

Wahrscheinlichkeitstheorie und Statistik

Michael Oberguggenberger

Skriptum zur Vorlesung
von Robert Eberle und Thomas Fetz
über Wahrscheinlichkeitstheorie und Statistik
im Sommersemester 2022

Bachelorstudien Mechatronik und Elektrotechnik

Arbeitsbereich für Technische Mathematik
Institut für Grundlagen der Technischen Wissenschaften
Universität Innsbruck

Inhaltsverzeichnis

1	Einführung	1
1.1	Die Rolle der Wahrscheinlichkeitstheorie und Statistik in den Technischen Wissenschaften	1
1.2	Grundsätzliches zur Methodik	9
1.3	Was ist Wahrscheinlichkeit?	10
2	Beschreibende Statistik eindimensionaler Daten	13
2.1	Datenerhebung und Darstellung	13
2.2	Datenbeschreibung: Lageparameter	16
2.3	Datenbeschreibung: Varianz	21
3	Beschreibende Statistik mehrdimensionaler Daten	25
3.1	Gemeinsame und Randhäufigkeiten	25
3.2	Kovarianz und Korrelation	27
3.3	Rangkorrelation	31
4	Eindimensionale Zufallsgrößen	33
4.1	Der mathematische Wahrscheinlichkeitsbegriff	33
4.2	Axiome der Wahrscheinlichkeit	34
4.3	Beispiele für diskrete Zufallsgrößen	35
4.4	Beispiele für kontinuierliche Zufallsgrößen	39
4.5	Wichtige Kennzahlen von Zufallsgrößen	43

4.6	Die Verteilungsfunktion einer Zufallsgröße	47
4.7	Funktionen von Zufallsgrößen	48
4.8	Anhang: Verteilungen und ihre Anwendungen	52
5	Mehrdimensionale Zufallsgrößen	56
6	Stichprobentheorie	62
6.1	Summen von Zufallsgrößen	62
6.2	Die mathematische Stichprobe	63
6.3	Die Gesetze der großen Zahlen	64
6.4	Der zentrale Grenzwertsatz	66
6.5	Monte-Carlo-Simulation	68
7	Schätzen von Parametern	72
7.1	Punktschätzungen	73
7.2	Konfidenzschätzungen	76
7.3	Hochwasserstatistik	78
7.3.1	Extremwertverteilungen	79
7.3.2	Hochwasserprognose	80
7.3.3	Konfidenzgrenzen für die Quantile	82
8	Statistische Testverfahren	87
8.1	Test für den Mittelwert μ einer Zufallsgröße	87
8.2	Der χ^2 -Anpassungstest	91
8.3	Die drei wichtigsten Testverteilungen	94
9	Stochastische Bemessung	98
10	Bayes'sche Konzepte	110
10.1	Bedingte Wahrscheinlichkeiten	110

10.2 Bayes'sche Parameterschätzung	111
10.3 Subjektive Wahrscheinlichkeiten	115
10.4 Ereignisbäume	118
11 Lineare Regression	123
11.1 Univariate lineare Regression	123
11.2 Rudimente der ANOVA	127
11.3 Multivariate lineare Regression	128
11.4 Statistik der Regression	130

Inhaltsangabe – Vorschau

1. Einführung

Wahrscheinlichkeitstheorie und Statistik in den Technischen Wissenschaften.

2. Beschreibende Statistik eindimensionaler Daten

Grundgesamtheit, Stichprobe, relative Häufigkeit, Histogramm, Mittelwert, Median, Quantile, Spannweite, Viertelweite, Ausreißer, Boxplot, Stichprobenvarianz, Summenhäufigkeit.

3. Beschreibende Statistik zweidimensionaler Daten

Gemeinsame Verteilung, Randverteilung, Unabhängigkeit, Kovarianz, Korrelation.

4. Eindimensionale Zufallsgrößen

Diskrete/kontinuierliche Zufallsgrößen, Wahrscheinlichkeit, Dichte, Verteilungsfunktion, Erwartungswert, Varianz; Alternativ-, Binomial-, Poisson-, Normal-, Beta-, Exponential-, Lognormalverteilung. Funktionen von Zufallsgrößen, Zentrierung der Normalverteilung.

5. Mehrdimensionale Zufallsgrößen

Gemeinsame Verteilung, Randverteilung, Kovarianz, Kovarianzmatrix, Korrelationskoeffizient; zweidimensionale Normalverteilung; Unabhängigkeit und Unkorreliertheit.

6. Summen von Zufallsgrößen

Erwartungswert, Varianz und Verteilungsdichte der Summe unabhängiger Zufallsgrößen; die mathematische Stichprobe, Erwartungswert und Varianz des Stichprobenmittels, die Gesetze der großen Zahlen, der zentrale Grenzwertsatz und seine Anwendung.

7. Schätzen von Parametern

Punktschätzung, Konfidenzschätzung, Konfidenzintervall für den Erwartungswert normalverteilter Größen bei bekannter und bei unbekannter Varianz.

8. Statistische Testverfahren

Null-, Alternativhypothese, Signifikanzniveau, Irrtumswahrscheinlichkeit, Test für den Erwartungswert einer normalverteilten Größe, χ^2 -Anpassungstest, die χ^2 -, t - und F -Verteilung.

9. Stochastische Bemessung

Widerstände und Belastungen als Zufallsgrößen, Versagenswahrscheinlichkeit p_f und Sicherheitsindex β , Bemessungspunkt, Teilsicherheitsfaktoren, der Sicherheitsnachweis nach den Normen.

10. Bayes'sche Konzepte

Bedingte Wahrscheinlichkeiten, Bayes'sche Parameterschätzung, der Begriff der subjektiven Wahrscheinlichkeit, Ereignisbäume.

11. Lineare Regression

Univariate lineare Regression, multivariate lineare Regression, Statistik der Regression.

Kapitel 1

Einführung

1.1 Die Rolle der Wahrscheinlichkeitstheorie und Statistik in den Technischen Wissenschaften

Das Einbeziehen von Risiko und Unschärfe in Projektplanung, Entwurf und Bemessung gewinnt zunehmend an Bedeutung in den Technischen Wissenschaften, nicht zuletzt wegen des probabilistischen Sicherheitskonzepts, das in den neuen Normen (Eurocode EN 1990:2002) festgeschrieben ist. Die Wahrscheinlichkeitstheorie bietet einen Rahmen, in dem Unsicherheiten und Unschärfen abgebildet und in mathematisch/technischen Berechnungsmodellen weiterverarbeitet werden können.

Eine ähnliche Entwicklung findet in den Naturwissenschaften statt, etwa in der Biologie und den Atmosphärenwissenschaften, unter dem Schlagwort „Uncertainty Quantification“, siehe etwa Smith(2014).

Andere Aspekte der Statistik und Modellierung mit Wahrscheinlichkeitstheorie treten in den Wirtschaftswissenschaften, der Soziologie, der Medizin und der Psychologie in den Vordergrund. Einen wieder anderen Blickwinkel hat die Physik in der Statistischen Mechanik und in der Quantenmechanik, wo der Wahrscheinlichkeitsbegriff Teil des physikalischen Modells ist.

Ziel dieser Vorlesung ist es, die allgemein gültigen Grundkonzepte zu vermitteln, wobei wir vornehmlich von Anwendungen in den Technischen Wissenschaften geleitet werden. In dieser Einführung sollen einige typische Anwendungsbereiche vorgeführt werden. (Die zu verwendenden Methoden werden in den nächsten Kapiteln erarbeitet.)

1. Messen, Datenerhebung: Beispiele dazu sind Vermessung in der Geodäsie, Materialkunde, Parameterbestimmung (zum Beispiel Druckfestigkeit, E-Modul), Qualitätsprüfung.

Eine Prüfung von 5 Betonwürfeln B30 einer Betonerzeugung habe etwa die folgenden Werte ergeben:

$$x = 28.5, \quad 26.0, \quad 30.2, \quad 26.6, \quad 29.9 \text{ [N/mm}^2\text{]}$$

Der Mittelwert ist

$$\bar{x} = 28.24 \text{ [N/mm}^2\text{]}$$

Wir fragen: Ist die Qualität des Betons als zureichend zu bewerten, wenn der Sollwert 30 [N/mm²] ist? Ist die Abweichung im Rahmen der zufälligen Schwankungen oder weist sie auf einen Qualitätsmangel hin?

2. Bemessen: Ein Einfeldträger unter zufällig variierender Belastung (zum Beispiel Schneelast) soll bemessen werden.

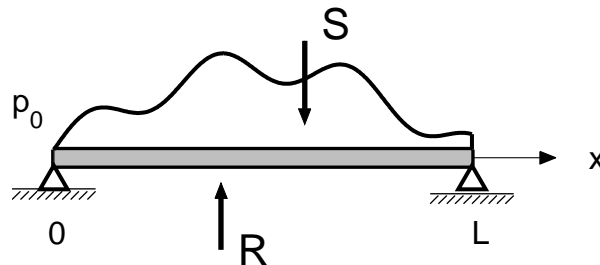


Abbildung 1.1: Balken unter variierender Streckenlast.

Der Träger ist einer Belastung S ausgesetzt, die zufällig schwankt. Die Belastbarkeit entspricht der Bruchlast R . Diese kann ebenfalls zufällig schwanken, etwa aufgrund unbekannter Materialfehler. Sicherheit liegt vor, wenn

$$R > S$$

erfüllt ist. In den alten Normen wurden zur Bemessung Sicherheitsbeiwerte aufgeschlagen, um Absicherung gegenüber Versagen zu gewährleisten. Die neuen Normen – mit dem (semi-)probabilistischen Sicherheitskonzept – setzen stattdessen ein wahrscheinlichkeitstheoretisches Modell an. Die Größen R und S werden als Zufallsgrößen aufgefasst. Dementsprechend gibt es keine absolute Sicherheit gegen Versagen, sondern eine *Versagenswahrscheinlichkeit*

$$p_f = P(S > R).$$

Das Bauwerk muss so bemessen werden, dass die Versagenswahrscheinlichkeit einen gewissen Wert nicht überschreitet, etwa

$$p_f \leq 0.000001.$$

Zufällige Lasten ergeben sich etwa durch Verkehrslasten, Wind, Wellen, Erdbeben usw.

3. Zusammenhänge ermitteln: Als ein Beispiel nehmen wir eine Messreihe des Durchmesser D [mm] und der Druckfestigkeit R [N/mm²] von Steinsalzkörnern. Die Messergebnisse sind in einem Streudiagramm dargestellt:

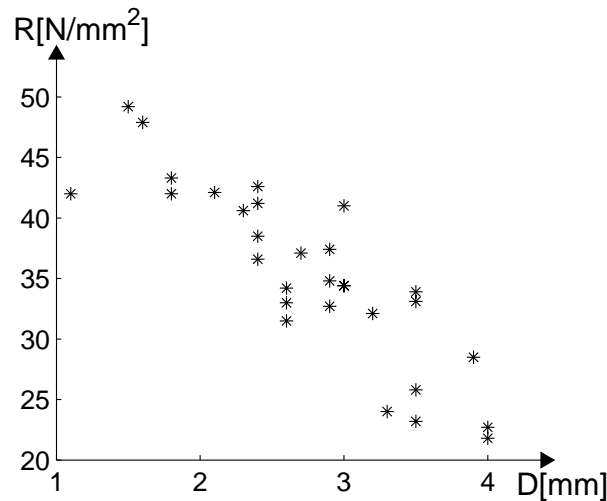


Abbildung 1.2: Durchmesser vs. Druckfestigkeit von Steinsalzkörnern, nach Stoyan(1993).

Offensichtlich besteht ein Zusammenhang zwischen den beiden Messgrößen. Die Statistik kann beantworten, wie stark die Größen *korreliert* sind. Sie kann aber auch bei der Entscheidung helfen, ob und was für ein funktionaler Zusammenhang vorliegt. In der Abbildung wurde eine lineare und eine quadratische Funktion mittels *Regression* an die Daten angepasst.

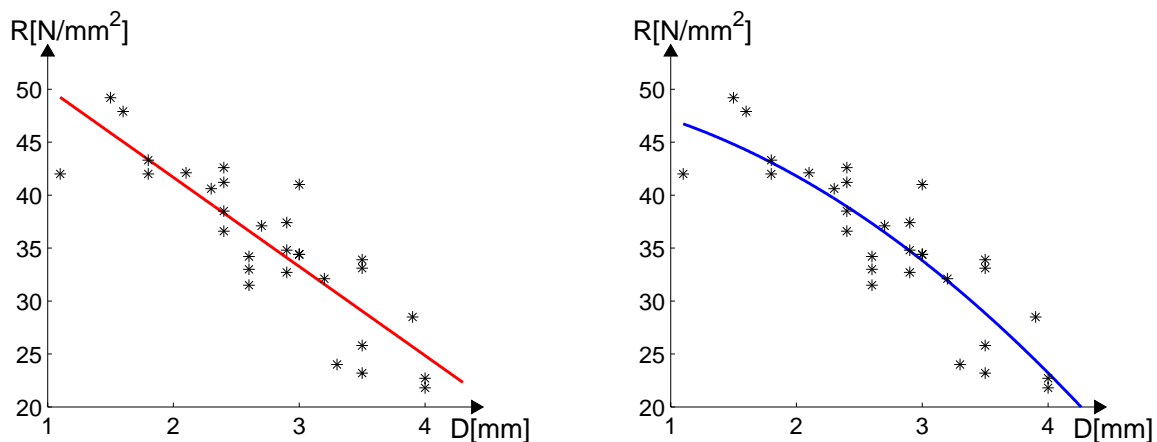


Abbildung 1.3: Regression mit linearer bzw. quadratischer Formfunktion.

Es handelt sich eigentlich um eine Aufgabe Ausgleichsrechnung, doch bietet die Statistik zusätzlich Indikatoren zur Modellwahl und Beurteilung der Anpassungsgüte an.

Ein weiteres typisches Anwendungsbeispiel der Regression im Bauwesen ist die Bestimmung von Materialparametern. Der E-Modul E etwa einer Stahlsorte im linear-elastischen Bereich ist durch das Hooke'sche Gesetz

$$\sigma = E\varepsilon$$

mit der Spannung σ und der Dehnung ε verbunden. Nimmt man einen Stahlstab der Länge L mit Querschnittsfläche A und legt eine Zugkraft F an, so bewirkt diese eine

Längenänderung ΔL . Der Zusammenhang mit Dehnung und Spannung ist

$$\varepsilon = \frac{\Delta L}{L}, \quad \sigma = \frac{F}{A}.$$

Legt man im Zugversuch mehrere Spannungen an, so ergibt sich eine Punktwolke von Messpunkten, die ungefähr auf einer Geraden liegen. Passt man nun mittels Regression eine Gerade durch den Ursprung an, so liefert deren Anstieg eine Schätzung für den Elastizitätsmodul E .

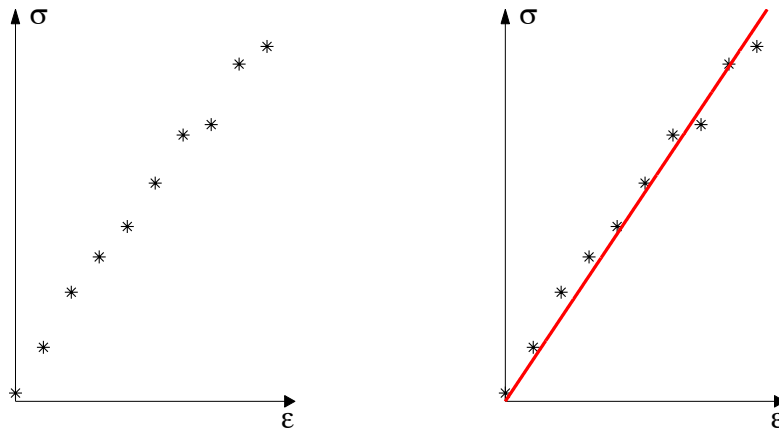


Abbildung 1.4: Ermittlung des E-Moduls als Anstieg der Regressionsgeraden.

4. Zeitreihen, Hochwasserstatistik: Die folgende Abbildung zeigt die Überschreitungsdauern des Inns 1971 – 1980, Pegel Innsbruck.

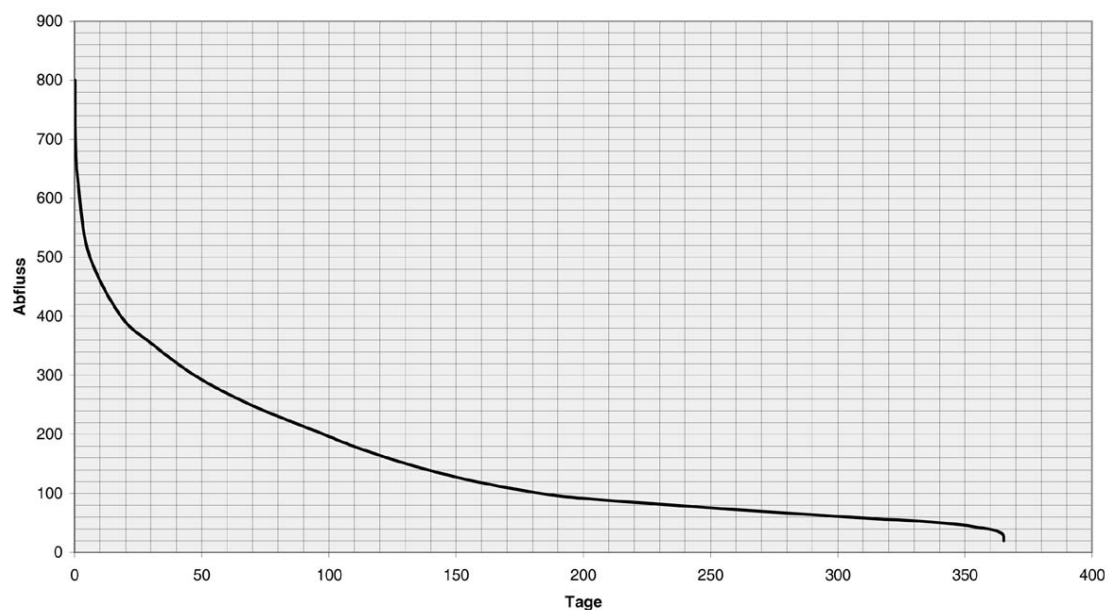


Abbildung 1.5: Überschreitungsdauern des Inns 1971 – 1980.

Wie kann daraus das 100- oder 1000-jährige Hochwasser abgeschätzt werden? Dies ist eine Aufgabe der Extremwertstatistik. Man verwendet dazu die Zeitreihe der Jahresmaxima HQ [m^3/s]:

Jahr	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
HQ	550	490	590	660	1030	445	713	797	702	778

Aus theoretischen Überlegungen kann die Reihe der Jahresmaxima durch eine so genannte Gumbelverteilung beschrieben werden. Die folgende Abbildung zeigt das Histogramm der zehn Jahresmaxima und die angepasste Dichte der Gumbelverteilung:

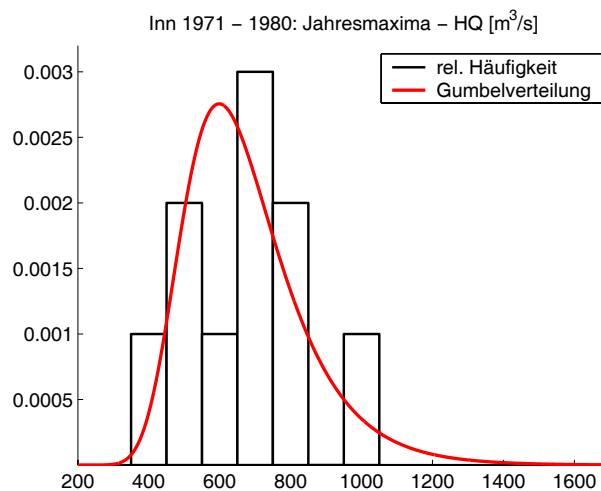


Abbildung 1.6: Jahresmaxima: Histogramm und Gumbelverteilung.

Trägt man die zugehörige Summenkurve (kumulative Verteilungsfunktion) auf, so kann daraus auf die Jährlichkeit geschlossen werden. Das 100-jährige Hochwasser erhält man als das 99%-Quantil der kumulativen Verteilungsfunktion, das 50-jährige als das 98%-Quantil usw.

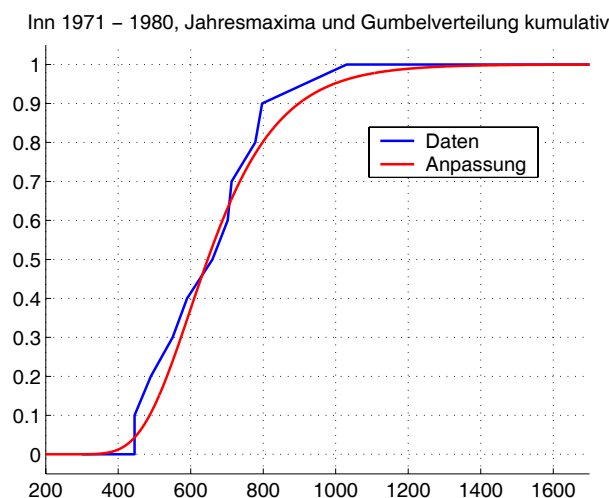


Abbildung 1.7: Jahresmaxima: Empirische Summenkurve, Verteilungsfunktion.

Am 23. August 2005 wurde am Pegel Inn die Spitze von $1540 \text{ m}^3/\text{s}$ gemessen. Die grobe Abschätzung mittels Gumbelverteilung ergibt dafür eine Jährlichkeit von 1150 [a].

5. Planung und Risikoabschätzung: Wahrscheinlichkeitstheoretische Modelle gewinnen in der Projektplanung und Abschätzung des Kostenrisikos zunehmend an Bedeutung. Bei Ausschreibungen von Großprojekten wird von den Bietern bereits häufig eine begleitende Risikoanalyse verlangt.

Betrachten wir als Beispiel die Abschätzung der Dauer eines einfachen Vorgangs der Schottergewinnung: Laden – Transportieren – Brechen. In einem deterministischen Netzplan sind die geschätzten Dauern (in Minuten) eingetragen:

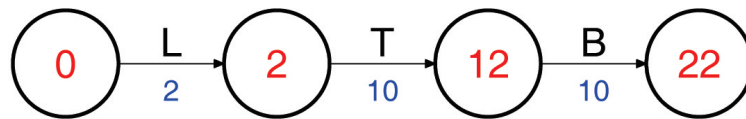


Abbildung 1.8: Netzplan Schottergewinnung.

Unter Vernachlässigung zufälliger Schwankung ergibt sich für die Gesamtdauer des Vorgangs 22 [min]. Tatsächlich wird diese Dauer variieren; eine zuverlässige Leistungsangabe sollte Aussagen über die Schwankungsbreite machen. Wir führen drei verschiedene Modellierungsmöglichkeiten für die Unschärfe der Einzeldauern vor.

Für den *Ladevorgang* wird eine Dreiecksverteilung postuliert.

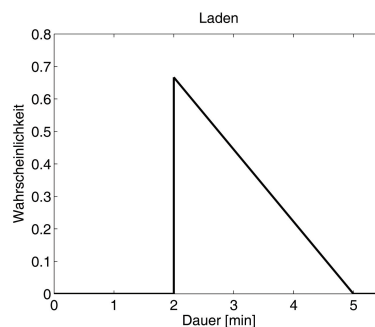


Abbildung 1.9: Laden: Dreiecksverteilung.

Für den *Transportvorgang* sei eine statistische Erhebung (Stichprobe vom Umfang 100) vorgenommen worden und eine Normalverteilung angepasst worden.

Für den *Brechvorgang* sei eine subjektiven Wahrscheinlichkeitsverteilung mittels Delphi-Methode (Expertenbefragung) erhoben worden, mit nachfolgender Anpassung einer Log-normalverteilung.

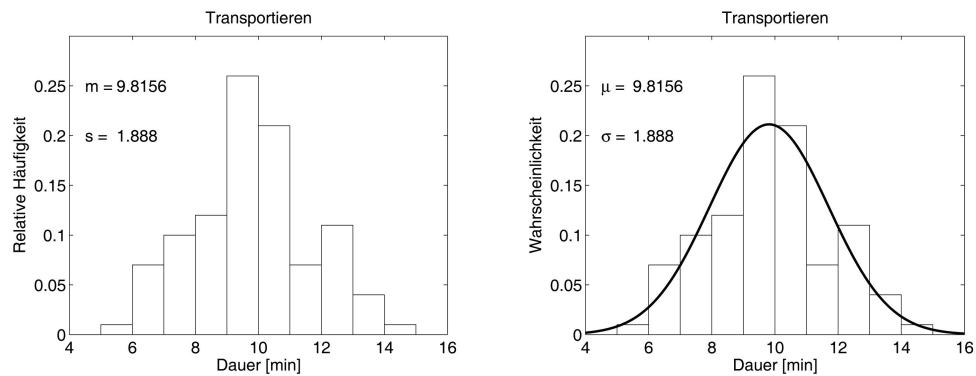


Abbildung 1.10: Transportieren: Histogramm und Normalverteilung.

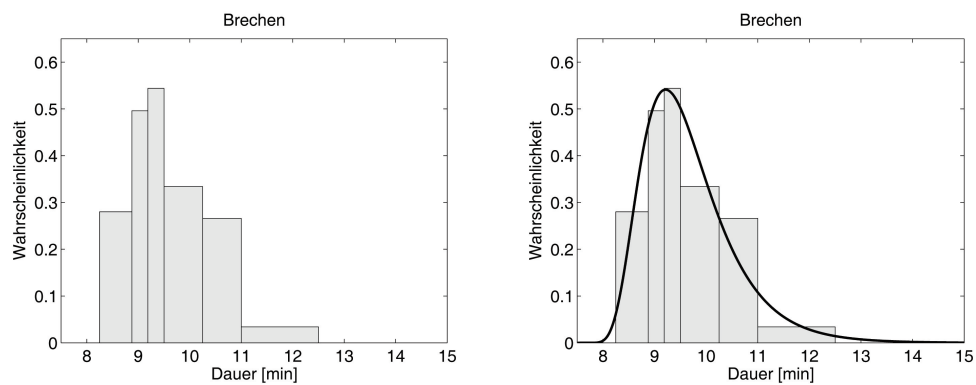


Abbildung 1.11: Brechen: Verteilung nach Expertenbefragung und Lognormalverteilung.

Die statistische Verteilung der Gesamtdauer kann mittels Monte-Carlo-Simulation numerisch berechnet werden:

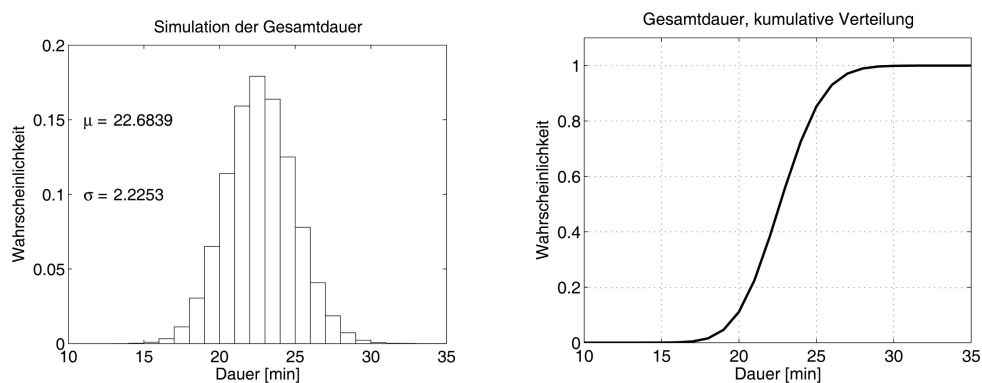


Abbildung 1.12: Simulation der Gesamtdauer.

In einem weiteren Anwendungsbeispiel soll das Schadensrisiko durch Hochwasser während des Baus eines Erddamms in einem Hochgebirgstal abgeschätzt werden, und zwar in seinen

Auswirkungen auf die Baukosten und auf die Bauzeit. Das Beispiel befasst sich mit der Baustelle Moosbach/Kartell (2004).

Für die Schüttphase des Damms wurde eine Bauwasserumleitung auf HQ50 dimensioniert. Betrachtete Risiken waren: Verklausung des Umleitungsstollens; Bauzustand des Damms (Kerngraben offen/nicht offen); Beschädigung der Dammkrone. Die kausalen Zusammenhänge wurden mit Hilfe eines Ereignisbaums dargestellt (Ausschnitt):

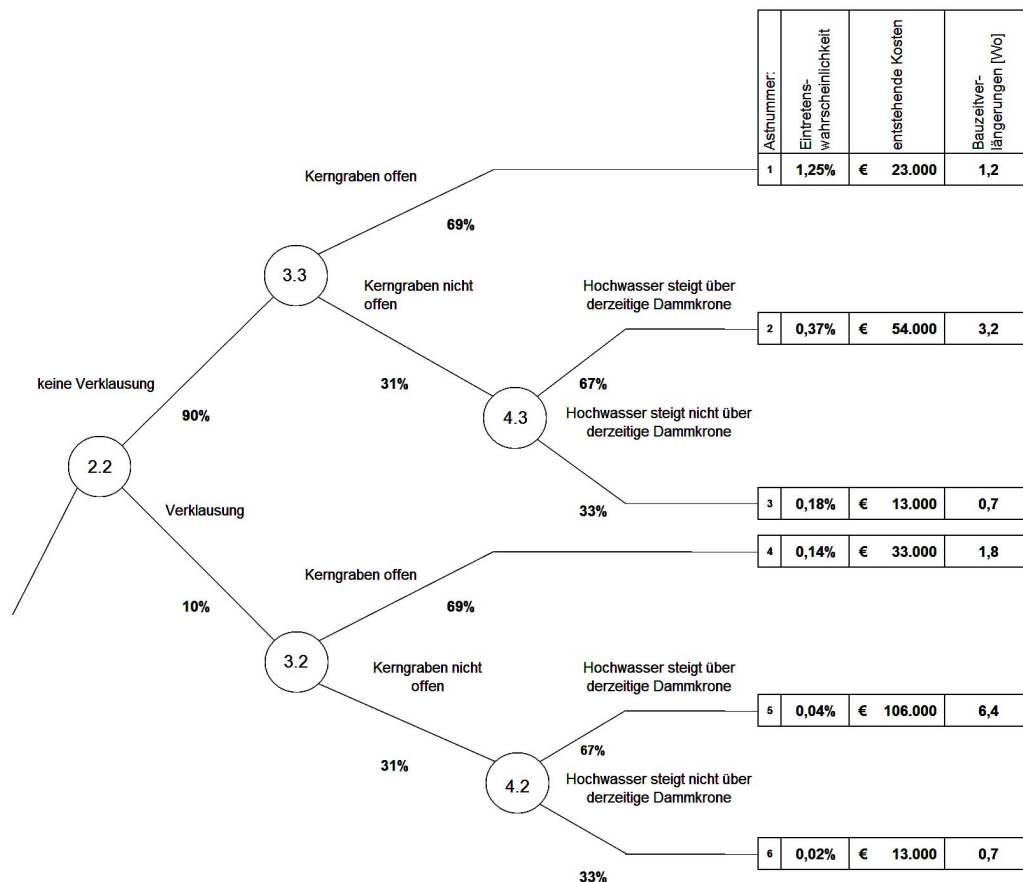


Abbildung 1.13: Ereignisbaum Hochwasserrisiko bei Dammschüttung (Ausschnitt)

Die Gesamtbewertung des Risikos erhält man durch statistische Auswertung der Kosten/Bauzeitverlängerung mit zugehörigen Wahrscheinlichkeiten. Erste Indikatoren sind der Erwartungswert der Schadenssumme/Bauzeitverlängerung und die Varianz. Ein deutlicheres Bild der Risikoverteilung liefert ein Histogramm.

6. Regionalisierte Variable: Regionalisierte Variable – ein Begriff aus der Geostatistik – sind „regional“ verteilte Zufallsgrößen. Genauer gesagt, wird jedem Punkt P eines Gebiets G eine Zufallsgröße X_P zugeordnet. Man spricht auch von einem räumlichen Zufallsfeld. Beispiele sind etwa die Niederschlagsmenge oder die Windstärke. Diese wird natürlich nur in einigen wenigen Punkten (Messstationen) gemessen, hat aber in jedem Punkt des betrachteten Gebiets einen Wert. Andere Beispiele sind die Schneehöhe, die Bodenfeuchte, der örtlich verteilte Erzgehalt von Gestein im Bergbau, die Konzentration einer Verunreinigung im Grundwasser, Bodenparameter in der Geotechnik u.a.m.

Eine der Aufgaben der regionalisierten Statistik ist die Extrapolation der Werte auf die Bereiche, in denen keine Messdaten vorliegen.

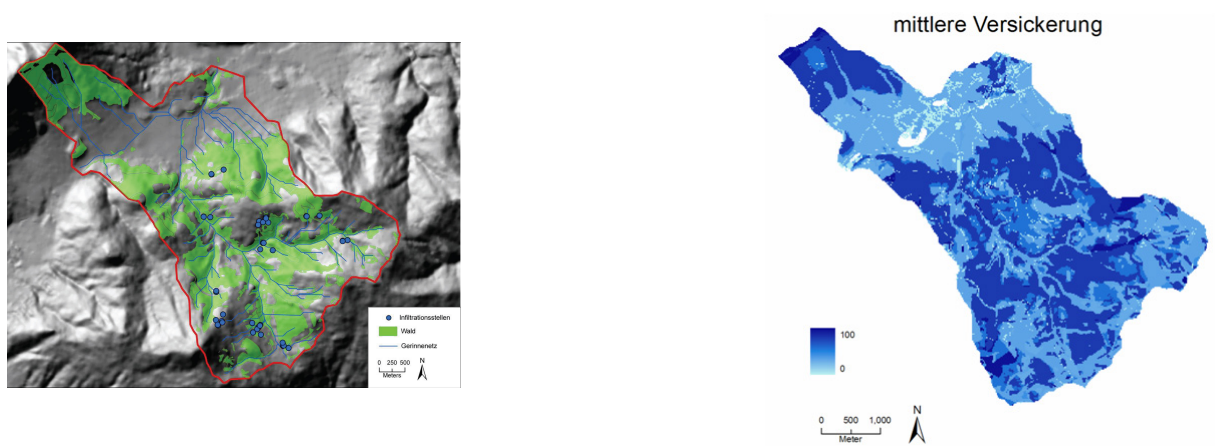


Abbildung 1.14: Aus dem alpS-Projekt A3.1, 2006: Bodenfeuchte und Messstellen, Extrapolation der mittleren Versickerung (Stampfangerbach, Tirol).

7. Spezielle probabilistische Modelle: Einige speziellere Anwendungen statistischer Methoden seien hier aufgelistet; diese sind jedoch nicht Gegenstand der Vorlesung.

- Verkehrszählung und Verkehrsplanung;
- Lagerhaltung;
- Zuverlässigkeitstheorie;
- Leistungsminderung durch Warteschlangen;
- stochastische Strukturmechanik (Erdbeben – Schwingungen von Bauwerken);
- statistische Zeitreihenanalyse von Brückenschwingungen zur Erkennung von Schädigungen;
- Sensitivitätsanalysen in der Strukturmechanik und Festigkeitslehre;
- Markow-Ketten in Simulation und Wettermodellen.

(Die Liste der aktuellen Anwendungen von wahrscheinlichkeitstheoretischen Modellen in den Technischen Wissenschaften ist damit keineswegs erschöpft ...)

1.2 Grundsätzliches zur Methodik

Die beschreibende Statistik erhebt Daten, stellt diese zusammenfassend dar und berechnet daraus Indikatoren. Dies reicht nicht aus, um Erklärungsmodelle, Prognosen und

Güteaussagen aufzustellen oder Simulationen vorzunehmen. Dazu braucht es ein mathematisches Modell für die Verteilung der Daten. Gerade in den Technischen Wissenschaften bildet daher das Zusammenspiel von Empirie (Datenerhebung und Expertenbeurteilung) und Theorie (mathematische Modellierung) den Kern der wahrscheinlichkeitstheoretischen Vorgangsweise. Erst die Kombination von Empirie und Theorie bringt aussagekräftige Ergebnisse. Die Vorgangsweise besteht im Wesentlichen aus fünf Schritten:

- Datenerhebung und Beschreibung;
- Aufstellen eines wahrscheinlichkeitstheoretischen Modells;
- Schätzung der Modellparameter aus den Daten;
- Überprüfung der Anpassungsgüte des Modells;
- Prognose und Simulation.

Ziel der Vorlesung ist es, die Begriffe, die für diese fünf Schritte benötigt werden, einzuführen und ihre Anwendung mit einigen Beispielen zu belegen, die Ihnen im Studium mit ziemlicher Sicherheit begegnen werden. In Anbetracht der zunehmenden Bedeutung der Probabilistik ist Kenntnis der richtigen Ausdrucksweise unverzichtbar geworden.

1.3 Was ist Wahrscheinlichkeit?

Die *Interpretation* oder *Semantik* der Wahrscheinlichkeitstheorie ist die Frage, wie die wahrscheinlichkeitstheoretischen Begriffe in der Realität zu interpretieren sind. Von den zahlreichen Interpretationen der Wahrscheinlichkeit sind für das Ingenieurwesen vor allem drei von großer Wichtigkeit: die *klassische*, die *frequentistische* und die *subjektivistische*.

Klassische Wahrscheinlichkeit:

Dieser Wahrscheinlichkeitsbegriff wird mit dem Namen *Laplace* verbunden und könnte auch als *kombinatorisch* bezeichnet werden. Kurz gesagt postuliert dieser Zugang, dass die Wahrscheinlichkeit eines Ereignisses A wie folgt anzugeben ist:

$$P(A) = \frac{\text{Anzahl günstige Fälle}}{\text{Anzahl mögliche Fälle}}$$

Beispiel: Wird beim Zahlenlotto 6 aus 45 das Ereignis des Auftretens eines Sechser mit A bezeichnet, so ist die Anzahl der günstigen Fälle gleich 1, die Anzahl der möglichen Fälle gleich der Anzahl aller möglichen Sechser unter 45 Zahlen, also gleich

$$\binom{45}{6} = \frac{45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 8145060.$$

Somit ist die Wahrscheinlichkeit für einen Sechser gleich

$$P(A) = \frac{1}{8145060} = 0.00000012 \dots$$

Die kombinatorische Wahrscheinlichkeit hat allerdings weitere zahlreiche Anwendungen, wie zum Beispiel den Entwurf von Qualitätsprüfverfahren oder die Herleitung der Exponentialverteilung für die Lebensdauer radioaktiver Partikel aus dem radioaktiven Zerfallsgesetz.

Frequentistische Interpretation:

Dabei wird die Wahrscheinlichkeit $P(A)$ eines Ereignisses A durch die relative Häufigkeit des Auftretens von A in einer Stichprobe hinreichend großen Umfangs angenähert. In der frequentistischen Interpretation nimmt man also folgende Entsprechung vor:

Wahrscheinlichkeit \approx relative Häufigkeit.

In der Idealvorstellung ist dazu eine große Zahl von Messwerten, die unter identischen Bedingungen erhoben wurden, zur Verfügung. Solche Daten sind gerade im Bauwesen meist nicht vorhanden oder nicht gewinnbar (man denke an Bodenproben). Daher muss die Datenlage in der Regel durch Expertenwissen (zum Beispiel des/der Geologen/Geologin) ergänzt werden; damit begibt man sich in den Bereich der so genannten subjektiven Wahrscheinlichkeit.

Subjektivistische Interpretation:

In dieser Interpretation wird die Wahrscheinlichkeit eines Ereignisses A als Vertrauensgrad in sein Auftreten gedeutet. Hier wird bewusst in Kauf genommen, dass dieser Vertrauensgrad eine subjektive Einschätzung bedeutet und daher von verschiedenen Experten verschiedene Werte zugewiesen bekommen kann. Selbstverständlich muss die Bewertung rational und unter Zuhilfenahme aller verfügbaren Information erfolgen. Um eine operative Methode zu erhalten, den subjektiven Vertrauensgrad zu quantifizieren, wurde vorgeschlagen, diesen als Indifferenzpreis für ein gedachtes Wettspiel anzusetzen. Diese Vorgangsweise wird *Introspektion* genannt. Wichtiger für die Risikoanalyse ist allerdings die Expertenbefragung, auch *Mediation*, *Delphi-Methode* oder auf Englisch *elicitation* genannt. Dabei werden Einzelwahrscheinlichkeiten und Verteilungsfunktionen in einem Diskussionsprozess festgelegt. Es können dabei unter anderem

- statistische Daten,
- Erfahrungswerte, Normwerte,
- Bewertungstabellen,
- Ereignisbäume

herangezogen werden.

Mathematisches Modell:

Das mathematische Modell ist in allen drei Interpretationen dasselbe und kann auf dem Begriff der Zufallsgröße und deren Verteilungsfunktion aufgebaut werden. Diesem Begriff

werden gewisse Axiome zu Grunde gelegt, eben die Axiome der Wahrscheinlichkeitstheorie, die besagen, wie Verknüpfungen vorzunehmen sind. Ein typisches Beispiel dafür ist die Additivität:

$$P(A) + P(A^c) = 1;$$

die Wahrscheinlichkeiten eines Ereignisses und dessen Komplementärereignisses müssen sich zu Eins addieren.

Für manche Anwendungen werden diese Axiome als zu strikt betrachtet, gerade wenn es um die Festlegung subjektiver Bewertungen geht. In neuerer Zeit sind einige Verallgemeinerungen von oder Alternativen zur Wahrscheinlichkeitstheorie entwickelt worden. Dazu gehören intervallwertige Wahrscheinlichkeiten (imprecise probabilities), zufällige Mengen (random sets) und unscharfe Mengen (fuzzy sets). Diese sind nicht Gegenstand der Vorlesung. Es sei dazu auf Fellin et al.(2005) oder Oberguggenberger(2012) verwiesen.

Kapitel 2

Beschreibende Statistik eindimensionaler Daten

2.1 Datenerhebung und Darstellung

Aufgabenstellung: Die Aufgabenstellung besteht in der Aufnahme und Analyse von Daten mit zufälligen Schwankungen. Dazu gehört die Schätzung der Parameter einer Grundgesamtheit aus einer Stichprobe. Die schließende Statistik baut darauf auf und erlaubt zum Beispiel die Prognose von Über- und Unterschreitenswahrscheinlichkeiten, das Erstellen eines Risikobildes sowie das Testen der Adäquatheit von Erklärungsmodellen.

Wir führen nun einige der grundlegenden Sprechweisen ein.

Grundgesamtheit: Diese kann real oder fiktiv sein; zum Beispiel eine Menge von *Individuen*, wie alle österreichischen Arbeitnehmer (real), oder alle potentiell möglichen Messungen der Druckfestigkeit von Beton B35 (fiktiv).

Merkmal: Jedes Element der Grundgesamtheit besitzt ein oder mehrere *Merkmale* mit zugehöriger *Ausprägung*. Über die Merkmalausprägung hinaus besitzen die Individuen keine Individualität. Einige Beispiele:

Grundgesamtheit	Merkmal	Ausprägung
Einwohner Österreichs	Geschlecht	m/w
	Partei	Ø / ÖVP / SPÖ / ...
	Alter*	0, 1, 2, 3, ...
Erwerbstätige Österreichs	Einkommen*	0, 1, 2, ..., 1000, ... Euro
Gemeinden Tirols	Einwohnerzahl*	1, 2, 3, ...
	Fläche	1, 2, 3, ... km ²
Versuche einer Messreihe	Messwert	Zahl der Messskala

* Zeitabhängige Merkmale, Zeitraum ist zusätzlich anzugeben, also etwa „Einkommen im Jahre 2016“.

Merkmaltypen: (1) *Qualitative Merkmale* z. B. Geschlecht, Parteibuch. Messung durch:

- (i) Nominalskala (= Benennung), z. B. m/w; ÖVP / SPÖ / ...
- (ii) Ordinalskala (= Anordnung) z. B. „Akzeptanz des EU-Beitrittes“

dagegen (4) – eher dagegen (3) – eher dafür (2) – dafür (1)

(2) *Quantitative Merkmale*, z. B.: Einwohnerzahl, Einkommen, physikalische Parameter. Die Ausprägung ist durch eine Zahl charakterisiert. Messung durch:

- (iii) Kardinalskala (= Zahlenwerte), z. B. 1000, 2000 km²; 1, 2, 3, 4 Millionen

Zufälliger Versuch: Herausgreifen eines Individuums und Feststellung des Merkmals bzw. Messung.

Stichprobe vom Umfang n : Die n -fache Wiederholung eines Versuches unter identischen Bedingungen (z. B. Erhebung unter 1000 Arbeitnehmern oder 10 Druckversuche an verschiedenen Würfeln derselben Betonsorte).

In der Folge beschränken wir uns auf quantitative Merkmale. Die *Ausprägungen* des Merkmals seien

$$\begin{array}{ll} x_1, \dots, x_k & \text{diskret oder} \\ x \in \mathbb{R} & \text{kontinuierlich.} \end{array}$$

Beispiel: Schülernoten in einer Klasse:

$$x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5.$$

Das Ergebnis der Stichprobe ist zunächst die *Urliste* ξ_1, \dots, ξ_n , zum Beispiel (mit $n = 31$)

1	///
2	////////
3	//////////
4	////////
5	//

Beispiel: Mittlere Tagestemperaturen im Jänner in Innsbruck (Tabelle 2.1). Die Temperatur besitzt die Ausprägung einer kontinuierlichen Größe; die Messdaten sind selbstverständlich diskret, werden aber zur besseren Darstellung und Auswertung in Klassen unterteilt (wie stets bei kontinuierlichen Größen). Beginnen wir mit dem Merkmal X = Temperatur Jänner 2001. Die Daten bilden zunächst die *Urliste* ξ_1, \dots, ξ_n (mit $n = 31$). Wir nehmen eine Klasseneinteilung der Temperatur vor, hier in Klassen von -10° bis 10° in Schritten von zwei Grad.

-5° bis -3°	4
-3° bis -1°	7
-1° bis $+1^\circ$	5
$+1^\circ$ bis $+3^\circ$	7
$+3^\circ$ bis $+5^\circ$	5
$+5^\circ$ bis $+7^\circ$	3

Die Wahl der Klassenbreite bleibt den Bearbeitenden überlassen. Ist die Klassenbreite zu gering, zerflattert die Darstellung; ist sie zu groß, verschwimmen die Eigenschaften der Datenverteilung.

Die Auswertung der Stichprobe liefert zunächst die Häufigkeitsverteilung als Grundlage für alles Weitere:

Absolute Häufigkeit:

$$H_j = \begin{cases} H(X = x_j) & \dots \text{Anzahl der Fälle, in denen } x_j \text{ aufgetreten ist;} \\ H(X \in \text{Klasse } j) & \dots \text{Anzahl der Fälle, in denen Klasse } j \text{ aufgetreten ist.} \end{cases}$$

Relative Häufigkeit:

$$h_j = \frac{1}{n} H_j \dots \text{Anteil des Auftretens von } x_j \text{ (bzw. der Klasse } j)$$

Beispiel Schülernoten:

x_j	1	2	3	4	5
H_j	3	7	10	9	2

Graphische Darstellungen: Man verwendet meist ein *Säulendiagramm* für die absoluten Häufigkeiten (längentreue Darstellung, Höhe \approx absolute Häufigkeit) und ein *Histogramm* für die relativen Häufigkeiten (flächentreue Darstellung, Fläche \approx relative Häufigkeit).

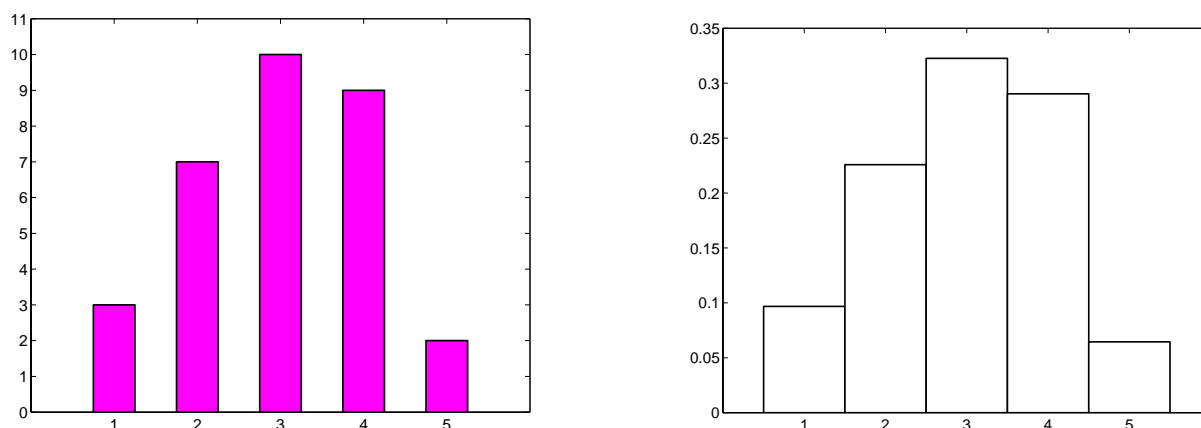


Abbildung 2.1: Säulendiagramm und Histogramm für das Schulnotenbeispiel.

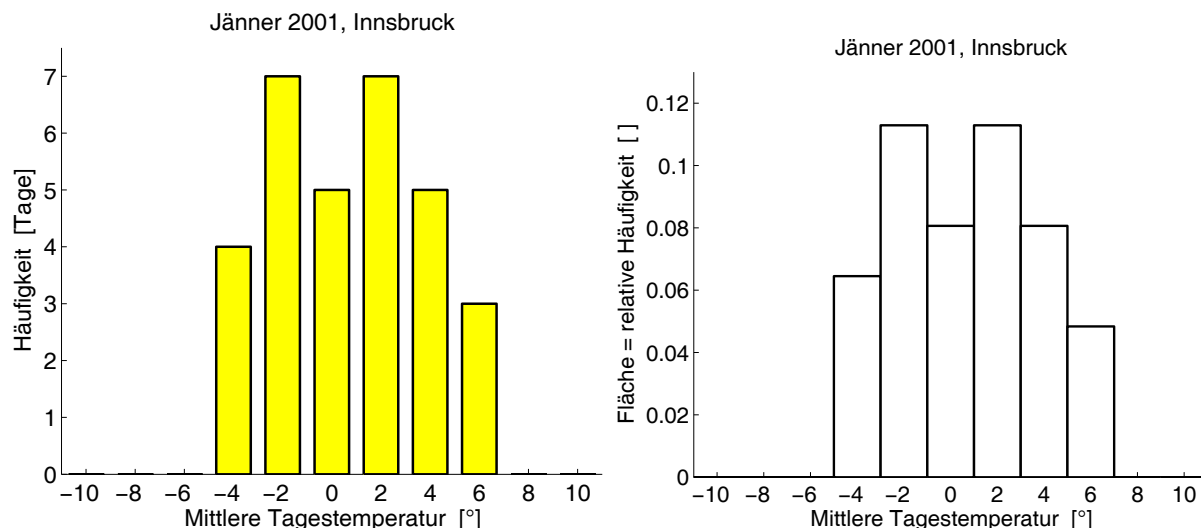


Abbildung 2.2: Säulendiagramm und Histogramm, mittlere Tagestemperatur Jänner 2001.

Die empirische Verteilungsfunktion/Summenhäufigkeit: Wir brauchen dazu das Konzept des *Ereignisses*.

Elementarereignisse: Auftreten eines Merkmales, also $X = x_j$. Die H_j sind dann die Auftretshäufigkeiten der Elementarereignisse.

Zusammengesetzte Ereignisse: Vereinigung von Elementarereignissen, zum Beispiel “Note zwischen 2 und 4”, $2 \leq X \leq 4$. Im Beispiel ist die absolute Häufigkeit davon $H(2 \leq X \leq 4) = 7 + 10 + 9 = 26$, die relative Häufigkeit $h(2 \leq X \leq 4) = 26/31$.

Die empirische Verteilungsfunktion, kumulierte- oder Summenhäufigkeit ist die für alle $x \in \mathbb{R}$ definierte Stufenfunktion

$$F_{\text{emp}}(x) = h(X \leq x) = \sum_{x_j \leq x} h_j = \frac{1}{n} \sum_{x_j \leq x} H_j.$$

Man summiert also einfach die Häufigkeiten links von jedem x_j . An der Stelle x_j macht $F_{\text{emp}}(x)$ einen Sprung der Höhe h_j . Bei Verwendung von Klasseneinteilungen wird x_j als Klassenmitte gewählt (siehe Abbildung 2.3).

Wir brauchen die empirische Verteilungsfunktion einerseits zur graphischen Kurvenanpassung, andererseits für das Konzept der Überschreitungswahrscheinlichkeiten.

2.2 Datenbeschreibung: Lageparameter

Die wichtigsten Lageparameter sind der Mittelwert (das Stichprobenmittel), Maximal- und Minimalwert, der Median und die Quantile.

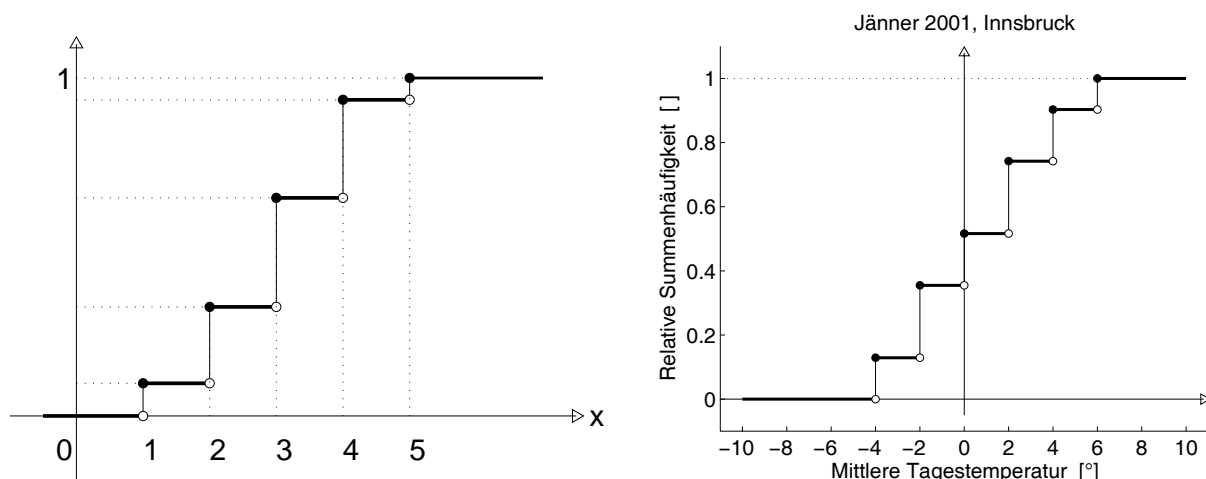


Abbildung 2.3: Summenhäufigkeit im Notenbeispiel (links) und für die mittleren Tagestemperaturen (rechts).

Der Mittelwert: Das arithmetische Mittel der aufgetretenen Werte,

$$\begin{aligned}\bar{x} &= \frac{1}{n}(\xi_1 + \xi_2 + \cdots + \xi_n) \\ &= \frac{1}{n}(x_1 H_1 + x_2 H_2 + \cdots + x_k H_k) \\ &= \frac{1}{n} \sum_{j=1}^k x_j H_j = \sum_{j=1}^k x_j h_j.\end{aligned}$$

Im Notenbeispiel:

$$\begin{aligned}\bar{x} &= \frac{1}{31}(1 \cdot 3 + 2 \cdot 7 + 3 \cdot 10 + 4 \cdot 9 + 5 \cdot 2) \\ &= \left(1 \cdot \frac{1}{31} + 2 \cdot \frac{7}{31} + 3 \cdot \frac{10}{31} + 4 \cdot \frac{9}{31} + 5 \cdot \frac{2}{31}\right) = 3\end{aligned}$$

Maximalwert und Minimalwert: x_{\max} , x_{\min} .

Median (Zentralwert): Gibt die Mitte der Urliste bei größenmäßiger Anordnung an. Zur Ermittlung stellt man die *Variationsreihe* (geordnete Urliste) auf,

$$\xi_1 \leq \xi_1 \leq \cdots \leq \xi_n.$$

Falls $n = 2\ell + 1$ ist, so ist der Median \tilde{x} gleich $\xi_{\ell+1}$; falls $n = 2\ell$ ist, so ist $\tilde{x} = (\xi_\ell + \xi_{\ell+1})/2$. Im Notenbeispiel ist die Variationsreihe

$$1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5$$

An 16. Stelle findet sich der Median $\tilde{x} = 3$.

Im Beispiel der mittleren Tagestemperaturen Jänner 2001 ergibt sich:

$$\bar{x} = 0.76, \quad x_{\max} = 5.80, \quad x_{\min} = -4.50.$$

Die Variationsreihe ist

-4.5, -4.5, -3.4, -3.4, -2.8, -1.6, -1.5, -1.5, -1.2, -1.1, -1.0, -0.6, -0.4, -0.2, 0.3, 0.4, 1.3, 1.3, 1.5, 1.8, 2.1, 2.1, 2.8, 3.7, 3.9, 4.0, 4.3, 5.0, 5.2, 5.8, 5.8

An 16. Stelle findet sich der Median $\tilde{x} = 0.40$.

Quantile: Das $q\%$ -Quantil Q ist dadurch gekennzeichnet, das $q\%$ der Werte unter Q und $(100 - q)\%$ darüber liegen. Der Median ist somit das 50%-Quantil.

Viertelwerte: V_u = unterer Viertelwert = 25%-Quantil, V_o = oberer Viertelwert = 75%-Quantil.

Die folgenden Größen sind eigentlich Streuungsmaße, leiten sich aber unmittelbar aus den Lageparametern her:

Spannweite: $\Delta = x_{\max} - x_{\min}$;

Viertelweite: $d_V = V_o - V_u$.

Im Notenbeispiel:

$$V_u = 2, \quad V_o = 4, \quad \Delta = 4, \quad d_V = 2.$$

Im Beispiel Temperatur Jänner 2001:

$$V_u = -1.4250, \quad V_o = 3.4750, \quad \Delta = 10.30, \quad d_V = 4.90.$$

Ausreißer sind Messwerte, die augenfällig deutlich abseits der Mehrzahl der übrigen Messwerte liegen. Der Begriff bleibt vage; ein Ausreißer kann auf einen Mess- oder Aufnahmefehler hinweisen oder in der breiten Streuung der Daten begründet sein. Dies muss in jedem Einzelfall untersucht werden; die statistische Ausreißertheorie geht über den Rahmen der Vorlesung hinaus, siehe etwa Barnett et al. (1994). In der Datenanalyse behilft man sich mit einer empirischen Definition von Ausreißern, indiziert durch das Überschreiten gewisser Grenzen im Vergleich zu den übrigen Lageparametern. Üblicherweise legt man diese Grenzen wie folgt fest.

Ausreißergrenzen: Obere Ausreißergrenze: $A_o = V_o + 1.5 d_V$; untere Ausreißergrenze: $A_u = V_u - 1.5 d_V$.

Vergleich Median-Mittelwert: Der Median ist stabil gegenüber Ausreißern, der Mittelwert nicht. Einzelne große Messwerte können den Mittelwert deutlich nach oben verschieben. Beispiel:

x_j	1	2	3	4	5	20	21
H_j	3	7	10	9	0	1	1

$$\tilde{x} = 3, \quad \bar{x} = \frac{1}{31} (1 \cdot 3 + 2 \cdot 7 + 3 \cdot 10 + 4 \cdot 9 + 20 \cdot 1 + 21 \cdot 1) = 4.$$

Darstellung im Boxplot: Das Boxplot besteht aus einem Rechteck, begrenzt durch V_u und V_o ; der Median wird durch einen Strich gekennzeichnet. Links und recht schließen sich

die “Whiskers” (Katzenhaare) an, die x_{\min} und x_{\max} andeuten. Im Falle von Ausreißern geben die Whiskers den minimalen bzw. maximalen Wert unter Entfernung der Ausreißer an, welche ihrerseits durch zusätzliche Ringe oder Sterne gekennzeichnet werden.

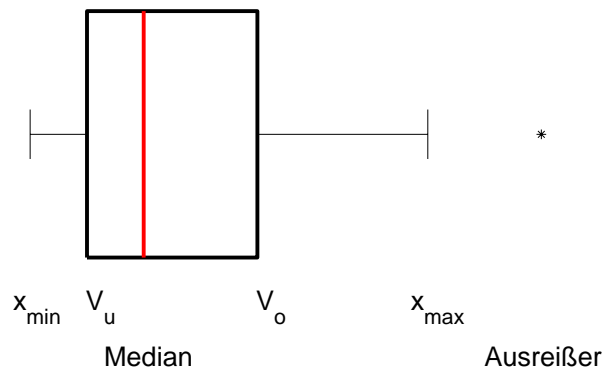


Abbildung 2.4: Schema eines Boxplots.

Boxplots in zwei Darstellungsmöglichkeiten für das Notenbeispiel und das Temperaturbeispiel sind der Abbildung 2.5 zu entnehmen.

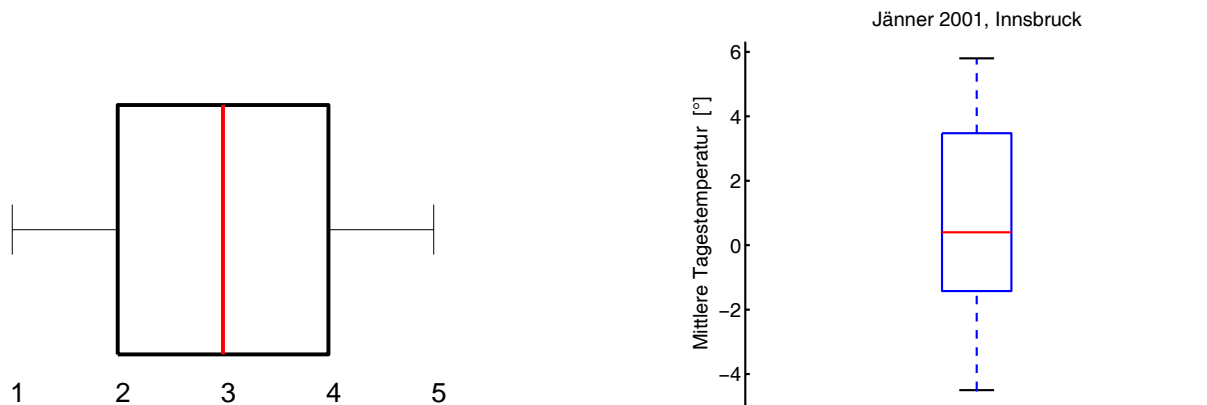


Abbildung 2.5: Boxplot für Notenbeispiel (links) und Temperaturen (rechts).

Boxplots sind auch zum Vergleich verschiedener Gesamtheiten von Interesse. Wir führen den Vergleich der Temperaturen im Jänner 2001 und im Jänner 2002 mit Hilfe verschiedener Darstellungsmethoden vor.

Zunächst die Darstellung der mittleren Tagestemperaturen im Jänner 2002:

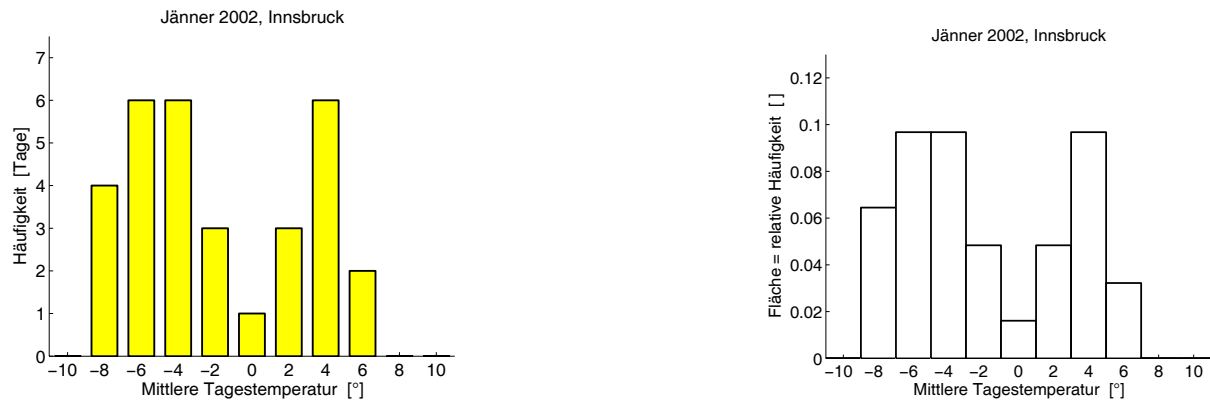


Abbildung 2.6: Tagestemperaturen Jänner 2002: Säulendiagramm und Histogramm.

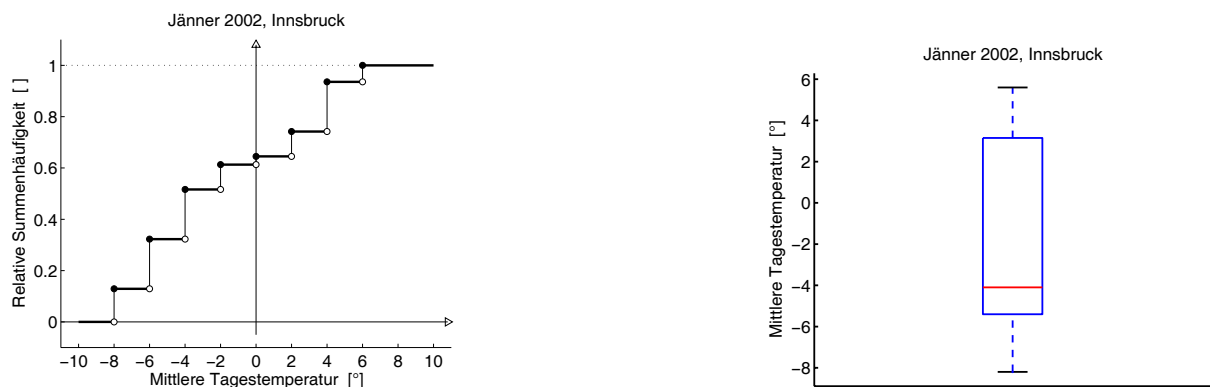


Abbildung 2.7: Tagestemperaturen Jänner 2002: Summenhäufigkeit und Boxplot.

Und hier die vergleichende Darstellung:

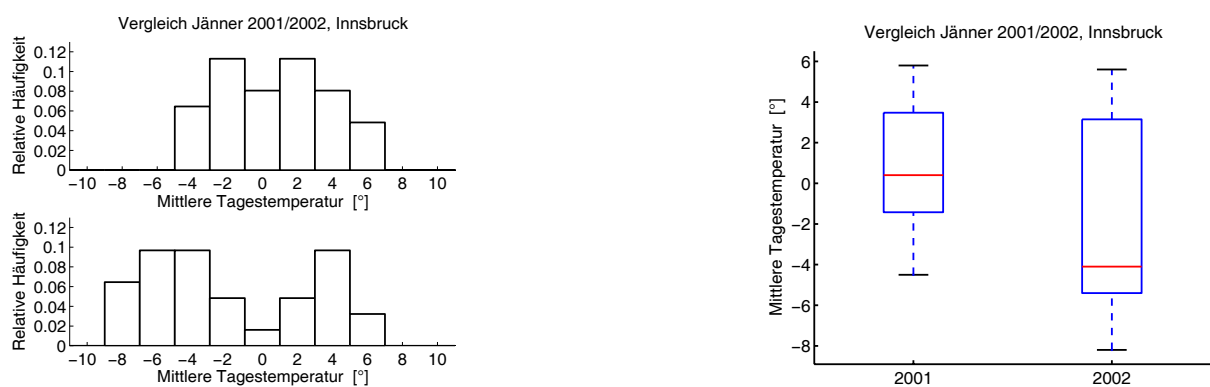


Abbildung 2.8: Vergleich 2001/2002: Histogramm und Boxplot.

2.3 Datenbeschreibung: Varianz

Die Stichprobenvarianz: Die Stichprobenvarianz s^2 ist ein Streuungsmaß und gibt an, wie sehr die Messwerte im Quadratmittel um den Mittelwert schwanken.

$(x_j - \bar{x})^2$... quadratische Abweichung des Messwertes x_j vom Mittelwert \bar{x}

Mittelwert der Abweichungsquadrate:

$$\begin{aligned} s^2 &= \frac{1}{n-1} \left((x_1 - \bar{x})^2 H_1 + (x_2 - \bar{x})^2 H_2 + \cdots + (x_k - \bar{x})^2 H_k \right) \\ &= \frac{1}{n-1} \sum_{j=1}^k (x_j - \bar{x})^2 H_j = \frac{n}{n-1} \sum_{j=1}^k (x_j - \bar{x})^2 h_j. \end{aligned}$$

Der Faktor $(n-1)$ statt n wird in der Schätztheorie erklärt (Erwartungstreue des Schätzers s^2 für die Varianz einer Zufallsgröße).

Die Standardabweichung Die Quadratwurzel s der Stichprobenvarianz heißt Standardabweichung.

Der Variationskoeffizient: Er gibt die relative Größe der Streuung in Bezug auf den Mittelwert an und ist definiert durch

$$v = s/\bar{x} \cdot 100 \, \%.$$

In der Ingenieurliteratur findet man auch die Bezeichnung CoV, was aber zur Verwechslung mit der Kovarianz (COV) führen kann.

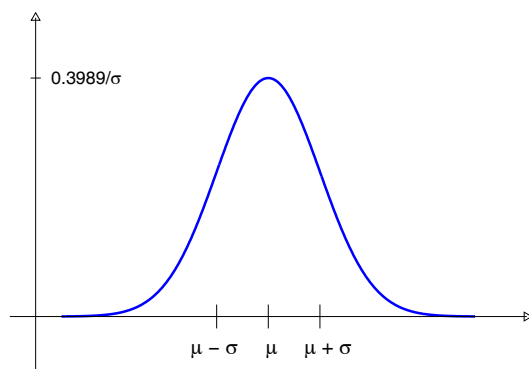
Beispiele zur Bedeutung der Standardabweichung, Notenverteilungen vom Umfang n :

$$\begin{array}{c|ccccc} x_j & 1 & 2 & 3 & 4 & 5 \\ \hline H_j & 2 & 2 & 2 & 2 & 2 \end{array} \quad \bar{x} = \tilde{x} = 3, \quad s \approx 1.5 \text{ (groß)}$$

$$\begin{array}{c|ccccc} x_j & 1 & 2 & 3 & 4 & 5 \\ \hline H_j & 0 & 1 & 8 & 1 & 0 \end{array} \quad \bar{x} = \tilde{x} = 3, \quad s \approx 0.67 \text{ (klein)}$$

Die Wichtigkeit der Stichprobenvarianz ergibt sich erst bei der Parameterschätzung und Anpassung von Verteilungen. Mit der Stichprobe allein kann man ja keine *schließende Statistik* machen. Dies geht erst nach Anpassung eines wahrscheinlichkeitstheoretischen Modells.

Vorschau: Wie die Anpassung mit Hilfe des Stichprobenmittels und der Stichprobenvarianz vorgenommen wird, können wir beispielhaft anhand des Modells einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ vorführen. Die Normalverteilung wird durch die Parameter μ (Erwartungswert) und σ^2 (Varianz) gesteuert.


 Abbildung 2.9: Dichte der Normalverteilung: Bedeutung der Parameter μ und σ .

Nach der *Momentenmethode* erfolgt die Anpassung durch Schätzung aus den entsprechenden Stichprobenparametern, also $\mu \approx \bar{x}$, $\sigma \approx s$. Im Falle der Temperaturen im Jänner 2001 erhält man die angepasste Normalverteilung $\mathcal{N}(0.76, 9.1085)$. Die Anpassungsgüte ist zufriedenstellend, wie aus dem Diagramm der kumulativen Verteilungsfunktionen ablesbar.

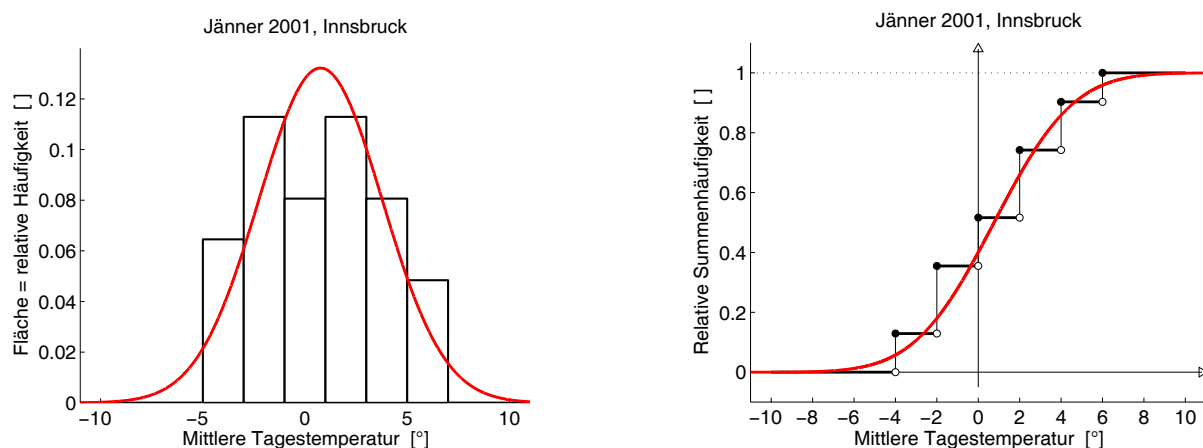


Abbildung 2.10: Temperaturdaten und angepasste Normalverteilung, Jänner 2001.

Als erste Anwendung können wir damit Wahrscheinlichkeiten für Temperaturbereiche näherungsweise berechnen (unter der Annahme der Gültigkeit des Normalverteilungsmodells), welche aus dem Histogramm nicht ablesbar sind.

Beispiel: Die Wahrscheinlichkeit, dass im Jänner eine Temperatur größer als 5° auftritt, ist nach dem Modell gleich 0.08, also 8%, und kann als Flächeninhalt der rot schraffierten Fläche abgelesen werden (eine robuste langjährige Aussage würde verlangen, Temperaturdaten im Jänner einer größeren Zahl von Jahren zu verwenden).

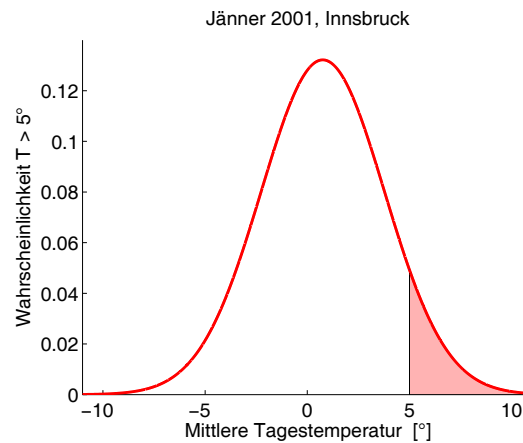


Abbildung 2.11: Darstellung einer Überschreitenswahrscheinlichkeit.

Ein Blick auf die kumulativen Verteilungsfunktionen zeigt übrigens, dass das Normalverteilungsmodell für Jänner 2002 nicht adäquat ist:

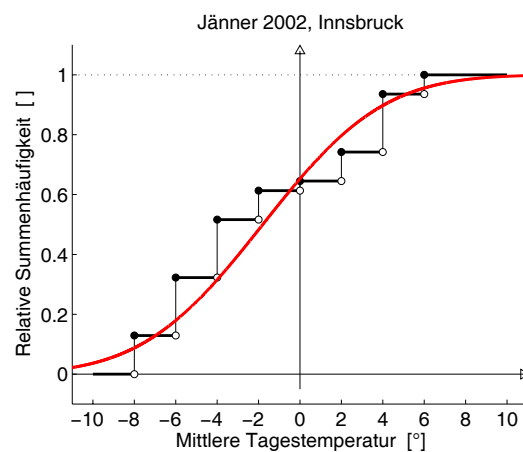


Abbildung 2.12: Summenhäufigkeit und angepasste Normalverteilung, Temperaturdaten Jänner 2002.

Tabellen

Datum	Temperatur 2001	Temperatur 2002	Bedeckung 2002
1	-3.4	-6.8	13
2	3.7	-7.1	0
3	2.8	-7.8	0
4	1.5	-8.2	10
5	4.0	-7.1	27
6	5.8	-5.7	10
7	5.0	-5.0	27
8	1.3	-4.5	10
9	0.4	-4.4	7
10	-0.4	-5.1	27
11	2.1	-6.0	73
12	1.8	-4.1	53
13	0.3	-2.7	33
14	-2.8	-1.5	10
15	-3.4	-5.5	10
16	-4.5	-4.5	33
17	-4.5	-2.8	7
18	-1.5	-4.4	90
19	-1.1	-4.7	73
20	-0.2	0.9	70
21	-0.6	2.4	53
22	2.1	1.9	57
23	4.3	3.0	97
24	5.8	5.2	53
25	5.2	5.6	50
26	3.9	4.1	97
27	1.3	4.3	100
28	-1.2	4.3	30
29	-1.6	4.1	40
30	-1.5	3.2	73
31	-1.0	3.2	40

Tabelle 2.1: Mittlere Tagestemperaturen [$^{\circ}$] und Bewölkungsgrad [%]. Innsbruck, Jänner 2001 und 2002 (aus: <http://www.zamg.ac.at>).

Kapitel 3

Beschreibende Statistik mehrdimensionaler Daten

3.1 Gemeinsame und Randhäufigkeiten

Aufgabenstellung: Es geht um die Analyse der gemeinsamen Verteilung von zwei oder mehreren Größen; wir beschränken uns hier auf den zweidimensionalen Fall und die zugehörigen Methoden der beschreibenden Statistik. Eine solche Situation tritt etwa in der Sozialstatistik auf, wenn jedes Individuum hinsichtlich zweier Merkmale (X, Y) untersucht wird, oder im Laborversuch, in dem zwei Größen (X, Y) gemessen werden. Ziel ist die Analyse der Abhängigkeit der beiden Größen. Beispiel: X = Schulnote in Mathematik, Y = Schulnote in Englisch; X = Durchmesser, Y = Druckfestigkeit von Steinsalz Körnern; X = Tagestemperatur, Y = Bewölkungsgrad.

Die Ausprägungen der Größe X seien x_1, \dots, x_k , die von Y seien y_1, \dots, y_l . Ein Messwert (x_i, y_j) stellt also einen Punkt in der (x, y) -Ebene dar. Die Menge $\{(x_i, y_j) : i = 1, \dots, k, j = 1, \dots, l\}$ bezeichnet man als *bivariaten Datensatz*. Es interessieren die gemeinsamen absoluten und relativen Häufigkeiten

$$H_{ij} = H(X = x_i, Y = y_j), \quad h_{ij} = \frac{1}{n} H_{ij}$$

beim Stichprobenumfang n ; bei Klasseneinteilung in X - bzw. Y -Klassen auch

$$H_{ij} = H(X \in X\text{-Klasse } i, Y \in Y\text{-Klasse } j).$$

Beispiel 3.1 Schülernoten, $n = 31$, X = Note Mathematik, Y = Note Englisch.

$X \setminus Y$	1	2	3	4	5	
1	2				1	3
2	4	2	1			7
3	2	5	2	1		10
4	4	4	1			9
5		1			1	2
	12	12	4	1	2	31

Dabei werden in der i -ten Zeile, j -ten Spalte aufgetragen:

$$H_{ij} = H(\text{Mathematik} = i, \text{Englisch} = j) = H(X = x_i, Y = y_j).$$

Die Randspalte bzw. Randzeile stellt die so genannte *Randverteilung* dar:

i -te Zeile:

$$H_{i.} = H(X = x_i, Y \text{ beliebig}) = \sum_{j=1}^l H_{ij}$$

j -te Spalte:

$$H_{.j} = H(X \text{ beliebig}, Y = y_j) = \sum_{i=1}^k H_{ij}$$

Im Allgemeinen hat eine zweidimensionale Häufigkeitstabelle die folgende Form:

$X \setminus Y$	y_j	
	\vdots	
x_i	$\dots H_{ij} \dots$	$H_{i.} \quad (\sum_j)$
	\vdots	
	$H_{.j}$	n
		(\sum_i)

Analog kann man die relativen Häufigkeiten darstellen:

$$h_{ij} = \frac{1}{n} H_{ij}, \quad h_{i.} = \frac{1}{n} H_{i.}, \quad h_{.j} = \frac{1}{n} H_{.j}$$

Die Randverteilungen geben sozusagen das Verhalten der Messgrößen X, Y einzeln wieder. Gegenüberstellung der Verteilungen im Boxplot:

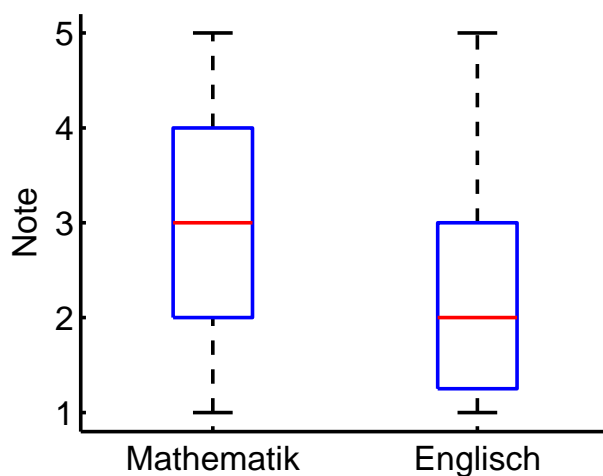


Abbildung 3.1: Boxplots für Notenbeispiel.

Berechnung der Parameter der Größen X und Y :

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^k x_i H_{i\cdot} &= 3 & \quad \bar{y} &= \frac{1}{n} \sum_{j=1}^l y_j H_{\cdot j} &= 2 \\ s_x^2 &= \frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 H_{i\cdot} &= \frac{36}{30} & \quad s_y^2 &= \frac{1}{n-1} \sum_{j=1}^l (y_j - \bar{y})^2 H_{\cdot j} &= \frac{38}{30}\end{aligned}$$

Was kann noch über die gemeinsame Verteilung gesagt werden? Es interessieren die wechselseitigen Abhängigkeiten. Zunächst der Extremfall der *Unabhängigkeit*. Dies sollte bedeuten, dass – ganz gleich, was der Wert für X ist – die Verteilung von Y stets dieselbe ist, und umgekehrt. Daraus folgt aber, dass die Zeilen und Spalten in der Häufigkeitsmatrix proportional sein müssen.

Beispiel 3.2 Ein Beispiel zweier unabhängiger Merkmale:

$X \setminus Y$	1	2	3	4	5	
1	1	2			1	4
2	3	6			3	12
3	2	4			2	8
4						0
5	1	2			1	4
	7	14	0	0	7	28

Insbesondere muss jede Zeile ein Vielfaches der Randverteilungszeile sein. Es gilt dann etwa für die erste Zeile:

$$h_{1j} = c \cdot h_{\cdot j}, j = 1, \dots, l.$$

Summation über j ergibt

$$h_{1\cdot} = \sum_{j=1}^l h_{1j} = \sum_{j=1}^l c \cdot h_{\cdot j} = c \cdot 1 = c,$$

und es folgt

$$h_{1j} = h_{1\cdot} \cdot h_{\cdot j}, j = 1, \dots, l$$

Das Entsprechende gilt für alle Zeilen. Damit ist geklärt, dass für die Unabhängigkeit notwendig und hinreichend ist, dass die relativen Häufigkeiten die Produkte der jeweiligen Randhäufigkeiten sind:

$$h_{ij} = h_{i\cdot} \cdot h_{\cdot j}, i = 1, \dots, k, j = 1, \dots, l.$$

Es ist klar, dass die Unabhängigkeit eine sehr starke Eigenschaft ist (bei Schülernoten wohl kaum jemals zutreffend, wie das erste Notenbeispiel zeigt). Es werden abgestufte Begriffe benötigt.

3.2 Kovarianz und Korrelation

Betrachten wir die Produkte der Abweichungen vom jeweiligen Mittelwert

$$(x_i - \bar{x})(y_j - \bar{y}).$$

Diese sind positiv, wenn gleichzeitig $x_i > \bar{x}, y_j > \bar{y}$ oder $x_i < \bar{x}, y_j < \bar{y}$ ist. Sie sind negativ, wenn $x_i > \bar{x}, y_j < \bar{y}$ oder $x_i < \bar{x}, y_j > \bar{y}$ ist. Das Ergebnis für das Schülernotenbeispiel (Beispiel 3.1) steht in der Tabelle unten (in Klammern sind die Häufigkeiten $\neq 0$ angegeben).

	y_i	1	2	3	4	5
	$y_j - \bar{y}$	-1	0	1	2	3
x_i	$x_i - \bar{x}$					
1	-2	2 (2)	0	-2	-4	-6 (1)
2	-1	1 (4)	0 (2)	-1 (1)	-2	-3
3	0	0 (2)	0 (5)	0 (2)	0 (1)	0
4	1	-1 (4)	0 (4)	1 (1)	2	3
5	2	-2	0 (1)	2	4	6 (1)

Im Falle der “Unkorreliertheit” sollten sich diese Produkte in Summe weg heben. Ist dies nicht der Fall, so treten also tendenziell Abweichungen von den Mittelwerten in derselben Richtung auf.

Die *empirische Kovarianz* ist definiert als

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^l (x_i - \bar{x})(y_j - \bar{y}) H_{ij}$$

Wenn $s_{xy} = 0$ ist, so heißen X, Y *unkorreliert*, andernfalls *korreliert*. Im Schülernotenbeispiel ist

$$s_{xy} = 4/30 = 0.1333,$$

die Leistungen in Englisch und Mathematik sind korreliert.

Um ein Maß für die Korrelation zu erhalten, das die Größenordnung der gesamten Schwankungsbreite herauskaliert, wird der *Pearson-Korrelationskoeffizient*

$$R = r_{xy} = \frac{s_{xy}}{s_x s_y}$$

eingeführt. Im Notenbeispiel ist

$$R = \frac{4/30}{\sqrt{36/30} \sqrt{38/30}} = 0.1081 \approx 0.11,$$

es liegt also eine schwach positive Korrelation vor. Die Extremfälle $R = 1$ bzw. $R = -1$ erhält man, wenn Y eine lineare Funktion von X mit positivem bzw. negativem Anstieg ist.

Bemerkung: *Unabhängige Größen sind unkorreliert.* Wir werden dies im Kapitel 5 beweisen. Die Umkehrung gilt jedoch nicht.

Beispiel 3.3 Ein Beispiel eines bivariaten Datensatzes, der unkorreliert, aber nicht unabhängig ist:

$X \setminus Y$	1	2	3	4	5	
1		2			1	3
2						0
3	4	2	9	4	3	22
4		4			2	6
5						0
	4	8	9	4	6	31

Beispiel 3.4 Wir betrachten den Bewölkungsgrad X (in %) und die mittlere Tagestemperatur Y (in °) im Jänner 2002 in Innsbruck (Tabelle 3.1). Die Abbildung zeigt das so genannte *Streudiagramm* der beiden Messgrößen.

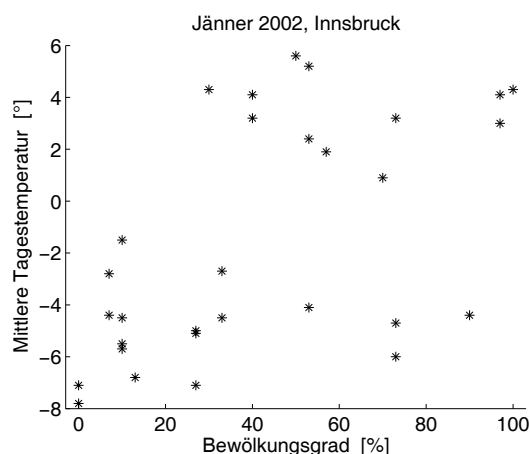


Abbildung 3.2: Streudiagramm Bewölkungsgrad/Temperatur, Innsbruck, Jänner 2002.

Offensichtlich gibt es eine Tendenz zu höheren Temperaturen bei höherem Bewölkungsgrad. Um diese Tendenz zu untermauern bzw. zu quantifizieren, berechnen wir den Pearson-Korrelationskoeffizienten. Sein Wert ist

$$R = 0.569$$

und belegt eine deutliche Korrelation im positiven Sinn.

Beispiel 3.5 Altersabhängige Unfallshäufigkeiten. Die Tabelle unten enthält die im Innsbrucker Straßenverkehr im Jahre 1992 Verunfallten nach Altersgruppe (Quelle: Statistisches Jahrbuch der Stadt Innsbruck 1992).

Altersgruppe	Gruppenmitte	Verunfallte	Gesamtzahl	Anteil
15 - 25	20	442	19296	0.02291
25 - 35	30	367	21492	0.01708
35 - 45	40	166	15671	0.01059
45 - 55	50	203	15621	0.01299
55 - 65	60	81	10557	0.00767
65 -	75	113	19083	0.00592

Um einen etwaigen Zusammenhang feststellen zu können, vergleichen wir die Größen X = Gruppenmitte und Y = Anteil. Dazu das Streubild:

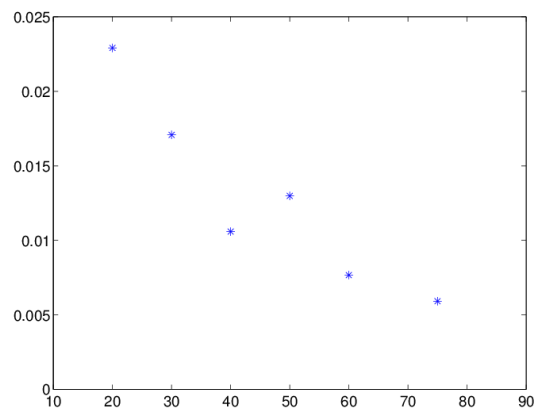


Abbildung 3.3: Altersgruppe gegen Unfallhäufigkeit.

Die statistischen Kennwerte sind:

$$s_x = 20.1039, \quad s_y = 0.0063, \quad s_{xy} = -0.1177, \quad r_{xy} = -0.9293.$$

Es liegt also tatsächlich eine stark negative Korrelation vor.

Beispiel 3.6 Aus der Geotechnik – Reibungswinkel und Kohäsion. Zur Ermittlung der Scherfestigkeit von Geschiebemergel wurden an vierzehn aus unterschiedlichen Bohrungen stammenden Sondierproben Kreisringversuche durchgeführt. Die Ergebnisse sind in der Tabelle unten angeführt (σ = Normalspannung [kPa], τ = Grenzscherspannung [kPa]; aus: B. Schuppener, Die Festlegung charakteristischer Bodenkennwerte - Empfehlungen des Eurocodes 7 Teil 1 und die Ergebnisse einer Umfrage. Geotechnik, Sonderheft 1999).

σ	150	200	300	100	150	250	300	100	150	250	100	150	200	250
τ	130	135	205	70	125	145	195	75	80	170	70	100	125	160

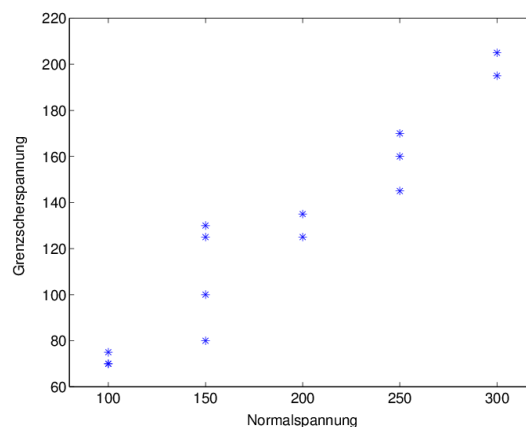


Abbildung 3.4: Normalspannung gegen Grenzscherspannung im Kreisringversuch.

Das Streubild zeigt eine deutliche Tendenz. Die statistischen Kennwerte sind:

$$s_x = 71.1947, \quad s_y = 44.8823, \quad s_{xy} = 3048.1, \quad r_{xy} = 0.9539.$$

Es besteht also tatsächlich eine stark positive Korrelation zwischen σ und τ . Dies ist im Einklang mit der im Mohr-Coulombschen Stoffgesetz postulierten linearen Beziehung $\tau = \sigma \tan(\varphi) + c$ (Reibungswinkel φ , die Kohäsion c).

3.3 Rangkorrelation

Die *Rangstatistik* setzt sich zum Ziel, den Einfluss der Skalenwahl zu entfernen, also zu erreichen, dass die statistischen Indikatoren nicht von den absoluten Zahlenwerten der Messungen abhängen. Dies erfolgt, indem die Messwerte durch ihre Rangnummern nach Größenordnung ersetzt werden. Die Ränge etwa der Messgrößen Bewölkungsgrad und Temperatur sind in Tabelle 3.1 berechnet worden. Nicht ganzzahlige Ränge ergeben sich durch Mittelbildung der Ränge bei Auftreten von *Bindungen* (gleichen Messwerten).

Die Vorgangsweise gewinnt erst bei zwei- und mehrdimensionalen Messgrößen an Bedeutung. Abhängigkeiten werden dann allein durch Beziehungen zwischen den Rängen dargestellt und können durch den *Spearman-Rangkorrelationskoeffizienten* ρ bewertet werden. Der Spearman-Koeffizient ρ ist nichts anderes als der Pearson-Korrelationskoeffizient R , berechnet aus den Rängen des Datensatzes. Er beträgt hier

$$\rho = 0.570$$

und unterscheidet sich in diesem Beispiel nicht wesentlich vom Pearson-Koeffizienten.

Tabellen

Bedeckung	Temperatur	Rang-Bedeckung	Rang-Temperatur
13	-6.8	10.0	5.0
0	-7.1	1.5	3.5
0	-7.8	1.5	2.0
10	-8.2	7.0	1.0
27	-7.1	12.0	3.5
10	-5.7	7.0	7.0
27	-5.0	12.0	10.0
10	-4.5	7.0	12.5
7	-4.4	3.5	14.5
27	-5.1	12.0	9.0
73	-6.0	26.0	6.0
53	-4.1	21.0	16.0
33	-2.7	15.5	18.0
10	-1.5	7.0	19.0
10	-5.5	7.0	8.0
33	-4.5	15.5	12.5
7	-2.8	3.5	17.0
90	-4.4	28.0	14.5
73	-4.7	26.0	11.0
70	0.9	24.0	20.0
53	2.4	21.0	22.0
57	1.9	23.0	21.0
97	3.0	29.5	23.0
53	5.2	21.0	30.0
50	5.6	19.0	31.0
97	4.1	29.5	26.5
100	4.3	31.0	28.5
30	4.3	14.0	28.5
40	4.1	17.5	26.5
73	3.2	26.0	24.5
40	3.2	17.5	24.5

Tabelle 3.1: Bewölkungsgrad, mittlere Tagestemperaturen und deren Ränge. Innsbruck, Jänner 2002.

Kapitel 4

Eindimensionale Zufallsgrößen

4.1 Der mathematische Wahrscheinlichkeitsbegriff

In diesem Kapitel werden wir das mathematische Modell der Wahrscheinlichkeit einführen. Die Begriffe *Grundgesamtheit*, *Merkmal*, *relative Häufigkeit*, *Mittelwert*, *Stichprobenvarianz*, *Summenhäufigkeit* usw. finden ihre Entsprechung in den Begriffen *Wahrscheinlichkeitsraum*, *Zufallsgröße*, *Wahrscheinlichkeit*, *Erwartungswert*, *Varianz*, *Verteilungsfunktion* usw.

Beginnen wir mit dem Begriff einer Zufallsgröße. Eine *Zufallsgröße* oder *Zufallsvariable* ist eine Größe, die zufällig verteilte Werte annimmt. Die *Verteilung* muss spezifiziert werden und gibt an, welche Werte mit welcher Wahrscheinlichkeit auftreten.

Eine Zufallsgröße X heißt *diskret*, wenn ihr Wertebereich eine diskrete Menge von reellen Zahlen $\{x_1, x_2, x_3, \dots\}$ ist; sie heißt *kontinuierlich*, wenn sie beliebige reelle Werte annehmen kann.

Schreibweise: Für die Zufallsgröße verwenden wir Großbuchstaben, etwa X , für ihre Realisierung (= aufgetretener Wert) den entsprechenden Kleinbuchstaben x .

Ereignisse: Wir sprechen auch hier wieder von Ereignissen. Ein Elementarereignis ist zum Beispiel $X = x_k$, also das Auftreten des Wertes x_k . Zusammengesetzte Ereignisse sind etwa $X \in \{x_1, x_2, x_3\}$ oder $a \leq X \leq b$.

Wahrscheinlichkeit: Jedem Ereignis wird eine Wahrscheinlichkeit P zugeordnet (P von *probability* aus dem Englischen), eine Zahl zwischen 0 und 1. In der frequentistischen Interpretation der Wahrscheinlichkeit, auf die wir uns hier beziehen, entspricht die Wahrscheinlichkeit eines Ereignisses der relativen Häufigkeit seines Auftretens in einer Stichprobe.

Die *Verteilung* einer diskreten Zufallsgröße ist durch die Angabe der Punktwahrscheinlichkeiten

$$p_k = P(X = x_k), \quad k = 1, 2, 3, \dots$$

spezifiziert. Kontinuierliche Zufallsgrößen X werden durch eine Dichtefunktion $f_X(x)$ beschrieben. Diese erlaubt es, die Intervallwahrscheinlichkeiten

$$P(a \leq X \leq b) = \int_a^b f_X(x) \, dx$$

als Integral anzugeben. Dabei muss $f(x) \geq 0$ und $\int_{-\infty}^{\infty} f(x) \, dx = 1$ sein. Daneben gibt es noch gemischte Verteilungen mit diskretem und kontinuierlichem Anteil.

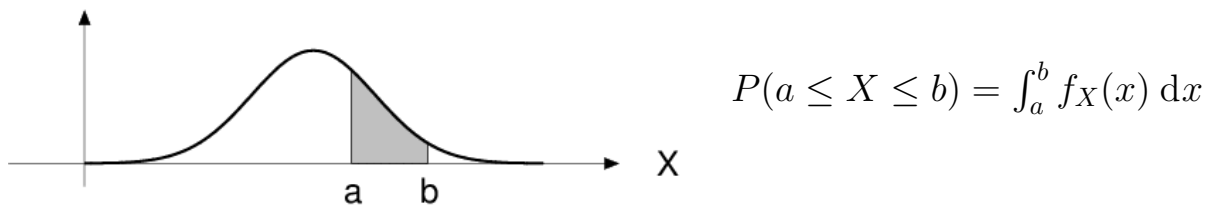


Abbildung 4.1: Intervallwahrscheinlichkeit für eine kontinuierliche Zufallsgröße.

Beispiel: Es bezeichne X die Augenzahl eines idealen Würfels. Der mögliche Wertebereich ist $X \in \{1, 2, 3, 4, 5, 6\}$. Alle Augenzahlen werden als gleich wahrscheinlich betrachtet. Die Punktwahrscheinlichkeiten sind demnach

$$P(X = 1) = \frac{1}{6}, \quad P(X = 2) = \frac{1}{6}, \quad \dots, \quad P(X = 6) = \frac{1}{6}.$$

Als Beispiel eines zusammengesetzten Ereignisses nehmen wir etwa

$$P(X = 1 \text{ oder } X = 2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Die Wahrscheinlichkeit der Vereinigung von Ereignissen, die sich gegenseitig ausschließen, ist die Summe der Wahrscheinlichkeiten.

4.2 Axiome der Wahrscheinlichkeit

Wir wollen nun die *Axiome der Wahrscheinlichkeit* etwas formaler einführen. Man geht von einer Grundmenge Ω aus (dem Wahrscheinlichkeitsraum) und sondert ein System von Teilmengen Σ aus, die Ereignisse (formal auch als *messbare Mengen* bezeichnet) und betrachtet eine Funktion

$$P : \Sigma \longrightarrow [0, 1].$$

P ordnet also jeder Teilmenge A von Ω , welche zu Σ gehört, eine Zahl $P(A)$ zwischen 0 und 1 zu. In der *statischen* Sprechweise ist $P(A)$ die *Wahrscheinlichkeit* oder das *Wahrscheinlichkeitsmaß* der Menge A . In der *dynamischen* Sprechweise denkt man mehr an das Auftreten eines Wertes $\omega \in \Omega$ (etwa als Ausgang einer Messung oder eines Experimentes) und spricht von dem *Ereignis*, dass ω in A zu liegen kommt. Dann ist $P(A) = P(\omega \in A)$ die Wahrscheinlichkeit dieses Ereignisses.

Wir besprechen zunächst das Verknüpfen von Ereignissen, die *Ereignisalgebra*.

Ereignisalgebra: Wir bezeichnen Ereignisse mit A, B, C usw.; wichtig sind die folgenden Kombinationen von Ereignissen:

$$A \cup B = A \text{ oder } B$$

$$A \cap B = A \text{ und } B$$

$$A^c = \text{das Komplementäreignis zu } A$$

Beispiel Würfeln: Wir betrachten etwa die Ereignisse $A : X \in \{1, 2\}$ und $B : X \in \{2, 3\}$. Dann gilt:

$$A \cup B : X \in \{1, 2, 3\}$$

$$A \cap B : X \in \{2\}, \text{ d.h. } X = 2$$

$$A^c : X \in \{3, 4, 5, 6\}$$

$$\emptyset : \text{unmögliches Ereignis}$$

$$\Omega : \text{stets eintretendes Ereignis, also } X \in \{1, 2, 3, 4, 5, 6\}$$

$$A \cap B = \emptyset : \text{disjunkte Ereignisse (schließen sich gegenseitig aus).}$$

Axiome der Wahrscheinlichkeit: Eine Mengenfunktion $P : \Sigma \rightarrow [0, 1]$ wird als *Wahrscheinlichkeitsmaß* bezeichnet, wenn sie folgende Eigenschaften hat:

$$P(\emptyset) = 0, \quad P(\Omega) = 1,$$

$$A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B).$$

Wenn Σ aus unendlich vielen Mengen besteht, ist allgemeiner zu fordern:

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

sofern die Mengen A_i zu Σ gehören und paarweise leeren Durchschnitt haben; dabei sind endliche und abzählbar unendliche Vereinigungen zugelassen. Es folgt dann zum Beispiel:

$$P(A^c) = 1 - P(A)$$

da $A \cap A^c = \emptyset$, $A \cup A^c = \Omega$ und damit $P(A) + P(A^c) = 1$ ist.

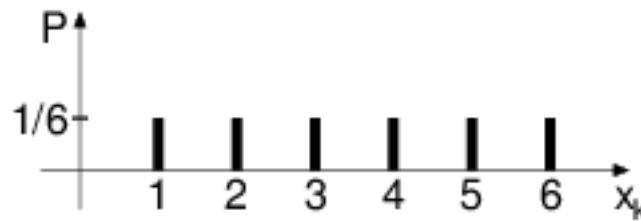
Eine *Zufallsgröße* wird formal präzisiert als eine Abbildung

$$X : \Omega \longrightarrow \mathbb{R}$$

mit der Eigenschaft, dass die Urbilder aller Abschnitte $(-\infty, a]$, $a \in \mathbb{R}$, Ereignisse sind (also zum ausgesonderten System von Teilmengen Σ gehören). In der Praxis arbeitet man allerdings kaum je mit dem zu Grunde liegenden Raum Ω , sondern definiert die Punkt- oder Intervallwahrscheinlichkeiten der auftretenden Werte direkt.

4.3 Beispiele für diskrete Zufallsgrößen

Beispiel 4.1 (Die diskrete Gleichverteilung) Eine gleichverteilte diskrete Zufallsgröße X nimmt k Werte x_1, x_2, \dots, x_k mit gleicher Wahrscheinlichkeit $P(X = x_k) = 1/k$ an. Ein typisches Beispiel ist der ungezinkte Würfel mit $k = 6$.

Abbildung 4.2: Diskrete Gleichverteilung, $k = 6$.

Beispiel 4.2 (Die Alternativverteilung) Es handelt sich um die diskrete Verteilung mit zwei Ausgängen $x_0 = 0, x_1 = 1$. Sie ist vollständig charakterisiert durch die *Erfolgswahrscheinlichkeit*

$$P(X = 1) = p,$$

da dann notwendigerweise

$$P(X = 0) = 1 - p$$

ist. Die Darstellung als Säulendiagramm ist aus der Abbildung unten ersichtlich.

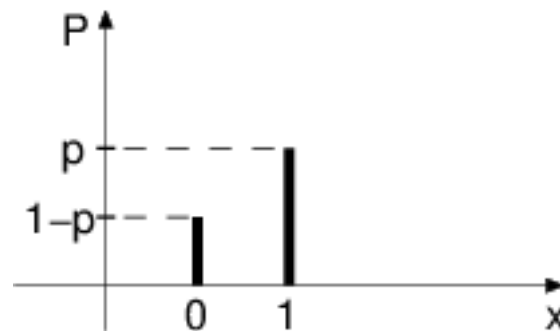


Abbildung 4.3: Wahrscheinlichkeitsfunktion einer Alternativverteilung.

Weitere Anwendungsbeispiele für Alternativverteilungen sind der Münzwurf und Umfragen mit zwei Antwortmöglichkeiten:

Münzwurf: $p = \frac{1}{2}$

Umfrage: $p \dots$ ja-Anteil; $1 - p \dots$ nein-Anteil

Beispiel 4.3 (Die Binomialverteilung) Diese diskrete Verteilung ergibt sich durch n -malige Beobachtung einer alternativverteilten Größe. Ihre Werte sind $x_0 = 0, x_1 = 1, \dots, x_n = n$. Sie gibt die Anzahl der Erfolge (Ausgang 1) bei n Beobachtungen an.

Im Würfelbeispiel ist ein Erfolg das Werfen eines Sechсers; dessen Wahrscheinlichkeit ist $p = 1/6$. Misserfolg tritt sonst auf; die Wahrscheinlichkeit dafür ist $1 - p = 5/6$. Bei

zweimaligem Würfeln ($n = 2$) ergeben sich folgende Wahrscheinlichkeiten:

$$P(X = 2) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

$$P(X = 1) = \frac{1}{6} \cdot \frac{5}{6} + \frac{5}{6} \cdot \frac{1}{6} = \frac{10}{36}$$

$$P(X = 0) = \frac{5}{6} \cdot \frac{5}{6} = \frac{25}{36}$$

Allgemein, bei beliebigem n , ist die Wahrscheinlichkeit einer Serie aus k Erfolgen und $n - k$ Misserfolgen gerade gleich $p^k(1 - p)^{n-k}$. Die Anzahl derartiger Serien entspricht der Anzahl der Möglichkeiten, k Elemente aus n auszuwählen und ist daher gleich

$$\binom{n}{k} = \frac{n(n-1) \cdots (n-k+1)}{1 \cdot 2 \cdots k}.$$

Somit erhält man für die Wahrscheinlichkeit von k Erfolgen unter n Versuchen

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

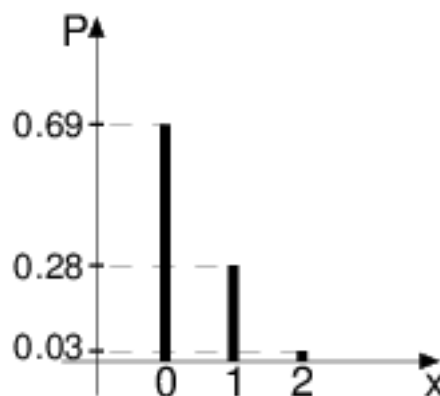


Abbildung 4.4: Binomialverteilung mit $n = 2$, $p = 1/6$.

Wir werden im Abschnitt 4.5 herleiten, dass eine binomialverteilte Zufallsgröße im Mittel den Wert $\mu = np$ annimmt (genauer: als Erwartungswert besitzt).

Beispiel 4.4 (Die Poissonverteilung) Ist von einem wiederkehrenden Ereignis eine *mittlere Rate* bekannt, so erlaubt die Poissonverteilung, eine Aussage über die Zahl der zu erwartenden Eintritte zu machen. Man erhält die Poissonverteilung durch Grenzübergang $n \rightarrow \infty$ in der Binomialverteilung unter Konstanthalten von $\mu = np$.

Zur Herleitung diene das Beispiel der Wahrscheinlichkeit von Verbrüchen im Tunnelvortrieb. Nehmen wir an, eine Geologin oder ein Geologe erwarte μ Verbrüche pro Kilometer Tunnel (also eine Rate). Was ist die Wahrscheinlichkeit, dass auf einer Länge von einem Kilometer 0, 1, 2, 3, ... Verbrüche auftreten?

Stellen wir uns vor, dass der Tunnel in Sektionen der Länge 100 m unterteilt ist. Die mittlere Anzahl von Verbrüchen in einer Sektion von 100 Metern ist nach Angabe dann $\mu/10$. Ein Verbruch kann in der ersten 100-Meter-Sektion oder in der zweiten 100-Meter-Sektion oder in der dritten auftreten, usw. Das Ergebnis ist eine Binomialverteilung für die Anzahl der „Erfolge“ in den 100-Meter-Sektionen

$$P(X = k) = \binom{10}{k} \left(\frac{\mu}{10}\right)^k \left(1 - \frac{\mu}{10}\right)^{10-k}.$$

Wir wiederholen die Rechnung mit n Sektionen und erhalten

$$P(X = k) = \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}.$$

Im Grenzwert ergibt sich

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} &= \lim_{n \rightarrow \infty} \frac{\mu^k}{k!} \frac{n(n-1) \cdots (n-k+1)}{n \cdot n \cdots n} \left(1 - \frac{\mu}{n}\right)^{-k} \left(1 - \frac{\mu}{n}\right)^n \\ &= \frac{\mu^k}{k!} \cdot 1 \cdot 1 \cdots 1 \cdot e^{-\mu} = \frac{\mu^k}{k!} e^{-\mu}. \end{aligned}$$

Bei einer Rate von μ Verbrüchen pro Kilometer ergibt sich die Wahrscheinlichkeit dafür, dass tatsächlich k Verbrüche auf einem Kilometer erfolgen, zu

$$P(X = k) = \frac{\mu^k}{k!} e^{-\mu},$$

$k = 0, 1, 2, 3, \dots$ Das ist die allgemeine Formel einer Poissonverteilung mit Parameter μ .

Beispiel einer Poissonverteilung mit $\mu = 0.25$: $P(X = 0) = 0.78$, $P(X = 1) = 0.195$, $P(X = 2) = 0.024$, $P(X = 3) = 0.002 \dots$

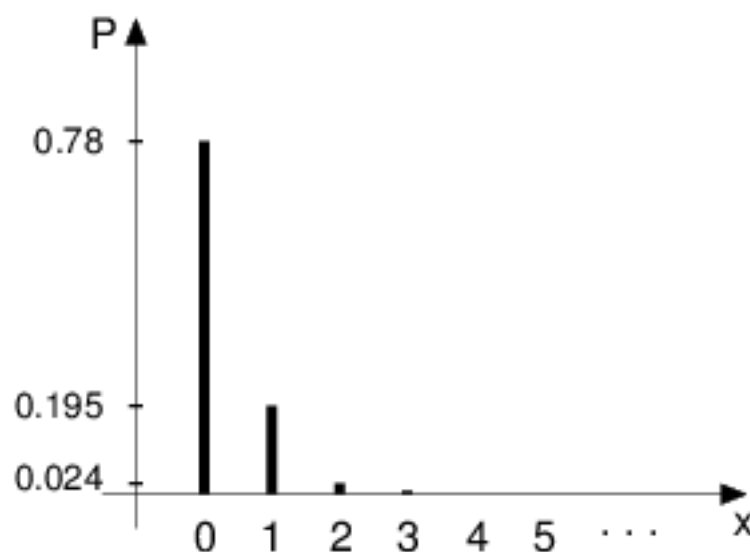


Abbildung 4.5: Poissonverteilung mit Parameter $\mu = 0.25$.

Bemerkung: Wir schreiben oft

$$p_k = P(X = x_k)$$

Dabei muss gelten: $\sum_k p_k = 1$, da dies die Wahrscheinlichkeit des sicheren Ereignisses ist (der Vereinigung aller möglichen Ergebniswerte).

Beispiel 4.5 (Die geometrische Verteilung) Sie beschreibt die Anzahl der Misserfolge bis zum ersten Erfolg bei Wiederholung eines 0-1-Versuches. Es gilt:

$$P(X = k) = (1 - p)^k p, \quad p = 0, 1, 2, 3, \dots$$

Beispiel 4.6 (Die hypergeometrische Verteilung) Sie ist eine typische Verteilung für die *Qualitätskontrolle*. Unter N Bauteilen seien M defekt ($0 \leq M \leq N$). Wir entnehmen eine Probe von K Stück. X bezeichne die Anzahl der defekten Teile unter den K entnommenen. Dann gilt:

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{K-k}}{\binom{N}{K}}, \quad 0 \leq k \leq K$$

Die Entnahme geschieht dabei ohne Zurücklegen.

Würde man jedes entnommene Teil wieder zurücklegen, so entspräche das der K -maligen Wiederholung eines Versuches mit Erfolgswahrscheinlichkeit $p = \frac{M}{N}$, also einer Binomialverteilung $P(X = k) = \binom{K}{k} \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{N}\right)^{1-k}$.

4.4 Beispiele für kontinuierliche Zufallsgrößen

Beispiel 4.7 (Die Normalverteilung) Nach dem zentralen Grenzwertsatz sind Summen unabhängiger Zufallsgrößen asymptotisch normalverteilt (das heißt, mit wachsender Anzahl der Summanden nähert sich die Verteilung der Summe einer Normalverteilung). Insbesondere sind die Mittelwerte von Stichproben beliebiger Zufallsgrößen bei hinreichend großem Umfang näherungsweise normalverteilt. Die Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ wird durch die zwei Parameter μ und σ spezifiziert. Ihre Dichtefunktion ist die Gauß'sche Glockenkurve

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

Die Spitze des Graphen (der *Modalwert*) liegt bei $x = \mu$, die Wendepunkte bei $x = \mu \pm \sigma$.

Die Substitution $\frac{x-\mu}{\sqrt{2}\sigma} = y$ zeigt, dass die Gesamtfläche gleich Eins ist:

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 1.$$

Die Bedeutung des Streuparameters σ kann genau angegeben werden (Abbildung 4.7).

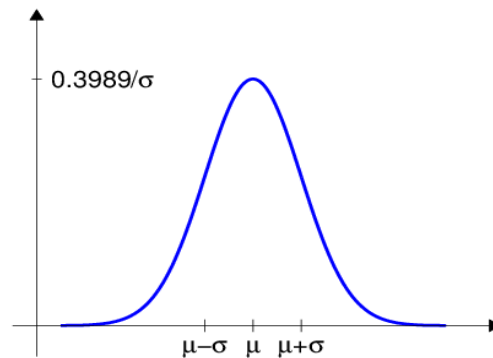


Abbildung 4.6: Gauß'sche Glockenkurve – Dichte der Normalverteilung.

$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$ (In 68% der Fälle liegt X zwischen $\mu - \sigma$ und $\mu + \sigma$)

$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$

$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9975$

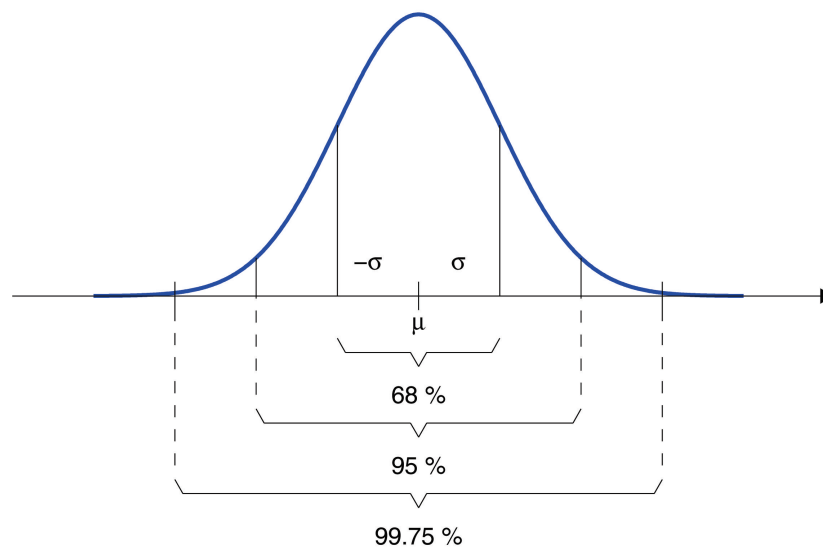


Abbildung 4.7: Gauß'sche Glockenkurve – Prozentanteile und Streuung.

$\mathcal{N}(0, 1)$ heißt *Standardnormalverteilung*. Das so genannte Zentrieren ermöglicht die Rückführung jeder normalverteilten Größe auf die Standardnormalverteilung:

$$X \sim \mathcal{N}(\mu, \sigma^2) \iff Y = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Die Standardnormalverteilung $\mathcal{N}(0, 1)$ mit $\mu = 0$ und $\sigma = 1$ ist tabelliert. Für die Herleitung der Umrechnung siehe Abschnitt 4.7.

Beispiel 4.8 (Kontinuierliche Gleichverteilung) X hat Werte in einem Intervall $[a, b]$. Dort sind diese „gleichmäßig verteilt“. Daraus ergibt sich die Dichtefunktion

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{sonst} \end{cases}$$

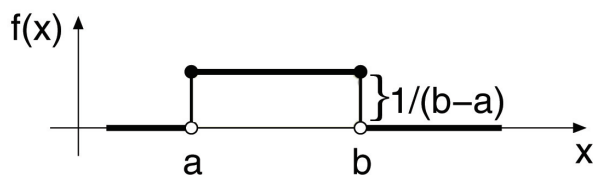


Abbildung 4.8: Kontinuierliche Gleichverteilung.

Beispiel 4.9 (Dreiecksverteilung) Die Dichte hat eine Dreiecksform:

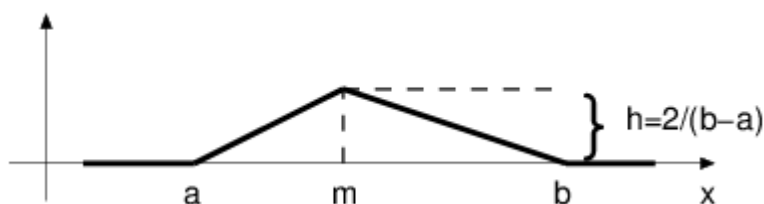


Abbildung 4.9: Dreiecksverteilung.

Aus der Forderung, dass die Gesamtfläche gleich Eins ist, ergibt sich $1 = h \cdot \frac{b-a}{2}$ und damit $h = \frac{2}{b-a}$.

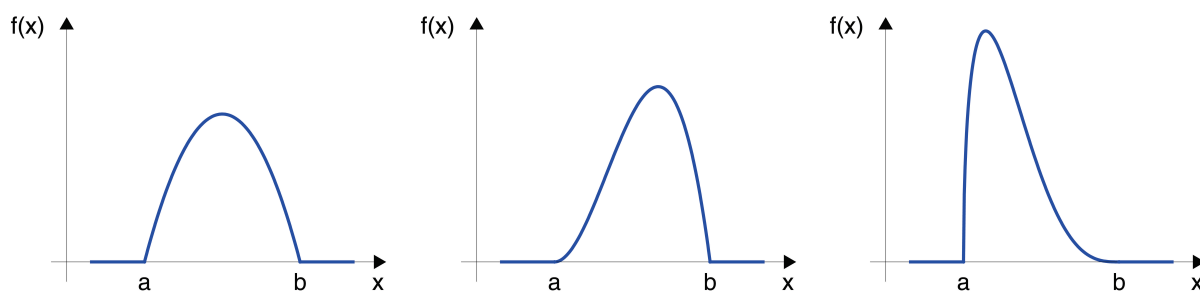
Beispiel 4.10 (Betaverteilung) Sie ist im Intervall $[a, b]$ konzentriert und hat eine (eventuell unsymmetrische) Hutform, Formparameter $r, s > 0$:

$$f(x) = \frac{(x-a)^{r-1}(b-x)^{s-1}}{(b-a)^{r+s-1}B(r,s)}, \quad a \leq x \leq b.$$

Die Zahl $B(r, s) = \int_0^1 y^{r-1}(1-y)^{s-1} dy$ wird – in Abhängigkeit von r und s – als *Betafunktion* bezeichnet. Sie muss hier eingeführt werden, damit $\int_a^b f(x) dx = 1$ ist: Mittels der Substitution $y = \frac{x-a}{b-a}$ gilt nämlich

$$\int_a^b (x-a)^{r-1}(b-x)^{s-1} dx = \int_0^1 y^{r-1}(1-y)^{s-1} dy \cdot (b-a)^{r-1+s-1+1} = (b-a)^{r+s-1} B(r, s).$$

Für ganzzahliges $r, s \geq 0$ ist das Integral explizit berechenbar. Sonst ist es eine höher transzendente Funktion von r, s und kann numerisch approximiert werden.

Abbildung 4.10: Betaverteilungen mit $r = 2, s = 2$; $r = 3, s = 2$ und $r = \frac{3}{2}, s = 4$.

Durch Nullsetzen der Ableitung erhält man die Lage des Modalwerts:

$$m = \frac{(s-1)a + (r-1)b}{r+s-2}.$$

Beispiel 4.11 (Die Exponentialverteilung) Diese einparametrische kontinuierliche Verteilung dient zur Modellierung von Lebenszeiten und Zeitlücken zwischen wiederholten zufälligen Ereignissen. Ihre Dichtefunktion ist

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0, \\ 0, & t < 0. \end{cases} \quad (\text{Parameter } \lambda \geq 0)$$

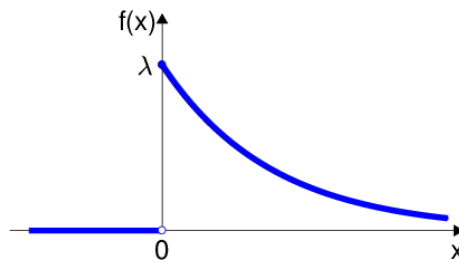


Abbildung 4.11: Dichte einer Exponentialverteilung.

Der Urtyp einer exponentialverteilten Größe ist die Lebensdauer T eines radioaktiven Atoms. Wir können die Verteilung aus dem radioaktiven Zerfallsgesetz herleiten. Es bezeichne $N(t)$ die Anzahl der Atome einer radioaktiven Substanz zum Zeitpunkt t . Die Zerfallsrate ist proportional zur jeweiligen aktuellen Zahl:

$$N'(t) = -\lambda N(t).$$

Die Lösung dieser Differentialgleichung ist

$$N(t) = N_0 e^{-\lambda t}.$$

Der Parameter λ kann aus der Halbwertszeit τ berechnet werden:

$$N(\tau) = \frac{1}{2} N_0, \quad N_0 e^{-\lambda \tau} = \frac{1}{2} N_0; \quad \lambda \tau = \ln 2, \quad \lambda = \frac{\ln 2}{\tau}.$$

Für das Einzelteilchen ergibt sich

$$\begin{aligned} P(T > t) &= P(\text{Teilchen noch nicht zerfallen zum Zeitpunkt } t) \\ &= \frac{\text{Anzahl nicht zerfallener Teilchen}}{\text{Anzahl aller Teilchen}} \\ &= \frac{N(t)}{N_0} = e^{-\lambda t}. \end{aligned}$$

Es ist aber $P(T > t) = \int_t^\infty f(s) ds$. Aus dem Hauptsatz der Integralrechnung folgt

$$f(t) = -\frac{d}{dt} \int_t^\infty f(s) ds = -\frac{d}{dt} e^{-\lambda t} = \lambda e^{-\lambda t}.$$

Beispiel 4.12 (Die Gumbelverteilung) Die *Gumbel-Verteilung* ist eine der Standardverteilungen zur Beschreibung von Extremwerten. Sie beschreibt näherungsweise die Verteilung des Stichprobenmaximums, also des größten Werts einer Stichprobe beliebigen Umfangs n . Das heißt, nimmt man N Stichproben vom Umfang n und ist X_j der größte Wert der Stichprobe Nummer j , so sind die X_1, \dots, X_N ungefähr Gumbel-verteilt. Die Gumbelverteilung wird zum Beispiel in der Hochwasserstatistik angewendet. Dabei wird etwa die Liste der Monatswerte für HQ eines Jahres als Stichprobe vom Umfang $n = 12$ betrachtet. Beobachtet man N Jahre, so erhält man für Jahr Nummer j das Jahresmaximum X_j . Diese Jahresmaxima können durch eine Gumbel-Verteilung näherungsweise beschrieben werden. Das 100-jährige Hochwasser erhält man als das 99%-Quantil der kumulativen Verteilungsfunktion, das 50-jährige als das 98%-Quantil usw. Die Dichte der Gumbelverteilung wird durch zwei Parameter λ und x_0 beschrieben:

$$f_X(x) = \lambda e^{-\lambda(x-x_0)} e^{-(e^{-\lambda(x-x_0)})}.$$

Die Transformation $y = \lambda(x - x_0)$ ergibt die Standardform

$$f_Y(y) = e^{-y} e^{-(e^{-y})}.$$

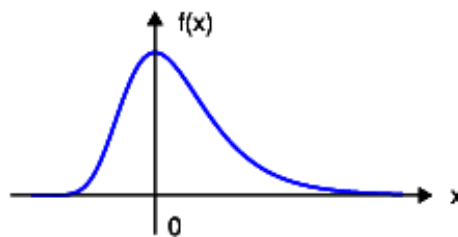


Abbildung 4.12: Dichte der Standardform der Gumbelverteilung.

4.5 Wichtige Kennzahlen von Zufallsgrößen

Die wichtigsten Kennzahlen von Zufallsgrößen sind der Erwartungswert und die Varianz. In der frequentistischen Deutung der Wahrscheinlichkeit sollen diese dem Stichprobenmittel und der Stichprobenvarianz entsprechen. Wir erinnern an die Definition des Stichprobenmittels

$$\bar{x} = \frac{1}{n} \sum_{j=1}^k x_j H_j = \sum_{j=1}^k x_j h_j$$

mit den relativen Häufigkeiten h_j . Für *diskrete Zufallsgrößen* X definieren wir den Erwartungswert $E(X)$ gemäß der Entsprechung $h_j \approx P(X = x_j)$:

$$E(X) = \sum_j x_j P(X = x_j)$$

Beispiel – die Alternativverteilung: Hier ist $P(X = 1) = p$ und $P(X = 0) = 1 - p$, daher

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p,$$

und wir erkennen die wichtige Tatsache, dass der Erwartungswert der Alternativverteilung gerade gleich der Wahrscheinlichkeit des Ereignisses $X = 1$ ist.

Um den Erwartungswert für *kontinuierliche Zufallsgrößen* zu definieren, gehen wir von der Beobachtung aus, dass

$$\sum x_j h_j = \frac{\sum x_j h_j}{\sum h_j}$$

im Histogramm den Schwerpunkt der Rechtecksflächen darstellt, wenn man für x_j die Klassenmitten wählt. Bei Verfeinerung der Klasseneinteilung nähert sich der so berechnete Flächenschwerpunkt dem Stichprobenmittel \bar{x} .

Dies legt nahe, den Erwartungswert als den Schwerpunkt der Fläche unter der Dichtekurve zu definieren:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

Da $\int_{-\infty}^{\infty} f(x) dx = 1$ ist, ist das erste Moment tatsächlich gleich dem Schwerpunkt.

Beispiel – die Normalverteilung $\mathcal{N}(\mu, \sigma^2)$: Es ergibt sich

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \left((x-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \mu e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) dx = 0 + \mu \cdot 1 = \mu;$$

der Erwartungswert stimmt mit dem Modalwert der Dichtekurve überein.

In derselben Weise wollen wir nun die *Varianz* als Gegenstück zur Stichprobenvarianz

$$s^2 = \frac{n}{n-1} \sum_{j=1}^k (x_j - \bar{x})^2 h_j \approx \sum_{j=1}^k (x_j - \bar{x})^2 h_j$$

eingeführen. Die Stichprobenvarianz ist die mittlere quadratischen Abweichung vom Mittelwert. Entsprechend definieren wir die Varianz $V(X)$ einer Zufallsgröße X als den Erwartungswert der quadratischen Abweichung von $E(X)$, also:

$$V(X) = E\left((X - E(X))^2\right)$$

Für diskrete Zufallsgrößen ergibt sich demnach

$$V(X) = \sum_j (x_j - E(X))^2 P(X = x_j),$$

für kontinuierliche Zufallsgrößen:

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx.$$

Wir fassen zusammen: Der *Erwartungswert* einer Zufallsgröße X ist das erste Moment oder der Massenmittelpunkt der Verteilung:

$$E(X) = \mu_X = \begin{cases} \sum_j x_j p_j, & X \text{ diskret,} \\ \int_{-\infty}^{\infty} x f_X(x) dx, & X \text{ kontinuierlich.} \end{cases}$$

Die *Varianz* ist das zentrierte zweite Moment oder der Erwartungswert der quadratischen Abweichung von μ_X :

$$V(X) = \sigma_X^2 = \begin{cases} \sum_j (x_j - \mu_X)^2 p_j, & X \text{ diskret,} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx, & X \text{ kontinuierlich.} \end{cases}$$

Die *Standardabweichung* oder *Streuung* σ_X ist die Quadratwurzel der Varianz, der *Variationskoeffizient* v_X der Quotient aus Standardabweichung und Erwartungswert:

$$\sigma_X = \sqrt{V(X)}, \quad v_X = \frac{\sigma_X}{\mu_X}.$$

Beispiel 4.13 (Alternativverteilung) Es gilt $P(X = 1) = p$, $P(X = 0) = 1 - p$ und

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p, \quad V(X) = (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p = p(1 - p).$$

Beispiel 4.14 (Diskrete Gleichverteilung) Falls die Werte einer gleichverteilten diskreten Zufallsgröße die natürlichen Zahlen $x_1 = 1, \dots, x_k = k$ sind (mit Wahrscheinlichkeiten $P(X = j) = 1/k$), kann man den Erwartungswert und die Varianz explizit berechnen:

$$E(X) = \sum_{j=0}^k j \cdot \frac{1}{k} = \frac{(k+1)k}{2} \cdot \frac{1}{k} = \frac{k+1}{2},$$

$$V(X) = \sum_{j=0}^k j^2 \cdot \frac{1}{k} - \frac{(k+1)^2}{4} = \frac{k(k+1)(2k+1)}{6} \cdot \frac{1}{k} - \frac{(k+1)^2}{4} = \frac{k^2 - 1}{12},$$

letzteres unter Verwendung der Formel $V(X) = E(X^2) - (E(X))^2$ (Anwendung 4.26).

Beispiel 4.15 (Binomialverteilung) Ihr Erwartungswert ergibt sich aus

$$\begin{aligned} E(X) &= \sum_{j=0}^n j P(X = j) = \sum_{j=0}^n j \binom{n}{j} p^j (1-p)^{n-j} \\ &= \sum_{j=0}^n j \cdot \frac{n(n-1)\cdots(n-j+1)}{1 \cdot 2 \cdots (j-1) \cdot j} p^j (1-p)^{n-j} \\ &= np \cdot \sum_{j=1}^n \binom{n-1}{j-1} p^{j-1} (1-p)^{n-j} \\ &= np \sum_{l=0}^{n-1} \binom{n-1}{l} p^l (1-p)^{n-1-l} = np(p + 1 - p)^{n-1} \\ &= np. \end{aligned}$$

Analog erhält man für die Varianz

$$V(X) = \sum_{j=0}^n (j - np)^2 \binom{n}{j} p^j (1-p)^{n-j} = np(1-p).$$

Beispiel 4.16 (Poissonverteilung) Es gilt:

$$\begin{aligned} E(X) &= \sum_{j=0}^{\infty} j \frac{\mu^j}{j!} e^{-\mu} = \mu \sum_{j=1}^{\infty} j \frac{\mu^{j-1}}{j!} e^{-\mu} = \mu \sum_{i=0}^{\infty} \frac{\mu^i}{i!} e^{-\mu} = \mu \\ V(X) &= \sum_{j=0}^{\infty} (j - \mu)^2 \frac{\mu^j}{j!} e^{-\mu} = \mu \end{aligned}$$

Bei der Poissonverteilung stimmen also Erwartungswert und Varianz überein und sind gleich dem Parameter μ .

Beispiel 4.17 (Normalverteilung $\mathcal{N}(\mu, \sigma^2)$) Den Erwartungswert haben wir oben zu

$$E(X) = \mu$$

berechnet. Für die Varianz ergibt sich

$$V(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$$

Die Parameter μ, σ^2 der Normalverteilung bedeuten also gerade Erwartungswert und Varianz.

Beispiel 4.18 (Kontinuierliche Gleichverteilung) Simple Integrationen ergeben die folgenden Formeln:

$$\begin{aligned} E(X) &= \int_a^b x \cdot \frac{1}{b-a} dx = \frac{x^2}{2} \Big|_a^b \cdot \frac{1}{b-a} = \frac{b^2-a^2}{2} \cdot \frac{1}{b-a} = \frac{a+b}{2} \\ V(X) &= \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a} dx = \frac{1}{(b-a)} \cdot \frac{1}{3} \left(x - \frac{a+b}{2}\right)^3 \Big|_a^b \\ &= \frac{1}{3(b-a)} \left(\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3 \right) = \frac{1}{12} \cdot \frac{(b-a)^3}{(b-a)} = \frac{(b-a)^2}{12} \end{aligned}$$

Beispiel 4.19 (Exponentialverteilung) Partielle Integration ergibt:

$$\begin{aligned} E(X) &= \int_0^{\infty} \lambda x e^{-\lambda x} dx = \lambda \left[-\frac{x}{\lambda} e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} \frac{1}{\lambda} e^{-\lambda x} dx \right] = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda} \\ V(X) &= \int_0^{\infty} \lambda \left(x - \frac{1}{\lambda}\right)^2 e^{-\lambda x} dx = \frac{1}{\lambda^2}. \end{aligned}$$

Der Erwartungswert einer exponentialverteilten Größe ist der Kehrwert des Parameters λ .

Beispiel 4.20 (Dreiecksverteilung) Mit den Bezeichnungen von Beispiel 4.9 gilt:

$$E(X) = \frac{1}{3}(a + m + b), \quad V(X) = \frac{1}{18}(a^2 + m^2 + b^2 - ma - mb - ab).$$

Beispiel 4.21 (Betaverteilung) Hier gilt

$$E(X) = \frac{sa + rb}{r + s}, \quad V(X) = \frac{rs}{(r + s)^2(r + s + 1)} (b - a)^2.$$

Beispiel 4.22 (Gumbelverteilung) Durch Rückführung auf Standardform und Integration erhält man:

$$E(X) = x_0 + \gamma/\lambda, \quad V(X) = \pi^2/(6\lambda^2).$$

Dabei ist

$$\gamma \approx 0.577216 \dots = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k} - \log n$$

die so genannte *Euler-Mascheroni'sche Konstante*.

4.6 Die Verteilungsfunktion einer Zufallsgröße

Die *kumulative Verteilungsfunktion* oder kurz *Verteilung* einer Zufallsgröße X ist die reelle Funktion

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

Für die Verteilungsfunktion einer diskreten Zufallsgröße gilt

$$F_X(x) = P(X \leq x) = \sum_{x_j \leq x} p_j;$$

diese ist stückweise konstant und springt jeweils beim Wert x_j um p_j . Die Abbildung 4.13 zeigt die Einzelwahrscheinlichkeiten und die Verteilungsfunktion einer Alternativverteilung.



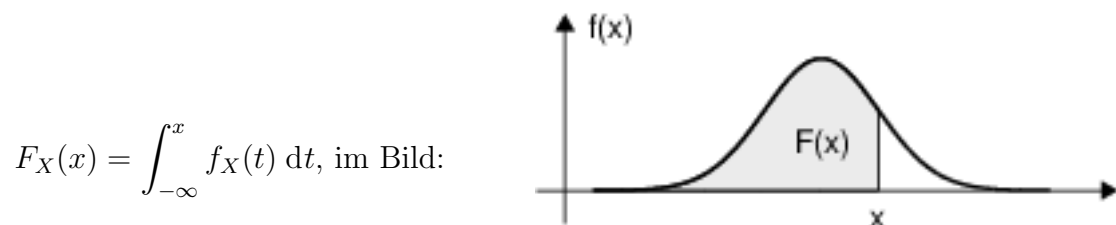
Abbildung 4.13: Verteilungsfunktion einer diskreten Zufallsgröße.

An der Verteilungsfunktion lassen sich die Einzelwahrscheinlichkeiten bei den Sprungstellen ablesen:

$$\begin{aligned} P(X = x_j) &= \text{Sprung von } F(x) \text{ bei } x = x_j \\ &= F(x_j+) - F(x_j-). \end{aligned}$$

Kenntnis der Verteilungsfunktion ist also gleichbedeutend mit Kenntnis der Einzelwahrscheinlichkeiten.

Für die Verteilungsfunktion einer kontinuierlichen Zufallsgröße gilt:



Aus dem Hauptsatz der Differential- und Integralrechnung folgt die Beziehung

$$\frac{d}{dx} F_X(x) = f_X(x);$$

die Dichtefunktion kann durch Ableiten der Verteilungsfunktion gewonnen werden.

Die Abbildung 4.14 zeigt die Dichte und Verteilungsfunktion der Standardnormalverteilung.



Abbildung 4.14: Verteilungsfunktion einer kontinuierlichen Zufallsgröße.

Zusammenfassend halten wir fest, dass eine Zufallsgröße durch Angabe ihrer Verteilungsfunktion vollständig charakterisiert ist.

4.7 Funktionen von Zufallsgrößen

Die Aufgabe, Funktionen von Zufallsgrößen auszuwerten, tritt häufig auf, zum Beispiel bei der Ermittlung der Verteilung der Antwort eines mechanischen Systems, wenn die Eingabewerte Zufallsgrößen sind. Man spricht auch von Transformationen von Zufallsgrößen. Wir beginnen mit dem *diskreten Fall*.

Sei X eine Zufallsgröße mit Werten x_1, \dots, x_k und $u = U(x)$ eine bijektive Funktion (eindeutige Zuordnung). Die Werte von $U = U(X)$ sind dann

$$u_1 = U(x_1), \dots, u_k = U(x_k).$$

Im einfachen Beispiel der Quadratfunktion ist $u = U(x) = x^2$; die funktionale Zuordnung ist im Bereich $x \geq 0$ eindeutig, vergleiche Abbildung 4.15.

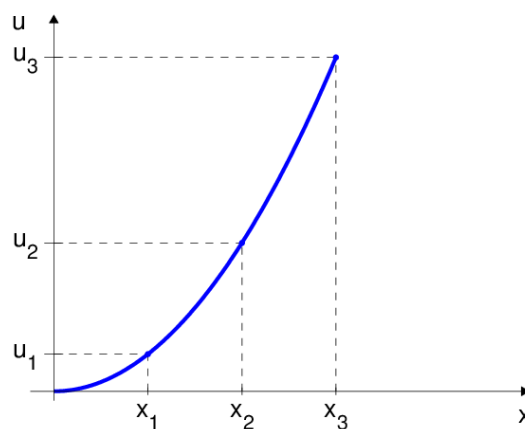


Abbildung 4.15: Transformation einer Zufallsgröße.

Wir suchen die Einzelwahrscheinlichkeiten $P(U = u_j)$. Diese lassen sich leicht mit Hilfe der Umkehrfunktion ausdrücken:

$$P(U = u_j) = P(U(X) = u_j) = P(X = U^{-1}(u_j)) = P(X = x_j).$$

Der Erwartungswert der transformierten Zufallsgröße U ergibt sich zu

$$E(U) = \sum_{j=1}^k u_j P(U = u_j) = \sum_{j=1}^k U(x_j) P(X = x_j).$$

Im Beispiel der Quadratfunktion $u = x^2$ erhält man das zweite Moment

$$E(X^2) = \sum_{j=1}^k x_j^2 P(X = x_j)$$

Auf diese Weise erhält man auch eine neue Interpretation der Varianz als Erwartungswert der transformierten Größe $U = (X - \mu)^2$, wobei $E(X) = \mu$ den Erwartungswert der ursprünglichen Größe bedeutet:

$$V(X) = E((X - \mu)^2) = \sum_{j=1}^k (x_j - \mu)^2 P(X = x_j).$$

Im *kontinuierlicher Fall* erfolgt die Berechnung der Verteilung von $U = U(X)$ am besten mit Hilfe der Verteilungsfunktionen:

$$F_U(u) = P(U \leq u) = P(U(X) \leq u) = P(X \leq U^{-1}(u)) = \int_{-\infty}^{U^{-1}(u)} f_X(x) dx = F_X(U^{-1}(u)).$$

Die Dichte $f_U(u)$ erhält man durch Differenzieren:

$$f_U(u) = \frac{d}{du} F_U(u) = \frac{d}{dx} F_X(U^{-1}(u)) \cdot \frac{d}{du} U^{-1}(u) = f_X(U^{-1}(u)) \frac{d}{du} U^{-1}(u)$$

Mit Hilfe der Substitution

$$x = U^{-1}(u), \quad dx = \frac{d}{du} U^{-1}(u) du, \quad u = U(x)$$

lässt sich der Erwartungswert berechnen:

$$E(U) = \int_{-\infty}^{\infty} u f_U(u) du = \int_{-\infty}^{\infty} u f_X(U^{-1}(u)) \frac{d}{du} U^{-1}(u) du = \int_{-\infty}^{\infty} U(x) f_X(x) dx.$$

Ähnlich wie im diskreten Fall erhält man die Varianz als Erwartungswert der transformierten Größe $U = (X - \mu)^2$ mit $\mu = E(X)$:

$$V(X) = E(U) = E((X - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx.$$

Anwendung 4.23 (Zentrierung der Normalverteilung) Sei X nach $\mathcal{N}(\mu, \sigma^2)$ verteilt. Dann ist

$$U = \frac{X - \mu}{\sigma}$$

nach $\mathcal{N}(0, 1)$ verteilt. Zur Herleitung dieser wichtigen Tatsache betrachten wir die Verteilungsfunktion

$$\begin{aligned} P(U \leq u) &= P\left(\frac{X - \mu}{\sigma} \leq u\right) = P(X \leq \mu + \sigma u) \\ &= \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\mu + \sigma u} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-y^2/2} dy, \end{aligned}$$

wobei im letzten Schritt die Substitution $y = \frac{x - \mu}{\sigma}$ verwendet wurde. Das Ergebnis ist aber gerade die Verteilungsfunktion von $\mathcal{N}(0, 1)$.

Die übliche Bezeichnungsweise für die Standardnormalverteilung ist

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \quad \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-y^2/2} dy.$$

In Tabelle 4.1 findet man ihre Werte für $0 \leq u \leq 3.69$.

Beispiel 4.24 (Zentrierung und Tabellenrechnung) Sei X nach $\mathcal{N}(\mu, \sigma^2)$ verteilt. Wir wollen die Wahrscheinlichkeit berechnen, dass ein Wert in einfacher Streudistanz vom Erwartungswert μ auftritt, das ist die Wahrscheinlichkeit

$$P(\mu - \sigma \leq X \leq \mu + \sigma).$$

Umformen zeigt, dass diese Wahrscheinlichkeit gleich

$$P(-\sigma \leq X - \mu \leq \sigma) = P(-1 \leq \frac{X - \mu}{\sigma} \leq 1) = P(-1 \leq U \leq 1)$$

ist, wobei die Größe U nach $\mathcal{N}(0, 1)$ verteilt ist. In Tabelle 4.1 sind nur einseitige Wahrscheinlichkeiten angegeben. Wir machen uns die Symmetrie der Normalverteilung zunutze:



Abbildung 4.16: Symmetrischer einfacher Streubereich der Standardnormalverteilung.

Die Abbildung zeigt, dass gilt

$$\begin{aligned} P(-1 \leq U \leq 1) &= P(-\infty < U \leq 1) - P(-\infty < U < -1) \\ &= P(-\infty < U \leq 1) - P(U > 1) \\ &= P(U \leq 1) - (1 - P(U \leq 1)) \\ &= 2P(U \leq 1) - 1 = 2 \cdot 0.8413 - 1 = 0.683 \approx 68\%. \end{aligned}$$

wobei die Zahlenwerte der Tabelle zu entnehmen sind. In ähnlicher Weise berechnet man die Prozentanteile in zwei- und dreifacher Streudistanz (Abbildung 4.7).

Als konkretes Beispiel nehmen wir die mittlere Tagestemperatur im Jänner 2001 in Innsbruck. In Abschnitt 2.3 haben wir eine Normalverteilung mit Parametern $\mu = 0.76$, $\sigma^2 = 9.1085$, $\sigma = 3.018$ angepasst. Unter der Annahme, dass dies ein zutreffendes Modell für die Jännertemperaturen ist, können wir die Wahrscheinlichkeit ausrechnen, dass die mittlere Tagestemperatur größer als 5° ist:

$$\begin{aligned} P(X \geq 5) &= 1 - P(X \leq 5) = 1 - P(X - 0.76 \leq 4.24) = 1 - P\left(\frac{X-0.76}{3.018} \leq \frac{4.24}{3.018}\right) \\ &= 1 - P\left(U \leq \frac{4.24}{3.018}\right) = 1 - P(U \leq 1.4049) \approx 1 - \Phi(1.405). \end{aligned}$$

Wir entnehmen der Tabelle: $\Phi(1.40) = 0.9192$, $\Phi(1.41) = 0.9207$ und mittels Interpolation $\Phi(1.405) \approx 0.92$. Insgesamt also

$$P(X \geq 5) = 1 - \Phi(1.405) = 0.08 \approx 8\%,$$

wie vor Abbildung 2.11 behauptet.

Anwendung 4.25 (Die Lognormalverteilung) Diese wird häufig zur Beschreibung rechts-schiefer unimodaler Verteilungen positiver Größen verwendet. Eine Zufallsgröße U ist lognormalverteilt, wenn ihr (natürlicher) Logarithmus normalverteilt ist. Man verwendet dann auch die Parameter der assoziierten Normalverteilung zur Spezifikation:

$$U \sim \mathcal{LogN}(\mu, \sigma^2) \quad \Longleftrightarrow \quad X = \log U \sim \mathcal{N}(\mu, \sigma^2).$$

Umgekehrt gilt natürlich ebenso: ist X normalverteilt nach $\mathcal{N}(\mu, \sigma^2)$, so ist $U = e^X$ lognormalverteilt nach $\mathcal{LogN}(\mu, \sigma^2)$. Ist $f_X(x)$ die Dichtefunktion der Normalverteilung, so erhält man die Dichtefunktion der Lognormalverteilung zu

$$\begin{aligned} f_U(u) &= f_X(U^{-1}(u)) \frac{d}{du} U^{-1}(u) = f_X(\log u) \frac{1}{u} \\ &= \begin{cases} \frac{1}{\sqrt{2\pi\sigma u}} e^{-(\log u - \mu)^2 / 2\sigma^2}, & u > 0, \\ 0, & u \leq 0. \end{cases} \end{aligned}$$

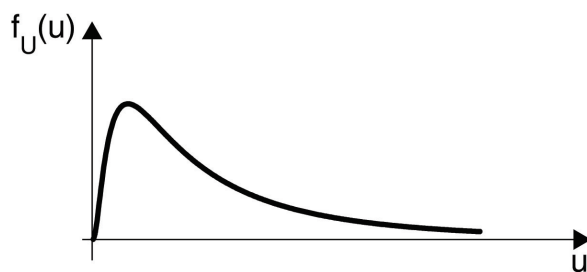


Abbildung 4.17: Dichte der Standardlognormalverteilung.

Erwartungswert und Varianz der Lognormalverteilung berechnen sich zu

$$E(U) = e^{\mu + \sigma^2/2}, \quad V(U) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1).$$

Anwendung 4.26 (Eine nützliche Formel) Die Varianz einer Zufallsgröße X lässt sich wie folgt berechnen:

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

Um das zu sehen, fassen wir die Varianz – wie oben dargelegt – als Funktion der Zufallsgröße X auf und können weiter umrechnen, mit $\mu = E(X)$:

$$\begin{aligned} V(X) &= E((X - \mu)^2) = E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2 = E(X^2) - (E(X))^2. \end{aligned}$$

Wir haben dabei verwendet, dass das Bilden des Erwartungswertes eine lineare Operation ist, also mit Addition und Skalarmultiplikation vertauschbar ist, wie die Formeln am Beginn des Abschnitts zeigen.

4.8 Anhang: Verteilungen und ihre Anwendungen

In diesem Anhang werden einige weitere Verteilungstypen vorgestellt und typische Anwendungen der besprochenen Verteilungen zusammengestellt.

Diskrete Verteilungen

Diskrete Gleichverteilung: Zufallsgrößen, die endlich viele Werte mit gleicher Wahrscheinlichkeit annehmen. Modell des ungezinkten Würfels. Anwendung bei statistischen Tests, zum Beispiel Test auf Trend einer Zeitreihe.

Alternativverteilung: X nimmt die Werte 1 (Erfolg) mit Wahrscheinlichkeit p und 0 (Misserfolg) mit Wahrscheinlichkeit $1 - p$ an. Anwendung: Ja-Nein-Entscheidungen, Modellierung des Eintretens oder Nichteintretens eines gegebenen Ereignisses.

Binomialverteilung: Bernoulli-Experiment: n -malige unabhängige Wiederholung eines 0-1-Versuches mit Erfolgswahrscheinlichkeit p . Eine binomialverteilte Größe X nimmt die Werte $0, 1, 2, \dots, n$ an und beschreibt die Anzahl der Erfolge bei n 0-1-Versuchen.

Geometrische Verteilung: Beschreibt die Anzahl der Misserfolge bis zum ersten Erfolg bei Wiederholung eines 0-1-Versuches.

Negative Binomialverteilung: Beschreibt die Anzahl der 0-1-Versuche, die erforderlich sind, um $r \geq 1$ Erfolge zu erzielen. Für $r = 1$ reduziert sie sich auf die geometrische Verteilung. Anwendung in der Versicherungsmathematik.

Hypergeometrische Verteilung: Anwendung in der Qualitätskontrolle. Gegeben: N Werkstücke, davon sind M defekt. Es wird eine Probe von K Stück entnommen. Die hypergeometrische Verteilung beschreibt die Anzahl der defekten Stücke unter den K entnommenen.

Poissonverteilung: Beschreibt die Anzahl von Ereignissen, die in einem gegebenen Zeitintervall eintreten. Die Wartezeit zwischen den Ereignissen ist exponentialverteilt. Anwen-

dungen: Anzahl eintreffender Anrufe in einer Telefonzentrale, Anzahl eintreffender Jobs auf einem EDV-Server, Anzahl von Sturmfluten oder Regenfronten in einem Jahr, Anzahl von Unfällen eines Versicherungsnehmers im Jahr.

Kontinuierliche Verteilungen

Kontinuierliche Gleichverteilung: Grundverteilung bei der Erzeugung von Zufallszahlen. Kontinuierliche Verteilung mit geringstem Informationsgehalt (Verwendung in der Bayes-Statistik als Priorverteilung).

Dreiecksverteilung: Wahrscheinlichkeitstheoretisches Modell einer Dreipunktschätzung (minimaler, häufigster, maximaler Wert).

Betaverteilung: Vierparametrische Verteilung, die unimodale Zufallsgrößen mit Werten in einem Intervall $[a, b]$ modelliert. Die beiden anderen Parameter erlauben es, verschiedene Glattheits- und Schiefegrade zu erreichen.

Normalverteilung: Eine weit anwendbare symmetrische Verteilung. Summen unabhängiger Zufallsgrößen hinreichend hoher Summandenzahl sind näherungsweise normalverteilt. Stichprobenmittel sind näherungsweise normalverteilt. Summen und Vielfache normalverteilter Größen sind normalverteilt.

Lognormalverteilung: Eine Zufallsgröße ist lognormalverteilt, wenn ihr natürlicher Logarithmus normalverteilt ist. Modellierung nicht negativer, rechtsschiefer Verteilungen mit *fat tails*, also langsamem Abfall im Unendlichen.

Exponentialverteilung: Kombinatorisches Modell für die Lebensdauer eines radioaktiven Partikels. Wird allgemein zur Modellierung von Lebensdauern (z. B. von Geräten) und Wartezeiten verwendet. Modelliert die Wartezeit zwischen zwei aufeinanderfolgenden Ereignissen eines Poissonprozesses.

Erlangverteilung: Eine Summe von exponentialverteilten Größen ist Erlang-verteilt.

Gammaverteilung: Allgemeine Verteilung, die als Spezialfälle die Exponentialverteilung, die Chi-Quadrat-Verteilung und die Erlangverteilung enthält. Für ganzzahlige Potenzen p modelliert sie die Wartezeit bis zum p -ten Ereignis in einem Poissonprozess.

Laplaceverteilung: Besteht aus zwei um Null gespiegelten, aneinandergefügt Exponentialverteilungen.

Gumbelverteilung: Verteilung der Extremwerte (Maxima und Minima) einer Größe, deren Verteilung exponentiell im Unendlichen abklingt. Anwendung: Hochwasserstatistik.

Weibullverteilung: Verallgemeinerung der Exponentialverteilung. Anwendungen in der Extremwertstatistik und zur Beschreibung von Lebensdauern.

Paretoverteilung: Beschreibt nichtnegative Zufallsgrößen, deren kleinen Werte mit sehr hoher und deren große Werte mit kleiner Wahrscheinlichkeit auftreten. Anwendungsbeispiel: Vermögensverteilung in einer Bevölkerung, Größenverteilung nach Einwohnerzahl

von Städten.

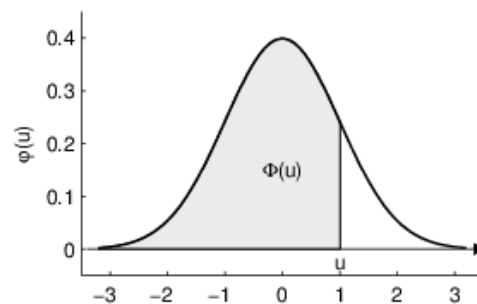
Cauchyverteilung: Verteilung einer Zufallsgröße, deren extreme Werte eine sehr hohe Wahrscheinlichkeit besitzen. Der Erwartungswert des Betrags $|X|$ einer Cauchy-verteilten Größe X ist ∞ . Anwendung in der Simulation.

Maxwellverteilung: Eine Wahrscheinlichkeitsverteilung der Statistischen Physik. Beschreibt die Verteilung der Teilchengeschwindigkeiten in einem idealen Gas.

Chi-Quadrat-Verteilung: Wichtige Testverteilung. Die Summe der Quadrate von n unabhängigen, standardnormalverteilten Größen ist χ^2 -verteilt mit n Freiheitsgraden. Lineare Abhängigkeiten unter den Größen führen zur Reduktion der Freiheitsgrade. Die Stichprobenvarianz ist χ^2 -verteilt mit $n - 1$ Freiheitsgraden.

t-Verteilung: Wichtige Testverteilung. Ist X standardnormalverteilt und unabhängig von der mit n Freiheitsgraden χ^2 -verteilten Größe Y , so besitzt $X/\sqrt{Y/n}$ eine t -Verteilung mit n Freiheitsgraden. Wird auch als Student-Verteilung bezeichnet.

F-Verteilung: Wichtige Testverteilung. Sind V_m und V_n unabhängig und χ^2 -verteilt mit m bzw. n Freiheitsgraden, so ist nV_m/mV_n F-verteilt mit (m, n) Freiheitsgraden. Wird auch als Fisher-Verteilung bezeichnet.

Tabelle der Standardnormalverteilung $N(0, 1)$ 

u	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	$\phi(u)$
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	0.3989
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.3970
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.3910
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	0.3814
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	0.3683
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	0.3521
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	0.3332
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	0.3123
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	0.2897
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	0.2661
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	0.2420
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	0.2179
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	0.1942
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	0.1714
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	0.1497
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	0.1295
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	0.1109
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	0.0940
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	0.0790
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	0.0656
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0.0540
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0.0440
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0.0355
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0.0283
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0.0224
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0.0175
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0.0136
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0.0104
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0.0079
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0.0060
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990	0.0044
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993	0.0033
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995	0.0024
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997	0.0017
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	0.0012
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.0009
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.0006

Tabelle 4.1: Werte der kumulativen Standardnormalverteilung.

Kapitel 5

Mehrdimensionale Zufallsgrößen

Eine zweidimensionale Zufallsgröße besteht aus einem Paar (X, Y) von Zufallsgrößen. Ziel dieses Abschnitts ist es zu zeigen, wie die *gemeinsame Verteilung* der Größen X und Y modelliert werden kann. Allgemeiner können auch mehrdimensionale Größen behandelt werden, die durch einen Vektor (X_1, \dots, X_n) dargestellt werden.

Der diskrete Fall: Die beiden diskreten Zufallsgrößen mögen die folgenden Werte annehmen:

Werte für X : x_1, x_2, \dots, x_k ;

Werte für Y : y_1, y_2, \dots, y_l .

Analog zu den gemeinsamen relativen Häufigkeiten der beschreibenden Statistik definieren wir die gemeinsamen Wahrscheinlichkeiten

$$p_{ij} = P(X = x_i, Y = y_j).$$

Die *Randwahrscheinlichkeiten* sind gegeben durch

$$\begin{aligned} p_{i\cdot} &= \sum_{j=1}^l p_{ij} = \sum_{j=1}^l P(X = x_i, Y = y_j) = P(X = x_i) \\ p_{\cdot j} &= \sum_{i=1}^k p_{ij} = \sum_{i=1}^k P(X = x_i, Y = y_j) = P(Y = y_j) \end{aligned}$$

Die gesamte Information kann in einer Wahrscheinlichkeitstabelle zusammengefasst werden:

$X \setminus Y$	y_j			
x_i	\dots	p_{ij}	\dots	$p_{i\cdot}$
		\vdots		
		\vdots		
		$p_{\cdot j}$		1

(\sum_j)

(\sum_i)

Wir können zunächst wieder die Erwartungswerte und Varianzen der beiden Zufallsgrößen einzeln berechnen:

$$\begin{aligned}\mu_X &= E(X) = \sum_{i=1}^k \sum_{j=1}^l x_i p_{ij} \\ \mu_Y &= E(Y) = \sum_{i=1}^k \sum_{j=1}^l y_j p_{ij} \\ \sigma_X^2 &= V(X) = \sum_{i=1}^k \sum_{j=1}^l (x_i - \mu_X)^2 p_{ij} \\ \sigma_Y^2 &= V(Y) = \sum_{i=1}^k \sum_{j=1}^l (y_j - \mu_Y)^2 p_{ij}\end{aligned}$$

Der wichtigste Parameter der gemeinsamen Verteilung ist die *Kovarianz*

$$\sigma_{XY} = \text{COV}(X, Y) = \sum_{i=1}^k \sum_{j=1}^l (x_i - \mu_X)(y_j - \mu_Y) p_{ij}.$$

Die Bildung der *Kovarianzmatrix*

$$C(X, Y) = \begin{bmatrix} V(X) & \text{COV}(X, Y) \\ \text{COV}(X, Y) & V(Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}.$$

wird sich als nützlich erweisen. Der *Korrelationskoeffizient* ist

$$\varrho(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{COV}(X, Y)}{\sqrt{V(X)V(Y)}}.$$

Wenn die Kovarianz – oder der Korrelationskoeffizient – zweier Größen X, Y gleich Null ist, so heißen X, Y *unkorreliert*.

Der kontinuierliche Fall: Die gemeinsame Verteilung wird durch eine *gemeinsame Dichtefunktion* $f(x, y)$ beschrieben. Für eine Menge $M \subset \mathbb{R}^2$ ist die Wahrscheinlichkeit, dass das Ereignis $(x, y) \in M$ eintritt, gerade das Volumen unter der Fläche $z = f(x, y)$, wie in Abbildung 5.1 dargestellt.

Insbesondere gilt für das Gesamtvolumen:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1.$$

Die *gemeinsame Verteilungsfunktion* ist

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(\xi, \eta) \, d\eta \, d\xi.$$

Die *Randverteilungen* sind gegeben durch

$$\begin{aligned}F_X(x) &= P(X \leq x) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(\xi, \eta) \, d\eta \, d\xi, \\ F_Y(y) &= P(Y \leq y) = \int_{-\infty}^y \int_{-\infty}^{\infty} f(\xi, \eta) \, d\xi \, d\eta\end{aligned}$$

$$P((X, Y) \in M) = \iint_M f(x, y) \, dx \, dy$$

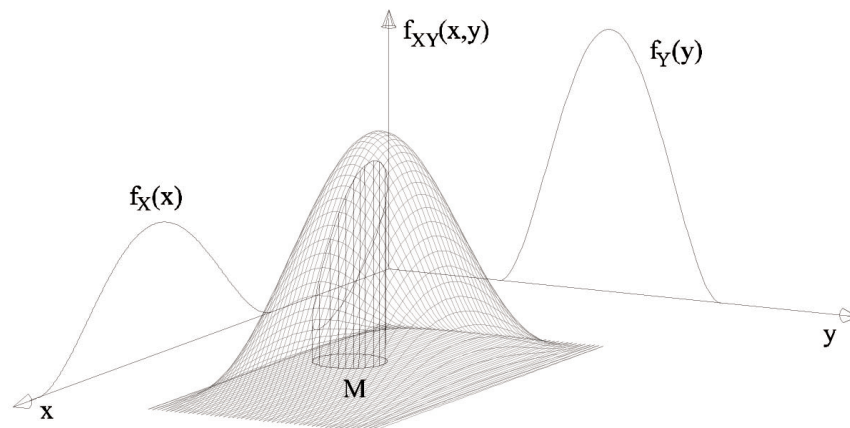


Abbildung 5.1: Dichte einer zweidimensionalen Zufallsgröße mit Darstellung der Randdichten.

mit den *Randdichten*

$$f_X(x) = \int_{-\infty}^{\infty} f(x, \eta) \, d\eta, \quad f_Y(y) = \int_{-\infty}^{\infty} f(\xi, y) \, d\xi.$$

Die Erwartungswerte der beiden Zufallsgrößen erhält man durch

$$\mu_X = E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) \, dx \, dy,$$

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) \, dx \, dy,$$

die Varianzen durch

$$\sigma_X^2 = V(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x, y) \, dx \, dy,$$

$$\sigma_Y^2 = V(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_Y)^2 f(x, y) \, dx \, dy$$

und die Kovarianz durch

$$\sigma_{XY} = \text{COV}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) \, dx \, dy.$$

Die Kovarianzmatrix und der Korrelationskoeffizient sind wie im diskreten Fall definiert durch

$$C(X, Y) = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}, \quad \varrho(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Wir wollen nun die Inverse der Kovarianzmatrix berechnen. Dazu schreiben wir $\mu_1, \mu_2, \sigma_1, \sigma_2$ und σ_{12} statt $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ und σ_{XY} . Wir fassen das Paar (x, y) zu einem Vektor $\mathbf{x} = [x \ y]^\top$ zusammen und ebenso die beiden Erwartungswerte: $\boldsymbol{\mu} = [\mu_1 \ \mu_2]^\top$. Aus $\varrho = \sigma_{12}/\sigma_1\sigma_2$ folgt $\sigma_{12} = \varrho \sigma_1\sigma_2$ und wir erhalten

$$C = \begin{bmatrix} \sigma_1^2 & \varrho \sigma_1\sigma_2 \\ \varrho \sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \quad \det(C) = \sigma_1^2\sigma_2^2(1 - \varrho^2)$$

und weiter

$$C^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \varrho^2)} \begin{bmatrix} \sigma_2^2 & -\varrho \sigma_1\sigma_2 \\ -\varrho \sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1 - \varrho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\varrho}{\sigma_1\sigma_2} \\ -\frac{\varrho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}.$$

Beispiel 5.1 (Die zweidimensionale Normalverteilung) Ihre Dichte lässt sich am einfachsten mit Hilfe der Kovarianzmatrix schreiben:

$$f(x, y) = \frac{1}{2\pi\sqrt{\det C}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top C^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

Setzt man die eben hergeleiteten Formeln für die Inverse der Kovarianzmatrix ein, so erhält man für $f(x, y)$ weiter

$$\begin{aligned} &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \varrho^2}} \exp\left(\frac{-1}{2(1 - \varrho^2)} [x - \mu_1 \quad y - \mu_2] \cdot \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\varrho}{\sigma_1\sigma_2} \\ -\frac{\varrho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix} \cdot \begin{bmatrix} x - \mu_1 \\ y - \mu_2 \end{bmatrix}\right) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \varrho^2}} \exp\left(\frac{-1}{2(1 - \varrho^2)} \left(\frac{(x - \mu_1)^2}{\sigma_1^2} - \frac{2\varrho}{\sigma_1\sigma_2}(x - \mu_1)(y - \mu_2) + \frac{(y - \mu_2)^2}{\sigma_2^2}\right)\right). \end{aligned}$$

Integration nach y bzw. x zeigt, dass die *Randverteilungen* eindimensionale Normalverteilungen sind:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x - \mu_1)^2/2\sigma_1^2}, \quad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(y - \mu_2)^2/2\sigma_2^2}.$$

Im *unkorrelierten* Fall $\varrho = 0$ ergibt sich

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma_1^2} - \frac{(y - \mu_2)^2}{2\sigma_2^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x - \mu_1)^2/2\sigma_1^2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(y - \mu_2)^2/2\sigma_2^2}, \end{aligned}$$

also

$$f(x, y) = f_X(x)f_Y(y). \quad (5.1)$$

Zwei *normalverteilte* Zufallsgrößen X, Y sind genau dann unkorreliert, wenn die gemeinsame Dichte gleich dem Produkt der Randdichten ist.

Unabhängigkeit von Zufallsgrößen: In Anlehnung an den entsprechenden Begriff aus der beschreibenden Statistik (Abschnitt 3.1) führen wir nun den Begriff der Unabhängigkeit zweier Zufallsgrößen ein.

Diskreter Fall: Die Größen X, Y heißen *unabhängig*, wenn stets gilt:

$$p_{ij} = p_{i \cdot} p_{\cdot j}$$

das heißt, die gemeinsamen Wahrscheinlichkeiten ergeben sich als Produkt der Randwahrscheinlichkeiten:

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j).$$

Kontinuierlicher Fall: Die Größen X, Y sind unabhängig, wenn gilt:

$$f(x, y) = f_X(x)f_Y(y).$$

Die gemeinsame Dichte ist also das Produkt der Randdichten.

Wir wollen nun die Beziehung zwischen Unabhängigkeit und Unkorreliertheit klären. Wir erinnern, dass zwei Zufallsgrößen (X, Y) *unkorreliert* sind, wenn $\text{COV}(X, Y) = 0$ ist.

Charakterisierung der Unkorreliertheit: Zwei Zufallsgrößen (X, Y) sind genau dann *unkorreliert*, wenn gilt:

$$E(XY) = E(X)E(Y).$$

Dies zeigt die folgende Rechnung:

$$\begin{aligned} \text{COV}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) = E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y). \end{aligned}$$

Somit ist $\text{COV}(X, Y) = 0$ gleichbedeutend mit $E(XY) = E(X)E(Y)$.

Unabhängige Zufallsgrößen sind auch unkorreliert. Dies folgt, jedenfalls im kontinuierlichen Fall, aus der Überlegung

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) \, dx \, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_X(x)f_Y(y) \, dx \, dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} yf_Y(y) \, dy \right) xf_X(x) \, dx \\ &= \int_{-\infty}^{\infty} E(Y) xf_X \, dx = E(X)E(Y). \end{aligned}$$

Eine ähnliche Rechnung ergibt dasselbe Ergebnis im diskreten Fall.

Das folgende Beispiel zeigt, dass unkorrelierte Größen im Allgemeinen *nicht* unabhängig sind.

Beispiel 5.2 Die Wahrscheinlichkeitstabelle zweier Zufallsgrößen X mit Werten $-1, 0, 1$ und Y mit Werten $0, 1$ sei

$X \setminus Y$	0	1	
-1	0.1	0.2	0.3
0	0.3	0.1	0.4
1	0.1	0.2	0.3
	0.5	0.5	

Die möglichen Werte für das Produkt XY sind dann $-1, 0, 1$ mit den Wahrscheinlichkeiten

XY	-1	0	1
	0.2	0.6	0.2

Wir erhalten daraus

$$\begin{aligned}
 E(XY) &= (-1) \cdot 0.2 + 0 \cdot 0.6 + 1 \cdot 0.2 = 0 \\
 E(X) &= (-1) \cdot 0.3 + 0 \cdot 0.4 + 1 \cdot 0.3 = 0 \\
 E(Y) &= 0 \cdot 0.5 + 1 \cdot 0.5 = 0.5
 \end{aligned}$$

Somit gilt $E(XY) = 0 = E(X)E(Y)$, die Größen sind also unkorreliert. Sie sind jedoch nicht unabhängig, da zum Beispiel $p_{11} \neq p_{1\cdot} \cdot p_{\cdot 1}$ ist.

Im Falle normalverteilter Größen ist die Lage allerdings einfacher:

Zwei normalverteilte Zufallsgrößen sind genau dann unabhängig, wenn sie unkorreliert sind.

Wir haben nämlich in Gleichung (5.1) gesehen, dass normalverteilte Größen genau dann unkorreliert sind, wenn ihre gemeinsame Dichte das Produkt der Randdichten ist, und dies bedeutet gerade die Unabhängigkeit.

Kapitel 6

Stichprobentheorie

6.1 Summen von Zufallsgrößen

Gegeben seien zwei Zufallsgrößen X, Y mit gemeinsamer Verteilung $F_{XY}(x, y)$. Wir fragen nach der Verteilung $F_Z(z)$ der Summe $Z = X + Y$.

Beispiel 6.1 Ein typisches Anwendungsbeispiel aus der Ablaufplanung ist die Dauer von Vorgängen. Werden zwei Vorgänge mit Dauern X, Y hintereinander ausgeführt, so ist die Gesamtdauer gerade $Z = X + Y$.

Satz 6.2 Für Erwartungswert und Varianz der Summe zweier Zufallsgrößen gilt:

$$E(X + Y) = E(X) + E(Y),$$

$$V(X + Y) = V(X) + V(Y) + 2 \operatorname{COV}(X, Y).$$

Sind X und Y unkorreliert, so gilt

$$V(X + Y) = V(X) + V(Y).$$

Bei kontinuierlichen Zufallsgrößen erhält man die Begründung leicht aus der Additivität des Integrals:

$$\begin{aligned} E(X + Y) &= \iint (x + y) f_{XY}(x, y) \, dx \, dy \\ &= \iint x f_{XY}(x, y) \, dx \, dy + \iint y f_{XY}(x, y) \, dx \, dy = E(X) + E(Y) \end{aligned}$$

und analog für diskrete Größen mit Hilfe der Summendarstellung. Die Formel für die Varianz erhält man aus

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

Merkregel: Der Erwartungswert der Summe ist die Summe der Erwartungswerte. Im Falle unkorrelierter Zufallsgrößen ist die Varianz der Summe gleich der Summe der Varianzen.

Es ist möglich, eine allgemeine Formel für die Verteilungsfunktion der Summe zweier Zufallsgrößen anzugeben. Diese Formel kann man wie folgt herleiten.

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \iint_M f_{XY}(x, y) \, dx \, dy$$

mit

$$M = \{(x, y) : x + y \leq z\} = \{(x, y) : y \in \mathbb{R}, x \leq z - y\}.$$

Daher folgt weiter

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{XY}(x, y) \, dx \, dy.$$

Durch Ableiten erhält man daraus die Dichte

$$f_Z(z) = \frac{d}{dz} P(Z \leq z) = \int_{-\infty}^{\infty} f_{XY}(z - y, y) \, dy.$$

Im Falle der Unabhängigkeit gilt also

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) \, dy = f_X * f_Y(z).$$

Dieses Integral nennt man die *Faltung* der Dichten f_X, f_Y . Im Allgemeinen ist dieses Integral nicht so einfach explizit auszurechnen; man behilft sich dann oft mit einer numerischen Auswertung durch Monte-Carlo-Simulation. Im Spezialfall zweier Gauß'scher Dichten kann man jedoch zeigen, dass die Faltung wieder eine Gauß'sche Dichte ergibt. Daher ist die Summe zweier Normalverteilungen ebenfalls eine Normalverteilung. Dasselbe gilt für die Differenz.

Satz 6.3 Ist X nach $\mathcal{N}(\mu_1, \sigma_1^2)$ und Y nach $\mathcal{N}(\mu_2, \sigma_2^2)$ verteilt und sind X, Y unabhängig, so ist $X + Y$ nach $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ und $X - Y$ nach $\mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$ verteilt.

6.2 Die mathematische Stichprobe

Gegeben sei eine Zufallsgröße X mit Verteilung $F_X(x)$. Wir wollen nun das *mathematische Modell* einer Stichprobe vom Umfang n mit den Realisierungen ξ_1, \dots, ξ_n entwerfen. Wir erinnern daran, dass die Ergebnisse ξ_1, \dots, ξ_n der Stichprobe durch n unabhängige, zufällige Versuchen unter identischen Bedingungen erhalten werden.

Wir fassen diese als Realisierungen von n unabhängigen Zufallsgrößen auf:

$$X_1, X_2, \dots, X_n,$$

die alle dieselbe Verteilung $F_X(x)$ wie X haben.

Das Ergebnis etwa des ersten Versuchs ist eben selbst eine Zufallsgröße X_1 . Man kann sich das so erklären: macht man mehrere Stichproben vom Umfang n , dann wird die erste Messung X_1 jeweils unterschiedliche Werte haben. Deren Verteilung ist aber identisch mit der Verteilung von X und auch mit der von X_2, X_3, \dots, X_n .

Ebenso ist das Stichprobenmittel

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

eine Zufallsgröße und wird bei verschiedenen Stichproben verschieden ausfallen.

Wir können daher vom *Erwartungswert* und der *Varianz* des Stichprobenmittels sprechen und erhalten:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n}(E(X_1) + E(X_2) + \dots + E(X_n)) = \frac{1}{n} \cdot n \cdot E(X) = E(X),$$

weil ja alle X_i identisch wie X verteilt sind.

Die Eigenschaft

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

nennt man *Erwartungstreue*. Für die Varianz folgt unter Zuhilfenahme der Unabhängigkeit:

$$\begin{aligned} V(\bar{X}) &= V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} n V(X) = \frac{1}{n} V(X). \end{aligned}$$

Es gilt daher die wichtige Formel

$$V(\bar{X}) = \frac{1}{n} V(X).$$

Insbesondere geht die Varianz des Stichprobenmittels mit wachsendem Stichprobenumfang gegen 0.

6.3 Die Gesetze der großen Zahlen

Gegeben sei Zufallsgröße X mit Erwartungswert $E(X) = \mu$ und Varianz $V(X) = \sigma^2$. Für eine mathematische Stichprobe X_1, \dots, X_n mit dem Stichprobenmittel $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ gilt, wie eben gezeigt:

$$E(\bar{X}) = \mu.$$

Der Erwartungswert des Stichprobenmittels ist allerdings ein Begriff des mathematischen Modells. In der realen Stichprobe wird das Stichprobenmittel \bar{x} erhoben, nicht dessen Erwartungswert. Die Frage – und diese liegt der frequentistischen Interpretation der Wahrscheinlichkeitstheorie zu Grunde – ist, ob sich die Stichprobenmittel bei wachsendem

Stichprobenumfang dem Erwartungswert μ der gegebenen Zufallsgröße annähern. Die Gesetze der großen Zahlen besagen, dass dies jedenfalls auf die mathematische Stichprobe zutrifft:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mu. \quad (6.1)$$

Die Frage bleibt, in welchem Sinne die Konvergenz in (6.1) zu verstehen ist.

Satz 6.4 (Starkes Gesetz der großen Zahlen) Die Konvergenzaussage (6.1) gilt *fast sicher*, das heißt,

$$P \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mu \right) = 1.$$

Der Nachweis dieses Gesetzes bedarf tieferer Methoden der Wahrscheinlichkeitstheorie. Insbesondere muss gesagt werden, um was für eine Wahrscheinlichkeit P es sich handeln soll.¹ Leichter zu zeigen (und zu interpretieren) ist:

Satz 6.5 (Schwaches Gesetz der großen Zahlen) Die Konvergenz in (6.1) erfolgt *in der Wahrscheinlichkeit*, das heißt, es gilt für alle $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right) = 0.$$

Wir können das schwache Gesetz der großen Zahlen leicht beweisen, wenn wir annehmen, dass \bar{X} eine Verteilungsdichte $f_{\bar{X}}(x)$ besitzt. Dann ist

$$P(|\bar{X} - \mu| > \alpha) = \frac{\alpha^2}{\alpha^2} \left(\int_{\mu+\alpha}^{\infty} f_{\bar{X}}(x) \, dx + \int_{-\infty}^{\mu-\alpha} f_{\bar{X}}(x) \, dx \right).$$

Ist aber $x > \mu + \alpha$, so auch $(x - \mu)^2 > \alpha^2$. Für $x < \mu - \alpha$ ist $x - \mu < -\alpha$, also ebenfalls $(x - \mu)^2 > \alpha^2$. Wir erhalten die Abschätzung

$$\begin{aligned} P(|\bar{X} - \mu| > \alpha) &\leq \frac{1}{\alpha^2} \left(\int_{\mu+\alpha}^{\infty} (x - \mu)^2 f_{\bar{X}}(x) \, dx + \int_{-\infty}^{\mu-\alpha} (x - \mu)^2 f_{\bar{X}}(x) \, dx \right) \\ &\leq \frac{1}{\alpha^2} \int_{-\infty}^{\infty} (x - \mu)^2 f_{\bar{X}}(x) \, dx = \frac{1}{\alpha^2} V(\bar{X}) \\ &= \frac{\sigma^2}{\alpha^2 n} \rightarrow 0 \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Es sei nun A ein Ereignis. Wir definieren die Zufallsgröße X als

$$X = \begin{cases} 1, & A \text{ tritt auf (Erfolg)} \\ 0, & A \text{ tritt nicht auf.} \end{cases}$$

¹Es handelt sich um die so genannte *Produktwahrscheinlichkeit* auf dem unendlichen Produkt des der Zufallsvariablen X zu Grunde liegenden Wahrscheinlichkeitsraums.

Wir setzen $P(X = 1) = p$, $P(X = 0) = 1 - p$ und erhalten somit eine Alternativverteilung für das Eintreten oder Nichteintreten des Ereignisses A . Früher haben wir schon gezeigt, dass $E(X) = p$ ist. In einer Stichprobe vom Umfang n ist

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{\sharp(1-\text{er})}{n} = h_1^{(n)}$$

gerade die relative Häufigkeit des Ereignisses A . Wenden wir darauf das schwache Gesetz der großen Zahlen an, so erhalten wir das *Gesetz der großen Zahlen von Bernoulli*.

Satz 6.6 (Gesetz der großen Zahlen von Bernoulli) Die relative Häufigkeit $h_1^{(n)}$ eines Ereignisses konvergiert gegen dessen Wahrscheinlichkeit p in dem Sinne, dass für alle $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} P(|h_1^{(n)} - p| > \varepsilon) = 0.$$

Dies ist also die mathematische Rechtfertigung unserer Übersetzung

$$\text{relative Häufigkeit} \sim \text{Wahrscheinlichkeit}$$

Wohlgemerkt, dies gilt für die mathematische Stichprobe! Ob die reale Stichprobe dem näherungsweise entspricht, ist eine andere Frage. Trotzdem werden diese Sätze zur Rechtfertigung der *frequentistischen Interpretation* der Wahrscheinlichkeitstheorie herangezogen.

6.4 Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz besagt, dass Summen unabhängiger Zufallsgrößen näherungsweise normalverteilt sind, wenn die Anzahl der Summanden groß ist, und dass das Stichprobenmittel bei großem Stichprobenumfang näherungsweise normalverteilt ist. Um diesen Sachverhalt zu präzisieren, nehmen wir unabhängige Zufallsgrößen X_1, X_2, X_3, \dots mit $E(X_k) = \mu_k$, $V(X_k) = \sigma_k^2$.

Satz 6.7 (Zentraler Grenzwertsatz) Unter geeigneten Bedingungen an die Varianzen σ_k^2 gilt

$$\frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{\sqrt{\sum_{k=1}^n \sigma_k^2}} = \frac{\sum_{k=1}^n X_k - E(\sum_{k=1}^n X_k)}{\sqrt{V(\sum_{k=1}^n X_k)}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1),$$

genauer, die Verteilungsfunktion der zentrierten Summe konvergiert für jedes $x \in \mathbb{R}$ gegen die Standardnormalverteilung $\Phi(x)$:

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy.$$

Bevor wir weitere Folgerungen ziehen, wollen wir den zentralen Grenzwertsatz etwas anschaulicher umformulieren. Wir schreiben

$$\tilde{\mu}_n = \sum_{k=1}^n \mu_k, \quad \tilde{\sigma}_n = \sqrt{\sum_{k=1}^n \sigma_k^2}.$$

Folgerung 6.8 Summen unabhängiger Zufallsgrößen sind asymptotisch normalverteilt. Es gilt

$$\frac{\sum_{k=1}^n X_k - \tilde{\mu}_n}{\tilde{\sigma}_n} \sim \mathcal{N}(0, 1)$$

beziehungsweise

$$\sum_{k=1}^n X_k \sim \mathcal{N}(\tilde{\mu}_n, \tilde{\sigma}_n^2), \quad n \rightarrow \infty.$$

Da viele reale Merkmale sich additiv aus anderen zufälligen Beiträgen zusammensetzen, erklärt dies das häufige Auftreten der Normalverteilung in der Praxis.

Folgerung 6.9 Sind X_1, X_2, X_3, \dots unabhängig und identisch verteilt mit $E(X_k) = \mu$, $V(X_k) = \sigma^2$, dann ist $E(\sum_{k=1}^n X_k) = n\mu$, $V(\sum_{k=1}^n X_k) = n\sigma^2$ und es gilt

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \sim \mathcal{N}(0, 1), \quad n \rightarrow \infty.$$

Nach Kürzen des Bruches durch n erhalten wir:

Folgerung 6.10 Das Stichprobenmittel ist für hinreichend große n näherungsweise normalverteilt:

$$\frac{\frac{1}{n} \sum_{k=1}^n X_k - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1), \quad n \rightarrow \infty,$$

das heißt,

$$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n), \quad n \rightarrow \infty.$$

Folgerung 6.11 (Satz von Moivre-Laplace) Besitzen die Größen X_k eine Alternativverteilung mit Erfolgswahrscheinlichkeit p , so gilt $E(X_k) = p$, $V(X_k) = p \cdot (1 - p)$ und

$$\frac{\sum_{k=1}^n X_k - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0, 1), \quad n \rightarrow \infty.$$

Insbesondere gilt für den Anteilswert $h_1^{(n)} = \frac{1}{n} \sum_{k=1}^n X_k$ der Erfolge bei n Versuchen:

$$\frac{h_1^{(n)} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1), \quad n \rightarrow \infty.$$

Die Gesetze der großen Zahlen und der zentrale Grenzwertsatz bilden die Grundlage der statistischen Schätz- und Testmethoden, die wir in den nächsten Kapiteln behandeln werden.

6.5 Monte-Carlo-Simulation

Die Monte-Carlo-Simulation (MCS) ist ein Näherungsverfahren zur numerischen Berechnung von Verteilungsfunktionen. Die MCS beruht auf dem Gesetz der großen Zahlen von Bernoulli, wonach die relative Häufigkeit eines Ereignisses in einer Stichprobe großen Umfangs eine Annäherung an dessen Wahrscheinlichkeit ist. Bei der MCS wird diese Stichprobe künstlich mittels Zufallszahlen erzeugt.

Die numerischen Modelle des Ingenieurwesens sind in der Regel Input-Output-Modelle. Zu gegebenen Eingangswerten (Stoffparameter, Dimensionen, Dauern, Kosten, ...) sind Ausgangswerte zu berechnen (Verschiebungen, Spannungen, Gesamtdauern, Gesamtkosten). Sind die Eingangswerte deterministische Einzelwerte, so errechnet das Modell einen deterministischen Ausgangswert. Schwanken die Eingangswerte, so werden auch die Ausgangswerte schwanken. Modelliert man nun die Eingangswerte als Zufallsgrößen, so ist auch der Ausgangswert eine Zufallsgröße, und man möchte dessen Verteilung oder zumindest wichtige Parameter, wie Erwartungswert und Varianz kennen. Prinzipiell determinieren die Eingangsverteilung und das Input-Output-System die Verteilung der Ausgangswerte. Jedoch ist es schon bei einfachsten System praktisch undurchführbar, die Ausgangsverteilung explizit zu berechnen. Sie muss daher numerisch angenähert werden. Dazu kann die Monte-Carlo-Simulation dienen.

Das Grundprinzip der Monte-Carlo-Simulation besteht darin, eine künstliche Stichprobe der Eingangswerte herzustellen, etwa für jede Eingangsgröße N Werte, welche den gegebenen Eingangsverteilungen gehorchen. Anschließend wird das Input-Output-Modell N -mal durchgerechnet. Die N Werte für die Ausgangsgröße werden dann statistisch ausgewertet: Das Histogramm ist eine Näherung an die Ausgangsverteilung, das Stichprobenmittel eine Schätzung für den Erwartungswert der Ausgangsgröße, usw.

Erzeugung von Zufallszahlen: Da heute jede mathematische Software Zufallszahlengeneratoren für praktisch jede gebräuchliche Verteilung schon einprogrammiert hat, ist es nicht mehr notwendig, selbst Zufallszahlen herzustellen. Daher nur ein paar Andeutungen, wie dies im Prinzip funktioniert. Zufallsgeneratoren beruhen auf der Konstruktion von *gleichverteilten Folgen*. Betrachten wir die Zahlenfolge

$$a_n = n\alpha - \lfloor n\alpha \rfloor, \quad n \geq 1.$$

Dabei bedeutet $\lfloor x \rfloor$ die größte ganze Zahl kleiner gleich x , sodass also a_n als Vielfaches einer festen Zahl α entsteht, wobei aber nur die Nachkommastellen erhalten werden. Zum Beispiel ist die Folge für $\alpha = 1/3$

$$\left(\frac{1}{3}, \frac{2}{3}, 1, \frac{1}{3}, \frac{2}{3}, 1, \frac{1}{3}, \frac{2}{3}, 1, \dots\right)$$

klarerweise periodisch, wie immer, wenn α eine rationale Zahl ist. Nimmt man allerdings α irrational, dann ist die Folge a_n für $n \rightarrow \infty$ im Intervall $[0, 1]$ gleichverteilt, das heißt,

die relative Anzahl der Folgenglieder, die in ein Teilintervall von $[0, 1]$ fallen, ist näherungsweise gleich der Teilintervalllänge. In Abbildung 6.1 wurden die ersten $n = 1000$ Folgenglieder der Folge mit $\alpha = \sqrt{2}$ erzeugt. Man sieht, dass ganz gut eine Gleichverteilung entstanden ist. Mittelwert und Standardabweichung dieser 1000 Zahlenwerte sind übrigens gleich 0.4999 und 0.2887, was recht gut mit den theoretischen Werten einer Gleichverteilung (0.5 und $1/\sqrt{12}$) übereinstimmt.

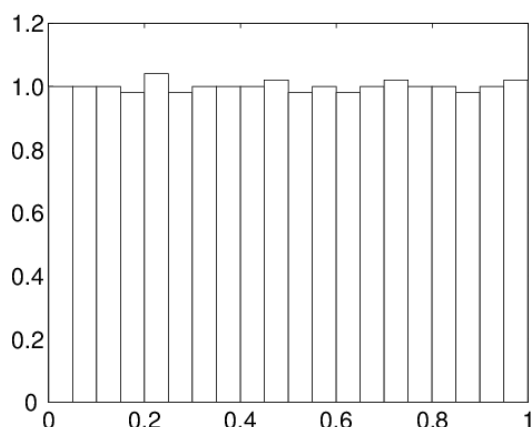


Abbildung 6.1: Histogramm von 1000 Zufallszahlen aus einer Gleichverteilung.

Da mittels Computerarithmetik nur endliche Dezimalzahlen, also rationale Zahlen, erzeugt werden können, ist die Folge der Zufallszahlen, sofern von einem Computer berechnet, tatsächlich periodisch – allerdings mit riesiger Periodenlänge. Man spricht daher oft von *Pseudozufallszahlen*.

Es gibt zahlreiche Methoden, aus gleichverteilten Zufallszahlen solche mit anderen Verteilungen zu berechnen. Die wichtigste ist wohl die *Inversionsmethode*. Sei F eine Verteilungsfunktion, gemäß derer Zufallszahlen erzeugt werden sollen. Hat man N gleichverteilte Zufallszahlen x_1, \dots, x_N , so besagt die Inversionsmethode, dass die Zufallszahlen

$$F^{-1}(x_1), \dots, F^{-1}(x_N)$$

die durch F definierte Verteilung besitzen.

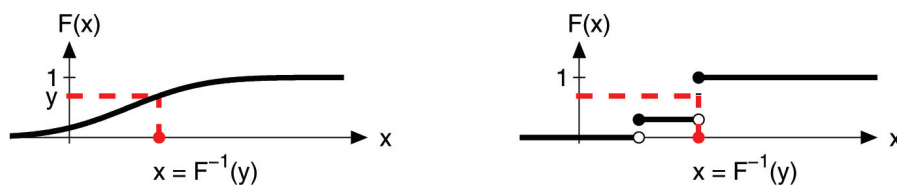


Abbildung 6.2: Illustration der Inversionsmethode.

In der Tat, die Verteilungsfunktion einer gleichverteilten Größe X ist

$$F_X(x) = \begin{cases} x, & 0 \leq x \leq 1, \\ 0, & x < 0 \text{ oder } x > 1. \end{cases}$$

Wir setzen $Y = F^{-1}(X)$. Dann ist

$$F_Y(y) = P(Y \leq y) = P(F^{-1}(X) \leq y) = P(X \leq F(y)) = F_X(F(y)) = F(y).$$

Somit stimmt die Verteilungsfunktion von Y tatsächlich mit der gegebenen Verteilungsfunktion F überein.

In Abbildung 6.3 wurden standardnormalverteilte Zufallszahlen aus den obigen 1000 gleichverteilten Zufallszahlen mittels der Transformation $Y = \Phi^{-1}(X)$ erzeugt. Das Ergebnis ist zufrieden stellend; auch Mittelwert und Standardabweichung stimmen mit -0.0012 und 0.9984 mit den theoretischen Werten 0 und 1 gut überein.

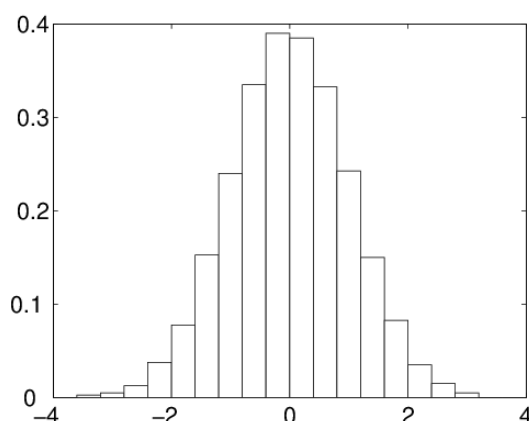
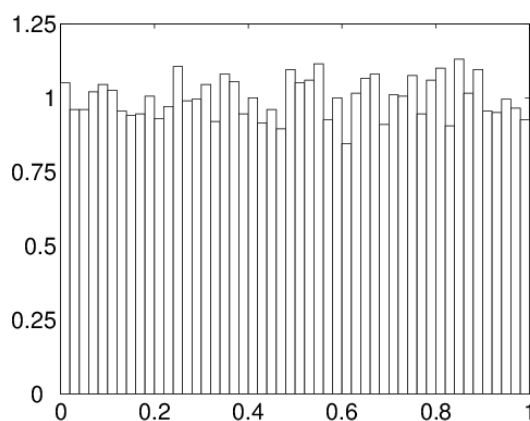
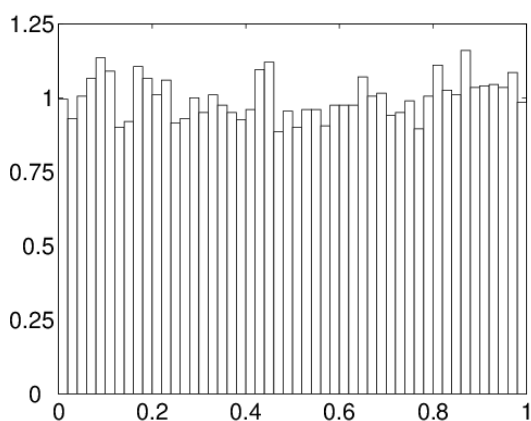


Abbildung 6.3: Histogramm von 1000 Zufallszahlen aus einer Normalverteilung.

Ein Simulationsbeispiel: Gegeben seien drei unabhängige Zufallsgrößen X, Y, Z , welche eine Gleichverteilung im Intervall $[0, 1]$ besitzen mögen. Gesucht ist die Verteilung der Größe $V = XY + Z$. Mittels Monte-Carlo-Simulation gehen wir folgendermaßen vor. Wir erzeugen durch einen Zufallsgenerator je $N = 10000$ gleichverteilte Zufallszahlen x, y, z . Wir berechnen N -mal $v = xy + z$; die empirische Verteilung der N Werte v ist eine Näherung an die Verteilung der Zufallsgröße V . Die Abbildung 6.4 zeigt Histogramme der 10000 Realisierungen der drei gleichverteilten Input-Größen X, Y, Z und der Output-Größe V .



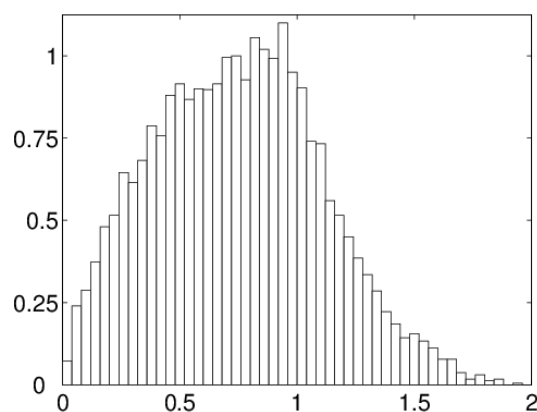
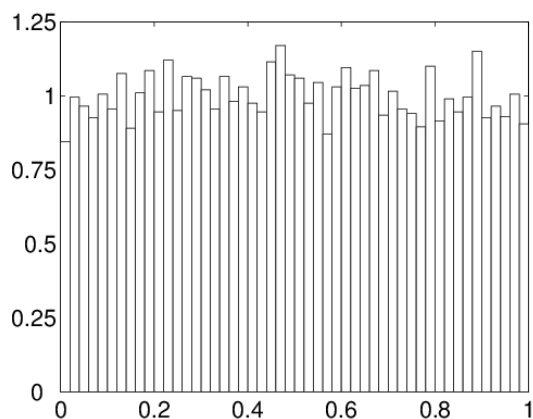


Abbildung 6.4: Histogramm von 1000 Zufallszahlen aus drei Gleichverteilungen sowie deren Verknüpfung.

Abbildung 6.5 zeigt beispielhaft zwanzig (mit Excel erzeugte) Realisierungen der Zufallsgrößen X, Y, Z und die zugehörigen Werte von $V = XY + Z$.

Nummer	X	Y	Z	$X \cdot Y + Z$
1	0,194525	0,328623	0,252541	0,316466
2	0,401898	0,990875	0,212134	0,610365
3	0,786340	0,241859	0,164678	0,354861
4	0,556688	0,312937	0,506485	0,680693
5	0,261147	0,324778	0,650594	0,735408
6	0,124577	0,843837	0,275552	0,380674
7	0,672597	0,523484	0,988769	1,340863
8	0,607105	0,005921	0,658742	0,662336
9	0,819849	0,475570	0,098575	0,488470
10	0,936827	0,366314	0,547624	0,890797
11	0,187567	0,436232	0,523881	0,605703
12	0,242042	0,198401	0,334086	0,382108
13	0,400525	0,817591	0,107547	0,435013
14	0,943693	0,828272	0,962127	1,743762
15	0,799005	0,269845	0,322336	0,537944
16	0,040498	0,946104	0,488113	0,526428
17	0,268868	0,656758	0,950407	1,126989
18	0,253792	0,622181	0,686758	0,844662
19	0,467971	0,247749	0,893124	1,009064
20	0,438032	0,555010	0,864193	1,107305

Abbildung 6.5: Illustration der numerischen Berechnung mittels Monte-Carlo-Simulation.

Kapitel 7

Schätzen von Parametern

In diesem Kapitel führen wir die Methoden der beschreibenden Statistik und der wahrscheinlichkeitstheoretischen Modellbildung zusammen und gehen damit den entscheidenden Schritt, der die *schließende Statistik* möglich macht.

Ziel ist der Schluss von der Stichprobe auf die Grundgesamtheit (etwa in der Sozial- und Wirtschaftsstatistik) oder auf die Verteilung der untersuchten Größe (etwa bei Boden- oder Materialparametern in den Ingenieurwissenschaften).

Gegenstand der Untersuchung ist eine Zufallsgröße X mit bekanntem oder zunächst postulierten Verteilungstyp, z.B. eine Normalverteilung $\mathcal{N}(\mu, \sigma^2)$, eine Alternativverteilung mit Parameter p , eine Exponentialverteilung mit Parameter λ . Die Parameter sind unbekannt und sollen aus einer Stichprobe X_1, \dots, X_n geschätzt werden.

Beispiel 7.1 (Die Sonntagsfrage) Es soll erhoben werden, welcher Anteil p der wahlberechtigten Österreicher (Grundgesamtheit) eine Partei XXX wählen würde. Es handelt sich offenbar um eine Alternativverteilung mit Ausgängen *ja* – 1 und *nein* – 0. Der Verteilungstyp ist in diesem Beispiel bekannt. Der Wert p soll aus einer Stichprobe von n wahlberechtigten Personen geschätzt werden.

Beispiel 7.2 (Mittelwert einer Grundgesamtheit) Zu Vergleichszwecken ist oft verlangt, den Mittelwert als Kenngröße einer Grundgesamtheit zu ermitteln, etwa das mittlere Bruttoeinkommen unselbständig Beschäftigter oder das mittlere Körpergewicht von Schülern. Wir fassen diesen Mittelwert als Parameter μ der Grundgesamtheit auf und wollen ihn aus einer Stichprobe kleineren Umfangs schätzen. Für diese Schätzung spielt der Verteilungstyp der Grundgesamtheit keine Rolle, wie wir sehen werden.

Beispiel 7.3 (Lebensdauer von Bauteilen) Aus theoretischen Überlegungen kann hergeleitet werden, dass Lebensdauern von technischen Geräten und Bauteilen näherungsweise exponentialverteilt sind. Die Geräteausfälle, über einen längeren Zeitraum beobachtet, können als Stichprobe herangezogen werden, um den Parameter λ der Exponentialverteilung zu schätzen.

Beispiel 7.4 (Hochwasserstatistik) Zur Bemessung von Dämmen an einem Fließgewässer muss das 100- oder 1000-jährige Hochwasser abgeschätzt werden. Man nimmt oft eine

Gumbelverteilung an, für die es eine gewisse plausible Herleitung gibt. Andere Extremwertverteilungen sind ebenfalls in Gebrauch. Jedenfalls wird aus einer Reihe von gemessenen Jahresmaxima auf die Parameter der Extremwertverteilung geschlossen. Daraus können die Jährlichkeiten als Quantile abgelesen werden.

Beispiel 7.5 (Aus der Bodenmechanik) Zur Bemessung eines Fundaments nach den probabilistischen Normen sind die statistischen Kennwerte der Bodenparameter zu erheben. Die Quantile der Verteilung gehen direkt in die Bemessung nach Norm ein. In Tabelle 7.2 sind die Ergebnisse für den Reibungswinkel von $n = 20$ Rahmenscherversuchen angegeben. Für die Bemessung interessiert der Reibungskoeffizient $\nu = \tan \varphi$. Um die Quantile abschätzen zu können, wird in der Regel angenommen, dass der Reibungskoeffizient normalverteilt ist. Die Parameter μ, σ^2 der Normalverteilung werden aus der Stichprobe geschätzt.

In den letzten beiden Beispielen ist der Verteilungstyp ein plausibles Postulat. Es sollte überprüfbar sein, ob solche Postulate zutreffend sind. Die schließende Statistik liefert auch dazu Entscheidungskriterien: aus der Stichprobe kann abgeleitet werden, ob die postulierte Verteilung akzeptabel ist: dies ist Teil der im nächsten Kapitel zu behandelnden statistischen Testtheorie.

7.1 Punktschätzungen

Im Abschnitt 4.5 haben wir die Bedeutung der Parameter der wichtigsten Verteilungen beschrieben. Wir verwenden nun die Entsprechungen Modell – Stichprobe mit

$$E(X) \approx \bar{x}, \quad V(X) \approx s^2$$

der frequentistischen Sichtweise, um so genannte *Schätzer* für die Verteilungsparameter zu gewinnen. In statistischer Tradition werden die Schätzer mit einem *Hut* gekennzeichnet, um sie von den Modellparametern zu unterscheiden. Einige wichtige Beispiele sind:

- (i) Parameter $\mu = E(X)$
Schätzung $\hat{\mu} = \bar{x}$, das Stichprobenmittel;
- (ii) Parameter $\sigma^2 = V(X)$
Schätzung $\hat{\sigma}^2 = s^2$, die Stichprobenvarianz;
- (iii) Alternativverteilung, Parameter $p = P(X = 1)$
Schätzung $\hat{p} = \text{Anteil der 1-Ausgänge}$;
- (iv) Exponentialverteilung, Parameter $\lambda = 1/E(X) = 1/\sqrt{V(X)}$
Schätzung $\hat{\lambda} = 1/\bar{x}$ oder $\hat{\lambda} = 1/s$.

Beispiel 7.6 (Begleitendes Beispiel) Wir verwenden in der Folge als Illustrationsbeispiel eine Erhebung der Körpergröße von 44 männlichen Bauingenieurstudenten aus 1998 (Tabelle 7.1). Wir fassen diese Erhebung als Stichprobe vom Umfang $n = 44$ für das Merkmal *Körpergröße männlicher österreichischer Studenten* auf. Wir postulieren (und werden dies im nächsten Kapitel überprüfen), dass diese Größe X einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ unterliegt.

Gesucht ist der Erwartungswert μ (der Mittelwert der Grundgesamtheit aller österreichischen Studenten) und die Varianz σ^2 .

Die Stichprobe vom Umfang $n = 44$ ergab:

$$\sum_{i=1}^{44} x_i = 7915, \quad \bar{x} = \frac{7915}{44} = 179.9,$$

$$\sum (x_i - \bar{x})^2 = 2459, \quad s^2 = \frac{2459}{43} = 57.2, \quad s = 7.6.$$

Somit ist die Punktschätzung

$$\hat{\mu} = 179.9, \quad \hat{\sigma} = 7.6.$$

Die *geschätzte Verteilung* ergibt sich somit als Normalverteilung $\mathcal{N}(179.9, 57.2)$. In der Abbildung 7.1 ist der ein- und zweifache Streubereich ablesbar.

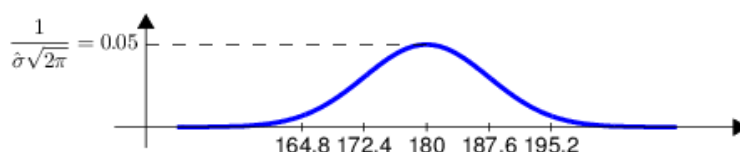


Abbildung 7.1: Verteilungsdichte der Körpergröße von Studenten, Anpassung gemäß Stichprobe.

Gütekriterien für Punktschätzungen: Es gibt meistens verschiedene Schätzer $\hat{\theta}$ für denselben Parameter θ , zum Beispiel für den Parameter λ einer Exponentialverteilung (Punkt (iv) oben) oder für die Normalverteilung, bei der man auch den Median der Stichprobe als Schätzer für den Zentralwert μ nehmen könnte. Somit stellt sich die Frage, welcher Schätzer vorzuziehen ist. Dazu stellt die Schätztheorie *Gütekriterien* auf. Einige dieser Gütekriterien sind die Erwartungstreue, die Effizienz und die Plausibilität:

(i) *Erwartungstreue:* $E(\hat{\theta}) = \theta$. Das heißt, wird $\hat{\theta}$ aus vielen Stichproben wiederholt bestimmt, so sollte im Mittel θ herauskommen.

Beispiel 7.7 Ist $\mu = E(X)$, so ist $\hat{\mu} = \bar{X}$ erwartungstreu, da nach Abschnitt 6.2 gilt:

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

Beispiel 7.8 Ist $\sigma^2 = V(X)$, so liefert die Stichprobenvarianz

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

einen erwartungstreuen Schätzer. Aus diesem Grund erscheint der Faktor $(n-1)$ in der Formel für die Stichprobenvarianz.

Zur Begründung machen wir die folgende Rechnung:

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{n-1} \sum_{i=1}^n E((X_i - \bar{X})^2) = \frac{1}{n-1} \sum_{i=1}^n E((X_i - \mu - (\bar{X} - \mu))^2) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E((X_i - \mu)^2) - 2 \sum_{i=1}^n E\left((X_i - \mu) \left(\frac{1}{n} \sum_{j=1}^n X_j - \mu\right)\right) + \sum_{i=1}^n E((\bar{X} - \mu)^2) \right). \end{aligned}$$

Als Realisierungen in der Stichprobe sind die X_i unabhängig und identisch wie X verteilt, daher gilt

$$E((X_i - \mu)^2) = \sigma^2, \quad E((X_i - \mu)(X_j - \mu)) = 0 \text{ für } i \neq j, \quad E((\bar{X} - \mu)^2) = V(\bar{X}) = \frac{\sigma^2}{n}$$

und somit weiter

$$E(\hat{\sigma}^2) = \frac{1}{n-1} \left(n \cdot \sigma^2 - 2n \cdot \frac{1}{n} \sigma^2 + n \cdot \frac{\sigma^2}{n} \right) = \frac{1}{n-1} (n\sigma^2 - 2\sigma^2 + \sigma^2) = \sigma^2.$$

(ii) *Effizienz*: Die Schätzvarianz

$$V(\hat{\theta}) = E((\hat{\theta} - \theta)^2)$$

sollte möglichst klein sein.

Beispiel 7.9 $\hat{\mu} = \bar{X}$ ist der effizienteste (lineare) Schätzer für $\mu = E(X)$.

(iii) *Plausibilität (Maximum Likelihood)*: Der Schätzer $\hat{\theta}$ sollte jenen Parameterwert θ treffen, für den die vorliegende Stichprobe die größte Wahrscheinlichkeit hat.

Beispiel 7.10 Das Stichprobenmittel \bar{X} ist Maximum-Likelihood-Schätzer für μ . Für σ^2 ist jedoch $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ der Maximum-Likelihood-Schätzer.

Parameterschätzungen sind mit zwei Fehlerquellen behaftet.

1. *Modellfehler*: Annahme oder Vorgabe des Verteilungstyps. Dies liegt in der Verantwortung der Bearbeiter. Im nächsten Kapitel werden wir sehen, wie die Berechtigung einer Verteilungsannahme überprüft werden kann.
2. *Zufälliger Fehler der Stichprobenparameter*: Zum Beispiel wird das Stichprobenmittel \bar{X} bei verschiedenen Stichproben verschieden sein. Dieser Fehler kann statistisch abgeschätzt werden und führt auf den Begriff der *Konfidenzschätzung*.

7.2 Konfidenzschätzungen

Die so genannten Konfidenzschätzungen ergänzen die Punktschätzungen durch Angabe einer Schwankungsbreite, die den statistischen Fehler abgrenzt, der durch die zufällige Wahl der Stichprobe entsteht.

Gegeben sei eine Zufallsgröße mit Erwartungswert und Varianz

$$E(X) = \mu, \quad V(X) = \sigma^2.$$

Aus einer Stichprobe vom Umfang n werde der Parameter μ durch das Stichprobenmittel

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$$

geschätzt. Wie oben erwähnt, würde das Ziehen einer anderen Stichprobe einen anderen Wert für \bar{X} ergeben. Somit ergibt sich die Frage nach dem Fehler des Schätzers \bar{X} – wie weit ist \bar{X} von μ entfernt? Um eine Abschätzung des Fehlers herzuleiten, erinnern wir daran, dass gilt:

$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n}.$$

Nach dem zentralen Grenzwertsatz ist \bar{X} für großes n näherungsweise $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ -verteilt (falls X nach $\mathcal{N}(\mu, \sigma^2)$ verteilt ist, sogar exakt). Die zentrierte Größe

$$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

ist demnach $\mathcal{N}(0, 1)$ -verteilt.

Frage: Gibt es eine Schranke β , sodass mit 95%-Sicherheit (oder auch 90%, 99% ...) der tatsächliche Wert μ zwischen $\bar{X} - \beta$ und $\bar{X} + \beta$ liegt? So eine Schranke liefert das 95%-Konfidenzintervall $[\bar{X} - \beta, \bar{X} + \beta]$.

Das 95%-Konfidenzintervall gibt eine *Bereichsschätzung* für μ an mit der Zusatzaussage, dass es in $\sim 95\%$ aller Fälle (aller Stichproben vom Umfang n) den tatsächlichen Wert μ enthält. Mögliche Ergebnisse von solchen Intervallen für fünf Stichproben sind in Abbildung 7.2 dargestellt.

Berechnung von β : Wir bestimmen zunächst α so, dass gilt:

$$P(-\alpha \leq U \leq \alpha) = 0.95.$$

Wegen der Symmetrie der Gaußkurve ist dies gleichbedeutend mit

$$P(-\infty < U \leq \alpha) = 0.975,$$

man vergleiche dazu Abbildung 7.3. Aus der Tabelle erhält man $\alpha = 1.96$. Also gilt mit $\sim 95\%$ Sicherheit

$$-1.96 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 1.96.$$

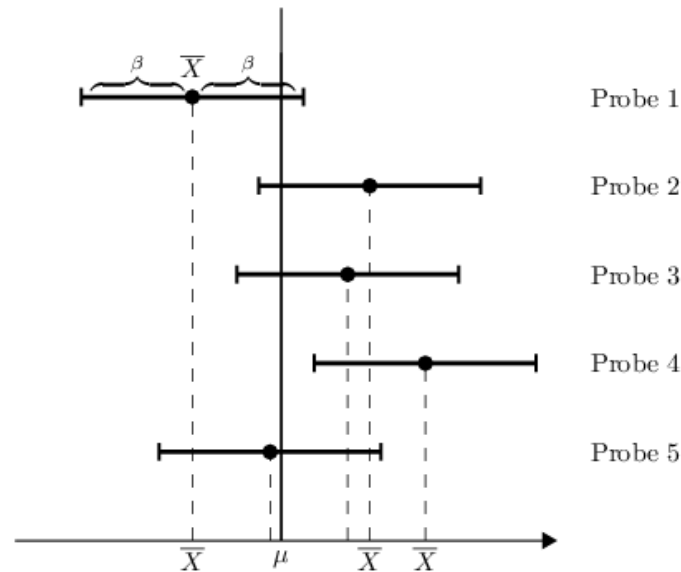


Abbildung 7.2: Verschiedene Stichprobenmittel derselben Zufallsgröße und zugehörige Konfidenzintervalle.

Weiteres Auflösen der Ungleichung führt auf

$$-1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}}, \quad \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}.$$

Damit ist die Schranke

$$\beta = 1.96 \frac{\sigma}{\sqrt{n}}$$

und das 95%-Konfidenzintervall $[\bar{X} - 1.96 \sigma / \sqrt{n}, \bar{X} + 1.96 \sigma / \sqrt{n}]$ gefunden.

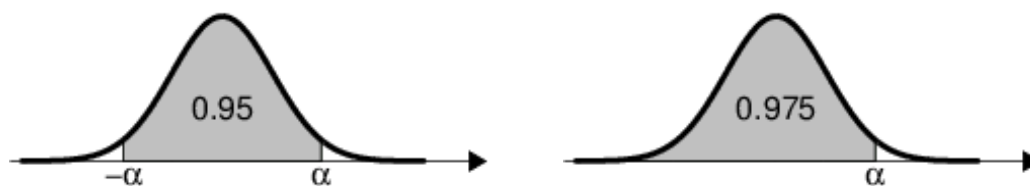


Abbildung 7.3: Quantile der Standardnormalverteilung; Symmetrieeigenschaft.

Wie geht man nun vor, wenn man eine Stichprobe vom Umfang n genommen hat? Zunächst wird das Stichprobenmittel \bar{x} berechnet. Ist die Streuung σ bekannt, so kann man $[\bar{x} - \beta, \bar{x} + \beta]$ berechnen. Ist σ unbekannt, so kann man für großes n ($n \geq 50$) stattdessen die Stichprobenvarianz nehmen, also $\sigma \approx s$ setzen (für kleines n siehe unten).

Beispiel 7.11 Wir setzen das begleitende Beispiel fort (Körpergröße X männlicher Studierender) und berechnen das 95%-Konfidenzintervall für den Erwartungswert μ . Aus den Stichprobenparametern

$$\bar{x} = 179.9, \quad s = 7.6, \quad n = 44$$

erhalten wir

$$\beta = 1.96 \frac{7.6}{\sqrt{44}} \approx 2.2.$$

Also gilt mit 95% Sicherheit (Irrtum nur bei 5% aller Stichproben) für den Erwartungswert:

$$179.9 - 2.2 \leq \mu \leq 179.9 + 2.2, \quad \text{d.h.} \quad 177.7 \leq \mu \leq 182.1.$$

Bemerkung: (a) Das Konfidenzintervall wird schmaler, wenn n größer wird. (b) Für erhöhte Sicherheit der Aussage, etwa 99%, wird es breiter. Für 99%ige Aussagesicherheit erhält man aus $P(U \leq \alpha) = 0.995$ die Lösung $\alpha = 2.575$ und

$$\beta = 2.575 \frac{7.6}{\sqrt{44}} \approx 3.0,$$

also das 99%-Konfidenzintervall

$$176.9 \leq \mu \leq 182.9.$$

Modifikation für kleine Stichproben: Falls X nach $\mathcal{N}(\mu, \sigma^2)$ verteilt ist, so besitzt die Größe

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

eine so genannte *t-Verteilung mit $(n - 1)$ Freiheitsgraden*. Für Stichproben vom Umfang $n < 50$ sollte man statt der Näherung durch die Normalverteilung die *t-Verteilung* verwenden und die Stichprobenvarianz einsetzen. Die Dichte der *t-Verteilung* ist glockenförmig, aber flacher als die Normalverteilung. Die Quantile sind tabelliert (vgl. Tabelle 7.4). Ansonsten verfährt man genauso wie oben.

Beispiel 7.12 Wir machen die Rechnung aus Beispiel 7.11 mit Hilfe der *t-Verteilung*. Bei $n = 44$ hat man 43 Freiheitsgrade. Man erhält $\alpha = t_{0.975} = 2.017$. Somit ist genauer

$$\beta = 2.017 \frac{s}{\sqrt{n}} = 2.017 \frac{7.6}{\sqrt{44}} \approx 2.3.$$

Unter Verwendung der *t-Verteilung* erhält man also als 95%-Konfidenzintervall

$$179.9 - 2.3 \leq \mu \leq 179.9 + 2.3, \quad \text{d.h.} \quad 177.6 \leq \mu \leq 182.2.$$

7.3 Hochwasserstatistik

Die Hochwasserstatistik mit ihrer Prognose der Jährlichkeiten beruht auf der so genannten Extremwertstatistik, welche die Verteilung maximaler oder minimaler Messwerte beschreibt. Wir beginnen zunächst mit einer theoretischen Einführung.

7.3.1 Extremwertverteilungen

Eine der grundlegenden Fragestellungen der Extremwertstatistik ist die Ermittlung der Verteilung des Stichprobenmaximums, also des größten Werts einer Stichprobe beliebigen Umfangs n . Das heißt, nimmt man N Stichproben vom Umfang n und ist X_j der größte Wert der Stichprobe Nummer j , so ist die Verteilung dieser größten Werte X_1, \dots, X_N gesucht.

Die *Gumbelverteilung* ist eine der Standardverteilungen zur Beschreibung der Stichprobenmaxima. Sie trifft asymptotisch zu, wenn die zu Grunde liegende Größe exponentialverteilt ist, wird aber häufig auch angewendet, wenn man annehmen kann, dass die Verteilungsdichte der zu Grunde liegenden Größe im Unendlichen exponentiell abfällt. Für anderes Abklingverhalten werden andere Extremwertverteilungen herangezogen; bekannt sind zum Beispiel die *Weibullverteilung* oder die *Pearson-3-Verteilung*.

Die Dichte der Gumbelverteilung wird durch zwei Parameter λ und x_0 beschrieben:

$$f_X(x) = \lambda e^{-\lambda(x-x_0)} e^{-(e^{-\lambda(x-x_0)})}.$$

Dabei ist x_0 der Modalwert und λ ein Skalierungsfaktor. Die Transformation $y = \lambda(x - x_0)$ ergibt die Standardform

$$f_Y(y) = e^{-y} e^{-(e^{-y})}.$$

Die Verteilungsfunktion der Gumbelverteilung ist

$$F_X(x) = e^{-(e^{-\lambda(x-x_0)})}.$$

Ihr Erwartungswert und ihre Varianz sind gegeben durch (vgl. Abbildung 4.12 für $x_0 = 0$, $\lambda = 1$):

$$\mu = E(X) = x_0 + \gamma/\lambda, \quad \sigma^2 = V(X) = \pi^2/(6\lambda^2); \quad \gamma \approx 0.577216 \dots$$

Diese Formeln bilden die Grundlage der *Parameterschätzung*, mit deren Hilfe die Spezifikation der an die Messdaten angepassten Gumbelverteilung vorgenommen wird. Es folgt nämlich aus den Formeln:

$$\lambda = \frac{\pi}{\sigma\sqrt{6}}, \quad x_0 = \mu - \frac{\gamma}{\lambda} \tag{7.1}$$

und μ bzw. σ lassen sich aus der Liste der Extremwerte als Mittelwert und Streuung schätzen.

Herleitung der Gumbelverteilung: Sei T exponentialverteilt mit Verteilungsfunktion

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0. \end{cases}$$

Die Verteilung des größten Werts $X = T_{\max}$ einer Stichprobe T_1, T_2, \dots, T_n vom Umfang n erhält man durch die Überlegung, dass X genau dann kleiner gleich einem gegebenen x ist, wenn alle T_1, T_2, \dots, T_n kleiner gleich x sind. Also:

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(T_1 \leq x, T_2 \leq x, \dots, T_n \leq x) \\ &= P(T_1 \leq x) \cdot P(T_2 \leq x) \cdots P(T_n \leq x) = (1 - e^{-\lambda x})^n \end{aligned}$$

für $x > 0$, da sich die Wahrscheinlichkeiten wegen der Unabhängigkeit der Größen T_1, T_2, \dots, T_n multiplizieren. Wir benützen nun die Euler'sche Formel $\lim_{n \rightarrow \infty} (1 + y/n)^n = e^y$ und erhalten asymptotisch für $n \rightarrow \infty$

$$(1 - e^{-\lambda x})^n \approx e^{-n(e^{-\lambda x})}.$$

Schreibt man nun n in der Form $n = e^{\lambda x_0}$, so ergibt sich nach Umformung des Exponenten die gesuchte Formel für die Gumbelverteilung

$$F_X(x) \approx e^{-(e^{-\lambda(x-x_0)})}.$$

7.3.2 Hochwasserprognose

In der Hochwasserstatistik betrachtet man die Liste der Monatswerte für HQ (Monatsmaxima) eines Jahres als Stichprobe vom Umfang $n = 12$. Beobachtet man N Jahre, so erhält man für Jahr Nummer j das Jahresmaximum X_j . Diese Jahresmaxima können durch eine Gumbelverteilung näherungsweise beschrieben werden.

Ihre Parameter erhält man, indem man den Erwartungswert aus einer mehrjährigen Reihe durch deren Mittelwert schätzt und analog die Varianz. Die Parameter der Gumbelverteilung lassen sich dann aus Formel (7.1) berechnen. An der kumulativen Verteilungsfunktion der Maxima lassen sich die Quantile ablesen. Dabei ist das $p \cdot 100\%$ -Quantil Q_p einer Zufallsgröße X mit Verteilung F_X definiert durch

$$F_X(Q_p) = P(X \leq Q_p) = p, \quad Q_p = F_X^{-1}(p).$$

Im Falle der Gumbelverteilung erhält man

$$Q_p = x_0 - \frac{1}{\lambda} \log(-\log p). \quad (7.2)$$

Das Ereignis, dass ein Messwert oberhalb des 99%-Quantils liegt, hat die Wahrscheinlichkeit von $0.01 = 1/100$. Dies ist die so genannte *Überschreitungswahrscheinlichkeit*, deren Kehrwert die *Jährlichkeit* darstellt. Man spricht dann von einem 100-jährigen Ereignis.

Wir wollen nun den Zusammenhang zwischen Quantilen, Überschreitungswahrscheinlichkeit und Jährlichkeit genauer ansehen.

Unterschreitungswahrscheinlichkeit: p ;

Überschreitungswahrscheinlichkeit: $p_{\text{ü}} = 1 - p$;

Jährlichkeit: $J = 1/p_{\text{ü}} = 1/(1 - p)$;

J -jähriges Hochwasser: $HQ_J = Q_p$.

Das 100-jährige Hochwasser erhält man als das 99%-Quantil der kumulativen Verteilungsfunktion, das 50-jährige als das 98%-Quantil usw., tabellarisch:

p	0.5	0.9	0.95	0.98	0.99	0.999
$p_{\text{ü}}$	0.5	0.1	0.05	0.02	0.01	0.001
J	2	10	20	50	100	1000
HQ	HQ ₂	HQ ₁₀	HQ ₂₀	HQ ₅₀	HQ ₁₀₀	HQ ₁₀₀₀
Q_p	$Q_{0.5}$	$Q_{0.9}$	$Q_{0.95}$	$Q_{0.98}$	$Q_{0.99}$	$Q_{0.999}$

Beispiel 7.13 Als Beispiel nehmen wir die Jahresmaxima 1971 – 1980 des Inns am Pegel Innsbruck nach Tabelle 7.3. Wir erhalten

$$\bar{x} = 675.5, \quad s = 171.234, \quad \hat{\lambda} = 0.0075, \quad \hat{x}_0 = 598.4.$$

Die angepasste Gumbelverteilung sowie die Prognose auf 1000 Jahre ist den Abbildungen 7.4 und 7.5 zu entnehmen.

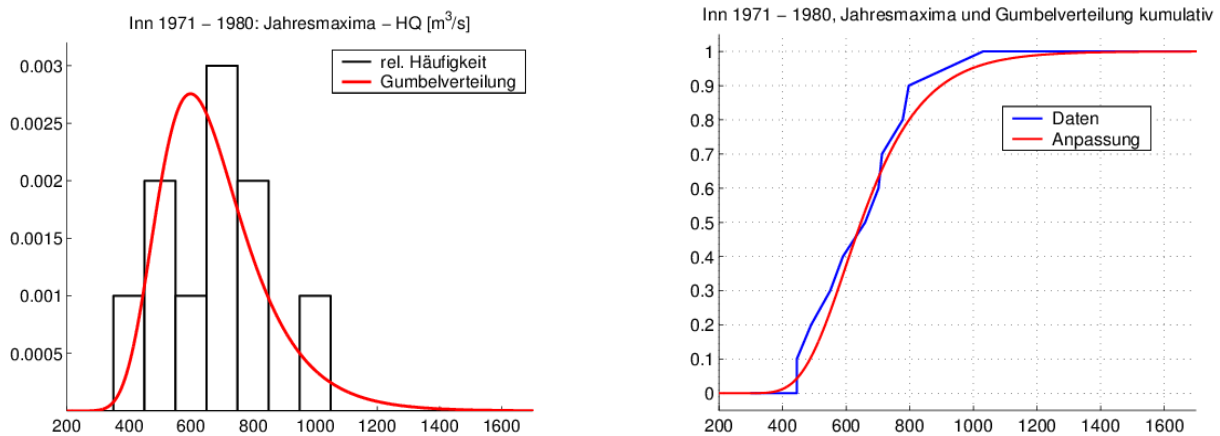


Abbildung 7.4: Jahresmaxima: Histogramm und Dichte (links), empirische Summenkurve und Verteilungsfunktion (rechts).

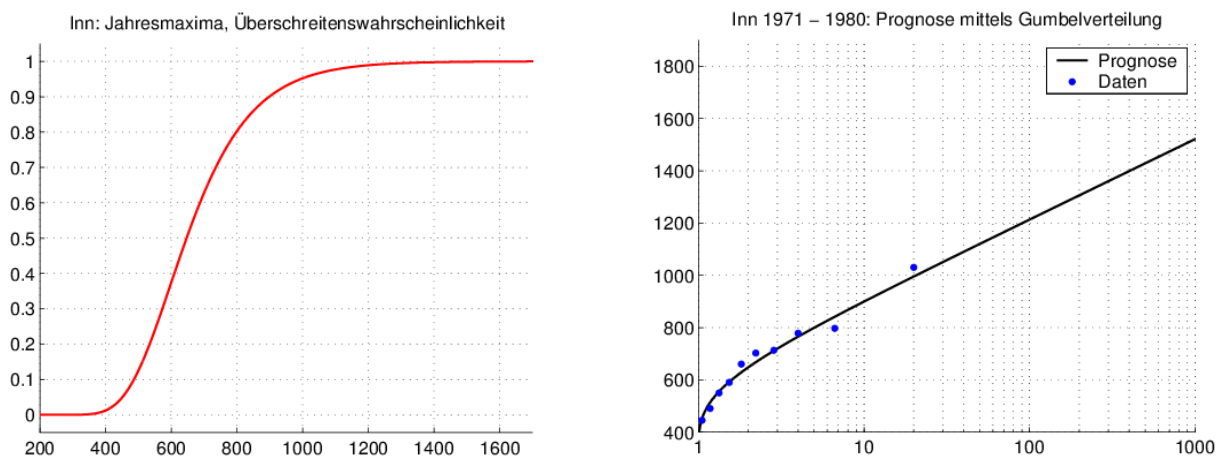


Abbildung 7.5: Jahresmaxima: Verteilungsfunktion der Gumbelverteilung (links) und Prognose der Jährlichkeit in halblogarithmischer Darstellung (rechts).

Die Quantile können mittels Formel (7.2) berechnet oder aus Abbildung 7.5 (links) abgelesen werden, zum Beispiel

$$Q_{0.9} = HQ_{10} = 899 \text{ m}^3/\text{s}, \quad Q_{0.95} = HQ_{20} = 995 \text{ m}^3/\text{s}, \quad Q_{0.99} = HQ_{100} = 1213 \text{ m}^3/\text{s}.$$

Im Wasserbau ist die halblogarithmische Darstellung der Jährlichkeit in Abbildung 7.5 (rechts) gebräuchlich. Man erhält sie, indem man J gegen HQ_J aufträgt.

Bemerkung: Die Datenpunkte werden in Abbildung 7.5 (rechts) mit Hilfe einer modifizierten empirischen Verteilungsfunktion eingefügt. Dazu ordnet man zunächst die $N = 10$ Daten für HQ aus Tabelle 7.3 der Größe nach. Dann weist man HQ_i die Unterschreitenswahrscheinlichkeit $p_i = (2i - 1)/2N$ und die Jährlichkeit $J_i = 1/(1 - p_i)$ zu, $i = 1, \dots, N$. Diese Punkte werden dann in halblogarithmischer Darstellung in der Abbildung geplottet. Der Grund für die Modifikation liegt darin, dass die undefinierten Fälle einer Jährlichkeit von $J = 1$ oder $J = \infty$, welche $p = 0$ bzw. $p = 1$ entsprechen, vermieden werden müssen.

7.3.3 Konfidenzgrenzen für die Quantile

Es erscheint etwas gewagt, von einer 10-jährigen Datenreihe auf 1000 Jahre zu schließen. Daher muss die Prognose mit einer Fehlerabschätzung versehen werden, die man aus den Konfidenzintervallen für das Quantil Q_p gewinnt. Unter der Annahme einer Normalverteilung für das Quantil erhält man für das 95%-Konfidenzintervall:

$$[Q_p - 1.96 \hat{\sigma}_{Q_p}, Q_p + 1.96 \hat{\sigma}_{Q_p}],$$

wobei $\hat{\sigma}_{Q_p}$ eine Schätzung für die Standardabweichung des Quantils darstellt.

Allgemein kann die Varianz des Quantils einer Zufallsgröße X wie folgt abgeschätzt werden. Die Größe X besitze den Erwartungswert μ , die Varianz σ^2 , den Schiefekoeffizienten C_s und die Kurtosis K (siehe unten). Das standardisierte Quantil ist

$$K_p = \frac{Q_p - \mu}{\sigma}.$$

Wird das Quantil Q_p aus einer Stichprobe vom Umfang n durch Anpassung der Verteilungsfunktion $F_X(x)$ geschätzt, so kann für seine Varianz folgende Näherungsformel angegeben werden; wir folgen hier Plate(1993):

$$\sigma_{Q_p}^2 = V(Q_p) = \frac{\sigma^2}{n} \left(1 + K_p C_s + \frac{K_p^2}{4} (K - 1) \right).$$

Im Spezialfall der Gumbelverteilung gilt (unabhängig von deren Parametern)

$$C_s = 1.1396, \quad K = 5.4.$$

Damit erhält man folgende Schätzung der Varianz des Quantils Q_p einer Gumbelverteilung:

$$\hat{\sigma}_{Q_p}^2 \approx \frac{s^2}{n} \left(1 + 1.1396 K_p + 1.1 K_p^2 \right)$$

mit

$$K_p \approx \frac{Q_p - \bar{x}}{s}, \quad Q_p = x_0 - \log(-\log p)/\hat{\lambda}.$$

Mit Hilfe dieser Formeln können wir ein Plot der 1000-jährigen Prognose mit Konfidenzintervallen erzeugen (Abbildung 7.6).

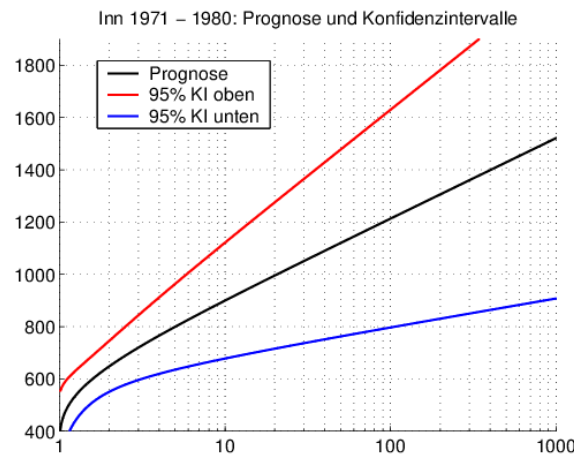


Abbildung 7.6: Jahresmaxima: Prognose der Jährlichkeit in halblogarithmischer Darstellung mit 95%-Konfidenzgrenzen.

Die Streuung der Prognose des 1000-jährigen Hochwassers ist enorm. Dies liegt an der Kürze der gemessenen Datenreihe. Macht man dieselbe Rechnung mit den Jahresmaxima des Inn von 1870 bis 2005, so erhält man die wesentlich aussagekräftigeren Ergebnisse von Abbildung 7.7 und 7.8.

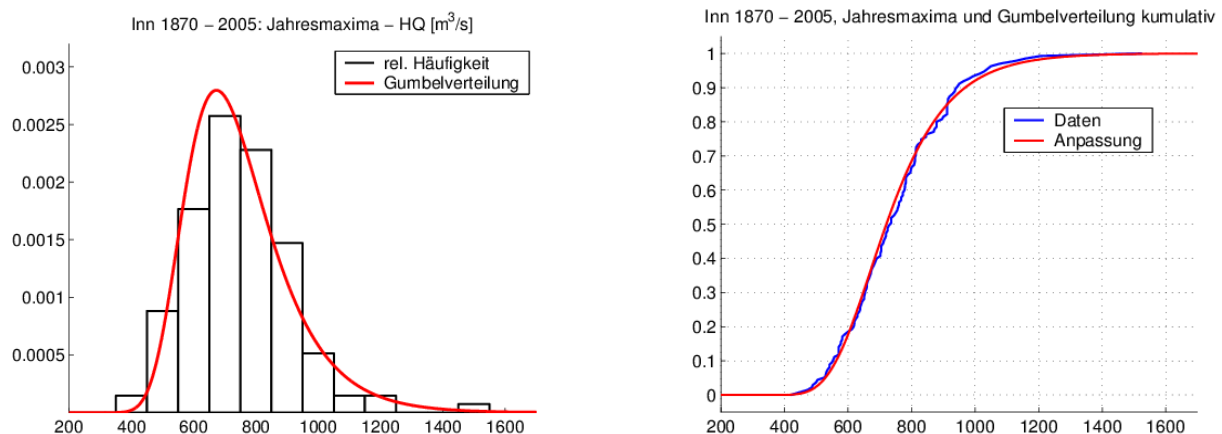


Abbildung 7.7: Jahresmaxima aus Datenreihe 1870 – 2005: Histogramm und Dichte (links), empirische Summenkurve und Verteilungsfunktion (rechts).

Erklärung von Schiefe und Kurtosis: Das k -te Moment einer Zufallsgröße ist definiert durch $m_k = E(X^k)$. Es gilt insbesondere $m_0 = 1$, $m_1 = \mu$. Das k -te zentrierte Moment

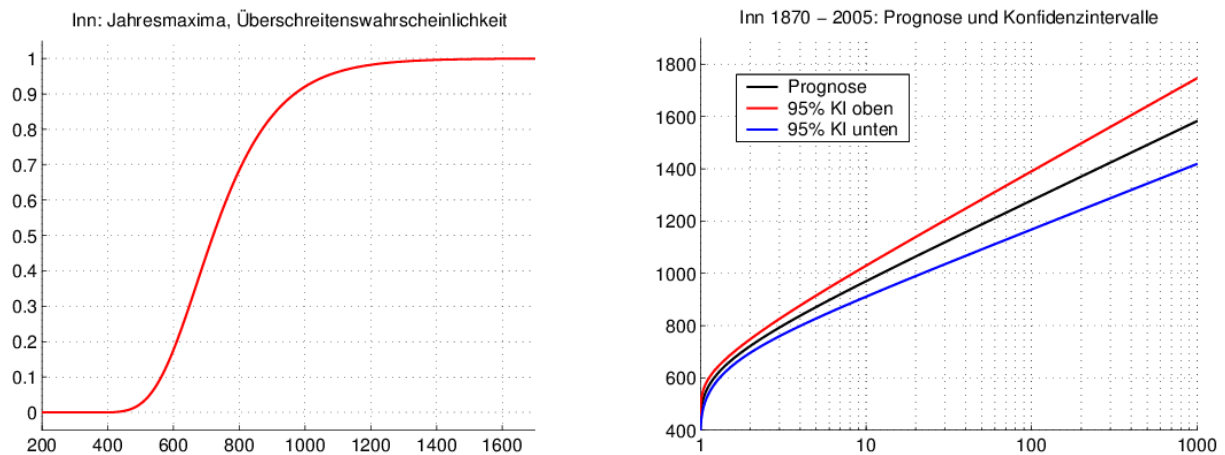


Abbildung 7.8: Jahresmaxima aus Datenreihe 1870 – 2005: Verteilungsfunktion der Gumbelverteilung (links) und Prognose der Jährlichkeit in halblogarithmischer Darstellung mit 95%-Konfidenzgrenzen (rechts).

einer Zufallsgröße ist definiert durch $m_{ck} = E((X - \mu)^k)$. Es gilt insbesondere $m_{c0} = 1$, $m_{c1} = 0$, $m_{c2} = \sigma^2$. Der *Schiefekoeffizient* einer Verteilung ist definiert durch $C_s = m_{c3}/\sigma^3$, die *Kurtosis* durch $K = m_{c4}/\sigma^4$.

Tabellen

Nr.	cm	kg	Nr.	cm	kg	Nr.	cm	kg
1	186	73	16	178	73	31	199	93
2	168	60	17	187	83	32	184	73
3	176	74	18	176	68	33	176	71
4	170	57	19	176	68	34	186	80
5	169	55	20	178	70	35	174	60
6	178	78	21	169	70	36	190	92
7	170	60	22	167	70	37	183	80
8	176	68	23	167	71	38	189	85
9	173	69	24	180	73	39	186	84
10	180	70	25	176	73	40	189	69
11	178	66	26	175	70	41	184	80
12	186	80	27	185	69	42	191	85
13	185	95	28	193	70	43	178	72
14	187	82	29	182	66	44	179	71
15	173	63	30	183	75			

Tabelle 7.1: Körpergröße und Gewicht von 44 Bauingenieurstudenten 1998.

25.6	25.5	24.0	26.0	24.1	24.0	28.5	25.3	23.4	26.5
23.2	25.0	22.0	24.0	24.9	30.0	27.0	24.4	24.3	29.5

Tabelle 7.2: Reibungswinkel $\varphi[^\circ]$ von $n = 20$ Rahmenscherversuchen an mittelplastischen Schluff. Nach: E. Bösing, Serienuntersuchungen zum Vergleich verschiedenener Rahmenschervergeräte, Institut für Bodenmechanik und Felsmechanik der Universität Karlsruhe 1996.

Jahr	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
HQ	550	490	590	660	1030	445	713	797	702	778

Tabelle 7.3: Jahresmaxima $[\text{m}^3/\text{s}]$ von 1971 bis 1980 des Inn am Pegel Innsbruck (Quelle: Hydrographischer Dienst Tirol).

Tabelle der t-Verteilung

FG	Wahrscheinlichkeit							
	0,75	0,875	0,90	0,95	0,975	0,99	0,995	0,999
1	1,000	2,414	3,078	6,314	12,706	31,821	63,657	318,309
2	0,817	1,604	1,886	2,920	4,303	6,965	9,925	22,327
3	0,765	1,423	1,638	2,353	3,182	4,541	5,841	10,215
4	0,741	1,344	1,533	2,132	2,776	3,747	4,604	7,173
5	0,727	1,301	1,476	2,015	2,571	3,365	4,032	5,893
6	0,718	1,273	1,440	1,943	2,447	3,143	3,707	5,208
7	0,711	1,254	1,415	1,895	2,365	2,998	3,499	4,785
8	0,706	1,240	1,397	1,860	2,306	2,896	3,355	4,501
9	0,703	1,230	1,383	1,833	2,262	2,821	3,250	4,297
10	0,700	1,221	1,372	1,812	2,228	2,764	3,169	4,144
11	0,697	1,214	1,363	1,796	2,201	2,718	3,106	4,025
12	0,695	1,209	1,356	1,782	2,179	2,681	3,055	3,930
13	0,694	1,204	1,350	1,771	2,160	2,650	3,012	3,852
14	0,692	1,200	1,345	1,761	2,145	2,624	2,977	3,787
15	0,691	1,197	1,341	1,753	2,131	2,602	2,947	3,733
16	0,690	1,194	1,337	1,746	2,120	2,583	2,921	3,686
17	0,689	1,191	1,333	1,740	2,110	2,567	2,898	3,646
18	0,688	1,189	1,330	1,734	2,101	2,552	2,878	3,611
19	0,688	1,187	1,328	1,729	2,093	2,539	2,861	3,579
20	0,687	1,185	1,325	1,725	2,086	2,528	2,845	3,552
21	0,686	1,183	1,323	1,721	2,080	2,518	2,831	3,527
22	0,686	1,182	1,321	1,717	2,074	2,508	2,819	3,505
23	0,685	1,180	1,319	1,714	2,069	2,500	2,807	3,485
24	0,685	1,179	1,318	1,711	2,064	2,492	2,797	3,467
25	0,684	1,178	1,316	1,708	2,060	2,485	2,787	3,450
26	0,684	1,177	1,315	1,706	2,056	2,479	2,779	3,435
27	0,684	1,176	1,314	1,703	2,052	2,473	2,771	3,421
28	0,683	1,175	1,313	1,701	2,048	2,467	2,763	3,408
29	0,683	1,174	1,311	1,699	2,045	2,462	2,756	3,396
30	0,683	1,173	1,310	1,697	2,042	2,457	2,750	3,385
40	0,681	1,167	1,303	1,684	2,021	2,423	2,704	3,307
50	0,679	1,164	1,299	1,676	2,009	2,403	2,678	3,261
60	0,679	1,162	1,296	1,671	2,000	2,390	2,660	3,232
70	0,678	1,160	1,294	1,667	1,994	2,381	2,648	3,211
80	0,678	1,159	1,292	1,664	1,990	2,374	2,639	3,195
90	0,677	1,158	1,291	1,662	1,987	2,368	2,632	3,183
100	0,677	1,157	1,290	1,660	1,984	2,364	2,626	3,174
200	0,676	1,154	1,286	1,653	1,972	2,345	2,601	3,131
500	0,675	1,152	1,283	1,648	1,965	2,334	2,586	3,107
∞	0,674	1,150	1,282	1,645	1,960	2,326	2,576	3,090

Tabelle 7.4: Quantile der t -Verteilung nach Freiheitsgraden und Wahrscheinlichkeiten.

Kapitel 8

Statistische Testverfahren

Bei statistischen Testverfahren soll eine Hypothese über einen Parameter einer Zufallsgröße auf ihre Berechtigung überprüft werden. Beispiele: Testen einer Zeitreihe auf Trend; Test, ob der Sollwert (etwa einer Produktionsserie) im Mittel erreicht wird; Test, ob einer der Koeffizienten einer Regressionsfunktion $= 0$ ist oder $\neq 0$ ist; Test, ob eine Verteilungsfunktion auf eine Stichprobe passt, genauer, ob der angenommene Verteilungstyp mit der Stichprobe verträglich ist. Es handelt sich hier weniger um eine mathematische Frage, als um eine Frage der *Entscheidungstheorie*. Entsprechend gibt ein Test keine Sicherheit, sondern nur einen plausiblen Indikator.

8.1 Test für den Mittelwert μ einer Zufallsgröße

Es soll überprüft werden, ob der Erwartungswert μ einer $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsgröße X (mit bekanntem σ) einen gegebenen Wert μ_0 hat oder nicht. Man formuliert statistische Tests in der Regel in Form von *Hypothesen*:

Nullhypothese $\mathcal{H}_0 : \mu = \mu_0$

Alternativhypothese $\mathcal{H}_1 : \mu \neq \mu_0$

Die Idee ist nun, dass – wenn die Nullhypothese zutrifft – die Zufallsgröße X eben nach $\mathcal{N}(\mu_0, \sigma^2)$ verteilt ist. Man gibt nun ein *Signifikanzniveau* vor, etwa 95%, und macht eine Stichprobe vom Umfang n . Das 95%-Konfidenzintervall ist laut Abschnitt 7.2 gleich

$$\text{KI} = \left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right].$$

In 95% aller Fälle liegt der tatsächliche Wert μ in diesem Intervall. Wir entscheiden über die Hypothese \mathcal{H}_0 :

- (a) Ist μ_0 außerhalb von KI, so wird die Nullhypothese verworfen. Wir können dann „mit einer Irrtumswahrscheinlichkeit“ von 5% sagen, dass μ *nicht gleich* μ_0 ist.

- (b) Ist μ_0 in KI enthalten, so wird die Nullhypothese nicht verworfen, sie ist mit der Stichprobe „verträglich“.

Im Falle (a) sagt man, dass \bar{x} *signifikant* (substantiell, nicht nur zufällig) von μ_0 abweicht.

Beispiel 8.1 Wir setzen das begleitende Beispiel aus dem vorigen Abschnitt fort und wollen es nutzen, um zu testen, ob die mittlere Größe männlicher Studierender von jener der männlichen Österreicher insgesamt abweicht. Der Mikrozensus des Statistischen Zentralamts im Jahr 1991 ergab, dass die durchschnittliche Größe aller männlichen österreichischen Erwachsenen (Alter ≥ 20 Jahre) gleich $\mu_0 = 175.5$ cm ist. Dagegen bezeichne μ die mittlere Größe männlicher Studierender. Wir setzen den Test, ob μ signifikant von μ_0 verschieden ist, wie folgt an:

$$\mathcal{H}_0 : \mu = 175.5$$

$$\mathcal{H}_1 : \mu \neq 175.5$$

Die Stichprobe vom Umfang $n = 44$ aus dem vorigen Abschnitt ergab als Punktschätzung für die mittlere Größe männlicher Studierender den Wert $\bar{x} = 179.9$. Das 95%-Konfidenzintervall war

$$\text{KI} = \left[179.9 - 1.96 \frac{7.6}{\sqrt{44}}, 179.9 + 1.96 \frac{7.6}{\sqrt{44}} \right] = [177.7, 182.1]$$

Der Wert $\mu_0 = 175.5$ liegt außerhalb von KI. Wir verwerfen daher die Nullhypothese \mathcal{H}_0 . Anders gesagt: Der gemessene Wert $\bar{x} = 179.9$ weist einen *signifikanten Unterschied* zum allgemeinen Durchschnittswert 175.5 auf.

Man kann natürlich an der Stelle von 95% auch andere Signifikanzniveaus wählen. Dem Signifikanzniveau von $(1 - p) \cdot 100\%$ entspricht die Irrtumswahrscheinlichkeit $p \cdot 100\%$.

99%: Die Irrtumswahrscheinlichkeit ist 1%, $p = 0.01$. Das Konfidenzintervall ist breiter und damit wird \mathcal{H}_0 eher angenommen.

90%: Die Irrtumswahrscheinlichkeit 10%, $p = 0.10$. Das Konfidenzintervall ist schmaler und damit wird \mathcal{H}_0 eher abgelehnt.

Über die Wahl von p kann mit Hilfe der Kosten von Ablehnung/Annahme von \mathcal{H}_0 entschieden werden. Zum Beispiel kann bei der Qualitätsprüfung einer Warenlieferung in Betracht gezogen werden, ob die Rücksendung fehlerhafter Waren oder die Akzeptanz eines Ausschusses teurer ist.

Die Bedeutung des Signifikanzniveaus kann mittels folgender Tabelle dargestellt werden.

\mathcal{H}_0	angenommen	abgelehnt
wahr	✓	Fehler 1. Art, Wahrscheinlichkeit p (gewählt)
falsch	Fehler 2. Art, Wahrscheinlichkeit q (unbekannt)	✓

Bei Zutreffen von \mathcal{H}_0 liegen nur $p \cdot 100\%$ der Fälle außerhalb von KI. Daher ist die Wahrscheinlichkeit des Fehlers 1. Art (\mathcal{H}_0 wird fälschlicherweise verworfen) gleich p . Der Fehler 2. Art kann nur schwer abgeschätzt werden, da dies davon abhängt, wie weit die falsche Nullhypothese von der Wirklichkeit abweicht. Im Rahmen der *Testtheorie* kann dessen Wahrscheinlichkeit q minimiert werden (*Trennschärfe*, *Gütefunktion* von Tests).

Das allgemeine Format des Tests kann wie folgt beschrieben werden. Zu gegebener Irrtumswahrscheinlichkeit von $p \cdot 100\%$ wähle man zuerst das Quantil α der Standardnormalverteilung, so dass für eine $\mathcal{N}(0, 1)$ -verteilte Größe U gilt:

$$P(-\alpha \leq U \leq \alpha) = 1 - p \quad \text{bzw.} \quad P(U < -\alpha \text{ oder } U > \alpha) = p.$$

Der Zusammenhang zwischen p und α ergibt sich aus der Symmetrie der Normalverteilung zu

$$\Phi(\alpha) = 1 - \frac{p}{2}, \quad p = 2(1 - \Phi(\alpha)).$$

(Für die Irrtumswahrscheinlichkeit von 5%, $p = 0.05$, ist $\alpha = 1.96$.)

Das $(1 - p) \cdot 100\%$ -Konfidenzintervall für \bar{X} ist dann $KI = [\bar{X} - \alpha \frac{\sigma}{\sqrt{n}}, \bar{X} + \alpha \frac{\sigma}{\sqrt{n}}]$. Das *Entscheidungskriterium* ist

- μ_0 liegt außerhalb von $[\bar{x} - \alpha \frac{\sigma}{\sqrt{n}}, \bar{x} + \alpha \frac{\sigma}{\sqrt{n}}]$... \mathcal{H}_0 wird verworfen;
- μ_0 liegt in $[\bar{x} - \alpha \frac{\sigma}{\sqrt{n}}, \bar{x} + \alpha \frac{\sigma}{\sqrt{n}}]$... \mathcal{H}_0 wird akzeptiert.

Dabei ist \bar{x} der aus der vorliegenden Stichprobe berechnete Mittelwert; bei hinreichend großem n kann σ durch die Stichprobenstreuung s geschätzt werden.

Varianten des Tests: (a) Umkehrung der Betrachtungsweise: man schlage das Entscheidungsintervall um μ_0 ab, ermittle also das Intervall $[\mu_0 - \alpha \frac{\sigma}{\sqrt{n}}, \mu_0 + \alpha \frac{\sigma}{\sqrt{n}}]$. Falls die Nullhypothese \mathcal{H}_0 zutrifft, so ist \bar{X} nach $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$ verteilt und daher

$$P(\mu_0 - \alpha \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu_0 + \alpha \frac{\sigma}{\sqrt{n}}) = 1 - p \quad \text{bzw.} \quad P(\bar{X} < \mu_0 - \alpha \frac{\sigma}{\sqrt{n}} \text{ oder } \bar{X} > \mu_0 + \alpha \frac{\sigma}{\sqrt{n}}) = p.$$

Das *Entscheidungskriterium* ist

- \bar{x} liegt außerhalb von $[\mu_0 - \alpha \frac{\sigma}{\sqrt{n}}, \mu_0 + \alpha \frac{\sigma}{\sqrt{n}}]$... \mathcal{H}_0 wird verworfen;
- \bar{x} liegt in $[\mu_0 - \alpha \frac{\sigma}{\sqrt{n}}, \mu_0 + \alpha \frac{\sigma}{\sqrt{n}}]$... \mathcal{H}_0 wird akzeptiert.

Dabei ist \bar{x} der aus der vorliegenden Stichprobe berechnete Mittelwert. Diese Form des Hypothesentests ist numerisch äquivalent zur eingangs beschriebenen Vorgangsweise.

Im obigen Beispiel ist $\mu_0 = 175.5$, $\alpha \frac{\sigma}{\sqrt{n}} = 1.96 \frac{7.6}{\sqrt{44}} = 2.2$ und das von μ_0 aus abgeschlagene Entscheidungsintervall gleich $[175.5 - 2.2, 175.5 + 2.2] = [173.3, 177.7]$. Das Stichprobenmittel $\bar{x} = 179.9$ liegt außerhalb, also wird die Nullhypothese \mathcal{H}_0 verworfen. Abbildung 8.1 zeigt das Konzept am Beispiel der Körpergrößen männlicher Österreicher.

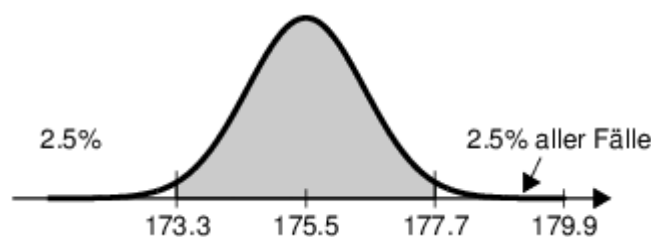


Abbildung 8.1: Variante des Hypothesentests mit μ_0 als Zentrum des Entscheidungsintervalls.

(b) Zentrieren der Testgröße: Falls die Nullhypothese \mathcal{H}_0 zutrifft, dann ist die zentrierte Größe

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

standardnormalverteilt. Es gilt daher wie oben $P(-\alpha \leq U \leq \alpha) = 1 - p$ beziehungsweise $P(U < -\alpha \text{ oder } U > \alpha) = p$. Es sei

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

die Realisierung der Testgröße U in der gegebenen Stichprobe. Das *Entscheidungskriterium* ist

- u liegt außerhalb von $[-\alpha, \alpha]$... \mathcal{H}_0 wird verworfen;
- u liegt in $[-\alpha, \alpha]$... \mathcal{H}_0 wird akzeptiert.

Dies ist allgemein gebräuchliche Form von Hypothesentests.

Im obigen Beispiel ist $u = \frac{197.9 - 175.5}{7.6/\sqrt{44}} = 3.84$ und liegt weit außerhalb des Intervalls $[-\alpha, \alpha] = [-1.96, 1.96]$. Die Nullhypothese wird verworfen.

(c) Der *P-Wert* oder die *Signifikanz* von \bar{X} . In den modernen Statistikprogrammen wird meistens der *P-Wert* ausgegeben. Dieser ist definiert als jene Irrtumswahrscheinlichkeit p^* , bei der \mathcal{H}_0 mit vorliegendem \bar{x} gerade noch akzeptiert würde.

Um den *P-Wert* herzuleiten, gehen wir von der Form (b) des Tests aus. \mathcal{H}_0 wird akzeptiert, wenn $u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ im Intervall $[-\alpha, \alpha]$ liegt. Wir fragen nach dem größtmöglichen α^* , sodass die gegebene Realisierung u der Testgröße U noch in $[-\alpha^*, \alpha^*]$ liegt. Offenbar ist das der Fall, wenn u auf einer der beiden Intervallgrenzen liegt, also $u = \pm\alpha^*$ ist. Das bedeutet aber $\alpha^* = |u|$. Den *P-Wert* p^* ermittelt man aus den Beziehungen

$$\Phi(\alpha^*) = 1 - \frac{p^*}{2}, \quad p^* = 2(1 - \Phi(\alpha^*)).$$

Je kleiner der *P-Wert*, umso eher ist die Nullhypothese abzulehnen bzw. umso höher ist die Signifikanz des Unterschieds von μ und μ_0 .

Im obigen Beispiel haben wir $u = 3.84$ berechnet. Somit ist $\alpha^* = 3.84$ und

$$p^* = 2(1 - \Phi(3.84)) = 0.000123 \dots \approx 0.01\%.$$

Wir schließen, dass die Nullhypothese \mathcal{H}_0 schon bei einem Signifikanzniveau von 99.99% verworfen würde, also mit Irrtumswahrscheinlichkeit 0.01%. Abbildung 8.2 erläutert die Situation des Beispiels.



Abbildung 8.2: Quantile der zentrierten Größe U . Der P -Wert im Beispiel entspricht der Fläche unter der Normalverteilungskurve über $(-\infty, -3.84)$ und $(3.84, \infty)$.

Teststatistiken: Für zahlreiche Situationen wurden geeignete Testgrößen entwickelt (mit entsprechenden Verteilungen). Hier ist auf die Literatur zu verweisen. Einige Beispiele werden wir in den folgenden beiden Abschnitten sehen.

8.2 Der χ^2 -Anpassungstest

Die Fragestellung ist, ob die Annahme eines gewissen Verteilungstyps gerechtfertigt ist. Dies soll an Hand einer vorliegenden Stichprobe entschieden werden. Der Verteilungstyp wird durch eine Verteilungsfunktion $F_0(x)$ beschrieben. Der Aufbau folgt dem Schema des Hypothesentests:

$$\mathcal{H}_0 : F(x) = F_0(x)$$

$$\mathcal{H}_1 : F(x) \neq F_0(x)$$

Zur Einführung beginnen wir mit dem diskreten Fall.

Diskreter Fall: Die Zufallsgröße X habe Ausprägungen x_1, \dots, x_k mit Wahrscheinlichkeiten $P(X = x_i) = p_i$. Der Test dient zur Feststellung, ob dies zutrifft. Eine Stichprobe vom Umfang n habe die Häufigkeiten H_i , $i = 1, \dots, k$ bzw. die relativen Häufigkeiten $h_i = H_i/n$ ergeben.

Nach dem zentralen Grenzwertsatz ist die Größe

$$\frac{h_i - p_i}{\sqrt{\frac{p_i(1-p_i)}{n}}} \approx \frac{h_i - p_i}{\sqrt{\frac{p_i}{n}}} = \frac{H_i - np_i}{\sqrt{np_i}}$$

näherungsweise nach $\mathcal{N}(0, 1)$ verteilt. Die Quadratsumme

$$Q = \sum_{i=1}^k \frac{(H_i - np_i)^2}{np_i}$$

ist daher näherungsweise (für $n \rightarrow \infty$) χ^2 -verteilt mit $(k - 1)$ Freiheitsgraden, falls die angenommenen Wahrscheinlichkeiten, also \mathcal{H}_0 , zutreffen. Die χ^2 -Verteilung werden wir in Abschnitt 8.3 unten besprechen. Wir testen, ob Q innerhalb etwa des 95%-Bereichs der χ^2 -Verteilung liegt oder nicht.

Beispiel 8.2 Würfel, $k = 6$, $p_i = \frac{1}{6}$. Eine Würfelserie $n = 100$ habe ergeben

$$H_1 = 18, H_2 = 12, H_3 = 16, H_4 = 6, H_5 = 18, H_6 = 30.$$

Um zu überprüfen, ob der Würfel ungezinkt ist oder nicht, berechnen wir

$$\begin{aligned} Q &= \frac{(18-100/6)^2}{100/6} + \frac{(12-100/6)^2}{100/6} + \frac{(16-100/6)^2}{100/6} \\ &+ \frac{(6-100/6)^2}{100/6} + \frac{(18-100/6)^2}{100/6} + \frac{(30-100/6)^2}{100/6} \cong 19. \end{aligned}$$

Das 95%-Quantil von $\chi^2_{(5)}$ ist 11.07, laut Tabelle 8.8 am Ende dieses Kapitels. Der Wert 19 liegt außerhalb. Daher wird \mathcal{H}_0 verworfen; wir entscheiden uns für die Ansicht, dass der Würfel gezinkt ist.

Kontinuierlicher Fall: Wir legen eine Verteilungsdichte $f(x, \lambda_1, \dots, \lambda_r)$ mit r Parametern zu Grunde. Zum Beispiel ist im Falle einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ die Anzahl der Parameter $r = 2$ und etwa $\lambda_1 = \mu, \lambda_2 = \sigma$.

Die Ausgänge der Stichprobe werden in k Klassen K_i unterteilt, $i = 1, \dots, k$. Theoretisch ist dann die Wahrscheinlichkeit der Klasse K_i gleich $p_i = \int_{a_i}^{b_i} f(x) dx$, $K_i = [a_i, b_i]$, siehe Abbildung 8.3. Die Parameter $\lambda_1, \dots, \lambda_r$ werden aus der Stichprobe geschätzt. Ferner sei

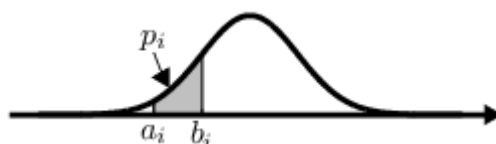


Abbildung 8.3: Theoretische Wahrscheinlichkeit der Klasse K_i .

H_i die Häufigkeit der Klasse K_i in der Stichprobe. Die Größe

$$Q = \sum_{i=1}^k \frac{(H_i - np_i)^2}{np_i}$$

ist näherungsweise χ^2 -verteilt mit $(k - r - 1)$ Freiheitsgraden (Reduktion der Freiheitsgrade, da die r Parameter von der Stichprobe abhängen), falls der Verteilungstyp $f(x)$ zutrifft.

Für die Wahl der Klasseneinteilung zieht man im Allgemeinen die *Faustregel*: $np_i \geq 4$ heran. Die Klassenbreiten sollten also groß genug gewählt werden.

Beispiel 8.3 Wir setzen das begleitende Beispiel der Körpergröße X männlicher Studierender fort. Mit Hilfe der Stichprobe vom Umfang $n = 44$ soll getestet werden, ob eine Normalverteilung vorliegt. Die geschätzten Parameter sind dann $\hat{\mu} = 179.9$, $\hat{\sigma} = 7.6$. Die zentrierte Größe $U = \frac{X-179.9}{7.6}$ ist dann nach $\mathcal{N}(0, 1)$ -verteilt, und wir können die Intervallwahrscheinlichkeiten aus der Normalverteilungstabelle entnehmen.

i	K_i	H_i	U_i	p_i	np_i	$\frac{(H_i - np_i)^2}{np_i}$
1	< 170	5	< -1.3026	0.0964	4.2394	0.1365
2	$[170, 175)$	5	$[-1.3026, -0.6447)$	0.1632	7.1807	0.6623
3	$[175, 180)$	13	$[-0.6447, 0.0132)$	0.2457	10.8108	0.4433
4	$[180, 185)$	7	$[0.0132, 0.6711)$	0.2437	10.7209	1.2914
5	$[185, 190)$	10	$[0.6711, 1.3289)$	0.1592	7.0031	1.2825
6	≥ 190	4	≥ 1.3289	0.0919	4.0450	0.0005

Tabelle 8.1: Rechentabelle zur Ermittlung des Wertes Q im Beispiel.

Damit ergibt sich

$$Q = \sum_{i=1}^6 \frac{(H_i - np_i)^2}{np_i} \approx 3.82.$$

Es liegt näherungsweise eine χ^2 -Verteilung mit $6 - 2 - 1 = 3$ Freiheitsgraden vor. Laut Tabelle 8.8 ist deren 95%-Quantil gleich 7.81. Da $3.82 < 7.81$ ist, liegt das Ergebnis im 95%-Bereich. Wir nehmen daher \mathcal{H}_0 an. Das heißt, die Hypothese einer Normalverteilung für die Körpergröße ist mit der Stichprobe verträglich.

Wir können unsere Einschätzung auch visuell überprüfen, indem wir das Histogramm und die empirische Verteilungsfunktion mit der angepassten Normalverteilung überlagern, siehe Abbildung 8.4:

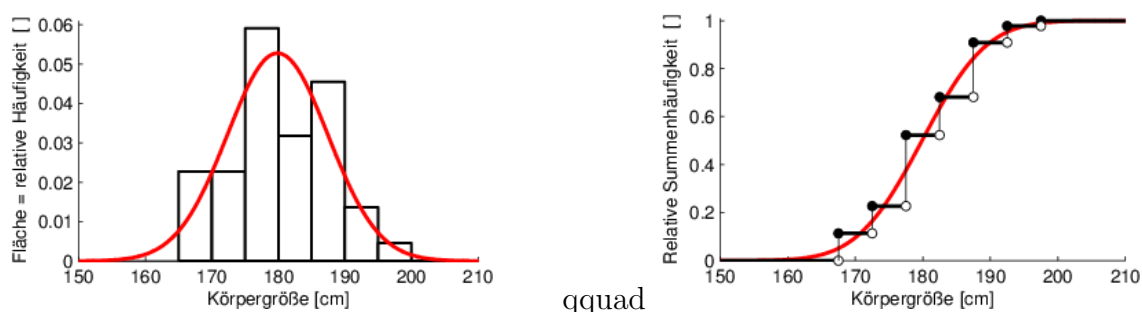


Abbildung 8.4: Histogramm/relative Summenhäufigkeit mit angepasster Dichte/Verteilungsfunktion.

Eine weitere rasche empirische Überprüfungsmöglichkeit liefert das *QQ-Plot*, bei dem die Stichprobenquantile gegen die Quantile der Standardnormalverteilung aufgetragen werden, siehe Abbildung 8.5. Falls die der vorliegenden Stichprobe zu Grunde liegende Zufallsgröße normalverteilt ist, so müssten die Quantile näherungsweise auf einer Geraden liegen.

Die Abbildungen 8.4 und 8.5 deuten an, dass im Falle der Körpergröße männlicher Studenten eine Normalverteilung mehr schlecht als recht passt; der χ^2 -Test schließt es jedenfalls nicht aus.

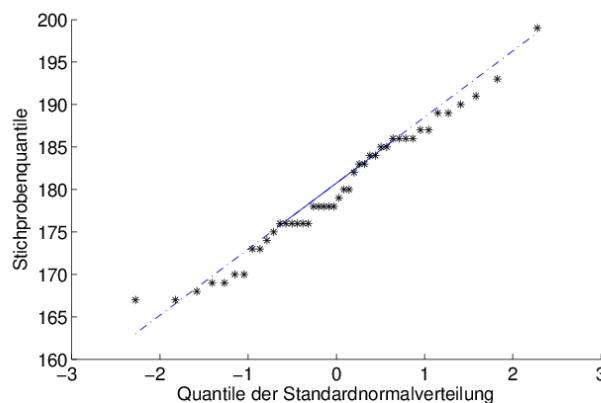


Abbildung 8.5: QQ-Plot zur Überprüfung der Normalverteilungshypothese für die Körpergröße.

8.3 Die drei wichtigsten Testverteilungen

Die χ^2 -Verteilung: Seien X_1, X_2, \dots, X_n unabhängige, $\mathcal{N}(0, 1)$ -verteilte Zufallsgrößen. Dann ist

$$\chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

χ^2 -verteilt mit n Freiheitsgraden. Die Dichte ist

$$f_{\chi_n^2}(x) = \begin{cases} C_n x^{n/2-1} e^{-x/2}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

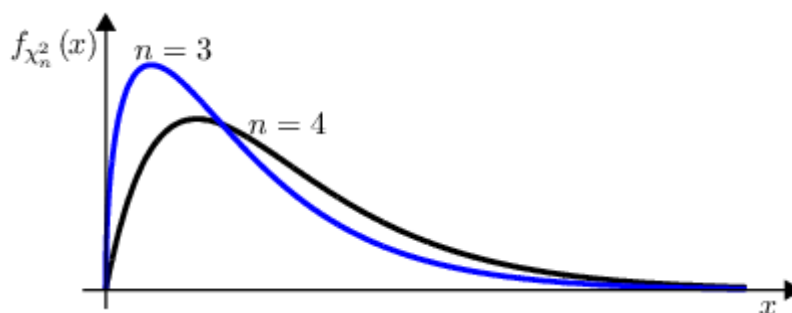
Die Normierungskonstante ist

$$C_n = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)}$$

mit der Gammafunktion $\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt$. Es gilt: $E(\chi_n^2) = n$, $V(\chi_n^2) = 2n$; der Verlauf der Dichtefunktion für verschiedene Werte von n ist Abbildung 8.6 zu entnehmen.

Anwendung auf die Stichprobenvarianz: Wir betrachten eine Stichprobe vom Umfang n . Die Werte X_1, \dots, X_n sind unabhängig; wir nehmen an, dass sie nach $\mathcal{N}(\mu, \sigma^2)$ verteilt sind. Das Stichprobenmittel und die Stichprobenvarianz sind

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Abbildung 8.6: Dichtefunktionen der χ^2 -Verteilung mit Freiheitsgraden $n = 3$ und $n = 4$.

Dann ist die Größe

$$Y = \frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

χ^2 -verteilt mit $(n-1)$ -Freiheitsgraden.

Begründung: $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2$ ist χ^2 -verteilt mit n Freiheitsgraden. Wird $X_i - \mu$ durch $X_i - \bar{X}$ ersetzt, so sind die Größen $\frac{X_i - \bar{X}}{\sigma}$ jedoch *nicht* unabhängig. Intuitiv kann man sich vorstellen, dass ein Freiheitsgrad von \bar{X} absorbiert wurde – es verbleiben nur mehr $(n-1)$ Freiheitsgrade. Einen rigorosen Beweis erhält man durch eine geschickte Transformation, nach der man $\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2$ in der Form $\sum_{i=1}^{n-1} Z_i^2$ schreiben kann mit unabhängigen und nach $\mathcal{N}(0, 1)$ verteilten Größen Z_i , $i = 1 \dots n-1$.

Die t-Verteilung (Student-Verteilung von W. Gosset): Die Größe X sei nach $\mathcal{N}(0, 1)$ verteilt, die Größe Y nach χ_n^2 -verteilt; X und Y seien unabhängig. Dann besitzt

$$t_n = \frac{X}{\sqrt{Y/n}}$$

eine t -Verteilung mit n Freiheitsgraden. Die Dichte ist

$$f_{t_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

Es gilt: $E(t_n) = 0$, $V(t_n) = \frac{n}{n-2}$. Der Verlauf der Dichtefunktion ist ähnlich der Standardnormalverteilung, jedoch etwas flacher und breiter (Abbildung 8.7).

Anwendung auf das zentrierte Stichprobenmittel: Wir haben im Abschnitt 7.2 erwähnt, dass das mittels Stichprobenvarianz zentrierte Stichprobenmittel

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

einer t -Verteilung genügt. Dies können wir nun herleiten. Sei X_1, \dots, X_n eine Stichprobe von $\mathcal{N}(\mu, \sigma^2)$ -verteilten Größen. Dann ist $X = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ nach $\mathcal{N}(0, 1)$ verteilt (Abschnitt 6.2)

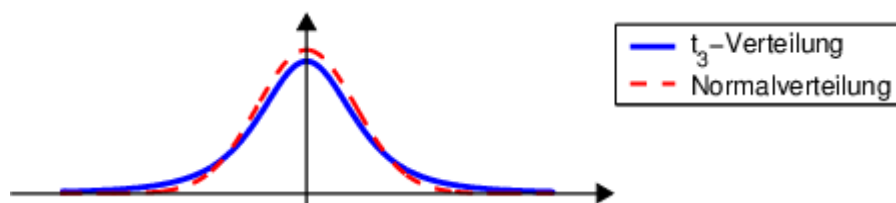


Abbildung 8.7: Dichte der t -Verteilung mit 3 Freiheitsgraden; Vergleich mit Standardnormalverteilung.

und $Y = \frac{(n-1)s^2}{\sigma^2}$ nach χ_{n-1}^2 -verteilt (siehe oben). Daher ist

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{X}{\sqrt{Y/(n-1)}}$$

t -verteilt mit $(n-1)$ -Freiheitsgraden. Wie in Abschnitt 7.2 gezeigt, kann dies zur Ermittlung von Konfidenzintervallen für den Erwartungswert einer Zufallsgröße bei kleinem Stichprobenumfang verwendet werden.

Die F-Verteilung (nach R.A. Fisher): Seien V_m, V_n unabhängige Zufallsgrößen, verteilt nach χ_m^2 bzw. nach χ_n^2 . Dann ist

$$F_{m,n} = \frac{V_m/m}{V_n/n}$$

F -verteilt mit (m, n) -Freiheitsgraden. Die F -Verteilung findet Anwendung zum Beispiel in der linearen Regression und im Test auf Trend einer Messreihe.

Tabelle der χ^2 -Verteilung

Freiheitsgrade	Wahrscheinlichkeit p										
	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995
1	0,00	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	4,57	5,58	10,34	17,28	19,68	21,92	24,73	26,76
12	3,07	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	14,34	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	15,34	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	18,34	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,34	14,04	21,34	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	24,34	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,65
28	12,46	13,56	15,31	16,93	18,94	27,34	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67
31	14,46	15,66	17,54	19,28	21,43	30,34	41,42	44,99	48,23	52,19	55,00
32	15,13	16,36	18,29	20,07	22,27	31,34	42,59	46,19	49,48	53,49	56,33
33	15,82	17,07	19,05	20,87	23,11	32,34	43,75	47,40	50,73	54,78	57,65
34	16,50	17,79	19,81	21,66	23,95	33,34	44,90	48,60	51,97	56,06	58,96
35	17,19	18,51	20,57	22,47	24,80	34,34	46,06	49,80	53,20	57,34	60,28
36	17,89	19,23	21,34	23,27	25,64	35,34	47,21	51,00	54,44	58,62	61,58
37	18,59	19,96	22,11	24,08	26,49	36,34	48,36	52,19	55,67	59,89	62,88
38	19,29	20,69	22,88	24,88	27,34	37,34	49,51	53,38	56,90	61,16	64,18
39	20,00	21,43	23,65	25,70	28,20	38,34	50,66	54,57	58,12	62,43	65,48
40	20,71	22,16	24,43	26,51	29,05	39,34	51,80	55,76	59,34	63,69	66,77

Abbildung 8.8: Quantile der Chi-Quadrat-Verteilung nach Wahrscheinlichkeiten p und Freiheitsgraden.

Kapitel 9

Stochastische Bemessung

Die Grundlagen der Bemessung im Bauwesen nach wahrscheinlichkeitstheoretischen Methoden wurden in den Fünfzigerjahren des 20. Jahrhunderts durch A. M. Freudenthal, V. V. Bolotin und andere gelegt. Die Auffassung, dass die Einflussparameter auf die Tragfähigkeit eines Bauwerks als Zufallsgrößen zu betrachten sind, führt zum *probabilistischen Sicherheitskonzept*. Ab den Achtzigerjahren diente dieser Ansatz als Grundlage und Verhandlungsrahmen für die neuen Normen. Die Eckpfeiler der neuen Normen sind:

- die Grenzzustandsfunktion für die Tragfähigkeit und für die Gebrauchstauglichkeit;
- die Teilsicherheitsbeiwerte;
- die Versagenswahrscheinlichkeit.

Diese Begriffe sollten die früheren Sicherheitsbeiwerte ersetzen, klarere Argumentationsgrundlagen schaffen und eine europäische Vereinheitlichung ermöglichen. In Österreich wurden die Grundsätze zuerst in der ÖNORM B4040 festgeschrieben. Etwas länger dauerte die europäische Normung in den Eurocodes. Die Grundlagen der Tragwerksplanung sind in der Norm EN 1990:2002 formuliert, welche 2005 die ÖNORM B4040 ersetzte. Da in den Normen das probabilistische Modell nur als Ansatz eingeht, die Teilsicherheitsbeiwerte jedoch auf Verhandlungsergebnissen in den Normungsausschüssen beruhen, spricht man auch vom *semiprobabilistischen Sicherheitskonzept*. In diesem Abschnitt sollen die Grundkonzepte erläutert werden.

Ein Bauwerk ist *Belastungen* ausgesetzt, wie zum Beispiel Eigengewicht, Verkehrslast, Wind usw. Wir fassen die Belastungen mit dem Buchstaben

S

(von englisch *stress*) zusammen. Die Norm EN 1990:2002 spricht allgemeiner von *Einwirkungen*. Demgegenüber steht die *Belastbarkeit*

R

(von englisch *resistance*). *Sicherheit* liegt vor, falls stets $R > S$ ist. *Versagen* tritt bei $R < S$ bzw. bei $R/S < 1$ auf. Bereits im deterministischen Sicherheitsansatz bezeichnete man

$$Z = R - S \quad \text{bzw.} \quad \nu = \frac{R}{S}$$

als *Sicherheitsabstand* bzw. *Sicherheitsfaktor*. Dementsprechend ist Versagen beim kritischen Zustand $Z < 0$ oder $\nu < 1$ zu erwarten.

Genauer ist ein Bauwerk ein System, das vielfältigen Belastungen S_1, \dots, S_k ausgesetzt ist und dessen Belastbarkeit durch verschiedene Widerstandsparameter R_1, \dots, R_l bedingt ist. Der Versagensfall wird durch die *Grenzzustandsfunktion* beschrieben:

$$g(R_1, \dots, R_l, S_1, \dots, S_k) < 0.$$

Im Spezialfall oben ist $g(R, S) = R - S$; die Normen unterscheiden nach Verlust der Tragfähigkeit und nach Verlust der Gebrauchstauglichkeit. Zusätzlich kommt die *Zeitabhängigkeit* ins Spiel, da sich die Belastungs- und Widerstandsparameter im Laufe der Zeit ändern:

$$\begin{aligned} R &= R(t), \quad t \geq 0 \\ S &= S(t), \quad t \geq 0. \end{aligned}$$

Die *Lebensdauer* eines Bauwerks ist jene Zeitspanne T , nach der erstmals der Fall $R(T) - S(T) < 0$ eintritt (vgl. Abbildung 9.1).

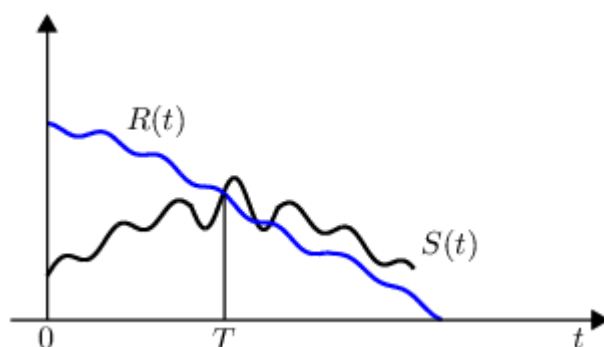


Abbildung 9.1: Zur Zuverlässigkeit und Lebensdauer eines Bauwerks.

Als *Zuverlässigkeit* bezeichnet man das Nichtversagen während der SOLL-Lebensdauer. Die Norm EN 1990:2002 definiert dazu Nutzungsdauerklassen (siehe Tabelle 9.2). Die Klassifizierung ist für die Wahl des so genannten *Zuverlässigkeitsindex* relevant, den wir weiter unten besprechen werden. In der Planung eines Projekts sind bei der Auslegung der Nutzungsdauer die *Schadenskosten* und die *Sicherheitskosten* zu berücksichtigen und ein *Qualitätsmanagement* vorzusehen.

Belastungen und Widerstände als Zufallsgrößen. Die Belastbarkeit R und die Belastungen S sind nicht konstant, sondern zufälligen Schwankungen unterworfen. Dies wird berücksichtigt, indem R und S als *Zufallsgrößen* aufgefasst werden. Der Zustand der Sicherheit $R \geq S$ kann in der Realität nicht zu 100% erreicht werden. Vielmehr ist die *Versagenswahrscheinlichkeit*

$$p_f = P(R < S)$$

Klasse der Nutzungsdauer	Planungsgröße der Nutzungsdauer (in Jahren)	Beispiele
1	10	Tragwerke mit befristeter Standzeit ^a
2	10–25	Austauschbare Tragwerksteile, z. B. Kranbahnträger, Lager
3	15–30	Landwirtschaftlich genutzte und ähnliche Tragwerke
4	50	Gebäude und andere gewöhnliche Tragwerke
5	100	Monumentale Gebäude, Brücken und andere Ingenieurbauwerke

^a ANMERKUNG Tragwerke oder Teile eines Tragwerks, die mit der Absicht der Wiederverwendung demontiert werden können, sollten nicht als Tragwerke mit befristeter Standzeit betrachtet werden.

Abbildung 9.2: Klassifizierung der Nutzungsdauer nach EN 1990:2002.

(von englisch *probability of failure*) möglichst gering zu halten. Das semiprobabilistische Sicherheitskonzept geht von der Versagenswahrscheinlichkeit aus und versucht daraus die *Sicherheitsbeiwerte* für die Bemessung herzuleiten,

Beispiel 9.1 (Grenzzustandsfunktion eines Balkens unter Gleichlast) Wir betrachten einen Balken der Länge ℓ [m], auf den eine Streckenlast konstanter Größe L [N/m] einwirkt (Abbildung 9.3). Die Bemessung soll nach dem maximalen Biegemoment erfolgen.

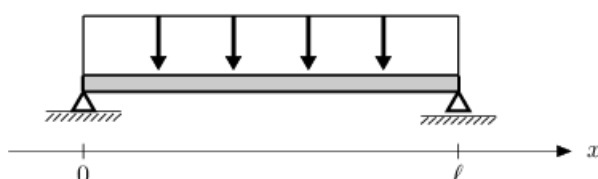


Abbildung 9.3: Balken unter Gleichlast.

Unter den gegebenen Auflagerbedingungen gilt für das Biegemoment

$$M''(x) = -L, \quad M(0) = M(\ell) = 0$$

und damit

$$M(x) = -\frac{L}{2}x^2 + L\ell\frac{x}{2}.$$

Das maximale Biegemoment tritt in der Balkenmitte auf und hat den Wert

$$M_{\max} = M\left(\frac{\ell}{2}\right) = \frac{L\ell^2}{8};$$

die *Belastung* ist somit $S = L\ell^2/8$. Um die Belastbarkeit herzuleiten, betrachten wir den Querschnitt mit der neutralen Faser (Abbildung 9.4). Mit dem Flächenträgheitsmoment $I = \int z^2 dA$ und der Normalspannung σ besagt das Elastizitätsgesetz (vgl. Abbildung 9.5)

$$\sigma(z) = \frac{M}{I} z.$$

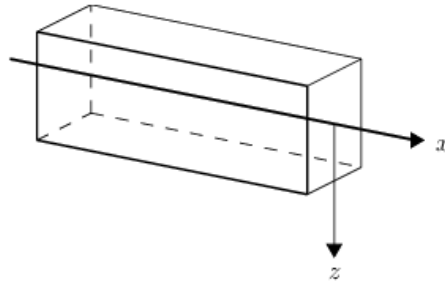


Abbildung 9.4: Querschnitt und neutrale Faser.

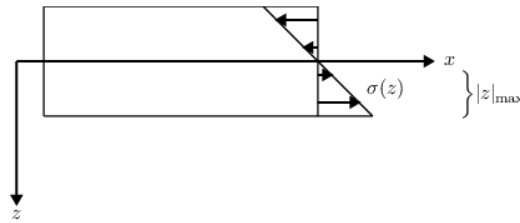


Abbildung 9.5: Zur Normalspannung.

Führen wir noch das Widerstandsmoment

$$W = \frac{I}{|z|_{\max}}$$

ein, so ergibt sich die maximale Normalspannung im Querschnitt zu

$$\sigma_{\max} = \frac{M_{\max}}{W}.$$

Ist der Grenzzustand durch eine zulässige Spannung σ_{zul} definiert, etwa die Fließspannung oder die Dehngrenze, so ist demnach

$$\sigma_{\max} \leq \sigma_{\text{zul}} \quad \text{bzw.} \quad M_{\max} \leq W \sigma_{\text{zul}}$$

zu fordern. Damit wird die *Belastbarkeit* durch $R = W \sigma_{\text{zul}}$ beschrieben. Sicherheit des Tragwerks

$$S < R$$

entspricht dann

$$M_{\max} < W \sigma_{\text{zul}} \quad \text{oder} \quad \frac{L \ell^2}{8} < W \sigma_{\text{zul}}.$$

Die Grenzzustandsfunktion des Balkens ist demnach

$$g(R, S) = W \sigma_{\text{zul}} - \frac{L \ell^2}{8}.$$

Widerstandsmomente für Rechtecks- und Kreisquerschnitt sind Abbildung 9.6 zu entnehmen. Betrachtet man die Länge ℓ und das Widerstandsmoment W als fest, so ist die

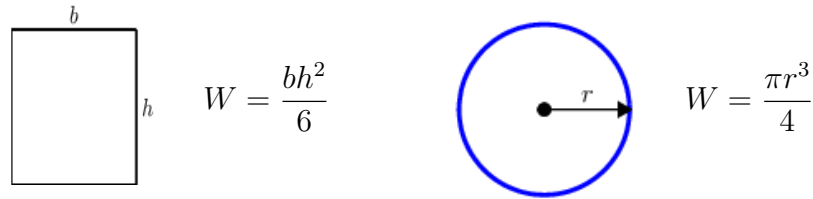


Abbildung 9.6: Widerstandsmomente für Rechtecks- und Kreisquerschnitt.

Grenzzustandsfunktion mit $R = W\sigma_{\text{zul}}$ und $S = L\ell^2/8$ von der simplen linearen Form

$$g(R, S) = R - S.$$

Eine kompliziertere nichtlineare Form kann leicht eintreten, wenn noch weitere Parameter als Zufallsgrößen behandelt werden, zum Beispiel der Radius bei einem Kreisquerschnitt (man denke an einen Baumstamm). Dann ergeben sich zwei Widerstandsgrößen $R_1 = \sigma_{\text{zul}}$ und $R_2 = r$; die Grenzzustandsfunktion erhält die nichtlineare Form

$$g(S, R_1, R_2) = \frac{\pi R_2^3}{4} R_1 - S.$$

Die Form der Grenzzustandsfunktion g ergibt sich aus den zur Bemessung herangezogenen Zufallsgrößen.

Stochastische Beschreibung in der Modellsituation. Wir wollen die Begriffe des probabilistischen Sicherheitskonzepts in der modellhaften Standardsituation erläutern, nämlich unter der Annahme, dass R und S normalverteilt und unabhängig sind und die Grenzzustandsfunktion linear ist: $Z = R - S$.

Genauer nehmen wir an, dass R nach $\mathcal{N}(\mu_R, \sigma_R^2)$ und S nach $\mathcal{N}(\mu_S, \sigma_S^2)$ verteilt ist. Die Graphen der Verteilungsdichten sind Abbildung 9.7 zu entnehmen. Die Verteilung von Z

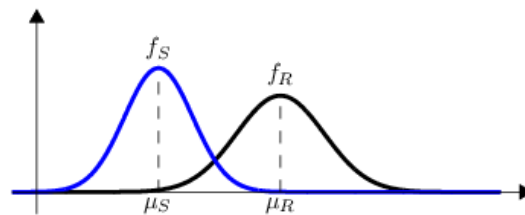


Abbildung 9.7: Verteilungsdichten für Belastung und Widerstand.

ist dann ebenfalls eine Normalverteilung, nämlich $\mathcal{N}(\mu_Z, \sigma_Z^2)$ mit

$$\begin{aligned}\mu_Z &= \mu_R - \mu_S \\ \sigma_Z^2 &= \sigma_R^2 + \sigma_S^2.\end{aligned}$$

Im Allgemeinen erhält man die Verteilungsdichte der Zufallsgröße Z durch die Faltung

$$f_Z(z) = \int_{-\infty}^{\infty} f_R(y) f_S(y - z) \, dy;$$

im Spezialfall der Normalverteilung ist diese ebenfalls eine Gauß'sche Glockenkurve (Abbildung 9.8).

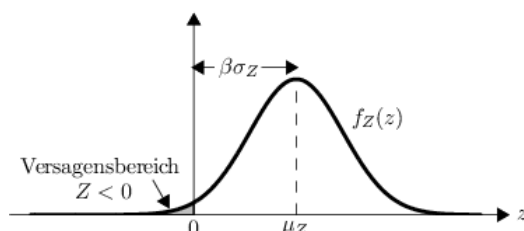


Abbildung 9.8: Dichte der Grenzzustandsfunktion Z ; Versagensbereich.

Die *Versagenswahrscheinlichkeit* ist

$$p_f = P(Z < 0) = \int_{-\infty}^0 f_Z(z) \, dz$$

und kann als Fläche unter der Dichtefunktion $f_Z(z)$ im Bereich $z < 0$ abgelesen werden.¹

Der *Sicherheitsindex*, in der Norm EN 1990:2002 auch als *Zuverlässigkeitsindex* bezeichnet, ist definiert als

$$\beta = \frac{\mu_Z}{\sigma_Z} = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}.$$

Der Sicherheitsindex steht in direktem Zusammenhang mit der Versagenswahrscheinlichkeit. Anschaulich ist klar (Abbildung 9.9), dass ein kleiner Sicherheitsabstand Z mit kleiner Streuung zur selben Versagenswahrscheinlichkeit führt wie ein großer Sicherheitsabstand Z , der weit streut.

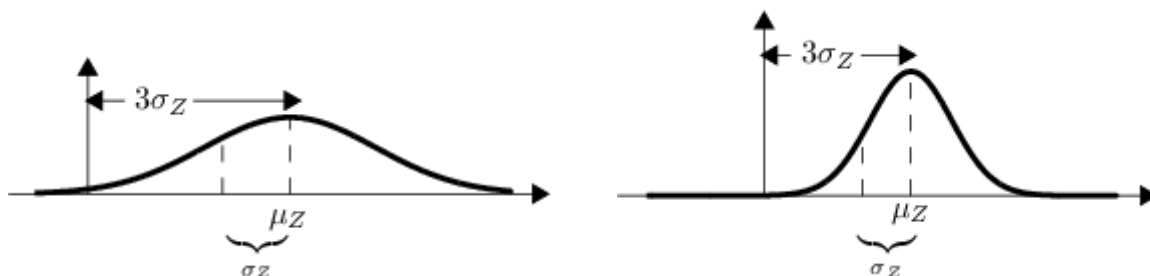


Abbildung 9.9: Zwei Fälle mit demselben Sicherheitsindex $\beta = 3$.

¹Achtung: Die Versagenswahrscheinlichkeit ist *nicht* die Fläche des Durchschnitts unter den Graphen von f_R und f_S in Abbildung 9.7!

Die Umrechnung von p_f nach β erfolgt durch

$$p_f = P(Z < 0) = P\left(\frac{Z - \mu_Z}{\sigma_Z} < -\frac{\mu_Z}{\sigma_Z}\right) = \Phi(-\beta).$$

Dabei ist Φ die kumulative Verteilungsfunktion der Standardnormalverteilung; man beachte, dass die zentrierte Größe $\frac{Z - \mu_Z}{\sigma_Z}$ nach $\mathcal{N}(0, 1)$ verteilt ist. Zum Beispiel entspricht einem Sicherheitsindex von $\beta = 3$ eine Versagenswahrscheinlichkeit von $p_f = \Phi(-3) = 0.0013$. Tabelle 9.1 listet weitere Entsprechungen auf.

β	p_f
1.282	0.1
2.326	0.01
3.090	0.001
3.719	0.0001
4.265	0.00001
4.753	0.000001

Tabelle 9.1: Sicherheitsindex und Versagenswahrscheinlichkeit.

Die Norm EN 1990:2002 empfiehlt die Verwendung verschiedener Werte von β je nach Bezugszeitraum und Zuverlässigkeitsklasse, welche wiederum durch die Schwere der Schadensfolgen charakterisiert wird, vgl. Tabelle 9.10. Die Bemessung eines Tragwerks kann

Zuverlässigkeits-Klasse	Mindestwert für β	
	Bezugszeitraum 1 Jahr	Bezugszeitraum 50 Jahre
RC 3	5,2	4,3
RC 2	4,7	3,8
RC 1	4,2	3,3

ANMERKUNG Die Bemessung nach EN 1990 mit den Teilsicherheitsbeiwerten nach Anhang A sowie nach EN 1991 bis EN 1999 führt in der Regel zu einem Tragwerk mit einer Mindestzuverlässigkeit $\beta \geq 3,8$ für einen Bezugszeitraum von 50 Jahren. Größere Zuverlässigkeitsklassen als RC 3 werden in diesem Anhang nicht weiter betrachtet, da für die betroffenen Bauteile Sonderuntersuchungen angestellt werden müssen.

Abbildung 9.10: Empfehlungen für die Mindestwerte des Zuverlässigkeitsindex β nach EN 1990:2002.

theoretisch auf dieser Stufe erfolgen: bei Kenntnis der statistischen Verteilung der Belastung S ist so zu dimensionieren, dass mit der resultierenden Belastbarkeit R eine geforderte Versagenswahrscheinlichkeit eingehalten wird, dass also $P(g(R, S) < 0) \leq p_f$ ist. Dabei ist zum Beispiel laut Norm EN 1990:2002 in der Zuverlässigkeitsklasse RC-2 der Zuverlässigkeitsindex $\beta = 4.7$ bzw. die Versagenswahrscheinlichkeit $p_f = 10^{-6}$ anzuwenden.

Die Eurocodes verlangen vom Statiker/von der Statikerin allerdings nicht, eine vollständig probabilistische Rechnung durchzuführen. Vielmehr werden Bemessungswerte, kritische

Werte und Teilsicherheitsbeiwerte eingeführt, sodass am Ende der Nachweis der Trag-sicherheit durch Einsetzen in Formeln geführt werden kann. Wir wenden uns nun der Herleitung dieser Größen zu.

Der Bemessungspunkt. Die gemeinsame Verteilung der – als unabhängig angenommenen – Basisvariablen (R, S) hat die Dichte

$$f_R(r)f_S(s) = \frac{1}{2\pi\sigma_R\sigma_S} \exp\left(-\frac{(r-\mu_R)^2}{2\sigma_R^2} - \frac{(s-\mu_S)^2}{2\sigma_S^2}\right).$$

Der *Bemessungspunkt* (R^*, S^*) ist jener Punkt auf der Grenzzustandsfunktion $g(r, s) = 0$, der die größte Wahrscheinlichkeit besitzt (Abbildung 9.11).²

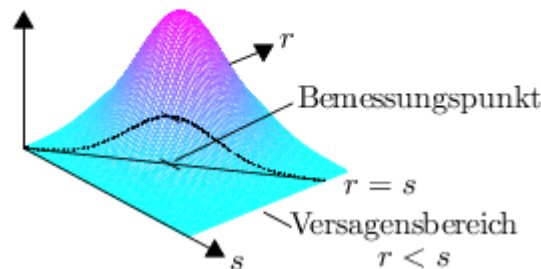


Abbildung 9.11: Definition des Bemessungspunkts.

Im Modellfall $g(r, s) = r - s$ liegt der Bemessungspunkt auf der Grenzgeraden $r = s$, erfüllt also $S^* = R^*$, und kann durch Nullsetzen der Ableitung der Dichtefunktion längs der Grenzgeraden berechnet werden:

$$0 = \frac{d}{dr} f_R(r)f_S(r) = \frac{1}{2\pi\sigma_R\sigma_S} \left(-\frac{(r-\mu_R)}{\sigma_R^2} - \frac{(r-\mu_S)}{\sigma_S^2} \right) \exp\left(-\frac{(r-\mu_R)^2}{2\sigma_R^2} - \frac{(r-\mu_S)^2}{2\sigma_S^2}\right).$$

Es muss also gelten

$$\frac{r-\mu_R}{\sigma_R^2} + \frac{r-\mu_S}{\sigma_S^2} = 0.$$

Die Lösung $r = R^*$ dieser Gleichung ist:

$$R^* = \frac{\sigma_S^2\mu_R + \sigma_R^2\mu_S}{\sigma_R^2 + \sigma_S^2} = S^*.$$

Mit Einführung der Wichtungsfaktoren

$$\alpha_R = \frac{\sigma_R}{\sqrt{\sigma_R^2 + \sigma_S^2}}, \quad \alpha_S = \frac{\sigma_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}$$

²Die Rolle des Bemessungspunkts ist im Anhang C.7 der Norm EN 1990:2002 beschrieben: *Die Bemessungswerte sollten so bestimmt werden, dass sie den Werten der Basisvariablen im Bemessungspunkt nach der Zuverlässigkeitsmethode 1. Ordnung entsprechen.*

lässt sich dies schreiben als

$$R^* = \mu_R - \alpha_R \beta \sigma_R, \quad S^* = \mu_S + \alpha_S \beta \sigma_S.$$

Die Wichtungsfaktoren α_R, α_S beschreiben also den Abstand des Bemessungspunkts zu den Mittelwerten (μ_R, μ_S) in Bezug auf $\beta \sigma_R$ bzw. $\beta \sigma_S$, vergleiche Abbildung 9.12. Durch die übliche Zentrierung der Normalverteilungen erhält man außerdem die Beziehungen

$$P(R < R^*) = \Phi(-\alpha_R \beta), \quad P(S > S^*) = \Phi(-\alpha_S \beta).$$

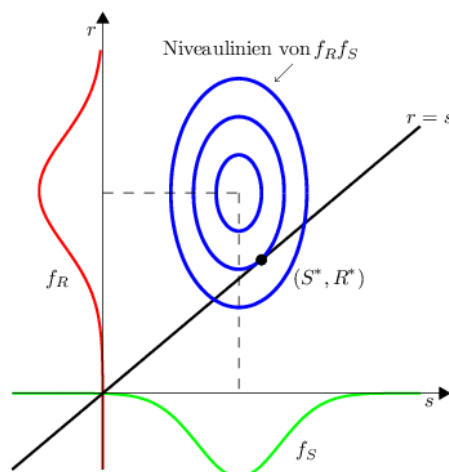


Abbildung 9.12: Lage des Bemessungspunkts.

Die charakteristischen Werte. Die so genannten *charakteristischen Werte* R_k, S_k werden als Quantilwerte eingeführt, wobei die Norm dann sagt, welcher %-Wert zu nehmen ist, je nach Sicherheitsklasse und Variationskoeffizient der Basisvariablen.

Abbildung 9.13 erinnert an die Definition des Quantils Q_p der Standardnormalverteilung. Es gilt zum Beispiel $Q_{0.95} = 1.645$, $Q_{0.97} = 1.88$.



Abbildung 9.13: Zur Definition des $p \cdot 100\%$ -Quantils Q_p der Standardnormalverteilung.

Für die Widerstände wird R_k so festgelegt, dass $p \cdot 100\%$ der Werte darüber liegen. Für die Belastungen sollen $q \cdot 100\%$ unter S_k liegen. Wie üblich erhält man diese Quantile aus der Zentrierung der Normalverteilung zu

$$R_k = \mu_R - Q_p \sigma_R, \quad S_k = \mu_S + Q_q \sigma_S,$$

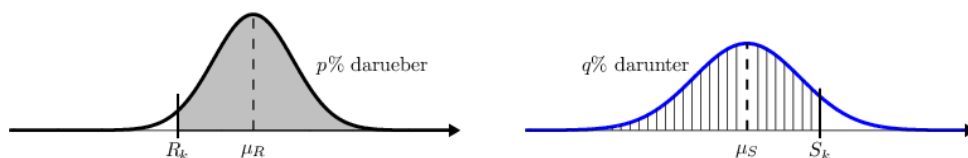


Abbildung 9.14: Zur Definition der charakteristischen Werte.

vergleiche Abbildung 9.14. Die Lage der charakteristischen Werte in Bezug auf die Verteilungsdichten von Belastung und Widerstand ist aus Abbildung 9.15 ersichtlich. Das Verhältnis R_k/S_k ist ein Maß für die Sicherheit.

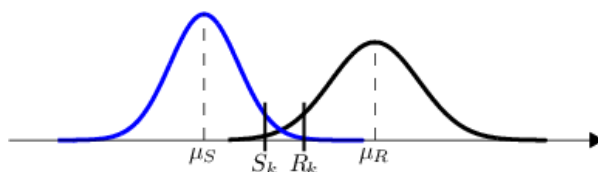


Abbildung 9.15: Lage der charakteristischen Werte und Mittelwerte.

Die Teilsicherheitsbeiwerte und Bemessungswerte. In den älteren Normen erfolgte der Sicherheitsnachweis mit Hilfe eines gemeinsamen *Nennsicherheitsbeiwerts* ν ; für die Bemessung war zu verlangen, dass $R_k \geq \nu S_k$ erfüllt war. In den neuen Normen wird der Nennsicherheitsbeiwert in einen Last- und Widerstandsanteil zerlegt: $\nu = \gamma_R \gamma_S$ mit so genannten *Teilsicherheitsbeiwerten* γ_R und γ_S . Schließlich werden noch die *Bemessungswerte*

$$R_d = \frac{R_k}{\gamma_R}, \quad S_d = \gamma_S S_k$$

eingeführt. Der Sicherheitsnachweis erhält damit das Format

$$R_d \geq S_d \quad \text{bzw.} \quad \frac{R_k}{\gamma_R} \geq \gamma_S S_k. \quad (9.1)$$

In der Modellsituation normalverteilter, unabhängiger Basisvariablen R, S ist es möglich, die Teilsicherheitsbeiwerte exakt so zu bestimmen, dass eine vorgegebene Versagenswahrscheinlichkeit $p_f^* = \Phi(-\beta^*)$ eingehalten wird. Dazu setzen wir

$$\gamma_S = \frac{1 + \beta^* \alpha_S v_S}{1 + Q_q v_S}, \quad \gamma_R = \frac{1 - Q_p v_R}{1 - \beta^* \alpha_R v_R} \quad (9.2)$$

wobei $v_R = \sigma_R/\mu_R$ und $v_S = \sigma_S/\mu_S$ die Variationskoeffizienten sind.

Satz 9.2 Werden γ_R und γ_S nach Formel (9.2) berechnet, so ist der Sicherheitsnachweis $R_d \geq S_d$ äquivalent zur Einhaltung der Versagenswahrscheinlichkeit $p_f^* = \Phi(-\beta^*)$.

Der Beweis dieser Tatsache erfolgt durch Einsetzen. Die Beziehung $R_d \geq S_d$ bedeutet ja $R_k/\gamma_R \geq \gamma_S S_k$. Nach Definition war $R_k = \mu_R - Q_p \sigma_R$ und $S_k = \mu_S + Q_q \sigma_S$. Einsetzen von (9.2) führt auf

die äquivalente Beziehung

$$\mu_R(1 - \beta^* \alpha_R v_R) \geq \mu_S(1 + \beta^* \alpha_S v_S)$$

beziehungsweise

$$\mu_R - \mu_S \geq \beta^*(\mu_R \alpha_R v_R + \mu_S \alpha_S v_S) = \beta^* \sqrt{\sigma_R^2 + \sigma_S^2},$$

das heißt

$$\beta = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \geq \beta^*,$$

was aber gleichbedeutend mit $p_f = \Phi(-\beta) \leq \Phi(-\beta^*) = p_f^*$ ist.³

Sicherheitsnachweis nach der Norm EN 1990:2002. Das Format des Sicherheitsnachweises ist durch (9.1) definiert. Im Unterschied zur Modellsituation werden allerdings die Teilsicherheitsbeiwerte γ_R, γ_S nicht berechnet, sondern stehen als Zahlenwerte in den Normen.

Zunächst wird nach *ständigen Einwirkungen* (G), wie Eigengewicht, *veränderlichen Einwirkungen* (Q), wie Nutzlasten, und *außergewöhnlichen Einwirkungen* (A), wie Explosionen, unterschieden (Abschnitt 4.1 in EN 1990:2002).

Beispiele für den Teilsicherheitsbeiwert für die Einwirkungen aus der Norm:

γ_S	= 1.35	...	ungünstige ständige Einwirkung
	= 1.0	...	günstige ständige Einwirkung
	= 1.5	...	ungünstige veränderliche Einwirkung
	= 0	...	günstige veränderliche Einwirkung

Die Werte für γ_R stehen in den einzelnen Fachnormen. Auch für die Wichtungsfaktoren α_R, α_S werden im Anhang C.7 der Norm EN 1990:2002 Werte vorgeschlagen, zum Beispiel

$$\alpha_R = 0.8, \quad \alpha_S = 0.7,$$

anzuwenden unter der Bedingung, dass $0.16 \leq \sigma_S/\sigma_R \leq 7.6$ eingehalten wird, mit entsprechend anderen Werten in den anderen Fällen.

Kurz zusammengefasst erfolgt der Sicherheitsnachweis nach der Norm in drei Schritten:

1. Ermittle μ_R, σ_R und μ_S, σ_S und daraus R_k, S_k .
2. Entnehme nach Klassifikation die Werte γ_R, γ_S aus der Norm.
3. Weise nach, dass $R_k/\gamma_R \geq \gamma_S S_k$ ist, bzw. dimensioniere entsprechend.

Tatsächlich verlangt die Norm eine wesentlich abgestufte Vorgangsweise. Zum Beispiel sind für die Kombination von Einwirkungen *Kombinationswerte* (ψ) heranzuziehen sowie weitere situationsabhängige Beiwerte. Schließlich sieht auch die Norm vor, verschiedene Verteilungstypen anzuwenden mit entsprechend angepassten Formeln für den Sicherheitsnachweis. Zum Beispiel werden folgende Verteilungstypen empfohlen (Zitat aus Anhang C.6):

³In der Modellsituation gilt sogar $R_d = R^*$ und $S_d = S^*$.

- lognormale Verteilung oder Weibull-Verteilung für Baustoffeigenschaften, Bauteilwiderstände und Modellunsicherheiten;
- Normalverteilung für Eigengewicht;
- für veränderliche Einwirkungen, ausgenommen für Ermüdungseinwirkungen, einfachheitshalber, die Normalverteilung, die Extremwertverteilung wäre angemessener.

Zur Interpretation der Versagenswahrscheinlichkeit: Der Festlegung, etwa für die Zuverlässigkeitsklasse RC-2, liegt der Zuverlässigkeitsindex $\beta = 4.7$ bzw. die Versagenswahrscheinlichkeit $p_f = 10^{-6}$ zu Grunde. Da aber die Teilsicherheitsbeiwerte auf Verhandlungsergebnissen in den Normenausschüssen beruhen, darf der Wert 10^{-6} *nicht* als relative Häufigkeit für das Versagen interpretiert werden. Man spricht von *operativer Wahrscheinlichkeit*, die eben nur eine Gewichtung in einem genormten Prozedere liefert. Auf diese Tatsache wird sogar im Anhang C.4 der Norm hingewiesen:

ANMERKUNG Die „Versagenswahrscheinlichkeit“ und der zugehörige Zuverlässigkeitsindex (siehe C.5) sind lediglich operative Werte, die nicht die wirklichen Versagensraten ausdrücken, sondern nur für die Kalibrierung der Normen und für Vergleiche der Zuverlässigkeitsniveaus verschiedener Tragwerke verwendet werden.

Vergleicht man verschiedene probabilistische Modellrechnungen, wobei die Dimensionierung nach der Norm vorgenommen wird, so stellt man Schwankungsbreiten der berechneten Versagenswahrscheinlichkeiten um mehrere Zehnerpotenzen fest, siehe etwa den Sammelband Fellin et al. (2005). Mit welcher Wahrscheinlichkeit das fertiggestellte Bauwerk versagen wird, kann in der Regel nicht überprüft werden, da Bauwerke Unikate sind (anders allerdings bei Serienanfertigungen von Bauteilen).

Als weiterführende Literatur zum Thema stochastische Bemessung verweisen wir auf Bucher(2009), Klingmüller et al. (1992), Schüller(1981).

Kapitel 10

Bayes'sche Konzepte

10.1 Bedingte Wahrscheinlichkeiten

Der Begriff der *bedingten Wahrscheinlichkeit* ist zentral in der Bayes-Statistik, aber auch bei Ereignisbäumen und Bayes'schen Netzwerken. Die bedingte Wahrscheinlichkeit eines Ereignisses A unter der Bedingung, dass Ereignis B eingetreten ist, ist

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

Dabei ist $P(B) \neq 0$ vorauszusetzen. (Wir erinnern an die Notationen $P(A \cap B)$, $P(A \cup B)$ für die Und-Verknüpfung bzw. Oder-Verknüpfung der beiden Ereignisse A und B sowie die Notation A^c für das Komplementärereignis aus Abschnitt 4.2.)

Es ergibt sich die *Multiplikationsformel*

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

Beispiel 10.1 Die Wahrscheinlichkeit eines Geräteschadens (Ereignis A) durch Hochwasserüberflutung einer Baugrube soll ermittelt werden. Die Baugrubensicherung sei auf ein fünfzigjähriges Hochwasser bemessen (Ereignis B , $P(B) = 2\%$). Geräteschaden tritt auf, wenn aufgrund fehlender Vorwarnung die Geräte nicht rechtzeitig aus der Baugrube entfernt werden können. Aus Erfahrung sei bekannt, dass der meteorologische Dienst in 60% der Hochwasserereignisse eine rechtzeitige Vorwarnung liefert. Somit ist die Wahrscheinlichkeit eines Geräteschadens, falls Hochwasser auftritt, gleich

$$P(A|B) = 0.4.$$

Wir erhalten für die Wahrscheinlichkeit des Ereignisses „Hochwasser mit Geräteschaden“ somit

$$P(A \cap B) = P(A|B) \cdot P(B) = 0.4 \cdot 0.02 = 0.008 \sim 0.8\%.$$

Der Bayes'schen Parameterschätzung liegt der einfache, aber wichtige *Satz von Bayes* zu

Grunde. Den Satz erhält man durch Kombination der beiden Beziehungen

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Satz 10.2 (Satz von Bayes) Für Ereignisse A, B mit $P(A) \neq 0$ gilt:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

Aus den Rechenregeln für die Ereignisalgebra folgt

$$P(A) = P(A \cap (B \cup B^c)) = P((A \cap B) \cup (A \cap B^c)) = P(A \cap B) + P(A \cap B^c).$$

Einsetzen der Multiplikationsformel oben ergibt:

Satz 10.3 (Satz von der totalen Wahrscheinlichkeit) Es gilt:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Schließlich kann die Multiplikationsformel auf mehrere Ereignisse ausgeweitet werden:

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ P(A \cap B \cap C) &= P(A|B \cap C)P(B \cap C) = P(A|B \cap C)P(B|C)P(C) \\ P(A \cap B \cap C \cap D) &= P(A|B \cap C \cap D)P(B|C \cap D)P(C|D)P(D) \quad \text{usw.} \end{aligned}$$

10.2 Bayes'sche Parameterschätzung

Der Bayes'sche Ansatz hat zum Ziel, vorhandene *Vorinformationen* (Expertenwissen) mit *aktuellen Messdaten* zu kombinieren, um eine verbesserte Schätzung für die Parameter einer Wahrscheinlichkeitsverteilung zu erzielen. Alle Wahrscheinlichkeiten werden *subjektivistisch*, also als Vertrauensgrade, aufgefasst.

Gegeben sei eine Zufallsgröße X mit einer Verteilungsfunktion, die von einem oder mehreren Parametern abhängt, etwa mit θ bezeichnet. Grundsatz des Bayesansatzes ist, dass *alles* als Zufallsgröße aufgefasst wird, somit auch der Parameter θ . Dieser besitzt also seinerseits eine Wahrscheinlichkeitsverteilung, die wir kurz mit $p(\theta)$ bezeichnen. (Eine präzisere Formulierung, unterschieden nach diskretem und kontinuierlichem Fall, folgt unten.) Die Verteilung $p(\theta)$ wird als *A-priori-Verteilung* bezeichnet und dazu verwendet, um die Vorinformationen zu beschreiben.

In einer Messung sei eine Realisierung ξ der Zufallsgröße X festgestellt worden. Wir wollen dieses Ergebnis mit der Vorinformation kombinieren, und erhalten damit die *A-posteriori-Verteilung* $p(\theta|\xi)$. Das Bayes'sche Theorem besagt:

$$p(\theta|\xi) = \frac{p(\xi|\theta)p(\theta)}{p(\xi)}$$

wobei sich $p(\xi)$ nachträglich durch die Forderung, dass $\int p(\theta|\xi) d\theta = 1$ sein muss, berechnen lässt. Man schreibt oft der Einfachheit halber

$$p(\theta|\xi) \propto p(\xi|\theta) p(\theta),$$

das heißt, die linke Seite ist proportional zur rechten Seite.

Die bedingte Wahrscheinlichkeit $p(\xi|\theta)$ ist die Wahrscheinlichkeit, dass X den Wert ξ annimmt, wenn der Verteilungsparameter Θ gleich θ ist. Sie wird als *Likelihoodfunktion* oder *Plausibilitätsfunktion* bezeichnet. Falls nicht nur ein Messwert sondern eine Stichprobe $\xi = (\xi_1, \dots, \xi_n)$ vorliegt, so sind die einzelnen Werte unabhängig und daher die Likelihoodfunktion das Produkt der einzelnen Wahrscheinlichkeiten:

$$p(\xi|\theta) = p(\xi_1|\theta)p(\xi_2|\theta) \dots p(\xi_n|\theta) = \prod_{j=1}^n p(\xi_j|\theta).$$

Wir präzisieren die Vorgangsweise im diskreten und kontinuierlichen Fall.

Fall 1: Die Zufallsgrößen X und Θ sind beide diskret. Dann handelt es sich in der Bayes'schen Formel um gewöhnliche Punktwahrscheinlichkeiten:

$$P(\Theta = \theta|X = \xi) \propto P(X = \xi|\Theta = \theta) P(\Theta = \theta)$$

bzw. im Falle des Vorliegens einer Stichprobe

$$P(\Theta = \theta|\xi) \propto \prod_{j=1}^n P(X = \xi_j|\Theta = \theta) P(\Theta = \theta).$$

Fall 2: X ist diskret und Θ kontinuierlich. Dann sind die $P(X = \xi|\Theta = \theta)$ Punktwahrscheinlichkeiten, aber $p(\theta|X = \xi)$ ist eine (so genannte bedingte) Wahrscheinlichkeitsdichte; auch die A-priori-Verteilung ist durch eine Dichte gegeben:

$$p(\theta|X = \xi) \propto P(X = \xi|\Theta = \theta) p(\theta)$$

bzw. im Falle des Vorliegens einer Stichprobe

$$p(\theta|\xi) \propto \prod_{j=1}^n P(X = \xi_j|\Theta = \theta) p(\theta).$$

Beispiel 10.4 Der Fall einer Alternativverteilung. Es sei zum Beispiel die Wahrscheinlichkeit des Ergebnisses „Kopf“ bei einem Münzwurf zu beurteilen. Die Zufallsgröße X hat demnach die Werte $X = 1$ mit Wahrscheinlichkeit θ und $X = 0$ mit Wahrscheinlichkeit $1 - \theta$. Der Parameter θ kann jedenfalls Werte zwischen 0 und 1 annehmen und ist daher eine kontinuierliche Zufallsgröße.

Bei einem einzelnen Münzwurf ($\xi = 0$ oder $\xi = 1$) erhalten wir folgende Likelihoodfunktionen:

$$P(X = 1|\Theta = \theta) = \theta \quad \text{bzw.} \quad P(X = 0|\Theta = \theta) = 1 - \theta.$$

Bei einer Stichprobe aus mehreren Würfeln ist die Likelihoodfunktion das Produkt der einzelnen Trefferwahrscheinlichkeiten. Ist zum Beispiel $\xi = (1, 0, 1, 1, 0)$ (mit $n = 5$), so ergibt sich

$$P(\xi|\Theta = \theta) = \prod_{j=1}^5 P(X = \xi_j|\Theta = \theta) = \theta^3(1 - \theta)^2.$$

Fall 3: Sowohl X als auch Θ sind kontinuierlich. Dann sind die Wahrscheinlichkeiten in der Bayes-Formel als Dichten zu interpretieren. Die Größen X und Θ besitzen eine gemeinsame Dichte $p(x, \theta)$ und Randdichten $p_X(x) = \int_{-\infty}^{\infty} p(x, \theta) d\theta$, $p_\Theta(\theta) = \int_{-\infty}^{\infty} p(x, \theta) dx$. Die *bedingten Dichten* sind definiert als

$$p(x|\Theta = \theta) = \frac{p(x, \theta)}{p_\Theta(\theta)}, \quad p(\theta|X = x) = \frac{p(x, \theta)}{p_X(x)}.$$

Die Bayes'sche Formel ist dann zu interpretieren als Aussage über die bedingten Dichten:

$$p(\theta|X = x) \propto p(x|\Theta = \theta) p(\theta) \quad \text{bzw.} \quad p(\theta|\boldsymbol{\xi}) \propto \prod_{j=1}^n p(\xi_j|\Theta = \theta) p(\theta).$$

Der Bayes-schätzer für den Parameter Θ : Der Bayes-schätzer für Θ ist der Erwartungswert der A-posteriori-Verteilung:

$$\begin{aligned} \hat{\theta} = E(\Theta|\boldsymbol{\xi}) &= \sum_{i=1}^k \theta_i P(\Theta = \theta_i|\boldsymbol{\xi}), \quad \dots \quad \Theta \in \{\theta_1, \dots, \theta_n\} \text{ diskret} \\ &= \int_{-\infty}^{\infty} \theta p(\theta|\boldsymbol{\xi}) d\theta, \quad \dots \quad \Theta \in \mathbb{R} \text{ kontinuierlich} \end{aligned}$$

Beispiel 10.5 Wir setzen Beispiel 10.4 fort und wollen die A-posteriori-Verteilung in verschiedenen Fällen ermitteln. Zunächst zur Wahl der A-priori-Verteilung. Erfahrungen mit Münzwürfen legen nahe, dass der Parameter θ , etwa bei Centmünzen, zwischen 0.4 und 0.6 liegt. Nehmen wir aber an, dass wir *keine* Vorinformation haben, so modellieren wir dies durch eine Gleichverteilung für den Parameter θ im Intervall $[0, 1]$ (andere Werte kommen nicht in Frage), also

$$p(\theta) = \begin{cases} 1, & 0 \leq \theta \leq 1, \\ 0, & \theta < 0 \text{ oder } \theta > 1. \end{cases}$$

Nehmen wir nun an, wir machen einen Münzwurf mit dem Ausgang $\xi = 1$. Nach dem Bayes'schen Theorem ergibt sich für die A-posteriori-Verteilung

$$p(\theta|X = 1) \propto p(X = 1|\theta)p(\theta) = \theta p(\theta),$$

das heißt,

$$p(\theta|X = 1) = \begin{cases} c\theta, & 0 \leq \theta \leq 1, \\ 0, & \theta < 0 \text{ oder } \theta > 1. \end{cases}$$

Die Konstante c erhält man aus der Bedingung

$$1 = \int_{-\infty}^{\infty} p(\theta|X = 1) d\theta = \int_0^1 c\theta d\theta = \frac{c}{2}$$

zu $c = 2$. Die A-posteriori-Verteilung für den Parameter θ ist somit eine Dreiecksverteilung.

Der Bayes-Schätzer für θ ist der Erwartungswert von Θ gemäß der A-posteriori-Verteilung, also

$$\hat{\theta} = \int_{-\infty}^{\infty} \theta p(\theta|X = 1) d\theta = \int_0^1 2\theta^2 d\theta = \frac{2}{3}.$$

Für den a-priori nicht vorinformierten Bayesianer ergibt sich also nach einem Münzwurf mit Ergebnis „Kopf“, dass die Münze als gezinkt zu betrachten ist mit Wahrscheinlichkeit von $2/3$ für einen Kopfwurf.

Die Formulierung ist etwas überspitzt, um das Grundprinzip herauszuarbeiten. Selbstverständlich wird auch ein Bayesianer die Münze mehrmals werfen, also seine A-posteriori-Verteilung aktualisieren („updating“). Ist die Münze ungezinkt und die Priorverteilung nicht einengend, so konvergieren die weiteren A-posteriori-Verteilungen zum frequentistischen Schätzer $\hat{\theta} = 1/2$. Der Vorteil der Bayes-Methode ist allerdings, dass sie im Gegensatz zur frequentistischen Methode schon bei *kleinen Stichproben* plausible Aussagen machen kann.

Beispiel, fortgesetzt: Nehmen wir an, die Münze wurde fünfmal geworfen mit Ergebnis $\xi = (1, 0, 1, 1, 0)$. Die A-posteriori-Dichte ist

$$p(\theta|\xi) = (1, 0, 1, 1, 0) = c\theta^3(1-\theta)^2 \quad \text{für } 0 \leq \theta \leq 1.$$

Die Konstante c erhält man zu $c = 60$ aus

$$1 = c \int_0^1 \theta^3(1-\theta)^2 d\theta = \frac{c}{60}.$$

Die Posteriorverteilung ist somit eine Betaverteilung mit Dichte

$$p(\theta|\xi) = (1, 0, 1, 1, 0) = 60\theta^3(1-\theta)^2, \quad 0 \leq \theta \leq 1.$$

Der Bayes-Schätzer für θ ist nunmehr

$$\hat{\theta} = 60 \int_0^1 \theta^4(1-\theta)^2 d\theta = \frac{4}{7}.$$

Um den Effekt der A-priori-Verteilung vorzuführen, nehmen wir (sinnvollerweise) eine Verteilung, die mehr in der Nähe von $\theta = 1/2$ konzentriert ist, ist zum Beispiel eine Betaverteilung $p(\theta) \propto \theta(1-\theta)$. Mit dem Stichprobenergebnis $\xi = (1, 0, 1, 1, 0)$ gilt dann

$$p(\theta|\xi) \propto p(\xi|\Theta = \theta) \cdot p(\theta) \propto \theta^3(1-\theta)^2 \cdot \theta(1-\theta) = \theta^4(1-\theta)^3.$$

Die A-posteriori-Verteilung ist demnach ebenfalls eine Betaverteilung mit Parametern $r - 1 = 4$, $s - 1 = 3$. Deren Erwartungswert ist nach einer allgemeinen Formel für Betaverteilungen gleich

$$E(\Theta|\xi) = \frac{r}{r+s} = \frac{5}{8}.$$

Der damit gewonnene Schätzer $\hat{\theta} = 5/8$ liegt näher bei $1/2$ als jener mit einer A-priori-Gleichverteilung.

Diese kurze Einführung in die mittlerweile hoch entwickelte Bayes-Statistik möge genügen. Für weitere Informationen siehe Viertl(2003).

10.3 Subjektive Wahrscheinlichkeiten

Im Zuge der Projektplanung, aber auch bei der Beurteilung von Eingangsgrößen in der Bemessung, sind die Eintrittswahrscheinlichkeiten oder die Verteilungen von Zufallsgrößen häufig durch persönliche Beurteilung durch die Bearbeiter oder Experten festzulegen. Dies trifft insbesondere auf die Wahl der A-priori-Verteilung bei der Bayes'schen Parameterschätzung zu.

In der Anfangsphase eines Projekts fehlen einerseits oft statistische Daten, etwa in Form von Bodenproben oder Laborversuchen, andererseits geht es in der Projektplanung um einen zukünftigen Zustand oder ein Risiko, das erst unter zukünftigen Bedingungen zum Tragen kommt. In solchen Fällen muss die Zuweisung von Wahrscheinlichkeitswerten – nach bestem Wissensstand – durch Abwägung und Diskussion erfolgen. Die Wahrscheinlichkeitstheorie spricht von *subjektiven Wahrscheinlichkeiten* und es sind eine Reihe Vorgangsweisen zu ihrer Interpretation und zu ihrer Ermittlung entwickelt worden.

In der subjektivistischen Interpretation wird die Wahrscheinlichkeit eines Ereignisses A als Vertrauensgrad in sein Auftreten gedeutet. Hier wird bewusst in Kauf genommen, dass dieser Vertrauensgrad eine subjektive Einschätzung bedeutet und daher von verschiedenen Experten verschiedene Werte zugewiesen bekommen kann. Selbstverständlich muss die Bewertung rational und unter Zuhilfenahme aller verfügbaren Information erfolgen. Um eine operative Methode zu erhalten, den subjektiven Vertrauensgrad zu quantifizieren, wurde von Savage(1954) vorgeschlagen, diesen als Indifferenzpreis für ein gedachtes Wettspiel anzusetzen. Das Spiel ergebe einen Gewinn von 1, falls A auftritt, und von 0, falls A nicht auftritt. Dann ist

$$\begin{aligned} P(A) &= \text{Indifferenzpreis für das Spiel} = \\ \overline{P}(A) &= \text{maximaler Kaufpreis} = \\ \underline{P}(A) &= \text{minimaler Verkaufspreis.} \end{aligned}$$

Soll zum Beispiel die Erfolgswahrscheinlichkeit einer Alternativverteilung beurteilt werden, also etwa bei einem Münzwurf die Wahrscheinlichkeit für das Ereignis $A = \text{„Kopf“}$, so würde das Wettspiel in einem Münzwurf bestehen mit der Gewinnausschüttung von 1.00 Euro, falls „Kopf“ auftritt. Sie wären vermutlich interessiert, diese Wette um einen Einsatz von 0.00, 0.10, 0.20 Euro zu kaufen; bei einem Kaufpreis von 1.00, 0.90, 0.80 Euro wären Sie wahrscheinlich nicht interessiert, einzusteigen. Schwieriger wird es schon bei einem Kaufpreis von 0.45 oder 0.55 Euro. Wenn Sie die Münze für ungezinkt halten, werden Sie das Kaufinteresse voraussichtlich genau bei 0.50 Euro verlieren, das heißt, es wird für Sie egal sein, ob Sie mitwetten oder nicht. Dies ist also Ihr Indifferenzpreis, und damit ist Ihre subjektive Wahrscheinlichkeit für das Auftreten von „Kopf“ gerade $P(A) = 0.5$.

Diese Vorgangsweise wird *Introspektion* genannt. Wichtiger für die Risikoanalyse ist allerdings die Expertenbefragung, auch *Mediation*, *Delphi-Methode* oder auf Englisch *elicitation* genannt. Dabei werden Einzelwahrscheinlichkeiten und Verteilungsfunktionen in einem Diskussionsprozess festgelegt. Es können dabei unter anderem

- statistische Daten,
- Erfahrungswerte, Normwerte,
- Bewertungstabellen,
- Ereignisbäume

herangezogen werden.

Ein Beispiel einer der Bewertungstabellen ist etwa die Sherman-Kent-Skala, mit deren Hilfe qualitativen Wahrscheinlichkeiten quantitative Wertebereiche zugeordnet werden können; Abbildung 10.3, aus Meyer et al. (2001):

Sherman Kent rating scale

Order Of Likelihood	Synonyms	Chances In 10	Percent (%)
Nearly Certain	Virtually (almost) We are convinced Highly probable Highly likely	9	99%
		8	80%
Probable	Likely We believe We estimate Chances are good It is probable	7	
		6	60%
Even Chance	Chances are slightly better than Chances are about even Chances are slightly less than	5	
		4	40%
Improbable	Probably not Unlikely We believe not	3	
		2	20%
Nearly Impossible	Almost impossible Only a slight chance Highly doubtful	1	10%

Abbildung 10.1: Die Bewertungsskala von Sherman und Kent.

Beispiel: Bestimmung einer kontinuierlichen Wahrscheinlichkeitsverteilung. Die folgende Vorgangsweise wurde von O'Hagan(1998) vorgeschlagen. Es soll die Verteilung einer Zufallsgröße, etwa einer Schadenssumme S , eruiert werden. Dazu müssen die Experten zunächst eine untere Grenze U und eine obere Grenze O sowie den Modalwert (häufigsten Wert) M festlegen. Anschließend werden sie angehalten, die Wahrscheinlichkeiten für die folgenden Intervalle anzugeben:

- (a) $[U, M]$,

- (b) $[U, (U + M)/2]$,
- (c) $[(M + O)/2, O]$,
- (d) $[U, (U + 3M)/4]$,
- (e) $[(3M + O)/4, O]$.

Es wurden zum Beispiel die folgenden Grenzen angegeben:

$$U = 165, \quad M = 190, \quad O = 250.$$

Die von O'Hagan vorgeschlagenen Intervalle wurden mit den folgenden Wahrscheinlichkeiten versehen:

- (a) $[165, 190] : P = 0.500$,
- (b) $[165, 177.5] : P = 0.175$,
- (c) $[220, 250] : P = 0.050$,
- (d) $[165, 183.75] : P = 0.330$,
- (e) $[205, 250] : P = 0.250$.

Daraus errechnen sich die Wahrscheinlichkeiten der nicht überlappenden Teilintervalle laut Tabelle unten:

Intervall	Klassenmitte z_j	Wahrscheinlichkeit p_j	Histogrammhöhe
$[165, 177.5]$	171.250	0.175	0.0140
$[177.5, 183.75]$	180.625	0.155	0.0248
$[183.75, 190]$	186.875	0.170	0.0272
$[190, 205]$	197.500	0.250	0.0167
$[205, 220]$	212.500	0.200	0.0133
$[220, 250]$	235.000	0.050	0.0017

Im nächsten Schritt wird nun eine Verteilungsfunktion angepasst. Wir nehmen in diesem Fall eine Lognormalverteilung $\mathcal{LogN}(\mu, \sigma^2)$. Wir berechnen dazu Mittelwert und Varianz des Logarithmus der Klassenmitten als Schätzung, also

$$\mu \approx \sum_j \log z_j p_j \approx 5.2609.$$

Ebenso ergibt sich die Varianz zu

$$\sigma^2 \approx \sum_j (\log z_j - \mu)^2 p_j; \quad \sigma \approx 0.0853.$$

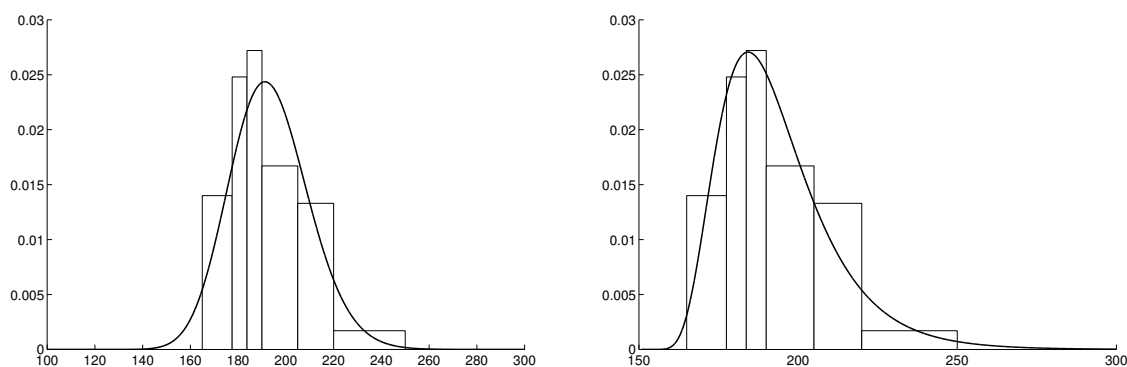


Abbildung 10.2: Angepasste zweiparametrische (links) und dreiparametrische (rechts) Lognormalverteilung.

Das Histogramm mit der angepassten Lognormalverteilung ist in Abbildung 10.3 (links) abgebildet. Da die angepasste Verteilung die Schiefeite des Histogramms nicht gut wiedergibt, machen wir noch einen Durchgang („stepping back“) und versuchen, eine verschobene Lognormalverteilung anzupassen. Das Histogramm legt nahe, um 150 nach links zu verschieben. Wir passen daher eine Lognormalverteilung an die verschobenen Klassenmittelpunkte $z_j - 150$ an; die Berechnung erfolgt wie oben und gibt nunmehr ein befriedigendes Modell der Verteilung der Schadenssumme – vgl. Abbildung 10.3 (rechts).

10.4 Ereignisbäume

Ereignisbäume ermöglichen die Berücksichtigung von Abhängigkeiten von Ereignissen untereinander, zum Beispiel kausaler Zusammenhänge oder zeitlicher Abfolgen. Der Baum besteht aus Knoten (den Ereignissen) und Kanten, die die Zusammenhänge darstellen. Die Kanten können mit Bewertungen versehen werden (zum Beispiel Schadenskosten), die Verzweigungen an den Knoten können mit Wahrscheinlichkeiten versehen werden. Ist der Baum aufgebaut, so können die Gesamtbewertungen an den Knospen (Enden) additiv, die dazugehörigen Eintrittswahrscheinlichkeiten multiplikativ berechnet werden (statistische Unabhängigkeit der Äste vorausgesetzt).

Beispiel 10.6 Es soll das Schadensrisiko durch Hochwasser während des Baus eines Erd Damms in einem Hochgebirgstal abgeschätzt werden, und zwar in seinen Auswirkungen auf die Baukosten und auf die Bauzeit. Das vorliegende Beispiel befasst sich mit der Baustelle Moosbach/Kartell und wurde 2004 bearbeitet¹.

Für die Schüttphase des Damms wurde eine Bauwasserumleitung auf HQ50 dimensioniert. Betrachtete Risiken waren:

¹Mathematische Methoden im Baubetrieb und Risikoanalyse, Universität Innsbruck, Bearbeitung durch M. Nitsch und A. Strolz.

- Verklausung des Umleitungsstollens;
- Bauzustand des Damms (Kerngraben offen/nicht offen);
- Beschädigung der Dammkrone.

Die kausalen Zusammenhänge wurden mit Hilfe eines Ereignisbaums dargestellt (gesamter Baum in Abbildung 10.6, Ausschnitt in Abbildung 10.6).

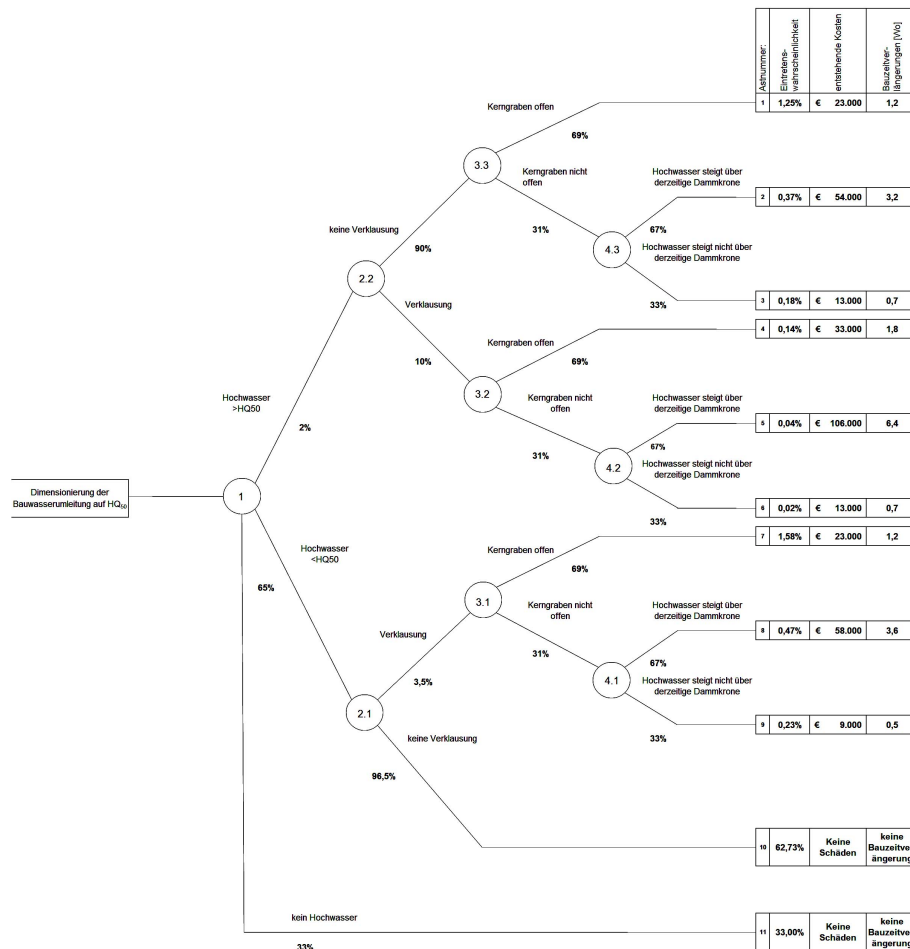


Abbildung 10.3: Vollständiger Ereignisbaum für Hochwasserrisiko/Erddamm.

Die Wahrscheinlichkeiten der einzelnen Verzweigungen an den jeweiligen Knoten müssen nun Schritt für Schritt festgelegt werden. Beispielhaft seien drei Knoten herausgegriffen. Im Knoten 1 tritt eine Verzweigung für das Eintreten von HQ50 auf; die Eintretenswahrscheinlichkeit davon ist selbstverständlich gleich 2%. Im Falle des Nichteintretens eines HQ50 wurde eine weitere Unterscheidung vorgenommen, mit subjektiv abgeschätzten Wahrscheinlichkeitswerten (siehe Abbildung 10.6). Die Summe der Wahrscheinlichkeiten am Knoten muss gleich 1 sein bzw. 100% entsprechen.

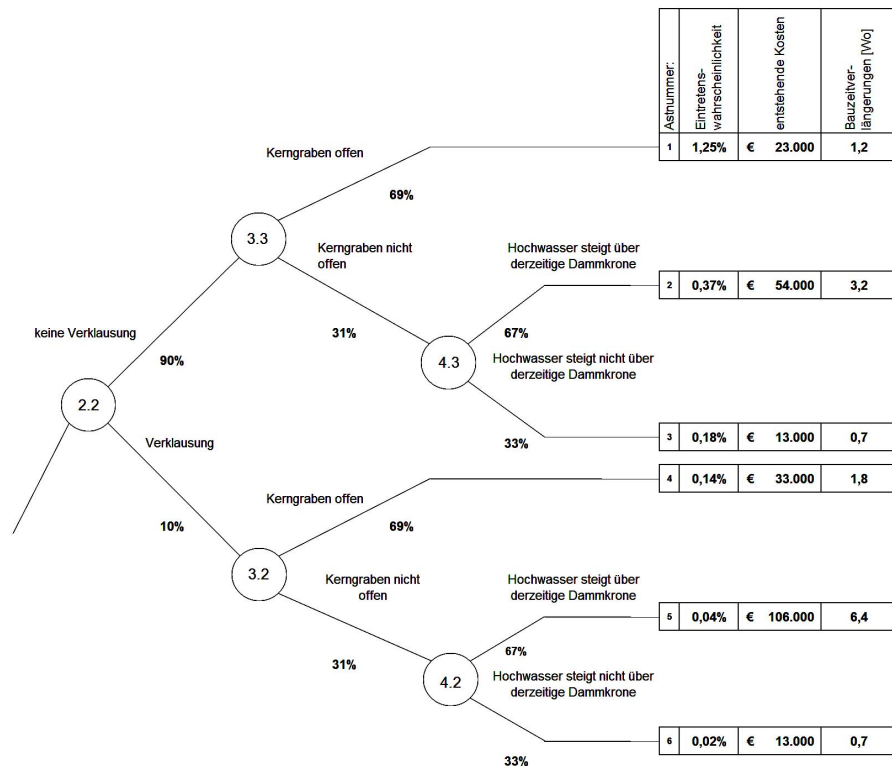


Abbildung 10.4: Ausschnitt aus dem Ereignisbaum.

Knotennummer: 1

HQ50= 6 m³/s

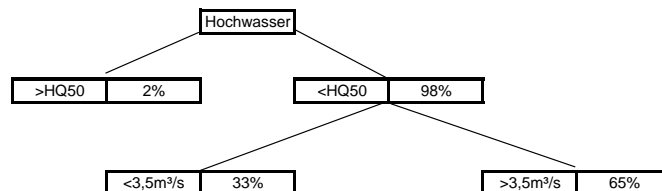


Abbildung 10.5: Ermittlung des Hochwasserrisikos am Knoten 1.

Am Knoten 2.2 wurden die Verkläusungswahrscheinlichkeiten mit Hilfe von Rängen ermittelt, die die Wichtigkeit der möglichen Ursachen bewertet (Abbildung 10.6). Die Bewertung des Bauzustands offener/verbauter Kerngraben wurde anhand der Bauzeitverteilung im Jahresablauf ermittelt (Abbildung 10.6).

Als Beispiel der Ermittlung der Auswirkung auf Kosten und Bauzeit sei beispielhaft der Knoten 3.1 herangezogen. Hier sind entsprechend baubetriebliche Grundsätze anzuwenden; dies wurde in der Tabelle in Abbildung 10.6 durchgeführt.

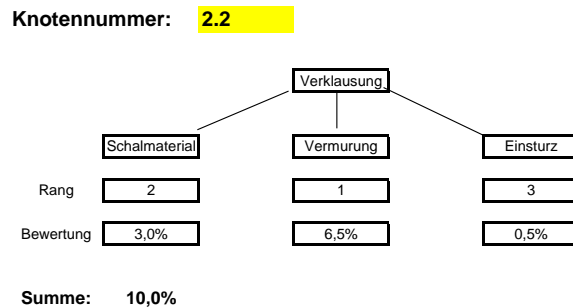


Abbildung 10.6: Ermittlung des Verklauungsrisikos am Knoten 2.2.

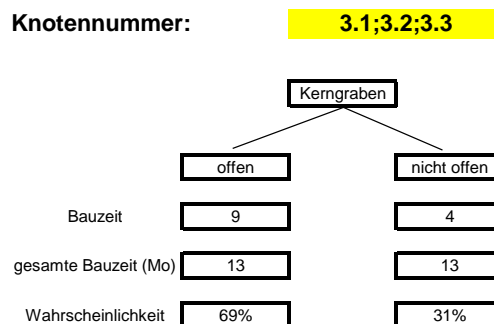


Abbildung 10.7: Wahrscheinlichkeit für offenen Kerngraben bei Hochwassereintritt.

Knoten:	3.1	Kerngraben offen	<HQ50
		nicht gesamte Länge des Kontrollganges in Massenberechnung enthalten (Bauablauf)	
Abgetragenes Volumen:			1200,00 m³
Beschädigtes Dammvolumen:			300,00 m³
Zu entfernendes und neu zu schüttendes Volumen:			1500,00 m³
	EHP	Masse	Pos.Preis
Ausbaggern+Transp.	3,16	1500,00	4740,00
Wiedereinbau	0,95	1500,00	1425,00
Vorhaltekosten	41914,99	0,30	12574,50
Regieanteile:			4000,00
	Schadenssumme: € 22.739,50		
Bauzeitverlängerung:		6,0 Tage	
		1,2 Wochen	

Abbildung 10.8: Ermittlung der Auswirkung des Risikos *offener Kerngraben* am Knoten 3.1.

Die Gesamtbewertung des Risikos erhält man durch statistische Auswertung der Kosten/Bauzeitverlängerung mit zugehörigen Wahrscheinlichkeiten. Erste Indikatoren sind der Erwartungswert der Schadenssumme/Bauzeitverlängerung und die Varianz. Ein deutlicheres Bild der Risikoverteilung liefert ein Histogramm.

Genauer: Die Verzweigungswahrscheinlichkeiten am Startknoten sind gewöhnliche Wahrscheinlichkeiten. An den späteren Knoten handelt es sich um bedingte Wahrscheinlichkeiten.

ten (jeweils unter der Bedingung, dass der entsprechende Knoten erreicht wurde). Verfolgt man einen Zweig i vom Startknoten bis zum Endknoten, so ergibt sich die Wahrscheinlichkeit p_i , den Endknoten zu erreichen, nach der Multiplikationsformel (Abschnitt 10.1) als Produkt der Wahrscheinlichkeiten längs des Zweigs. Die Bewertung x_i erhält man als Summe der Kantenbewertungen längs des Zweigs. Alle m Endknoten gemeinsam stellen eine diskrete Zufallsgröße X mit Ausprägungen x_1, \dots, x_m und zugehörigen Wahrscheinlichkeiten p_1, \dots, p_m dar. Deren Erwartungswert ist

$$E(X) = \sum_{i=1}^m x_i p_i.$$

Kapitel 11

Lineare Regression

Dieses Kapitel ist ein paar kurzen einführenden Bemerkungen zur linearen Regression gewidmet. Das umfangreiche Gebiet geht über den Rahmen der Vorlesung weit hinaus. Eine umfassende Darstellung findet man in Montgomery et al. (2012). Weitere Beispiele und Unterrichtsmaterial findet man in Kreyszig(1970) und Oberguggenberger et al. (2009).

11.1 Univariate lineare Regression

Gegeben seien Paare von Messdaten (x_i, y_i) , $i = 1, \dots, n$, geometrisch eine Punktwolke in der Ebene. Dabei können die x_i und y_i durchaus mehrfach auftreten, also auch zu gegebenem x_i mehrere Messwerte y_{i1}, \dots, y_{ip} vorliegen. Die Standardaufgabe der linearen Regression ist es, ein *lineares Modell*

$$y = \beta_0 + \beta_1 x$$

an die Messdaten anzupassen, also eine *beste Gerade* durch die Punktwolke zu legen.

Beispiel 11.1 Eine Stichprobe von $n = 44$ Studenten im Jahr 1998. Hier ist x die Körpergröße [cm] und y das Körpergewicht [kg], vergleiche die Datentabelle 7.1. Abbildung 11.1 stellt diese Werte im Streudiagramm dar. Die beste Gerade postuliert und beschreibt einen linearen Zusammenhang zwischen Größe und Gewicht.

Wir beginnen mit der Modellbildung des Regressionsansatzes. Das postulierte Modell für den Zusammenhang zwischen x und y ist das lineare Modell

$$y = \beta_0 + \beta_1 x$$

mit unbekannten Koeffizienten β_0 und β_1 . Die vorliegenden Daten liegen jedoch nicht exakt auf der entsprechenden Gerade, sondern zeigen Abweichungen ε_i , $i = 1, \dots, n$:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

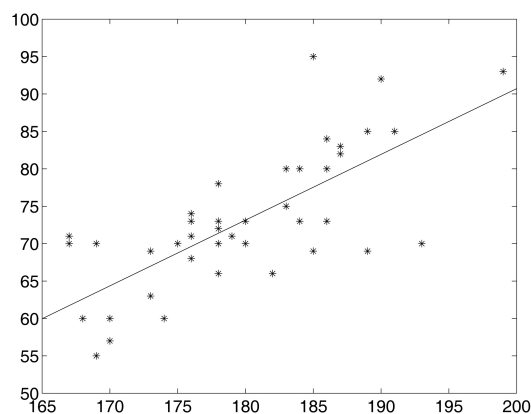


Abbildung 11.1: Streudiagramm Körpergröße/Körpergewicht mit Regressionsgerade.

Wir wollen aus den vorliegenden Daten Schätzwerte $\hat{\beta}_0, \hat{\beta}_1$ für β_0, β_1 gewinnen. Dies erfolgt durch Minimierung der Summe der Fehlerquadrate

$$L(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2,$$

sodass also $\hat{\beta}_0, \hat{\beta}_1$ Lösung des folgenden Minimierungsproblems ist:

$$L(\hat{\beta}_0, \hat{\beta}_1) = \min \left(L(\beta_0, \beta_1) : \beta_0 \in \mathbb{R}, \beta_1 \in \mathbb{R} \right)$$

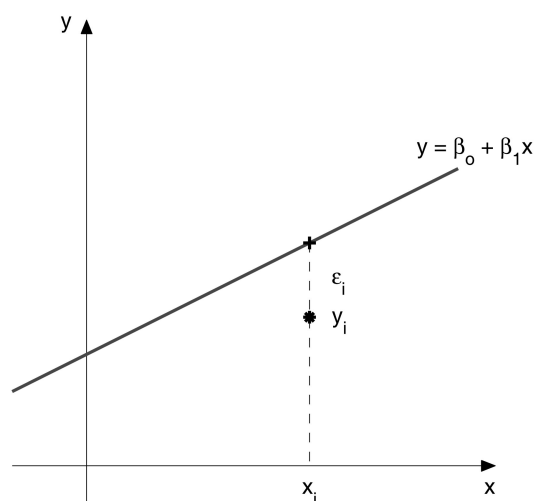


Abbildung 11.2: Lineares Modell und Fehler.

Wir erhalten $\hat{\beta}_0$ und $\hat{\beta}_1$, indem wir die partiellen Ableitungen von L nach β_0 und nach β_1 Null setzen:

$$\begin{aligned}\frac{\partial L}{\partial \beta_0}(\hat{\beta}_0, \hat{\beta}_1) &= -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ \frac{\partial L}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) &= -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0\end{aligned}$$

Dies führt auf das lineare Gleichungssystem für $\hat{\beta}_0, \hat{\beta}_1$ (die *Normalgleichungen*):

$$\begin{aligned}n \hat{\beta}_0 + \left(\sum x_i \right) \hat{\beta}_1 &= \sum y_i \\ \left(\sum x_i \right) \hat{\beta}_0 + \left(\sum x_i^2 \right) \hat{\beta}_1 &= \sum x_i y_i\end{aligned}$$

mit der Lösung

$$\begin{aligned}\hat{\beta}_0 &= \left(\frac{1}{n} \sum y_i \right) - \left(\frac{1}{n} \sum x_i \right) \hat{\beta}_1 \\ \hat{\beta}_1 &= \frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2}\end{aligned}$$

Das *Ergebnis der Regression* ist die *geschätzte Regressionsgerade*

$$y = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Die *durch das Modell prognostizierten Werte* sind dann

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i = 1, \dots, n.$$

Deren Abweichungen von den Messwerten y_i bezeichnet man als *Residuen*

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = 1, \dots, n.$$

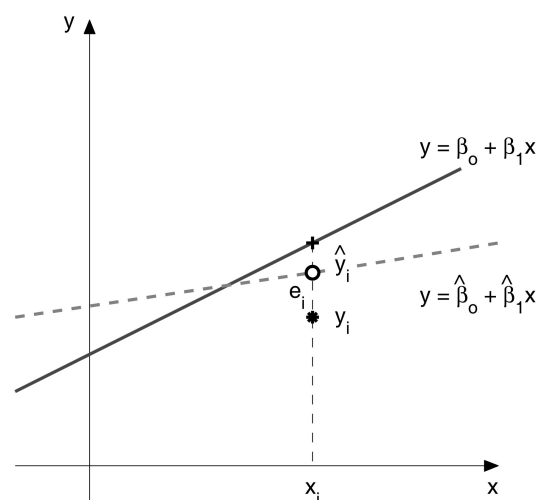


Abbildung 11.3: Modell, Schätzung, Fehler, Residuum.

Damit ist das *deterministische Regressionsmodell* spezifiziert. Im *statistischen Regressionsmodell* werden die Fehler ε_i als Zufallsgrößen mit Erwartungswert Null interpretiert. Unter weiteren Zusatzannahmen (Unabhängigkeit, konstante Varianz, Normalverteilung) wird das Modell dann statistischen Test- und Diagnoseverfahren zugänglich gemacht.

Umformulierung der Normalgleichungen zur besseren numerischen Handhabbarkeit. Wir führen die folgenden Vektoren und Matrizen ein:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

Die Relation Daten-Modell

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

schreibt sich dann in Matrixform einfacher zu

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

Weiters ist

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \cdot \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$\mathbf{X}^\top \mathbf{y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix},$$

sodass die Normalgleichungen sich als

$$\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$$

darstellen. Die Lösung ist somit

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y};$$

die Prognosewerte und Residuen sind

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}, \quad \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}.$$

Beispiel 11.2 Fortsetzung von Beispiel 11.1. Die Koeffizienten im Beispiel Größe/Gewicht ergeben sich zu $\hat{\beta}_0 = -85.0209$, $\hat{\beta}_1 = 0.8787$ mit der in der Abbildung dargestellten Regressionsgerade.

11.2 Rudimente der ANOVA

Die Abkürzung ANOVA steht für *ANalysis Of VAriance*. Ein erstes Indiz für die Anpassungsqualität des linearen Modells liefert die *Varianzanalyse*. Das arithmetische Mittel der y -Werte ist

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Die Abweichung des Messwertes y_i vom Mittelwert \bar{y} ist $y_i - \bar{y}$. Die *totale quadratische Abweichung* oder *Gesamtvariabilität* der Daten ist

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2.$$

Diese wird in zwei Teilkomponenten zerlegt, und zwar in

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Dies wird folgendermaßen interpretiert: $\hat{y}_i - \bar{y}$ ist die Abweichung des Prognosewertes vom Mittelwert, und

$$SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

die *durch die Regression beschriebene Datenvariabilität* (Sum of Squares - Regression). Andererseits sind $e_i = y_i - \hat{y}_i$ gerade die Residuen, und

$$SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

die Quadratsumme der Residuen, welche als *Restvariabilität* interpretiert wird, die durch das lineare Modell unerklärt verbleibt (Sum of Squares-Error). Wir haben also die Varianzzerlegung

$$S_{yy} = SS_R + SS_E$$

erhalten. Die Größe

$$R^2 = \frac{SS_R}{S_{yy}}$$

wird als *Bestimmtheitsmaß* bezeichnet und misst den Anteil der durch die Regression erklärten Variabilität an der Gesamtvariabilität. Im Grenzfall einer exakten Anpassung,

wenn die Regressionsgerade genau durch alle Datenpunkte geht, ist $SS_E = 0$ und damit $R^2 = 1$. Ein kleines R^2 ist ein Indiz dafür, dass das lineare Modell weniger gut an die Daten passt. $R = \sqrt{R^2}$ ist der empirische Korrelationskoeffizient von \mathbf{x} und \mathbf{y} .

Bemerkung: Im Falle der Regression mittels einer Geraden durch den Ursprung, also für das konstantenfreie Modell $y = \beta_1 x$, ist die Varianzzerlegung ungültig.

Beispiel 11.3 Fortsetzung von Beispiel 11.1. Es ergibt sich

$$S_{yy} = 3535.9, \quad SS_E = 1637.7, \quad SS_R = 1898.2$$

und

$$R^2 = 0.5368, \quad R = 0.7327,$$

also eine eher schlechte Anpassung (bzw. ein Indiz dafür, dass Körpergröße und Gewicht eher keinen linearen Zusammenhang haben).

11.3 Multivariate lineare Regression

In der multivariaten linearen Regression hängt die Variable y nicht nur von einer Regressorvariable x ab, sondern von mehreren Variablen, etwa x_1, x_2, \dots, x_k . Man beachte den Notationswechsel: x_i bezieht sich auf die i -te Regressorvariable. Die Messwerte der i -ten Regressorvariablen werden nun mit zwei Indizes versehen, nämlich $x_{i1}, x_{i2}, \dots, x_{in}$, sodass nunmehr $k \times n$ Messwerte vorliegen. Gesucht ist wieder ein lineares Modell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

mit noch unbekannten Koeffizienten $\beta_0, \beta_1, \dots, \beta_k$.

Das allgemeine multiple lineare Modell $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$ beinhaltet als Spezialfall die einfache lineare Regression mit mehreren nichtlinearen Formfunktionen, also

$$y = \beta_0 + \beta_1 \varphi_1(x) + \beta_2 \varphi_2(x) + \dots + \beta_k \varphi_k(x),$$

wobei $x_1 = \varphi_1(x), x_2 = \varphi_2(x), \dots, x_k = \varphi_k(x)$ als Regressorvariablen betrachtet werden. Insbesondere kann man polynomiale Ansätze

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k$$

zulassen.

Die Messdaten (je n Stück) für die einzelnen Variablen stellen sich schematisch wie folgt dar:

Variable	y	x_1	x_2	\dots	x_k
Messung Nr. 1	y_1	x_{11}	x_{21}	\dots	x_{k1}
Messung Nr. 2	y_2	x_{12}	x_{22}	\dots	x_{k2}
\vdots	\vdots	\vdots	\vdots		\vdots
Messung Nr. n	y_n	x_{1n}	x_{2n}	\dots	x_{kn}

Jeder Wert y_i ist zu approximieren durch

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad i = 1, \dots, n$$

mit den Abweichungen ε_i . Die geschätzten Koeffizienten $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ werden wieder als Lösung der Minimierungsaufgabe

$$L(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \varepsilon_i^2 \rightarrow \min$$

gewonnen. Unter Verwendung der Vektor- und Matrixnotation

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

lautet das lineare Modell an die Daten in Kurzform wieder

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Die bestangepassten Koeffizienten erhält man wieder nach der Formel

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

mit den Prognosewerten und Residuen

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}, \quad \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}.$$

Die Varianzzerlegung

$$S_{yy} = SS_R + SS_E$$

ist weiterhin gültig; das *multiple Bestimmtheitsmaß*

$$R^2 = SS_R / S_{yy}$$

ist ein Indikator der Anpassungsgüte des Modells.

Das adjustierte R^2 und der residuelle mittlere quadratische Fehler: Mit zunehmender Variablenzahl k wächst das Bestimmtheitsmaß R^2 , bis der Wert 1 bei $k = n$ erreicht wird. Um diesen Effekt auszugleichen, kann man einen Strafterm einführen, der verhindert, dass eine zu große Zahl an Regressorvariablen in das Modell aufgenommen wird. Man erreicht dies durch Einführen des *residuellen mittleren quadratischen Fehlers*

$$MS_E = \frac{SS_E}{n - k - 1}.$$

Da SS_E mit wachsendem k fällt, hat die Funktion MS_E in Abhängigkeit von k ein Minimum bei moderatem k . Das könnte als Stoppkriterium dienen: keine Aufnahme weiterer Variablen ins Modell. Das *adjustierte Bestimmtheitsmaß* kann aus MS_E berechnet werden, hat aber an der Stelle ein Maximum, wo MS_E minimal ist:

$$R_{\text{adj}}^2 = 1 - \frac{SS_E / (n - k - 1)}{S_{yy} / (n - 1)} = 1 - \frac{n - 1}{n - k - 1} (1 - R^2).$$

Kategorielle Variable in der linearen Regression: Kategorielle Variable, also solche, die endlich viele Werte annehmen, können in einem linearen Regressionsmodell mit Hilfe so genannter *Dummy-Variablen* berücksichtigt werden.

Beispiel: Gehen wir von einem Regressionsmodell $y = \beta_0 + \beta_1 x_1$ aus mit einer zusätzlichen kategoriellen Variable x_2 , die die Werte 0 oder 1 annehmen kann (z. B. Szenario 0, Szenario 1). Dann setzt man das Regressionsmodell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

an und macht eine gewöhnliche multivariate Regression mit diesem Modell. Dies entspricht den beiden Modellen

$$y = \beta_0 + \beta_1 x_1 \quad \text{bzw.} \quad y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1$$

je nachdem, ob $x_2 = 0$ oder $x_2 = 1$ ist, lässt also für die beiden Szenarien unterschiedliche Achsenabschnitte und Anstiege zu. Im Falle dreier Kategorien braucht man zwei Dummies, etc.

11.4 Statistik der Regression

Stufe 1 – nichtparametrische Annahmen: Es wird angenommen, dass die Fehler ε_i , $i = 1, \dots, n$, unabhängig und identisch verteilt sind, mit Mittelwert Null:

$$E(\varepsilon_i) = 0, \quad V(\varepsilon_i) = \sigma^2.$$

Mit einigen kurzen Rechnungen erhält man daraus für den Erwartungswert und die Kovarianzmatrix

$$E(\hat{\beta}) = \beta, \quad C(\hat{\beta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}.$$

Insbesondere ist $\hat{\beta}$ ein erwartungstreuer Schätzer für β . Für die einzelnen Varianzen erhält man

$$V(\hat{\beta}_j) = \sigma^2 \Gamma_{jj}, \quad \text{COV}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 \Gamma_{ij}$$

mit $\Gamma = (\mathbf{X}^T \mathbf{X})^{-1}$. Weiters kann gezeigt werden, dass $E(SS_E) = (n - k - 1)\sigma^2$ gilt. Daher ist

$$\hat{\sigma}^2 = \frac{1}{n - k - 1} SS_E$$

ist erwartungstreuer Schätzer für σ^2 .

Stufe 2 – Normalverteilungshypothese: Es wird angenommen, dass die Fehler ε_i (unabhängig und) normalverteilt sind nach $\mathcal{N}(0, \sigma^2)$. Bei festem Datenpunkt (x_{i1}, \dots, x_{ik}) ist der zugehörige Funktionswert

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i$$

ebenfalls normalverteilt, mit Mittelwert $\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ und Varianz σ^2 . Weiter sind alle Schätzwerte sowie alle Residuen

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad \hat{\mathbf{y}} = \mathbf{X} \hat{\beta}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$$

Linearkombinationen normalverteilter Größen, also selbst normalverteilt.

Wir werden verschiedene statistische Tests zusammenbauen; dazu müssen die Verteilungen von S_{yy} , SS_E und SS_R ermittelt werden. Es gilt:

S_{yy}/σ^2 hat eine χ^2 -Verteilung mit $(n - 1)$ Freiheitsgraden.

SS_E/σ^2 hat eine χ^2 -Verteilung mit $(n - k - 1)$ Freiheitsgraden.

Wegen $S_{yy} = SS_R + SS_E$ hat SS_R gerade k Freiheitsgrade.

Test auf Signifikanz der Regression: Die Regression wird als *signifikant* erachtet, falls wenigstens einer der Koeffizienten β_1, \dots, β_n verschieden von Null ist. Wir formulieren das Gegenteil als Nullhypothese \mathcal{H}_0 :

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$\mathcal{H}_1 : \beta_j \neq 0 \text{ für mindestens ein } j \in \{1, \dots, k\}$$

Wir wissen aus Stufe 1 und 2, dass

$$E(\hat{\beta}) = \beta, \quad E(\hat{y}) = E(\mathbf{X}\hat{\beta}) = \mathbf{X}\beta = E(\mathbf{y})$$

gilt. Falls die Nullhypothese zutrifft, dann ist

$$\beta = \begin{bmatrix} \beta_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{X}\beta = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix},$$

somit $E(\hat{y}_i) = E(y_i) = E(\bar{y})$ und schließlich $E(\hat{y}_i - \bar{y}) = 0$ für alle i . Es folgt, dass SS_R/σ^2 eine χ^2 -Verteilung mit k Freiheitsgraden besitzt.

Andererseits wissen wir bereits, dass SS_E/σ^2 eine χ^2 -Verteilung mit $(n - k - 1)$ Freiheitsgraden hat.

Daher hat

$$F_0 = \frac{SS_R/k}{SS_E/(n - k - 1)}$$

eine F -Verteilung mit $(k, n - k - 1)$ Freiheitsgraden. Der zugehörige P -Wert

$$P_0 = P(F \geq F_0)$$

ist die Wahrscheinlichkeit, dass das Ergebnis der F -Statistik größer oder gleich dem berechneten Wert F_0 ist, falls die Nullhypothese zutrifft. Üblicherweise verwirft man die Nullhypothese und erachtet die Regression als signifikant, falls $P_0 \leq 0.05$ ist.

Test auf Signifikanz einzelner Koeffizienten: Der Test sucht zu überprüfen, ob ein gegebenes β_j verschieden von Null ist, $j = 0, 1, \dots, k$. Wieder wird das Gegenteil als Nullhypothese formuliert:

$$\mathcal{H}_0 : \beta_j = 0$$

$$\mathcal{H}_1 : \beta_j \neq 0$$

Wir wissen bereits, dass $\hat{\beta}_j$ eine Normalverteilung mit Varianz $V(\hat{\beta}_j) = \sigma^2 \Gamma_{jj}$ besitzt. Falls die Nullhypothese zutrifft, dann muss $E(\hat{\beta}_j) = \beta_j = 0$ sein. Dann hat die Größe $\hat{\beta}_j / \sqrt{\sigma^2 \Gamma_{jj}}$ eine Standardnormalverteilung $\mathcal{N}(0, 1)$.

Andererseits wissen wir, dass $SS_E / \sigma^2 = \hat{\sigma}^2 (n - k - 1) / \sigma^2$ eine χ^2 -Verteilung mit $(n - k - 1)$ Freiheitsgraden besitzt. Es folgt, dass

$$T_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 \Gamma_{jj}}} = \frac{\hat{\beta}_j / \sqrt{\sigma^2 \Gamma_{jj}}}{\sqrt{\frac{\hat{\sigma}^2 (n - k - 1)}{\sigma^2} / (n - k - 1)}}$$

eine t -Verteilung mit $(n - k - 1)$ Freiheitsgraden hat. Man macht üblicherweise einen zweiseitigen Test, daher ist der P -Wert gleich $P_0 = P(|T| \geq |T_0|)$. Die Nullhypothese wird verworfen, falls $P_0 \leq 0.025$ ist, und β_j wird als signifikant verschieden von Null erachtet.

Im Fall $P_0 > 0.025$ ist die Annahme $\beta_j = 0$ mit den Daten verträglich und die Variable x_j kann nicht aus dem Modell entfernt werden.

Ein weiterer wichtiger Schritt bei der Erstellung eines Regressionsmodells ist die *Überprüfung der Modellannahmen*. Wir gehen hier nur kurz auf die Frage des Verteilungstyps ein.

Annahme der Gaußverteilung: Falls die Normalverteilungshypothese zutrifft, sind die Fehler ε_i sowie die Residuen $e_i = y_i - \hat{y}_i$ normalverteilt. Dies kann leicht mit Hilfe eines QQ-Plots überprüft werden. Falls die e_i sich als nicht normalverteilt herausstellen, ist eine mögliche Abhilfe eine Transformation der unabhängigen Variable y .

Sehr häufig wird dabei die Box-Cox-Transformation $y \rightarrow y^{(\lambda)}$ verwendet. Diese hängt vom Parameter λ wie folgt ab:

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda Y^{\lambda-1}}, & \lambda \neq 0 \\ Y \log y, & \lambda = 0 \end{cases} \quad \text{mit } Y = \sqrt[n]{y_1 y_2 \cdots y_n}$$

Der Parameter λ wird so bestimmt, dass im entstehenden Regressionsmodell SS_R minimal wird.

Literatur zur Vorlesung

W. Assenmacher: Deskriptive Statistik. 2. Aufl., Springer, Berlin 1998.

V. Barnett, T. Lewis: Outliers in Statistical Data. Third Ed., John Wiley & Sons, Chichester 1994.

F. Beichelt: Stochastik für Ingenieure. Teubner, Stuttgart 1995.

O. Beyer, H. Hackel, V. Pieper, J. Tiedge: Wahrscheinlichkeitsrechnung und mathematische Statistik. Teubner, Leipzig 1991.

C. Bucher: Computational Analysis of Randomness in Structural Mechanics. CRC Press/Balkema, Leiden 2009.

F. Eisenführ, M. Weber: Rationales Entscheiden. 4. Aufl., Springer-Verlag, Berlin 2003.

Europäisches Komitee für Normung: EN 1990:2002. Eurocode: Grundlagen der Tragwerksplanung. CEN, Brüssel 2002.

W. Fellin, H. Lessmann, M. Oberguggenberger, R. Vieider (Eds.): Analyzing Uncertainty in Civil Engineering. Springer-Verlag, Berlin 2005.

E. J. Gumbel: Statistics of Extremes. Columbia University Press, New York 1958.

O. Klingmüller, U. Bourgund: Sicherheit und Risiko im Konstruktiven Ingenieurbau. Verlag Vieweg, Braunschweig 1992.

E. Kreyszig: Statistische Methoden und ihre Anwendungen. 3. Aufl., Vandenhoeck & Ruprecht, Göttingen 1970.

M. Lefebvre: Applied Probability and Statistics. Springer, New York 2016.

M. A. Meyer, J. M. Booker: Eliciting and Analyzing Expert Judgement. SIAM, Philadelphia PA 2001.

D. C. Montgomery, G. C. Runger: Applied Statistics and Probability for Engineers. Wiley, New York 1994.

D. C. Montgomery, E. A. Peck, G. G. Vining: Introduction to Linear Regression Analysis. Fifth Ed., Wiley-Interscience, Hoboken 2012.

M. Oberguggenberger: Combined methods in nondeterministic mechanics. In: I. Elishakoff, C. Soize (Eds.), Nondeterministic Mechanics. CISM International Centre for Mechanical Sciences, Vol. 539. Springer, Wien 2012, 263 - 356.

M. Oberguggenberger, A. Ostermann: Analysis für Informatiker. 2. Aufl., Springer-Verlag, Berlin 2009.

A. O'Hagan: Eliciting expert beliefs in substantial practical applications. The Statistician **47** (1998), 21-35.

J. Pfanzagl: Allgemeine Methodenlehre der Statistik I, II. 5. Aufl., de Gruyter, Berlin 1978.

E. J. Plate: Statistik und angewandte Wahrscheinlichkeitslehre für Bauingenieure. Ernst & Sohn, Berlin 1993.

L. Sachs: Angewandte Statistik. 4. Aufl., Springer, Berlin 1974.

L. J. Savage: The Foundations of Statistics. Wiley, New York 1954.

G. I. Schuëller: Einführung in die Sicherheit und Zuverlässigkeit von Tragwerken. Ernst & Sohn, Berlin 1981.

R. C. Smith: Uncertainty Quantification. Theory, Implementation, and Applications. SIAM, Philadelphia 2014.

R. Storm: Wahrscheinlichkeitsrechnung, Mathematische Statistik, Statistische Qualitätskontrolle. 10. Aufl., Fachbuchverlag Leipzig – Köln 1995.

D. Stoyan: Stochastik für Ingenieure und Naturwissenschaftler. Akademie Verlag, Berlin 1993.

R. Viertl: Einführung in die Stochastik – mit Elementen der Bayes-Statistik und Ansätzen für die Analyse unscharfer Daten. 3. Aufl., Springer, Wien 2003.

W. Voß u. a.: Taschenbuch der Statistik. Fachbuchverlag Leipzig im Carl Hanser Verlag, München 2000.

Hydrographischer Dienst Tirol:

<http://www.tirol.gv.at/themen/umwelt/wasser/wasserkreislauf/hydro-online/>

Hydrographisches Jahrbuch von Österreich. Abteilung Wasserhaushalt, Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien 1954 - 2014:

https://www.bmlfuw.gv.at/wasser/wasser-oesterreich/wasserkreislauf/hydrographische_daten/jahrbuecher.html