 Kombinationen mit Wiederholung

Die Formel



Bei dieser Formel handelt es sich um die schwierigste bezüglich der Herleitung. Um das Problem zu lösen, führen wir einen neuen Begriff ein: Den Begriff **Duplikator des Vorgängerobjekts für den Platz k**. Dazu beachten wir, dass die gegebenen m unterscheidbaren Objekte in der Urne, aus welcher ausgewählt werden soll, immer mit Nummern versehen gedacht werden können. Und dass zu jeder Auswahl von k Objekten aus den m Objekten dann eindeutig bezüglich der Objektnummern eine Rangliste existiert.

Eine beliebige Auswahl kann so immer unzweideutig als Rangliste interpretiert werden. In dieser Rangliste existieren dann k Plätze, auf denen die Objekte ihren Nummern nach angeordnet gedacht werden können. Durch die Wiederholungen kann es dann passieren, dass auch einmal alles Objekte mit derselben Nummer ausgewählt werden. Wiederholte Objekte wollen wir uns hier immer als Duplikate eines Originals vorstellen. Das ändert die Anzahl der Möglichkeiten nicht. Man stellt aber fest, dass bei jeder möglichen Auswahl immer mindestens ein Originalobjekt dabei ist. Duplikate kann es damit höchstens $k - 1$ Stück geben. In den Ranglisten sind dann die Duplikate der gleichen Nummer nach dem Originalobjekt angeordnet, denn dieses Originalobjekt kann man sich zuerst gezogen denken, bevor Duplikate nachgeliefert werden. Statt der Duplikate denken wir uns nun Geräte, Objekte quasi mit dem Namen „Duplikator“, welche die Duplikate aus dem Vorgängerobjekt erzeugen. Diese Erzeugung ist eindeutig. Statt einem duplizierten Element auf Platz Nummer j kann man daher auch den zu dieser Platznummer j gehörenden Duplikator d_j ausgewählt denken. Dieser dupliziert das Objekt vom davor liegenden Platz mit der Nummer $j-1$ und überlässt ihm anschliessend den eigenen Platz mit der Nummer j . Dieser Duplikationsprozess ist eineindeutig. Man kann aus dem Duplikat und seiner Nummer in der Rangliste auch wieder den Duplikator zurückgewinnen. Damit haben wir das Problem, aus m Objekten in der Urne und den Duplikatoren für die Plätze 2, 3, ..., k für unsere Rangliste k Objekte auszuwählen, wobei dann zuerst die Duplikatoren auf ihre Plätze der Rangliste gesetzt werden und anschliessend die Objekte der Grösse ihrer Nummern oder Ränge nach. Damit ist die Rangliste perfekt. So haben wir das Problem, eine Auswahl von k Elementen aus m Objekten und $k - 1$ Duplikatoren zu treffen. Wir wählen also aus $m - k + 1$ Elementen k Elemente aus. Das ergibt bekanntlich $\binom{m + k - 1}{k}$ Möglichkeiten.

Satz:

$$\overline{C}(k, m) = \binom{m + k - 1}{k}$$

Beispiele

Zur Auswahl stehen 20 Typen von Häusern. Man soll damit Quartiere zu je 7 Häusern bauen. Wiederholungen desselben Haustyps sind erlaubt. Wie viele mögliche Siedlungen gibt es?

Lösung: $\overline{C}(7, 20) = \binom{20 + 7 - 1}{7} = 657800$.

C.3 Eine Aufgabe

Löse das Problem des Chevalier de Méré. (Siehe Seite 8.)

C.4 Links zur Fortsetzung

Links zum weiteren Eindringen in den Stoff findet man unter

<http://rowicus.ch/Wir/Scripts/Scripts.html>

Man beachte dort die Skripte „Wahrscheinlichkeit und Statistik, d“ sowie „Anhang zu Wahrscheinlichkeit und Statistik, d“ und auch „Kombinatorik, Teil 6, d “ .

Anhang D

Fehler von statistischen Kenngrössen und Standardfehler

Unter dem Begriff „Fehlerrechnung“ werden in der Literatur zwei verschiedene Problemtypen behandelt: Erstens das Problem der Verpfanzung von Messfehlern bei der Anwendung von Funktionen auf fehlerbehaftete Messgrössen. Und zweitens das Problem der Fehler von statistischen Kenngrössen (z.B. Mittelwert), welche meistens von einer grossen Datenmenge abhängen. Hier wollen wir weiter unten den zweiten Fehlertyp behandeln. Der erste ist im getrennt herausgegebenen Skript-Hauptteil zu diesem Anhang besprochen worden. Wir wollen ihn weiter unten zum Vergleich mit dem zweiten Typ nochmals wiedergeben.

D.1 Fehler von statistischen Kenngrössen

D.1.1 Die Fragestellung

Gegeben sei eine Serie von Messwerten inklusive ihrer Messfehler von der Form $a_k \pm \Delta a_k$, $k = 1, \dots, n$.

Frage: Was ist dann z.B. der aus den Messfehlern $\pm \Delta a_k$ resultierende Fehler für den Mittelwert $\bar{a} = \frac{1}{n} \sum_{k=1}^n a_k = m(a_1, \dots, a_n)$. Entsprechend stellt sich die für die Varianz $var(a_1, \dots, a_n)$, die Standardabweichung $s(a_1, \dots, a_n)$ und so fort. Wir haben es hier demnach mit Funktionen von n Variablen zu tun, deren Belegung oder Ausprägung durch die Messwerte gegeben sind.

Hier verwenden wir daher die Formel für den maximalen Fehler eines Funktionswertes $h = f(a_1, \dots, a_n)$: $|\Delta f| \leq D_1 |f'_{a_1}(\vec{a}_0)| + \dots + D_n |f'_{a_n}(\vec{a}_0)| := \Delta f_{max}$, $D_k = \pm \Delta a_k$ (Toleranz des Messwerts a_k). Dabei

setzen wir $\vec{a}_0 = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$.

Bemerkung:

Aus der Formel geht hervor, dass damit nur abhängige Grössen in Form von Funktionen $h = f(a_1, \dots, a_n)$ behandelt werden können, welche in allen ihren Variablen differenzierbar sind.

D.1.2 Fehler des Mittelwerts

Gegeben seien die Daten

$$\{a_1 \pm \Delta a_1, a_2 \pm \Delta a_2, \dots, a_n \pm \Delta a_n\}$$

Der Mittelwert \bar{a} berechnet sich dann zu $\bar{a} = \frac{1}{n} \sum_{k=1}^n a_k$.

Dann wird der Fehler $\Delta \bar{a}$ des Mittelwerts

$$\Delta \bar{a}_{max} = |\Delta a_1| \left| \left(\frac{1}{n} \sum_{k=1}^n a_k \right)'_{a_1} \right| + \dots + |\Delta a_n| \left| \left(\frac{1}{n} \sum_{k=1}^n a_k \right)'_{a_n} \right| = |\Delta a_1| \left| \frac{1}{n} \cdot 1 \right| + \dots + |\Delta a_n| \left| \frac{1}{n} \cdot 1 \right| = \frac{1}{n} \cdot \sum_{k=1}^n |\Delta a_k|$$

Satz:

Der maximale Fehler des Mittelwerts ist der Mittelwert der Beträge der Einzelfehler: $\Delta \bar{a}_{max} = \frac{1}{n} \cdot \sum_{k=1}^n |\Delta a_k|$

D.1.3 Fehler der Standardabweichung

$$\text{Für die Standardabweichung gilt: } s = \sqrt{\frac{1}{n-1} \cdot \sum_{k=1}^n (a_k - \bar{a})^2} = \left(\frac{1}{n-1} \cdot \sum_{k=1}^n (a_k - \bar{a})^2 \right)^{\frac{1}{2}}$$

$$\begin{aligned} \text{Nun wird der Fehler } \Delta \bar{s} \text{ der Standardabweichung} \quad & \Delta \bar{s}_{max} = \\ |\Delta a_1| \left| \left(\frac{1}{n-1} \sum_{k=1}^n (a_k - \bar{a})^2 \right)^{\frac{1}{2}} \right)'_{a_1} \right| + \dots + |\Delta a_n| \left| \left(\frac{1}{n-1} \sum_{k=1}^n (a_k - \bar{a})^2 \right)^{\frac{1}{2}} \right)'_{a_n} \right| + \\ & |\Delta \bar{a}| \left| \left(\frac{1}{n-1} \sum_{k=1}^n (a_k - \bar{a})^2 \right)^{\frac{1}{2}} \right)'_{\bar{a}} \right| \\ \left(\frac{1}{n-1} \sum_{k=1}^n (a_k - \bar{a})^2 \right)'_{a_k} &= \frac{1}{2} \cdot \frac{1}{\sqrt{\frac{1}{n-1} \cdot \sum_{k=1}^n (a_k - \bar{a})^2}} \cdot \frac{2}{n-1} \cdot (a_k - \bar{a}) = \frac{1}{\sqrt{(n-1) \cdot \sum_{k=1}^n (a_k - \bar{a})^2}} \cdot (a_k - \bar{a}), \\ \left(\frac{1}{n-1} \sum_{k=1}^n (a_k - \bar{a})^2 \right)'_{\bar{a}} &= \frac{1}{2} \cdot \frac{1}{\sqrt{\frac{1}{n-1} \cdot \sum_{k=1}^n (a_k - \bar{a})^2}} \cdot \frac{-2}{n-1} \cdot \sum_{j=1}^n (a_j - \bar{a}) = \\ & \frac{-1}{\sqrt{(n-1) \cdot \sum_{k=1}^n (a_k - \bar{a})^2}} \cdot \sum_{j=1}^n (a_j - \bar{a}) \\ \Rightarrow \Delta \bar{s}_{max} &= \frac{1}{\sqrt{(n-1) \cdot \sum_{k=1}^n (a_k - \bar{a})^2}} \cdot \left(\sum_{k=1}^n |a_k - \bar{a}| \cdot |\Delta a_k| + \underbrace{\left| \sum_{k=1}^n (a_k - \bar{a}) \right| \cdot |\Delta \bar{a}|}_{=0} \right) \text{ wegen} \\ & \left| \sum_{k=1}^n (a_k - \bar{a}) \right| = \left| \left(\sum_{k=1}^n a_k \right) - \left(\sum_{k=1}^n \bar{a} \right) \right| = n \cdot \bar{a} - n \cdot \bar{a} = 0. \end{aligned}$$

Daraus folgt:

Satz:

$$\Delta \bar{s}_{max} = \frac{\sum_{k=1}^n |a_k - \bar{a}| \cdot |\Delta a_k|}{\sqrt{(n-1) \cdot \sum_{k=1}^n (a_k - \bar{a})^2}} = \frac{\sum_{k=1}^n |a_k - \bar{a}| \cdot |\Delta a_k|}{(n-1) \cdot s}$$

Hinweis: Die obigen Summen von Produkten lassen sich auch elegant mit Hilfe von Skalarprodukten schreiben. Die Ausführung sei dem Leser überlassen. Ebenso die Herleitung weiterer solcher Formeln.

Bemerkung:

Obige Formeln lassen sich für den Kundigen alle noch mit Hilfe des Skalarprodukts übersichtlicher schreiben.

D.1.4 Beispiel

Bsp.: Gegeben sind 8 Messwerte (Zugversuch Holz, Fichte)

$$a_k \in Fichte = \{95.53, 81.93, 83.57, 54.82, 73.83, 58.48, 59.15, 83.29\}$$

mit je einem Fehler von $\Delta a_k = 0.01$. $\rightsquigarrow \bar{a} = 73.825, s = 14.8004 \approx 14.80$.

Wie gross ist der Fehler des Mittelwerts \bar{a} und der Standardabweichung s ?

Für die Berechnung des Fehlers des Mittelwerts $\Delta \bar{a}_{max}$ ist der Mittelwert selbst nicht wesentlich. Da alle Δa_k gleich sind, erhält man:

$$\Delta \bar{a}_{max} = \frac{1}{n} \cdot (n \cdot \Delta a_k) = \frac{1}{8} \cdot (8 \cdot \Delta a_k) = \Delta a_k = 0.01.$$

Weiter gilt hier bei $s = 14.8004$:

$$\Delta \bar{s}_{max} = \frac{\sum_{k=1}^8 |a_k - \bar{a}| \cdot |\Delta a_k|}{(n-1) \cdot s} = \left[\frac{\sum_{k=1}^8 |(Fichte[k] - Mean[Fichte]) \Delta a_k|}{(n-1) StandardDeviation[Fichte]} \right] = 0.0946403$$

D.2 Der Standardfehler von statistischen Kenngrössen

D.2.1 Die Fragestellung

Gegeben sei wieder eine Serie mit der Nummer i von je k Messwerten (**Stichproble** mit der Nummer i , darin Messwerte mit der Nummer k) inklusive ihrer Messfehler von der Form $a_{i,k} \pm \Delta a_{i,k}$, $k = 1, \dots, n$, wobei wir die $\Delta a_{i,k}$ vorläufig nicht ins Zentrum der Betrachtung stellen wollen. Gegeben ist damit ein Vektor \vec{a}_i aus einer Messserie $i = 1, \dots, m$

Der **Standardfehler**, **Stichprobenfehler** oder **Schätzfehler** ist bekanntlich ein Streuungsmaß für die vorhandene **Verteilung** einer solchen oder solcher **Stichproben**. Z.B. der **Standardfehler des Mittelwerts** wird definiert als die Wurzel aus der Varianz der Verteilung **der Stichproben-Mittelwerte** (also die **Standardabweichung** der gemessenen Mittelwerte) von m gleichgroßen Stichproben, welche aus einer gegebenen Grundgesamtheit stammen. Der Standardfehler des Mittelwerts hat damit in der Regel nichts zu tun mit dem Messfehler des Gesamtmittelwerts $\bar{a} = \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{k=1}^n \Delta a_{i,k}$. Entsprechend verhält es sich mit anderen Kenngrössen, z.B. mit der Varianz. Dabei stellen sich Fragen wie die folgenden:

Fragen: Wie kann man den gesamten Mittelwert mit Hilfe der Mittelwerte der einzelnen Messserien schätzen? — Dies scheint eine einfache Frage zu sein. Die nachstehende Anschlussfrage jedoch ist dies

keineswegs: Wie kann man die Varianz der Grundgesamtheit mit Hilfe der Varianzen der einzelnen Messerien schätzen?

D.2.2 Einige Begriffe

Punktschätzer, Schätzer, Schätzwert

Hier geht es um statistische **Schätzmethoden**, z.B. um **Aussagen** über eine **unbekannte Verteilungsfunktion** einer Grundgesamtheit oder deren **Parameter** erhalten zu können.

Statt von **Punktschätzern**, **Schätzern** oder **Schätzwerten** bzw. **Näherungswerten** reden wir auch von **Punktschätzungen**, **Schätzungen** oder **Schätzfunktionen**.

Wenn es hingegen um Aussagen über eine Genauigkeit oder Sicherheit eines Näherungswertes geht, reden wir von **Konfidenzschätzung** oder von einem **Schätzintervall** bzw. einer **Bereichsschätzung** — oder auch manchmal auch von **Toleranzschätzung**. Punktuelle Werte, kurz Punkte, stehen hier also Intervallen gegenüber.

Z.B. können wir zur Schätzung eines Mittelwerts einer unbekannten Grundgesamtheit unterschiedliche Stichprobenmittelwerte heranziehen wie das arithmetische oder das geometrische Mittel, den empirischen Median, im unimodalen Fall den empirischen Modalwert, u.s.w.

Frage: Was ist hier der geeignetste Wert für eine Schätzung? Dabei möchte man eine möglichst „gute“ Annäherung an den unbekannten Parameter der Grundgesamtheit erhalten. Doch was soll man unter einer „guten Annäherung“ verstehen?

Wir nehmen an, dass wir eine Stichprobe $M_S = \{x_1, \dots, x_n\}$ aus einer unbekannten Grundgesamtheit mit einer unbekannten Verteilungsfunktion F vor uns haben, welche von einen unbekannten Parameter Θ abhängt, z.B. $\Theta = \mu$.

Nun sollten wir Θ mit Hilfe von M_S schätzen. Sei dazu etwa eine Funktion T verwendet (man denke z.B. an die Funktion zur Berechnung des Mittelwerts der x_k). Wir schreiben $t = T(x_1, \dots, x_n)$. Dann wäre t ein Wert oder eine Realisierung von $T = T(X_1, \dots, X_n)$. Dabei hängt hier die Funktion T ab vom Variablenvektor $\vec{X}^T = (X_1, X_2, \dots, X_n)$. Neben der **wirklichen Grösse** Θ der Grundgesamtheit erhalten wir dadurch eine **Schätzgrösse** $\hat{\Theta} = T(X_1, \dots, X_n)$, also eine Wahrscheinlichkeitsfunktion mit der **Realisierung** $\hat{\Theta}_0 = t = T(x_1, \dots, x_n)$. Damit haben wir die verschiedenen Ebenen „Grundgesamtheit“, „Stichprobenfunktion (Schätzfunktion)“ und „Stichprobenfunktionswert“ gegeneinander abgetrennt.

Erwartungswert und Erwartungstreue

Gegeben sei ein Zufallsexperiment mit einer Zufallsvariablen X . Z.B. kann man dabei an das Würfeln denken mit dem Ergebnis $X = x_k = \text{Augenzahl beim Wurf Nr. } k$.

Als den **Erwartungswert** dieser Zufallsvariablen X bezeichnet man jenen Wert, welcher sich bei relativ oftmaligem Wiederholen des Experiments als **Mittelwert** der Ergebnisse ergibt resp. sich ergeben sollte. Diesen Wert **erwartet** man also, wenn einem nicht der unberechenbare Zufall einen Strich durch die Rechnung macht. Der Erwartungswert bestimmt die Lage der Verteilung. Man kann dabei also an das empirische arithmetische Mittel einer gemessenen Häufigkeitsverteilung denken. Wegen dem Gesetz der großen Zahlen sollte dann z.B. der Stichprobenmittelwert mit wachsender Stichprobengröße gegen den Erwartungswert konvergieren — so erwarten wir dies.

Schreibweise: Für den Erwartungswert der Schätzfunktion $\hat{\Theta}$ schreiben wir $E(\hat{\Theta})$.

Statt vom **Erwartungswert** spricht man auch oft vom **Grundgesamtheitsmittelwert**. Dieser lässt sich aber nur im Falle von **exakt erfassbar definierbaren Grundgesamtheiten** auch berechnen. Bei **konzeptionellen Grundgesamtheiten** fehlt manchmal die Berechnungsvorschrift.

Wir nennen die Schätzung $\hat{\Theta}$ **erwartungstreu** oder **unverzerrt**, wenn der zu schätzende Parameter gleich dem Erwartungswert ist. Wenn also gilt:

$$E(\hat{\Theta}) = \Theta$$

Man redet von **asymptotisch erwartungstreu**, wenn die Gleichung $E(\hat{\Theta}) = \Theta$ nur im Limes für $n \rightarrow \infty$ gilt.

Wie die asymptotische Erwartungstreue gibt es in diesem Zusammenhang noch weitere gebräuchliche Begriffe (etwa schwach konsistent, die Wirksamkeit, die Effizienz oder die Suffizienz), welche diesen Rahmen hier sprengen.

D.2.3 Die Berechnung des Erwartungswerts

Für eine gegebene Funktion $f(\vec{X}^T)$ einer Zufallsvariablen $\vec{X}^T = (X_1 \dots, X_n)$ wollen wir bei einer Stichprobe $S = \vec{X}_0^T = (x_1, \dots, x_n)$ im diskreten resp. entsprechend im kontinuierlichen Fall den Erwartungswert erklären. Sei $G = \Omega$ die Grundgesamtheit der Mächtigkeit $|\Omega|$. Wir definieren:

Definition:

$$E(f(X)) := \sum_{k=1}^n f(x_k) \cdot p(x_k) \quad (\text{diskreter Fall, } n = |\Omega|)$$

$$E(f(X)) := \int_{-\infty}^{\infty} f(x) \cdot p(x) dx \quad (\text{kontinuierlicher Fall})$$

Für \vec{X}^T bauen wir $E(f(\vec{X}^T))$ später rekursiv auf $E(f(X))$ auf. Wir betrachten daher vorläufig $E(f(X))$ an Stelle von $E(f(\vec{X}^T))$.

Weiter ist $p(x_k)$ die Wahrscheinlichkeit von x_k . Entsprechend ist $p(x)$ ist die Wahrscheinlichkeitsdichte von x .

Für $f = \text{identische Abbildung}$ gilt $f(X) = X$. Dann haben wir im diskreten Fall: $E(X) := \sum_{k=1}^n x_k \cdot p(x_k)$. Im Fall dass alle x_k **gleich wahrscheinlich** sind, muss bei $n = |\Omega|$ Teilversuchen gelten:

$$p(x_k) = \frac{1}{n} \quad \text{und} \quad \sum_{k=1}^n p(x_k) = \sum_{k=1}^n \frac{1}{n} = \frac{n}{n} = 1.$$

$$\text{Dann folgt: } E(X) := \sum_{k=1}^n x_k \cdot \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n x_k = \bar{x} = \text{Stichprobenmittelwert, } n = |\Omega|$$

Somit folgt der Satz, dass im diskreten Fall der Erwartungswert einer Zufallsvariablen X mit *gleich verteilter Wahrscheinlichkeit* tatsächlich der Stichprobenmittelwert ist:

Satz:

Vor.:

Wie eben beschrieben: X gleich verteilt.

Beh.:

$$E(X) = \bar{x}$$

Allgemein haben wir aber nicht in der Regel eine gleich verteilte Wahrscheinlichkeit. Daher ist $E(X) = \bar{x}$ nur im Falle $N \rightarrow \infty$ resp. $N \rightarrow N_{max}$ zu erwarten. Wir schreiben aus diesem Grunde für den theoretischen Mittelwert μ :

$$\mu_X := E(X)$$

Verwendet man die empirische Wahrscheinlichkeitsdefinition mit Hilfe der relativen Häufigkeit ($\lim_{N \rightarrow \infty} p(X_k) = \frac{N_k}{N}$ oder $p(X_k) \approx \frac{N_k}{N}$, $N \gg 1$), so ist der Erwartungswert „wie erwartet“ der Mittelwert, den man bei grossen Stichproben etwa erwartet.

D.2.4 Formeln für den Erwartungswert

Kann bei einer Stichprobe nur einmal a gezogen werden, so ist $X = a$, $n = 1$. Daher gilt $p(a) = 1$ resp. $\int_{-\infty}^{\infty} a \cdot p(x) dx = a \cdot \int_{-\infty}^{\infty} p(x) dx = a \cdot 1 = a$. (Falls a n mal gezogen wird, gilt $p(a) = \frac{1}{n}$, weil die gesamte Wahrscheinlichkeit aller möglichen Werte 1 sein muss. Dann ist das Resultat dasselbe.) Damit erhalten wir:

Formel: $E(a) = a$

Weiter gilt: $E(a \cdot X) = \sum_{k=1}^n a \cdot x_k \cdot p(x_k) = a \cdot \sum_{k=1}^n x_k \cdot p(x_k) = a \cdot E(X)$, $n = |\Omega|$.

Entsprechend im kontinuierlichen Fall. Damit gilt:

Formel: $E(a \cdot X) = a \cdot E(X)$

Es ist: $E(a + X) = \sum_{k=1}^n (a + x_k) \cdot p(x_k) = \sum_{k=1}^n a \cdot p(x_k) + \sum_{k=1}^n x_k \cdot p(x_k) = a \cdot \sum_{k=1}^n p(x_k) + E(X) = a \cdot 1 + E(X)$. ($n = |\Omega|$.) Entsprechend im kontinuierlichen Fall. Damit gilt:

Formel: $E(a + X) = a + E(X)$

Seien weiter X und Y **unabhängige Zufallsvariablen**. Dann gilt mit $n = |\Omega_X|$, $m = |\Omega_Y|$:

$$\begin{aligned} E(X + Y) &= \sum_{k=1}^n \sum_{j=1}^m (x_k + y_j) \cdot p(x_k) p(y_j) = \sum_{k=1}^n p(x_k) \sum_{j=1}^m (x_k + y_j) \cdot p(y_j) \\ &= \sum_{k=1}^n p(x_k) \left(\sum_{j=1}^m x_k \cdot p(y_j) + \sum_{j=1}^m y_j \cdot p(y_j) \right) = \sum_{k=1}^n p(x_k) \left(x_k \underbrace{\sum_{j=1}^m p(y_j)}_{=1} + E(Y) \right) = \sum_{k=1}^n p(x_k) (x_k + E(Y)) \\ &= \underbrace{\sum_{k=1}^n p(x_k) x_k}_{=E(X)} + \underbrace{\sum_{k=1}^n p(x_k) E(Y)}_{=E(Y) \cdot \sum p(x_k) = E(Y) \cdot 1} = E(X) + E(Y). \end{aligned}$$

Entsprechend im kontinuierlichen Fall. Damit gilt:

Formel: $E(X + Y) = E(X) + E(Y)$ bei unabhängigem X und Y .

Kombinieren wir $E(X + Y) = E(X) + E(Y)$ mit $E(a \cdot X) = a \cdot E(X)$, so folgt:

Formel: $E(aX + bY) = aE(X) + bE(Y)$ bei unabhängigem X und Y .

D.2.5 Formeln für die Varianz

Definition: $\sigma^2 = \sigma_X^2 := E((X - \mu)^2) = E((X - E(X))^2)$, $\sigma_X := \text{var}(X)$

Für eine endliche Grundgesamtheit $G = \Omega$ mit $|\Omega| = N$, $P((X - \mu)^2) = P((X_k - \mu)^2) = \frac{1}{N}$, gilt:

$$\sigma^2 = E((X - \mu)^2) = \sum_{k=1}^N (X_k - \mu)^2 P((X_k - \mu)^2) = \sum_{k=1}^N (X_k - \mu)^2 \frac{1}{N} = \frac{1}{N} \sum_{k=1}^N (X_k - \mu)^2.$$

Formel: $G = \Omega$ mit $|\Omega| = N$, $P((X - \mu)^2) = P((X_k - \mu)^2) = \frac{1}{N}$
 $\Rightarrow \sigma^2 = \frac{1}{N} \sum_{k=1}^N (X_k - \mu)^2$.

Es gilt: $\sigma^2 := E((X - \mu)^2) = E(X^2 - 2X\mu + \mu^2) = E(X^2) - E(2X\mu) + E(\mu^2)$
 $= E(X^2) - 2E(X)\mu + \mu^2 E(1) = E(X^2) - 2\mu\mu + \mu^2 = E(X^2) - \mu^2 = E(X^2) - (E(X))^2$

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

Sein nun $X = X_1 + X_2$. Dann folgt:

$$\mu_X = E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = \mu_{X_1} + \mu_{X_2} \quad \text{und}$$

$$\begin{aligned} \sigma_X^2 &= E((X - E(X))^2) = E((X_1 + X_2 - E(X_1 + X_2))^2) = E((X_1 + X_2 - E(X_1) - E(X_2))^2) \\ &= E(((X_1 - E(X_1)) + (X_2 - E(X_2)))^2) = E((X_1 - E(X_1))^2 - 2(X_1 - E(X_1))(X_2 - E(X_2)) + (X_2 - E(X_2))^2) \\ &= \underbrace{E((X_1 - E(X_1))^2)}_{=\sigma_{X_1}^2} - 2 \underbrace{E((X_1 - E(X_1))(X_2 - E(X_2)))}_{=\text{cov}(X_1, X_2)} + \underbrace{E((X_2 - E(X_2))^2)}_{=\sigma_{X_2}^2}. \end{aligned}$$

Hier ist $\text{cov}(X_1, X_2)$ die **Kovarianz** von X_1 und X_2 . Für sie gilt:

$$\begin{aligned} \text{cov}(X_1, X_2) &= E((X_1 - E(X_1))(X_2 - E(X_2))) = E(X_1 \cdot X_2 - X_1 E(X_2) - E(X_1) X_2 + E(X_1) E(X_2)) \\ &= E(X_1 \cdot X_2) - E(X_1) E(X_2) - E(X_1) E(X_2) + E(X_1) E(X_2) = E(X_1 \cdot X_2) - E(X_1) E(X_2) \end{aligned}$$

Damit erhalten wir die Formeln:

Formel: $\sigma_{X_1+X_2}^2 = \sigma_{X_1}^2 - 2 \text{cov}(X_1, X_2) + \sigma_{X_2}^2$

Für $\text{cov}(X_1, X_2) = 0$ (Fall der Unkorreliertheit) gilt:

Formel: $\sigma_{X_1+X_2}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2$ (Fall der Unkorreliertheit von X_1 und X_2).

$$\sigma_{a \cdot X}^2 = E((aX - E(aX))^2) = E((aX - aE(X))^2) = E(a^2(X - E(X))^2) = a^2 E((X - E(X))^2) = a^2 \sigma_X^2$$

Formel: $\sigma_{a \cdot X}^2 = a^2 \cdot \sigma_X^2$

D.2.6 Formel für den Erwartungswert des Mittelwerts

Sei \bar{X} = Zufallsvariable für einen Mittelwert: $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$, $n \leq |\Omega_X|$. Dabei nehmen wir an, dass $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$ eine Wertbelegung von $\frac{1}{n} \sum_{k=1}^n X_k$ ist und dass X bei jeder Wertrealisation als eine Realisation eines einzelnen Wertes für eine einzelne, zu allen andern gleiche Zufallsvariable X_k auftritt. Dann ist $\mu_X = \mu_{X_k}$ für alle k . Damit gilt:

$$\mu_{\bar{X}} = \mu_{\bar{X}_n} = E(\bar{X}) = E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} E\left(\sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{1}{n} \sum_{k=1}^n \mu_X = \frac{n}{n} \mu_X = \mu_X.$$

Formel: $\mu_{\bar{X}} = \mu_{\bar{X}_k} = \mu_X$ oder $E(\bar{X}) = E(X)$, $n \leq |\Omega_X|$

D.2.7 Nochmals Formeln für die Varianz

Es gilt wie oben hergeleitet: $\sigma^2 = var(X) = E((X - E(X))^2)$ und weiter

$$var(a) = E((a - E(a))^2) = E((a - a)^2) = E(0) = 0 \text{ sowie}$$

$$var(a + X) = E((a + X - E(a + X))^2) = E((a + X - a - E(X))^2) = E((X - E(X))^2) = var(X) = \sigma_X^2.$$

σ_X^2 kann dabei geschätzt werden durch eine Realisierung $\hat{\sigma}_X^2 = s^2 = StD^2$. Wir notieren:

Formel: $var(a) = 0$ und $var(a + X) = var(X) = \sigma_X^2$

D.2.8 Formel für die Varianz des Mittelwerts

Im Falle der Unkorreliertheit gilt:

$$\begin{aligned} \sigma_{\bar{X}}^2 = var(\bar{X}) &= var\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \left(\frac{1}{n}\right)^2 var\left(\sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n var(X_k) = \frac{1}{n^2} \sum_{k=1}^n var(X) = \frac{n}{n^2} var(X) \\ &= \frac{1}{n} var(X) = \frac{1}{n} \sigma_X^2 \end{aligned}$$

Formel: $\sigma_{\bar{X}}^2 = var(\bar{X}) = \frac{1}{n} var(X) = \frac{1}{n} \sigma_X^2$ oder

$$\sigma_{\bar{X}} = \frac{1}{\sqrt{n}} \sigma_X \quad (\text{Fall der Unkorreliertheit, } \bar{X} = \bar{X}_n, n \leq |\Omega_X|)$$

Konsequenz: $0 \leq \sigma_X \leq c \in \mathbb{R}^+ \Rightarrow \lim_{n \rightarrow \infty} \sigma_{\bar{X}} = 0$

D.3 Messfehler und Standardfehler

Wird bei einer Messung (Werte $\{a_1, \dots, a_n\}$) kurzum ein hypothetischer Messfehler in der Form der absoluten Abweichung vom hypothetischen Mittelwert genommen, so können wir diese Abweichung mittels eines guten Schätzers (hier $\hat{\Theta} = \bar{a}$) für den Mittelwert angeben. Damit wird $|\Delta a_i| = |\bar{a} - a_i| \approx |\mu - a_i|$.

Für die maximale Streuung (vgl. Seite 24) folgern wir damit:

$$\Delta \bar{s}_{max} = \frac{\sum_{k=1}^n |a_k - \bar{a}| \cdot |\Delta a_k|}{(n-1) \cdot s} = \frac{\sum_{k=1}^n (a_k - \bar{a})^2}{(n-1) \cdot s} = \frac{1}{s} \cdot \frac{1}{n-1} \sum_{k=1}^n (a_k - \bar{a})^2 = \frac{s^2}{s} = s$$

Formel: $|\Delta a_i| \approx |\mu - a_i| \Rightarrow \Delta \bar{s}_{max} = s$

D.4 Einige Links

<http://de.wikipedia.org/wiki/Erwartungstreue>

<http://de.wikipedia.org/wiki/Standardfehler>

<http://de.wikipedia.org/wiki/Erwartungswert>

<http://de.wikipedia.org/wiki/Grundgesamtheitsmittelwert> (Erwartungswert)

http://de.wikipedia.org/wiki/Bedingter_Erwartungswert

<http://de.wikipedia.org/wiki/Varianz>

D.5 Ein Beispiel

In einem Betrieb hat man über eine übliche Produktionsperiode von 20 Tagen den Schmiermittelverbrauch in Litern an einer Produktionslinie protokolliert. Die letzte Ziffer der Werte ist jeweils als Rundung zu verstehen. Die betrachtete Periode wird als repräsentativ angenommen, da keine besonderen Vorkommnisse zu verzeichnen waren und sie sich durch keine Auffälligkeit von anderen Perioden unterschieden hat. Die Maschinen waren in dieser Produktionszeit wie üblich im Dauereinsatz. Nun liegt die folgende Messreihe des Verbrauchs vor (20 Datenelemente):

$$M = \{12.1, 12.0, 12.3, 12.6, 12.2, 12.3, 12.3, 12.8, 11.6, 12.5, 12.3, 12.4, 12.2, 12.9, 12.4, 12.6, 12.5, \\ 12.3, 12.4, 12.4\}$$

Die Daten sind mit einem Fehler von ± 0.1 behaftet. Auf der Grundlage dieser Messreihe soll versucht werden, in Form von Schätzern Aussagen über die Grundgesamtheit G des Verbrauchs dieser Produktionslinie zu gewinnen. Dabei wird angenommen, dass an dieser Linie keine Änderungen der Parameter vorgenommen wurden respektive stattgefunden haben. Auf der Grundlage früherer Erfahrungen weiß man auch, dass die Verteilung der Grundgesamtheit G im wesentlichen Bereich annähernd eine Normalverteilung ist.

1. Berechne den Mittelwert der Stichprobe.
2. Berechne den Messfehler des Mittelwerts der Stichprobe.

3. Berechne einen einfachen Schätzer für den Mittelwert der Grundgesamtheit.
4. Berechne einen einfachen Schätzer für die Streuung (Standardabweichung) des Mittelwerts der Grundgesamtheit.
5. Berechne auf der Grundlage der Normalverteilung annähernd ein 95 %-Vertrauensintervall für den Mittelwerts der Grundgesamtheit.
6. Berechne auf der Grundlage der Normalverteilung annähernd ein 99.75 %-Vertrauensintervall für den Mittelwert der Grundgesamtheit.

Zur Lösung:

1. Für den Mittelwert gilt: $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k = \frac{1}{20} (12.1 + 12.0 + 12.3 + \dots + 12.4 + 12.4) \approx 12.355$

2. Für den Messfehler des Mittelwert gilt:

$$\Delta \bar{x}_{max} = \frac{1}{n} \cdot \sum_{k=1}^n |\Delta x_k| \approx \frac{1}{20} \cdot (|0.1| + |0.1| + \dots + |0.1|) = \frac{20}{20} \cdot 0.1 = 0.1$$

3. Schätzer des Mittelwerts der Grundgesamtheit: $\mu_{\bar{X}} = \mu_X \approx \bar{x} = 12.355$

4. Schätzer der Streuung (Mittelwert): $\sigma_{\bar{X}} = \frac{1}{\sqrt{n}} \sigma_X \approx \frac{1}{\sqrt{20}} s_X = \frac{1}{\sqrt{20}} 0.279991 = 0.0626078$

5. Ein 95 %-Vertrauensintervall für den Mittelwerts der Grundgesamtheit ist gegeben durch

$$\mu_{\bar{X}} \pm 2 \sigma_{\bar{X}} = \mu_X \pm 2 \frac{1}{\sqrt{n}} \sigma_X \approx 12.355 \pm 2 \cdot 0.0626078 = 12.355 \pm 0.125216$$

6. Ein 99.75 %-Vertrauensintervall für den Mittelwerts der Grundgesamtheit ist annähernd gegeben durch das 99.73 %-Vertrauensintervall

$$\mu_{\bar{X}} \pm 3 \sigma_{\bar{X}} = \mu_X \pm 3 \frac{1}{\sqrt{n}} \sigma_X \approx 12.355 \pm 3 \cdot 0.0626078 = 12.355 \pm 0.187823$$

D.6 Der Erwartungswert eines Funktionswerts

Gegeben sei eine Funktion $Y = f(X)$ einer Zufallsvariablen X mit dem Erwartungswert $E(X)$.

Frage: $E(Y) = ?$ Und wie kann man $E(Y)$ aus $E(X)$ berechnen?

Wir nehmen hier an, dass es zur Funktion f eine **Taylorapproximation** mit abschätzbarem Restglied gibt. Insbesondere ist das dann der Fall, wenn f in eine **Potenzreihe** entwickelbar ist. Dort gilt:

$$Y = f(X) = \sum_{k=0}^n \frac{f^{(k)}(X_0)(X - X_0)^k}{k!} + R(X, X_0, n) \quad \text{mit} \quad \lim_{X \rightarrow X_0} R(X, X_0, n) = 0$$

1. In der 1. Approximation hätte man dann $Y = f(X) \approx f(X_0) + f'(X_0)(X - X_0)$.

2. Und in der 2. Approximation $Y = f(X) \approx f(X_0) + f'(X_0)(X - X_0) + \frac{f''(X_0)(X - X_0)^2}{2}$.

Sei jetzt $X_0 = E(X) = \mu_X$. Weil $X_0 = \mu_X$ sowie $f^n(X_0)$ und auch $n!$ Konstanten sind, gilt damit:

$$\mu_Y = E(Y) = E(f(X)) \approx E\left(\sum_{k=0}^n \frac{f^{(k)}(\mu_X)(X - \mu_X)^k}{k!}\right) = \sum_{k=0}^n \frac{f^{(k)}(\mu_X) E((X - \mu_X)^k)}{k!}$$

$$= f(\mu_X) + f'(\mu_X) \underbrace{(\mu_X - \mu_X)}_{=0} + \frac{f''(\mu_X)}{2} \underbrace{E((X - \mu_X)^2)}_{=\sigma_X^2} + \dots \text{ Damit erhalten wir:}$$

Formel: 1. Approximation: $\mu_Y = E(Y) = E(f(X)) \approx f(\mu_X)$

Formel: 2. Approximation: $\mu_Y = E(Y) = E(f(X)) \approx f(\mu_X) + \frac{f''(\mu_X)}{2} \sigma_X^2$

Analog finden wir für Funktionen mit mehreren Variablen $Y = f(X_1, \dots, X_j)$ den Erwartungswert $\mu_Y = E(Y) = E(f(X_1, \dots, X_j))$ wie folgt:

Formel: 1. Approximation: $\mu_Y = E(Y) = E(f(X)) \approx f(\mu_{X_1}, \dots, \mu_{X_j})$

Formel: 2. Approximation: $\mu_Y = E(Y) = E(f(X)) \approx f(\mu_{X_1}, \dots, \mu_{X_j}) + \frac{\partial^2 f}{\partial X_1^2}(\mu_{X_1}, \dots, \mu_{X_j}) \sigma_{X_1}^2 + \dots + \frac{\partial^2 f}{\partial X_j^2}(\mu_{X_1}, \dots, \mu_{X_j}) \sigma_{X_j}^2$

Damit lassen sich Vertrauensintervalle berechnen, indem man für die μ_{X_k} und die σ_{X_k} geeignete Schätzer verwendet.

Weiter gilt in der 1. Approximation: $\mu_Y \approx f(\mu_X)$ und $Y = f(X) \approx f(\mu_X) + f'(\mu_X)(X - \mu_X)$

Daraus folgt:

$$\begin{aligned} \sigma_Y^2 &= E((Y - E(Y))^2) = E((Y - \mu_Y)^2) \approx E((Y - f(\mu_X))^2) \\ &= E((f(\mu_X) + f'(\mu_X)(X - \mu_X) - f(\mu_X))^2) = E((f'(\mu_X)(X - \mu_X))^2) \\ &= (f'(\mu_X))^2 E((X - E(X))^2) = (f'(\mu_X))^2 \sigma_X^2. \end{aligned}$$

Damit erhalten wir:

Formel: 1. Approximation: $\sigma_Y^2 \approx (f'(\mu_X))^2 \sigma_X^2$

Entsprechend ergibt sich bei mehreren Variablen, $Y = f(X_1, \dots, X_j)$:

Formel: 1. Approximation:

$$\sigma_Y^2 \approx \left(\frac{\partial^2 f}{\partial X_1^2}(\mu_{X_1}, \dots, \mu_{X_j}) \right)^2 \sigma_{X_1}^2 + \dots + \left(\frac{\partial^2 f}{\partial X_j^2}(\mu_{X_1}, \dots, \mu_{X_j}) \right)^2 \sigma_{X_j}^2$$

Größere Anwendungsbeispiele sollen nicht im Rahmen der Grundlagen dargestellt werden. Sie sind Gegenstand allfälliger Übungen.

D.7 Standardabweichung und Tschebyschow–Ungleichung

D.7.1 Zur Standardabweichung

Die **Standardabweichung** σ_X ist definiert als $\sqrt{E((X - E(X))^2)}$. Oft gibt man im diskreten Fall dafür als **Schätzer** den Wert $s_X^* = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{x})^2}$ an, weil $(s^*)^2$ ein erwartungstreuer Schätzer

für die Varianz σ_X^2 ist. s_X^* ist aber kein erwartungstreuer Schätzer für die Standardabweichung σ_X . s_X^* überschätzt in vielen Fällen die Standardabweichung σ_X der Grundgesamtheit. Statt s_X^* verwendet man

dann vielfach den Maximum-Likelihood-Schätzer $s_X^{ML} = \sqrt{\frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^2}$.

Für spezielle damit verbundene Fragen möge der Leser die Fachliteratur konsultieren.

Beispiel:

Gegeben sei die Datenmenge (Grundgesamtheit) $G = \Omega = \{1, 1, 1, -1, -1, -1\}$. Man sieht sofort, dass hier $E(X) = \mu = 0$ ist.

Für die Varianz $\sigma^2 = E((X - E(X))^2)$ erhalten wir damit

$$\sigma^2 = E((X - E(X))^2) = E(X^2) = \frac{1}{6}((1)^2 + (1)^2 + (1)^2 + (-1)^2 + (-1)^2 + (-1)^2) = \frac{1}{6} \cdot 6 = 1.$$

Damit wird $\sigma_X = \sqrt{E((X - E(X))^2)} = 1$. Sein nun „Stichprobe = G “:

Rechnen wir jetzt mit der Formel $s_X^* = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{x})^2}$, so erhalten wir die Überschätzung

$$s_X^* = \sqrt{\frac{1}{6-1} ((1)^2 + (1)^2 + (1)^2 + (-1)^2 + (-1)^2 + (-1)^2)} = \sqrt{\frac{6}{5}} = \sqrt{1.2} \approx 1.09545 > 1 = \sigma.$$

Hingegen liefert hier $s_X^{ML} = \sqrt{\frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^2}$ den richtigen Wert 1.

Generell gilt jedoch (vgl. dazu Seite 102):

1. Sei $s = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}$ die **empirische Standardabweichung**. $\{x_1, \dots, x_n\}$ sei stochastisch unabhängig. Dann ist s ein **biasfreier Schätzer** für σ .

2. Sei $\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2}$ die Standardabweichung der Stichprobe als Grundgesamtheit, $\{x_1, \dots, x_n\}$ stochastisch unabhängig. $\hat{\sigma}$ ist **kein biasfreier Schätzer** für σ .

D.h. wegen $s > \hat{\sigma}$ **unterschätzt** $\hat{\sigma}$ den Wert σ in der Regel.

Jedoch gilt für $n \gg 1$ bald $\frac{1}{n-1} \approx \frac{1}{n}$ und damit $s \approx \hat{\sigma}$.

D.7.2 Die Tschebyschow–Ungleichung

Wir betrachten eine Grundgesamtheit mit $N \in \mathbb{N}$ Elementen. Für diese gelte $P(x_k) = \frac{1}{N}$.

Sei $x_k \notin [\mu - \Delta x, \mu + \Delta x]$, $\Delta x > 0$.

Dann stellen wir fest: $|x_k - \mu| > \Delta x \Rightarrow (x_k - \mu)^2 > \Delta x^2 \Rightarrow \frac{(x_k - \mu)^2}{\Delta x^2} > 1$.

$$\begin{aligned} \text{Somit ist: } P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) &= \sum_{x_k \notin [\mu - \Delta x, \mu + \Delta x]} P(x_k) = \sum_{x_k \notin [\mu - \Delta x, \mu + \Delta x]} \frac{1}{N} \\ &\leq \sum_{x_k \notin [\mu - \Delta x, \mu + \Delta x]} \frac{1}{N} \cdot \frac{(x_k - \mu)^2}{\Delta x^2} \leq \sum_{k=1}^N \frac{1}{N} \cdot \frac{(x_k - \mu)^2}{\Delta x^2} = \frac{1}{\Delta x^2} \cdot \left(\frac{1}{N} \sum_{k=1}^N (x_k - \mu)^2 \right) = \frac{1}{\Delta x^2} \cdot \sigma_X^2 = \frac{\sigma^2}{\Delta x^2}. \end{aligned}$$

Diese Ungleichung gilt auch allgemeiner. Man formuliert sie wie folgt:

Satz: (Tschebyschow- oder (engl.) Chebyshev-Ungleichung)

Vor.:

X sei Zufallsgrösse mit Verteilungsfunktion F und $\mu, \sigma \in \mathbb{R}$ (endlich).

Beh.: $P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) \leq \frac{\sigma^2}{\Delta x^2}$

Damit kann man also die Wahrscheinlichkeit abschätzen, dass x_k ausserhalb eines zu μ symmetrischen Intervalls der Breite $2\Delta x$ liegt. Das hat eine grosse **praktische Bedeutung**, wenn die **Verteilungsfunktion unbekannt** ist, man aber μ oder vielleicht σ kennt!

Konsequenz: Für alle Wahrscheinlichkeitsmodelle gilt unter den eben beschriebenen Voraussetzungen:

1. Sei $\Delta x = 2\sigma \Rightarrow P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) \leq \frac{\sigma^2}{(2\sigma)^2} = \frac{1}{4} = 0.25.$
2. Sei $\Delta x = 3\sigma \Rightarrow P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) \leq \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9} = 0.11\bar{1} \dots$
3. Sei $\Delta x = 4\sigma \Rightarrow P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) \leq \frac{\sigma^2}{(4\sigma)^2} = \frac{1}{16} = 0.0625.$

Damit lassen sich also, wie man sieht, Wahrscheinlichkeiten von Werten abschätzen, welche „unvernünftig“ weit weg vom Erwartungswert μ liegen. Sei zum Beispiel σ bekannt. Dann ist die Wahrscheinlichkeit, dass ein Wert x_k ausserhalb des Intervalls $I_{3\sigma} = [\mu - 3\sigma, \mu + 3\sigma]$ liegt, nie grösser als $0.11\bar{1} \dots$

Folgerung:

Damit lässt sich aussagen, dass bei in einer beliebigen Grundgesamtheit mit endlichem μ und σ immer ca. 89 % der Werte im Intervall $I_{3\sigma}$ liegen!

Ausreisser sind demnach wenig wahrscheinlich.

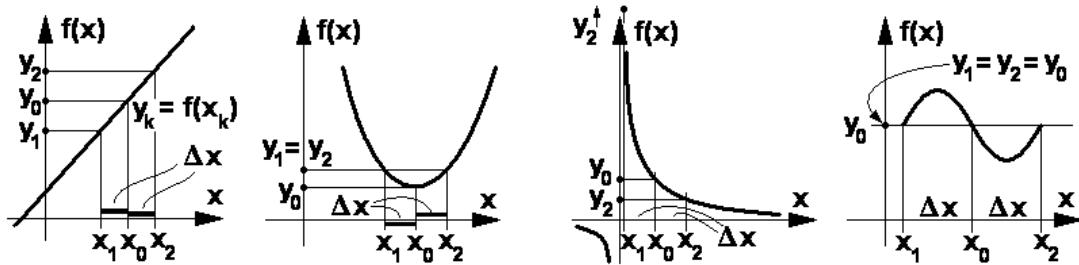
D.8 Messfehler (funktionale Abhängigkeit)

Auszug aus: Wirz, Bibl. A15, Script ◇ Math ◇ Ing, ◇ Analysis ◇ Analyse ◇, kurz & bündig. Mit Ergänzung

D.8.1 Das Problem der Verpflanzung — Le problème de la dépendance

Situation: (a): $y = f(x)$

Gemessen wird $x \pm \Delta x \rightsquigarrow y \pm \Delta y = ?$



$$y_0 = f(x_0), \quad y_1 = f(x_1), \quad y_2 = f(x_2), \quad \Delta y_1 = |y_1 - y_0| = ?, \quad \Delta y_2 = |y_2 - y_0| = ?$$

Situation: (b): $y = f(x_{10}, x_{20}, \dots, x_{n0})$

Gemessen werden die Werte $x_{10}, x_{20}, \dots, x_{n0}$ der Variablen x_1, x_2, \dots, x_n . Bei kontinuierlichen Messwerten gibt es immer Ableseungenauigkeiten, die aber abschätzbar sind. Diese zugehörigen „Messfehler“ betragen $\Delta x_1, \Delta x_2, \dots, \Delta x_n$. Die „exakten Werte“ $x_k^*, k = 1, \dots, n$ liegen daher in den Intervallen $[x_{k0} - \Delta x_k, x_{k0} + \Delta x_k]$. Zudem sei eine Funktion $f(x_1, x_2, \dots, x_n)$ gegeben, mit deren Hilfe eine weitere Grösse berechnet werden muss.

Problem: In welchem Intervall liegt der „wahre“ Wert $f(x_1^*, x_2^*, \dots, x_n^*)$?

D.8.2 Verwendung des totalen Differentials — Appliquer la différentielle totale

Sei $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\vec{x}_0 = \begin{pmatrix} x_{10} \\ \vdots \\ x_{n0} \end{pmatrix}$, $f(\vec{x}) := f(x_1, x_2, \dots, x_n)$, $D_k \geq |\Delta x_k|$ (D_k ist eine bezifferbare Schranke.)

Aus der Theorie des **totalen Differentials** weiss man:

$$\Delta f = f(\vec{x}_0 + \Delta \vec{x}) - f(\vec{x}_0) = \Delta x_1 f'_{x_1}(\vec{x}_0) + \dots + \Delta x_n f'_{x_n}(\vec{x}_0) + O[2] \\ (O: \text{Glieder höherer Ordnung})$$

$$\rightsquigarrow \Delta f \approx \Delta x_1 f'_{x_1}(\vec{x}_0) + \dots + \Delta x_n f'_{x_n}(\vec{x}_0)$$

$$\rightsquigarrow |\Delta f| \leq |\Delta x_1| |f'_{x_1}(\vec{x}_0)| + \dots + |\Delta x_n| |f'_{x_n}(\vec{x}_0)| \leq D_1 |f'_{x_1}(\vec{x}_0)| + \dots + D_n |f'_{x_n}(\vec{x}_0)| := \Delta f_{max}$$

Satz:**Vor.:**

Messsituation wie oben beschrieben ,
 $f \in \mathcal{D}^{(1)}$

Beh.:

$$|\Delta f| \leq D_1 |f'_{x_1}(\vec{x}_0)| + \dots + D_n |f'_{x_n}(\vec{x}_0)| := \Delta f_{max}$$

Konsequenz:

$$f(\vec{x}^*) = f(x_1^*, x_2^*, \dots, x_n^*) \in [f(\vec{x}_0) - \Delta f_{max}, f(\vec{x}_0) + \Delta f_{max}]$$

Definition:

$|\Delta f|$ heisst **absoluter Fehler** ,

$|\frac{\Delta f}{f(\vec{x}_0)}|$ heisst **relativer Fehler** .

1. Beispiel: $f(x, y) = x \pm y \Rightarrow \Delta f_{max} = D_x |1| + D_y |1| = D_x + D_y$

2. Beispiel: $f(x, y) = x \cdot y \Rightarrow \Delta f_{max} = D_x |y_0| + D_y |x_0|$

3. Beispiel: $f(x, y) = \frac{x}{y} \Rightarrow \Delta f_{max} = D_x |\frac{1}{y_0}| + D_y |\frac{x_0}{y_0^2}|$

4. Beispiel: $f(x, y) = x^y \Rightarrow \Delta f_{max} = D_x |y_0 \cdot x_0^{y_0-1}| + D_y |x_0^{y_0} \ln(x_0)|$

5. Beispiel:

$$f(x) = x^2 - 2x + 4 - \sin(x) + \ln(x) \Rightarrow \Delta f_{max} = D_x |2x_0 - 2 - \cos(x_0) + \frac{1}{x_0}|$$

Bemerkung:

Diese Beispiele zeigen, dass die oft geäusserte Meinung, es genüge mit den extremen Werten zu rechnen, wohl äusserst falsch sein muss.

6. Beispiel:

Messwerte:

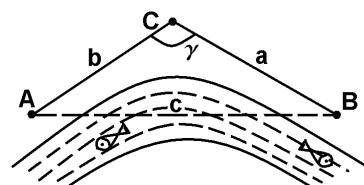
$$a = 364.76 \pm 0.05m$$

$$b = 402.35 \pm 0.05m$$

$$\gamma \hat{=} 68^\circ 14' \pm 4'$$

$$\rightsquigarrow \gamma \approx 1.1909 \pm 0.002$$

$$c = ?$$



$$\rightsquigarrow c = \sqrt{a^2 + b^2 - 2ab \cos(\gamma)} \approx 431.38$$

$$\Rightarrow \Delta c_{max} = D_a \cdot \left| \frac{\partial c}{\partial a} \right| + D_b \cdot \left| \frac{\partial c}{\partial b} \right| + D_\gamma \cdot \left| \frac{\partial c}{\partial \gamma} \right| =$$

$$= 0.05 \cdot \left| \frac{2a - 2b \cos(\gamma)}{2\sqrt{a^2 + b^2 - 2ab \cos(\gamma)}} \right| + 0.05 \cdot \left| \frac{2b - 2a \cos(\gamma)}{2\sqrt{a^2 + b^2 - 2ab \cos(\gamma)}} \right| + 0.002 \cdot \left| \frac{2ab \sin(\gamma)}{2\sqrt{a^2 + b^2 - 2ab \cos(\gamma)}} \right|$$

$$\approx 0.02498 + 0.03096 + \underbrace{0.36762}_{!!!} \approx 0.424 \Rightarrow c \pm \Delta c_{max} = 431.38 \pm 0.424$$

Anhang E

Monte-Carlo, Resampling und anderes

E.1 Monte-Carlo-Simulationen

E.1.1 Berechnung von Pi, Messeexperiment

Schulexperiment: Vielleicht haben Sie in der Schule anlässlich der Einführung der Zahl π einmal das nachfolgend beschriebene Experiment machen müssen: Der Lehrer stellt die Hausaufgabe, bei ca. 10 runden Gegenständen den Umfang und den Durchmesser mit Hilfe eines Messbandes zu messen. Die Resultate sollten in eine Tabelle eingetragen werden. Zu jedem Masspaar soll zudem das Verhältnis von Umfang zu Durchmesser berechnet werden und ebenfalls eingetragen werden. Es kommt jeweils eine Zahl zwischen ca. 3.0 und 3.3 heraus, denn die Schüler sind noch nicht so flink im exakten Messen. Dann wird der Mittelwert über die Klasse genommen. Das Resultat ist dann etwa 3.14. Der Lehrer erklärt: „Diese Zahl heisst π “. Ein Schüler ist sichtlich erstaunt, denn seine Messresultate weichen vom befohlenen „exakten“ Wert ab. Er sieht nicht ein, dass sein Resultat jetzt falsch sein soll. Der Lehrer hat sich zwar grosse Mühe gegeben, jedoch missachtet, dass die Schüler noch nicht in Ähnlichkeitsgeometrie unterrichtet worden sind. π wird hier mit der beschriebenen Methode als geschätzter Ähnlichkeitsfaktor gewonnen. Als Schätzer dient der Mittelwert.

Was hat nun dieses Experiment mit einem Zufallsexperiment zu tun? Sofort ist klar, dass die Auswahl der Gegenstände aus der Sicht des Lehrers zufällig erfolgt sind. Aus der Sicht der Schüler gilt dies jedoch kaum, denn jeder hat die nächstbesten Gegenstände vermessen: Büchsen, Blumentöpfe, Tassen, Teller, Rohre, schön rund oder auch verzogen. Jedoch keine Eisenbahnräder, keine Wellen von Schiffsmotoren u.s.w.. Zudem sind die Messfehler aus der Sicht des Lehrers zufällig, denn man kann kaum voraussehen, wer wann welche Fehler macht. Dagegen ist der errechnete Mittelwert ein Schätzer für den theoretischen Wert. Mit mehr Messungen wird er besser.

Hier haben wir es mit einem alltäglichen physikalischen Experiment zu tun. Eine Schätzung einer Grösse wird statistisch mit Hilfe eines Datensatzes gewonnen. Die obligatorische Fehlerrechnung fehlt noch im Bewusstsein. Der Wahrscheinlichkeitsbegriff spielt bei diesem einfachsten solchen Experiment nur am Rande eine Rolle, vielleicht qualitativ, jedoch nicht quantitativ.

In den beiden nachstehend beschriebenen Experimenten jedoch steht der klassische Wahrscheinlichkeitbegriff nach Laplace im Zentrum: Wir reden hier von **Monte-Carlo-Methoden**. Wiederum geht es hier um die Berechnung von π .

E.1.2 Das Konzept der Monte–Carlo–Simulation resp. –Studie

Die Idee ist nun, numerische Lösungen von mathematischen oder mathematisch beschriebenen Problemen via einen bekannten Zusammenhang zu Wahrscheinlichkeiten durch Zufallsexperimente zu gewinnen. Die Basis, auf die wir uns hier stützen, ist das Laplace–Experiment oder der Wahrscheinlichkeitsbegriff nach Laplace. Wir gehen von Elementarereignissen mit gleicher Wahrscheinlichkeit aus. Dann können wir die z.B. durch geometrische Zusammenhänge beschriebene Wahrscheinlichkeit eines Ereignisses A als Grenzwert der experimentell bestimmbarer relativen Häufigkeit $h(A)$ gewinnen, gestützt auf den bekannten Zusammenhang:

$$p(A) = \lim_{N \rightarrow \infty} \frac{H(A)}{N} = \lim_{N \rightarrow \infty} h(A)$$

N ist dabei die Anzahl der im Experiment gemachten einzelnen Versuche (Gesamtzahl der Ereignisse) und $H(A)$ die absolute Häufigkeit für das Eintreffen des günstigen resp. erwarteten Ereignisses. Dabei lässt sich praktisch der Grenzwert nur annähern, jedoch nicht exakt bestimmen. Das kann aber je nach Fall trotzdem von grossem Nutzen sein.

Bezüglich der Experimente können wir zwei Methoden unterscheiden:

1. Reales Experiment, z.B. mit dem Galton–Brett (vgl. <http://de.wikipedia.org/wiki/Galtonbrett> oder *Kurs Wahrscheinlichkeit & Statistik, Spezielle stetige Verteilungen, Bemerkung zum Zufall*).
2. Computer–Simulation: Hier erzeugt man die Zufallsresultate unter Benützung eines „Zufallsgenerators“ wie in den unten aufgeführten Beispielen.

E.1.3 Berechnung von Pi, Monte–Carlo–Flächensimulation

In der nebenstehenden Skizze hat das Quadrat

den Inhalt $I_Q = 1 \cdot 1 = 1$ und der Viertelskreis

$$\text{den Inhalt } I_V = 1^2 \cdot \pi \cdot \frac{1}{4} = \frac{\pi}{4}.$$

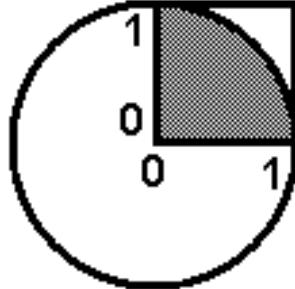
Auf einem beliebig fein gerasterten und möglichst exakten horizontal liegenden Plan der gezeigten Situation mit dem Quadrat werfen wir nach dem Zufallsprinzip einen feinen Gegenstand so, dass kein Zielpunkt vor den andern beim Wurf ausgezeichnet ist.

Das zu notierende Resultat kann entweder sein: „im Viertelskreis liegend“ oder „nicht im Viertelskreis, aber im Quadrat liegend“. Resultate ausserhalb des Quadrates notieren wir nicht. Da alle möglichen Zielpunkte gleich wahrscheinlich sind, bilden die Punkte im Quadrat das Mass für die möglichen Fälle und die Punkte im Viertelskreis das Mass für die günstigen Fälle.

$$\rightsquigarrow \frac{\text{Günstige Fälle}}{\text{Mögliche Fälle}} = \frac{I_V}{I_Q} = \frac{\pi}{4} \approx h(A) = \frac{H(A)}{N} \rightsquigarrow \pi \approx 4 \cdot \frac{H(A)}{N}.$$

Nachstehend sind die Resultate einer Simulation mit einem *Mathematica*–Programm aufgeführt:

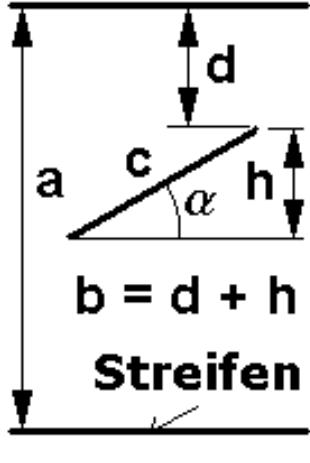
```
(* Programm *) nN = 10000;
f[x_,y_]:=0 /;(y> Sin[ArcTan[y/x]]); f[x_,y_]:=1 /;(y<=Sin[ArcTan[y/x]]);
t[k_]:=Table[f[Random[],Random[]],{n,1,nN}];
u[k_]:=4 (Plus@@t[k])/Length[t[k]] //N;
Print[Table[u[k],{k,1,10}]];
Mean[Table[u[k],{k,1,10}]]
```



```
Output: {3.1316, 3.1612, 3.1476, 3.15, 3.1364,
         3.1096, 3.1428, 3.1272, 3.184, 3.156}
Mean (Mittelwert): 3.13932
```

E.1.4 Berechnung von Pi, Buffon, Nadelexperiment

Eine andere Methode zur Bestimmung von π mit Hilfe eines Monte-Carlo-Experimentes ist das Nadelexperiment von Buffon.



Eine Nadel der Länge c wird nach dem Zufallsprinzip auf einen Plan geworfen, auf dem in gleichen Abständen a parallele Streifen angebracht sind. Dabei sei Ereignis A definiert als das Eintreffen des Falls, dass die Nadel einen Streifen trifft. Wir fragen nach der Wahrscheinlichkeit $p(A)$, dass A eintrifft. Hier hat man eine geometrische Situation und folglich eine „geometrische Wahrscheinlichkeit“. d sei die Distanz der Nadel zum nächsten oberen Streifen, c die Nadellänge, $h = c \cdot \sin(\alpha)$ und $b = d + h$ die Distanz der unteren Nadelspitze zum nächsten oberen Streifen. Wir lesen aus der Skizze ab, dass die Nadel genau dann einen Streifen trifft, wenn die Bedingung $0 \leq b \leq h$ erfüllt ist ($-h \leq d \leq 0$).

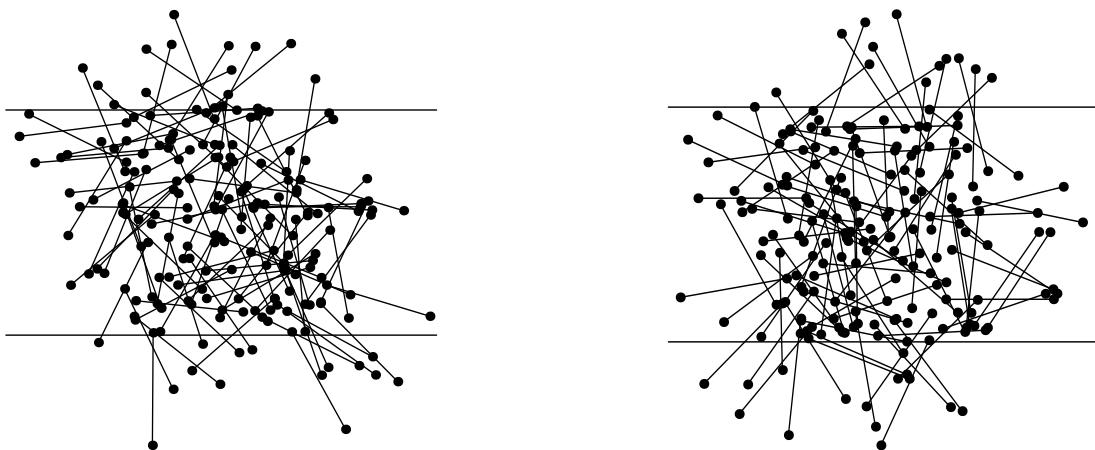
Die Geometrie hier diktiert uns die Bedingungen: $d \in [0, a]$ und $\alpha \in [0, \pi]$. Daher finden wir die Menge der möglichen Fälle als Paarmenge $A_m = \{(d, \alpha) \in [0, a] \times [0, \pi]\}$ und die Menge der günstigen Fälle $G = \{(b, \alpha) \in [0, h(\alpha)] \times [0, \pi]\}$. Das Mass (Flächeninhalt) von A_m ist das Produkt $a \cdot \pi$. Wir schreiben kurz $A_m = a \cdot \pi$. Das Mass von A_g hingegen wegen der Abhängigkeit $b \leq h(\alpha)$ das Integral (Flächeninhalt unter

der Kurve $y = h(\alpha)$): $A_g = \int_0^{\pi} h(\alpha) d\alpha = \int_0^{\pi} c \sin(\alpha) d\alpha = c \cdot (-\cos(\alpha)) \Big|_0^{\pi} = 2c \Rightarrow p(A) = \frac{A_g}{A_m} = \frac{2c}{a \cdot \pi}$

Wählen wir nun die Nadel halb so lang wie a , d.h. $a = 2c$, so wird $p(A) = \frac{1}{\pi}$ und $\frac{1}{p(A)} = \frac{A_m}{A_g} = \pi$.

Damit lässt sich π bestimmen. Denn es folgt: $\pi = \frac{A_m}{A_g} \approx \frac{N}{H(A)}$.

Eine Simulation mit Hilfe von *Mathematica* ergibt Wurfbilder wie die nachfolgend gezeigten Beispiele:



Mathematica-Programm:

```
k=100; (* 100 Würfe *)
p[n_]:={Random[],Random[]};
tp=Table[p[n],{n,1,k}];
a[n_]:=2 Pi Random[];
ta=Table[a[n],{n,1,k}];
q[n_]:=tp[[n]]+0.5 {Sin[ta[[n]]],Cos[ta[[n]]]};
tq=Table[q[n],{n,1,k}];
poi[n_]:=Point[tp[[n]]];
qoi[n_]:=Point[tq[[n]]];
lin[n_]:=Line[{tp[[n]],tq[[n]]}];
Show[Graphics[Prepend[
Union[
Flatten[Table[{poi[n],qoi[n],lin[n]},{n,1,k}]],
{Line[{{-0.5,0},{1.5,0}}],Line[{{-0.5,1},{1.5,1}}]}]
,PointSize[0.02]],AspectRatio->Automatic];
```

Simulation von π :

```
k=1000;
tabOut=Table[
p[n_]:={Random[],Random[]};
tp=Table[p[n],{n,1,k}];
a[n_]:=2 Pi Random[];
ta=Table[a[n],{n,1,k}];
q[n_]:=tp[[n]]+0.5 {Cos[ta[[n]]],Sin[ta[[n]]]};
tq=Table[q[n],{n,1,k}];
ty=Table[{Min[{tp[[n]][[2]],tq[[n]][[2]]}],Max[{tp[[n]][[2]],tq[[n]][[2]]}]},{n,1,k}];
tw=Table[{Floor[Min[{tp[[n]][[2]],tq[[n]][[2]]}]],Floor[Max[{tp[[n]][[2]],tq[[n]][[2]]}]]},{n,1,k}];
tz=Table[{Floor[Min[{tp[[n]][[2]],tq[[n]][[2]]}]]^2,Floor[Max[{tp[[n]][[2]],tq[[n]][[2]]}]]^2},k/Apply[Plus,Flatten[tz]]//N,
{u,1,10}];
Print[tabOut];
Print["Mittelwert = ",Apply[Plus,tabOut]/Length[tabOut]]
```

Output:

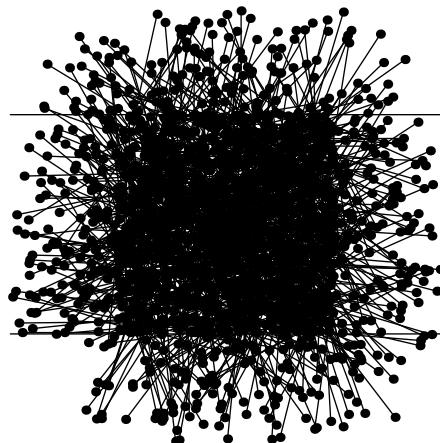
```
{2.99401, 3.18471, 2.85714, 3.04878, 3.10559, 2.98507,
3.1746, 3.30033, 3.26797, 3.47222}
```

Output:

```
Mittelwert = 3.13904
```

Eine Simulation mit 1000 Versuchen:

„Wie man sieht, sieht man nichts mehr...“



Literatur vgl. Links:

<http://de.wikipedia.org/wiki/Monte-Carlo-Simulation>

http://de.wikipedia.org/wiki/Kreiszahl#Buffonsches_Nadelproblem

http://www.mathematik.ch/anwendungenmath/wkeit/buffon/Buffon_Nadelproblem.php

<http://www.fh-friedberg.de/users/jingo/mathematics/buffon/buffonneedle.html>

u.s.w.

E.2 Resampling-Methoden (Einführung)

E.2.1 Die Idee von Bootstrap und Jackknife

Als blosse Demonstration, d.h. ohne auf die mathematischen Fundamente tiefer einzugehen, wollen wir hier auf die beiden für den Gebrauch wesentlichsten Typen von Resampling eingehen: Auf Bootstrap und Jackknife.

Zur Idee

Gegeben sei z.B. eine Stichprobe in Form eines Vektors: $\vec{x}^T = (x_1, \dots, x_n)$. Aus der Komponentenmenge von \vec{x}^T ziehen wir nun *mit Zurücklegen* neue, ev. kleinere Stichprobenvektoren $\vec{x}_k^T = (x_{1,k}, \dots, x_{j,k})$ mit $1 \leq j \leq n, 1 \leq k \leq m, j, k, m \in \mathbb{N}$.

Diese neuen Vektoren $\{\vec{x}_1^T, \dots, \vec{x}_m^T\}$ benutzen wir zur Schätzung von statistischen Kenngrößen wie **Bias** (Differenz zwischen Erwartungswert eines Schätzers und dessen wahrer, vorliegenden Approximation), Standardfehler der Schätzer, Mittelwert, Median, Quantile, Korrelationskoeffizient u.s.w., Konfidenzintervalle, Grundlagen für Testhypothesen, u.s.w.. Die zugrunde liegende theoretische Verteilung $F(x)$ muss dabei nicht bekannt sein. Wir ersetzen hier sogar die theoretische Verteilung der x_i durch die empirische Verteilung $F_N(x)$ der Stichprobe.

Vorgehen

Methode für Bootstrap: Ziehe aus \vec{x}^T heraus m Stichproben $\vec{x}_k^T, k = 1, \dots, m$ vom Umfang (resp. der Länge) $j = n, m$ beliebig aber fix.

Bei Bootstrap hängen somit die weiter zu erzielenden Resultate von m ab. Es werden hier aus einem Datensatz \vec{x}^T neue Datensätze gleicher Länge generiert, denen die genau gleichen, meist aber unbekannten Verteilungen zugrunde liegen wie dem Ausgangsdatensatz \vec{x}^T . Wenn wir z.B. in jedem der neu generierten Datensätze \vec{x}_k^T den Median berechnen, können wir auch den Mittelwert und die Varianz dieser Mediane berechnen. Und aus den gewonnenen und geordneten m „Sampels“ können wir dann Konfidenzintervalle ableiten. Haben wir z.B. $m = 200$ gesetzt, also 200 Sampels generiert, so lesen wir ein 95 %-Konfidenzintervall zwischen den Stellen 5 und 195 ab. Will man effizient arbeiten, so benötigt hier zwar einen Computer, hat aber dagegen nur einen minimalen Aufwand an Verständnis zu bewältigen.

Methode für Jackknife: Wir gehen vor wie bei Bootstrap, jedoch mit Stichprobenumfang $j = n - 1, m = n$. D.h. wir lassen bei jeder Stichprobe eines der Elemente x_i weg.
 $\rightsquigarrow \vec{x}_i^T = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$.

Bei Jackknife entsteht im Gegensatz zu Bootstrap keine Abhängigkeit von m .

Anwendung: Berechne für die Menge $\{\vec{x}_i^T \mid i = 1, \dots, j\}$, (j je nach Bootstrap oder Jackknife gewählt) die im Interessen stehende Grösse $\Theta(\vec{x}_i^T)$ und approximiere damit die Zielgrösse $\Theta(\vec{x}^T)$.

Die Verteilung von $\Theta(\vec{x}^T)$ wird damit also durch die empirische Verteilung der $\Theta(\vec{x}_i^T)$ approximiert.

Bemerkung:

Oft hat man das Problem, bestimmte Kenngrößen einer unbekannten Verteilung zu schätzen und daraus z.B. mit Hilfe von Bootstrap eine Verteilung der geschätzten Grösse zu generieren. Speziell ist dies so, wenn keine Anhaltspunkte über die Grenzverteilung vorliegen.

Statt Annahmen über die Parameter zu treffen, werden hier also statistische Veränderungen zwischen Untermengen der Stichprobe zur Schätzung von Größen verwendet.

Schätzer und Bias

Definition: Als **Bias** bezeichnen wir die Differenz zwischen Erwartungswert eines Schätzers und dessen wahrer, vorliegenden Approximation.

Sei $\hat{\Theta} = \hat{\Theta}(\vec{x}^T)$ ein Schätzer für $\Theta = \Theta(\vec{X}^T)$ und sei $\hat{\Theta}_B = \frac{1}{m} \sum_{k=1}^m \hat{\Theta}(\vec{x}_k^T)$ der Mittelwert der aus den Bootstrapping-Vektoren (Sampels, \vec{x}_k^T) gewonnenen Werte zur gegebenen Funktion $\hat{\Theta}$ (Parameter, Prüfgröße oder andere abhängige Variable). Dann gilt bei Bootstrapping (ohne Herleitung):

Formel:

Bootstrapping:

$$\widehat{Bias}_{B,\Theta} = \hat{\Theta}_B - \hat{\Theta}.$$

Bei Jackknife verhält es sich damit etwas anders. Bedeutet z.B. $\hat{\Theta}(\vec{x}_k^T)$ der Mittelwert des k -ten Sampels, d.h. $\hat{\Theta}(\vec{x}_k^T) = \frac{1}{n-1} \sum_{i=1, i \neq k}^n x_i$, so kommt in der Summe $\sum_{k=1}^n \hat{\Theta}(\vec{x}_k^T)$ jedes x_i genau $n-1$ mal vor, und es ist $\frac{n-1}{n} \sum_{k=1}^n \hat{\Theta}(\vec{x}_k^T) = \frac{n-1}{n} \sum_{k=1}^n \frac{1}{n-1} \left(\sum_{i=1, i \neq k}^n x_i \right) = \frac{n-1}{n} \sum_{k=1}^n x_k = (n-1) \hat{\Theta}(\vec{x}^T)$.

Man hat also hier Übereinstimmung mit der Formel für den Mittelwert: $(\frac{1}{n} \sum_{k=1}^n \hat{\Theta}(\vec{x}_k^T)) - \hat{\Theta}(\vec{x}^T) = 0$

$$\begin{aligned} \text{Sei } \hat{\Theta}_J := (\frac{1}{n} \sum_{k=1}^n \hat{\Theta}(\vec{x}_k^T)) \Rightarrow (n-1)(\hat{\Theta}_J - \hat{\Theta}(\vec{x}^T)) = 0 &= (n-1)\hat{\Theta}_J - (n-1)\hat{\Theta}(\vec{x}^T) \\ &\Rightarrow \tilde{\Theta}_J = n\hat{\Theta}_J - (n-1)\hat{\Theta}(\vec{x}^T). \end{aligned}$$

Allgemeiner ist der *Jackknife-korrigierte Schätzer*: $\tilde{\Theta}_J = n\hat{\Theta}_J - (n-1)\hat{\Theta}(\vec{x}^T)$

Damit gilt (ohne Herleitung):

Formel:

Jackknife:

$$\widehat{Bias}_{J,\Theta} = \tilde{\Theta}_J - \hat{\Theta}_J = (n-1)(\hat{\Theta}_J - \hat{\Theta}(\vec{x}^T))$$

$$\tilde{\Theta}_J = n\hat{\Theta}_J - (n-1)\hat{\Theta}(\vec{x}^T)$$

Weiter gilt z.B. für den Jackknife Standardfehlerschätzer:

Formel:

Jackknife:

$$SD_{\tilde{\Theta}} = \sqrt{\frac{n-1}{n} \sum_{k=1}^n (\hat{\Theta}(\vec{x}_k^T) - \hat{\Theta}_J)^2}$$

E.2.2 Resampling, Beispiele

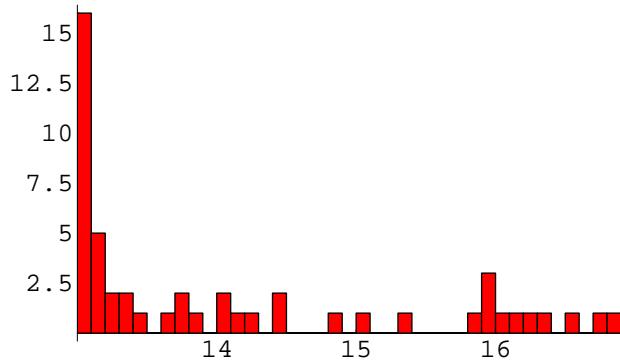
1. Beispiel: Eine Simulation einer Jackknife-Anwendung mit *Mathematica*:

```
(* Add-ons zuladen *)
<<Statistics`DescriptiveStatistics`;
<< Graphics`Graphics`;

(* Datengenerierung *)
m=50;
ind[j_]:=Table[12.+Floor[40 (Random[]^3)]/10.+1 ,{k,1,m}];
tabx = ind[0]

{13.1, 13., 13.3, 13.4, 14.4, 13., 14., 13., 13., 15.9, 14.4, 13.8, 13.6, 16.3, 13.2, 15.3,
14.8, 13., 15.9, 16.8, 13.1, 13.1, 14.2, 14., 13., 16., 13., 16.2, 16.5, 13., 16.1, 14.1,
15.8, 13.2, 13., 13.1, 13.3, 15., 13., 13.7, 13.1, 15.9, 13., 13., 16.7, 13., 13.7, 13.}

(* Histogramm der generierten Daten *)
Histogram[tabx,HistogramCategories -> Table[k,{k,12,20,0.1}]];
```



```
(* Mittelwert der generierten Daten *)
meanTabx=Mean[tabx]
```

14.06

```
(* Resampling: Jackknife *)
sampl[k_]:= Drop[tabx,{k}];
meanSamplek[k_]:=Mean[sampl[k]];
tabMean=Table[meanSamplek[k],{k,1,Length[tabx]}]
```

```
{14.0796, 14.0816, 14.0755, 14.0735, 14.0531, 14.0816, 14.0612, 14.0816, 14.0816, 14.0816,
14.0224, 14.0531, 14.0653, 14.0694, 14.0143, 14.0776, 14.0347, 14.0449, 14.0816, 14.0224,
14.0041, 14.0796, 14.0796, 14.0571, 14.0612, 14.0816, 14.0204, 14.0816, 14.0163, 14.0102,
14.0816, 14.0184, 14.0592, 14.0245, 14.0776, 14.0816, 14.0816, 14.0796, 14.0755, 14.0408,
14.0816, 14.0673, 14.0796, 14.0224, 14.0816, 14.0816, 14.0061, 14.0816, 14.0673, 14.0816}
```

```
(* Mean Jackknifedaten zur Kontrolle *)
meanJack=Mean[tabMean]
```

14.06

```
meanTabx == meanJack
```

True

```
(* Deskriptive Statistik: Report 1 *)
LocationReport[tabMean]

{Mean → 14.06, HarmonicMean → 14.06, Median → 14.0745}

(* Deskriptive Statistik: Report 2 *)
DispersionReport[tabMean]

{Variance → 0.000684749, StandardDeviation → 0.0261677,
SampleRange → 0.077551, MeanDeviation → 0.0222204,
MedianDeviation → 0.00714286, QuartileDeviation → 0.0204082}

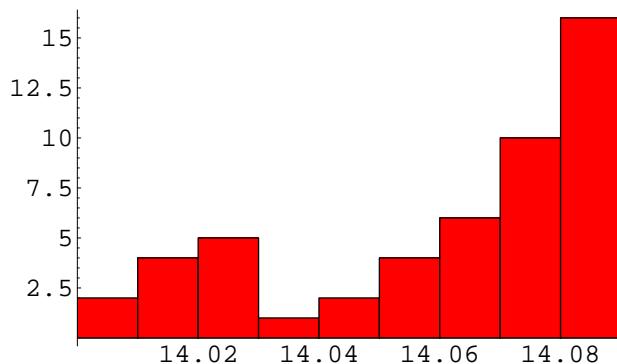
(* Deskriptive Statistik: Report 3 *)
ShapeReport[tabMean]

{Skewness → -0.893652, QuartileSkewness → -0.65, KurtosisExcess → -0.754031}

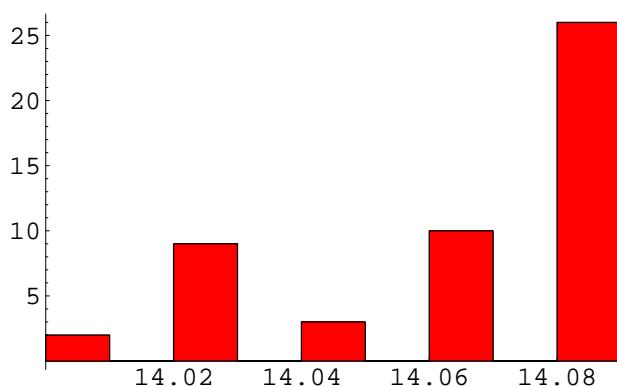
(* Rundungen der generierten Mittelwerte *)
tabMeanW=Round[50 tabMean]/50.

{14.08, 14.08, 14.08, 14.08, 14.06, 14.08, 14.06, 14.08, 14.08, 14.02, 14.06,
14.06, 14.06, 14.02, 14.08, 14.04, 14.04, 14.08, 14.02, 14., 14.08, 14.08, 14.06, 14.06,
14.08, 14.02, 14.08, 14.02, 14.02, 14.08, 14.02, 14.06, 14.02, 14.08, 14.08, 14.08,
14.08, 14.08, 14.04, 14.08, 14.06, 14.08, 14.02, 14.08, 14.08, 14., 14.08, 14.06, 14.08}
```

(* Histogramm der generierten Mittelwerte *)
Histogram[tabMean];



(* Histogramm der generierten gerundeten Mittelwerte *)
Histogram[tabMeanW];



```
(* Jackknife Standardfehlerschätzer *)
SDtabMean=Sqrt[(Length[tabx]-1)/Length[tabx] *
Apply[Plus, Table[(tabMean[[k]]-Mean[tabMean])^2 ,{k,1,Length[tabx]}]]]
0.181333

(* Standardabweichung der generierten Mittelwerte *)
StandardDeviation[tabMean]

0.0261677
```

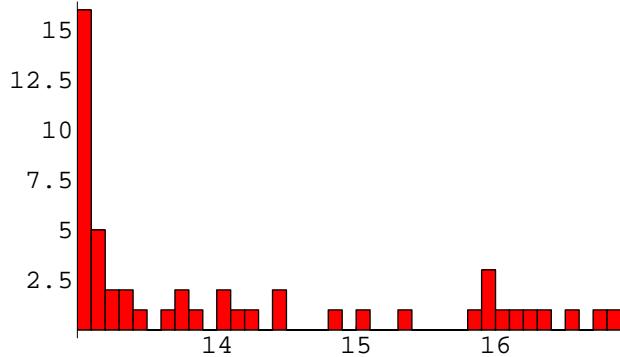
2. Beispiel: Eine Simulation einer Bootstrap-Anwendung mit *Mathematica*:

```
(* Add-ons zuladen *)
<<Statistics`DescriptiveStatistics`;
<< Graphics`Graphics`;

(* Daten vo Jackknife übernommen, hier gleiche Session *)
tabx

{13.1, 13., 13.3, 13.4, 14.4, 13., 14., 13., 13., 15.9, 14.4, 13.8, 13.6, 16.3, 13.2, 15.3,
14.8, 13., 15.9, 16.8, 13.1, 13.1, 14.2, 14., 13., 16., 13., 16.2, 16.5, 13., 16.1, 14.1,
15.8, 13.2, 13., 13.1, 13.3, 15., 13., 13.7, 13.1, 15.9, 13., 13., 16.7, 13., 13.7, 13.}

(* Histogramm der generierten Daten *)
Histogram[tabx, HistogramCategories -> Table[k,{k,12,20,0.1}]];
```



```
(* Mitelwert der generierten Daten *)
meanTabx=Mean[tabx]

14.06

(* Resampling: Bootstrap *)
bootstrap=50; (* Anzahl Samples *)
randInt[k_]:=Table[Random[Integer,{1,m}],{j,1,m}];
Table[randIntFix[k]=randInt[k],{k,1,bootstrap}];
sampl[k_]:= Table[tabx[[randIntFix[k][[j]]]],{j,1,m}];
meanSamplek[k_]:=Mean[sampl[k]];
tabMean=Table[meanSamplek[k],{k,1,bootstrap}]

{13.808, 14.126, 14.274, 13.934, 14.136, 14.082, 13.986, 14.082, 14.07, 14.016, 14.206, 14.078,
13.914, 13.984, 14.044, 14.14, 14.276, 13.958, 14.004, 14.376, 13.954, 14.054, 14.09, 14.308, 13.986,
14.26, 13.904, 13.972, 13.784, 13.734, 14.292, 14.348, 13.834, 14.03, 13.842, 14.34, 14.012, 13.744,
13.942, 13.878, 13.966, 14.528, 13.98, 13.874, 14.124, 14.07, 13.916, 13.822, 14.106, 14.168}
```

```
(* Mean Bootstrapdaten zur Kontrolle *)
meanBootstrap=Mean[tabMean]
```

14.0471

```
meanTabx==meanBootstrap
```

False

```
(* Deskriptive Statistik: Report 1 *)
LocationReport[tabMean]
```

$\{Mean \rightarrow 14.0471, HarmonicMean \rightarrow 14.045, Median \rightarrow 14.023\}$

```
(* Deskriptive Statistik: Report 2 *)
DispersionReport[tabMean]
```

$\{Variance \rightarrow 0.0308715, StandardDeviation \rightarrow 0.175703, SampleRange \rightarrow 0.794,$
 $MeanDeviation \rightarrow 0.13801, MedianDeviation \rightarrow 0.105, QuartileDeviation \rightarrow 0.101\}$

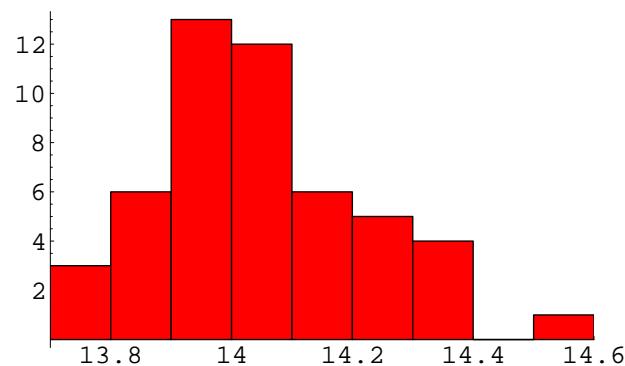
```
(* Deskriptive Statistik: Report 3 *)
ShapeReport[tabMean]
```

$\{Skewness \rightarrow 0.492396, QuartileSkewness \rightarrow 0.118812, KurtosisExcess \rightarrow -0.0869509\}$

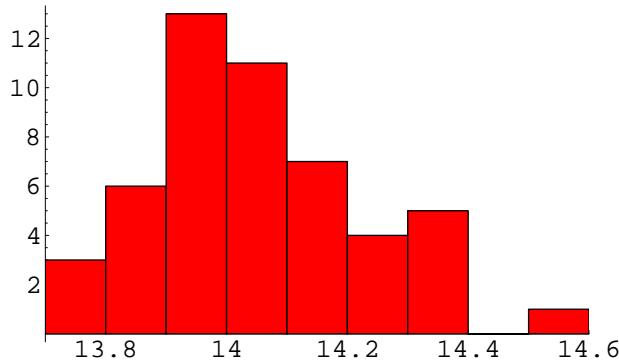
```
(* Rundungen der generierten Mittelwerte *)
tabMeanW=Round[50 tabMean]/50.
```

$\{13.8, 14.12, 14.28, 13.94, 14.14, 14.08, 13.98, 14.08, 14.08, 14.02, 14.2, 14.08,$
 $13.92, 13.98, 14.04, 14.14, 14.28, 13.96, 14., 14.38, 13.96, 14.06, 14.1, 14.3, 13.98,$
 $14.26, 13.9, 13.98, 13.78, 13.74, 14.3, 14.34, 13.84, 14.04, 13.84, 14.34, 14.02, 13.74,$
 $13.94, 13.88, 13.96, 14.52, 13.98, 13.88, 14.12, 14.06, 13.92, 13.82, 14.1, 14.16\}$

```
(* Histogramm der generierten Mittelwerte *)
Histogram[tabMean];
```



```
(* Histogramm der generierten gerundeten Mittelwerte *)
Histogram[tabMeanW];
```



```
(* Standardabweichung der generierten Mittelwerte *)
StandardDeviation[tabMean]
```

0.175703

```
(* Gesamtmittelwert und Standardabweichungen bei wiederholtem Bootstrap *)
tabRes[w_]:=Table[
randInt[k_]:=Table[Random[Integer,{1,m}],{j,1,m}];
Table[randIntFix[k]=randInt[k],{k,1,bootstrap}];
sampl[k_]:= Table[tabx[((randIntFix[k][[j]])]],{j,1,m}];
meanSamplek[k_]:=Mean[sampl[k]];
tabMean=Table[meanSamplek[k],{k,1,bootstrap}];
meanBootstrapN[h]=Mean[tabMean];
StandardDeviationN[h]=StandardDeviation[tabMean];
Print["Nr. = ",h," Mean = ",meanBootstrapN[h]," SD = ",StandardDeviationN[h]];
If[h==w,Print["Wiederholbar mit verändertem h!"," Stop"]],{h,1,w}];
tabRes[15]
```

Nr. = 1, Mean = 14.0541, SD = 0.168073

Nr. = 2, Mean = 14.0939, SD = 0.197083

Nr. = 3, Mean = 14.0633, SD = 0.166843

Nr. = 4, Mean = 14.0446, SD = 0.165617

Nr. = 5, Mean = 14.0573, SD = 0.204615

Nr. = 6, Mean = 14.0756, SD = 0.176742

Nr. = 7, Mean = 14.0595, SD = 0.188485

Nr. = 8, Mean = 14.0603, SD = 0.195219

Nr. = 9, Mean = 14.0214, SD = 0.204079

Nr. = 10, Mean = 14.0478, SD = 0.170797

Nr. = 11, Mean = 14.0648, SD = 0.184593

Nr. = 12, Mean = 14.1158, SD = 0.209445

Nr. = 13, Mean = 14.0629, SD = 0.155767

Nr. = 14, Mean = 14.0468, SD = 0.184151

Nr. = 15, Mean = 14.057, SD = 0.164297

Wiederholbar mit verändertem h! Stop

Achtung:

Bei Bootstrap ändern die Sample-Daten und damit die statistischen Kenngrößen bei Wiederholung des Samplings! \leadsto Die daraus zu ziehenden Schlüsse sind dem Anwender überlassen. (Eine weitere Strategie hängt auch von den Rohdaten ab.)

Links zum Thema

[http://en.wikipedia.org/wiki/Resampling_\(statistics\)](http://en.wikipedia.org/wiki/Resampling_(statistics))
<http://de.wikipedia.org/wiki/Bootstrapping>
[http://en.wikipedia.org/wiki/Bootstrapping_\(statistics\)](http://en.wikipedia.org/wiki/Bootstrapping_(statistics))
[http://en.wikipedia.org/wiki/Resampling_\(statistics\)#Bootstrap](http://en.wikipedia.org/wiki/Resampling_(statistics)#Bootstrap)
<http://people.revoledu.com/kardi/tutorial/Bootstrap/index.html>
<http://de.wikipedia.org/wiki/Varianzschätzung>
<http://de.wikipedia.org/wiki/Bias>
<http://www.statistik.uni-muenchen.de/institut/ag/biostat/teaching/compint2004/vorlesung/slides5.pdf>
<http://www.ufz.de/data/Dormann2004Statsskript1625.pdf>
<http://www.univet.hu/users/jreiczig/dortmund2005/Dortmund-Resampling-Folien-Boot1.pdf>

Anhang F

Eine Bootstrap–Anwendung Schritt für Schritt (mit *Mathematica*)

F.1 Aufgabenstellung und Konzept

F.1.1 Das praktische Problem der Verteilungsfunktion

Bei nicht verteilungsfreien statistischen Methoden stösst man oft zwangsläufig auf das meist schwierige Problem der Modellierung der benötigten Verteilungsfunktionen. Wer das Problem kennt, weiss davon ein Liedchen zu singen. Oft ist man gezwungen Vermutungen über unbekannte Verteilungstypen auszusprechen und die Vermutungen dann eventuell noch überzeugend zu testen, wofür man oft wieder andere Vermutungen zu Hilfe nehmen muss, vielleicht nur explorative Daten hat und eventuell nochmals teuere Versuche durchführen muss. Ein anderer Zugang wird möglich, wenn man mittels Bootstrap–Methoden aus den vorhandenen Daten „empirische“ Verteilungsfunktionen gewinnt und dann diese verwendet. Das wiederum beruht auf einem theoretischen Fundament, das im Rahmen dieses Skriptums nicht behandelt werden kann. Es wird darauf in einem gesonderten Skript über *Grundzüge statistischer Datenanalyse* auf dem Niveau einer Bestandesaufnahme eingegangen.

F.1.2 Eine mögliche Aufgabenstellung

Aus einer Anzahl von Messwerten (Zufallsvariable X) soll der Mittelwert μ geschätzt und dazu ein 60 %-Vertrauensintervall für diesen Mittelwert angegeben werden. \bar{X} sei die dazu gehörige Zufallsvariable. Die Messwerte sollen in den hier gezeigten Beispielen behelfsmässig als Zufallszahlen generiert werden. Dabei kann man lernen, wie das geht.

Als Schätzer für den Mittelwert verwenden wir das arithmetische Mittel der Daten $\hat{\mu}$. Und als Modell für die Verteilungsfunktion $F(\bar{X})$ benutzen wir die aus den Daten durch eine Bootstrapping–Simulation gewonnene tatsächlichen Verteilungsfunktion $\hat{F}(\bar{X})$. Am damit gewonnenen Graphen von \hat{F} können wir dann das 60 %-Vertrauensintervall direkt ablesen (oder numerisch mit dem Computer eingrenzen).

F.2 Beispiel 1: Generierung einer Zahlenmenge als Messwerte

Die Funktion „Random“ (zur Wahl einer Zufallszahl zwischen 0 und 1)

Mathematica-Programm:

```
Remove["Global`*"];
Random[]
```

Output:

```
| 0.471221 |
```

Wahl einer ganzen Zufallszahl zwischen 1 und 12

Mathematica-Programm:

```
| Random[Integer, {1, 12}] |
```

Output:

```
| 7 |
```

Wahl von 20 ganzen Zufallszahlen zwischen 1 und 12

Mathematica-Programm:

```
| Table[Random[Integer, {1, 12}], {n, 1, 20}] |
```

Output:

```
| {9, 4, 4, 3, 3, 5, 10, 4, 7, 2, 5, 11, 4, 7, 1, 12, 9, 8, 9, 6} |
```

Wahl von 50 Zufallszahlen zwischen -10 und 10

Mathematica-Programm:

```
| M = Table[20 Random[] - 10, {n, 1, 50}] |
```

Output:

```
{8.79888, -5.83327, -9.996, -0.836214, 4.51456, 1.19145, -6.32316,
-7.70834, -5.77326, 0.728093, -2.14945, -1.26558, -3.28883, 5.49731,
1.25834, -7.98332, 1.09272, 3.94322, 0.580776, 6.4726, 0.0734423,
8.51339, -1.16745, -2.95183, 1.27456, 4.34666, -1.17145, 7.88439,
6.76, -6.84479, -4.8483, 5.59272, 2.53326, 2.42712, 7.30115, -3.1417,
-4.17792, 6.92981, -3.95719, -5.15838, 4.72936, -7.01341, 5.46204,
-1.63098, -5.34408, -5.5268, -3.37051, -8.67915, 3.38135, 0.126539}
```

Kontrolle:

Mathematica-Programm:

```
| M |
```

Output:

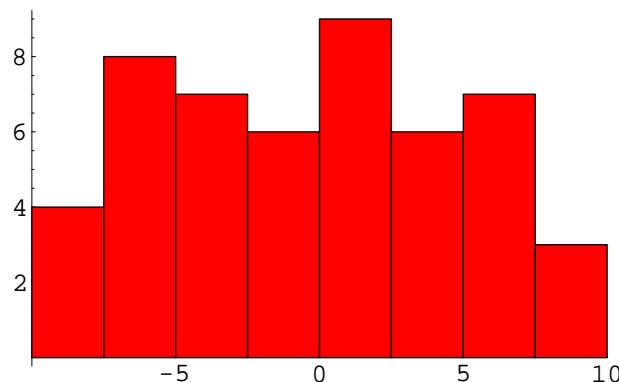
```
{8.79888, -5.83327, -9.996, -0.836214, 4.51456, 1.19145, -6.32316,
-7.70834, -5.77326, 0.728093, -2.14945, -1.26558, -3.28883, 5.49731,
1.25834, -7.98332, 1.09272, 3.94322, 0.580776, 6.4726, 0.0734423,
8.51339, -1.16745, -2.95183, 1.27456, 4.34666, -1.17145, 7.88439,
6.76, -6.84479, -4.8483, 5.59272, 2.53326, 2.42712, 7.30115, -3.1417,
-4.17792, 6.92981, -3.95719, -5.15838, 4.72936, -7.01341, 5.46204,
-1.63098, -5.34408, -5.5268, -3.37051, -8.67915, 3.38135, 0.126539}
```

F.3 Beispiel 1: Bearbeitung der Menge von Messwerten

Histogramm und Mittelwert von M

Mathematica-Programm:

```
<< Graphics`Graphics`;
Histogram[M];
M // Mean
```



Output:

```
| -0.294552 |
```

Die 10-te Zahl aus M

Mathematica-Programm:

```
| z1 = M[[10]] |
```

Output:

```
| 0.728093 |
```

Kontrolle:

Mathematica-Programm:

```
| z1 |
```

Output:

```
| 0.728093 |
```

Die n-te Zahl aus M, n = ganze Zufallszahl zwischen 1 und 50, $|M| = 50$

Mathematica-Programm:

```
| n1 = Random[Integer, {1, 50}]; z2 = M[[n1]]; {n1, z2} |
```

Output:

```
| {8, -7.70834} |
```

Kontrolle:

Mathematica-Programm:

```
| z2 |
```

Output:

```
| -7.70834 |
```

Eine Funktion u, die mit einer Nummer j eine Menge von 50 Zufallszahlen aus M so auswählt wie es eben geschehen ist (Exemplar Nummer j mit 50 Zahlen aus M)

Mathematica-Programm:

```
u[j_] := Table[M[[Random[Integer, {1, 50}]]], {k, 1, 50}];
tWork1 = u[1]
```

Output:

```
-3.95719, -3.28883, 3.38135, 5.59272, 0.0734423, -4.8483, 8.79888,
0.0734423, 5.49731, 7.88439, 3.38135, 1.25834, 8.51339, 4.34666,
-1.17145, 1.25834, 6.92981, 5.49731, -7.70834, 6.76, -1.63098, 3.94322,
5.46204, -5.77326, 4.51456, 8.79888, 8.51339, -2.95183, -3.37051,
1.19145, -8.67915, 2.53326, -5.77326, -9.996, 7.30115, -6.84479,
0.580776, 3.94322, -1.17145, 1.19145, -5.83327, 0.0734423, -3.28883,
2.42712, 6.4726, -7.70834, 0.0734423, 4.34666, 7.88439, 1.09272
```

Kontrolle:

Mathematica-Programm:

```
tWork1
```

Output:

```
-3.95719, -3.28883, 3.38135, 5.59272, 0.0734423, -4.8483, 8.79888,
0.0734423, 5.49731, 7.88439, 3.38135, 1.25834, 8.51339, 4.34666,
-1.17145, 1.25834, 6.92981, 5.49731, -7.70834, 6.76, -1.63098, 3.94322,
5.46204, -5.77326, 4.51456, 8.79888, 8.51339, -2.95183, -3.37051,
1.19145, -8.67915, 2.53326, -5.77326, -9.996, 7.30115, -6.84479,
0.580776, 3.94322, -1.17145, 1.19145, -5.83327, 0.0734423, -3.28883,
2.42712, 6.4726, -7.70834, 0.0734423, 4.34666, 7.88439, 1.09272
```

Kontrolle der Anzahl der gewählten Zahlen

Mathematica-Programm:

```
Length[u[1]]
```

Output:

```
50
```

Mit $u(j)$, $j = 1$ bis 80 , wird in eine Tabelle mit dem Namen „ t “ 80 mal der Mittelwert von 50 Zufallszahlen aus M geschrieben: Das sind jetzt Mittelwerte von 80 Bootstrap-Kopien!

Mathematica-Programm:

```
| tWork2 = Table[Mean[u[j]], {j, 1, 80}] |
```

Output:

```
0.459226, -0.810515, -1.19051, -1.11871, -0.864426, -0.647294, -0.787226,
0.293086, -0.798827, 0.725485, -0.451151, -0.464175, -1.27947, -0.196644,
-0.206662, 0.263628, 0.295626, -0.27978, 0.89629, -0.462274, 1.05565,
0.325698, -0.858681, 0.488327, -0.386761, -0.338291, -0.353974, -0.120557,
0.326703, 0.209404, -0.931436, -1.49898, -0.594813, -0.324888, 0.804326,
-1.10232, -0.507406, -0.093157, -0.73157, -0.320648, -0.871774,
-0.807057, -0.207304, -0.924659, -0.504472, -0.290555, 0.309056,
-0.464312, 0.362335, 0.202175, -0.943627, -1.65005, -0.513785, 0.302006,
-0.240725, 0.0436444, -0.597013, -0.315092, -1.06526, 0.209233,
-0.321134, -0.105876, 0.670831, -0.358651, 0.833696, 0.216187, -0.965687,
-0.300275, -0.804799, 0.0738544, 0.110673, 0.0166167, -1.18572, -0.763453,
-1.99561, -0.811618, -0.11997, -0.618394, -0.349117, -0.0513129
```

Kontrolle:

Mathematica-Programm:

```
| tWork2 |
```

Output:

```
0.459226, -0.810515, -1.19051, -1.11871, -0.864426, -0.647294, -0.787226,
0.293086, -0.798827, 0.725485, -0.451151, -0.464175, -1.27947, -0.196644,
-0.206662, 0.263628, 0.295626, -0.27978, 0.89629, -0.462274, 1.05565,
0.325698, -0.858681, 0.488327, -0.386761, -0.338291, -0.353974, -0.120557,
0.326703, 0.209404, -0.931436, -1.49898, -0.594813, -0.324888, 0.804326,
-1.10232, -0.507406, -0.093157, -0.73157, -0.320648, -0.871774,
-0.807057, -0.207304, -0.924659, -0.504472, -0.290555, 0.309056,
-0.464312, 0.362335, 0.202175, -0.943627, -1.65005, -0.513785, 0.302006,
-0.240725, 0.0436444, -0.597013, -0.315092, -1.06526, 0.209233,
-0.321134, -0.105876, 0.670831, -0.358651, 0.833696, 0.216187, -0.965687,
-0.300275, -0.804799, 0.0738544, 0.110673, 0.0166167, -1.18572, -0.763453,
-1.99561, -0.811618, -0.11997, -0.618394, -0.349117, -0.0513129
```

Jetzt wird 20'000 mal der Mittelwert von 50 Zufallszahlen aus M in eine Tabelle geschrieben und damit ein Histogramm der Mittelwerte erstellt

Mathematica-Programm:

```
| t = Table[Mean[u[j]], {j, 1, 20000}]; Histogram[t]; |
```

Das Histogramm ist hier nicht spezielle wiedergegeben, da auf Seite 62 nochmals ein entsprechendes Histogramm abgebildet ist. (Siehe <http://rowicus.ch/Wir/MathematicaPackages/Bootstrap.pdf>)

Wir wiederholen das Programm und erstellen nochmals ein Histogramm (andere Bootstrapkopien, anderes Histogramm, ähnliche Form, aber verschoben...)

Mathematica-Programm:

```
| t = Table[Mean[u[j]], {j, 1, 20000}]; Histogram[t]; |
```

Das Histogramm ist hier nicht spezielle wiedergegeben, da auf Seite 62 nochmals ein entsprechendes Histogramm abgebildet ist. (Siehe <http://rowicus.ch/Wir/MathematicaPackages/Bootstrap.pdf>)

Die Form des Histogramms bleibt bei 20'000 mal „bootstrapen“ beim zweiten Mal etwa ähnlich. Daher kann man einer daraus gewonnenen Verteilungsfunktion vertrauen.

Erstellung einer Verteilungsfunktion auf der Grundlage von Bootstrap-Kopien (hier mit 1000 Kopien)

Mathematica-Programm:

```
<< Statistics`DataManipulation`;
ttt = Sort[Table[Mean[u[j]], {j, 1, 1000}]];
freq = Union[Frequencies[ttt]];
F[x_] := Apply[Plus, Table[Take[freq, Length[Select[Table[freq[[k1, 2]],
{k1, 1, Length[freq]}], # == x &]]][[k1]][[1]],
{k1, 1, Length[Take[freq, Length[Select[Table[freq[[k, 2]],
{k, 1, Length[freq]}], # == x &]]]]} ]]/Length[ttt];
Plot[{F[x], 1}, {x, -3, 3}];
```

Der Graph ist hier nicht spezielle wiedergegeben, da auf Seite 62 nochmals ein entsprechender Graph abgebildet ist. (Siehe <http://rowicus.ch/Wir/MathematicaPackages/Bootstrap.pdf>)

Output:

```
freq = Union[Frequencies[t]];
(* Ergibt 20'000 Zahlenpaare! Diese werden des Umfangs wegen nicht
ausgegeben. Daher der ";" an Ende des Befehls. *)
```

F.4 Beispiel 2: Einlesen und bearbeiten von Messwerten, neue Messreihe

Einlesen resp. Erzeugung der Messwerte und Setzung der Parameter. Hier wird erst eine neue Messreihe erzeugt.

Mathematica-Programm:

```
nN = 12;
nHistogram = 20000;
nBootstrap = 1000;
M1 = Table[Random[]/10 + 20, {n, 1, nN}];
```

Bearbeitung mit Plug–In–Methode (Bootstrapping)

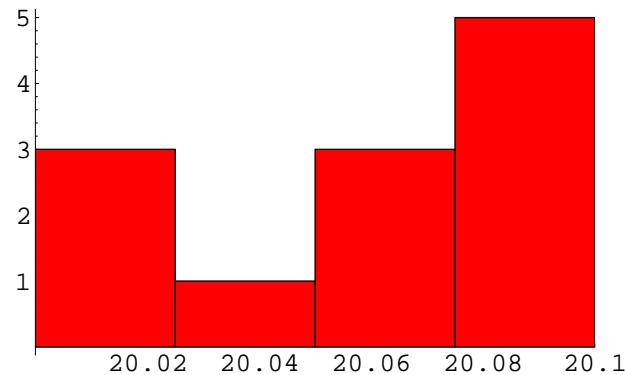
Mathematica-Programm:

```
<< Graphics`Graphics`;
<< Statistics`DataManipulation`;
M1 = Table[Random[]/10 + 20, {n, 1, nN}];
Print["Messwerte: ", M1];
Print["Mittelwert: ", M1 // Mean];
Histogram[M1];
u1[j_] := Table[M1[[Random[Integer, {1, nN}]]], {k, 1, 50}];
Print["Kontrolle Anzahl: ", Length[u1[1]] == nN];
Print[nBootstrap, " mal bootstrapen"];
t1 = Table[Mean[u1[j]], {j, 1, nBootstrap}]; Histogram[t1];
Print["Nochmals ", nBootstrap, " mal bootstrapen"];
t1 = Table[Mean[u1[j]], {j, 1, nBootstrap}]; Histogram[t1];
Print[nHistogram, " mal bootstrapen"];
t1 = Table[Mean[u1[j]], {j, 1, nHistogram}]; Histogram[t1];
Print["Nochmals ", nHistogram, " mal bootstrapen"];
t1 = Table[Mean[u1[j]], {j, 1, nHistogram}]; Histogram[t1];
ttt1 = Sort[Table[Mean[u1[j]], {j, 1, nBootstrap}]];
Print[nBootstrap, " Bootstrap-Kopien der Mittelwerte= ", ttt1];
MeanMin = Min[ttt1];
MeanMax = Max[ttt1];
Print["Mittelwert Min. bis Max. = ", "[", MeanMin, ", ", MeanMax, "]"];
freq = Union[Frequencies[
  ttt1]]; F[x_] := Apply[Plus, Table[Take[freq,
  Length[Select[Table[freq[[k1, 2]], {k1, 1, Length[freq]}],
  # = x &]]][[k1]][[1]], {k1, 1, Length[Take[freq, Length[
  Select[Table[freq[[k, 2]], {k, 1,
  Length[freq]}], # = x &]]]]}]]/Length[ttt1] ;
Plot[{F[x]}, {x, MeanMin, MeanMax}];
```

Output:

```
Messwerte: 20.0606, 20.0902, 20.0149, 20.0314, 20.0913,
20.0513, 20.0205, 20.0806, 20.0553, 20.0191, 20.093, 20.0804

Mittelwert: 20.0574
```

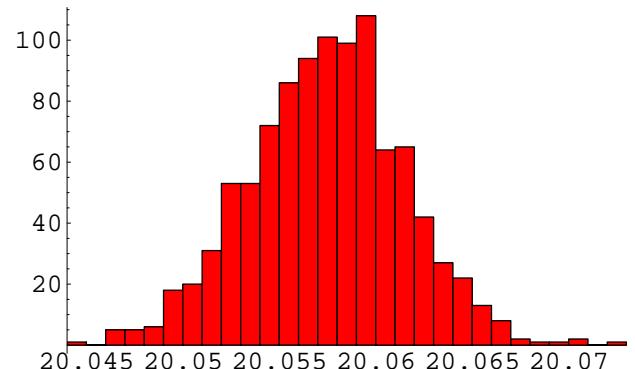
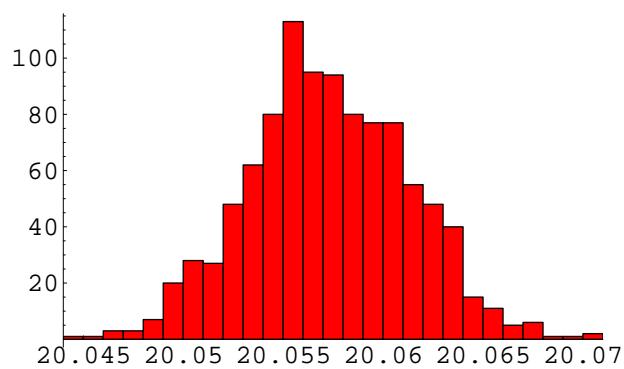


Output:

```
Kontrolle Anzahl: 50 == nM

1000 mal bootstrapen

Nochmals 1000 mal bootstrapen
```

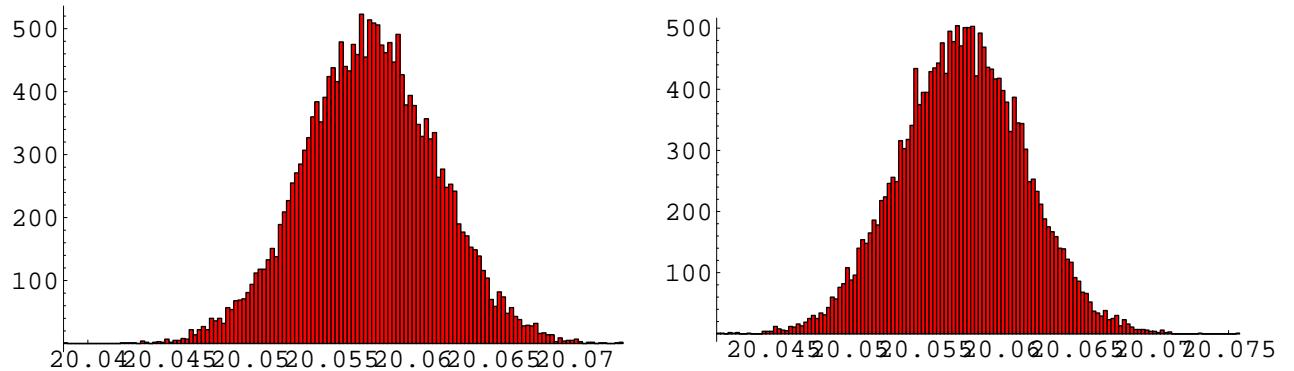


Output:

```
Kontrolle Anzahl: 50 == nM

20000 mal bootstrapen

Nochmals 20000 mal bootstrapen
```

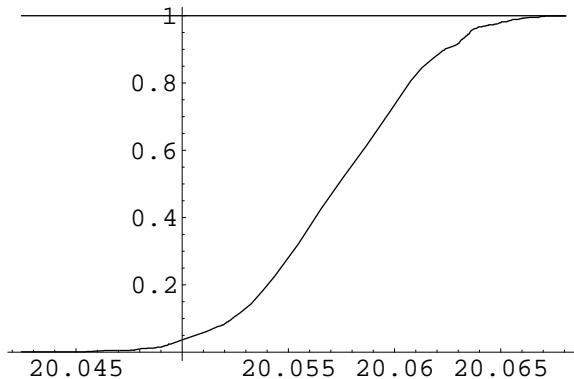


Die beiden Histogramme sehen ähnlich aus. Berücksichtigt man die Histogramme von 1000 Bootstrap–Kopien von vorhin, so kann man vermutlich davon ausgehen, dass bei Erhöhung der Bootstrap–Kopien–Anzahl die grobe Histogrammform nicht mehr stark ändert. Von der Datengewinnung wissen wir zudem, dass diese Daten hier keine indexabhängige Korrelation aufweisen: Ein Datenexemplar hängt nicht vom etwaigen vorhergehenden ab. Man kann daher mit gutem Gewissen dieses Datenmaterial dazu verwenden, eine empirische Verteilungsfunktion durch Simulation zu erzeugen.

Output:

```
Bootstrap-Kopien der Mittelwerte =
.... (des Umfangs wegen nicht wiedergegeben)
Mittelwert Min. bis Max. = [20.0424, 20.0681]
```

(Siehe <http://rowicus.ch/Wir/MathematicaPackages/Bootstrap.pdf>)



Vertrauensintervall ablesen

Die gewonnene Verteilungsfunktion ist eine ganz feine Treppenfunktion mit fast unsichtbaren Treppenstufen, also vermeintlich stetig. Man kann sie gebrauchen. Aus dem Graphen lesen wir auf der x –Achse das 60 %–Vertrauensintervall bei den y –Marken 0.2 und 0.8 ab:

$$I \approx [20.0540, 20.0618].$$

Anhang G

Anhang: Datensatzänderung

G.1 Die Problemstellung

Situation:

Aus einer Publikation (Archiv) kennt man statistische Kenngrößen eines Datensatzes, der aus $n = a$ Messungen eines Merkmals bestanden hat (a steht für „alte“). Die einzelnen Messungen stehen nicht mehr zur Verfügung. Mit diesem Datensatz ist jetzt ein Problem entstanden: Man hat erfahren, dass von den $m = a$ Messungen deren $m_1 = f$ garantiert falsch eingetragen waren, wobei man die verwechselten Werte kennt (f steht für „falsche“). An einer Messvorrichtung ist ein systematischer Fehler entdeckt worden. Weiter hat man herausgefunden, dass $m_2 = v$ Werte in der Statistik nicht erfasst worden sind, denn die Daten von einer Messvorrichtung sind damals verloren gegangen, jetzt aber wieder aufgetaucht (v steht für „verlorene“).

Problem:

Welche statistischen Kenngrößen kann man jetzt im Nachhinein korrigieren und wie stellt man dies an?

G.1.1 Beurteilung einer Korrekturmöglichkeit

Minimum, Maximum, Spanne

Wir stellen uns eine Strichliste oder ein Histogramm der damals vorhandenen Daten vor. Anhand einer beliebig angenommenen Konstellation sehen wir, dass Minimum, Maximum und Spanne von einzelnen Daten abhängen. Anhand der Korrekturdaten können wir feststellen, ob die genannten Kenngrößen richtig oder falsch sind.

Konsequenz:

Minimum, Maximum und Spanne können in der gegebenen Situation korrigiert werden.

Median und Modus

Da wir die einzelnen Daten nicht mehr kennen und damit auch über ihre Rangliste und ihre Frequenztafel nichts wissen, kann über den Median und den Modus nichts ausgesagt werden. Dazu sind jetzt die Grundlagen nicht mehr vorhanden.

Mittelwert, Varianz und Standardabweichung

Für diese Kenngrößen ist eine Korrektur sehr wohl möglich. Nachfolgend werden wir das Vorgehen dazu mittels einer vertieften Diskussion der Sache erhellern.

Methode:

Der Einfachheit halber gehen wir nach den Korrekturen (1) $m_1 = f$ und (2) $m_2 = v$ getrennt und unabhängig vor. In der Praxis kann man dann z.B. zuerst die Korrektur (1) vornehmen und die so erhaltene Situation dann als Ausgangsbasis (mit $n = a$) für die Korrektur (2) verwenden.

G.2 Die Minimaleigenschaft des Mittelwerts

G.2.1 Varianz, Standardabweichung und Wahrscheinlichkeitsfunktion

Bemerkung:

Nachfolgend benützen wir die Abkürzung **StD** für die **Standardabweichung** (StandardDeviation), $\text{StD} := \hat{\sigma}$

Aus der Literatur ersieht man, dass allgemein als **erwartungstreuer Schätzer der Standardabweichung** σ

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

die folgende Grösse (empirische Standardabweichung) Verwendung findet ($V(X) = \text{Varianz}$):

$$\hat{\sigma} = \text{StD} = s_X^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}}, \quad (\text{var} = s_X^*)^2 = \hat{\sigma}^2$$

Der vermutlich von „Uneingeweihten“ vorerst als StD erwartete Maximum–Likelihood–Schätzer (für eine Normalverteilung)

$$s_{ML}^* = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}}$$

weicht für grosse n normalerweise nicht sehr viel von s_X^* ab. Dabei ist jeweils $(s^*)^2 = \text{var}$ die empirische Varianz.

Wir führen die nachfolgenden Betrachtungen mit s_X^* aus. Mit s_{ML}^* wäre das Vorgehen und die Resultate analog.

Wir betrachten nun die zur empirischen StD gehörige Wahrscheinlichkeitsfunktion mit der Variablen \bar{X} :

$$f(\bar{X}) = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{\frac{1}{2}} \Rightarrow \text{StD} = s_X^* = f(\bar{x}).$$

Dabei ist hier $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ der arithmetische Mittelwert des Datensatzes $DS_n := \{x_1, x_2, \dots, x_n\}$.

G.2.2 Minimaleigenschaft des Mittelwerts

Satz: $\bar{X} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ minimiert $f(\bar{X})$ und $(f(\bar{X}))^2$, $n > 1$.

Beweis:

$$\begin{aligned} \bar{X} = \bar{x} \Rightarrow \sum_{i=1}^n (x_i - \bar{X}) &= (\sum_{i=1}^n x_i) - n \cdot \bar{X} = n \cdot \bar{x} - n \cdot \bar{X} = n \cdot \bar{x} - n \cdot \bar{x} = 0 \\ \Rightarrow f'(\bar{X}) &= \frac{1}{2} \cdot \left(\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{-\frac{1}{2}} \cdot \frac{1}{n-1} \cdot \sum_{i=1}^n 2 \cdot (x_i - \bar{X}) = \\ &\underbrace{\left(\frac{1}{\left(\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{\frac{1}{2}} \cdot (n-1)} \right)}_{\neq 0} \cdot \underbrace{\sum_{i=1}^n (x_i - \bar{X})}_{=0, \bar{X}=\bar{x}} = 0 \\ \rightsquigarrow \Rightarrow f'(\bar{X}) &= 0 \text{ für und nur für } \bar{X} = \bar{x}. \end{aligned}$$

Da $f(\bar{X})$ in X unter der Wurzel quadratisch ist mit positiver Parabelöffnung, ist in \bar{x} und nur in \bar{x} ein Minimum von $f(\bar{X})$ vorhanden. q.e.d.

Daraus folgt, dass $f(\bar{X})$ für ein anderes $x^* \neq \bar{x}$ grösser wird als $f(\bar{x}) = StD$.

Korollar: $x^* \neq \bar{x} \Rightarrow f(x^*) > f(\bar{x}) = StD$

G.3 Verkleinerung der StD bei Erweiterung eines Datensatzes

G.3.1 Verkleinerung

Wir wollen hier zeigen, dass man einen gegebenen Datensatz $\{x_1, x_2 \dots x_n\}$ durch hinzufügen neuer Daten x_1', x_2', \dots, x_m' so erweitern kann, dass der neue Datensatz $\{x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_m'\}$ eine kleinere StD besitzt. Die x_k' wählen wir geeignet.

$$\text{Seien } f_n(\bar{X}) = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - \bar{X})^2 \right)^{\frac{1}{2}}, \quad f_{n+m}(\bar{X}) = \frac{1}{n+m-1} \left(\underbrace{\left(\sum_{i=1}^n (x_i - \bar{X})^2 \right) + \left(\sum_{k=1}^m (x_k' - \bar{X})^2 \right)}_{:= \sum_{i=1}^{n+m} (x_i - \bar{X})^2} \right)^{\frac{1}{2}}$$

$$\text{Wähle } x_k' := \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \rightsquigarrow \sum_{k=1}^m (x_k' - \bar{x})^2 = \sum_{k=1}^m (\bar{x} - \bar{x})^2 = 0$$

$$\rightsquigarrow f_n(\bar{X}) \geq f_n(\bar{x}) = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} > \frac{1}{n+m-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + 0 \right)^{\frac{1}{2}} =$$

$$\begin{aligned}
&= \frac{1}{n+m-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{k=1}^m (x_k' - \bar{x})^2 \right)^{\frac{1}{2}} = \frac{1}{n+m-1} \left(\sum_{i=1}^{n+m} (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \\
\rightsquigarrow StD_n = f_n(\bar{x}) &= \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} > \frac{1}{n+m-1} \left(\sum_{i=1}^{n+m} (x_i - \bar{x})^2 \right)^{\frac{1}{2}} = f_{n+m}(\bar{x}) = StD_{n+m}
\end{aligned}$$

Die Standardabweichung StD_{n+m} des erweiterten Datensatzes ist im betrachteten Falle also kleiner als die Standardabweichung StD_n des ursprünglichen Datensatzes. Die gegebene Erweiterung mit $x_k' = \bar{x}$ ist in jedem Falle möglich.

Satz:

Zu einem Datensatz existiert immer eine Erweiterung mit kleinerer Standardabweichung.

G.3.2 Konsequenzen für den Mittelwert

$$\begin{aligned}
\text{Seien } \bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i, \quad x_k' = \bar{x}_n \quad &\Rightarrow \bar{x}_{n+m} = \frac{1}{n+m} \left(\left(\sum_{i=1}^n x_i \right) + \left(\sum_{k=1}^m x_k' \right) \right) = \\
&= \frac{1}{n+m} \left((n \cdot \bar{x}_n) + (m \cdot \underbrace{\bar{x}_m'}_{=\bar{x}_n}) \right) = \frac{1}{n+m} ((n \cdot \bar{x}_n) + (m \cdot \bar{x}_n)) = \frac{1}{n+m} (n+m) \cdot \bar{x}_n = \bar{x}_n \\
\text{Sei } \bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{x}_m' := \frac{1}{m} \sum_{k=1}^m x_k' = \bar{x}_n \quad &\Rightarrow \bar{x}_{n+m} = \frac{1}{n+m} \left(\left(\sum_{i=1}^n x_i \right) + \left(\sum_{k=1}^m x_k' \right) \right) = \\
&= \frac{1}{n+m} \left((n \cdot \bar{x}_n) + (m \cdot \underbrace{\bar{x}_m'}_{=\bar{x}_n}) \right) = \frac{1}{n+m} ((n \cdot \bar{x}_n) + (m \cdot \bar{x}_n)) = \frac{1}{n+m} (n+m) \cdot \bar{x}_n = \bar{x}_n
\end{aligned}$$

Korollar:

Vor.:

In einem Datensatz mit dem Mittelwert \bar{x}_n werden neue Daten der Form x_1', x_2', \dots, x_m' hinzugefügt.

Dabei gelte: $\bar{x}_m' := \frac{1}{m} \sum_{k=1}^m x_k' = \bar{x}_n$. (Spezialfall: $x_k' = \bar{x}_n$)

Beh.:

Dann bleibt der gesamte Mittelwert unverändert:

$$\bar{x}_n = \bar{x}_{n+m}$$

G.4 Vergrösserung der StD bei Erweiterung eines Datensatzes

Man kann einen Datensatz auch so erweitern, dass gilt: $StD_n < StD_{n+m}$. Wir konstruieren dazu ein Beispiel. Dabei wählen wir nochmals den oben behandelten Fall mit $\bar{x}_n = \bar{x}_{n+m} := \bar{x}$.

$$\begin{aligned}
\text{Abkürzungen: Seien } QDS_n &:= \sum_{i=1}^n (x_i - \bar{x})^2, \quad QDS'_m := \sum_{i=1}^m (x_i' - \bar{x})^2, \\
QDS_{n+m} &:= \sum_{i=1}^{n+m} (x_i - \bar{x})^2 := QDS_n + QDS'_m \quad (QDS := „Quadratdifferenzsumme“).
\end{aligned}$$

Wir verlangen:

$$StD_n^2 = \frac{1}{n-1} \underbrace{\sum_{i=1}^n (x_i - \bar{x})^2}_{=QDS_n} \leq StD_{n+m}^2 = \frac{1}{n+m-1} \left(\underbrace{\sum_{i=1}^n (x_i - \bar{x})^2}_{=QDS_n} + \underbrace{\sum_{k=1}^m (x_k' - \bar{x})^2}_{=QDS_m'} \right) = StD_{n+m}^2$$

Daraus müsste folgen:

$$(n+m-1) QDS_n = (n-1) QDS_n + m QDS_n \leq (n-1)(QDS_n + QDS_m') = (n-1) QDS_n + (n-1) QDS_m'$$

$$\text{Resultat: } \Rightarrow \underbrace{m \cdot QDS_n \leq (n-1) \cdot QDS_m'}_{\text{gegeben}} \quad \Rightarrow \quad \frac{m}{n-1} \cdot QDS_n \leq QDS_m' = \underbrace{\sum_{i=1}^m (x_k' - \bar{x})^2}_{x_k' \text{ wählbar}}$$

Wenn wir die x_k' hier so wählen, dass $\frac{1}{m} \sum_{k=1}^m x_k' = \bar{x}_n$ gilt, wo können wir obige Ungleichung erfüllen.

Es ist keine wesentliche Einschränkung, wenn wir z.B. $x_k' = x_{2j}' = \bar{x}_n + z_j$, $x_k' = x_{2j+1}' = \bar{x}_n - z_j$ wählen und ein allfälliges letztes $x_k' = x_m' = x_{2j+1}' = \bar{x}_n$, also hier das entsprechende $z_j' = z_m' = 0$ setzen.

Dann gilt: $\sum_{i=1}^m (x_k' - \bar{x})^2 = 2 \sum_{i=1}^{[\frac{m}{2}]} z_j^2$. (Hier ist $[\frac{m}{2}]$ der Gauss-Klammer-Ausdruck, siehe Analysis I.)

Da wir nun die z_j beliebig gross wählen können, kann QDS_m' und damit auch QDS_{n+m} einen beliebigen positiven Wert annehmen. Daher kann man bei fixen n und m sowie QDS_n , also bei fixem StD_n^2 wegen

$$StD_n^2 \leq \frac{1}{n+m-1} QDS_{n+m} = StD_{n+m}^2,$$

durch die Wahl der z_j den Ausdruck StD_{n+m}^2 so vergrössern, wie man will.

Satz: Man kann einen gegebenen Datensatz immer auch so erweitern, dass die Standardabweichung grösser wird.

Weiter oben haben wir schon festgestellt: Wählt man die $z_j = 0$, so wird StD_{n+m}^2 kleiner als StD_n^2 . Da nun StD_{n+m}^2 stetig von den z_j abhängt, kann man letztere auch so wählen, dass $StD_n^2 = StD_{n+m}^2$ gilt. Kleiner als $\frac{n-1}{n+m-1} StD_n^2$ kann man die Standardabweichung jedoch durch eine Erweiterung des Datensatzes nicht drücken, da bei einer Erweiterung zu QDS_n keine negativen Werte werden entstehen können (Quadratsumme). Daraus folgt:

Korollar: Bei einer Erweiterung eines Datensatzes kann StD_{n+m}^2 alle Werte grösser gleich $\frac{n-1}{n+m-1} StD_n^2$ annehmen.

Bemerkung: Analoge Betrachtungen kann man für eine Verkleinerung eines Datensatzes anstellen. Dies sei dem Leser überlassen.

G.5 Umrechnung des Mittelwerts und der empirischen StD

G.5.1 Eine Formel für den Mittelwert

A. Bei Veränderung von Datenwerten

Problem: Gegeben sei ein Datensatz $DS_n := \{x_1, x_2, \dots, x_n\}$. Die Werte x_1, x_2, \dots, x_j sollen nun in x_1', x_2', \dots, x_j' geändert werden. Was passiert dann mit dem Mittelwert \bar{x}_n ?

Man berechnet den neuen Mittelwert \bar{x}_n' , indem man im alten Mittelwert die zu ersetzenen Daten subtrahiert und anschliessend die neuen, korrigierten Daten addiert:

$$\text{Formel: } \bar{x}_n' = \frac{1}{n} \left(n \cdot \bar{x} - \sum_{i=1}^j x_i + \sum_{i=1}^j x_i' \right)$$

B. Bei Hinzunahme von weiteren Datenwerten

Problem: Gegeben sei ein Datensatz $DS_n := \{x_1, x_2, \dots, x_n\}$. Neu hinzukommen sollen nun die Werte x_1', x_2', \dots, x_m' . Das ergibt den neuen Datensatz $DS_{n+m} := \{x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_m'\}$. Was passiert mit dem Mittelwert \bar{x}_n ? D. h. wie berechnet sich der neue Mittelwert \bar{x}'_{n+m} ?

Man berechnet den neuen Mittelwert \bar{x}'_{n+m} , indem man im alten Mittelwert die neuen, korrigierten Daten addiert und die Dimension anpasst:

$$\text{Formel: } \bar{x}'_{n+m} = \frac{1}{n+m} \left(n \cdot \bar{x} + \sum_{i=1}^m x_i' \right)$$

G.5.2 Eine Formel für die empirische Varianz

Es gilt: $var = \hat{\sigma}^2 = StD_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$. Dabei ist

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) = \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 = (\sum_{i=1}^n x_i^2) - 2 \cdot \bar{x} \cdot n \cdot \bar{x} + n \cdot \bar{x}^2$$

$$\rightsquigarrow \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \quad \Rightarrow \quad var = \hat{\sigma}^2 = StD_n^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 \right) - \frac{n}{n-1} \bar{x}^2$$

$$\text{Satz: } var = \hat{\sigma}^2 = StD_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 \right) - \frac{n}{n-1} \bar{x}^2$$

$\sum_{i=1}^n x_i^2$ ist hier die quadrierte Länge des Datenvektors $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $\rightsquigarrow \sum_{i=1}^n x_i^2 = |\vec{x}|^2$, wie man aus dem Skalarprodukt des Vektors mit sich selbst sieht.

$$\text{Korollar: } var = \hat{\sigma}^2 = StD_n^2 = \frac{1}{n-1} |\vec{x}|^2 - \frac{n}{n-1} \bar{x}^2$$

Bemerkung:

Wird ein Datensatz erweitert, so wird der Vektor \vec{x} länger. Ebenfalls wird seine Dimension n grösser. Damit wird $\frac{1}{n-1}$ kleiner und $|\vec{x}|^2$ grösser. $\frac{n}{n-1} \bar{x}^2$ muss dabei nicht sehr ändern. Ob das zu einer Verkleinerung oder zu einer Vergrösserung von $var = \hat{\sigma}^2 = StD_n^2$ führt, hängt sehr von den einzelnen Werten x_k' ab. Eine allgemeine und markante Aussage in eine gewisse Richtung kann hier damit nicht abgelesen werden.

G.5.3 Anwendung: Berechnung einer veränderten StD

Aus der letzten Formel lässt sich eine Möglichkeit der Berechnung einer veränderten StD eines veränderten Datensatzes gewinnen:

$$\text{Sei } QS_n := \sum_{i=1}^n x_i^2 \text{ (Quadratsumme). } \rightsquigarrow QS_n := \sum_{i=1}^n x_i^2 = (n-1) StD_n^2 + n \bar{x}^2.$$

$$\text{Satz: } QS_n := \sum_{i=1}^n x_i^2 = (n-1) StD_n^2 + n \bar{x}^2$$

Anwendungsmethoden:

$$\text{Wir stützen uns auf die Formel } StD_n = \left(\frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 \right) - \frac{n}{n-1} \bar{x}_n^2 \right)^{\frac{1}{2}} = \left(\frac{1}{n-1} QS_n - \frac{n}{n-1} \bar{x}_n^2 \right)^{\frac{1}{2}}.$$

Diese Formel gewinnen wir durch Wurzelziehen im Satz auf Seite 68. In dieser Formel müssen wir jetzt die QS_n sowie den Mittelwert \bar{x}_n anpassen.

A. Bei Veränderung von Datenwerten

Ändern wir im Datensatz $DS_n := \{x_1, x_2, \dots, x_n\}$ die Werte x_1, x_2, \dots, x_j in x_1', x_2', \dots, x_j' , so begegnen uns folgende Veränderungen:

1. Änderung der QS_n : Die Quadratsumme ändert sich in $QS_n' = QS_n - \sum_{i=1}^j x_i^2 + \sum_{i=1}^j (x_i')^2$.

2. Änderung von \bar{x}_n : Der Mittelwert ändert sich in $\bar{x}_n' = \frac{1}{n} \left(n \cdot \bar{x} - \sum_{i=1}^j x_i + \sum_{i=1}^j x_i' \right)$

3. Damit wird die veränderte Standardabweichung wie folgt:

$$\text{Formel: } StD_n' = \left(\frac{1}{n-1} QS_n' - \frac{n}{n-1} (\bar{x}_n')^2 \right)^{\frac{1}{2}}$$

B. Bei Hinzunahme von weiteren Datenwerten

Nehmen wir im Datensatz $DS_n := \{x_1, x_2, \dots, x_n\}$ die Werte x_1', x_2', \dots, x_m' neu hinzu, so begegnen uns folgende Veränderungen:

1. Änderung der QS_n : Die Quadratsumme ändert sich in $QS_{n+m} = QS_n + \sum_{k=1}^m (x_k')^2$.

2. Änderung von \bar{x}_n : Der Mittelwert ändert sich in $\bar{x}_{n+m} = \frac{1}{n+m} \left(n \cdot \bar{x} + \sum_{k=1}^m x_k' \right)$

3. Damit wird die veränderte Standardabweichung wie folgt:

Formel: $StD_{n+m} = \left(\frac{1}{n+m-1} QS_{n+m} - \frac{n+m}{n+m-1} (\bar{x}_{n+m})^2 \right)^{\frac{1}{2}}$

G.6 Ein Beispiel

Gegeben ist von einem Datensatz mit $n = 500$ Werten der gerundete Mittelwert $\bar{x}_{alt} = 149$ und die gerundete Standardabweichung $StD = s_{alt} = 22$. Weiter ist bekannt, dass die Messungen in Klassen eingeteilt worden waren. Es handelt sich also hier approximativ um Mittelwert und Standardabweichung von mittleren Klassenwerten.

Nun ist intern bekannt geworden, dass jemand vom Marketing aus konkurrenztechnischen Gründen $j = 90$ Werte gefälscht hat. Es wurden die Messwerte $x_{j,falsch} = 95$ an Stelle von $x_{j,wahr} = 120$ publiziert. Das Labor findet, dass Fälschungen aus diversen Gründen nicht tolerierbar sind und dass sein wissenschaftlicher Ruf auf dem Spiel steht. Daher werden jetzt der richtige Mittelwert \bar{x}_{wahr} und die richtige Standardabweichung $StD = s_{wahr}$ berechnet.

Wie geht der mit der Sache beauftragte Angestellte bei der Berechnung vor?

1. Korrektur des Mittelwerts:

$$\bar{x}_{alt} = 149 = \frac{1}{500} \sum_{k=1}^{500} x_{k,alt} = \frac{1}{500} \left(\left(\sum_{k=1}^{500-90} x_{k,wahr} \right) + (90 \cdot 95) \right) = \frac{1}{500} ((Sum_{wahr} + (90 \cdot 95)) \Rightarrow Sum_{wahr} = 65950$$

$$\bar{x}_{wahr} = \frac{1}{500} \sum_{k=1}^{500} x_{k,new} = \frac{1}{500} \left(\left(\sum_{k=1}^{500-90} x_{k,wahr} \right) + (90 \cdot 120) \right) = \frac{1}{500} ((Sum_{wahr} + (90 \cdot 120)) \Rightarrow \bar{x}_{wahr} = 153.5$$

2. Korrektur der Standardabweichung:

$$\begin{aligned} s_{alt}^2 &\approx \frac{n}{n-1} \sigma_a^2 \Rightarrow \frac{n-1}{n} s_{alt}^2 = \frac{n-1}{n} \sum_{k=1}^n (x_{k,alt} - \bar{x}_{alt})^2 \approx \sigma_a^2 = E((X - \mu_a)^2) = E(X^2) - (E(X))^2 \\ &= \frac{1}{n} \left(\sum_{k=1}^n (x_{k,alt})^2 \right) - \mu_{alt}^2 \approx \frac{1}{500} \left(\sum_{k=1}^n (x_{k,alt})^2 \right) - 149^2 = \frac{1}{500} \underbrace{\left(\left(\sum_{k=1}^{500-90} (x_{k,alt})^2 \right) + 90 \cdot 95^2 \right)}_{:=QS} - 149^2 \\ &\Rightarrow \frac{500-1}{500} 22^2 = \frac{1}{500} (QS + 90 \cdot 95^2) - 149^2 \Rightarrow QS \approx 10529766 \\ &\Rightarrow \sigma_w^2 = \frac{1}{n} \left(\sum_{k=1}^n (x_{k,wahr})^2 \right) - \mu_{wahr}^2 \approx \frac{1}{500} \left(\sum_{k=1}^n (x_{k,wahr})^2 \right) - 153.5^2 = \frac{1}{500} (QS + 90 \cdot 120^2) - 153.5^2 \\ &= \frac{1}{500} (10529766 + 90 \cdot 120^2) - 153.5^2 = 89.282 \\ &\Rightarrow s_{wahr}^2 \approx \frac{n}{n-1} \sigma_w^2 \approx \frac{500}{500-1} 89.282^2 = 89.4609 \Rightarrow s_{wahr} \approx 9.45838 \approx 9.5. \end{aligned}$$

Das Labor stellt fest, dass bei der notwendigen Korrektur zur Offenlegung der wahren Situation der Mittelwert von 149 nur wenig auf 153.5 ansteigt. Die Standardabweichung jedoch sinkt von 22

auf etwa 9.5. Man wundert sich daher heftig. Gilt nicht eine engere Streuung im Bewusstsein der Leser immer als vorteilhafter gegenüber einer weiten Streuung? Hatte der grössere Mittelwert das Marketing zu einer Angstreaktion veranlasst, sodass man unbewusst zur Lüge noch einen Marktanteil gesetzt hat? Im Labor sieht man lange Gesichter — und die Meinung über das Marketing ist dort gemacht. — Was denkt der Leser zu dieser Sache?

G.7 Übung

Übung 1

Gegeben ist von einem Datensatz mit $n = 4984$ Messungen der gerundete Mittelwert $\bar{x}_n = 652$ und die gerundete Standardabweichung $StD = 184$. Weiter ist bekannt, dass die Messungen in Klassen eingeteilt worden waren. Es handelt sich also hier approximativ um Mittelwert und Standardabweichung von mittleren Klassenwerten.

Nun ist noch bekannt geworden, dass eine Messvorrichtung, mit der $j = 196$ Werte gemessen worden sind, die Klassenwerte 650 statt richtig 670 geliefert hat. Ebenfalls sind bei dieser Messvorrichtung 212 Werte der Klassengrösse 750 als unsinnig abgetan und unterdrückt worden. n ist also zu klein eingerechnet. Berechne den korrigierten Mittelwert und die korrigierte Standardabweichung in 2 Stufen.

Bemerkung zur Lösung: Diese soll mit einem Rechner ausgeführt werden. Eine Musterlösung wird bei Gelegenheit unter dem folgenden Link bereitgestellt:

Link: http://rowicus.ch/Wir/ProblemsSolutBachelor/UEMAlg16_zus.pdf

Übung 2

Überlege dir, ob über Parameter wie die Schiefe u.s.w. ebenfalls entsprechende Aussagen möglich sind.

Anhang H

Anhang: Spezielle Wahrscheinlichkeitssituationen

H.1 Kreuztabellen, Beispiel

Viele Probleme lassen sich übersichtlich begreifen, indem man **Kreuztabellen** oder **Kontingenztafeln** verwendet. Damit gewinnt man eine Darstellungsart, welche die Übersichtlichkeit erhöht. Wir wollen die Sache an einem Beispiel studieren:

Problem: Ein Edelmetall verarbeitendes Unternehmen besitzt 90 gewöhnliche und 10 gepanzerte Lieferwagen, letztere für den Transport von Ware mit sehr hohem Wert. Die beiden Wagentypen kann man von aussen kaum voneinander unterscheiden. Im Mittel machen alle Wagen etwa gleich lange Tagesrouten. Die Firma beschäftigt gleichviele Fahrer und Fahrerinnen, zusammen soviele wie die Lieferwagen, wobei die Fahrer aus Sicherheitsüberlegungen vier mal soviel für Sicherheitstransporte in gepanzerten Fahrzeugen eingesetzt werden wie die Fahrerinnen. Die Beschäftigten wissen auf ihren Transporten nicht, was der Inhalt der transportierten verschlossenen Metallkisten ist. Nun studiert man in der Firma mitten im Frieden einen möglichen ersten Überfall. Infolge der Struktur der Belegschaft darf angenommen werden, dass niemand von der eigenen Firma am Überfall beteiligt sein kann. Alle Wagen sind im Einsatz.

1. Erstelle eine Kreuztabelle (Kontingenztafel) für die Situation an diesem Tag, an dem alle Wagen und Fahrer resp. Fahrerinnen im Einsatz sind.
2. Was ist die Chance, dass es im Falle eines ersten Überfalles an einem solchen Tag einen Sicherheitstransport trifft?
3. Was ist die Chance, dass im Falle eines ersten Überfalles ein Sicherheitstransport mit einer Frau am Steuer überfallen wird?
4. Was ist die Chance, dass im Falle eines ersten Überfalles ein Sicherheitstransport überfallen wird unter der Voraussetzung (resp. der Bedingung), dass eine Frau am Steuer sitzt? (Dies somit unter der Annahme, dass die Übeltäter gezielt Transporte mit Frauen auswählen.)

Lösung:

1. **Kontingenztafel oder Kreuztabelle:**

Wir benutzen nun eine rechteckige Tabelle mit den Zeileneingängen „Männer“ und „Frauen“ und den Spalteneingängen „gepanzert“ und „ungepanzert“. Dazu kommen die Totale.
Zuerst tragen wir dann jeweils die 50, die 100, die 90 und die 10 in die Tafel ein. Anschliessend

ermitteln wir die mittlere Anzahl Männer und Frauen in den gepanzerten Wagen. Da die Männer dort vier mal mehr eingesetzt werden, kommen auf eine Frau vier Männer. Total muss das 10 Personen ergeben. Somit erhalten wir 2 Frauen und $4 \cdot 2 = 8$ Männer, denn $2 + 8 = 10$. Die Zahlen sind hier so, dass es immer ganze Werte gibt, dass man sich also nicht über Bruchanteile von Personen wundern muss. Anschliessend können wir den Rest der Tabelle durch Ermittlung der fehlenden Differenzen ausfüllen.

Die damit erhaltene Tabelle heisst **Kontingenztafel oder Kreuztabelle**.

Fahrer	Wagen ungepanzert	Wagen gepanzert	Total Zeile
Männer: 50	$50 - 8 = 42$	$4 \cdot 2 = 8$	50
Frauen: 50	$50 - 2 = 48$	$1 \cdot 2 = 2$	50
Total Spalte: 100	90	10	100

Wenn wir eine Spalte summieren, so schreiben wir die Summe in die letzte Zelle der Spalte unten. Wenn wir eine Zeile summieren, so schreiben wir die Summe in die letzte Zelle der Zeile rechts. Wir haben in unserem Falle 2 mal 2 nicht summierte Zellen.

Aus dieser Tafel kann man jetzt Wahrscheinlichkeiten nach der Wahrscheinlichkeitsdefinition im Sinne von Laplace ablesen: $P = \frac{\text{Anzahl günstige Fälle}}{\text{Anzahl mögliche Fälle}}$.

2. Chance, dass im Falle dieses ersten Überfalles an einem solchen Tag ein Sicherheitstransport überfallen wird: Es hat 10 Sicherheitstransporte (günstige Fälle), aus denen einer gewählt wird. Dazu hat es total 100 Transporte (mögliche Fälle).

$$P = \frac{10}{100} = 0.1$$

3. Chance, dass im Falle eines ersten Überfalles ein Sicherheitstransport mit einer Frau am Steuer überfallen wird: Es hat 2 Sicherheitstransporte (günstige Fälle), aus denen einer gewählt wird. Total hat es dazu 100 Transporte (mögliche Fälle).

$$P = \frac{2}{100} = 0.02$$

4. Was ist die Chance, dass im Falle eines ersten Überfalles ein Sicherheitstransport überfallen wird unter der Voraussetzung (resp. der Bedingung), dass eine Frau am Steuer sitzt? Es hat 2 Sicherheitstransporte (günstige Fälle), aus denen einer gewählt wird und total 50 Transporte, bei denen eine Frau am Steuer sitzt (mögliche Fälle).

$$P = \frac{2}{50} = 0.04$$

H.2 Ein Beispiel mit einem Taxi, das Fragen aufwirft

Nachstehend ist ein Problem wiedergegeben, das von einem Kollegen in einem nicht genannten Buch entdeckt worden ist. Die vermutlich vom Autor angegebene Lösung hat dazu veranlasst, *über das Wesen von Wahrscheinlichkeit nachzudenken*. In diesem Problem wurde nach einer Wahrscheinlichkeit gefragt. Zuvor muss klar sein, in welchen Fällen man von Wahrscheinlichkeit reden kann. Dadurch sind Fragen aufgeworfen worden, die unten kurz angedeutet werden sollen. Hier folgt der Text der Aufgabe aus dem Buch, der angeblich von zwei Wirtschaftswissenschaftlern formuliert worden sein soll:

Problem:

„Ein Taxi streift in einer Winternacht ein anderes Auto. Es gibt in der Stadt zwei Taxigesellschaften: eine mit blauen Wagen, die andere mit grünen. Die mit grünen Wagen beherrscht 85 % des Marktes. Eine Zeugin sagt aus, das Unfalltaxi sei blau gewesen. Unabhängige Tests ergeben, dass sie in 80 % der Fälle eine richtige Aussage macht. Welche Farbe hatte das Taxi wirklich? Fast jeder wird glauben, dass es blau war, weil er sich auf die hohe Glaubwürdigkeit der Zeugin stützt. Die entscheidende Frage ist aber, wie weit ihre Glaubwürdigkeit die Tatsache beeinflusst, dass ein zufällig ausgewähltes Taxi mit 85-prozentiger Wahrscheinlichkeit grün ist. Kombiniert man die beiden Wahrscheinlichkeiten, ist die Chance, dass das fragliche Taxi *grün war*, 59 %. Die Wahrscheinlichkeit für grün ist also grösser als die für blau. Zu diesem Schluss können wir nie intuitiv kommen, dazu müssen wir *rechnen*.“

Eine andere Person hat dann sofort unter Verwendung einer Kreuztabelle eine Lösung präsentiert nach dem Muster des letzten Abschnitts. Denn ohne Kreuztabelle ist es schwieriger, sich der Sache zu nähern. Verglichen mit dem Beispiel im letzten Abschnitt handelt es sich um den dort erwähnten Fall: „Was ist die Chance, dass im Falle eines ersten Überfalles ein Sicherheitstransport überfallen wird unter der Voraussetzung (resp. der Bedingung), dass eine Frau am Steuer sitzt? (Dies unter der Annahme, dass die Übeltäter gezielt Transporte mit Frauen auswählen.)“

Nachfolgend ist die Kreuztabelle wiedergegeben. Die Konstruktion lässt sich einfach nachvollziehen. Dabei wird formal wie folgt gerechnet:

Zuerst nehmen wir an, wir hätten statt 100 % Marktanteil jetzt 100 Taxis. Damit können wir einfacher argumentieren. Dann gäbe es 85 grüne und 15 blaue Taxis. Dagegen ist nichts einzuwenden.

Weiter argumentieren wir: Die Zeugin, so heisst es, sagt mit einer Wahrscheinlichkeit $P = 0.8$ jeweils das Richtige. Wenn sie daher sagt „das Taxi sei blau gewesen“, ergibt das im Falle der 15 blauen Taxis $15 \cdot P = 15 \cdot 0.8 = 12$. Das tragen wir ein. Da es hier nur 15 blaue Taxis gibt, muss die Zeugin daher in $15 - 12 = 3$ Fällen „grün“ sagen. Das tragen wir auch ein. Dass sie etwa gar keine Aussage macht, das schliessen wir einfach aus — ohne Grund natürlich.

Ebenso würden wir im Falle, dass die Zeugin behauptete, das Taxi sei grün gewesen, dies mit der Wahrscheinlichkeit $P = 0.8$ akzeptieren. Da es hier 85 grüne Taxis gibt, ergäbe das dann $85 \cdot P = 85 \cdot 0.8 = 68$. Auch das wird jetzt eingetragen. Die Differenz, d.h. der Fall dass die Zeugin blau sagt, das Taxi aber grün ist, ergibt sich zu $85 - 68 = 17 = 85 \cdot (1 - 0.8) = 85 \cdot 0.2 = 17$. Damit können wir die Tabelle fertig stellen und rechts noch die Zeilensummen 71 und 29 eintragen.

	Taxi ist grün	Taxi ist blau	Total
Zeugin sagt grün	68	3	71
Zeugin sagt blau	17	12	29
Total	85	15	100

Jetzt kann man die Wahrscheinlichkeit, dass „das Taxi grün ist im Falle dass die Zeugin sagt, es sei blau gewesen“, ablesen. Man erhält:

$$P = \frac{\text{Anzahl günstige Fälle}}{\text{Anzahl günstige Fälle}} = \frac{17}{29} = 0.59 \hat{=} 59\%$$

Man könnte aber auch wie folgt argumentieren: Die Wahrscheinlichkeit, dass die Zeugin „blau“ sagt, ist $0.15 \cdot 0.8 + 0.85 \cdot 0.2 = 0.12 + 0.17 = 0.29$. Die Wahrscheinlichkeit, dass die Zeugin bei einem grünen Taxi „blau“ sagt, ist $0.85 \cdot 0.2 = 0.17$. Hier ist gemeint, dass das Taxi, ausgewählt aus allen Taxis, ein grünes ist und das die Zeugin aus ihren möglichen Aussagen diejenige betreffend blau abgibt, womit $P(\text{blau} = 0.8)$ zu verwenden ist, da eine Aussage mit dieser Chance als richtig genommen wird. D.h. man lässt dazu noch die Wahrscheinlichkeiten 0.8 für Treffer sowie diejenige für Irrtümer mit $1 - 0.8 = 0.2$ gelten, dies bezüglich allen Taxis (was bestritten werden könnte, wie weiter unten gezeigt wird). Durch die Konjunktion „und“ hat man die Produktwahrscheinlichkeit für die Schnittmenge „{Aussage der Zeugin falsch} \cap {Taxi ist wirklich grün}“ gegeben, also $0.2 \cdot 0.85 = 0.17$. Als Wahrscheinlichkeit verwendet man hier jeweils die relative Häufigkeit.

Die Wahrscheinlichkeit, dass das Taxi grün ist und die Zeugin jedenfalls „blau“ sagt, ist

$$P(\text{Taxi grün eingeschränkt auf Fälle mit „Zeugin sagt blau“}) = \frac{P(\text{Taxi grün})}{P(\text{Zeugin sagt „blau“})} = \frac{0.17}{0.29} = 0.59.$$

Hier hat man somit die Wahrscheinlichkeit bei Einschränkung der Grundmenge auf die Teilmenge mit Zeugenaussagen „blau“ in einer Gesamtmenge, welche weiter nicht mehr interessiert: $P(\text{Taxi grün eingeschränkt auf Fälle mit „Zeugin sagt blau“})$. Man begegnet hier auch dem Problem, dass jene, welche nicht täglich mit dem „Fachchinesisch der Datenanalyse“ und seinen definierten Bezügen konfrontiert sind, den Text nicht selbstverständlich so verstehen können, wie es die Autoren vielleicht beabsichtigen. Das Problem dieser oft vorliegenden effektiven, manchmal von den Betroffenen gar nicht wahrgenommenen Unverständlichkeit, hat man in allen Fachrichtungen. Es ist eines der wahren Übel unserer Zeit. Denn es gibt im Hause der Fachgebiete inzwischen mehr solche als nur jene, in die ein Mensch in Anbetracht seiner beschränkten Zeit heute Einblick gewinnen kann.

Gegen die formale Herleitung dieses Resultats 0.59 ist nichts einzuwenden. Jedoch gibt die intentionale Ebene zu diskutieren: Es stellt sich das Problem der *Realitätstreue*, wenn man die konkrete Situation dieses Falles überdenkt.

Im Text steht: „Ein Taxi streift in einer *Winternacht* ein anderes Auto.“ Es muss daher dunkel gewesen sein. In der Dunkelheit kann ein erwachsener Mensch aber nur schwer grün von blau unterscheiden. Jeder kennt doch den Ausspruch: „In der Nacht sind alle Katzen grau.“ Also auch orange–rote, braungelbe oder gar weisse Katzen erscheinen grau, sofern sie nicht etwa durch genetische Veränderungen reflektierende Pigmente eingebaut haben...

Weiter liest man da: „Unabhängige Tests ergeben, dass sie in 80 % der Fälle eine richtige Aussage macht.“ Hat man daher mit der Zeugin wohl Reihenuntersuchungen gemacht? Dafür hätte man viel Zeit gebraucht und sie daher reich entlohnen müssen, was angesichts des Schadens bei einer Streifkollision wohl kaum denkbar ist. Oder hat man etwa Nachbarn über ihren Leumund befragt und daraus $p = 0.8$ errechnet? Dann wäre dieses 0.8 wohl soviel wert wie das Resultat einer Volksabstimmung unter Laien über die Existenz einer als Arbeitshypothese angenommenen unbekannten Lösung einer mathematischen Gleichung. In diesem Fall darf man schon die Unsinnsvermutung aufstellen!

Konkret ist hier zu fragen: Wie kann man der Richtigkeit einer Zeugenaussage eine Wahrscheinlichkeit beimessen? Man müsste dabei drei Fälle unterscheiden:

1. Die Wahrscheinlichkeit bezieht sich auf die Farb–Wahrnehmungsfähigkeit der Zeugin bei Dunkelheit.
2. Die Wahrscheinlichkeit bezieht sich auf die Aufrichtigkeit und die Wahrheitstreue der Zeugin bei der Abgabe einer Aussage im konkreten Fall.
3. Man hat eine Situation, in die beide vorgenannten Fälle gemischt eingehen.

Der Text sagt nichts zu den drei Fällen. Einfach den ersten obigen Fall zu postulieren wäre praktizierte Naivität, die das errechnete Resultat sehr fraglich erscheinen lässt. Man hätte die Zeit besser investieren können, als derart kopflos zu rechnen. Und ob die Verlässlichkeit des Farbsehens der Zeugin bei Dunkelheit anlässlich nachträglicher Tests ermittelt werden kann, ist ebenso fraglich. Es könnte ja sein, dass die

Zeugin abends damals etwas getrunken hätte, dies nur beiläufig, ohne es richtig zu werten, sich so nicht eines nachgelassenen Sehvermögens bewusst war, dass sie später aber dann anlässlich der Tests nüchtern war. Und so weiter.

Noch schlimmer ist es bei der Aufrichtigkeit und die Wahrheitstreue. Würde die Zeugin zugeben, dass ihr Schwager im Eventualfall der Fahrer des Taxis war, er in diesem Falle also ein grünes Taxi fuhr? Wer kennt nicht die hochgradig zuverlässigen Manager und Bankdirektoren, die sich während Jahrzehnten nichts zuschulden kommen liessen, dann aber plötzlich als Verbrecher da standen: „Und die Bank war Pleite.“ So liesst man am 23. 1. 2008 in Zeitungen die Schlagzeilen: „Ex-Chef der Deutschen Post Klaus Zumwinkel gesteht Steuerhinterziehung in Millionenhöhe“. Wohlverstanden nicht freiwillig, sondern nachdem sich der deutsche Geheimdienst Daten und Informationen über Konten bei lichtensteinischen Finanzunternehmen auf dem Schwarzmarkt zusammengekauft hatte... Zudem kann ein Mensch über Jahre konstant ein Verhalten zeigen, sich dann aber plötzlich aus freien Stücken zu einem anderen, mit dem vorhergehenden Verhalten nicht mehr zu vereinbaren neuen Verhalten mutieren: „Vertrauen ist gut, Kontrolle ist besser“.

Wenn man statistische Häufigkeiten als Wahrscheinlichkeiten akzeptieren will, so muss man die Streubreite und die Konstanz in der Entwicklung lanfristig untersuchen und gegebenenfalls korrigierend anpassen. Eine Zahl alleine wie 0.8 hat daher keinen akzeptablen, verlässlichen Wert. Damit kann man bei der obigen Geschichte die Zeugin nur als Einzelphänomen in die Überlegung einbeziehen. Man darf ihr also keinenfalls die Rolle der Lieferantin eines zentralen Zahlenwertes zubilligen, der ein rechnerisches Resultat wesentlich beeinflusst.

Weiter weiss man nicht (wie oben erwähnt), ob aus $P(\text{Zeugin sagt Wahrheit}) = 0.8$ folgt, dass $P(\text{Zeugin sagt Unwahrheit}) = 0.2$ ist. Denn die Zeugin könnte ja auch schweigen, d.h. die Aussage verweigern. Dieser Fall bleibt in der obigen Rechnung völlig ignoriert.

Wenn man aber die Zeugin nicht beachtet, so ist die Wahrscheinlichkeit, dass ein blaues Taxi daher zu fahren kommt, gleich 0.15. Auch wenn man in einem Test mit 100 Taxis, 85 grüne und 15 blaue, herausfähnde, dass die Zeugin in 29 Fällen „blau“ sagt und in den restlichen Fällen dann „grün“, falls sie überhaupt etwas sagt, lässt noch keinen Schluss auf den Unfallabend zu. Was bei der Sache schlimm erscheint, ist der selbstaufgeriegte Zwang, hier zu meinen etwas rechnen zu müssen. Vermutlich hat nicht zuletzt diese Haltung dazu beigetragen, dass in den letzten Monaten des Jahres 2008 in der Weltwirtschaft eine Dollarsumme vernichtet worden ist von mindestens der Höhe eines 4-stelligen Milliardenbetrages! Das war eine gigantische Katastrophe, wie sie ähnlich seit fast 100 Jahren nicht mehr vorkam.

Ein Grund dazu liegt in der ahnungslosen, unreflektierten, primitiv gutgläubigen, aus Gier und Naivität blinden Einschätzung des Wahrscheinlichkeitsbegriffs. Den Hintergrund finden wir in psychologischen Faktoren. Darauf handelt das nächste Kapitel. Es geht dort um die Frage „Wie alt ist der Kapitän?“ Diese Frage steht unter dem Scheinwerfer der Vernunft gesehen auch im Zusammenhang mit dem Wahrscheinlichen, dem wahr Scheinenden und dem wahr Seienden.

Zusammenfassend sind daher folgende Punkte bezüglich ihrer Realitätskonformität wesentlich und daher im Falle realitätsbezogener Wahrscheinlichkeitsaussagen relevant:

1. Im gestellten Problem wird voraussetzungslos gefragt, welche Farbe das Taxi wirklich hatte: „Welche Farbe hatte das Taxi wirklich?“ Bedinge Wahrscheinlichkeiten sind an Voraussetzungen gebunden. Eine Zeugenaussage ist dabei ein Eingriff in das Geschehen, ein quasi zufälliger Eingriff in die Voraussetzungen also. Hätte man andere Zeugen, würden diese vielleicht anders urteilen. Vielleicht gibt es sogar welche. Nur haben sie sich nicht gemeldet. Oder man wollte sie nicht, weil sie vielleicht nicht die Aussage machten, die man wünschte. Das wird jetzt vielleicht verschwiegen. Ganz sicher aber hängt eine berechnete Wahrscheinlichkeit von der Anzahl der Zeugen und ihrer Beschaffenheit ab.
2. Die Wahrscheinlichkeit für grün wird mit 0.59 und die für blau mit $1 - 0.59 = 0.41$ angegeben. Es handelt sich hier um bedingte Wahrscheinlichkeiten. So wird z.B. im Falle von grün die Wahrscheinlichkeit nur für jene Konstellation berechnet, die vorhanden ist in der Situation, in welcher die Zeugin „blau“ sagt.

3. Man stellt angeblich in Tests fest, dass die Zeugin mit einer Wahrscheinlichkeit 0.8 richtige Antworten gibt, also in $\frac{4}{5}$ der Fälle. Für diese Feststellung muss man zuerst minimal 5 Fälle gezählt haben. Man hat also wiederholt beobachten müssen. Die Umstände dieser Beobachtung sind unbekannt. Man weiss nicht, ob es sich um eine Wahrnehmungsuntersuchung oder um eine Anfälligkeit für situationsgebundene Lügen handelt. Letzteres kann man kaum experimentell behandeln, da die Vielfalt der möglichen Situationen das Erfassbare übersteigt. Daher ist die Bonität der Zahl 0.8 als sehr klein einzuschätzen.
4. Aus $P(\text{Antwort der Zeugin richtig}) = 0.8$ folgt nicht, dass $P(\text{Antwort der Zeugin falsch}) = 0.2$ gilt. Denn die Zeugin könnte die Antwort auch verweigern oder nicht finden. Weiter gibt zu denken:
5. Um Wahrscheinlichkeitsaussagen machen zu können, muss man eine gewisse *Reliabilität oder Zuverlässigkeit* haben: Bei Wiederholung des Experimentes unter gleichen Rahmenbedingungen sollte man das gleiche Messergebnis erzielen. Ebenso müsste man bürgen können für eine gewissen *Validität*, d.h. ein gewisses argumentatives Gewicht. Und nicht zuletzt sollte man eine gewisse Objektivität haben, d.h. das Ergebnis sollte unabhängig von der Beschreibung eines Sachverhalts durch den Beobachter sein.
6. Damit ein Text objektiv verstanden werden kann, muss er objektiviert gestaltet sein. Fehlende Angaben werden vom Leser nach seinen subjektiven kulturellen Gepflogenheiten eingefüllt und damit unterschoben. Kein Leser eines Textes kann sich beim Lesen einer Wertung entziehen, welche mit seiner situationsabhängigen Einschätzung von Verhältnismässigkeit auf seinen Erfahrungsschwerpunkten basiert. So kümmert sich z.B. ein Wirtschaftswissenschaftler kaum um die Herkunft und die Zuverlässigkeit von Wahrscheinlichkeitsaussagen, wie die Weltfinanzkrise vom Jahr 2008 beweist. Die Schreiber des behandelten Textes sind aber, so der Text, Wirtschaftswissenschaftler. Sie kennen die Geschehnisse aus der Optik über ihrem Schreibtisch, nicht aus der massgebenden Realität.
7. Das Verhalten eines Zeugen bei einer Zeugenaussage ist keine Massenerscheinung, der eine Wahrscheinlichkeit zukommt. Für solche Fälle sind keine gültigen oder verlässlichen Verteilungsfunktionen bekannt. Denn man wundert sich ja immer wieder, wie Personen in den höchsten staatlichen Positionen, denen am meisten Vertrauen zukommen sollte, plötzlich nochmals unverhofft öffentlich als Lügner entlarvt und gebrandmarkt werden. So lesen wir z.B. am 22. Jauar 2008 auf Swissinfo.ch:

„Schweizer Regierung kuschte vor den USA“

Der Bundesrat habe nicht souverän, jedoch feige und unter grober Missachtung der Rechtsstaatlichkeit gehandelt. — So kommentiert die Schweizer Presse den Bericht zur Aktenvernichtung im Fall der mutmasslichen Atomschmuggler Familie Tinner.

Der Bundesrat habe im Fall Tinner auf der ganzen Linie versagt, kommentiert die Basler Zeitung: „Er hat sich durch die Regierung der USA instrumentalisiert lassen. Er hat die Souveränität unseres Landes missachtet und durch mutwillige Sabotage eines Gerichtsverfahrens dem Ansehen des Rechtsstaates Schweiz geschadet.“

Nicht nur der damals federführende Justizminister Christoph Blocher mache „eine jämmerrliche Figur“, so die Basler Zeitung, der Bundesrat in Corpore habe sich „in dieser übeln Affäre fast wie das Regime einer Bananenrepublik“ benommen.

Die Süddeutsche Zeitung stellt fest, der Bericht der parlamentarischen Aufsicht zur Aktenvernichtung lasse die Schweiz „als eine Art Bananenrepublik am Gängelband der USA“ erscheinen.

Weiter aus der Presse: Vorher sei in öffentlichen Auftritten bezüglich dieses Themas im vollen Wissen um die Tatsachen immer alles abgestritten, beschönigt oder übergangen worden. Nicht nur das Volk (der Souverän also), auch das Parlament sei an der Nase herum und hinters Licht geführt worden...

Nun weiter zu den psychologischen Faktoren:

H.3 „Wie alt ist der Kapitän?“

Problem:

In einem dem Autor zugängig gemachten Text der Didaktik-Dozentin S. Müller von der Uni Münster stehen folgende Zeilen zu lesen:

Im Jahr 1989 erregte das Buch „Wie alt ist der Kapitän?“ von Stella Baruk großes Aufsehen. 76 von 97 befragten Zweit- und Drittklässlern (Anmerkung: Bei uns Primarschule) hatten die Aufgabe: „Auf einem Schiff befinden sich 26 Schafe und 10 Ziegen. Wie alt ist der Kapitän?“ „gelöst“, üblicherweise mit „36 Jahren“. Eine Ausweitung der (französischen) Untersuchung auf mehr Aufgaben und mehr Schüler auch anderer Jahrgangsstufen bestätigte das erschütternde Ergebnis.

Ebenso erwähnte Prof. Dr. G. Steiner, Ordinarius für Psychologie an der Uni Basel und vormals angeblich Mathematiklehrer, ähnliche Ergebnisse in einem vom Autor besuchten Didaktik–Weiterbildungskurs für Hochschul–Dozenten, wobei bei Prof. Steiner die Aufgabe etwas komplizierter gelautet hat. Frei nachempfunden etwa so: „Auf einem Schiff befinden sich der Kapitän, 12 Matrosen, 80 Pferde und 6 Hühner. Wie alt ist der Kapitän.“.

Es soll nun im Sinne der Mitteilung von Prof. Steiner versucht werden, in der Art der Schüler beim Experiment die Aufgabe zu lösen. Im Text genannt sind die Anzahlen 12, 80, 4 und dazu 1 für den Kapitän. Etwas Vorgehen eines durchschnittlichen Schülers:

1. Ein Schüler versucht zu rechnen: $12 + 80 + 4 + 1 = 97$. Diese Zahl ist zu gross. So alt kann er nicht sein, also muss man anders rechnen.
2. $12 \cdot 80 \cdot 4 \cdot 1 = 3'840$. Diese Zahl ist unmöglich. So alt kann er nicht sein, also muss man anders rechnen.
3. $80 - 12 - 4 - 1 = 63$. Das ist zwar alt, aber nicht unmöglich. Das ginge also. Aber vielleicht ist es falsch. Somit nochmals probieren!
4. $80 - 12 \cdot 4 - 1 = 31$ — das ginge ja sehr gut! Aber noch besser wäre folgendes:
5. $80 - 12 \cdot 4 + 1 = 33$ — das ist ausgezeichnet! Das ist ja so alt wie Christus. Dann muss es stimmen. Denn der Lehrer wählt immer Aufgaben, die ein einsichtiges Ergebnis haben. Also doppelt unterstreichen und abgeben. Dann ist man des Lobes gewiss.

Prof. Steiner hat auch mitgeteilt, dass etwa ein Schüler pro Klasse gemerkt hat, dass es sich bei der Aufgabe um Unsinn handelt, und dies dann als Resultat dem Lehrer mitgeteilt hat. Etwa einer hat also jeweils gewagt, den Sachverhalt der Aufgabe logisch zu erfassen. Die andern haben gleich, wie auf einen militärischen Befehl hin, zu rechnen begonnen: „Kopf einziehen und durch!“

Es fragt sich nun, wieso durchschnittlich intelligente Schüler, welche nach ca. 10 Jahren zu ehrbaren und vernunftsbegabten Bürgern herangewachsen sind, solchen Unsinn produzieren können. Prof. Steiners Erklärung war sinngemäß die folgende:

Der Schüler ist gewohnt, dass bei einer Mathematikaufgabe mit solchen Zahlen als Resultat eine Zahl gefunden werden muss, die einerseits in der Situation als brauchbar gelten kann und die andererseits dann doppelt unterstrichen werden muss. Das Resultat ist unbrauchbar, wenn es nicht doppelt unterstrichen ist oder wenn die Zahl fehlt. Dann nämlich ist die Aufgabe nicht gemacht. So etwas abzugeben wäre gefährlich. Die Erfahrung hat das ja gelehrt. Im schlimmsten Fall könnte man dafür Schläge einstecken. Weiter wird das Resultat als falsch bewertet, wenn die Zahl undenkbar ist, wie z.B. beim Alter von 3'840 Jahren für den Kapitän. Dafür gibt es nicht Schläge, sondern nur eine schlechte Note.

Wenn der Schüler jedoch den Lehrer kritisiert, ihm also mitteilt, diese von ihm gestellte Aufgabe sei Unsinn, so wäre das gewiss eine Beleidigung für den ehrbaren und unfehlbaren Lehrer. Der Schüler müsste mit Rausschmiss, Anzeige beim Rektor und bei den Eltern rechnen. Dann würden ihm zur Strafe vom Lehrer weiter zudem auch noch von den Eltern als Strafe die Ferien gestrichen, zusammen wohl mit einer Kürzung des Sackgelds. Das ist die allergefährlichste Variante!

Also wählt der Schüler die ungefährlichere Variante: Er macht lieber den Unsinn gleich mit.

Wir entdecken da, dass ein Strategieproblem vorliegt. Der Schüler strebt nicht nach der Wahrheit, hier also nicht nach einem vernünftigen Resultat zur gegebenen Frage. Sonder er strebt nach der Minimierung der Gefahr, dass ihm Unglück zustossen könnte. Damit entwickelt er nicht Lösungskonzepte für mathematische Aufgaben, sondern Strategiern zur Minimierung der Gefahren, die auf ihn lauern. Denn es geht ihm zur Hauptsache um das sorglose Überleben, damit also um den grössten eigenen Profit. Es geht ihm nicht um das Erkunden von Wahrheit in einem kleinen ihn betreffenden Teilbereich in Form einer Mathematikstunde unter vielen anders gearteten Stunden im Leben.

Das erklärt nun vieles. Es erklärt auch, wieso in einem Buch die von Wirtschaftswissenschaftlern erfundene und oben wiedergegebene Wahrscheinlichkeitsaufgabe gefunden werden kann...

Anhang I

Hinweise zur Datenanalyse

- *Ici, il y a pour le moment seulement le texte allemand à disposition. Momentanément, la traduction française manque encore.*

I.1 Grundfragen — Modellierungen

I.1.1 Wichtige Abläufe

Um was geht es beim Datensammeln oder bei der Datenanalyse? — Stichworte:

1. Statistische **Dokumentation** (deskriptive Statistik).
2. Versuchsplanung, Versuchsoptimierung, Kostensenkung („design of experiments“).
3. **Explorative Datenanalyse** (EDA, Konzeptbildung Planung von Datenerhebungen, „... was noch bezahlbar ist“).
4. **Affirmative Statistik** (mathematische Statistik, Aussagen mit belegbarer Genauigkeit).
5. Naturwissenschaftliche, humanwissenschaftliche oder wirtschaftswissenschaftliche **Modellbildung**: Man möchte Regeln oder gar Gesetze erkennen und diese statistisch abstützen resp. **validieren** oder **evaluieren**. Das heisst, man möchte die Behauptungen auf **Plausibilität** und auf **Gültigkeit** prüfen. Dasselbe gilt zuvor für die verwendeten Prüfverfahren.
Generell geht es hier um **Modellbildung** (z.B. um mathematische Modelle auf der Grundlage explorativer und später affirmativer Statistik).
6. **Optimierung**.
7. **Hypothesenprüfung**.
8. **Gesetz– oder Regelfindung**.
9. **Entscheidungsfindung**.
10. Erarbeitung **sicherheitstechnischer Grundlagen**.
11. **Prozessüberwachung**.
12. **Qualitätsüberwachung** und **Qualitätskontrolle**.

Dazu dienliche Instrumente, welche erst sein kurzem bzw. erst in der Neuzeit zur Verfügung stehen: Computer, Wahrscheinlichkeitsrechnung, Neuere statistische Methoden (z.B. verteilungsfreie Methoden, Simulationsmethoden, robuste Methoden)

Bemerkung: Als **robust** bezeichnet man z.B. solche Methoden, bei welchen Fehler in den Daten die Resultate praktisch nicht oder nur wenig beeinflussen. So können z.B. **Ausreisser** in den Rohdaten einen Mittelwert erheblich verändern. Der zugehörige Median hingegen wird durch Ausreisser praktisch nicht beeinflusst. Der **Median** ist daher **robust gegen Ausreisser**.

I.1.2 Vorgehen bei Modellierungen

Absicht: Man möchte z.B. im Labor durch Messungen eine naturwissenschaftliche Aussage oder gar ein Gesetz gewinnen. Oder man möchte Kennwerte einer nicht vollständig zugänglichen und damit nicht ganz erfassbaren Population erarbeiten. Wie geht man statistisch vor?

1. **Vorbereitung:** Erst definieren, welche statistische Kennzahlen gesucht werden sollen.
2. **Vor dem Experiment:** Für jede Angelegenheit ist ein **Wahrscheinlichkeitsmodell zu wählen**: Was kann man zu einer Verteilungsfunktion oder zu einer Häufigkeitsfunktion jetzt sagen? Kennt man die Verteilung und daher die beim Versuch involvierten Wahrscheinlichkeiten?
3. **Vor dem Experiment:** Lege die Art, die Grösse und die Erhebungsbedingungen (u.s.w.) der zu gewinnenden Stichprobe fest. D.h. definiere die **Stichprobe**.
4. **Vor dem Experiment:** Definiere, welche **Kennzahlen** wie geschätzt werden sollen. Lege also die **Schätzfunktionen** fest.
5. **Vor dem Experiment:** Stelle einen detaillierten **Versuchsplan** auf, welcher es ermöglicht, mit möglichst wenigen Beobachtungen ein maximal verlässliches Ergebnis zu erreichen. Denn Versuche können sehr teuer sein. Gegebenenfalls ist ein **Fragebogen** auszuarbeiten, welcher den Einsatzbedingungen maximal Rechnung tragen und auch eine **seriöse Datengewinnung** gewährleisten muss (Minimierung des schlechten schriftlichen Ausdrucks, elektronische Erhebung, ankreuzen u.s.w.). Wichtig sind auch **Kontrollfragen**, welche Informationen dafür liefern, ob das Ausfüllen seriös unternommen worden ist. — Oder ob jemand einfach zufällig angekreuzt hat.
6. **Während des Experiments:** Führe das Experiment wie geplant durch. Falls etwas nicht in Übereinstimmung mit der Planung funktioniert, so unterbreche oder breche ab und plane neu. Andernfalls: Erhebe damit die Daten so exakt wie möglich.
7. **Nach dem Experiment:** Rechnen! Das Heisst: **Berechne** für die gesuchten Kenngrössen die **Schätzwerte** beim gemachten Experiment.
8. **Nach dem Experiment:** Führe eine Fehlerrechnung durch. Bestimme damit die **Genauigkeit** der **Schätzung**.

Zum Beispiel scheint es plausibel, einen Erwartungswert durch einen Mittelwert zu schätzen. Oft ist man auch der **Plausibilität** aus **pragmatischen Gründen** ausgeliefert, weil man gar keine andere Wahl hat. Wenn ein Experiment eine Million Euro oder Franken kostet, wird man sich ein zweites sehr gut überlegen müssen.

Eine wichtige Frage ist auch die Frage nach der Parameterart oder auch der Beziehung zwischen den Messgrössen. So legen **Lageparameter** die Lage der Stichprobe (etwa in einem Diagramm) fest. **Streuparameter** geben ein Mass für die gegenseitigen Nähe der Werte. Damit kann man z.B. **zwei Stichproben als wesentlich verschieden einstufen**, wenn sie gleiche Streuung haben, die Zentralwerte aber so weit auseinander liegen, dass sich die durch die Streuung gegebenen Intervalle in keiner Weise überdecken.

I.1.3 Hypothesenmodellierung, Hypothesentest

Als Beispiel soll untersucht werden, ob eine neue Fabrikationsmethode F_2 gleiche Resultate liefert wie die bisherige, alte Methode F_1 . Man hat z.B. festgestellt, dass unter gleichen Fabrikationsbedingungen wie bisher die Festigkeit ν_k vieler der mittels der neuen Methode fabrizierten Stücke (Nr_k) unterhalb der bisherigen Normfestigkeit liegen. Bekannt ist der Erwartungswert mit der alten Methode μ_1 .

Vorgehen mittels Hypothesenformulierung und Hypothesentest:

1. Formuliere die **Nullhypothese** H_0 . Hier z.B.: „Die neue Fabrikationsmethode funktioniert korrekt. Die Festigkeit der neuen Stücke ist gleich gut oder besser wie bisher.“
2. Formuliere die **Alternativhypothese** H_1 . Hier z.B.: „Die neue Fabrikationsmethode funktioniert nicht korrekt. Die Festigkeit der neuen Stücke ist untragbar viel schlechter als bisher.“
3. Suche ein **Wahrscheinlichkeitsmodell**, das eine plausible Annahme für die Wahrscheinlichkeitsverteilung einer Zufallsvariablen ausdrückt, welche die Nullhypothese ausreichend charakterisiert. In unserem Beispiel: Wenn die neue Fabrikationsmethode korrekt funktioniert, so ist zu erwarten, dass etwa die Hälfte der damit fabrizierten Stücke eine grössere und die andere Hälfte eine kleinere Festigkeit haben als μ_1 . Gleiche Festigkeit $\nu_k = \mu_1$ wird wegen den Messtoleranzen ausgeschlossen. Man müsste sich also für grösser oder kleiner entscheiden. Die Wahrscheinlichkeit, dass ein neues Stück Nr_k eine Festigkeit $\nu_k < \mu_1$ hat, müsste demnach $P(\nu_k < \mu_1) = 0.5$ sein.
4. **Die Stichprobendefinition:** Aus Kostengründen entschliesst man sich, eine Messerie von 10 Stücken mit der neuen Methode herzustellen und zu testen. Die Sache ist deshalb kostspielig, weil das Testverfahren zu einer **Zerstörung der Stücke** führt (Zugversuch, Messung der Fliessgrenze und der Bruchlast).
5. Definition der **Testgrösse** und der **Positivität** (Erfüllung der Nullhypothese): Als Testgrösse wird die Anzahl N der Stücke genommen mit $\nu_k \geq \mu_1$. Der Test verläuft positiv, wenn bei 10 Versuchen $N \geq 5$ herauskommt.
6. **Durchführung des Versuchs:** Es werden 10 Messungen gemacht. Man stellt jedoch leider fest, dass $N = 1$ ist. 9 Stücke fallen also in die unerwünschte Klasse mit $\nu_k < \mu_1$. Das würde für die Alternativhypothese sprechen.
7. Jetzt legen wir noch ein **Signifikanzniveau** für die Nullhypothese fest. Doch was aber bedeutet der Ausdruck „**Signifikanz**“? — Den Unterschied zwischen Messgrößen oder Variablen (hier dem Erwartungswert μ_1 und den neuen gemessenen Festigkeiten ν_k) nennen wir **signifikant** oder **wesentlich**, wenn die Wahrscheinlichkeit, dass sie durch Zufall derart zustande kommen können, nur sehr gering ist. Handelt es sich in unserem Fall um wesentliche Unterschiede? (Können wir sagen, dass $N \leq 1$ wesentlich verschieden von $N = 5$ ist?)
Wir legen hier das **Signifikanzniveau** $\alpha = 0.05 \hat{=} 5\%$ fest. α bedeutet die maximal zulässigen **Irrtumswahrscheinlichkeit** die Nullhypothese abzulehnen, obwohl sie richtig ist (**Fehler 1. Art**). $\beta = 1 - \alpha$ wäre dann die Wahrscheinlichkeit, die Alternativhypothese abzulehnen, obwohl diese wahr ist. (**Fehler 2. Art**).
8. Wir überprüfen die Signifikanz durch eine Abschätzung der **Irrtumswahrscheinlichkeit**: Wir berechnen jetzt, was bei **Annahme der Gültigkeit der Nullhypothese** unter diesen Umständen die Wahrscheinlichkeit ist, dass $N \leq 1$ sein kann. Zur Berechnung kann man den folgenden Modellvergleich heranziehen: Gegeben seien 10 schwarze Kugeln, zeitlich *nacheinander* aufgereiht wie die produzierten Werkstücke. Was ist die Anzahl Möglichkeiten, maximal eine der schwarzen Kugel zufällig durch maximal eine weisse zu ersetzen? (Jeweils 9 oder 10 Kugeln bleiben dann schwarz.) Die möglichen Fälle erhalten wir, wenn wir alle die Möglichkeiten von Ersetzungen durchrechnen. Bei den Ersetzungen handelt es sich um Auswählen. Man hat es also bei dieser Betrachtungsart mit Kombinationen ohne Wiederholung zu tun.
Mit Hilfe der hier passenden Binomialverteilung ergibt sich:

$$\left(\binom{10}{0} + \binom{10}{1}\right) \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^1 = \left(\binom{10}{0} + \binom{10}{1}\right) \left(\frac{1}{2}\right)^{10} \approx 0.0107.$$

Ebenso ist die Anzahl der günstigen Fälle $g = \sum_{k=0}^{10} \binom{10}{k} = 1024$. Und diejenige der möglichen Fälle ist: $m = \left(\binom{10}{0} + \binom{10}{1}\right) = 11$. Damit wird $\frac{g}{m} = \frac{11}{1024} \approx 0.0107$.

- 9. Formulierung des Testresultats:** Die reale Situation ($N \leq 1$) tritt bei Annahme der Nullhypothese mit einer Wahrscheinlichkeit von 0.0107 (ca. 1%) auf. Es ist $0.0107 \leq \alpha = 0.05$. Die Wahrscheinlichkeit, dass der betrachtete Realfall bei einer richtigen Nullhypothese eintreffen kann, ist mit 0.0107 sehr gering und kleiner als α . Das deutet auf eine Richtigkeit der Alternativhypothese, denn die Nullhypothese muss damit verworfen werden. Die Wahrscheinlichkeit für einen Fehler 1. Art ist damit sehr gering: Der Test „schlägt aus!“

I.1.4 Das Problem der Wechselwirkungen zwischen Variablen

Beispiel In einer Untersuchung soll festgestellt werden, ob der Konsum K eines gewissen elektronischen Produkts H vom Alter A des Konsumenten abhängt, dass also vermutlich junge Leute infolge ihrer eher draufgängerischen Art mehr H konsumieren. Und andererseits ob der Konsum vom Geschlecht G abhängt. Die Resultate sollen eine gezielte Werbung ermöglichen. Diese lassen sich in $A \times K$ - und $G \times K$ -Diagrammen mittels der **Bandbreite der Streuung** darstellen und bearbeiten.

Es könnte auch sein, dass zwischen A und G bezüglich des maximalen oder des mittleren Konsums ein Zusammenhang besteht. Ob eine solcher Vermutung richtig sein könnte, lässt sich als erster Schritt ebenfalls mittels Diagrammen abschätzen.

I.2 ♣ Wirklichkeit, mathematische Modelle, Wesentlichkeit der Entscheide

♣ ♣ ♣ Die nachfolgenden Ausführungen sind dem Kurs „Einführung“ entnommen.

♣ ♣ ♣ Siehe <http://rowicus.ch/Wir/Scripts/KursEinf.pdf>.

I.2.1 ♣ Das Problem der sinnvollen Frage

Hier wollen wir fragen:
 Was sind denn die Beziehungen zwischen Mathematik einerseits und Technik sowie Naturwissenschaften andererseits?
 Wieso darf man mit Hilfe von Mathematik z.B. die Konstruktionsweise einer Brücke vorausberechnen, die Brücke dann bauen und schliesslich sicher sein, dass sie hält, wenn wir mit einem schweren Lastwagen darüberfahren?

Um die Problematik zu erhellen, wollen wir uns eine alte naturwissenschaftliche Frage stellen. Sie lautet:

Frage 1 Wieso fällt der Apfel vom Baum?

Wie soll man da antworten? — Natürlich ist klar, dass er nicht ewig oben bleibt. Ein Kind hat geantwortet: „Weil die Erde ihn liebt!“ Diese Antwort wirkt auf den zweiten Blick vielleicht gar nicht so dumm. Das Kind spricht ehrlich und unverbildet. Doch weiter hilft uns das ja nicht.

Anscheinend liefert folgende Aussage eine Erklärung: „Es existiert halt eine Anziehungskraft.“ So scheint

man festzustellen. Doch hilft uns das weiter? Entdeckt man denn im Leben den Apfel nicht viel früher als die Kraft? Ja, der Begriff *Kraft* ist doch sehr abstrakt und nicht jederman ganz klar. Seine Durchdringung erfordert einiges an Aufwand. Und ist man einmal so weit, so müssen wir doch eingestehen, dass durch das Wort *Kraft* die Frage noch nicht ganz beantwortet ist. Kraft steht hier als Name für ein dahinter verborgenes Naturphänomen, das wieder seine Ursachen hat. Wir werden dann einwenden, die Masse des Apfels und die der Erde sei die Ursache der Kraft. Doch was ist dann die Ursache der Masse? Bei jeder Ursache können wir ja wieder nach der vorausgehenden Ursache fragen, nach der Ursache der Ursache und so fort. Schliesslich landen wir so in einer unendlichen Kette von Ursachen, was uns in einem endlichen Leben nie zu einem Ende führt. So kommen wir also nicht weiter zu einem abschliessenden Verständnis. Verrückt, nicht? Endlich müssen wir einsehen: Aus dieser unangenehmen Lage können wir uns nur so befreien, dass wir halt eingestehen, dass es nicht möglich ist, zu einer abschliessenden Antwort zu kommen. Die eingangs gestellte Frage lässt sich irgendwann weiter nicht mehr durch Angabe weiterer Ursachen beantworten und bringt uns daher nicht viel weiter. Wir müssen uns fragen, ob die eingangs gestellte Frage wissenschaftlich weiter überhaupt sinnvoll ist, oder ob man besser anders frägt, um zu weiteren Resultaten zu kommen. Konsultieren wir nun die Physik und ihre Gesetze, so sehen wir, dass wohl eine andere Frage eher beantwortet werden kann und daher viel sinnvoller wäre. Nämlich:

Frage 2 Wie fällt der Apfel vom Baum?

Diese Frage ist es ja, die im Fallgesetz ihre Antwort findet. Also merken wir uns:

In exakten Wissenschaften ist es manchmal viel schwieriger, zu einem Phänomen die exakte richtige Frage zu finden als die exakte richtige Antwort.

♣ Dazu eine Anekdote:

I.2.2 ♣ Galilei und Archimedes

In der Physik gewinnt man neue Erkenntnisse, indem man einerseits die Natur mit Hilfe von reproduzierbaren Experimenten beobachtet: *durch den experimentellen Ansatz*. Mit „reproduzierbar“ meinen wir, dass das Experiment unter den gegebenen Bedingungen, d.h. im festgelegten Messrahmen, immer wieder zu gleichen Resultaten führt. Die Resultate analysiert man dann mit mathematischen Methoden und gewinnt so Aussagen in der Sprache der Mathematik: *die Theorie*. Doch andere gehen manchmal genau umgekehrt vor. Aus theoretischen Überlegungen heraus hat jemand eine Vermutung. Um zu entscheiden, ob diese richtig oder falsch ist, benutzt er jetzt die experimentelle Verifikation, die Überprüfung im Labor (*theoretischer Ansatz*). Und früher oder später irgendwann beobachtet er bei beiden Methoden die Natur mit Hilfe des Experiments.

Aristoteles (384 – 322 v.Chr., etwa Zeit der Nachklassik) hat nun beobachtet, dass beim Fallenlassen zweier Körper der schwerere schneller fällt als der leichtere. Klar, denn jedes Kind sieht doch, dass eine Vogelfeder langsamer fällt als ein Stein. Wir halten fest:

Satz I.1 (Aristoteles) *Leichte Körper fallen langsamer als schwere.*

Bis zu Galilei (1564 – 1642, italienisches Barockzeitalter) war das bei uns im Abendlande so etwas wie ein Glaubenssatz. Wieso war das so? Wer sich in der Geschichte auskennt, der weiss, dass an den

Hochschulen der Antike die Sprache der Gebildeten nicht Latein war, sondern Griechisch. Nur im Westteil des Römerreiches und beim Staat hat man Latein gesprochen. Daher gerieten die griechischen wissenschaftlichen Schriften der Antike nach dem Untergang des alten weströmischen Reiches, d.h. nach der Völkerwanderung, im Abendland ausser Reichweite und daher in Vergessenheit. Der Kontakt zum griechischen Ostrom brach ziemlich ab, nicht zuletzt auch wegen der religiösen Prioritätsstreitigkeiten. Im Westen kam das scholastisch geprägte Mittelalter herein mit seiner Geissel namens *Ideologie*. „Denken“ wurde durch „glauben“ ersetzt, „fragen“ durch Auslegung der offenen Heils Wahrheit. Wehe dem, der noch zu fragen wagte. Aus dem fragenden Suchen nach Wahrheit, nach Erkenntnis, war die Lehre von der richtigen Wahrheit geworden: Schulung — Scholastik. *Und die Lehre war jetzt die Doktrin der Doktoren.* Über dem Lichte der Wahrheit ward es dunkel.

Im Spätmittelalter jedoch brach plötzlich sehr mächtig die Realität herein über Europa. In Südspanien erlebten die dortigen Kalifen eine Zeit der Hochblüte. Die Araber hatten das altgriechische Wissen bewahrt und der unvermeidliche Kontakt mit ihnen in Spanien führte bei den Christen zu Veränderungen, denn eine höherstehende Kultur verdrängt und überlagert die minderwertigere beim Kontakt. Die Kreuzzüge taten das ihre. Man kam nicht um den geistigen Austausch mit den Arabern herum. Aber wie dieser andern Kultur begegnen? Hat man doch im Westen damals alles auf die Lehre, auf den christlichen Glauben bezogen, eine Grundlage, die die Araber nicht gelten liessen. Man musste sich daher auf die gemeinsamen Wurzeln besinnen. Und die fand man bei den altgriechischen Philosophen, speziell bei Aristoteles, der so viele Fundamente der Wissenschaften gelegt hat. So ist es nicht verwunderlich, dass die Werke des Aristoteles erst um etwa 1200 n.Chr. in Toledo (Spanien) aus dem Arabischen erstmals ins Latein übersetzt worden sind. Latein war die Sprache der Gelehrten des Abendlandes bis ans Ende des 19. Jahrhunderts. Und dem Aristoteles hat man dann im Westen streng nach scholastischer Manier geglaubt. Wehe dem, der es wagte zu zweifeln! Galilei, sensibilisiert durch den Kontakt mit Kopernikus, dessen Anhänger er geworden war, hat es nun gewagt zu zweifeln an so vielen Dingen. Erst im Jahre 1993 hat die römische Kirche das Urteil des Inquisitionsprozesses gegen ihn revidiert. Mit Galilei beginnt wissenschaftsgeschichtlich die Neuzeit, denn er hat es als erster geschafft einen Begriff zu bilden, der das Abstraktionsvermögen in der Antike übersteigen: die abstrakte Größe *Beschleunigung* als *Veränderung* der abstrakten Größe *Geschwindigkeit*. Auch verwundert es nicht, dass es erst einen Galilei geben konnte, nachdem sich in Italien die Renaissance breit gemacht hatte. Dies nach dem Kreuzzug gegen das griechische Byzanz, verstärkt durch dessen Fall an die Türken 1453, da dann viele griechische Gelehrten nach Italien ausgewandert sind und so der Kontakt mit der Antike stärker aufblühen konnte. *Was können wir nun von Galilei lernen?* Was hat er genau herausgefunden? Und wie ist er vorgegangen? Während Aristoteles wahrscheinlich experimentell vorgegangen ist, darf man bei Galilei den theoretischen Ansatz vermuten. Von ihm wird nämlich das unten folgende *Gedankenexperiment* überliefert. (Aus „Gedanken“ schliessen wir hier auf „theoretisch“.) Dazu das *Fallgesetz*:

Satz I.2 (Galilei, Fallgesetz 1) *Alle Körper hier auf Erden fallen gleich schnell.*

In vielen Museen wird heute demonstriert, wie Galilei dann dieses Gesetz experimentell verfeinert hat und eine mathematische Beziehung fand:

Satz I.3 (Galilei, Fallgesetz 2) *Es gilt das Gesetz:*

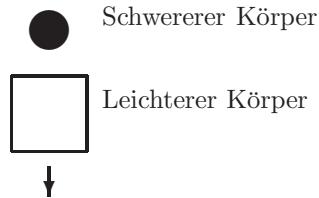
$$s \sim t^2.$$

Später haben andere dann mit noch feineren Messmethoden festgestellt, dass auch die Höhe über Meer in der Formel eine Rolle spielen muss. Und man hat dann das Gesetz entsprechend verfeinert (vgl. *Gravitationsgesetz der Physik* von Newton). Doch wie kam Galilei überhaupt dazu, an Aristoteles zu zweifeln?

Gedankenexperiment von Galilei: (Vgl. Abb. I.1)

Wir lassen erst den leichteren Körper fallen — und dann gerade anschliessend den schwereren in derselben Bahn. Wenn der schwerere Körper schneller fällt, so muss er bald den leichteren einholen. Nun geschieht's: Einerseits muss jetzt beim Zusammenprall der leichtere Körper den schwereren abbremsen,

Abbildung I.1: Das Gedankenexperiment von Galilei



denn Aristoteles muss ja recht haben. Zusammen sind die beiden Körper dann *langsamer* als der schwerere Körper. Andererseits sind sie zusammen doch schwerer als der schwerere Einzelkörper. So müssen sie zusammen *schneller fallen* als der schwerere, denn Aristoteles muss ja recht haben. Zusammen fallen sie also einerseits langsamer, andererseits schneller. Und das genau ist ein *Widerspruch!* — Wo also steckt nun der Fehler? Der Gedankengang war doch richtig! — Nur etwas haben wir nie überprüft: die Voraussetzung nämlich, dass Aristoteles recht haben muss. Das kann also nicht sein. Somit müssen beide Körper gleich schnell fallen.

I.2.3 ♣ Extrapolation und mathematisches Modell

Um nun zum zweiten Gesetz zu kommen, gab es nur eine Möglichkeit: Messungen, d.h. das Experiment. Wie oft hat Galilei wohl gemessen? — Wir wissen es nicht. Vielleicht 10 mal, vielleicht 100 mal, vielleicht 1000 mal. Doch eines ist ganz sicher: *Man kann nur endlich viele Male messen*. Doch wie ist es dann überhaupt möglich, ein Gesetz als richtig zu verkaufen, wenn man es im Vergleich zu den möglichen Fällen doch nur in einigen wenigen Fällen experimentell überprüft hat? Hier hat uns Descartes (Kartesius, 1596 – 1650) einen Rat gegeben, dem wir folgen wollen:

Axiom I.1 (Descartes) *Will man auf Grund experimenteller Ergebnisse, d.h. endlich vieler Messungen, ein allgemeingültiges Naturgesetz ableiten, so soll man immer die einfachst mögliche Form für das Gesetz wählen und diese gelten lassen, solange nicht feinere Messungen zu einer Verfeinerung des Gesetzes zwingen.*

So schliesst man von der besonderen Situation, etwa von n Messungen, auf die allgemeine Situation, die unendlich vielen Möglichkeiten. Man sagt, man schliesse *vom Besonderen auf das Allgemeine*. Man *extrapoliert* also. Aber wie können wir das rechtfertigen? Liegt hier nicht nur eine eigenartige, jedoch zufällige Verhaltensweise von uns Menschen vor? — Da müssen wir als unkundig geborene, aber geistig stets wachsende Denker eben zugeben, dass uns eigentlich gar keine andere vernünftige Möglichkeit übrigbleibt. Und mit „vernünftig“ meinen wir, dass wir diese Verhaltensweise aus der *Erfahrung* lernen! Ja, die Erfahrung (und nichts anderes) sagt uns, dass dieses Vorgehen gut so ist. Denn bisher hatten wir damit Erfolg! Nur so können wir „das Vernünftige“ begründen.

So messen wir wie schon Galilei vielleicht 100 mal und finden in jedem gemessenen Fall $s \sim t^2$. Wir vermuten dann, dass der Zusammenhang $s \sim t^2$ allgemein gilt in allen unendlich vielen möglichen Fällen. Doch unendlich oft können wir ja nicht messen. Dass es aber trotzdem nicht falsch ist, in allen diesen unendlich vielen Fällen das Gesetz vorläufig für richtig anzunehmen, begründen wir mit dem Axiom von Descartes. Wir beschreiben demnach die Realität durch ein *Gesetz in der Sprache der Mathematik*, durch eine mathematische Beziehung, etwa durch Gleichungen, Formeln oder dergleichen, *die nur gültig sind innerhalb der Messgenauigkeit und unter den Messbedingungen*, d.h. im Rahmen des Experiments. Insofern stimmen unsere Gesetze nicht, sie sind ungenau. Genauere Messungen zwingen uns zu einer genaueren, verfeinerten Formulierung des Gesetzes.

Hat also Aristoteles zu unrecht sein Gesetz behauptet? Nach Descartes nicht, denn damals wusste man

noch nichts von Gasen, von Luftwiderstand und solcherlei. Seine Aussage ist im damals für die Beobachtungen gegebenen Rahmen ja richtig. Und Galilei hat daher die Aussage eben nur verfeinert, in einen genaueren Rahmen gestellt. So ist man vorwärts gekommen und kommt weiter vorwärts. Wir halten also als gültig fest:

Methode I.1 (Naturwissenschaften) *Wir beschreiben die Wirklichkeit durch mathematische Gesetze, die innerhalb der Messgenauigkeit gültig sind. Wir erklären nicht letzte Ursachen der Natur, sondern beschreiben nur ihr Verhalten. Legitimiert wird dieses Vorgehen durch die positive Erfahrung.*

Wir beantworten demnach nicht die Frage „*Warum...?*“ sondern die Frage „*Wie...?*“ — Man muss eben die **sinnvolleren oder vernünftigeren Fragen stellen!**

Das hier beschriebene naturwissenschaftliche Vorgehen, der Schluss von endlich vielen gemessenen Einzelfällen auf ein in unendlich vielen Fällen gültiges Gesetz, nennt man *induktiv*. Diese Art Schlussweise nennen wir auch (naturwissenschaftliche) *Induktion* (vom Besonderen zum Allgemeinen). Dagegen steht in der Mathematik die *Deduktion*, die genauer in der mathematischen *Logik* besprochen wird¹.

Wie verhalten sich somit Natur und Mathematik zueinander? Man schaut z.B. im Experiment, wie die Natur funktioniert und nimmt dazu aus der Mathematik das Gesetz, das z.B. quantitativ den gleichen Sachverhalt spiegelt. Soll etwa die Formel, d.h. die Sprache der Mathematik, also passen, so muss sie sich verhalten wie die Natur. So ein Verhalten „so wie hier ... so auch da“, nennt man *analog* und spricht von einem *Modell*:

Axiom I.2 (Mathematisches Modell) *Die Natur verhält sich analog zu den entsprechenden mathematischen Gesetzen. Diese Gesetze geben dann ein Modell der Natur.*

Modellhaft wird also die Natur durch die Mathematik beschrieben, so exakt wie nur möglich auf Grund unexakter Messungen, bis eben exaktere Messungen nach einem *exakteren Modell* verlangen. Das Modell verhält sich zur Realität so etwa wie die Modelleisenbahn zur wirklichen Eisenbahn. Man kann wohl mit der Modelleisenbahn Rangierprobleme studieren, was billiger ist als dies mit der wirklichen Eisenbahn zu tun. Doch mit der Modelleisenbahn eine Reise zu unternehmen, das geht eben nicht. Ein Modell stimmt immer nur in einem gegebenen Rahmen: dem Messrahmen. Es ist nie selbst die Wirklichkeit, es ist uns immer nur *Denkhilfe* oder auch *Denkkrücke*.

Wir merken uns:

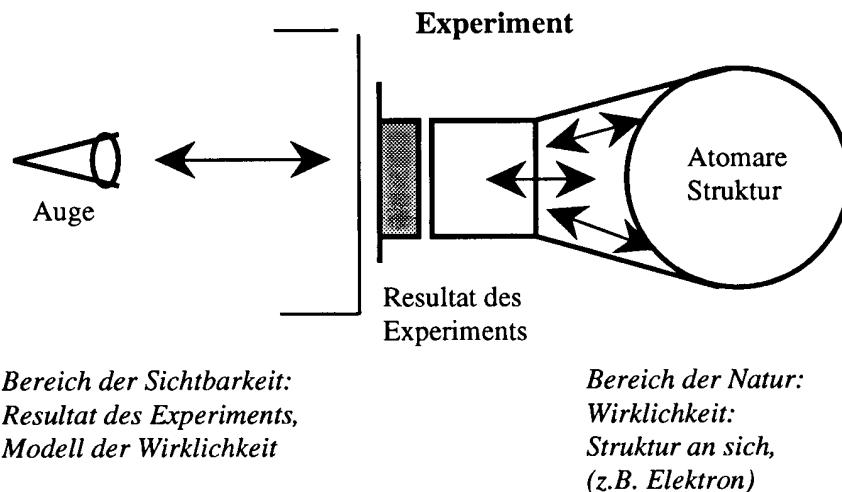
Wir erleben Natur wie Mathematik in ihrem Einzelverhalten kausal.
Natur und Mathematik zueinander aber verhalten sich analog.

I.2.4 ♣ Wozu ein Modelle?

In Abb. I.2 ist die Beziehung zwischen *sichtbarer Wirklichkeit* und *Wirklichkeit an sich* schematisch dargestellt. Die Brille, durch die man sich feinere Strukturen in der Natur gewahr werden kann, die ausserhalb der durch die Sinne erreichbaren Grössenordnungen und Möglichkeiten liegen, ist das *Experiment*. Z.B. hat wohl noch niemand ein Elektron direkt „vorbeifliegen“ sehen. So ein Teilchen selbst

¹Vgl. dazu Teil 2, vom Besonderen zum Allgemeinen, oder auch die Bemerkung zu Euklid auf Seite ??, Abschnitt ??.

Abbildung I.2: Schematische Darstellung der Beziehung zwischen Mensch und Natur via Experiment



hat ja keine Farbe, keinen Geschmack, keinen Geruch u.s.w.. Was also zeigt das Experiment? Als sichtbares Resultat erscheint natürlich nur die Wirkung des Experiments auf diese atomare resp. partikuläre Struktur. Es zeigt uns ein Modell, das beim selben untersuchten Gegenstand abhängig von der Experimentieranordnung sehr verschieden ausfallen kann. So erweist sich ein Elektron einmal als *Partikel*, ein anderes Mal als *Welle*. Die dahintersteckende eigentliche Wirklichkeit jedoch bleibt uns verborgen. Sie ist anders geschaffen als die Welt, in der wir leben: Die Welt des Menschen, der psychischen Wirklichkeit unseres Bewusstseins, die nicht mit der äußeren Realität gleichgesetzt werden soll.

Was taugen nun solche mathematischen Modelle? Sie sind eine tragende Säule unserer Kultur und den Menschen ein Segen. Denn seit es die Möglichkeit einer mathematischen Beschreibung der Natur gibt, sind Konstruktionen an die Stelle von Experimenten getreten. Ein Hochhaus zu bauen stellt kein Wagnis mehr dar. Wir wissen im voraus, dass es halten wird, wie es berechnet worden ist. Über eine neue Brücke müssen wir nicht zuerst einen Versuchskarren schicken, um herauszufinden, ob sie hält, wenn die Rechnungen stimmen und alle Gefahren bei der Planung berücksichtigt worden sind. Wir können über die Schranken der Zeit hinweg das Verhalten der Natur so vorausberechnen und beherrschen in diesem Sinne die Zeit. Also:

Methode I.2 (Vorausberechenbarkeit des Verhaltens der Natur) Durch Verwendung mathematischer Modelle wird das Verhalten der Natur im jeweiligen Fall vorausberechenbar. Der Mensch wird hier so zum Herr der Zeit.

Dass das nicht immer so gewesen ist, zeigt die Geschichte. Z.B. vor ca. 5000 Jahren stürzte die erste große Pyramide von Medium beim Bau zusammen. Diese riesige Baukatastrophe muss vermutlich Tausende von Opfern gefordert haben. Man hat eben versucht — das erste Mal ohne Erfolg ...

Beweise sind auch nicht zuletzt daher äußerst wichtig, weil sonst die darauf fassenden Modelle falsch sein könnten und die Wirklichkeit damit falsch beschrieben würde, was in der Anwendung katastrophale Folgen haben kann. Auch für denjenigen, der gerechnet hat...



I.2.5 Wirklichkeit, Wahrscheinlichkeitsmodell und Entscheidung

Durch statistische Messungen soll ein Phänomen der Wirklichkeit beobachtet werden. Die Ziele des Experiments sind festgelegt, die aufzufindenden Kennzahlen spezifiziert, das Experiment kann beginnen. Was geschieht nun? — Wir sehen den folgenden Ablauf, der sich zu einem Loop entwickeln kann:

1. **Beobachtung** der Wirklichkeit, quantitative Messungen, Zählungen oder Registrierung von qualitativen Ausprägungen von Phänomenen.
2. Eventuelle Eingabe in den **Computer** und **Aufbereitung der Daten**.
3. **Präsentation** der Daten.
4. **Berechnungen** von Wahrscheinlichkeiten auf Grund eines gewählten und explizit bezeichneten Wahrscheinlichkeitsmodells.
5. **Schlussfolgerungen**: Auffinden von zufälligen oder unwahrscheinlichen Phänomenen.
6. **Entscheidungsfindung**: Erarbeitung von Entscheiden mittels des gewählten Modells.
7. Eventuell **neue Experimente**: Ist eine bessere Validierung des gewählten Modells durch neue Daten notwendig? — Können damit Fehler 1. oder 2. Art verkleinert werden u.s.w.?
8. Eventuelle **Relativierung** der Resultate. (Spezielle Umstände u.s.w.)

I.3 Mathematische Modelle und Parameter

I.3.1 Zur Grundgesamtheit

Zum Auffinden oder Schätzen von Kennzahlen von Grundgesamtheiten benötigen wir bekanntlich Wahrscheinlichkeitsmodelle, mit welchen alle möglichen Messwerte modelliert werden können. Als Modelle haben wir bis jetzt schon viele Wahrscheinlichkeitsfunktionen kennengelernt, etwa diejenige für einen Mittelwert, um einen Erwartungswert zu schätzen. Bei solchem Unterfangen ging es meistens um den Schluss von Stichprobenfunktionswerten auf Kennzahlen der Grundgesamtheit. Denn die statistische Arbeit hat ja zum Ziel, die **Grundgesamtheit** durch geeignete Parameter und Modelle zu erfassen, da es oft nicht möglich ist, **Vollerhebungen** durchzuführen (Beispiel: **Prognose** von Wahlergebnissen von den Wahlen). Dabei hat man **zwei Typen** von Grundgesamtheiten zu unterscheiden:

1. **Endliche Grundgesamtheiten** wie etwa eine fixe **Population**, eine variable, jedoch theoretisch dennoch erfassbare Population, ein **Bestand von Dingen** oder **Ereignissen** wie ein Lagerbestand oder die ehemaligen Wahlerfolge einer Partei in einem aktuell nicht mehr existierenden Staat.
2. **Konzeptionelle** oder **praktisch beziehungsweise theoretisch mögliche Grundgesamtheiten** wie etwa vergangene zusammen mit auch künftige **mögliche Ereignisse** in einer Produktionskette, oder einem **Herstellungsprozess**, mögliche **Wasserzusammensetzungen** in den Weltmeeren, mögliche Stoffanteile in der Atmosphäre u.s.w. infolge vorhandener und möglicher Abgasentwicklungen u.s.w.

Wichtig ist, dass die **Grundgesamtheit** vor der Durchführung eines Experiments **genau spezifiziert** ist. Man denke z.B. an Wahlprognosen oder Verkaufsprognosen. Man muss also genau abgrenzen, welche Teile oder Ausprägung einer Bevölkerung an Wahlen teilnehmen wird bzw. als Käufer in Frage kommt und was etwa die Mächtigkeit der Grundgesamtheit ist, bevor man an die Planung der Erhebung einer Stichprobe gehen kann. Denn dazu muss der Stichprobenumfang und die Erhebungsmodalitäten (Versuchsplan) festgelegt sein.

I.3.2 Zur Stichprobe

Zu einer Stichprobenerhebung sind die zu bestimmenden Parameter im Versuchsplan zentral. Denn es gibt in der Realität verschiedene Stichprobenwerte–Arten oder **Kategorien**:

1. Erhebungsresultate können **skaliert (geordnet)** oder **nicht skalierbar** sein — ein Ordnungsprinzip wäre da nicht sehr einfach einzurichten. Beispiel: Die Länge eines Gegenstandes ist skaliert. Man kann sie in eine Skala eintragen. Man kann Längen ordnen. Fragt man jedoch nach der Lieblingsfarbe einer Dame, so lässt sich das Resultat kaum vernünftig in eine eindimensionale Skala einordnen und dann dort verrechnen. Solche Resultate nennen wir **kategorial**.
2. **Messresulata** sind oft **reelle Zahlen**, also **stetig**. Ein Mittelwert könnte also als Messresultat meistens auch vorkommen. Messresultate sind praktisch immer mit **Messfehlern (Toleranzen)** behaftet (**Messunsicherheit**).
3. **Zählresulata** sind oft **natürliche Zahlen**, also **diskret**. Hier kann ein Mittelwert nicht immer als Zählresultat vorkommen. Man denke an den Mittelwert beim Würfeln.
4. Beim Erheben von Messresultaten können gleichzeitig mehrere Größen oder **Merkmale** erhoben werden. Z.B. die Länge und das Gewicht eines Werkstücks. Messungen mit nur einem Merkmal nennen wir **univariat**, solche mit zwei Merkmalen **bivariat**, solche mit mehreren Merkmalen **multivariat**.
5. Dagegen sind **Stichprobenfunktionen** Funktionen, welche Stichprobewerte x_1, \dots, x_n etwa in eine Zahl oder einen Vektor abbilden. **Beispiele** sind der Mittelwert oder die Standardabweichung, wobei hier verschiedene Bezeichnungen in Gebrauch sind:

$$\bar{x} = \hat{\mu} = \frac{x_1 + \dots + x_n}{n}, \quad s = \widehat{StD} = \widehat{SD} = \sqrt{\frac{1}{n-1} ((x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)}$$

I.3.3 Wahrscheinlichkeit, Verteilungsfunktion, Wahrscheinlichkeitsmodell

Bekanntlich nennen wir eine Variable X , welche uns die Grundgesamtheit $G = \Omega$ zufällig in die Bildmenge der Messresultate B abbildet, eine **Zufallsfunktion** oder **Zufallsvariable**: $X : G \mapsto B$

Zu einer Zufallsvariable, welche Zahlenwerte abbildet, gehört gewöhnlich eine **Verteilungsfunktion**:

$$F_X(x) = P(X \leq x), \quad x \in G = \Omega, \quad 0 \leq F_X(x) \leq 1.$$

Verteilungsfunktionen auf Grundgesamtheiten mit diskreten Messwerten sind **Treppenfunktionen**. Die zugehörigen **Wahrscheinlichkeitsfunktionen** ($f_X(x_k) = F_X(x_k) - F_X(x_{k-1})$) nennen wir wegen der Ähnlichkeit ihrer Stabdiagramme zu den Massenverteilungen am Balken in der Statistik auch **Massenfunktionen**.

Grundgesamtheiten mit stetigen Messwerten sind gewöhnlich **stetige Funktionen**. Dazu gibt es aber auch Ausnahmen.

Bei stetigen Verteilungsfunktionen ist $P(X \leq x) = F_X(x) = \int_{-\infty}^x f(u) du$, wobei $f(u)$ jetzt eine **Wahrscheinlichkeitsdichtefunktion** ist. Hier gilt dann: $P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1)$.

Beispiel: Ein **idealer Zufallsgenerator** genügt dem Gesetz der folgenden Funktion:

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x \leq 1 \\ 1 & 1 < x \end{cases}$$

Wir nennen somit eine **Verteilungsfunktion** einer fixen diskreten Grundgesamtheit (resp. einer Variablen eines diskreten Merkmals der Grundgesamtheit) **diskret**, insofern sie dort nur diskrete Werte annehmen kann. Eine Verteilungsfunktion eines stetigen Merkmals einer konzeptionell unendlichen Grundgesamtheit nennen wir **stetig**.

Diskrete Verteilungsfunktionen auf diskreten, endlichen Grundgesamtheiten kann man oft einfach konstruieren, wenn man Laplace'sche Wahrscheinlichkeiten nachweisen kann. Man denke z.B. an eine Bernoulli- resp. eine Binomialverteilung. Anders ist es oft bei unendlichen Grundgesamtheiten. Hier muss man dann mit **Modellannahmen** arbeiten und als Verteilungsfunktion **parametrische Modelle** benutzen, deren Parameter sich einfach schätzen lassen. Man denke hier als Beispiel an eine Normalverteilung oder an eine Exponentialverteilung.

Eine stetige Zufallsgrösse mit ihrer Verteilungsfunktion $F_X(x)$ heisst **symmetrisch**, wenn für alle Δx gilt: $P(X \leq Med - \Delta x) = P(X \geq Med + \Delta x)$. Für stetige symmetrische Verteilungsfunktionen gilt dann natürlich: $Med_X = \mu_X$

Ein **Wahrscheinlichkeitsmodell** für ein **Merksmal** oder eine **Messgrösse** aufgefasst als **Zufallsgrösse** bezüglich einer bezeichneten **Grundgesamtheit** ist durch eine solche **Verteilungsfunktion** gegeben. Diese beschreibt das Wahrscheinlichkeitsverhalten der Messgrösse bezüglich der Grundgesamtheit.

Das **Problem** ist nun, dass eine **Verteilungsfunktion** oft **unbekannt** ist und daher unter Annahmen und auf Grund von Wahrscheinlichkeitsbetrachtungen konstruiert werden muss. Einfacher wird es, wenn man für die notwendigen statistischen Betrachtungen Methoden verwendet, welche unabhängig sind von Verteilungsfunktionen. Man spricht hier auch von **nicht-parametrischen Methoden**. Bei der Betrachtung der Hypothese H_0 aus Seite 83 haben wir eine solche Methode verwendet.

Bei gegebener Verteilungsfunktion kann eine Zufallsgrösse (Zufallsvariable) X in der Regel vollständig durch die Verteilungsfunktion und die Grundgesamtheit beschrieben werden. Leider ist, wie eben ausgeführt, die Verteilungsfunktion oft unbekannt. Oder ihre Bestimmung erfordert einen enormen Aufwand, zumal man es oft mit Wahrscheinlichkeiten zu tun hat, welche sich nur mittels relativer Häufigkeiten abschätzen lassen. **Daher verzichtet man oft auf die Verteilungsfunktion und begnügt sich mit Schätzungen von Parametern oder Kennzahlen.** Eine Kennzahl Θ , welche zu einem (vielleicht unbekannten) Wahrscheinlichkeitsmodell resp. zu der damit verbundenen Verteilungsfunktion einer Zufallsgrösse X der Grundgesamtheit Ω gehört, nennen wir auch **Estimand** (**Schätzparameter** oder **Parameter** $\Theta = fct(F_X)$).

Beispiele solcher Parameter haben wir schon genügend kennen gelernt: Spezielle gefragte **Wahrscheinlichkeiten** $P(x_k)$ oder $P(X \leq x_k)$, **Lageparameter** wie den **Erwartungswert** oder den **Mittelwert** μ der Grundgesamtheit, den **Median** Med (auch **Zentralwert**) oder α -Quantile. ($q_\alpha := \text{Min}(x \mid F(x) \geq \alpha)$). Weiter die **Streuparameter** wie die **Standardabweichung** $\sigma_X = SD_X = StD_X$ oder die **Quartilsdifferenz** $d = q_{0.75} - q_{0.25}$. Dabei gilt $P(q_{0.25} < X \geq q_{0.75}) = 0.5$.

Achtung: Bei unendlichen Grundgesamtheiten können **Erwartungswerte** auch **unendlich sein**. Oder man kann sie manchmal gar nicht berechnen. Der Median hingegen ist endlich.

Hinweis: Auf Seite 33 haben wir auch erfahren, dass der Schätzer $\sqrt{\frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{x})^2}$ für die Standardabweichung σ_X der Grundgesamtheit problematisch ist. Oft verwendet man auch den Maximum-Likelihood-Schätzer $s_X^{ML} = \sqrt{\frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^2}$, mit dem man σ_X weniger überschätzt.

I.3.4 Prozesskontrolle, Prozesstauglichkeit

Da aus Qualitätsgründen Produktionsprozesse meistens kontrolliert ablaufen müssen, überwacht man oft die Produktion mittels Messgrößen, die durch Zufallsfunktionen wie X gegeben werden können. Die konkreten registrierten Werte x_k trägt man dann in **Kontrollkarten** ein, in denen sichtbar gemacht werden kann, ob die genannten Werte x_k über eine gewisse Schranke UCL (upper control limit, **obere Kontrollgrenze**) oder unter einer gewisse Schranke LCL (lower control limit **untere Kontrollgrenze**) zu liegen kommen. Oft gibt man diese Limiten an durch Formeln wie $LCL = \mu - L \cdot \sigma$ und $UCL = \mu + L \cdot \sigma$. Dabei ist L ein Wert, welcher den Qualitätsstandard definiert und daher bekanntgegeben werden muss. Als **Prozesstauglichkeit** C_p von X definiert man dann den Wert

$$C_p = \frac{\text{Erforderliche Spezifikation von } X}{(\sigma)}.$$

Die erforderliche Spezifikation von X kann z.B durch ein toleriertes $|x_{max} - \mu|$ gegeben sein:

$$|x_{max} - \mu| = |UCL - \mu| = |\mu + L \cdot \sigma - \mu| = L \cdot \sigma$$

Z.B. mit $L = L_0 = 2$ wäre dann die Produktion bei $|x_k - \mu| \geq 2\sigma$ zu stoppen.

I.3.5 Zu den Box–Whisker–Plots

In den Box–Whisker–Plots werden bekanntlich der Median, $q_{0.25}$, $q_{0.75}$ sowie x_{Min} und x_{Max} angegeben. Damit haben wir Lage– und Streuparameter, welche im Diagramm sichtbar gemacht werden und an denen, falls es sich um erfasste Grundgesamtheiten handelt, Wahrscheinlichkeitsgrenzen für X abzulesen sind: 0 %, 25 %, 50 %, 75 %, 100 %. Bei Stichproben haben wir statt exakte Werte nur Schätzer vor uns. Damit lassen sich dann Stichproben vergleichen. Überdecken sich die Intervalle bei zwei Stichproben kaum, so stammen diese Proben vermutlich nicht aus derselben Grundgesamtheit.

<http://de.wikipedia.org/wiki/Boxplot>

I.3.6 Tschebyschow–Ungleichung

Auf Seite 34 haben wir die Tschebyschow–Ungleichung kennen gelernt. Wir übernehmen den wichtigsten Teil des dort gesehenen Textes hier, weil die Sache bei Datenerhebungen sehr wichtig ist:

Satz: (Tschebyschow– oder Chebyshev–Ungleichung)

Vor.:

X sei Zufallsgrösse mit Verteilungsfunktion F und $\mu, \sigma \in \mathbb{R}$ (endlich).

Beh.: $P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) \leq \frac{\sigma^2}{\Delta x^2}$

Damit kann man also die Wahrscheinlichkeit abschätzen, dass x_k ausserhalb eines zu μ symmetrischen Intervalls der Breite $2\Delta x$ liegt. Das hat eine grosse **praktische Bedeutung**, wenn die **Verteilungsfunktion unbekannt** ist, man aber μ oder vielleicht σ kennt!

Konsequenz: Für alle Wahrscheinlichkeitsmodelle gilt unter den eben beschriebenen Voraussetzungen:

$$1. \text{ Sei } \Delta x = 2\sigma \Rightarrow P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) \leq \frac{\sigma^2}{(2\sigma)^2} = \frac{1}{4} = 0.25.$$

$$2. \text{ Sei } \Delta x = 3\sigma \Rightarrow P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) \leq \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9} = 0.11\bar{1}\dots$$

$$3. \text{ Sei } \Delta x = 4\sigma \Rightarrow P((X < \mu - \Delta x) \vee (X > \mu + \Delta x)) \leq \frac{\sigma^2}{(4\sigma)^2} = \frac{1}{16} = 0.0625.$$

Damit lassen sich also, wie man sieht, Wahrscheinlichkeiten von Werten abschätzen, welche „beachtlich“ weit weg vom Erwartungswert μ liegen.

Sei zum Beispiel σ bekannt. Dann ist die Wahrscheinlichkeit, dass ein Wert x_k ausserhalb des Intervalls $I_{3\sigma} = [\mu - 3\sigma, \mu + 3\sigma]$ liegt, nie grösser als $0.11\bar{1}\dots$

Folgerung: So lässt sich aussagen, dass in einer beliebigen Grundgesamtheit mit endlichem μ und σ immer ca. 89 % der Werte im Intervall $I_{3\sigma}$ liegen! — **Ausreisser sind demnach wenig wahrscheinlich.**

I.4 Zu den Datenerhebungskonzepten

I.4.1 Wo und wieso Daten erheben?

Grundsatz: Datenmaterial sollte sehr vertrauenswürdig sein und eine maximal hohe Qualität aufweisen. Denn Fehler können zu falschen Abschätzungen, Entscheiden oder bei darauf folgenden falschen Sicherheitsentscheiden gar zu Verlusten von Leben führen.

Daten werden in vielen Bereichen gesammelt,

1. um die **Effizienz** von Produktionsmethoden, von Produkten (man denke z.B. an Medikamente oder an Fahrzeuge) zu belegen,
2. um die **Sicherheit** von Produkten zu garantieren (wieder denke z.B. an Medikamente oder an Fahrzeuge),
3. um **Prozesse** zu kontrollieren (Produktionsprozesse, chemische Prozesse, ökonomische Prozesse u.s.w.),
4. um Kennzahlen für die Planung zu erhalten (man denke z.B. an die Bereitstellung von Lebensmittel bei Grossverteilern oder auch an die Öffnung von Schaltern und das termingerechte Aufbieten von Schalterpersonal beim Betrieb einer grossen Warenmesse),
5. um die **Qualität** im Griff zu haben (man denke an die Einhaltung von Toleranzen bei technischen Halbfabrikaten u.s.w.),
6. um das psychologische Verhalten eines Kollektivs **vorhersagen** zu können, z.B. anlässlich von Wahlen,
7. um an die **Risikoabschätzung** bei Versicherungen oder auch bei Naturgefahren,
8. um in einem Geschäft über den Geschäftsgang **orientiert zu sein** und **Entscheide** fällen zu können.

Abstrakter gefasst könnte man sagen: Wir sammeln Daten

1. zum Verständnis einer aktuellen oder einer historischen Situation,
2. zur Analyse von Beziehungen zwischen Grössen einer Gesamtheit,
3. zur Kontrolle und Regulierung von Prozessen,

4. zur Beschaffung von Entscheidungsgrundlagen,
5. zur Beschaffung von Beweismaterial zur Belegung von Sicherheit,
6. zur Beschaffung von Beweismaterial zur Belegung von Effizienz und Qualität.

Die letzten beiden angedeuteten Forderungen gelten auch für das Datensammeln selbst: **Effizienz und Qualität auch vor allen Entscheiden und Beweisen beim Datensammeln selbst angesagt!** Die Kosten verlangen dies. Also keinen **Datenmüllhaufen** erzeugen und die **finanzielle Grundlage** nicht ruinieren: **Nicht für den Abfallhaufen produzieren!**

I.4.2 Zur Datenqualitätssteuerung

Bei der Planung der Datenerhebung kann die **Datenqualität gesteuert** werden.

1. Wenn möglich sollen **messbare Daten** erhoben werden. Das Resultat besteht dann aus **Zahlen**. Daraus errechnet man dann gewöhnlich Schätzungen für Lageparameter und Streuparameter oder Kombinationen davon.
2. Falls **kategoriale Daten** erhoben werden, soll wenn möglich eine **Ankreuzliste** präsentiert werden, die logisch konsistent ist. Hier ist unabdingbar, dass der Listenersteller das logische Denken auch de facto beherrscht. Dieses wird vor allem in der Mathematik geübt und nicht unbedingt durch betriebsinterne Beförderungen erworben.
3. Falls die Datenkategorien nicht bis zu einer Ankreuzliste heruntergebrochen werden können, wie es etwa in der medizinischen Anamnese der Fall ist, müssen **Fallbeschreibungen** erstellt werden. Das ist daher aufwändig.

Daten können aber oft nicht einfach erhoben werden. Es verbietet einem zwar niemand, die Breite einer Brücke oder die Geschwindigkeit eines fremden Fahrzeuges zu vermessen. Dagegen wäre es gegen die Regeln der Ethik, wenn ein Lehrer in einer Schulkasse plötzlich beginnen würde, medizinische Versuche durchzuführen. Solche unterstehen ethischen Regeln (hier die Helsinki-Konventionen u.s.w.). Allgemein ist es oft auch in andern Gebieten so, dass ein Versuch erst einer Regelung und einer Zulässigkeitskontrolle bedarf. Beispiele von Feldern mit Regelungsbedarf bei Experimenten:

1. Notwendige Abmachungen und Offenlegung von Information zwischen Produzent und Konsument.
2. Notwendige Abmachungen und Offenlegung von Information zwischen Arzt und Patient.
3. Notwendige Abmachungen und Offenlegung von Information zwischen Pharmaindustrie und Registrierungsbehörden oder zwischen Vertriebsstellen, Labors und Kontrollstellen.
4. Notwendige Abmachungen und Offenlegung von Information zwischen Lebensmittelproduzenten und Kontrollbehörden.
5. Notwendige Abmachungen und Offenlegung von Information zwischen Giftstoffverkaufstellen und Kontrollbehörden. — u.s.w.

Bei Datenerhebungen sind auch die **Einflussfaktoren der Daten** genaustens zu recherchieren und zu beachten. Man denke dabei an die früher verbreitete Idee der „Montagsautos“. Autos, so sagte man, welche von Hand am Montag produziert worden waren, hatten allerlei kostenverursachende Mängel. Einmal streikte der Vergaser, ein anderes Mal viel ein Rad ab. Als Grund sah man den freien Sonntag, wo viele der Arbeiter, so sagte man bösartigerweise, nichts besseres gewusst hätten als sich zu besaufen. Am Montag hatten diese dann ihren Kater. Daher hätten einige von ihnen ungeheuer schlechte Arbeit geleistet. Und ein Fehler genügt schon, damit ein Auto in der Garage bleiben muss... Zum Glück leisten heute die Maschinen die Arbeit, welche im Mittel am Montag dieselbe Qualität hat wie am Samstag. So hat diese gewiss übertriebene üble Nachrede ein Ende gefunden.

Wichtig ist aber folgende Feststellung: In **Systemen** oder **Prozessen**, an welchen man Daten erfassen sollte, sind **kontrollierbare** oder **bekannte** und **nicht kontrollierbare Faktoren (Rauschfaktoren, unbekannte Faktoren)** wirksam. Solche Faktoren können die Messgrößen beeinflussen. Man denke etwa an Luftfeuchtigkeit oder Temperatur in einem Labor als bekannte Faktoren.

Den **Einfluss konstanter Faktoren** kann man dadurch **minimieren**, indem man **konstante Messbedingungen** schafft. Dass man also im Beispiel des Labors die Luftfeuchtigkeit oder die Temperatur konstant hält (klimatisierte Räume, bekannt von der 1. Computergeneration her...)

Bekannt und sehr unerwünscht sind nicht bekannte Faktoren vor allem bei Medikamenten. In klinischen Studien in der Pharmaindustrie werden daher bei kleinsten Vermutungen Unsummen von Geld aufgewendet, um negative Faktoren auf die Spur zu kommen, welche den Tod eines Konsumenten des Medikamentes verursachen könnten. Hier geht es vor allem um **Interaktionen** zwischen verschiedenen Medikamenten, zwischen Medikament und chemischen Stoffen in gewissen Nahrungsmitteln oder zwischen Medikamenten und veränderten körperlichen Umständen. Man denke an das Schlafmittel Kontergan, welches bei schwangeren Frauen an den Fötten enorme genetische Schäden verursacht. Oder an Medikamente, welche in Interaktion mit Stoffen wie z.B. Tyramin (konzentriert vorhanden in gewissen Käsesorten) sehr rasch zu hohem Blutdruck führt — und damit zu Schlaganfall und damit zum Tod.

Man versucht daher den Faktoren auf die Spur zu kommen, indem man die Messgrößen als Funktion von unbekannten Faktoren auffasst. Hat man einmal einen Faktor etwas lokalisiert, so kann dieser wieder eine Funktion anderer Faktoren sein. Ein statistisches Instrument um Faktoren sichtbar zu machen ist die **Faktorenanalyse**, z.B. mittels Kreuztabellen und Tests.

Ein Problem kann es auch sein, wenn man die Experimentierumgebung zu stark einschränkt. Wenn man die Faktoren alle fix einstellt und man so den **Wirkungsraum des Experiments** klein hält. Dann erhält man Daten nur für einen kleinen Wirkungsraum, aus dem unter Umständen gar keine greifenden Schlüsse gezogen werden, da man dann sowieso nur einen Spezialfall kennt. Es geht also bei statistischen Untersuchungen auch darum, alle wesentlichen Einflussfaktoren zu bestimmen und zu untersuchen. Man denke dabei an ein Medikament: Ein solches sollte in allen Lebensumständen, in jedem Alter, von jedem Geschlecht, von jeder funktionell unterscheidbaren Schicht, in jedem Klima, zu jeder Tageszeit u.s.w. eingenommen werden können... In der Praxis benutzt man zur Untersuchung von Einflussfaktoren oft **faktorielle Pläne** und **Ursache–Wirkungs–Diagramme, Pareto–Diagramme** (nach Vilfredo Pareto) u.s.w.

In einem Pareto–Diagramm kann man den geschätzten Effekt von Faktoren auf die gemessene Größe sowie von ihren Interaktionen sichtbar machen. Da es nicht möglich ist, auf diesem engen Raum hier dem Umfang der Sache gerecht zu werden, vergleiche man dazu die Fachliteratur. Hinweise findet man z.B. unter folgenden Links:

<http://de.wikipedia.org/wiki/Paretodiagramm>

<http://www.meistersite.de/pdfs/methodenkoffer/Pareto-Diagramm.pdf>

Oder nur intern zugänglich:

<http://rowicus.ch/Wir/Scripts/restricted/MasterIndex.html> unter (1)

<http://de.wikipedia.org/wiki/Experimentator-Effekt>

I.4.3 Placebo–Effekt, Doppelblindstudien, Monitoring

Neben den äusseren Einflussfaktoren gibt es auch die innern, bei den Experimentatoren liegenden Faktoren. In diesem Zusammenhang kennt man auch den **Experimentator-Effekt**. Man redet hier auch von **systematischen Fehlern**. Solche Fehler können die erhobenen Daten wertlos machen. Wichtig ist daher eine vorausgehende genaueste Analyse der inneren Logik in Fragebögen oder die richtige Kalibrierung von Messgeräten.

Bekannt sind z.B. auch die **psychischen Einflussfaktoren**, welche in der Medizin vermutlich ihren Anteil am **Placebo–Effekt** haben. Darunter versteht man eine statistisch belegbare Besserung im Krankheitsverlauf oder eine entsprechende Schmerzlinderung nach Verabreichung oder Einnahme von wirkstofffreien Präparaten, welche dann aber trotz fehlenden Wirkstoffen die Ursache für die Besserung sein müssen. Hier liegt in den Augen des Wirkstoffproduzenten ein „systematischer Fehler“ im Patienten vor, der aber vom Patienten aus gesehen kein Fehler sein kann, da er sich ja mit allen Mitteln um eine Besserung bemüht. Nur für den Produzenten ist dies schlecht, da ihm dadurch die Einnahmen aus dem Wirkstoffverkauf entgehen. Also: „*Wenn es dir schlecht geht, so nehme erst mal Placebo! Und wenn das dann nichts nützt, dann nimm eben Placebo forte!*“

Zum Nachweis der Wirkung eines Stoffes in der Medizin werden häufig sogenannte **Doppelblind-Studien** gemacht — wenn dies die Ethik zulässt. Bei solchen Studien werden die Testpersonen oder Probanden in zwei Gruppen eingeteilt, wobei weder der Experimentator noch der Proband seine Gruppenzuteilung kennt. Dann wird allen in der gleichen Darbietungsform eine Dosis verabreicht. In der einen Gruppe aber ist es der Wirkstoff, blind natürlich, da niemand weiß, was er bekommt. In der andern Gruppe ist es Placebo. Da Placebo üblicherweise auch wirkt, kann man die Wirkung des Stoffes nur erhärten, wenn sie die Placebowirkung wesentlich übersteigt.

Solche Studien müssen jedoch oft aus ethischen Gründen unterlassen werden. Denn man kann einem Patienten nicht Placebo verabreichen, wenn es um die lebensrettende Wirkung eines neuen Medikaments geht, die schon sehr stark zu vermuten, jedoch statistisch in hohem Masse noch nicht erwiesen ist. Denn solches wäre ja fahrlässige Tötung oder gar vorsätzlicher Mord.

Ähnliches wie für die Medizin gilt auch für die Biologie, für die Landwirtschaft oder die Nahrungsmittelindustrie.

Damit man eine gewisse Garantie für gute Datenqualität hat, ist ein **Monitoring** oder eine vor-Ort-Überwachung des Experimentablaufs notwendig. Insbesondere betrifft das das Ausfüllen von Fragebögen. Trotz des hohen Durchschnittlichen Bildungsstandes gibt es hier für gute Qualität erfahrungsgemäß keine Garantie. Vielfach ist einiges gar nicht ausgefüllt oder der Abbruch ist nicht kommentiert. Natürlich muss nach dem Tod eines Patienten das Ausfüllen eines Fragebogen abgebrochen werden, doch sollte man eben die Urdsachen kennen, da dies für eine Studie wesentlich ist.

Auch hat man immer wieder Examinatoren entdeckt, die gewisse Resultate favorisieren und dann dafür sorgen, dass ihre Wuschresultate auch eintreffen. Die Ralität wird dadurch verzerrt. Man redet von **verzerrten Resultaten**.

I.4.4 Grundfragen und Datenerhebungsprinzipien

Zuerst einige **Ratschläge** aus der Kochkiste der Vernunft:

1. Datenerheben heisst **protokollieren**. Protokolliere daher zuerst in oberster Absicht die **Versuchsziele**, aus denen **Schlüsse** gezogen werden sollen. Spezifiziere die Schlüsse.

2. **Definiere** die **Grundgesamtheit**, die gesuchten Parameter oder die Arten von Modellen und die Merkmale.
3. Erarbeite einen **Versuchsplan**.
4. Bestimme allfällige **Einflussfaktoren** und dazu die Niveaus.
5. **Optimierte** den Versuch unter Hochhaltung der **Qualität**.
6. Ziehe die **Schlüsse** aus den Daten gemäss den Prinzipien der abgestützten Methoden und der **Logik**.

Folgende **Grundfragen** stellen sich immer beim Entwurf einer datenbasierten Studie:

1. Wie soll man **randomisieren**, d.h wie soll man dafür sorgen, dass man der Natur einer Zufallsvariablen gerecht wird? Wie sorgt man für eine **Zufallsauswahl der Daten** (Zufallsstichprobe, Zufallsreihenfolge der Daten, **stochastische Unabhängigkeit** der Erhebungsresultate), so dass nicht systematisch Einflussfaktoren favorisiert werden? Wie charakterisiert man überhaupt *stochastische Unabhängigkeit* von Zufallsvariablen?
2. Wie kann man einen vernünftigen **Stichprobenumfang** berechnen, sodass einerseits die Aussagen repräsentativ gestützt werden können und andererseits nicht unnötig für viel Geld zuviel Daten gesammelt werden?
3. Wie **minimiert** man die Effekte der unerkannten Einflussfaktoren oder Raschfaktoren? Vgl dazu auch:
<http://de.wikipedia.org/wiki/Kreuztabelle> und
<http://de.wikipedia.org/wiki/Faktorenanalyse>

Erklärungen zu diesen Grundfragen:

1. **Die Randomisierung:** Zwei Zufallsvariablen (Zufallsgrößen) X_j und X_k nennen wir **stochastische unabhängig (stoch. unabh.)**, wenn die beobachteten Werte der einen Variablen die andere nicht beeinflusst. Das kann man wie mittels einer mathematisch gefassten Beziehung ausdrücken:

$$X_j, X_k \text{ stoch. unabh.} \Leftrightarrow P(X_j \leq x_j \Rightarrow X_k \leq x) = P(X_k \leq x).$$

Dabei ist z.B. $P(X_k \leq x)$ zu verstehen als $P(\text{Relation } X_k \leq x \text{ ist erfüllt})$.

Die Menge der Variablen $\{X_1, \dots, X_n\}$ ist entsprechend stoch. unabh., wenn die Elemente paarweise stoch. unabh. sind.

Die Erfahrung zeigt, dass man qualitativ die besten Resultate bekommt, wenn man eine Stichprobe **randomisiert**, d.h. die Elemente der Stichprobe **zufällig** und **kontrolliert** zieht. Kontrolliert heisst hier, dass man alle erfassbaren systematischen Einflüsse ausschaltet. Bei Personenbefragungen ergeben sich jedoch oft unverhinderbare Einflüsse, weil die einbringbaren Antworten meistens auf Freiwilligkeit beruhen. Man erfasst daher nur Personen mit einer gewissen Offenheit, welche Freiwilligkeit akzeptieren.

Mögliche Randomisierungsmodell: Die Randomisierung gelingt z.B. anlässlich einer Zufallsauswahl ohne Wiederholung von Elementen n bei einer gegebenen endlichen Grundmenge mit N Elementen, indem man zuerst die Grundmenge nach irgend einem Prinzip durchnummert. Anschliessend wählt man mittels eines validierten Zufallsgenerator n Zahlen aus einem Zahlenbereich $\{1, 2, \dots, N\}$ aus. Damit hat man in der ausgewählten Reihenfolge die Nummern der Elemente aus der Grundmenge, mit welchen man jetzt die Stichprobe definiert.

Von einem Zufallsgenerator reden wir, wenn nach menschlichem Ermessen die Wahrscheinlichkeit alle je ausgewählten Zahlen gleich gross ist.

Über Randomisierungsmodelle lässt sich übrigens lange streiten. Wir wollen das hier unterlassen. Akzeptable Zufallsgeneratoren sind übrigens heute in allen einigermassen vernünftigen Mathematikprogrammen für PC implementiert.

2. Oft kann eine Stichprobe daher nicht in einem Randomisierungsschritt erzeugt werden. Sind in einer Stichprobe der Art wie im vorhin geschilderten Fall in der gezogenen Placebogruppe Schwerkranke drin, welche bei Vorenthaltung eines wirksamen Medimanets zu Schaden kämen, so muss man sie aussortieren und klassisch behandeln. Aus der verbleibenden Menge wird dann in einem zweiten Randomisierungsschritt eine endgültige Stichprobe gezogen u.s.w. So entstehen **mehrstufig gezogene Stichproben**.

Weiter kennt man **schichtweise** oder **strukturebenenweise** gezogene Stichproben. Bei einem Produktionsprozess wird man z.B. in täglichen Schichten Stichproben ziehen.

3. **Der Stichprobenumfang:** Manchmal ist der Grundgesamtheitsmittelwert (Erwartungswert) nicht bekannt. Man denke an bevorstehende Wahlen, wo die Grundgesamtheit vorher nicht fixierbar ist. Man kann hier von einer konzeptionellen Grundgesamtheit reden. Denn es können Wähler wegsterben u.s.w.. In solchen Situationen ist es oft üblich, diesen Erwartungswert den Mittelwert einer Stichprobe zu schätzen: $\mu_X = \mu_{\bar{X}} \approx \bar{X}$, $\hat{\bar{X}} = \frac{x_1 + \dots + x_n}{n}$, $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} \approx \frac{StD(X)}{\sqrt{n}}$.

Dazu ein **Gedankenexperiment**: Wir denken an eine Volksabstimmung mit ca. $10^6 = 1'000'000$ Abstimmenden. Wir wollen eine Umfrage für eine Abstimmungsprognose machen. Wir bestimmen dazu eine randomisierte Stichprobe von n Personen, welche zu „Ja / Nein“ befragt werden. k Personen entscheiden sich für „Ja“, $n - k$ Personen für „Nein“. Wir legen fest: X ist die Zufallsvariable für die Anzahl Personen, welche sich für „Ja“ resp. 1 entscheidet. (Für Nein notieren wir 0). Wir schätzen damit die gesamte Wahrscheinlichkeit für „Ja“: $\mu_{\bar{X}} : 10^6 = \mu_X : 10^6 = P(X \leq \mu) \approx \frac{k}{n}$.

Grob schätzend argumentieren wir wie folgt: Es ist $2 \cdot \sigma_X \leq 10^6$, $n = \frac{\sigma_X^2}{\sigma_X^2} = \frac{(10^6)^2}{(10^4)^2} = \frac{10^{12}}{4 \cdot 10^8} = 2.5 \cdot 10^3$. Wir möchten etwa erreichen, dass gilt $\sigma_{\bar{X}} \approx 10^4$, um dann $\mu_{\bar{X}} \pm \sigma_{\bar{X}}$ als einigermassen valable Prognose verwenden zu können. Dann wird $n_{Max} \approx \frac{(10^6)^2}{4 \cdot (10^4)^2} = \frac{10^{12}}{4 \cdot 10^8} = 2.5 \cdot 10^3$.

Die Stichprobenumfänge bei Abstimmungsprognosen liegen etwa um einen Faktor 2 tiefer. Doch die Größenordnung ist mit dieser Abschätzung getroffen.

Hinweis: Um genauere Abschätzungen zu bekommen, sollte man mehr über die Stichprobe einbringen können. Nützlich ist die Verteilungsfunktion. Dazu bedarf es aber viel Erfahrung mit der konkreten Sache. Oft werden Stichprobengrößen auch in Normen festgelegt. Eine „Faustregel“ lautet: $n \approx \sqrt{N}$, $N = |\Omega|$. Das würde hier $n \approx 10^3$ ergeben.

Ein Link zur Sache:

<http://de.wikipedia.org/wiki/Stichprobe>

4. Die Methode mit der **Block-, Schicht- oder Clusterbildung**: Hier fasst man Elemente der Grundgesamtheit resp. der Stichprobe zu **Blöcken** oder **Clustern** zusammen, welche in welchen die Elemente wesentliche, für die Untersuchung relevante Eigenschaften gemeinsam haben. Wenn man z.B. so vorgehen will wie in einer Doppelblindstudie, so kann man durch Zufallsauswahl jedem Block die Hälfte der Probanden einer Gruppe 1 zuteilen und die andere Hälfte einer Gruppe 2. Damit entstehen 2 Zufallsgruppen, mit welchen man einen Wirkstoff gegen Placebo testen kann. Auch könnte man unter geeigneten Bedingungen mehrere Studien zu Clustern zusammenfassen und so die Fehlerwahrscheinlichkeit senken. Vgl. dazu auch:

<http://de.wikipedia.org/wiki/Clusteranalyse>

5. Bemerkung zur **stochastischen Unabhängigkeit**: Da in statistischen Methoden oft die stochastische Unabhängigkeit vorausgesetzt wird, sollte man diese auch prüfen. Dazu kann man z.B. Stichprobendaten in der Reihenfolge ihrer Gewinnung (Urliste!) plotten und so Trends ausmachen. Vgl. dazu auch die folgenden Links:

http://de.wikipedia.org/wiki/Stochastische_Unabhängigkeit (ersetze ae durch ä)

http://de.wikibooks.org/wiki/Mathematik:_Statistik:_Gemeinsame_Wahrscheinlichkeit_mehrerer_Ereignisse

I.4.5 Statistiken

Hier geht es um **Schätzgrößen** und **Testgrößen**. Wenn man keine Vollerhebungen machen kann, sich also auf Stichproben stützen muss, so ist man auf Schätzungen der Parameter der Grundgesamtheit angewiesen.

Gegeben sei eine Stichprobe von Messwerten einer Zufallsgröße oder Zufallsfunktion X , etwa $\{x_1, \dots, x_n\}$. Eine **Statistik** ist dann eine empirische Realisierung einer Stichprobenfunktion, also eine Zahl wie z.B. die Lageparameter, Streuparameter oder andere Größen. Solche wichtige „Statistiken“ sind uns längst bekannt: **Mittelwert** \bar{x} , **empirische Standardabweichung** $s = SD = StD$, **Median** x_{Med} u.s.w..

Mit „Statistiken“ lassen sich also Größen der Grundgesamtheit schätzen. Mit einem realen Würfel wurde z.B. ermittelt: $\bar{x} = \hat{\mu} = 3.428 \approx \mu = 3.5$.

Pearson hatte 24000 mal eine Münze geworfen und 12012 mal dabei Kopf erhalten. Er stellte fest: $P(X = \text{Kopf}) \approx 0.5005$. Das ist eine Statistik. (Also eine Funktionswert, berechnet aus der Stichprobe, d.h. eine Realisierung einer Funktion, hier $P = \frac{\text{Anzahl Kopf}}{N}$.)

Weitere **Beispiele**:

1. Gegeben sei eine Stichprobe $\Upsilon = \{x_1, \dots, x_n\}$ (Urliste). Daraus erstellen wir mittels der Sortier- oder Ordnungsfunktion für Listen neu eine **Rangliste**: Wir nennen sie $\Xi = \{x_{1,r}, \dots, x_{n,r}\}$ mit $x_{1,r} \leq x_{2,r} \leq \dots \leq x_{n,r}$. ($n = \text{Stichprobenumfang}$).
2. In der selbe Stichprobe ist der **empirische Mittelwert** $\bar{x} = \frac{x_1 + \dots + x_n}{n}$. Man verwendet diesen empirischen Mittelwert oft als Punktschätzer für den Mittelwert μ der Grundgesamtheit. \bar{x} ist bekanntlich nicht robust.
3. Der **empirische Median** ist gleich

$$\widehat{\text{Median}} = \widehat{\text{Med}} = \begin{cases} x_{(n+1)/2} & n \text{ gerade} \\ \frac{1}{2}(x_{n/2} + x_{(n/2)+1}) & n \text{ ungerade} \end{cases}$$

Man verwendet diesen empirischen Median oft als Punktschätzer für den Median Med der Grundgesamtheit. Der empirische Median ist bekanntlich ein robuster Schätzer.

4. Um den Mittelwert etwas robuster zu machen und ihn zur Schätzung des Erwartungswerts gebrauchen zu können, benutzt man oft das **gestutzte** oder das **winsorisierte Mittel**. Um diese zu bilden **modifizieren** wir die Stichprobe. Wir wollen diese Mittel nachstehend kurz erklären.

Vgl. dazu auch <http://de.wikipedia.org/wiki/Mittelwert>

Gestutztes und winsorisiertes Mittel:

Sei $\beta \in [0, 1]$. Damit bilden wir $\beta \cdot 100\%$ und $\beta \cdot 50\%$. Gegeben sei zudem eine Stichprobe Υ .

Wir entfernen nun aus Υ die $\beta \cdot 100\%$ grössten oder die $\beta \cdot 100\%$ kleinsten Werte oder wir entfernen die $\beta \cdot 50\%$ grössten sowie auch die $\beta \cdot 50\%$ kleinsten Werte, je nach Ausreissersituation. Dann erhalten wir mittels der verbleibenden Werte das β -gestutzte Mittel \bar{x}_β .

Wir entfernen wieder aus Υ die $\beta \cdot 100\%$ grössten oder die $\beta \cdot 100\%$ kleinsten Werte oder wir entfernen die $\beta \cdot 50\%$ grössten sowie auch die $\beta \cdot 50\%$ kleinsten Werte. Darauf ersetzen wir die entfernten Werte durch die verbleibenden naheliegenden grössten resp. kleinsten Werte. Dann erhalten wir mittels der neuen Werte das β -winsorisierte Mittel \bar{W}_β .

I.4.6 Konstruktion von Wahrscheinlichkeitsmodellen mit Statistiken

Beispiel:

1. In einer grossen Kasse befindet sich viel Geld (Grundgesamtheit K): 12 Stücke zu 1 CHF, 18 Stücke zu 2 CHF und 15 Stücke zu 5 CHF: Total also 45 Stücke. Damit kann man die Verteilungsfunktion für G (Geldstücke) von K konstruieren. Z.B. ist $P(G = 2) = \frac{18}{45}$. **Man hat also hier ein Wahrscheinlichkeitsmodell für die Zufallsgrösse G gewählt.**

Frage: Wie kommt man nun zu einem Schätzer für \bar{G} , ohne ihn direkt zu berechnen? —

Anhand dieser einfachen Situation können wir studieren, wie man Schätzer bilden kann und was für Situationen damit entstehen.

2. Jemand darf jetzt zufällig 3 Stücke ziehen. Damit entsteht jeweils eine Stichprobe Υ von 3 Elementen. **Damit ist erklärt, was hier die Stichprobe ist.**
3. Mit der Stichprobe können wir den Mittelwert \bar{g} als Schätzer für \bar{G} bilden. **Damit ist die Schätzung für die gesuchte Kennzahl festgelegt.**
4. Wie ist nun die zugehörige Wahrscheinlichkeitsfunktion zu wählen, d.h. wie reich wird man mit welcher Wahrscheinlichkeit bei einem Zug? Was ist also die **Wahrscheinlichkeitsfunktion** der Stichprobe $\hat{\Upsilon}$ (Mittelwerte)?

Hier sind die einzelnen Wahrscheinlichkeiten von Laplace'schen Typ, denn man kann die günstigen und die möglichen Fälle konstruieren. Wenn die Münzen nacheinander gezogen werden, so erhalten wir Kombinationen und können abzählen. Der niedrigste Betrag ist 3 mal 1 CHF, also total 3 CHF. Um das einzubringen hat man eine einzige Möglichkeit. Der nächstniedrigste Betrag ist 2 mal 1 CHF plus 1 mal 2 CHF gleich 4 CHF. Um diese einzubringen hat man 3 Möglichkeiten. Dann kommt der Fall mit dem nächstgrösseren Betrag 2 mal 2 CHF plus 1 mal 1 CHF. Um diese einzubringen hat man wieder 3 Möglichkeiten. Und so fort. Alle die so gewonnenen einzelnen günstigen Möglichkeiten zusammen ergeben die Anzahl der total möglichen Fälle.

Daraus erhält man wie gewohnt für jeden Fall die Laplace'sche Wahrscheinlichkeit und damit eine Verteilungsfunktion für die Mittelwerte. Diese ist ebenfalls eine Treppenfunktion. Sie hat viel mehr Treppenstufen als die Verteilungsfunktion von K ! Und man kann an dieser Stichprobenmittelwertsfunktion Wahrscheinlichkeiten schätzen — die Frage ist nur, wie genau. Wir reden daher hier von der Verteilungsfunktion der Schätzung. Diese **Verteilungsfunktion** lässt sich übrigens mit Hilfe der **Monte–Carlo–Methode simulieren**, was natürlich in unserem einfachen Fall nicht der einfachste Weg ist zur Gewinnung der Verteilungsfunktion. Doch in komplexeren Fällen kann die Methode nützlich sein. Dabei stützt man sich auf das Gesetz der grossen Zahlen von Bernoulli. Danach müsste die Wahrscheinlichkeit sehr klein sein, dass die mittlere Abweichung der Zufallsvariable X_i vom Erwartungswert $E(X_i)$ bei grossem n nicht so sehr klein wird. Siehe auch:

http://de.wikipedia.org/wiki/Gesetz_der_grossen_Zahlen

5. Nun müsste man die aus der Simulation gewonnene **Verteilungsfunktion der Schätzung** numerisch oder graphisch fassen. Bei einem Stichprobenumfang von $n = 3$ wirkt allerdings das Gesetz $\sigma_{\bar{G}} = \frac{\sigma_G}{\sqrt{n}}$ nicht besonders griffig. Erst bei grossem n gelingt es offensichtlich $\sigma_{\bar{G}}$ klein zu machen und damit \bar{G} genauer zu erhalten.

6. Anwendungen:

Man redet je nach Anwendungsart von **Schätzstatistiken** oder von **Teststatistiken**.

Schätzstatistiken verwendet man für **Parameterschätzungen**. Z.B. wird oft der unbekannte Parameter μ_X der Grundgesamtheit durch \bar{x} geschätzt, Med_X wird durch \widehat{Med} geschätzt oder σ_X durch $s = SD = Std$.

Teststatistiken dienen dagegen zur Verwendung in statistischen Tests, d.h. um Entscheidungen zu begründen.

I.4.7 Zu den Fehlern bei Schätzungen

Bei Schätzungen durch Schätzstatistiken kann es zu zwei Arten von Fehlern kommen: Zu **systematischen Fehlern** und zu **Zufallsfehlern**.

Schätzt man z.B. μ mehrmals durch j verschiedene Stichproben und gewinnt man dadurch die Schätzungen $\bar{x}_1, \dots, \bar{x}_j$, so stellt man vielfach fest, dass die gewonnenen Schätzungen nicht alle identisch sind. Man kann an die Abweichungen $\Delta_k = \mu - \bar{x}_k$ vom momentan unbekannten Wert der Grundgesamtheit μ denken. Diese Δ_k nennen wir **zufällige Fehler**. Sie haben wiederum eine Verteilung und können durch einen mittleren Streuwert, etwa die **Standardabweichung** $\sigma_{\hat{\Delta}}$, charakterisiert sowie quantitativ gefasst werden. Allgemein sprechen wir bei einem Parameter Θ mit den Statistiken bzw. Schätzungen $\hat{\Theta}_k$ bei der Standardabweichung $\sigma_{\hat{\Theta}}$ vom **Standardfehler** (*standard error SE*) von $\hat{\Theta}$. Kurz: Standardfehler hier = $SE(\hat{\Theta})$.

Neben dem Standardfehler können wir es aber auch mit einem **eingebauten Fehler** zu tun haben. Man denke etwa an **systematische Fehler** wie bei der Verwendung eines zu stark einseitig winsorisierten Mittels. Hier läge damit ein **methodischer Fehler** mit einem **Bias** resp. einer **Verzerrung** vor. (Hier die Differenz zwischen dem Erwartungswert der Statistik und dem Schätzwert.)

Liegt **keine Verzerrung** vor, so reden wir von einem **erwartungstreuen** oder **verzerrungsfreien Schätzer** (auch Bias-frei). Bias-freie Schätzer sind erwünscht, weil sie in der Regel brauchbare Werte liefern. (Der Bias gilt auch als Mass für die **Zentrierung** einer Schätzung um den zu schätzenden Parameter.)

Erwartungstreu ist dabei ein Schätzer $\hat{\Theta}$ für den Parameter Θ , wenn für den Mittelwert von Statistiken (Schätzwerten) in der Regel gilt: $\lim_{j \rightarrow \infty} \frac{\hat{\Theta}_1 + \dots + \hat{\Theta}_j}{j} = \Theta$. „In der Regel“ bedeutet genauer, dass

Ausnahmen erlaubt sind, dass aber für die Wahrscheinlichkeit gilt: $P(\lim_{j \rightarrow \infty} \frac{\hat{\Theta}_1 + \dots + \hat{\Theta}_j}{j} = \Theta)$ ist sehr klein. Allgemein können wir daher formulieren:

$$\hat{\Theta} = \Theta + (\Delta\Theta)_{\text{systematisch, Bias}} + (\Delta\Theta)_{\text{Zufall}}$$

Ein weiterer oft verwendeter Begriff ist der **mittlere quadratische Fehler**, **MSE**: $MSE = SE^2 + Bias^2$.

Wir wünschen natürlich, dass gilt: $MSE(\Theta) \ll \Theta$. Im Falle des Medians hat man gewöhnlich einen grossen Standardfehler. $MSE(Med) \ll Med$ wäre daher kaum zu erwarten.

Weiter hätte man gerne den zufälligen Fehler quasi in Schranken gewiesen. Der Schätzer sollte also auch robust sein.

Sei eine Grundgesamtheit Ω mit $|\Omega| = N < \infty$, $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}$ gegeben. Dann kann **MSE** von \bar{x} explizit berechnet werden.

Der Bias resp. die Verzerrung von eines Schätzers \bar{x} für $\mu = E(X) = E(\bar{X})$ ist hier null. Das lässt sich im folgenden Fall einfach zeigen. Wir betrachten hier der Abkürzung halber nur die Situation $\bar{X} = \frac{X_1 + X_2}{2}$, wobei X_1 und X_2 zwei identische Kopien von X sind. Seien dabei $\{x_1, \dots, x_n\}$ die möglichen Messwerte von X . Dann erhalten wir beim Ziehen mit Zurücklegen die folgenden Realisierungen von \bar{X} :

$$\begin{aligned} \bar{x}_{1,1} &= \frac{x_1 + x_1}{2} & \bar{x}_{1,2} &= \frac{x_1 + x_2}{2} & \dots & \bar{x}_{1,N} &= \frac{x_1 + x_N}{2} \\ \bar{x}_{2,1} &= \frac{x_2 + x_1}{2} & \bar{x}_{2,2} &= \frac{x_2 + x_2}{2} & \dots & \bar{x}_{2,N} &= \frac{x_2 + x_N}{2} \\ &\dots &&\dots &&\dots &\\ \bar{x}_{N,1} &= \frac{x_N + x_1}{2} & \bar{x}_{N,2} &= \frac{x_N + x_2}{2} & \dots & \bar{x}_{N,N} &= \frac{x_N + x_N}{2} \end{aligned}$$

Damit lässt sich hier der Erwartungswert nach Laplace explizit berechnen:

$$E(\bar{X}) = \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N \bar{x}_{j,k} = \frac{1}{N^2} \left(\frac{x_1 + x_1}{2} + \dots + \frac{x_j + x_k}{2} + \dots + \frac{x_N + x_N}{2} \right)$$

Da in jedem Summanden der Faktor $\frac{1}{2}$ steckt, kann man diesen mit $\frac{1}{N^2}$ vor die Klammer schreiben. Weiter erhält man, wenn man z.B. im vorhin gezeigten Schema über die Zeilen summiert:

$$\begin{aligned} E(\bar{X}) &= 1/(2N^2) (N x_1 + (x_1 + \dots + x_N) + N x_2 + (x_1 + \dots + x_N) + \dots + N x_N + (x_1 + \dots + x_N)) \\ &= 1/(2N^2) (N(x_1 + \dots + x_N) + N(x_1 + \dots + x_N)) \\ &= (2N)/(2N^2) (x_1 + \dots + x_N) \\ &= (1/N)(x_1 + \dots + x_N) \end{aligned}$$

Damit wird $E(\bar{X}) = \frac{1}{N}(x_1 + \dots + x_N) = \mu$.

Ebenso lässt sich der Standardfehler hier mit Blick auf die $x_{1,1}, \dots, x_{N,N}$ berechnen. Es gilt:

$$(\frac{x_j + x_k}{2} - \mu)^2 = \frac{1}{4} (x_j + x_k - 2\mu)^2 = \frac{1}{4} ((x_j - \mu) + (x_k - \mu))^2 = \frac{1}{4} ((x_j - \mu)^2 + 2(x_j - \mu)(x_k - \mu) + (x_k - \mu)^2)$$

$$\begin{aligned} SE &= \sqrt{\frac{1}{N} \left(\left(\frac{x_1 + x_1}{2} - \mu \right)^2 + \left(\frac{x_1 + x_2}{2} - \mu \right)^2 + \dots + \left(\frac{x_N + x_N}{2} - \mu \right)^2 \right)} \Rightarrow \\ SE^2 &= \frac{1}{N} \left(\left(\frac{x_1 + x_1}{2} - \mu \right)^2 + \left(\frac{x_1 + x_2}{2} - \mu \right)^2 + \dots + \left(\frac{x_N + x_N}{2} - \mu \right)^2 \right) = \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N \left(\frac{x_j + x_k}{2} - \mu \right)^2 \end{aligned}$$

Setzen wir die Formel von drei Zeilen weiter oben ein, dann erhalten wir:

$$SE^2 = \frac{1}{4N} \underbrace{\left(\sum_{j=1}^N \sum_{k=1}^N (x_j - \mu)^2 + \sum_{j=1}^N \sum_{k=1}^N 2(x_j - \mu)(x_k - \mu) + \sum_{j=1}^N \sum_{k=1}^N (x_k - \mu)^2 \right)}_{=N\sigma^2}$$

$$\text{Dabei gilt mit } \bar{x} = \mu : \sum_{j=1}^N \sum_{k=1}^N 2(x_j - \mu)(x_k - \mu) = 2 \sum_{j=1}^N ((x_j - \mu) \cdot \underbrace{\sum_{k=1}^N (x_k - \mu)}_{N\bar{x} - N\mu = N\mu - N\mu}) = 0$$

Somit gilt:

$$SE^2 = \frac{1}{4N} 2N\sigma^2 = \frac{\sigma^2}{2} \Rightarrow SE = \frac{\sigma}{\sqrt{2}}$$

In ähnlicher Weise findet man bei $\bar{X} = \frac{X_1 + \dots + X_m}{m}$ die Formel $SE_{\bar{X}} = \frac{\sigma}{\sqrt{m}}$

Wenn man jedoch die Stichprobe durch zufällige Elementauswahl ohne Zurücklegen erzeugt, so ergibt sich ein Korrekturfaktor. Die Formel lautet dann: $SE_{\bar{X}} = \frac{\sigma}{\sqrt{m}} \cdot \sqrt{1 - \frac{m-1}{N-1}}$. Der Faktor $\sqrt{1 - \frac{m-1}{N-1}}$ ist der erwähnte **Korrekturfaktor**. Auch kann man den Limes für $N \rightarrow \infty$ bilden und die Grenzwertaussage machen. Wir wollen das zusammenfassen:

Resultat für den Standardfehler eines arithmetischen Mittelwertes:

Vor.:

Sei Ω eine Grundgesamtheit mit $\Omega = N \leq \infty$. X sei Zufallsvariable mit $E(X) = \mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}$. Sei $\bar{X} = \frac{X_1 + \dots + X_m}{m}$ ein Schätzer für μ .

Beh.:

Bei zufälligem Ziehen ohne Zurücklegen gilt:

$$SE_{\bar{X}} = \frac{\sigma}{\sqrt{m}} \cdot \sqrt{1 - \frac{m-1}{N-1}}$$

Bei zufälligem Ziehen mit Zurücklegen gilt:

$$SE_{\bar{X}} = \frac{\sigma}{\sqrt{m}}$$

Für $N \rightarrow \infty$ wird $\lim_{N \rightarrow \infty} \frac{\sigma}{\sqrt{m}} \cdot \sqrt{1 - \frac{m-1}{N-1}} = \frac{\sigma}{\sqrt{m}}$.

Hingegen für $m \rightarrow \infty$ (was praktisch allerdings nicht zu bewerkstelligen ist), ergibt sich $SE_{\bar{X}} = \frac{\sigma}{\sqrt{m}} \rightarrow 0$.

Hinweis: Für den Median kennt man keine vergleichbare Formel. „Gehandelt“ wird die Näherungsformel:

$$SE_{\widehat{med}} \approx \frac{q_{0.75} - q_{0.25}}{\sqrt{n}} \cdot 1.57$$

I.4.8 Verteilungsfreiheit und Vertrauensintervalle

Wir gehen von folgender Situation aus:

Studiere eine Messgrösse oder Zufallsvariable X mit $X : \Omega \mapsto \mathbb{R}$ (univariat).

Gesucht sind z.B. Median, Mittelwert... (Lageparameter, Streuparameter...), allgemein: Θ .

Methode: Der Parameter Θ soll möglich biasfrei mit Hilfe von Stichprobe(n) geschätzt werden.

Zu diesem Zweck wählen wir die Stichprobe $\mathcal{S} = \{x_1, x_2, \dots, x_n\}$, $|\mathcal{S}| = n$.

Zusammenhang zwischen Θ und $\hat{\Theta}$: $\Theta = \hat{\Theta} + \Delta\Theta_{Zufall}$.

Das übliche Problem: Die Verteilungsfunktion $F(X)$ ist unbekannt, das Wahrscheinlichkeitsmodell (ω, X, F) ist unbekannt.

Ausweg: Wir benützen eine Methode, mit der F nur qualitativ gefasst resp. anhand der Daten nur angenähert wird. Also eine Methode, die nicht von bestimmten Wahrscheinlichkeitsmodellen abhängt. Dafür werden dann die Fehler größer abgeschätzt.

Primäres Vorhaben oder Ziel: Abschätzen von **Vertrauensintervallen**.

Das Beispiel des Medians

Sei $\mathcal{S} = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ geordnet: $x_1 \leq x_2 \leq \dots \leq x_n$. Der Median $Med \in \Omega$ ist unbekannt und soll durch \mathcal{S} geschätzt werden. Folgende **Hypothese** scheint **plausibel**: $x_1 \leq Med \leq x_n$, d.h. $Med \in I = [x_1, x_n]$. Für die zugehörigen Zufallsvariablen $X_{Min}, X_{Max}, \mathcal{I}$ müssten wir sinnvollerweise postulieren: $Med \in [\mathcal{I} = X_{Min}, X_{Max}]$.

Gesucht ist daher die zur Variablen \mathcal{I} gehörige Verteilung sowie $\gamma = P(Med \in \mathcal{I})$. Für $P(Med \in \mathcal{I})$ sagen wir auch: „ \mathcal{I} überdeckt“ den Median. Für grosse γ (d.h. $\gamma \approx 1$) kann man vermutlich **vertrauen** in die Methode haben und den Median durch I abschätzen. I wäre daher ein **Vertrauensintervall** für den Median zum Niveau γ . (In der Praxis trifft man für γ Werte wie 0.9, 0.95, 0.99, 0.975 u.s.w.)

Das Problem ist aber, dass man in der Regel γ wohl fordern kann, dagegen $P(Med \in \mathcal{I})$ nicht kennt, da man auch die Verteilungsfunktion zu \mathcal{I} nicht kennt. Daher wählt man in der Praxis oft ein Verfahren, in dem als zentralem Punkt von der **konkreten Annahme eines Wahrscheinlichkeitsmodells** ausgegangen wird. Alle daraus folgenden Schlüsse lassen sich nur unter der **Voraussetzung der Annahme der Gültigkeit dieses Modells** rechtfertigen. Hier liegt eine grosse Gefahr des **Missbrauchs** oder des infantilen Umgangs mit der Sache, wie dies in Zeiten von Wirtschaftskrisen immer wieder zu Tage tritt, wo die Gültigkeit zwecks Realisierung unedler Absichten nur vorgetäuscht ist. (Mit solchen Dingen, gepaart mit den verbreiteten mathematischen Unkenntnissen und der Wundergläubigkeit, lassen zweifellos und offensichtlich ungestraft dunkle und sehr einträgliche Geschäfte machen: Durch bewusst riskierte Täuschung, indem vorgetäuschte oder auch empirisch gewonnene Wahrscheinlichkeit mit **verdeckten Parametern** wie Sicherheit dargeboten wird, obwohl Hochrisiko besteht.)

Das Binärmodell für den Median — Methodenbeschrieb im Falle des Medians:

1. Wähle für $F(X)$ ein **Modell** mit $P(X = Med) = 0$. Das heisst, der Median kommt unter den Messwerten nicht vor. (Das lässt sich oft praktisch einrichten: Man arbeitet mit einer geraden Anzahl von einzeln unterscheidbaren Messwerten mit genügend einflusslosen Nachkommastellen, falls notwendig künstlich angefügt...) Dadurch wird die Menge der Messwerte durch den Median in zwei gleich grosse Untermengen geteilt: $\Omega_u = \{x_k \in \Omega \mid x_k < Med\}$ und $\Omega_o = \{x_k \in \Omega \mid x_k > Med\}$. Damit gilt im gewählten Modell: $P(X < Med) = 0.5 = P(X > Med)$. So erhält man zwei binär codierbare Aussagen oder Fälle. Wir haben ein **Binärmodell**.
2. Die Wahl der **Codierung**: Gilt $x_k < Med$, so ersetzen wir x_k durch 0. Gilt $x_k > Med$, so ersetzen wir x_k durch 1. Aus \mathcal{S}_n entsteht damit \mathcal{C}_n .
3. Lege im Rahmen der Möglichkeiten die **Stichprobengrösse** fest und ziehe eine Stichprobe nach dem Prinzip der Zufallsauswahl resp. der Randomisierung: $\mathcal{S} = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$. Wir wählen für die Diskussion z. B. $n = 10$.
4. Ordne die Stichprobe. Z.B sei $\mathcal{S} = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$ geordnet. Dabei ist ein Wert x_k eine Realisierung der Zufallsvariablen X_k . Alle X_k sind **unabhängige** identische Kopien von X .
5. Wähle die **Schätzstatistik**. Definiere dazu z.B.:
 $I_1 = [x_{Min}, x_{Max}] = [x_1, x_n], \quad I_2 = [x_{1+1}, x_{n-1}], \quad \dots \quad I_k = [x_{1+k-1}, x_{n-k}] = [x_k, x_{n-k}], \dots$
6. Nun definieren wir die **Überdeckungsbedingung**, um anschliessend damit eine Wahrscheinlichkeitsverteilung mit Hilfe einer Simulation zu konstruieren, falls die konkrete Rechnung der vielen Möglichkeiten wegen zu aufwändig wäre. Wir wählen für die Diskussion z. B. $\mathcal{I} = \mathcal{I}_2 = [x_2, x_9]$. Dieses Intervall codieren wir: Gilt $x_2 < Med$, so ersetzen wir x_2 durch 0. Gilt $x_2 > Med$, so ersetzen wir x_2 durch 1. Entwprechend für x_9 . Dann erhalten wir ein codiertes Intervall der Art $\mathcal{C}_2 = [0, 0]$ oder $\mathcal{C}_2 = [0, 1]$ oder $\mathcal{C}_2 = [1, 1]$. Die **Bedingung**, dass \mathcal{I}_2 den **Median überdeckt**, können wir jetzt übersetzt durch das Vorliegen der Codierung $\mathcal{C}_2 = [0, 1]$ ausdrücken.
7. Um zu einer **Verteilungsfunktion** zu kommen, können wir fiktive Ziehungen z.B. mit Hilfe der Monte-Carlo-Methode simulieren und die Resultate danach codieren, so wie eben beschrieben.

Dann muss für die Wahrscheinlichkeit γ (Wahrscheinlichkeit dass \mathcal{I} den Median überdeckt) gelten:

$$\gamma = \frac{\#(\mathcal{C}_2 = [0, 1])}{\# \text{ Simulationen}}$$

($\#$ steht hier wie üblich für „Anzahl“.)

Die Problematik des Resultats:

Beim Gebrauch dieser Metode muss man sich immer die folgenden Punkte kritisch vor Augen halten, um damit mögliche Probleme mit genügend grosser Wahrscheinlichkeit oder allenfalls Plausibilität entschärfen zu können:

1. Die Stichprobe beruht auf einer Zufallsauswahl aus einer als unbekannt vorausgesetzten Grundgesamtheit. Der kleinste gewählte Wert könnte daher viel kleiner noch sein oder aber auch grösser, grösser als der Median der Grundgesamtheit. Analog für den grössten Wert, was die Anwendung der Methode auf die gezogene Stichprobe sinnlos machen würde. Ebenso könnten Werte der Grundgesamtheit gehäuft vorkommen, ohne dass sie in der Auswahl erscheinen u.s.w. . Wenn man über etwas nichts zu sagen weiss, so kann man darüber nicht reden, auch über Chancen nicht. Dann müsste man darüber schweigen. Oder eben Information in Erfahrung bringen, die Eingrenzungen zulässt. Diese Information kann heutistischer oder auch sachlogischer Natur sein. Jedoch ist sie nicht allgemeiner Natur, sodass sie in Formeln einfließen könnte.
2. Dasselbe gilt für den Bias. Verdeckte systematische Fehler können unerkannt bleiben, besonders dann wenn man sie gar nicht zu ahnen vermag, oder auch wenn ihre Möglichkeit gegen die herrschende Ideologie verstösst...
3. Das Wahrscheinlichkeitsmodell ist aus Plausibilitätsgründen angenommen, jedoch nicht gesichert. Denn die Grundgesamtheit wird hier in unserer Situation als unbekannt vorausgesetzt. Im praktischen Fall wäre eine Begründung notwendig, die auch heutistisch und damit pragmatisch sein kann, wenn andere Gründe nicht aufzutreiben sind. Dadurch entstehen trotz vieler Rechnungen nie wirkliche Wahrscheinlichkeit — oder gar Sicherheit, sondern höchstens nur Plausibilität, das heisst schön gewollte Vermutung.
4. Zur Stichprobengrösse sind auch noch Fragen zu stellen. Je kleiner die Stichprobe ist, desto weniger Möglichkeiten hat man bei der Wahl von \mathcal{C}_1 und desto kleiner wird auch die Wahrscheinlichkeit des Erfolgs bei \mathcal{C}_1 . Je grösser die Stichprobe ist, desto mehr kann man mit dem k bei I_k spielen, um den Median genauer eingrenzen zu können. Die Genauigkeit geht aber immer auf Kosten der Grösse der Wahrscheinlichkeit.
5. Zum Schluss hat man immer noch eine Wahrscheinlichkeitsaussage für die Überdeckung des Medians durch ein Intervall und dazu unter der Annahme eines Modells, für deren Gültigkeit wir keine Wahrscheinlichkeit oder gar Sicherheit haben. Wahrscheinlichkeit im Modell ist aber nicht Wahrscheinlichkeit, dass mit dem Modell mit der gewählten Stichprobe der gesuchte Parameter überhaupt erfasst wird. Die beiden Begriffe sollte man nie verwechseln. Wahrscheinlichkeit ist risikobehaftet, Sicherheit dagegen nicht.
6. Weiter ordnet man ein Interaktionsproblem: Einer Zufallsauswahl (Stichprobe) wird ein Wahrscheinlichkeitsmodell verpasst, das für die Grundgesamtheit mutmasslich Gültigkeit hat. Die Grundgesamtheit ist unbekannt, das Modell jedoch ist bekannt. Andererseits ist die Stichprobe bekannt, das dort zu adaptierende Modell aber unsicher. Der Median der Stichprobe kann wesentlich verschieden sein vom Median der Grundgesamtheit. Es könnte sogar sein, dass die gesamte Stichprobe in Ω_u oder in Ω_o fällt. Die Wahrscheinlichkeit für diese Situation nimmt jedoch mit der Stichprobengrösse ab, allerdings nur, wenn die Grundgesamtheit nicht unendlich, uneingrenzbar oder unbeschränkt ist. Im andern Fall könnte die Wahrscheinlichkeit abschätzbar, wenn man $|\Omega|$ kennt.

7. Wählt man verschiedene Stichproben, so erhält man in der Regel bei jeder Stichprobe zu einer gegebenen Wahrscheinlichkeit ein anderes Vertrauensintervall. Das heisst mit anderer Lage und einer anderen Länge. Das Vertrauensintervall wird ja mit einem Zufallsintervall \mathcal{I} modelliert.
8. Kann man jedoch beurteilen, dass die gegebene **Stichprobe repräsentativ** für Ω ist, so ändert sich die angebrachte Kritik grundlegend zugunsten der gewählten Methode. Unter Umständen kann man dafür sorgen, dass man eine repräsentative Stichprobe erhält, etwa durch Auswahl aus verschiedenen Klassen.

Ausweitung der Methode für Quantile

Beim **Median** haben wir Ω in zwei Hälften geteilt:

$\Omega_u = \{x_k \in \Omega \mid x_k < Med\}$ und $\Omega_o = \{x_k \in \Omega \mid x_k > Med\}$. Damit gilt im gewählten Modell: $P(X < Med) = 0.5$, $P(X > Med) = 0.5$. Ebenso kann man bei **Quantilen** vorgehen: Wir betrachten statt $Med = q_{0.5}$ nun z.B. das **Quartil** $q_{0.25}$ oder allgemeiner das $\alpha - Quantil q_\alpha$.

Dann gilt: $\Omega_{u,\alpha} = \{x_k \in \Omega \mid x_k < \alpha\}$ und $\Omega_{o,\alpha} = \{x_k \in \Omega \mid x_k > \alpha\}$. Die Wahrscheinlichkeiten im Vergleich zum Median müssen nun korrigiert werden:

$P(X < q_\alpha) = \alpha$, $P(X > q_\alpha) = 1 - \alpha$. Sonst beilbt die Methode gleich wie beim Median beschrieben. Nicht zu vergessen ist die Annahme für die Wahrscheinlichkeitsfunktion: $P(X = q_\alpha) = 0$ (entsprechend wie beim Median $P(X = Med = q_{0.5}) = 0$.)

Das Problem mit der Methode beim Erwartungswert

Beim Erwartungswert $\mu = E(X)$ kann man $\Omega_u = \{x_k \in \Omega \mid x_k < \mu\}$ und $\Omega_o = \{x_k \in \Omega \mid x_k > \mu\}$ mit $P(x_k < \mu) = P(x_k > \mu)$ nicht bilden wie beim Median, denn in vielen Fällen ist μ wesentlich verschieden vom Median. Das arithmetische Mittel ist bekanntlich nicht robust und somit stark beeinflussbar durch grosse Ausreisser. Daher wäre die vorhin verwendete Verteilungsfunktion hier nicht sehr korrekt. Dazu kann man weiter Folgendes bemerken:

1. Wenn hingegen bekannt ist, dass X symmetrisch um den Median verteilt ist, so gilt $\mu = Med$. Daher ist hier das Vertrauensintervall für den Median auch für den Erwartungswert übertragbar.
2. Kann man bei μ approximativ von einer Normalverteilung ausgehen, so weiss man, dass \bar{X} ein Schätzer für μ ist und damit \bar{x} einer repräsentativen Stichprobe vom Umfang n ein Schätzwert wäre. Weiter weiss man, dass bei dieser Verteilung ca. 95.4% der Werte im Intervall $\mu \pm 2 \frac{\sigma}{\sqrt{n}}$ liegen. Damit wäre bei $P(\mathcal{I} \text{ überdeckt } \mu) = \gamma = 0.95$ das Vertrauensintervall annähernd durch

$$I = [\bar{x} - 2 \frac{\sigma}{\sqrt{n}}, \bar{x} + 2 \frac{\sigma}{\sqrt{n}}]$$

gegeben. Die Approximation wird, wie man um so besser sieht, je grösser n ist.

I.4.9 Nochmals zu statistischen Tests

Grundsätzlich können wir feststellen, dass der Sinn von **Vertrauensintervallen** meistens darin liegt, einen gegebenen Parameter in gewissen Schranken **einzuschliessen**.

Anders ist die Sachlage bei **statistischen Tests**: Oft möchte man hier gewisse Werte für einen Parameter **ausschliessen**.

Solche Tests funktionieren, wie wir früher gesehen haben, auf der Basis von **Nullhypothese** und **Alternativhypothese**. Z.B. sei die Nullhypothese: $\mu \in [a b]$, wobei $[a b]$ als Vertrauensintervall vorgeschlagen worden ist. Man **nimmt** hier **an**, dass die **Nullhypothese wahr** ist und **berechnet** dazu die **Wahrscheinlichkeit** P , dass unter dieser Annahme die **real vorgefundene Situation eintreffen** kann. Stellt man dann fest, dass P **unter** einer gewissen vorher festgelegten Wahrscheinlichkeitsschranke liegt (**Signifikanzniveau** α , $P < \alpha \ll 1$), so **verwirft** man die Nullhypothese. Das **Risiko** eines falschen

Entscheides in der Folge und einer **fälschlicherweise Verwerfung der Nullhypothese** ist hier durch α abgeschätzt. Folgende Punkte sind dabei wesentlich:

1. Die Definition des Testparameters Θ .
2. Die Formulierung der Nullhypothese H_0 .
3. Die Formulierung der Alternativhypothese H_1 resp. H_A zu H_0 . (Achtung: Hier ist formallogische und sachlogische Korrektheit unabdingbar!) Die Alternativhypothese definiert die **Richtung** des statistischen Tests.
4. Die Definition des Signifikanzniveaus unter der Annahme, dass H_0 wahr ist.
5. Die Ermittlung einer Wahrscheinlichkeitsfunktion (Testfunktion) für die Berechnung der Wahrscheinlichkeit einer realen Stichprobe unter der Bedingung, dass H_0 wahr ist.
6. Die Ermittlung einer Stichprobe.
7. Die Durchführung der numerischen Rechnung und Feststellung, ob der Test „ausschlägt“.
8. Die Entscheidungsphase, oft mit dem Ziel, die Nullhypothese zugunsten der Alternativhypothese zu verwerfen.
9. Schlägt der Test nicht aus, d.h. kann die Nullhypothese nicht verworfen werden, so kann man versuchen, den Fehler der 2. Art (siehe unten) zu berechnen.
10. Statt dem Fehler der 2. Art kann man auch versuchen, ein Vertrauensintervall für den Testparameter Θ zu bestimmen.

Beim Entscheid sind verschiedene Situationen und auch zwei Fehler möglich.

Zu den Möglichkeiten und den Fehlern die folgende Übersicht:

1. **Ablehnung** der Nullhypothese, obwohl sie **wahr** ist. (**Fehler 1. Art**, α -Fehler.)
 $\alpha = P(\text{Ablehnung der Nullhypothese, obwohl sie wahr ist})$.
2. **Keine Ablehnung** der Nullhypothese im Falle dass sie **wahr** ist (und die Alternativhypothese falsch ist). Anschliessend **Entscheid für die Nullhypothese** oder **kein Entscheid** bezüglich der betrachteten Hypothesen: Man bleibt im Ungewissen. (Was entschieden wird, hängt von weiteren Umständen ab. — Wenn man eine Hypothese nicht ablehnen kann, dann ist damit nicht immer automatisch ihre Gültigkeit erwiesen. Es kann auch sein, dass die Entscheidungsgrundlage noch fehlt.)
Dazu ein Bild: Wenn ein Beschuldigter vor Gericht mangels Beweisen nicht verurteilt werden kann, dann ist seine Unschuld noch nicht bewiesen. Er könnte trotzdem der Delinquent sein und man könnte es ihm bloss nicht nachweisen, bliebe also im Ungewissen... .
3. **Annahme** der Nullhypothese, obwohl sie **falsch** ist. (**Fehler 2. Art**, β -Fehler.)
 $\beta = P(\text{Annahme der Nullhypothese, obwohl sie falsch ist})$.
4. **Keine Annahme** der Nullhypothese im Falle, dass sie **falsch** ist (und die Nullhypothese wahr ist): Anschliessend **Entscheid für die Alternativhypothese** oder **kein Entscheid** bezüglich der betrachteten Hypothesen: Man bleibt je nach Fall im Ungewissen.

Die Berechnung von α bei bekannten Verteilungsfunktionen und insbesondere die Berechnung von β kann recht kompliziert sein. Hier sei auf die Spezialliteratur zu den Teststatistiken verwiesen. Vgl. dazu auch:

http://de.wikipedia.org/wiki/Statistischer_Test
<http://de.wikipedia.org/wiki/Irrtumswahrscheinlichkeit>

Ein **parameterfreier** oder **nicht-parametrischer** oder **verteilungsfreier Test** ist der im Hauptskript besprochene **Vorzeichentest**. Dazu sei auf das Hauptskript Statistik und die einschlägige Literatur verwiesen:

<http://rowicus.ch/Wir/Scripts/KursWahrschStatistd.pdf> (deutsch) oder
<http://rowicus.ch/Wir/Scripts/KursWahrschStatistdf.pdf> (deutsch-französisch) oder
<http://rowicus.ch/Wir/Scripts/KursWahrschStatistf.pdf> (französisch) oder dann

<http://de.wikipedia.org/wiki/Vorzeichentest> sowie
<http://de.wikipedia.org/wiki/Wilcoxon-Vorzeichen-Rang-Test>

Beispiel zur Situation mit den Fehlern:

Vor uns haben wir 1'000 produzierte Wellen. Jede dieser Wellen könnten Ausschuss oder auch in Ordnung, also gute Qualität sein. Durch lange und sehr exakte Evaluationen von Kontrollmethoden und deren Anwendung hat man festgestellt, dass durchschnittlich eine von 1'000 Wellen Ausschuss ist.

Bei der vorgesehenen automatischen Qualitätskontrolle wird der Druchmesser an 5 verschiedenen Orten an der Welle gemessen. Damit ist natürlich nicht die ganze Welle erfasst. Es besteht also ein Risiko.

Würde man aber die Qualitätskontrolle in Form von 50 Messungen pro jeweilige Welle durchführen, so könnte man das Risiko der falschen Erkennung erheblich senken.

Nun wird ein Test in Form der beschriebenen Art der Qualitätskontrolle eingerichtet. Die Irrtumswahrscheinlichkeit in beiden Richtungen, also Ausschuss oder Qualität falsch zu erkennen, ist 1%. 99 von 100 Wellen sollten daher mit dem Test richtig erkannt werden. Die Toleranz ist „1 von 100 falsch zu erkennen“. So wurde es mit dem Kunden verabredet. Das heisst konkret: Von 100 Ausschusswellen wird eine an den Kunden ausgeliefert, 99 aber werden erkannt und aus dem Lieferlos entfernt. Andererseits ist die Ausschussrate nach alter Erfahrung 1 Promill. D.h. zu ca. 100 getesteten und als Ausschuss erkannten Wellen werden wegen diesem Promill $100 \cdot 999 = 99'900$ Wellen produziert, welche qualitativ in Ordnung sind. Man muss somit die Betrachtung also mit einem Gesamttotal von ca. 100'000 Wellen anstellen. (Zu diesem Beispiel siehe auch Wikipedia, Irrtumswahrscheinlichkeit.) Nun werden aufgrund dieser Erkenntnis 100'000 Wellen näher untersucht.

Man erhält daher die folgende Daten (Übersichtstabelle):

******	Wellen tatsächlich Ausschuss	Wellen tatsächlich in Ordnung
Total 100'000	100	99'900
Als Ausschuss erkannt 1096 Entsorgt	98 richtig	998 Fehler 1. Art
Als in Ordnung erkannt 98'904 Ausgeliefert	2 Fehler 2. Art	98'902 richtig

Man beachte, dass in dieser Tabelle oben immer die **Spaltensummen** und links die **Zeilensummen** erscheinen. Eine solche Tabelle nennt man auch **Kreuztabelle** oder **Kontingenztafel**. Sie ist eine Matrix mit „Zeileneingängen“ und „Spalteneingängen“.

In der Tabelle sieht man, dass 98 von 1096 beim Test dafür gehaltenen Ausschusswellen auch wirklich Ausschusswellen waren. 998 waren keine Ausschusswellen und sind daher entsorgt worden, obwohl sie hätten verkauft werden können. Die Wahrscheinlichkeit, hier bei automatischer Registrierung von Ausschuss mittels Test auch tatsächlich Ausschuss zu sein, ist nur $\frac{98}{1096} \hat{=} 8.94161\%$

Und man sieht auch, dass 98'902 von 98'904 für qualitativ in Ordnung befundenen Wellen auch in

Wirklichkeit in Ordnung waren und man sie ausgeliefert hatte. 98'904 sind total ausgeliefert worden. Die Wahrscheinlichkeit, hier bei automatischer Registrierung von guter Qualität auch tatsächlich gute Qualität zu sein, ist sehr gross: $\frac{98'902}{98'904} \hat{=} 99.998\%$

Die totale Irrtumswahrscheinlichkeit ist $\frac{998 + 2}{100'000} = \frac{1'000}{100'000} \hat{=} 1.0\%$, wie auf der Grundlage des Tests erwartet.

Folgerung: Der Produzent ist jetzt der Geschädigte. Der Konsument oder Kunde hat bei dieser Situation einen im Vergleich zum Produzenten nur kleinen Schaden erfahren.

Die **Problematik** liegt darin, dass der Irrtum zwischen Produzent und Konsument hier zu sehr ungleichen Stückzahlteilen aufgeteilt wird. Man hat ein grosses vorliegendes **Produzentenrisiko** und ein nur kleines vorliegendes **Konsumentenrisiko**. Man staunt vermutlich über diese Situation. Die berechneten Risikowahrscheinlichkeiten sind ohne die Berechnung nicht leicht zu erraten. Das liegt daran, dass es sich da um **bedingte Wahrscheinlichkeiten** handelt. Generell kann man bemerken, dass man in **Kreuztabellen** bedingte Wahrscheinlichkeiten ausmachen kann.

I.4.10 Bemerkung zu Darstellungsmethoden der explorativen Statistik

Dazu eine nicht abschliessende Aufzählung:

1. Oben sind die recht nützlichen **Kreuztabellen** erwähnt worden. Man nennt sie auch **Kontingenztafeln**. Siehe dazu auch:
<http://de.wikipedia.org/wiki/Kreuztabelle>
2. **Stamm–Blatt–Diagramme**. Siehe dazu:
<http://de.wikipedia.org/wiki/Stamm–Blatt–Diagramm> sowie auf dieser Seite unten
http://de.wikibooks.org/wiki/Mathematik:_Statistik unter deskriptiver Statistik,
 Häufigkeitsverteilungen.
3. Weitere **Diagramme**, speziell **Strichlisten**, siehe
<http://rowicus.ch/Wir/Scripts/KursWahrschStatistd.pdf> oder
<http://rowicus.ch/Wir/Scripts/KursWahrschStatistdf.pdf> oder
<http://rowicus.ch/Wir/Scripts/KursWahrschStatistf.pdf>
4. **Modus** und **Verteilungstypen** siehe auch unter
[http://de.wikipedia.org/wiki/Modus_\(Statistik\)](http://de.wikipedia.org/wiki/Modus_(Statistik))
<http://de.wikipedia.org/wiki/Bimodal>
5. **Streudiagramme** siehe auch unter
<http://de.wikipedia.org/wiki/Streudiagramm>
[http://de.wikipedia.org/wiki/Trend_\(Statistik\)](http://de.wikipedia.org/wiki/Trend_(Statistik))
6. Zu **empirischen Verteilungsfunktionen** siehe unter
http://de.wikipedia.org/wiki/Empirische_Verteilungsfunktion oder
<http://de.wikipedia.org/wiki/Histogramm>

Für empirische Verteilungsfunktionen (ableitbar auch aus den zugehörigen Balkendiagrammen oder Histogrammen) gilt der **Satz von Gliwenko–Cantelli**:

Satz:

Vor.:

Seien X_1, \dots, X_n identisch verteilte Zufallsvariablen mit der Verteilungsfunktion $F(x)$. Die X_k kann man als unabhängige Kopien einer Zufallsvariablen X auffassen. ($\text{Sup} = \text{Supremum}$.)

$$\hat{F}_n(x) := \frac{1}{n} \cdot |\{X_k \leq x \mid 0 \leq k \leq n\}|, \quad d_n := \text{Sup}_x (\hat{F}_n(x) - F_n(x))$$

Beh.:

$$P(\lim_{n \rightarrow \infty} d_n = 0) = 1$$

Hier handelt es sich also um eine Aussage über die **Annäherung einer empirischen Verteilungsfunktion einer Stichprobe an die Verteilungsfunktion der Grundgesamtheit**. Wir bemerkten, dass die Annäherung für $n \rightarrow \infty$ gilt.

Speziell ist $|(x_1, \dots, x_n)| = 6, 7, \dots, 10, \dots, 100, \dots, 100'000 \dots \neq \infty$.

Siehe auch unter

<http://de.wikipedia.org/wiki/Gliwenko-Cantelli-Satz>

Im Diagramm der empirischen Verteilungsfunktionen kann man die Anzahl derjenigen Stichprobenwerte ablesen, welche durch einen bestimmten betrachteten Wert begrenzt sind.

Weiter sieht man sofort: Je steiler der **Graph der empirischen Verteilungsfunktion verläuft**, desto **weniger streuen** dort die Werte.

7. Das Problem der **Ausreisser** und der **aussergewöhnlichen Werte**:

Unter diesem Titel fasst man zwei Sorten von Werten oder Beobachtungsausprägungen zusammen:

- (a) **Logisch unmögliche** Ausprägungen wie z.B. eine Anzahl „schwangerer Männer“, welche grösser als null ist. Solche Daten müssen nach allen möglichen erfolglosen Abklärungen aus der Datenbank entfernt werden, insofern man sich damit nur lächerlich macht.
- (b) Von der Erfahrung her **seltsame, seltene**, aber **nicht unmögliche** Ausprägungen wie z.B. eine Anzahl „Menschen mit einer Körperhöhe grösser als 2.5 Meter“. Solche Daten sind selten, aber können durchaus die Realität wiedergeben. Sie können nicht einfach aus der Datenbank entfernt werden.

Ausreisser werden in der **robusten Statistik** behandelt. Über ihre Definition herrscht noch keine überall akzeptierte Einigkeit.

Sei \hat{d} die **empirische Quartilsdifferenz**: $\hat{d} = \hat{q}_{0.75} - \hat{q}_{0.25}$, ermittelt aus einer Datenerhebung. Dann nennt man oft einen Wert x einen **Ausreisser**, wenn gilt:

$$x \in (\hat{q}_{0.25} - 3\hat{d}, \hat{q}_{0.25} + 1.5\hat{d}) \quad \text{oder} \quad x \in (\hat{q}_{0.75} + 1.5\hat{d}, \hat{q}_{0.75} + 3\hat{d})$$

x heisst aber **Extremwert** oder **extremer Ausreisser**, wenn eine noch grössere Abweichung vorhanden ist:

$$x \leq \hat{q}_{0.25} - 3\hat{d} \quad \text{oder} \quad x \geq \hat{q}_{0.75} + 3\hat{d}$$

8. Eine rasche Übersicht über die Lage und Streuung eines Datensatzes kann man sich bekanntlich mit Hilfe der **Box-Wkisker-Plots** oder **Boxplots** machen. Darin kann man den **Median** $q_{0.5}$, die **Quartile** $q_{0.25}$, $q_{0.75}$ und das **Minimum** $q_{0.0}$ sowie das **Maximum** $q_{1.0}$ eines Datensatzes ablesen. Oft wird in solchen Plots auch noch ein **approximatives Vertrauensintervall** für den Median dargestellt. Man kann somit die Überdeckung von Datensätzen und damit ihre Verschiedenheit beurteilen.
 Siehe im Hauptskript zur Statistik oder auch in
<http://de.wikipedia.org/wiki/Boxplot> oder
<http://de.wikipedia.org/wiki/Box-Whisker-Plot>
9. Bemerkung zu **Stabdiagrammen** und **Histogrammen**: Diese beiden Begriffe werden heute noch nicht überall im gleichen Sinne verwendet. Oft benutzt man jedoch den Begriff Stabdiagramm zur Darstellung **diskreter Daten**, etwa analog einer Strichliste. Dagegen findet der Begriff Histogramm oft bei **stetigen Daten** Anwendung. Die Unterscheidung ist aber auch oft **obsolet**, da eine Maschine den diskreten Daten ja nicht unbedingt ansieht, ob es sich nicht um spezielle stetige Daten handelt. Bei Histogrammen wählt man vielfach die Flächeninhalte der Balken als Masseinheit für die Anzahl Beobachtung, wobei die Breite variieren kann.

10. Der Fantasie bei Darstellungen sind oft keine Grenzen gesetzt. Siehe dazu:

<http://de.wikipedia.org/wiki/Diagramm>
 oder etwa in Google ~ Bildsuche ~ Statistik Diagramm eingeben.

11. **Stochastik** — Siehe dazu unter:

<http://de.wikipedia.org/wiki/Stochastik>

I.5 Nochmals zu den Bootstrap–Methoden

Wie bekannt ist, kann man mit Computersimulationen z.B. das Problem der approximativen Schätzung von Vertrauensintervallen bei unbekannten Verteilungsfunktionen von Parametern einer Messgrösse umgehen. Man denke etwa an μ oder an σ . Man umfährt damit das Problem der Modellwahl für die Verteilungsfunktionen.

I.5.1 Zum Percentil–Lemma

Im besagten Lemma geht der Begriff der pivotalen Grösse ein. Um was handelt es sich dabei?
 Eine **pivotale Grösse** oder eine **pivotale Zufallsfunktion** ist eine Zufallsfunktion, die nicht von „unbekannten“ Parametern abhängt (wobei andererseits das Unbekannte nicht beurteilbar ist).

Siehe dazu auch: http://en.wikipedia.org/wiki/Pivotal_quantity

Sei nun $\hat{\Theta}$ eine Schätz–Statistik für Θ , wobei der Parameter Θ zu irgend einer gegebenen Messgrösse gehöre. $\hat{\Theta} - \Theta$ ist die Abweichung zwischen der Schätz–Statistik aus einer Stichprobe und dem Wert Θ der Grundgesamtheit.

Zu $\gamma \in [0, 1]$ bilden wir $\gamma_1 = \frac{1-\gamma}{2}$ und $\gamma_2 = \frac{1+\gamma}{2}$. Dann gilt ebenfalls: $\gamma_1, \gamma_2 \in [0, 1]$. Seien nun $q_1 = q_{\gamma_1}$ und $q_2 = q_{\gamma_2}$ die q_{γ_1} – bzw. q_{γ_2} –Quantile zu $\hat{\Theta}$.

Satz: **Percentil–Lemma**

Vor.:

Sei $\hat{\Theta} - \Theta$ pivotal und symmetrisch verteilt um den Ursprung.

Beh.:

$I = [q_{\gamma_1}, q_{\gamma_2}]$ ist ein Vertrauensintervall zu Θ zum Niveau γ

(**Percentil–Vertrauensintervall**).

In Fällen, wo $\hat{\Theta} - \Theta$ unsymmetrisch um den Ursprung verteilt und nicht pivotal ist, kann man eine **Transformation** finden, welche zu einer symmetrischen Verteilung und einer pivotaler Form führt. Nach der Transformation kann man das Percentil–Lemma anwenden und nachher wieder zurück transformieren.

Wir ziehen nun eine Stichprobe aus einer im Detail unbekannten Grundgesamtheit und bestimmen damit einen **Schätzer** $\hat{\Theta}$ eines unbekannten Parameters Θ . Mit Hilfe der **Bootstrap–Methode** kann man dann durch Ziehung von Bootstrap–Kopien aus der Stichprobe eine **empirische Verteilungsfunktion** $\hat{F}(x)$ einer dadurch erzeugten empirischen Grundgesamtheit $\hat{\Omega}$ gewinnen. An dieser empirischen Verteilungsfunktion können wir das Vertrauensintervall zu einem gegebenen Niveau γ ablesen. Zur Ablesung vgl. Beschreibung Seite 62.

Wenn wir „unendlich“ lang simulieren (wenn wir dies könnten), so reden wir von **idealem Bootstrap**. Idealer Bootstrap ist also eine hypothetische Sache in der Form eines unerreichbaren Grenzfalls. Für die Beurteilung der Genauigkeit eines Bootstrap–Percentil–Vertrauensintervalls kann der folgende Satz hilfreich sein:

Satz:

Sei n der Stichprobenumfang und C eine Konstante.

Bei idealem Bootstrap gilt dann für die Genauigkeit des Bootstrap–Percentil–Vertrauensintervalls für einen Parameter Θ zum Niveau γ :

$$P(I_{\text{Bootstrap}} \text{ überdeckt } \Theta) = \gamma \pm \frac{C}{n}$$

Folgerung:

Für $n \rightarrow \infty$ wird $P(I_{\text{Bootstrap}} \text{ überdeckt } \Theta) = \gamma$

An Stelle einer Methodenkritik wollen wir hier zur Kenntnis nehmen, was unter anderem in der Praxis als Bedingungen und Empfehlungen „gehandelt“ werden:

1. Die Bootstrap–Metode kann verwendet werden zur Bestimmung von Lageparametern, Streuparametern und auch Wahrscheinlichkeitsparametern wie z.B. Verhältnisse von Wahrscheinlichkeiten.
2. Die Stichprobenwerte sollen unabhängig sein.
3. Für den Stichprobenumfang n soll gelten: $n \geq 6$.
4. Der Graph der simulierten Verteilungsfunktion soll keine hohen Treppenstufen mehr aufweisen.
5. Man soll solange simulieren, bis sich die Form des Graphen praktisch nicht mehr ändert.

Ein nur intern zugänglicher Link zu weiterer Literatur zur Bootstrap-Methode (auch **Plug-in-Methode** genannt) findet man auf:

<http://rowicus.ch/Wir/Scripts/restricted/MasterIndex.html> unter (1).

Und weiter öffentlich:

[http://en.wikipedia.org/wiki/Bootstrapping_\(statistics\)](http://en.wikipedia.org/wiki/Bootstrapping_(statistics))
[http://de.wikipedia.org/wiki/Bootstrapping_\(Statistik\)](http://de.wikipedia.org/wiki/Bootstrapping_(Statistik))

Interessant sind die folgenden Gebiete: Punktschätzung, Konfidenzintervalle, Tests, Permutationstests und Modelle.

Anhang J

Zusammenfassungen und Ausblicke

J.1 Parametertests — Signifikanztests

J.1.1 Das Problem des Lernenden

Unter Signifikanztests (Parametertests) versteht man meist sehr komplexe Methoden, welche dazu dienen, die Wahrscheinlichkeit oder Unwahrscheinlichkeit einer Annahme (Hypothese) oder Behauptung zu prüfen und damit das Vertrauen in die Hypothese oder Behauptung zu erhärten oder zu erweisen. In der Literatur werden solche Tests andererseits infolge ihrer Komplexität oft eher verwirrend als erklärend dargestellt, da zu einer klaren Kommunikation ein strenges logisches Gerüst in der Argumentation und im Begriffsaufbau wie auch ein profundes Verständnis des Lehrenden notwendig ist, der sich nicht leisten sollte, die Menge an Details so wiederzugeben, wie sie zufällig bei ihm aus der Erinnerung auftauchen. Denn dadurch entsteht selten Klarheit und kostet dann dem Zuhörer enorm Zeit zur Entwirrung, wenn er nicht schon vorher das Vertrauen in die Konstellation verloren und aufgegeben hat. Man kann sich ja oft auch mit Hilfe von Auswendiglernen ohne viel verstehen zu müssen durchwursteln.

J.1.2 Das Denkgerüst hinter einem Test

Um was geht es nun genauer bei der erwähnten Angelegenheit der Signifikanztests? — Kurz kann man es so sagen: Man möchte eine Wahrscheinlichkeitsaussage über die vermutete Grösse oder die Lage eines Parameters gewinnen. Z.B. stellen wir uns als Parameter den Mittelwert einer verteilten Grösse einer Grundgesamtheit $\mu \in [a, b]$ vor. Denken wir dabei an einen Durchmesser, etwa von Schrauben. Statt μ messen wir dazu die Mittelwerte \bar{x}_k von Stichproben mit dem Index k .

Zur Gewinnung der Aussage gehen wir wie folgt vor:

1. Zuerst formuliert man die vorhandene Vermutung über den Parameter als „Nullhypothese“ H_0 . Dazu muss man weiter eine vom studierten Parameter (z.B. gegeben durch die \bar{x}_k) abhängige Prüfgrösse (Variable Z) definieren, welche man dann einfach mathematisch behandeln kann. Bedenken wir z.B. dass statt mit dem Mittelwert \bar{x}_k einer Stichprobe, für den wir die Variable \bar{X} verwenden, einfacher mit dem transformierten Parameter $Z = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$ gearbeitet werden kann (Transformation auf Standardverhältnisse, n = Stichprobenumfang, S = Variable für die Streuung). Die Nullhypothese könnte hier z.B. in der Form $z = z_0$ (= fixer Wert z_0) formuliert sein. Dieser Hypothese kommt dann eine gewisse Wahrscheinlichkeit p zu, denn es handelt sich ja nicht um eine Hypothese und nicht um eine gesicherte Erkenntnis. Die Realisierungen oder Werte z der Variablen Z durch die Stichproben folgen somit dem Verhalten einer Wahrscheinlichkeitsfunktion, welche erst noch gefunden werden muss. Schliesslich sind zuvor die Stichprobenwerte auch nach einer gewissen Wahrscheinlichkeitsfunktion verteilt. Und ebenso die daraus gerechneten Mittelwerte der einzelnen Stichproben.

2. Zur Nullhypothese H_0 gibt es nun die Alternativhypothese $H_1 = \neg H_0$ („nicht H_0 “) mit der Wahrscheinlichkeit $q = 1 - p$. Hier trifft man das Problem, dass vielen Menschen die logische Negation einer Sache nicht sehr geläufig ist. Die Negation von „jetzt regnet es“ ist nicht etwa „jetzt ist schönes Wetter“, sondern „jetzt regnet es nicht“. Wenn für die Nullhypothese H_0 eine grosse Wahrscheinlichkeit erwartet wird, so wird für folglich die Alternativhypothese H_1 eine kleine Wahrscheinlichkeit erwartet.
3. Eine Teststrategie ist nun die Folgende: Man versucht zu zeigen, dass die Realisierungen der Alternativhypothese H_1 in den Belegen mittels Stichproben nur mit sehr, sehr kleiner Wahrscheinlichkeit vorkommen. Daraus kann man dann schliessen, dass nicht auch die Nullhypothese H_0 nur mit einer sehr kleinen Wahrscheinlichkeit (oft α , Irrtumswahrscheinlichkeit genannt) vorhanden sein kann und daher die Wahrscheinlichkeit von H_0 grösser ist, so dass man H_0 also nicht ablehnen oder verworfen kann. Da H_0 dann also nicht verworfen werden kann, muss diese Hypothese stehen gelassen werden.
4. Eine andere Teststrategie wäre es, die Nullhypothese H_0 durch das vorhandene Datenmaterial erwiesenermassen als sehr unwahrscheinlich erscheinen zu lassen. Damit müsste also dann der Alternativhypothese H_1 eine grössere Wahrscheinlichkeit zukommen, womit man die Nullhypothese nicht mit gutem Glauben annehmen könnte usw.
5. Das Problem bei solchen Tests ist allemal die Konstruktion der für den vorhandenen Parameter Z am ehesten passenden Wahrscheinlichkeitsfunktion, weil man oft die zugrunde liegende Verteilung überhaupt nicht kennt und sich daher auf Annahmen stützen muss. Gefürchtet sind dabei Dinge wie die Einflüsse unbekannter Parameter usw. Daher zieht man manchmal parameterfreie Verfahren vor wie Bootstrapping-Simulationen usw. vor.

J.2 Parameterschätzung und Maximum–Likelihood–Methode

Siehe auch <http://de.wikipedia.org/wiki/Maximum-Likelihood-Methode>

J.2.1 Die Idee

Bei der Maximum–Likelihood–Methode handelt es sich um ein Verfahren zur Bestimmung von Schätzfunktionen, mit welchem optimale Werte von Parametern oder Punktschätzern geschätzt werden können.

Gegeben sei eine Stichprobe vom Umfang n mit einer Nummer k . Darin seien die Werte $S_k = \{x_{k,1}, x_{k,2}, \dots, x_{k,n}\}$, kurz $S = \{x_1, x_2, \dots, x_n\}$. Statt bei S von n -fachen Realisationen oder Belegungen einer Variablen X zu sprechen, kann man bekanntlich auch von n einfachen Realisationen oder Belegungen von n identisch verteilten Variablen X_1, X_2, \dots, X_n ausgehen. Sei nun Y ein zu bestimmender Parameter (man denke z.B. an den Mittelwert der n Werte der Stichprobe). Zu solchen Parametern Y gehört, wie aus der Verschiedenheit von Stichproben hervorgeht, je eine Verteilungsfunktion, welche man immer, wie nachstehend am Beispiel gezeigt, zu einer Funktion mit mehreren Variablen zusammenfassen kann. Sei nun $f(X, Y)$ zu S resp. $f(X_k, Y)$ zu S_k resp. $f(X_1, X_2, \dots, X_n, Y)$ (zu S_k) mit der Realisierung $f(x_1, x_2, \dots, x_n, y)$ die jeweils zugehörige zusammengefasste Verteilungsfunktion. $Y = y$ soll nun optimal geschätzt werden. Das Resultat bezeichnen wir dann mit \hat{y} .

Trick: Wir definieren zu x_1, x_2, \dots, x_n, y die zugehörige Maximum–Likelihood–Funktion $L(y)$ resp. $L(Y = y)$ resp. $L(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, Y = y)$ wie folgt:

$$L(y) = L(Y = y) := f(X = X_1 = x_1, y) \cdot f(X = X_2 = x_2, y) \cdot \dots \cdot f(X = X_n = x_n, y),$$

$$\text{kurz: } L(y) := f(x_1, y) \cdot f(x_2, y) \cdot \dots \cdot f(x_n, y).$$

Dabei bedeutet $f(X = X_1 = x_1, y) = f(x_1, y)$ die Wahrscheinlichkeit für das Eintreffen von x_1 und y usw., zusammen $L(y)$ also eine durch L definierte Wahrscheinlichkeit für das Eintreffen von x_1, x_2, \dots, x_n, y . Der Schätzwert \hat{y} soll nun die Eigenschaft haben, dass für $y = \hat{y}$ die Wahrscheinlichkeit $L(\hat{y})$ maximal gross ist. Daher gilt es, das Maximum der $L(\hat{y})$ von $L(y)$ zu suchen. Wenn die Faktoren $f(X = X_k = x_k, y) = f(x_k, y)$ differenzierbar sind, kann man zur Auffindung von \hat{y} die notwendige Maximierungsbedingung $\frac{dL(y)}{dy} = 0$ verwenden. \leadsto Darin besteht der Trick.

Oft erreicht man eine Vereinfachung durch eine Transformation wie folgt: Statt $L(y)$ zu maximieren, maximiert man $L_{ln}(y) := \ln(L(y))$. Das ist problemlos, weil der \ln auf dem Wertebereich $(0, 1)$ definiert ist und dort streng monoton wächst. Man hat dann die Maximierungsbedingung $\frac{dL_{ln}(y)}{dy} = 0$.

Analog kann man argumentieren, wenn man mehr als einen Parameter maximieren resp. optimieren will. Im Falle von y und z sei z.B.

$$L(y, z) := f(x_1, y, z) \cdot f(x_2, y, z) \cdot \dots \cdot f(x_n, y, z)$$

Notwendig für ein Maximum ist dann $\frac{\partial L(y, z)}{\partial y} = \frac{\partial L(y, z)}{\partial z} = 0$ resp. $\frac{\partial L_{ln}(y, z)}{\partial y} = \frac{\partial L_{ln}(y, z)}{\partial z} = 0$

J.2.2 Ein Beispiel zur Veranschaulichung

Durch $f(x, \mu) = \frac{\mu^x}{x!} e^{-\mu}$ ist eine Poisson-Verteilung gegeben. An dieser lässt sich das Maximum-Likelihood-Verfahren besonders schön illustrieren. Es soll daher hier der Parameter μ optimal geschätzt und damit $\hat{\mu}$ bestimmt werden. Zuerst definieren wir L :

$$\leadsto L(\mu) := f(x_1, y) \cdot \dots \cdot f(x_n, y) = \frac{\mu^{x_1}}{x_1!} e^{-\mu} \cdot \dots \cdot \frac{\mu^{x_n}}{x_n!} e^{-\mu} = \frac{\mu^{x_1+x_2+\dots+x_n}}{x_1! \cdot x_2! \cdot \dots \cdot x_n!} e^{-\mu \cdot n}$$

Verwendet man hier $L_{ln}(\mu)$ statt $L(\mu)$, so wird das Differenzieren danach einfacher: $L_{ln}(\mu) = \ln(f(x_1, y) \cdot \dots \cdot f(x_n, y)) = \ln\left(\frac{\mu^{x_1}}{x_1!} e^{-\mu} \cdot \dots \cdot \frac{\mu^{x_n}}{x_n!} e^{-\mu}\right) = \ln(\mu^{x_1+x_2+\dots+x_n}) - \ln(x_1! \cdot x_2! \cdot \dots \cdot x_n!) + \ln(e^{-\mu \cdot n}) = (x_1 + x_2 + \dots + x_n) \cdot \ln(\mu) - \ln(x_1! \cdot x_2! \cdot \dots \cdot x_n!) - \mu \cdot n$ \leadsto einfach zu differenzieren!

$$\leadsto \frac{dL(\mu)}{d\mu} = (x_1 + x_2 + \dots + x_n) \cdot \frac{1}{\mu} - 0 - n = 0 \Rightarrow \mu = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{x}, \text{ wie man für } \mu = \hat{\mu} \text{ erwartet.}$$

$$\leadsto \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = f(X_1, X_2, \dots, X_n) \text{ ist die optimale Schätzfunktion für } \hat{\mu}.$$

J.2.3 Erwartungstreue, Konsistenz, Effizienz

Wichtig für eine Schätzfunktion f sind die früher schon erwähnten Begriffe Erwartungstreue, Konsistenz und Effizienz oder Wirksamkeit. Hier eine kurze Erinnerung:

1. Wir nennen einen Parameter Y **erwartungstreu** für die Schätzfunktion f , wenn für den Erwartungswert E gilt: $E(f(X_1, X_2, \dots, X_n)) = Y$
2. Wir nennen eine Schätzfunktion f **konsistent**, wenn ihre Varianz mit zunehmendem Stichprobenumfang abnimmt.
3. Wir nennen eine Schätzfunktion f im Vergleich zu anderen konkurrenzierenden Schätzfunktionen **effizient**, wenn sie im Vergleich mit den anderen die kleinste Varianz besitzt.

Eine **optimale Schätzfunktion** soll die genannten Eigenschaften erwartungstreu, konsistent und effizient aufweisen, damit man von einer seriösen Schätzung reden darf.

J.2.4 Wie weiter?

Falls sich eine erdrückende Notwendigkeit ergibt, wird diese Sammlung bei Gelegenheit ergänzt — der Autor bittet um Geduld.

Ein anderer weiterführender Weg findet man im Weiterstudium mittels eintauchen in die überaus reich vorhandene Literatur. Dabei stellt sich das Problem, dass weiterführende Literatur immer dem momentan vorhandenen Niveau angepasst sein soll, um Frust zu vermeiden. Gerade in der Statistik existiert überaus viel Literatur auf niedrigstem, ja geradezu banalem Niveau, denn fast jedermann schlägt sich mit Statistiken herum. Dieses Niveau kann nun hier nicht mehr befriedigen. Ebenso existiert viel Literatur auf höchstem Niveau, denn manche möchten ja zeigen, dass sie an der „Forschungsfront“ tätig sind. Von dieser „Front“ sie hier abgeraten, denn der notwenige Sprung nach ganz oben ist von hier aus zu gross. Somit entsteht die Schwierigkeit, „etwa in der Mitte zu fischen“, da wo der Anstieg der Anforderungen sehr steil ist. Am besten fährt man hier, wenn man jene nach geeigneter Literatur fragt, die gerade erst denselben Schritt getan haben, welchen man jetzt tun möchte. Daher ist es immer sehr ratsam, gute Beziehungen zu den „oberen Semestern“ zu pflegen. Dort findet man jene, welche einem am effizientesten weiterhelfen können sollten, falls man nicht alleine ganze Fachbibliotheken durchwühlen möchte.

J.3 Ausblick: Weitere Links zu Stochastik und Statistik

Nicht behandelt wurden hier die folgenden interessanten und weiterführenden Themen, da diese den hier gesetzten Rahmen sprengen:

1. Umfassende klassische Tests
2. Multivariate Statistik
3. Varianzanalyse
4. Regressionsanalyse
5. Zeitreihenanalyse
6. Prozesse
7. u.s.w.

Man konsultiere dazu die Fachliteratur. Erste Informationen findet man unter nachfolgenden Links, unter denen man weitere Links zu hier nicht geannnten, jedoch statistisch relevanten Themen findet. Man vergesse bei der Anwendung die genaue Abklärung der Gültigkeit der Voraussetzungen nicht. Sonst kann man keinen Erfolg erwarten.

<http://de.wikipedia.org/wiki/Statistik>

http://de.wikipedia.org/wiki/Multivariate_Statistik

<http://de.wikipedia.org/wiki/Regressionsanalyse>

<http://de.wikipedia.org/wiki/Zeitreihen>

<http://de.wikipedia.org/wiki/Varianzanalyse>

Bemerkung: Varianzanalyse, engl. *Analysis of variance* \leadsto **ANOVA**. Dazu gehören auch:

<http://de.wikipedia.org/wiki/F-Verteilung>

<http://de.wikipedia.org/wiki/F-Test>

<http://de.wikipedia.org/wiki/Diskriminanzanalyse>

<http://de.wikipedia.org/wiki/Kontingenzanalyse>

http://de.wikipedia.org/wiki/Logistische_Regression

<http://de.wikipedia.org/wiki/Conjoint-Analyse>

<http://de.wikipedia.org/wiki/Faktorenanalyse>

<http://de.wikipedia.org/wiki/Clusteranalyse>

http://de.wikipedia.org/wiki/Multidimensionale_Skalierung

<http://de.wikipedia.org/wiki/Korrespondenzanalyse>

<http://de.wikipedia.org/wiki/Korrespondenzanalyse>

http://de.wikipedia.org/wiki/Neuronale_Netze

<http://de.wikipedia.org/wiki/Hauptkomponentenanalyse>

http://de.wikipedia.org/wiki/Statistische_Prozesslenkung

<http://de.wikipedia.org/wiki/Qualitaetskontrolle>

\leadsto Ersetze nötigenfalls „ae“ durch „ä“

<http://de.wikipedia.org/wiki/Stochastik>

http://de.wikipedia.org/wiki/Stochastischer_Prozess

<http://de.wikipedia.org/wiki/Markow-Kette>

<http://de.wikipedia.org/wiki/Wahrscheinlichkeitstheorie>

<http://de.wikipedia.org/wiki/Wahrscheinlichkeitstheorie>

http://de.wikipedia.org/wiki/Statistisches_Schaetzverfahren

\leadsto Ersetze nötigenfalls „ae“ durch „ä“

Ende