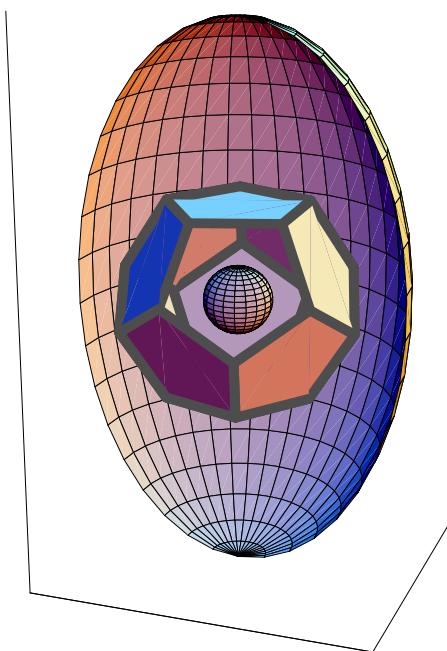


Script ◊ Math ◊ Ing
◊ Wahrscheinlichkeit & Statistik ◊
kurz & bündig ◊



Scripta bilingua

von

Rolf Wirz

Berner Fachhochschule BFH ◊ TI und AHB

V.1.12.29 / df / 20. Dezember 2011 **!Draft! Erweiterte deutsche Version!**

Enthält Teil 6 (Kombinatorik) eines Repetitoriums und Textbuchs zur Begleitung und Ergänzung des Unterrichts.
Produziert mit LaTeX auf NeXT-Computer/ PCTEX WIN98.
Einige Graphiken sind auch mit *Mathematica* entstanden.

1999 hat der Autor einen Computer-Crash erlebt. Infolge des dadurch provozierten Systemwechsels haben einige Graphiken sehr gelitten. Sie werden neu erstellt, sobald die Zeit dafür vorhanden ist.

Glück hilft manchmal, Arbeit immer ...

Brahmanenweisheit

Aktuelle Adresse des Autors (2007):

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

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

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

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

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 1 | Organisatorisches — Quant à l'organisation | 1 |
| 2 | Grundfragen der Statistik — Questions fondam. de la statist. | 3 |
| 2.1 | Einführung — Introduction | 3 |
| 2.1.1 | Stochastik und Statistik — Stochastique et statistique | 3 |
| 2.1.2 | Das Wesen der Statistik — La nature de la statistique | 3 |
| 2.1.3 | Problemkreise — Domaines de problèmes | 4 |
| 2.1.4 | Arten von Massenerscheinungen — Types d'événements de masse | 5 |
| 2.1.5 | Stichproben — Echantillons | 5 |
| 2.2 | Arbeitsweise der math. Statistik — Façon de travailler d. l. stat. math. | 6 |
| 2.3 | Beschreibende Statistik — Statistique descriptive | 7 |
| 2.3.1 | Fragen — Questions | 7 |
| 2.3.2 | Häufigkeitsverteilungen — Répartition de fréquences | 7 |
| 2.3.3 | Häufigkeitsfunktion — Fonction de fréquences | 8 |
| 2.3.4 | Darstellungstechniken — Techniques de représentation | 9 |
| 2.3.5 | Zur Klassenbildung — Quant à la formation ce classes | 13 |
| 2.4 | Masszahlen einer Stichprobe — Mesures d'un échantillon | 14 |
| 2.4.1 | Mittelwert und empirische Varianz — Moyenne et variance empirique | 14 |
| 2.4.2 | Vereinfachungsmeth. bei Berechnungen — Méthodes de simplif. aux calculs | 16 |
| 2.4.3 | Berechnungen, Häufigkeitsfunktion — Calculs, fonction de fréquence | 17 |
| 2.4.4 | Häufigkeitsverteilung, Massenverteilung — Distr. de fréq. et de masse | 18 |
| 2.4.5 | Beispiel mit Mathematica — Exemple avec Mathematica | 18 |
| 2.5 | Auswertung: Beispiel — Exploitation: Exemple | 20 |
| 2.5.1 | Dateneingabe — Entrée de données | 20 |
| 2.5.2 | Kenngrößen — Caractéristiques | 20 |
| 2.5.3 | Darstellung mittels Kenngrößen — Représentat. par caract.: BoxWhiskerPlot | 21 |
| 2.5.4 | Andere statistische Plots — D'autres représentation statistiques | 22 |
| 2.6 | Weitere Kenngrößen — D'autres caractéristiques | 23 |
| 2.6.1 | Diverse Mittelwerte einer Verteilung — Certaines valeurs moyen. d'une distr. | 23 |
| 2.6.2 | Momente einer Verteilung — Moments d'une distribution | 25 |
| 2.6.3 | Die Schiefe einer Verteilung — Le biais d'une distribution | 26 |
| 2.6.4 | Kurtosis und Exzess — Kurtosis et excès | 26 |
| 2.6.5 | Sinn und Gefahr von Kenngrößen — Caractéristiques: Sens propre et danger | 27 |
| 3 | Kombinatorik — Analyse combinatoire | 29 |
| 3.1 | Einleitung — Introduction | 29 |
| 3.1.1 | Problemstellung — Problème | 29 |
| 3.1.2 | Fakultäten — Factorielles | 29 |
| 3.2 | Anordnungsprobleme — Problèmes d'arrangement | 30 |
| 3.2.1 | Permutationen ohne Wiederholung — Permutations sans répétition | 30 |
| 3.2.2 | Permutationen mit Wiederholung — Permutations avec répétition | 33 |

| | | |
|----------|---|-----------|
| 3.3 | Auswahlprobleme — Problèmes de choix | 35 |
| 3.3.1 | Die Fragestellungen — Les questions | 35 |
| 3.3.2 | Variation ohne Wiederholung — Arrangement sans répétition | 36 |
| 3.3.3 | Kombination ohne Wiederholung — Combinaison sans répétition | 38 |
| 3.3.4 | Variation mit Wiederholung — Arrangement avec répétition | 39 |
| 3.3.5 | Kombination mit Wiederholung — Combinaison avec répétition | 40 |
| 3.4 | Übungen — Exercices | 42 |
| 4 | Wahrscheinlichkeitsrechnung — Calcul des probabilités | 43 |
| 4.1 | Einführung — Introduction | 43 |
| 4.1.1 | Problemstellung — Problème | 43 |
| 4.1.2 | Anwendung — Application | 43 |
| 4.1.3 | Personen — Personnages | 44 |
| 4.2 | Zufallsexperiment, Ereignis — Expérience de hasard, événement | 44 |
| 4.2.1 | Zufallsprozesse, Häufigkeit — Processus aléatoire, fréquence | 46 |
| 4.3 | Ereignisalgebra — Algèbre des événements | 47 |
| 4.3.1 | Ereignisalgebra und Mengenalgebra — Alg. des événem. et alg. d. ensembles | 47 |
| 4.3.2 | Boolsche Algebren — Algèbres de Boole | 49 |
| 4.3.3 | Zur Mächtigkeit und Häufigkeit — Quant à la puissance et la fréquence | 50 |
| 4.3.4 | Ereignisbäume — Des arbres d'événements | 50 |
| 4.4 | Klass. Wahrscheinlichkeit n. Laplace — Probabilité class. d'a. Lapl. | 51 |
| 4.4.1 | Laplace-Experiment, Gleichwahrsch. — Exp. de Laplace, probab. identique | 51 |
| 4.4.2 | Wahrsch. als Mass für Gewinnchancen — Prob. comme mesure p. gagner | 52 |
| 4.5 | Axiomatischer Wahrsch'begriff — Probabilité axiomatique | 53 |
| 4.5.1 | Begriff, Axiome, Folgerungen — Notion, axiomes, conclusions | 53 |
| 4.5.2 | Der Begriff Wahrscheinlichkeitsraum — La notion espace de probabilité | 55 |
| 4.5.3 | Bedingte Wahrscheinlichkeit — Probabilité conditionnelle | 56 |
| 4.5.4 | Totale Wahrscheinlichkeit — Probabilité totale | 59 |
| 4.6 | Wahrscheinlichkeitsverteilungen — Fonctions de répartition | 61 |
| 4.6.1 | Zufallsvariablen — Variables aléatoires | 61 |
| 4.6.2 | Verteilungsfunktion — Fonction de répartition | 63 |
| 4.6.3 | Verteilungstypen — Types de répartitions | 64 |
| 4.6.4 | Diskrete Verteilung — Répartition discrète | 64 |
| 4.6.5 | Kontinuierliche Verteilung — Répartition continue | 68 |
| 4.7 | Mass- oder Kennzahlen einer Verteilung — Mesures de répartition | 70 |
| 4.7.1 | Allgemeines — Considérations générales | 70 |
| 4.7.2 | Mittelwert — Valeur moyenne | 70 |
| 4.7.3 | Erwartungswert — Valeur d'espérance | 71 |
| 4.7.4 | Symmetrische Verteilung — Distribution symétrique | 72 |
| 4.7.5 | Varianz, Standardabweichung — Variance, écart-type | 73 |
| 4.7.6 | Momente einer Verteilung — Moments d'une distribution | 74 |
| 4.7.7 | Schiefe einer Verteilung — Dissymétrie d'une distribution | 76 |
| 4.7.8 | Weitere Kenngrössen — Autres caractéristiques | 77 |
| 4.7.9 | Momentenerzeugende Funktion — Fonction char. génér. de moments | 78 |
| 4.7.10 | Laplace- und z-Transformation — Transf. de Laplace et en z | 79 |
| 4.8 | Spezielle diskrete Verteilungen — Distributions discrètes spéciales | 79 |
| 4.8.1 | Bernoulliverteilung — Distribution de Bernoulli | 79 |
| 4.8.2 | Gesetze für die Binomialverteilung — Lois pour la distribution de Bernoulli | 81 |
| 4.8.3 | Poissonverteilung — Distribution de Poisson | 82 |
| 4.8.4 | Pascalverteilung — Distribution de Pascal | 83 |
| 4.8.5 | Geometrische Verteilung — Distribution géométrique | 84 |
| 4.8.6 | Hypergeometrische Verteilung — Distribution hypergéométrique | 85 |
| 4.8.7 | Beispiel — Exemple | 87 |

| | | |
|--------|--|-----|
| 4.9 | Spezielle stetige Verteilungen — Distributions continues spéciales | 87 |
| 4.9.1 | Allgemeines — Généralités | 87 |
| 4.9.2 | Rechtecksverteilung — Distribution rectangulaire | 87 |
| 4.9.3 | Normalverteilung — Distribution normale ou de Gauss | 88 |
| 4.9.4 | Grenzwertsätze von Moivre Laplace — Théorèmes limites de Moivre Laplace | 90 |
| 4.9.5 | Lokaler Grenzwertsatz — Théorème limite locale | 90 |
| 4.9.6 | Grenzwertsatz von De Moivre/ Laplace — Théorème limite de De Moivre/ Laplace | 92 |
| 4.9.7 | Das Gesetz der grossen Zahlen von Bernoulli — La loi des grands nombres | 93 |
| 4.9.8 | Bemerkung zum Zufall — Remarques quant au hasard | 94 |
| 4.9.9 | Tschebyscheffsche Ungleichung — Inéquation de Tschebyscheff | 94 |
| 4.9.10 | Logarithmische Normalverteilung — Distribution normale logarithmique | 94 |
| 4.9.11 | Exponentialverteilung — Distribution exponentielle | 95 |
| 4.9.12 | Weibullverteilung — Distribution Weibull | 96 |
| 4.9.13 | Gammaverteilung — Distribution gamma | 97 |
| 4.9.14 | Ausblick — Autres distributions | 97 |
| 4.10 | Zufallsvektoren — Vecteurs aléatoires | 97 |
| 4.10.1 | Fragestellung, Begriffe — Question, notions | 97 |
| 4.10.2 | Der diskrete Fall — Le cas discret | 100 |
| 4.10.3 | Der stetige Fall — Le cas continu | 101 |
| 4.11 | Mehrdimensionale Erwartung — Espérance multidimensionnelle | 102 |
| 4.11.1 | Erwartung, Mittelwert — Espérance, moyenne | 102 |
| 4.11.2 | Varianz, Kovarianz, Korrelation — Variance, covariance, corrélation | 103 |
| 4.11.3 | Der diskrete Fall — Le cas discret | 106 |
| 4.11.4 | Der stetige Fall — Le cas continu | 106 |
| 4.12 | Mehrdimensionale Verteilungen — Répartitions multidimensionnelles | 107 |
| 4.12.1 | Zweidimensionale Normalverteilung — Distribution normale bidimensionnelle | 107 |
| 4.12.2 | Stichprobenfunkt., Testverteilungen — Fonct. d'échant., distrib. de test | 108 |
| 4.12.3 | Chi-Quadrat-Verteilung — Distribution du Khi-deux | 111 |
| 4.12.4 | Sätze zur Chi-Quadrat-Verteilung — Théorèmes sur la distribution du Khi-deux | 113 |
| 4.12.5 | t-Verteilung von Student — Distribution de Student | 113 |
| 4.12.6 | F-Verteilung von Fisher — Distribution de Fisher | 114 |
| 4.13 | Anhang I: Einige Beweise — Annexe I: Certaines preuves | 115 |
| 4.13.1 | Formel zur Gammafunktion — Formule pour la fonction gamma | 115 |
| 4.13.2 | Dichte der Chi-Quadrat-Verteilung — Densité de la distribution Khi-deux | 116 |
| 4.13.3 | Dichte der Student-Verteilung — Densité de la distribution de Student | 117 |
| 4.13.4 | Beweis Tschebyscheffsche Ungleichung — Preuve d'inéquation de Tschebyscheff | 118 |
| 4.14 | Anhang II: Ergänzungen — Annexe II: Suppléments | 119 |
| 4.14.1 | Quadr'summe contra Betrags. — Somme d. carrés com. à la somme d. val. abs. | 119 |
| 4.14.2 | Verteilungstreue u.s.w. — Conformité du type de répartition etc. | 120 |
| 4.14.3 | Zentraler Grenzwertsatz — Théorème limite central | 122 |
| 4.14.4 | Lineare Transformationen — Transformations linéaires | 123 |
| 5 | Math. Statistik — Statist. mathématique | 125 |
| 5.1 | Qualitätskontrolle — Contrôle de qualité | 125 |
| 5.1.1 | Allgemeines, SQC — Généralités, SQC | 125 |
| 5.2 | SQC1: Prozesskontrolle — SQC1: Contrôle de processus | 125 |
| 5.2.1 | Problemstellung — Problème | 125 |
| 5.2.2 | Beispiel Mittelwert — Exemple moyenne | 126 |
| 5.2.3 | Überw'techn. m. Kontr'karten — Techn. de surv. à l'aide d. cartes de contr. | 128 |
| 5.3 | SQC2: Annahmekontrolle — SQC2: Contrôle d'acceptation | 128 |
| 5.3.1 | Hypergeom. Verteil., Urnenmodell — Répart. hypergéom., modèle d'urne | 129 |
| 5.3.2 | Annahmevertrag, Prüfplan — Contrat d'accept., plan d'échant. | 129 |
| 5.3.3 | Binomialmodell, Urnenmodell — Modèle binomial, modèle d'urne | 131 |

| | | |
|----------|--|------------|
| 5.3.4 | Produzenten- u. Konsumentenrisiko — Risque du producteur e.d. consommateur | 132 |
| 5.4 | Poisson-Prozess, Warteschlange — Files d'attente, processus de Poisson | 134 |
| 5.4.1 | Poisson-Prozess — Processus de Poisson | 134 |
| 5.4.2 | Warteschlangenmodelle — Modèles de files d'attente | 137 |
| 5.5 | Schätzungen — Estimations | 138 |
| 5.5.1 | Schätzer, Punktschätzer — Estimations ponctuelles | 138 |
| 5.5.2 | Erwartungstreue — Propriété d'être sans biais | 138 |
| 5.5.3 | Konsistenz — Consistance | 140 |
| 5.5.4 | Vertrauensintervalle I — Intervalles de confiance I | 142 |
| 5.5.5 | Wichtige Testfunktionen — Fonctions de test importantes | 144 |
| 5.5.6 | Vertrauensintervalle II — Intervalles de confiance II | 149 |
| 5.5.7 | Vertrauensintervalle III — Intervalles de confiance III | 152 |
| 5.5.8 | Vertrauensintervalle IV — Intervalles de confiance IV | 155 |
| 5.5.9 | Automat. Bestimm. der Vertrauensber. — Calcul automat. des domaines de conf. | 155 |
| 5.6 | Signifikanztests — Tests de signification | 156 |
| 5.6.1 | Hypothesen — Hypothèses | 156 |
| 5.6.2 | Zweiseitige Alternative, t -Test — Alternative bilatérale, test de Student | 156 |
| 5.6.3 | Einseitige Alternative, t -Test — Alternative unilatérale, test de Student | 158 |
| 5.6.4 | Testrisiken — Risques (aléas) aux tests | 160 |
| 5.6.5 | Chi-quadrat-Test für die Varianz — Test khi deux pour la variance | 161 |
| 5.6.6 | Automatisches testen — Tester automatiquement | 161 |
| 5.6.7 | Testrezept im Überblick — Vue d'ensemble d'une recette de test | 163 |
| 5.6.8 | Vergleichstest für Mittelwerte I — Test de compar. de moyennes I | 164 |
| 5.6.9 | Vergleichstest für Mittelwerte II — Test de compar. de moyennes II | 165 |
| 5.7 | Weitere Testverfahren — Autres méthodes de test | 167 |
| 5.7.1 | Eine Typeneinteilung von Tests — Une classification des tests | 167 |
| 5.7.2 | Parameterfreier Test: Vorz'test — Test libre de param.: Test du signe | 167 |
| 5.7.3 | Chi-Quadrat-Anpassungstest — Test d'adaptation de Khi deux | 168 |
| 5.7.4 | Zum Fishertest — Quant au test de Fisher | 169 |
| 5.7.5 | Kontingenztafeln — Tableaux de contingence | 170 |
| 6 | Fehlerech., Regr., Korr. — Calc. de l'err., régr., corr. | 173 |
| 6.1 | Fehlerrechnung (Abhängigkeit) — Calcul de l'erreur (dépendance) | 173 |
| 6.1.1 | Das Problem der Verpfanzung — Le problème de la dépendance | 173 |
| 6.1.2 | Verwendung des totalen Differentials — Appliquer la différentielle totale | 174 |
| 6.2 | Regression — Régression | 175 |
| 6.2.1 | Der Begriff — La notion | 175 |
| 6.2.2 | Methode der kleinsten Quadrate — Méthode des carrés minimaux | 176 |
| 6.2.3 | Korrelation — Corrélation | 179 |
| 6.2.4 | Korrelationskoeff.: Bedeutung — Coeff. de corrélation: Signification | 181 |
| 6.2.5 | Rechnen mit Punktschätzern: Probleme — Calculer avec des estimateurs: Probl. | 182 |
| 6.3 | Zum Schluss — Quant à la fin | 183 |
| A | Aus dem DIYMU: Link | 187 |
| B | Fehler von statistischen Kenngrössen und Standardfehler: Link | 189 |
| C | Monte-Carlo, Resampling und anderes: Link | 191 |
| D | Eine Bootstrap-Anwendung Schritt für Schritt: Link | 193 |
| E | Datensatzänderung: Link | 195 |
| F | Spezielle Wahrscheinlichkeitssituationen: Link | 197 |

| | |
|---|------------|
| G Hinweise zur Datenanalyse: Link | 199 |
| H Crashkurs Kombinatorik, Wahrscheinlichkeit: Link | 201 |

Vorwort zum Teil Kombinatorik

Probleme mit ganzen Zahlen

Liebe Leserin, lieber Leser,

Das Thema *Kombinatorik* ist ein klassischer Bestandteil des Mittelschullehrplans. Auch an Berufsmittelschulen sollte es eigentlich behandelt werden. Doch was, wenn ein Student aus irgendwelchen Gründen gerade diesem Stoff an der Schule nie begegnet ist — oder ihn vergessen hat? Dann heisst es eben nacharbeiten und repetieren. Daher ist dieser Text als *Repetitorium* und als *Ausbau* gedacht.

Die Wichtigkeit der Kombinatorik für den Weg durch die weitere Mathematik ist unbestritten. Sie ist ein Werkzeug zur Lösung von Problemen, die manchmal unverhofft an einem herantreten. Geradezu grundlegend ist das Thema aber für das Wissensgebiet „Wahrscheinlichkeitsrechnung und Statistik“.

Dieser Text ist in Skriptform abgefasst. Das bedeutet, dass er in äusserst knapper Fassung nur das wesentliche Skelett des zu lernenden Stoffes wiedergibt. Für weitere, ausführliche Erklärungen, Beispiele, exakte Beweise und ergänzende Ausführungen ergeht daher an den Studenten der Rat, ein oder mehrere Lehrbücher beizuziehen. Studieren bedeutet zu einem wesentlichen Teil, sein Wissen selbstständig mit Hilfe der Literatur zu erweitern, streckenweise sogar selbstständig zu erarbeiten, zu festigen und anzuwenden. Ein Skript ist dabei nur ein Wegweiser und nie ein Lehrbuchersatz. Welche Lehrbücher jemand verwenden will, ist jedem freigestellt. Das Thema Kombinatorik findet man in praktisch allen Unterrichtswerken für die klassische Gymnasialstufe. Beziiglich der Fachhochschulliteratur sei auf das Beispiel Bibl. A5 (Brenner, Lesky, Mathematik für Ingenieure und Naturwissenschaftler, Band 1) verwiesen.

Im Sommer 1996

Der Autor

Kapitel 1

Organisatorisches

Kurze Übersicht

1. Organisation, Rahmen
2. Stoff
3. Ziel, Weg, Methoden, Feedback, Team
4. Übungen, Selbststudium
5. Lerntechnik, Arbeitstechnik, Selfmanagement
6. Rechte und Pflichten des Studenten und der Schule
7. Prinzipien, Grundsätze
8. Rechner, Computer, Mathematiksoftware
9. Semesterorganisation Mathematik (Anzahl Noten, Prüfungsreglement, Prüfungsplan, Prüfungsrahmen, erlaubte Unterlagen, formale Anforderungen, Benotungskriterien, Benotung der Übungen und Projekte, Arbeitsnachweismappe, Klassensprecher, Klassenbetreuer, Kopierchef, Sprechstunden)
10. Hilfsmittel (Bibliothek, Taschenrechner, Mathematiksoftware, Literatur)
11. Zeitplanung
12. Einführung

Kapitel 2

Grundfragen der Statistik

2.1 Einführung

2.1.1 Stochastik

Stochastik gilt als Oberbegriff. Dieser Name ist vom altgriechischen Wort „stochastike techne“ abgeleitet. Lateinisch bedeutet dies „ars coniectandi“, das heisst „Ratekunst“, Kunst des Vermutens. Die Stochastik ist eigentlich ein Teilgebiet der Mathematik, obwohl man sie heute in eigens für dieses Gebiet eingerichteten Studiengängen lehrt. Man kann also an einigen Universitäten als „Statistiker“ abschliessen, ist dann in Wirklichkeit aber „Stochastiker“. Unter dem Oberbegriff „Stochastik“ werden die Gebiete „Wahrscheinlichkeitstheorie“ und „Statistik“ zusammengefasst.

Unter **Statistik** dagegen versteht man heute die Zusammenfassung bestimmter Methoden zur Analyse von empirischen Daten. Manchmal werden aber die Begriffe Stochastik und Statistik nicht so genau getrennt, da der Begriff „Stochastik“ nicht jedermann bekannt und daher wenig werbewirksam ist. Viele Autoren betiteln ihre Werke mit „Statistik“, meinen damit aber „Stochastik“. Dieser Begriffsgebrauch wird dann in den Werken oft so fortgesetzt.

Das Wort Statistik leitet sich vermutlich vom lateinischen „statisticum“ ab und bedeutet „den Staat betreffend“. Andererseits gibt es in der italienischen Sprache das Wort „statista“, womit Staatsmann resp. Politiker gemeint ist. Im 18. Jahrhundert kam im deutschen Sprachbereich das Wort „Statistik“ in Gebrauch. Man bezeichnete damit die Lehre von den Daten über den Staat (Staatstheorie). Die Wortbedeutung von Statistik als sammeln und auswerten von Daten stammt vermutlich erst aus dem 19. Jahrhundert.

2.1.2 Das Wesen der Statistik

Die grosse Bedeutung der Statistik liegt in ihrer Natur: Die Statistik dient zur Beurteilung von **Massenerscheinungen**. Massenerscheinungen entsprechen unserem Zeitgeist. Neue Methoden machen die Statistik sehr leistungsfähig. Alte empirische Verfahren der Handwerker genügen nicht mehr.

Man unterscheidet aktuell zwischen folgenden Arten von Statistik:

1. **Beschreibende, deskriptive oder empirische Statistik**
2. **Schliessende, induktive oder mathematische Statistik**, auch **Interferenzstatistik**
3. **Hypothesen generierende, explorative oder datenschürfende Statistik**, auch **Datenschürfung** oder **Data mining**

Die deskriptive oder beschreibende Statistik ist mathematisch harmlos. Es geht hier um die Darstellung oder Präsentation von statistischem Material.

Bei der mathematischen Statistik (beurteilende, induktive Statistik) will man zu affirmativen Resultaten gelangen. Es geht da um Schätzverfahren und Testverfahren für Hypothesen. Die Resultate sind Wahrscheinlichkeitsaussagen. Hier fliesst also die Wahrscheinlichkeitsrechnung ein. Gesucht werden Zusammenhänge oder Unterschiede in Datenbeständen. Auch sucht man das Mass an Sicherheit von Vermutungen, hypothetischen Ergebnissen oder Modellen. Als **statistisch gesichert** erachtet man ein Resultat erst dann, wenn im Rahmen von vorausschauenden, prospektiven Versuchsplanungen Hypothesen mit einer sehr grossen Wahrscheinlichkeit bestätigt werden können. Zu diskutieren geben hier allemal die Form der Fragestellung in Bezug auf Alternativen sowie das akzeptierte Wahrscheinlichkeitsmass. Die explorative Statistik (hypothesen-generierende Statistik oder Datenschürfung) ist von den Methoden her eine Zwischenform der beiden vorher genannten Statistikarten. Wegen dem zunehmenden Anwendungsbedarf und ihrer damit wachsenden Bedeutung erleben wir hier zur Zeit (Jahrhundertwende) die Entwicklung einer eigenständigen Anwendungsform.

Empirische oder beschreibende Statistik kennen wir schon vom alten Ägypten her (Zwiebelstatistik beim Pyramidenbau ...). Heute geht es aber sehr oft darum, aus einer durch Zahlen beschriebenen Situation durch eine „mathematische Auswertung“ zu einer Beurteilung zu kommen, die mit andern Beurteilungen vergleichbar ist. Die Methoden zur Gewinnung und Auswertung von statistischem Material liefert die **mathematische Statistik**.

Von Statistiken fordert man folgende fünf Eigenschaften, was leicht nachvollziehbar ist:

1. **Objektivität** (Unabhängigkeit vom Standpunkt des Erstellers der Statistik)
2. **Reliabilität** (Verlässlichkeit)
3. **Validität** (Über den Kontext hinausgehende Gültigkeit)
4. **Signifikanz** (Das Mass an Bedeutung oder des Zutreffens, die Wesentlichkeit)
5. **Relevanz** (Wichtigkeit)

2.1.3 Problemkreise der mathematischen Statistik

Bezüglich des Vorgehens bei der mathematischen Statistik können wir folgende Problemkreise unterscheiden:

1. Art und Methoden der Datensammlung
2. Gliederung, Zusammenfassung
3. Analyse: Auswertung, Schlüsse, Hypothesenprüfung
4. Bewertung der Zuverlässigkeit
5. Beurteilung
6. Aufzeigung statistischer Gesetzmässigkeiten

Anwendung: Überall, wo Massenerscheinungen eine Rolle spielen. Z.B. Reparaturstatistik, Verkaufsstatistik, Lager- oder Materialstatistik, Unfallstatistik, Geburten-, Sterbe-, Bevölkerungstatistik, Gesundheitsstatistik, Klimastatistik, Statistiken in der industriellen Fertigung zwecks Qualitätskontrolle u.s.w..

Der Gewinn aus der Statistik ist die Erarbeitung strategischer Zahlen als Grundlage strategischer Entscheide. Die Absicht ist oft die Risikoverminderung, die Kostensenkung, die Gewinnerhöhung oder die Vermeidung von Verlusten u.s.w..

2.1.4 Arten von Massenerscheinungen

Man kann zwischen **zwei grundsätzlich verschiedenen Arten von Massenerscheinungen** unterscheiden:

1. Grosse Zahl von **Individuen** (Zählstatistiken). Das Zahlenmaterial besteht aus Anzahlen ($\in \mathbb{N}$).
2. Grosse Zahlen von **Einzelerscheinungen** oder meist physikalischen Messungen. Das Zahlenmaterial besteht aus Messwerten, d.h. meist reelle Zahlen ($\in \mathbb{R}$).

Beim beobachteten Material (Individuen, Einzelerscheinungen) kann es sich allgemeiner um örtliche, zeitliche oder sonst irgendwie geartete unabhängige Dinge handeln, welche wir hier **Beobachtungseinheiten** oder **Merkmale** nennen wollen, die sich auf verschiedene Arten manifestieren oder verschiedene **Ausprägungen** zeigen können. Wir wollen folgende Ausprägungsarten klassifizieren:

1. **Qualitative** Merkmale, d.h. nicht quantifizierbare Merkmale.

Beispiel: Etwas wird als „schön“, „weniger schön“, „unschön“ oder „scheusslich“ qualifiziert. Eine Übertragung auf eine Skala ist hier immer problematisch, da es im vorliegenden Fall um menschliche Gefühle geht, welche nicht gut gegeneinander abgrenzbar und bei verschiedenen Einschätzern nicht unbedingt vergleichbar sind. Dieses Problem taucht oft bei einer Notengebung auf. Mit den hier vorliegenden Ausprägungen zu „rechnen“ scheint unsinnig. Man kann sich ja fragen, ob schön plus scheusslich dann noch schön, weniger schön, unschön oder gar schon scheusslich sein soll. Dasselbe gilt für eine etwaige Multiplikation statt der Addition u.s.w..

2. **Quantitative** Merkmale. Hier kann man zwei Gruppen unterscheiden:

- (a) **Diskret** quantifizierbare Merkmale, z.B. Anzahlen von Individuen. Man denke etwa an die Anzahl Menschen in einem Lift. Hier werden Nachkommastellen absurd, wenn es sich nicht um Kenngrößen wie etwa einen Mittelwert handelt. Wir werden kaum eine Liftkabine finden, in der sich z.B. „4.852 Menschen“ befinden...
- (b) **Stetig** quantifizierbare Merkmale, z.B. Messresultate, die an einer Skala abgelesen werden sind. So habe man etwa die Länge eines Stabes gemessen. Dieser misst 3.952 m , wobei die letzte Stelle ungenau ist ($\pm 0.001\text{ m}$). Dabei ist hier zu beachten, dass bei Naturbeobachtungen keine unendlich grosse Genauigkeit möglich ist. „Stetig“ ist also hier in einem weniger exakten, d.h. nicht mathematisch exakten Sinn zu verstehen.

Oft zwingt die Komplexität des Materialumfangs zu einer vorgängigen **Klasseneinteilung** der vorliegenden Zahlen und daher zu einer ersatzweisen Klassendarstellung oder Auswertung.

Im Gegensatz zu den Massenerscheinungen treten bei Individuen oder Einzelerscheinungen zufällige Unregelmässigkeiten auf. „Statistische Gesetzmässigkeiten“ fehlen.

2.1.5 Stichproben

Häufig handelt es sich bei den zu untersuchenden Massenerscheinungen um Erscheinungen, die sich aus einer sehr grossen Zahl von Einzelerscheinungen zusammensetzen, die man oft unmöglich alle erfassen kann. Daher ist man darauf angewiesen, die Zahl der betrachteten Einzelerscheinungen zu reduzieren

und auf eine repräsentative Art und Weise eine Auswahl zu treffen. Eine solche repräsentative Auswahl nennen wir **Stichprobe** (Extrakt). Ihr Umfang heisst **Stichprobenumfang**.

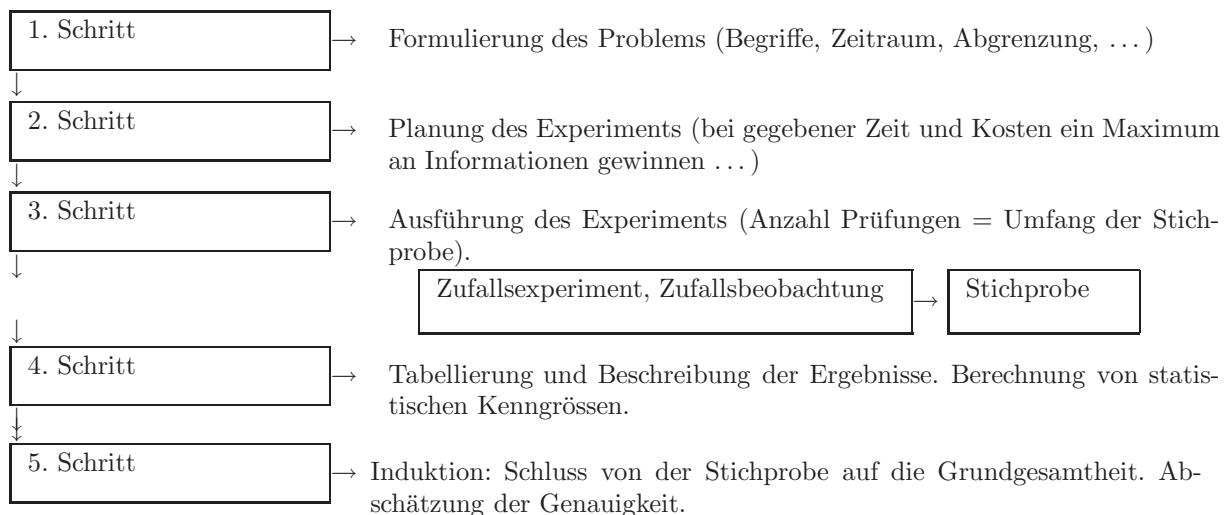
Die erhobenen **Stichprobenwerte** werden in der Regel in einer **Urliste** erfasst (Messprotokoll, Fragebogen u.s.w.). Sie stammen aus einer **Grundgesamtheit** von Ausprägungen, die normalerweise wegen den real vorhandenen Erfassungseinträgen als bekannt oder als im Rahmen der Erfassung bekannt und gegebenenfalls als abgesteckt erweiterbar vorausgesetzt werden kann. So könnte man z.B. bei einer Erfassung einer Biodiversität eine neue Pflanzenart entdecken, die bisher noch in keiner Liste bekannt war. Die Grundgesamtheit wird dadurch jetzt beschränkt erweitert.

Urlisten kann man natürlich auch gleich ordnen. Dadurch entstehen **geordnete Urlisten**, z.B. eine **Rangliste**. Bei der Erstellung von Ranglisten geht aber schon Information verloren, da dabei die Reihenfolge der Datenerhebung oder der Messungen verändert wird. Diese Reihenfolge kann aber einen systematischen Einfluss auf die Daten haben. Daher liegt das Kapital in der Urliste. Man sollte diese unter allen Umständen als wichtigste Quelle immer aufbewahren, solange an der statistischen Aufbereitung der Daten noch zu arbeiten ist.

Eine **Stichprobe** besteht aus **Beobachtungseinheiten** (z.B. Individuen), welche **Merkmale** tragen (Variablen). Diese nehmen in der Regel bei jedem Individuum eine persönliche **Ausprägung** (Wert) an. Die gesammelte Wertemenge nennen wir **Stichprobendaten**.

2.2 Arbeitsweise der mathematischen Statistik

In der Praxis sind die statistischen Fragestellungen vielfältig. Bearbeitungsschritte und Methoden sind jedoch vielfach gleich:



Wichtig:

1. Die Schlüsse von der Stichprobe auf die Grundgesamtheit fussen auf der Mathematik der Wahrscheinlichkeitsrechnung. Vollkommen sichere Schlüsse dieser Art existieren nicht.
2. Schlüsse auf die Grundgesamtheit sind nur gültig, wenn wirklich ein **Zufallsexperiment** vorliegt (vgl. 4.2).
3. Ein Experiment kann zur Zerstörung des Artikels führen. Damit wird der Artikel der Zweckbestimmung entzogen.

2.3 Beschreibende Statistik

2.3.1 Fragen

Stichproben sind in der Regel nur dann sinnvoll, wenn sie aus Zahlen bestehen oder durch Zahlen dargestellt werden können.

Problem:

1. Wie können die gesammelten Daten sinnvoll dargestellt werden?
2. Wie kann man die Stichproben durch wenige Zahlen kennzeichnen um diese dann so vergleichen zu können? (\leadsto Mittelwert, Varianz ...)

Wichtige Begriffe sind **absolute** und **relative Häufigkeit**, **Häufigkeitsfunktion** $\tilde{f}(x)$, **Summenhäufigkeits-** oder **Verteilungsfunktion** $\tilde{F}(x)$ sowie **Kenngrößen** wie etwa ein Mittelwert. Kenngrößen dienen der „Vermessung“ von Daten. Das Ziel ist ihre Darstellung durch wenige Zahlen z.B. zur einfachen Vergleichbarkeit verschiedener Datensätze.

Bemerkung:

$\tilde{f} \leadsto$ engl. distribution function
 $\tilde{F} \leadsto$ engl. cumulative distribution function on the sample

2.3.2 Häufigkeitsverteilungen

Die Ergebnisse „Zählung“ oder „Messung“ müssen irgendwie festgehalten oder protokolliert werden. So entsteht eine **Urliste** oder das **Protokoll**. Misst man z.B. an jedem Individuum je zwei Eigenschaften, so hat man schliesslich n Zahlenpaare und daher den Stichprobenumfang n .

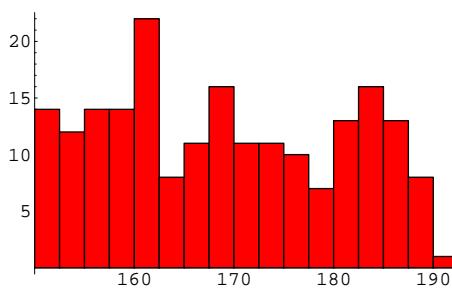
Bsp.:

Urliste: Grösse von 200 Pilotenschülern, Messung. Damit die Daten besser darstellbar sind, hat man die Messungen gleich in 40 Klassen (von ca. 150 bis ca. 190) eingeteilt.

| | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 162. | 174.4 | 182.8 | 153.2 | 175.2 | 162.4 | 187.6 | 153.6 | 184.8 | 185.6 |
| 162.4 | 178.4 | 173.6 | 178.8 | 167.2 | 180. | 162. | 155.6 | 181.6 | 186.8 |
| 183.2 | 155.2 | 150.8 | 187.6 | 171.2 | 170.8 | 157.6 | 184.4 | 186. | 158.4 |
| 160.4 | 180.4 | 151.2 | 163.2 | 188. | 152. | 167.6 | 174. | 170.8 | 162. |
| 155.6 | 168.4 | 179.2 | 165.2 | 162.4 | 163.2 | 178.4 | 167.6 | 180.8 | 182. |
| 170.8 | 173.6 | 184.8 | 173.6 | 160.4 | 182.8 | 183.6 | 160.8 | 162.4 | 180.8 |
| 166. | 176.8 | 181.6 | 168.8 | 160. | 158.4 | 152.4 | 153.6 | 188. | 185.2 |
| 164. | 175.6 | 157.2 | 153.2 | 183.2 | 152.4 | 162.4 | 169.6 | 173.2 | 159.2 |
| 168.8 | 158.8 | 160.8 | 168.4 | 153.2 | 172.4 | 168.8 | 190. | 182.8 | 164. |
| 166.4 | 186.4 | 184.8 | 168.8 | 152.4 | 160.4 | 178. | 165.6 | 158.8 | 158.4 |
| 165.6 | 186. | 176. | 188.8 | 186.8 | 177.2 | 165.2 | 170.4 | 183.6 | 154.8 |
| 186.4 | 170.4 | 150.8 | 180.8 | 170. | 174. | 156. | 162.4 | 167.6 | 163.6 |
| 168. | 186.8 | 158.8 | 155.2 | 152.4 | 150.8 | 172.8 | 156.4 | 155.6 | 163.6 |
| 157.6 | 176. | 162.4 | 158.8 | 161.2 | 155.6 | 161.6 | 167.6 | 181.2 | 171.6 |
| 155.6 | 155.6 | 163.6 | 158. | 177.6 | 158.4 | 154.8 | 152.4 | 175.2 | 157.6 |
| 172. | 186. | 169.6 | 184. | 164.4 | 160. | 157.2 | 175.2 | 153.2 | 154.4 |
| 185.6 | 157.6 | 162. | 172.8 | 180. | 152. | 188.4 | 165.2 | 152.4 | 183.6 |
| 183.6 | 162.4 | 167.2 | 175.6 | 161.6 | 166.4 | 187.6 | 181.6 | 187.2 | 156.4 |
| 180.4 | 156.4 | 183.6 | 152. | 184.4 | 189.2 | 171.6 | 168.8 | 154.4 | 187.2 |
| 172.8 | 154. | 152. | 153.6 | 179.2 | 181.6 | 174.8 | 168. | 167.6 | 165.2 |

Darstellung der Stichprobe: \rightsquigarrow Der Grösse nach geordnete Werte: **Frequenztabelle** (manchmal auch **Strichliste** wie beim Jassen). Die angegebenen Anzahlen heissen **absolute Häufigkeiten**. Z.B. hat der Wert 155.6 die absolute Häufigkeit 6.

| | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|
| (150.8, 3) | (155.6, 6) | (160.4, 3) | (165.2, 4) | (170., 1) | (174., 2) | (178.4, 2) | (183.2, 2) |
| (151.2, 1) | (156., 1) | (160.8, 2) | (165.6, 2) | (170.4, 2) | (174.4, 1) | (178.8, 1) | (183.6, 5) |
| (152., 4) | (156.4, 3) | (161.2, 1) | (166., 1) | (170.8, 3) | (174.8, 1) | (179.2, 2) | (184., 1) |
| (152.4, 6) | (157.2, 2) | (161.6, 2) | (166.4, 2) | (171.2, 1) | (175.2, 3) | (180., 2) | (184.4, 2) |
| (153.2, 4) | (157.6, 4) | (162., 4) | (167.2, 2) | (171.6, 2) | (175.6, 2) | (180.4, 2) | (184.8, 3) |
| (153.6, 3) | (158., 1) | (162.4, 8) | (167.6, 5) | (172., 1) | (176., 2) | (180.8, 3) | (185.2, 1) |
| (154., 1) | (158.4, 4) | (163.2, 2) | (168., 2) | (172.4, 1) | (176.8, 1) | (181.2, 1) | (185.6, 2) |
| (154.4, 2) | (158.8, 4) | (163.6, 3) | (168.4, 2) | (172.8, 3) | (177.2, 2) | (181.6, 4) | (186., 3) |
| (154.8, 2) | (159.2, 1) | (164., 2) | (168.8, 5) | (173.2, 1) | (177.6, 1) | (182., 1) | (186.4, 2) |
| (155.2, 2) | (160., 2) | (164.4, 1) | (169.6, 2) | (173.6, 3) | (178., 1) | (182.8, 3) | (186.8, 3) |



Statt eine Frequenz- oder Strichliste zu machen, kann man die Daten auch gleich in ein Histogramm oder in ein Punktdiagramm eintragen. \rightsquigarrow Punkt pro Messung, Punkte statt Balken.

Wenn man die Häufigkeit als Bruchteil des ganzen Stichprobenumfangs bestimmt, spricht man von der **relativen Häufigkeit**. Durch diese Grösse lassen sich Stichproben mit verschiedenem Umfang vergleichen.

Definition:

Sei $H(a)$ die absolute Häufigkeit der Beobachtung A mit dem Wert $x = a$ und n der Stichprobenumfang.

\rightsquigarrow **relative Häufigkeit**

$$h(a) = \frac{H(a)}{n}$$

Es gilt selbstverständlich:

Satz:

$$0 \leq h(a) \leq 1$$

2.3.3 Häufigkeitsfunktion

Geg.:

m verschiedene Werte $\{x_1, x_2, x_3, \dots, x_m\}$ mit den relativen Häufigkeiten $h(x_1), h(x_2), \dots, h(x_m)$, kurz h_1, h_2, \dots, h_m . Die Werte fallen also in m Klassen $\{[x]_1, \dots, [x]_m\}$. Der Stichprobenumfang sei n .

Definition:

Häufigkeitsfunktion:

$$\tilde{f}(x) = \begin{cases} h_i & x = x_i \\ 0 & x \neq x_i \end{cases}$$

Definition:**Verteilungsfunktion:** (Diskrete Verteilung)

$$\tilde{F}(x) = \sum_{x_i \leq x} f(x)$$

Die Häufigkeitsfunktion der Stichprobe zeigt, wie die Häufigkeiten verteilt sind. Sie bestimmen also die **Häufigkeitsverteilung**. Folgende Beziehungen sind unmittelbar einsichtig:

Satz:

1. $\sum_{i=1}^m H(x_i) = \sum_{i=1}^m H_i \leq \sum_{i=1}^n H(x_i) = n$
2. $\sum_{i=1}^m h(x_i) = \sum_{i=1}^m h_i = \sum_{i=1}^m \tilde{f}(x) = \tilde{F}(x_m) \leq 1 = \tilde{F}(x_n)$

2.3.4 Darstellungstechniken

Beispiel einer Datenmenge

Geg.:

Resultate einer Zählung von defekten Komponenten an 40 Apparaten eines gegebenen Typs, die zur Reparatur eingeschickt worden sind. Notiert sind die Anzahlen x_i der Komponenten.

0, 0, 0, 1, 1, 2, 1, 5, 5, 6,
 3, 4, 2, 6, 8, 9, 7, 3, 4, 1,
 2, 3, 5, 7, 5, 3, 2, 4, 6, 8,
 9, 7, 5, 6, 4, 2, 2, 1, ·

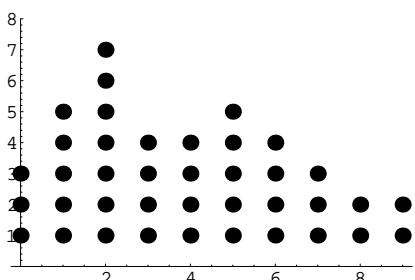
Bemerkung:

Der Punkt (·) bedeutet **missing value**: Der Wert wurde nicht notiert. Damit wird hier $n = 39$.

~> Frequenztabelle:

| x_i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------|---|---|---|---|---|---|---|---|---|---|----|
| H_i | 3 | 5 | 7 | 4 | 4 | 5 | 4 | 3 | 2 | 2 | — |

Punktdiagramm

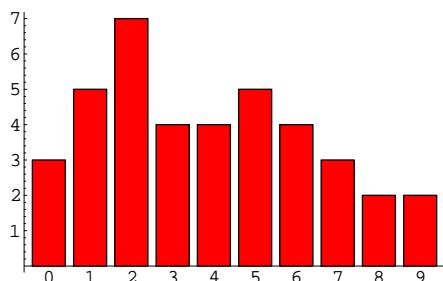


Punktdiagramm (statt Strichliste)

Bemerkung:

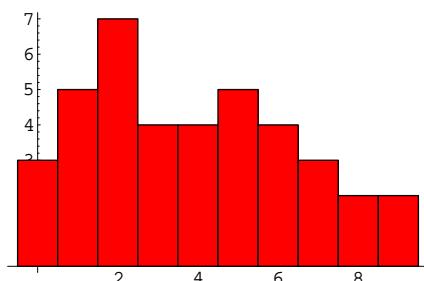
Statt mit Strichlisten kann man auch mit **Stamm-Blatt-Diagrammen** arbeiten. Dabei werden z. B. bei Messungen die vorkommenden Ziffern ohne die letzte in eine Rangliste (Spalte) eingetragen. Die letzten Ziffern werden statt einem Strich wie bei einer Strichliste dahinter in die zugehörige Zeile geschrieben. Das macht Sinn, wenn die Messungen sich häufig nur in der letzten Ziffer unterscheiden.

Stabdiagramm



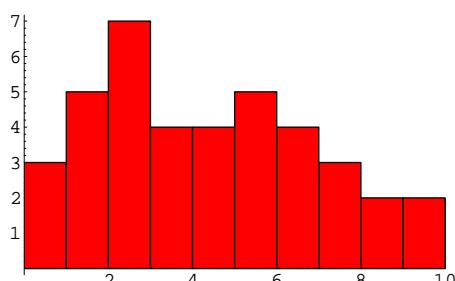
Stabdiagramm oder Balkendiagramm

Generalisiertes Stabdiagramm

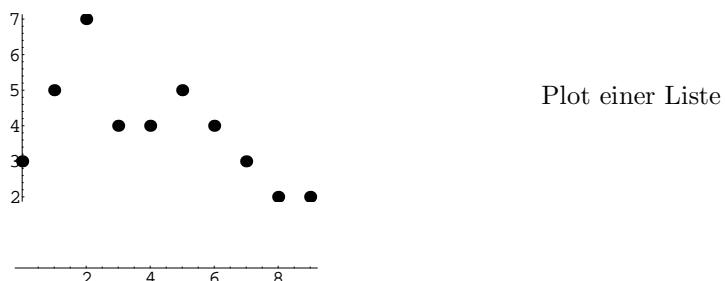
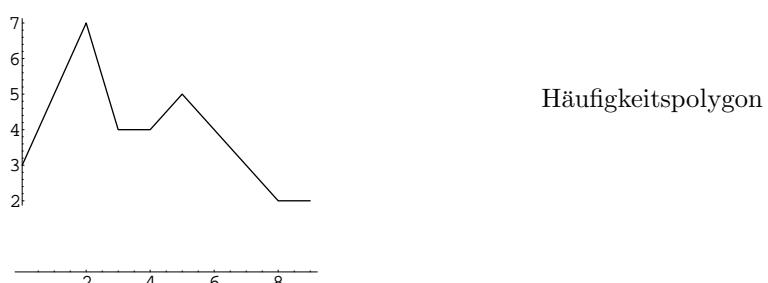
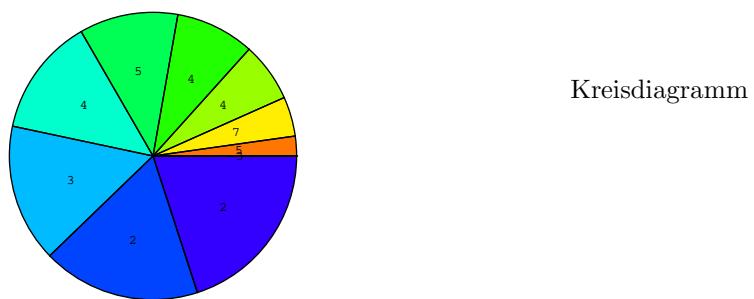


Generalisiertes Stabdiagramm (hier Balkenbreite grösser)

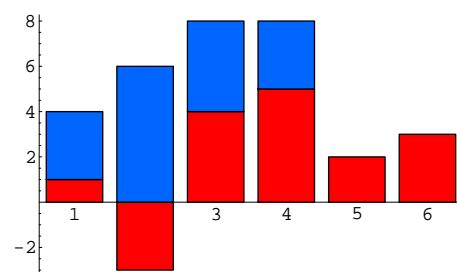
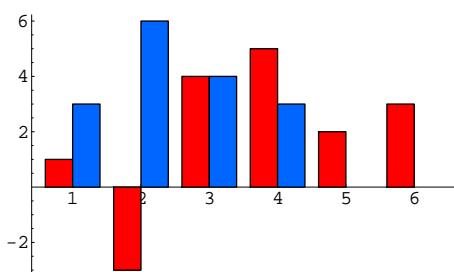
Histogramm

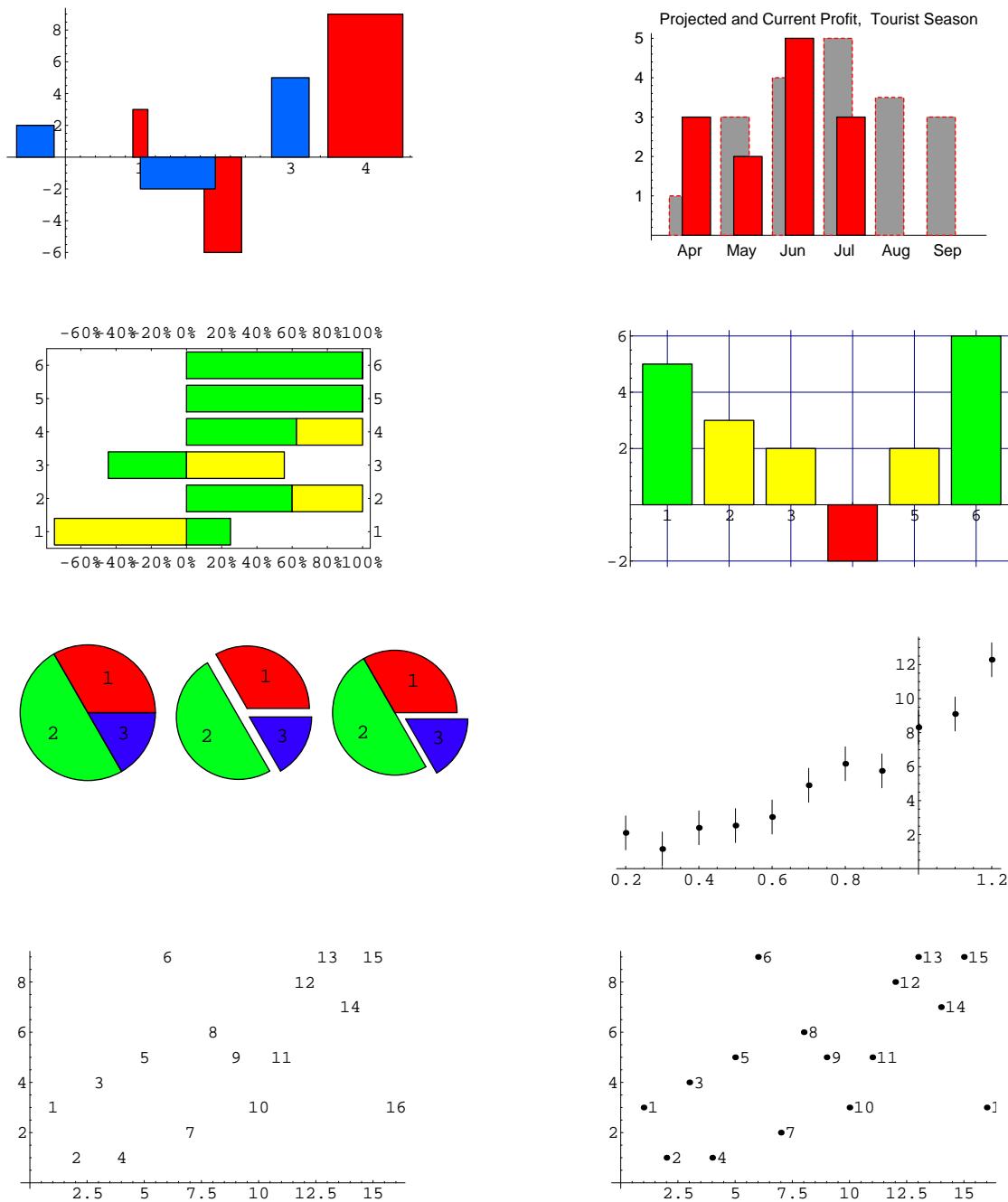


Histogramm oder Staffelbild (Treppenfunktion), in der Regel bei einer Zusammenfassung der Daten in Klassen

ListPlot**Häufigkeitspolygon****Kreisdiagramm****Gegliederte Diagramme**

Den folgenden Beispielen liegen Datenmengen zugrunde, die von obigen Daten abweichen. Die Natur der Sache macht dies notwendig. Die Beispiele sind selbsterklärend.

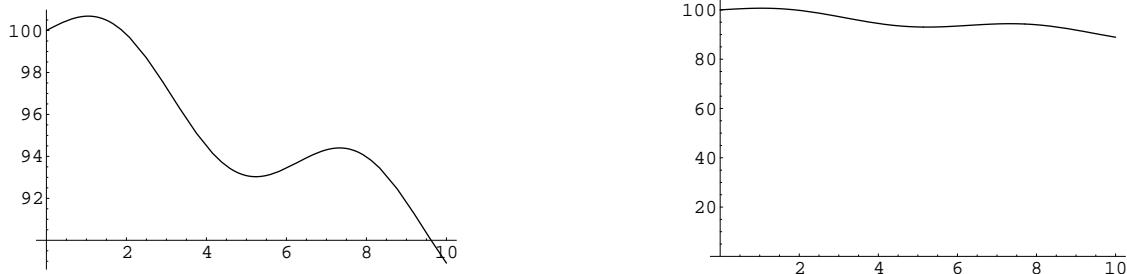




In technischen und ökonomischen Publikationen findet man eine Vielfalt weiterer Darstellungsarten. Erwähnenswert sind **Sankey-Diagramme**, die an Flussdiagramme erinnern und **Körperschaubilder** oder **3D-Darstellungen**.

Achtung Schwindel

Folgendes Beispiel spricht für sich:



2.3.5 Zur Klassenbildung

Bei grösseren Datenmengen wird eine Tabelle oder Graphik rasch unübersichtlich. Durch Gruppierung der Daten in Klassen (**Klassenbildung**) kann man die Übersicht wieder gewinnen und die Sache vereinfachen. Der gemachte Fehler ist höchstens von der Grösse der halben Klassenbreite.

Dabei wird das Intervall, in dem alle Stichprobenwerte liegen, in **Klassenintervalle** (Teilintervalle) unterteilt. Die Mitten dieser Intervalle heissen **Klassenmitten**. Die in einem Teilintervall liegenden Werte der Stichprobe ist eine Teilstichprobe und heisst **Klasse von Werten**. Die Anzahl der Werte, die in einer Klasse liegen, nennen wir die **zur Klasse gehörige absolute Häufigkeit** resp. **Klassenhäufigkeit**. Dazu gehört:

Definition:

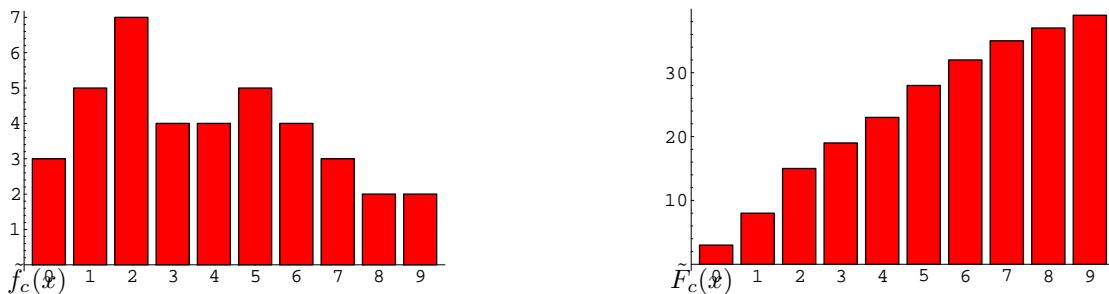
$$\text{Relative Klassenhäufigkeit} := \frac{\text{absolute Klassenhäufigkeit}}{\text{Stichprobenumfang}}$$

Bemerkung:

Wir nehmen jetzt an, dass alle Werte einer Klasse in der Klassenmitte liegen. (Diese Annahme wird meistens so gemacht.) Das erzeugt einen Fehler, der aber höchstens so gross wie die halbe Klassenbreite ist.

Die Häufigkeit in Abhängigkeit von der Klassenmitte wird beschrieben durch die Häufigkeitsfunktion $\tilde{f}_c(x)$ der nun neu in Klassen eingeteilten Stichprobe. Dazu gehört die Summenhäufigkeitsfunktion oder Verteilungsfunktion $\tilde{F}_c(x)$. $\tilde{F}_c(x)$ ist eine Treppenfunktion, die monoton wächst.

Beispiel: Die folgenden Diagramme zeigen $n \tilde{f}_c(x)$ und $n \tilde{F}_c(x)$ dargestellt durch Stabdiagramme. Diese Darstellung ist zwar nicht ganz korrekt, dafür heute aber oft leicht auf dem Computer erzeugbar.



Es ist unmittelbar klar:

Satz:

$$x_1 \leq x_2 \Rightarrow 0 \leq \tilde{F}_c(x_1) \leq \tilde{F}_c(x_2) \leq 1$$

Bemerkung:

1. $\tilde{f}_c(x)$ und $\tilde{F}_c(x)$ sind konstant außer für $x = \text{Klassengrenze}$.
2. $\tilde{f}_c(x)$ und $\tilde{F}_c(x)$ springen an den Klassengrenzen. Die Sprunghöhe von $\tilde{F}_c(x)$ ist gleich $\tilde{f}_c(x)$.

2.4 Masszahlen einer Stichprobe

Eine Stichprobe kann einerseits durch die Häufigkeitsfunktion oder durch die Verteilungsfunktion beschrieben werden. Andererseits kann man sie auch durch gewisse Masszahlen charakterisieren. Diese Masszahlen sind dann mit den Masszahlen anderer Stichproben zahlenmäßig vergleichbar. Das ist ein Vorteil. Der Nachteil ist, dass bei den Masszahlen viel Information nicht mehr direkt sichtbar ist, welche ursprünglich noch da war. Wichtige Masszahlen sind der Mittelwert und die Varianz.

Solche Masszahlen teilt man in verschiedene Kategorien ein. So ist der Mittelwert oder auch das Minimum und das Maximum einer Stichprobe ein **Lagemass**. Die Streuung dagegen ist ein **Streumass** u.s.w.

2.4.1 Mittelwert und empirische Varianz

Mittelwert

Der Mittelwert zeigt die durchschnittliche Grösse der Stichprobenwerte oder Klassen an. Er ist das arithmetische Mittel der Stichprobenwerte oder Klassenwerte:

Definition:

Mittelwert:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Problem: Die Stichproben 1, 2, 3, 4, 5 und 2.7, 3.0, 3.1, 3.2 haben denselben Mittelwert, obwohl sie wesentlich verschieden sind. Gesucht ist also eine Masszahl, die die Abweichungen der Stichprobenwerte vom Mittelwert angibt.

Der Mittelwert ist nicht **robust** (wenig empfindlich) gegen **Ausreisser** (z.B. sonderbare, unerklärlich grosse oder kleine Stichprobenwerte). Ein Ausreisser kann den Mittelwert sehr verfälschen.

Statt dem gewöhnlichen arithmetischen Mittel sind auch andere Mittelwerte in Gebrauch. Solche anderen Mittelwerte gibt es viele: Z.B. das **geometrische**, das **harmonische** oder das **gewichtete arithmetische Mittel** (entsprechend einer gemittelten Momentensumme). Dazu konsultiere man die Literatur.

Empirische Varianz

Eine Möglichkeit wäre es, die Summe der Differenzen $\sum(x_i - \bar{x})$ zu nehmen. Das funktioniert aber nicht, wie man sofort nachrechnet:

$$\sum_{i=1}^n (x_i - \bar{x}) = \left(\sum_{i=1}^n x_i \right) - (n \cdot \bar{x}) = n \cdot \frac{\sum_{i=1}^n x_i}{n} - n \cdot \bar{x} = n \cdot \bar{x} - n \cdot \bar{x} = 0$$

$x_i - \bar{x}$ kann negativ oder positiv oder auch null werden. In der Summe heben sich die Werte dann auf. Das könnte man umgehen, indem man die Beträge $|x_i - \bar{x}|$ nimmt. Mathematisch einfacher und

gewinnbringender ist es aber, statt dessen die Quadrate zu nehmen. In Anlehnung an den Mittelwert der so denkbaren quadratischen Abweichung der Stichprobenwerte x_i vom Stichprobenmittelwert \bar{x} definieren wir:

Definition: $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$ heisst **Varianz** der Stichprobe $\{x_1, \dots, x_n\}$. $s = \sqrt{s^2}$ heisst **Standardabweichung**.

Varianz und Standardabweichung werden auch als **Streuung** bezeichnet.

Bemerkung:

Interessanterweise steht bei der empirischen Varianz im Nenner $n-1$ und nicht n , wie das bei einem Mittelwert zu erwarten wäre. Einerseits hat das bei grossen n praktisch keinen Einfluss auf den Wert von s^2 . Andererseits aber erreicht man damit, dass s^2 eine „erwartungstreue Schätzung der Streuung σ^2 “ der Grundgesamtheit ist. (Vgl. dazu die Ausführungen in Storm, Bibl. A12.)

Die Varianz und die Standardabweichung sind nicht robust gegen Ausreisser.

1. Beispiel:

$$M_1 = \{1, 2, 3, 4, 5\} \Rightarrow \bar{x} = 3.0, s^2 \approx 2.5, s \approx 1.6$$

2. Beispiel:

$$M_1 = \{2.7, 3.0, 3.1, 3.2\} \Rightarrow \bar{x} = 3.0, s^2 \approx 0.05, s \approx 0.22$$

Definition: $[\bar{x} - s, \bar{x} + s]$ heisst **Standardintervall**

Im letzten Beispiel ist das Standardintervall ca. $[2.78, 3.22]$.

Weitere Masszahlen

Definition: **Spannweite** der Stichprobe := grösster minus kleinster Stichprobenwert.

Median (\tilde{x} , Zentralwert) der Stichprobe := Wert in der Mitte bei einer ungeraden Anzahl von Stichprobenwerten, ansonst arithmetisches Mittel der beiden Werte in der Mitte. (Der Median ist robust gegen Ausreisser!)

Quartile, Quantile: Vgl. Box-Whisker-Plot, Seite 21.

Modus (D, Dichtemittel) der Stichprobe := Wert mit der grössten Häufigkeit (wird manchmal auch lokal genommen).

Schiefe, Kurtosis und diverse Mittelwerte vgl. Seite 23.

2.4.2 Vereinfachungsmethoden bei Berechnungen

Es gilt:

$$\begin{aligned}
 s^2 &= \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) = \frac{1}{n-1} \cdot (\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2) = \\
 &= \frac{1}{n-1} \cdot (\sum_{i=1}^n x_i^2 - 2\bar{x} n \bar{x} - n \bar{x}^2) = \frac{1}{n-1} \cdot ((\sum_{i=1}^n x_i^2) - n \bar{x}^2) = \frac{1}{n-1} \cdot \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}^2 = \\
 &= \frac{1}{n-1} \cdot \sum_{i=1}^n x_i^2 - \frac{1}{n(n-1)} (\sum_{i=1}^n x_i)^2 = \frac{1}{n-1} \cdot (\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2)
 \end{aligned}$$

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

Bei grossen Zahlen x_i ist eine Koordinatentransformation $x_i = c + x_i^*$ vortheilhaft. (c ist so gewählt, dass die x_i^* kleinen Zahlen werden.)

$$\rightsquigarrow x_i^* = c - x_i$$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{n} \cdot \sum_{i=1}^n (c + x_i^*) = \frac{1}{n} \cdot \sum_{i=1}^n x_i^* + \frac{1}{n} \cdot \sum_{i=1}^n c = \frac{1}{n} \cdot \sum_{i=1}^n x_i^* + \frac{1}{n} \cdot \sum_{i=1}^n c = \bar{x}^* + \frac{n}{n} c = \bar{x}^* + c$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n (c + x_i^* - \bar{x}^* - c)^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i^* - \bar{x}^*)^2 = (s^*)^2 \Rightarrow s^2 = (s^*)^2$$

Satz:

Vor.:

$$x_i = c + x_i^*$$

Beh.:

$$1. \bar{x} = \bar{x}^* + c$$

$$2. s^2 = (s^*)^2$$

$$3. s = s^*$$

Manchmal ist sogar eine **lineare Transformation** nützlich:

$$\text{Sei } x_i = c_1 x^* + c_2, \quad x^* = \frac{1}{c_1} x_i - c_2$$

$$\rightsquigarrow \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{n} \cdot \sum_{i=1}^n (c_1 x^* + c_2) = (\frac{c_1}{n} \cdot \sum_{i=1}^n x^*) + \frac{n}{n} \cdot c_2 = c_1 \cdot \bar{x}^* + c_2$$

$$\begin{aligned}
 \rightsquigarrow s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n (c_1 x^* + c_2 - (c_1 \cdot \bar{x}^* + c_2))^2 = \frac{1}{n-1} \sum_{i=1}^n (c_1 x^* - c_1 \cdot \bar{x}^*)^2 = \\
 &= \frac{c_1^2}{n-1} \cdot \sum_{i=1}^n (x^* - \bar{x}^*)^2 = c_1^2 \cdot \frac{1}{n-1} \cdot \sum_{i=1}^n (x^* - \bar{x}^*)^2 = c_1^2 \cdot (s^*)^2
 \end{aligned}$$

Satz:**Vor.:**

$$x_i = c_1 x^* + c_2$$

Beh.:

1. $\bar{x} = c_1 \cdot \bar{x}^* + c_2$
2. $s^2 = c_1^2 \cdot (s^*)^2$

2.4.3 Berechnungen mittels der Häufigkeitsfunktion

Sei $\bar{x} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$, Klassen $\{[x]_1, \dots, [x]_m\}$, $x_i \in [x]_j$, $i \leq j$.

Sei n_k die Häufigkeit in der Klasse k . Für die Klassen gilt dann:

$$\begin{aligned} \overline{[x]} &= \frac{1}{n} \cdot \sum_{k=1}^j n_k \cdot [x]_k, \quad h([x]_k) = \tilde{f}(x_k) = \frac{n_k}{n} \Rightarrow n_k = n \cdot \tilde{f}(x_k) \\ \Rightarrow \overline{[x]} &= \frac{1}{n} \cdot \sum_{k=1}^j n \cdot \tilde{f}(x_k) \cdot [x]_k = \sum_{k=1}^j \tilde{f}(x_k) \cdot [x]_k \end{aligned}$$

Bemerkung:

$$x_i \in [x]_k \Rightarrow x_i \approx x_k \wedge \overline{[x]} \approx \bar{x}$$

Fehler: Max. halbe Klassenbreite.

Satz:

$$\overline{[x]} = \sum_{k=1}^j \tilde{f}(x_k) \cdot [x]_k \approx \bar{x}$$

Es gilt:

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^j (x_i - \bar{x})^2 \approx (\overline{[s]})^2 = \frac{1}{n-1} \sum_{i=k}^j ([x]_k - \overline{[x]})^2 \cdot n_k \\ &\approx \frac{1}{n-1} \sum_{i=k}^j ([x]_k - \overline{[x]})^2 \cdot n \cdot \tilde{f}([x]_k) = \frac{1}{n-1} \sum_{i=k}^j ([x]_k - \overline{[x]})^2 \cdot n \cdot \tilde{f}([x]_k) \end{aligned}$$

Auf Seite 16 haben wir berechnet:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^j (x_i - \bar{x})^2 = \left(\frac{1}{n-1} \cdot \sum_{i=1}^n x_i^2 \right) - \left(\frac{n}{n-1} \cdot \bar{x}^2 \right) = \frac{1}{n-1} \cdot \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot (\sum_{i=1}^n x_i)^2 \right)$$

Analog schliessen wir auch hier:

Satz:

$$\begin{aligned} \text{Sei } s_K^2 &:= (\overline{[s]})^2: s^2 \approx s_K^2 = \frac{1}{n-1} \sum_{i=k}^j ([x]_k - \overline{[x]})^2 \cdot n \cdot \tilde{f}([x]_k) \\ &= \frac{1}{n-1} \cdot \sum_{i=k}^j ([x]_k \cdot n \tilde{f}([x]_k) - \frac{1}{n} \cdot (\sum_{i=k}^j x_k n \tilde{f}([x]_k))^2) \\ &= \frac{1}{n-1} \cdot (\sum_{i=k}^j ([x]_k)^2) \cdot n \cdot \tilde{f}([x]_k) - n \cdot (\overline{[s]})^2 \end{aligned}$$

2.4.4 Häufigkeitsverteilung und Massenverteilung

Idee:

Die Masse 1 sei längs einer Gerade verteilt.

| | | | | |
|-------------|----------------|----------------|-------|----------------|
| Werte: | x_1 | x_2 | | x_m |
| Häufigkeit: | $h_1 = f(x_1)$ | $h_2 = f(x_2)$ | | $h_m = f(x_m)$ |

Analogie:

| | | | | |
|--------|-------|-------|-------|-------|
| Punkt: | x_1 | x_2 | | x_m |
| Masse: | h_1 | h_2 | | h_m |

$$\Rightarrow \tilde{F}(x_1) = \sum_{x_i \leq x} h_i = \sum \text{Massen links von } x$$

Konsequenz:

Der **Schwerpunkt** x_S entspricht dem **Mittelwert** \bar{x} :

$$x_S \hat{=} \bar{x} \cdot 1 = \sum x_i \cdot h_i$$

Das **Trägheitsmoment** Φ entspricht bis auf einen Faktor der **Varianz** s^2 :

$$\Phi \hat{=} \frac{n}{n-1} \cdot s^2$$

2.4.5 Beispiel mit Mathematica

Packages laden — **Mathematica-Programm**:

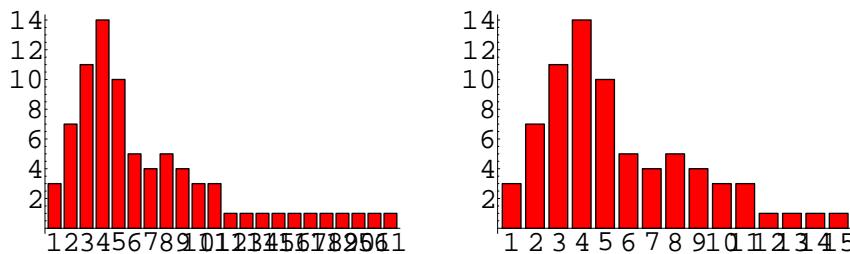
```
<< Statistics`DataManipulation`;
<< Statistics`DescriptiveStatistics`;
<< Graphics`Graphics`
```

Daten — **Mathematica-Programm**:

```
d = {0, 0, 0, 1, 1, 7, 7, 10, 14, 1, 1, 2, 2, 2, 2, 2, 1, 5, 5, 6, 3, 4, 2, 6,
  8, 9, 7, 13, 3, 4, 1, 2, 3, 5, 7, 5, 3, 3, 3, 3, 3, 2, 4, 6, 8, 11,
  12, 4, 3, 3, 3, 3, 4, 4, 4, 4, 4, 9, 7, 5, 6, 4, 2, 8, 2, 2, 1, 8, 9,
  10, 15, 16, 17, 18, 19, 10, 50, 60} + 1;
d1 = {0, 0, 0, 1, 1, 7, 7, 10, 6, 6, 1, 1, 2, 2, 2, 2, 1, 5, 5, 6, 3, 4, 2,
  6, 8, 9, 7, 3, 4, 1, 2, 3, 5, 7, 5, 3, 3, 3, 3, 2, 4, 6, 8, 4, 3,
  3, 3, 3, 4, 4, 4, 4, 9, 7, 5, 6, 4, 2, 8, 2, 2, 1, 8, 9, 10, 10} + 1
```

Manipulation — **Mathematica-Programm**:

```
f = Frequencies[d]; f1 = Frequencies[d1]; p1 = BarChart[f]; p2 = BarChart[f1];
Show[GraphicsArray[{p1, p2}]];
```



Mittelwerte — **Mathematica-Programm:**

```
| {"d", LocationReport[d] // N, "d1", LocationReport[d1] // N}
```

Output:

```
| {"d", {Mean -> 7.6125, HarmonicMean -> 4.13919, Median -> 5.},
  "d1", {Mean -> 5.16901, HarmonicMean -> 3.73698, Median -> 5.} }
```

Bemerkung:

Der Median ist unempfindlich (robust) gegen Ausreißer!

Manipulation — **Mathematica-Programm:**

```
| s = Sort[d];
 {s, "Median->", s[[Length[d]/2]], s[[Length[d]/2 + 1]], "Length->", Length[d]}
```

Output:

```
| { {1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4,
  4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6,
  7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10, 10, 10, 10, 11, 11, 11, 11, 12, 13,
  14, 15, 16, 17, 18, 19, 20, 51, 61},
  Median->, 5, 5, Length->, 80}
```

Streuungen — **Mathematica-Programm:**

```
| {"d", DispersionReport[d] // N, "d1", DispersionReport[d1] // N}
```

Output:

```
| {d, {Variance -> 80.215, StandardDeviation -> 8.95628, SampleRange -> 60.,
  MeanDeviation -> 4.92688, MedianDeviation -> 2., QuartileDeviation -> 3.},
  d1, {Variance -> 6.97103, StandardDeviation -> 2.64027, SampleRange -> 10.,
  MeanDeviation -> 2.15791, MedianDeviation -> 2., QuartileDeviation -> 2.}}
```

2.5 Auswertung: Beispiel

2.5.1 Dateneingabe

Wir verwenden *Mathematica*:

In:

```
Fichte={95.53,81.93,83.57,54.82,73.83,58.48,59.15,83.29};

Buche={113.14,156.61,169.96,142.20,162.85,196.08,125.03,88.47};

allData=Union[Fichte, Buche]
```

Out:

```
{54.82, 58.48, 59.15, 73.83, 81.93, 83.29, 83.57, 88.47, 95.53, 113.14,
125.03, 142.2, 156.61, 162.85, 169.96, 196.08}
```

2.5.2 Kenngrößen

In:

```
<<Statistics`DescriptiveStatistics`;
data=Fichte; {LocationReport[data], DispersionReport[data], ShapeReport[data]}
```

Out:

```
{{Mean -> 73.825, HarmonicMean -> 71.1504, Median -> 77.88},
{Variance -> 219.052, StandardDeviation -> 14.8004, SampleRange -> 40.71,
MeanDeviation -> 12.2563, MedianDeviation -> 11.67, QuartileDeviation -> 12.3075},
{Skewness -> -0.0521399, QuartileSkewness -> -0.549055, KurtosisExcess -> -1.38182}}
```

In:

```
data=Buche; {LocationReport[data], DispersionReport[data], ShapeReport[data]}
```

Out:

```
{{Mean -> 144.293, HarmonicMean -> 136.328, Median -> 149.405},
{Variance -> 1185.56, StandardDeviation -> 34.432, SampleRange -> 107.61,
MeanDeviation -> 27.0825, MedianDeviation -> 22.465, QuartileDeviation -> 23.66},
{Skewness -> -0.176882, QuartileSkewness -> -0.281488, KurtosisExcess -> -0.844303}}
```

In:

```
data=allData; {LocationReport[data], DispersionReport[data], ShapeReport[data]}
```

Out:

```
{{Mean -> 109.059, HarmonicMean -> 93.5017, Median -> 92.},
{Variance -> 1979.66, StandardDeviation -> 44.4934, SampleRange -> 141.26,
MeanDeviation -> 37.8073, MedianDeviation -> 32.94, QuartileDeviation -> 35.7625},
{Skewness -> 0.521187, QuartileSkewness -> 0.605173, KurtosisExcess -> -0.996434}}
```

2.5.3 Darstellung mittels Kenngrössen: BoxWhiskerPlot

Der **Box-Whisker-Plot** ist von unschätzbarem Wert um einen schnellen Überblick über einen numerischen Datensatz zu gewinnen. Er bekommt seine Form durch das Rechteck, das durch die Distanz zwischen zwei Quartilen um den Median herum gegeben wird. Üblicherweise sind es das 25% und das 75% Quartil. Normalerweise schliessen „whiskers“, d.h. Linien wie Backenbärte (Schnurrhaare) die Rechtecke ein, welche die Ausdehnung des Datensatzes entweder ganz oder ohne die Ausreisser markieren. **Ausreisser** sind Punkte jenseits einer Distanz von $3/2$ der zwischenquantilen Spanne weg vom Rechteck. **Grosse Ausreisser** sind Punkte jenseits von 3 mal diese **Spanne**.

Der **Quantilabstand (Quartilsdifferenz)** ist ein Mass der Streuung (Höhe des Rechtecks). Weiter ist der Median in der Box eingezeichnet, welcher durch seine Lage innerhalb der Box einen Eindruck von der Schiefe der den Daten zugrunde liegenden Verteilung vermittelt. Diese Quartile (Quantile) sowie der Median sind robust gegen Ausreisser: Sie ändern nicht enorm, wenn man die Ausreisser weglässt, im Gegensatz zum arithmetischen Mittelwert und damit der Standardabweichung.

Wie schon der **Median** ein **robustes Lagemass** war, so ist somit auch die **Quartilsdifferenz** ein **robustes Streumass gegen Ausreisser**.

Auszug aus dem Help-File von *Mathematica*:

„The box-and-whisker plot is invaluable for gaining a quick overview of the extent of a numeric data set. It takes the form of a box that spans the distance between two quantiles surrounding the median, typically the 25 % quantile to the 75 % quantile. Commonly, „whiskers“, lines that extend to span either the full data set or the data set excluding outliers, are added. Outliers are defined as points beyond $3/2$ the interquartile range from the edge of the box; far outliers are points beyond three times the interquartile range.“

Bemerkung: Bei den **Quartilen** benutzt man die folgende Definition:

Definition: **Quartil** q_j :

Sei $n = 4k + r$, $k, r \in \mathbb{N}$, x_k = Messwert Nummer k in der Rangliste der Messwerte \rightsquigarrow

$$\begin{aligned} r = 0 &\Rightarrow q_{0.25} = 0.25 x_k + 0.75 x_{k+1} \\ r = 1 &\Rightarrow q_{0.25} = x_{k+1} \\ r = 2 &\Rightarrow q_{0.25} = 0.75 x_{k+1} + 0.25 x_{k+2} \\ r = 3 &\Rightarrow q_{0.25} = 0.5 x_{k+1} + 0.5 x_{k+2} \end{aligned}$$

$$r = \dots \Rightarrow q_{0.25} = \tilde{x}$$

$$\begin{aligned} r = 0 &\Rightarrow q_{0.75} = 0.75 x_{3k} + 0.25 x_{3k+1} \\ r = 1 &\Rightarrow q_{0.75} = x_{3k+1} \\ r = 2 &\Rightarrow q_{0.75} = 0.25 x_{3k+1} + 0.75 x_{3k+2} \\ r = 3 &\Rightarrow q_{0.75} = 0.5 x_{3k+2} + 0.5 x_{3k+3} \end{aligned}$$

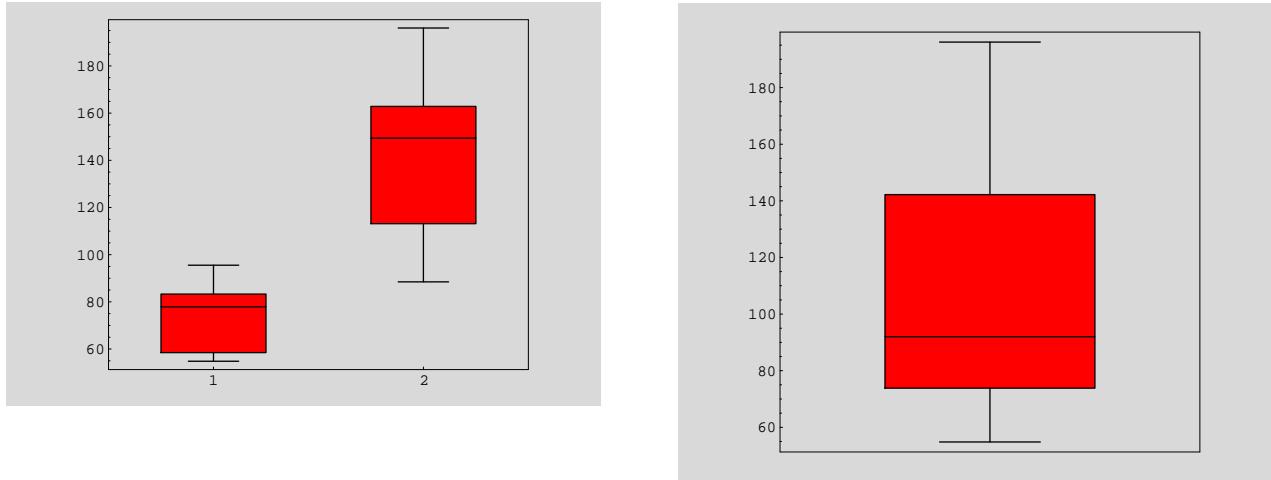
$Q_1 = q_{0.25} = \mathbf{1. Quartil}$, $Q_2 = q_{0.5} = \mathbf{2. Quartil} = \text{Median}$, $Q_3 = q_{0.75} = \mathbf{3. Quartil}$, entsprechend für **Quintile** u.s.w.

Das erste Quartil ist also der Wert bei etwa einem Viertel der Messwerte, das zweite Quartil der Wert bei etwa der Hälfte der Messwerte und das dritte Quartil der Wert bei etwa dreien Vierteln der Messerte.

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

In:

```
<<Statistics`StatisticsPlots`  
BoxWhiskerPlot[Transpose[{Fichte, Buche}]] ;      BoxWhiskerPlot[Transpose[{allData}]] ;
```

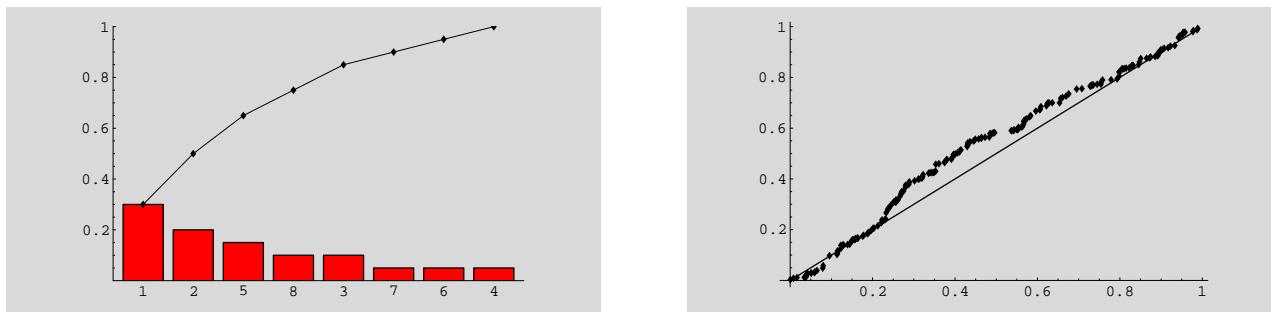


2.5.4 Andere statistische Plots

Andere statistische Plots: Vgl. Erklärungen in *Mathematica*.

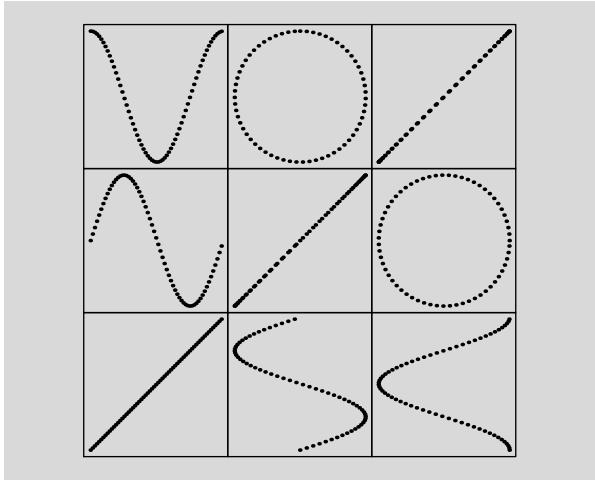
In:

```
data={1,1,1,2,2,3,2,4,1,5,5,6,3,1,1,7,2,8,8,5};  
ParetoPlot[data] ;      QuantilePlot[Table[Random[],{200}],Table[Random[],{200}]] ;
```



In:

```
data = Table[{x, Sin[x], Cos[x]}, {x, 0, 2 Pi, 0.1}] ;      PairwiseScatterPlot[data] ;
```



In:

StemLeafPlot[data]:

Siehe Help in *Mathematica*.

2.6 Weitere Kenngrößen

2.6.1 Diverse Mittelwerte einer Verteilung

Definitionen und Folgerungen:

1. Arithmetisches Mittel:

$$\bar{x}_a = \frac{1}{n} \sum_{j=1}^n x_j = \frac{x_1 + x_2 + \dots + x_n}{n}$$

2. Gewichtetes arithmetisches Mittel, Gewicht w_i zu x_i :

$$\bar{x}_{a, gew} = \frac{1}{\sum_{j=1}^n w_j} \cdot \sum_{j=1}^n w_j \cdot x_j = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n}{\sum_{j=1}^n w_j}$$

3. Geometrisches Mittel:

$$\bar{x}_g = \left(\prod_{j=1}^n x_j \right)^{\frac{1}{n}} = \sqrt[n]{\prod_{j=1}^n x_j} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}, \quad \ln(\bar{x}_g) = \frac{1}{n} \sum_{j=1}^n \ln(x_j) = \overline{\ln(x)}_a$$

4. Gewichtetes geometrisches Mittel:

$$\bar{x}_{g, gew} = \left(\prod_{j=1}^n x_j^{w_j} \right)^{\frac{1}{w}} = \sqrt[w]{\prod_{j=1}^n x_j^{w_j}}, \quad w = \prod_{j=1}^n w_j$$

5. Harmonisches Mittel:

$$\bar{x}_h = \frac{n}{\sum_{j=1}^n \frac{1}{x_j}}, \quad \sum_{j=1}^n \frac{1}{\bar{x}_h} = \frac{n}{\bar{x}_h} = \sum_{j=1}^n \frac{1}{x_j}$$

Für nur zwei Werte x_1 und x_2 gilt:

$$\bar{x}_h = \frac{\bar{x}_g^2}{\bar{x}_a}$$

6. **Gewichtetes harmonisches Mittel:**

$$\bar{x}_{h, gew} = \frac{\sum_{j=1}^n w_j}{\sum_{j=1}^n \frac{w_j}{x_j}}$$

7. **Logarithmisches Mittel von zwei positiven Werten:**

$$\bar{x}_{\ln, x_1, x_2} = \frac{x_1 - x_2}{\ln(x_1) - \ln(x_2)}$$

Dann gilt:

$$\bar{x}_g < \bar{x}_{\ln, x_1, x_2} < \bar{x}_a$$

8. **Potenzmittel mit positiven Werten:**

$$\bar{x}_{p, k} = \sqrt[k]{\frac{1}{n} \sum_{j=1}^n x_j^k} = \left(\frac{1}{n} \sum_{j=1}^n x_j^k \right)^{\frac{1}{k}}$$

Bemerkung:

Für $k = 1$ hat man das arithmetische Mittel, für $k = 2$ das quadratische Mittel \bar{x}_q , für $k = -1$ das harmonische Mittel u.s.w.

Weiter gilt mit dem r -ten statistischen Moment (siehe nächster Abschnitt):

$$m_r = (\bar{x}_{p, r})^r \text{ sowie } u < v \Rightarrow \bar{x}_{p, u} \leq \bar{x}_{p, v}$$

$$x_{min} \leq \bar{x}_h \leq \bar{x}_g \leq \bar{x}_a \leq \bar{x}_q \leq x_{max}$$

9. **Quasi-arithmetisches Mittel:** Sei f streng monoton und stetig, $w_j \in [0, 1]$, $\sum_{j=0}^n w_j = 1$

$$\bar{x}_f = f^{-1} \left(\sum_{j=1}^n w_j \cdot x_j \right)$$

10. **Winsorisierte oder gestutzte Mittelwerte:** Wenn die Daten durch Ausreisser (einige wenige viel zu hohe oder viel zu niedrige Werte) verunreinigt sind und man diese Werte nicht berücksichtigen will, so „winsorisiert“ man. D.h. man sortiert die Beobachtungswerte nach aufsteigendem Rang oder Grösse. Dann streicht man am Anfang und am Ende der so entstehenden Rangliste eine gleiche Anzahl von Werten, die Ausreisser also, weg. Darauf berechnet man von den verbleibenden Werten den Mittelwert.

Beispiel: Ein 20 % winsorizierter Mittelwert wird erzeugt, indem man 10 % der Anzahl aller Werte am unteren Ende und 10 % dieser Werte am oberen Ende der Rangliste streicht.

11. Weitere interessante Mittelwerte, von denen man jedoch einzelne selten in der deskriptiven Statistik trifft, sind das **a -Mittel**, verschiedene Ausprägungen von **gleitenden Mittelwerten** oder das **arithmetisch-geometrische Mittel**. Beziiglich dieser Mittelwerte sei auf die einschlägige Fachliteratur verwiesen, da sie den Rahmen dieses Kurses sprengen.

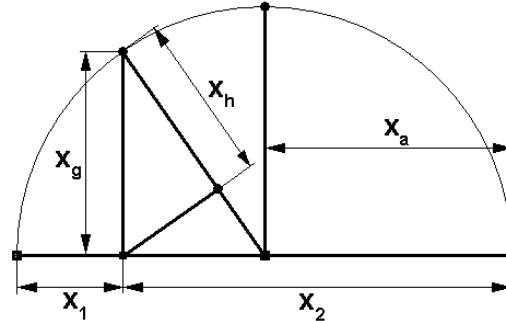
Hinweis:

Mit dem Höhensatz, dem Kathetensatz und

$$\bar{x}_h = \frac{\bar{x}_g^2}{\bar{x}_a^2}$$

sieht man aus nebenstehender Figur:

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x}_a$$



2.6.2 Momente einer Verteilung

Seien x_1, x_2, \dots, x_n Messwerte, die mehrmals vorkommen können. Wir nehmen an, dass, wenn wir nur jeweils ein Exemplar jedes vorkommenden Messwerts auflisten, wir die Liste $\{x_1, x_2, \dots, x_k\}$ bekommen. Seien $f_1 = n_1, f_2 = n_2, \dots, f_k = n_k$ die zur Liste gehörigen Frequenzen oder absolute Häufigkeiten. Dann gilt für die totale Anzahl der Messwerte $n = n_1 + n_2 + \dots + n_k = f_1 + f_2 + \dots + f_k$.

Analog zum **Mittelwert** definieren wir jetzt das **Moment r-ter Ordnung** $m_{r,a}$ der Messreihe in Bezug auf einen Ausgangspunkt oder Ausgangswert a . Wenn $a = \bar{x}$ ist, so schreiben wir kurz m_r und sprechen vom **Moment r-ter Ordnung**, ohne den Ausgangspunkt zu nennen.

Definition:

$$m_{r,a} = \frac{1}{n} \sum_{j=1}^k n_j (x_j - a)^r = \frac{n_1 (x_1 - a)^r + n_2 (x_2 - a)^r + \dots + n_k (x_k - a)^r}{n} = \overline{(x - a)^r}$$

Folgerungen:

Wenn $a = \bar{x}$ gilt, so ergibt sich:

$$m_r = \frac{1}{n} \sum_{j=1}^k n_j (x_j - \bar{x})^r = \frac{n_1 (x_1 - \bar{x})^r + n_2 (x_2 - \bar{x})^r + \dots + n_k (x_k - \bar{x})^r}{n} = \overline{(x - \bar{x})^r}$$

Wenn dann noch $f_1 = n_1 = f_2 = n_2 = \dots = f_k = n_k = 1$ gilt, so ergibt sich:

$$m_r = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^r = \frac{(x_1 - \bar{x})^r + (x_2 - \bar{x})^r + \dots + (x_n - \bar{x})^r}{n} = \overline{(x - \bar{x})^r}$$

Wenn $a = 0$ gilt, so ergibt sich:

$$m_{r,0} = \frac{1}{n} \sum_{j=1}^k n_j x_j^r = \frac{n_1 x_1^r + n_2 x_2^r + \dots + n_k x_k^r}{n} = \overline{x^r}$$

Daher gilt für den Mittelwert \bar{x} und die Standardabweichung s :

$$m_{1,0} = \bar{x}, \quad s^2 = \frac{n}{n-1} \cdot m_{2,0}$$

Weiter gelten die Zusammenhänge:

1. $m_{1,a} = \bar{x} - a$
2. $m_2 = m_{2,a} - m_{1,a}^2$
3. $m_3 = m_{3,a} - 3m_{1,a}m_{2,a} + 2m_{1,a}^3$
4. $m_4 = m_{4,a} - 4m_{1,a}m_{3,a} + 6m_{1,a}^2m_{2,a} - 3m_{1,a}^4$
5. ...

Bemerkung:

In der Literatur findet man eine grosse Fülle von Behandlungen der Momente. Zu erwähnen sind dabei „Charlier’s Probe“ und „Sheppard’s Korrektur“ sowie die „dimensionslose Form der Momentendefinition“ $M_r = \frac{m_r}{s^r}$. (Vgl. z.B. Murray R. Spiegel, Statistik, McGraw–Hill, Schaum–Lectures.)

2.6.3 Die Schiefe einer Verteilung

Die **Schiefe** soll die Stärke der Einseitigkeit oder Neigung einer Häufigkeitsfunktion oder einer Verteilung beschreiben. Schiefe ist also ein Mass für die Asymmetrie. In manchen Fachgebieten definiert man ein solches Maß nach den praktischen Bedürfnissen. Ein offensichtlicher physikalischer Zugang zur Schiefe ist nicht unmittelbar ersichtlich. Die Schiefe ist in der Regel keine Messgröße. Die Schiefe gibt an, ob (und wie stark) eine Verteilung nach rechts (positiv) oder nach links (negativ) geneigt ist. Es ist leicht nachvollziehbar, dass die nachfolgenden Definitionen für die deskriptive Statistik Sinnvolles leisten:

Definition:

Erster Pearson'scher Schiefekoeffizient:

$$P_{S,1} := \frac{\text{Mittelwert} - \text{Modus}}{\text{Standardabweichung}} = \frac{\bar{x} - x_{\text{modus}}}{s}$$

Zweiter Pearson'scher Schiefekoeffizient:

$$P_{S,1} := \frac{3 \cdot (\text{Mittelwert} - \text{Median})}{\text{Standardabweichung}} = 3 \cdot \frac{\bar{x} - x_{\text{Median}}}{s}$$

Ein Momentenkoeffizient als Schiefenmass:

$$a_3 := \frac{m_3}{s^3} = \frac{m_3}{m_2^{\frac{3}{2}}}$$

Hinweis: Der Median ist einfacher greifbar als ein Modus, vor allem wenn es sich um multimodale Verteilungen handelt.

2.6.4 Kurtosis und Exzess

Die **Kurtosis** und der **Exzess** beschreiben die Wölbung oder Spitzigkeit einer statistischen Verteilung. Schmale aber hohe Verteilungen nennt man „leptokurtisch“. Flache, niedere Verteilungen heißen „platykurtisch“ und mittlere Verteilungen, etwa wie eine schöne Normalverteilung, heißen „mesokurtisch“.

Der mit der Kurtosis verknüpfte Exzess beschreibt die Abweichung im Verlauf der Verteilung zum Verlauf einer Normalverteilung.

Definition:**Momenten–Kurtosis:**

$$\beta_2 := \frac{m_4}{s^4} = \frac{m_4}{m_2^2}$$

Quartil–Dezil–Kurtosis:

Analog zu den Quartilen Q_1, Q_2, Q_3 definiert man auch die Zentile P_1, P_2, \dots, P_{99} . Sei Q der Quantilabstand: $Q = \frac{1}{2}(Q_3 - Q_1)$. Dann definiert man die Quartil–Dezil–Kurtosis:

$$\kappa := \frac{Q}{P_{90} - P_{10}}$$

Exzess:

$$\gamma_2 = \beta_2 - 3$$

2.6.5 Sinn und Gefahr von Kenngrößen

Ein Beispiel mag das Problem der Sinnhaftigkeit von Kenngrößen illustrieren. In der Praxis trifft man oft die Situation, dass Messwerte vorhanden sind, die nun auf allerlei Art höchst kopflos verrechnet werden. Angenommen, es gehe um den Vergleich zwischen dem Wohlbefinden von zwei Personen, Hans und Franz. Hans liegt auf dem Bett. In seinem Zimmer misst man die Temperatur von $30^\circ C$. Die Temperatur stimmt. Für das Wohlbefinden von Hans gibt es bezüglich der Zimmertemperatur kein Problem. Der Mittelwert der Temperatur seiner beiden Hände ist $30^\circ C$.

In einem andern Haus in einem andern Zimmer wird Franz von zwei Gangstern festgehalten. Er wird gefoltert. Eine Hand hat man ihm in kochendes Wasser gedrückt. Wasser kocht bei $100^\circ C$. An der Handoberfläche, wo die Nervenzellen zur Temperabschätzung sitzen, herrscht somit $100^\circ C$. Die andere Hand hat man ihm in einem Eisklotz bei $-40^\circ C$ eingefroren. An jener Handoberfläche herrscht somit $-40^\circ C$. Der Durchschnitt, der Mittelwert also, ist $\frac{1}{2} \cdot (-40^\circ C + 100^\circ C) = 30^\circ C$. Die so ermittelte Durchschnittstemperatur bei Franz ist also gleich der bei Hans. Wer nichts von der Sache weiss, nur den jeweiligen Mittelwert von $30^\circ C$ kennt und sich auf diese eine statistische Aussage verlässt, muss also zum Schlusse kommen, dass die beiden, Hans und Franz, bezüglich Temperatur–Wohlbefinden sich nicht unterscheiden. Krass! Man bedenke, dass täglich mit statistischen Kenngrößen Politik und Werbung gemacht wird, ohne dass dabei jemals weitere Worte über die Datenbestände und ihre Umstände fallen! — !! ? !!

Kapitel 3

Kombinatorik

3.1 Einleitung

3.1.1 Problemstellung

Im Stoffgebiet *Kombinatorik* behandeln wir die 6 Typen der klassischen *Anzahlprobleme*. Dabei geht es um folgende Kategorie von Fragestellungen: Wir fragen nach der *Anzahl* der Möglichkeiten, aus einer endlichen Menge M nach einer gewissen Vorschrift Elemente *auszuwählen*, diese ev. *anzuordnen* oder die Menge *in Klassen einzuteilen*. Dabei können sich Elemente *wiederholen* – oder nicht. Da das Resultat y jeweils eine natürliche Zahl ist, reden wir auch von **Anzahlfunktionen** $M \longmapsto y$.

3.1.2 Fakultäten

In der Kombinatorik spielt der Begriff *Fakultät* eine grosse Rolle. Man definiert die *Fakultäten induktiv*¹ durch die folgende *Rekursion*:

Definition 3.1 (Fakultät:)

$$\begin{aligned} f(0) &= 0! := 1 && (\text{Verankerung}) \\ f(n) &= n! := n \cdot (n-1)! && (\text{Vererbung}) \end{aligned}$$

Bemerkungen:

1. Es gilt dann: $n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1 = \prod_{k=1}^n k$. (Siehe².)
Daraus ergibt sich: $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$ etc..
2. Der Begriff *Rekursion* hat sich heute in der *Informatik* sehr stark verbreitet. Man versteht dort darunter die *Definition einer Funktion oder eines Verfahrens durch sich selbst* (vgl. dazu z.B. Bibl. A6, (Claus, Schwill)). Man darf den Begriff *Rekursion* in diesem hier verwendeten einfachen Sinne jedoch nicht verwechseln mit den in der höheren Mathematik gebräuchlichen, etwas schwierigen Begriffen *allgemeine Rekursion*, *primitive Rekursion*, *rekursive Funktion* (in der Zahlentheorie), *rekursive Relation* (in verschiedenem Sinne in Logik und Mengenlehre). Vgl. dazu Fachlexikon a b c (Bibl.: A1), Iyanaga, Kawada (Bibl. A8) und Meschkowski, Bibl. A11.

Wir halten fest:

¹Nach dem Schema der vollständigen Induktion, vgl. Thema *natürliche Zahlen, Induktionsaxiom* (eines der Axiome aus dem System von Peano).

² \prod steht für „*Produkt*“.

Definition 3.2 (Rekursion (Informatik))

Unter Rekursion verstehen wir hier die Definition einer Funktion oder eines Verfahrens durch sich selbst.

Beispiel: Aus der Definition von $n!$ ergibt sich: $f(n) = n \cdot f(n - 1)$. Die Funktion an der Stelle n wird also durch die Funktion (also durch sich selbst) an der Stelle $n - 1$ definiert.

Die Werte $f(n) = n!$ werden sehr rasch sehr gross. z.B. ist:

$$40! = 815915283247897734345611269596115894272000000000 \approx 8.15915 \cdot 10^{47}.$$

Ein einfaches Programm auf einem Rechner kann daher schnell Probleme machen. Hier ist eine Formel von *Stirling* hilfreich (ohne Beweis):

Satz 3.1 (Formel von Stirling:)

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

e ist hier die Eulersche Zahl: $e \approx 2.71828182845904523536028747135$ (auf 30 Stellen)

3.2 Anordnungsprobleme

3.2.1 Permutationen ohne Wiederholung

Paradigma (Beispiel eines praktischen Problems)³

Problem 3.1 (Sitzmöglichkeiten:)

Situation:

In einem Klassenzimmer befindet sich nichts ausser 26 nummerierten Stühlen. Die Nummern gehen von 1 bis 26. Pulte, Bänke und Tische hat man nach draussen gebracht. Vor der Tür warten 26 Studenten. Zur besseren Unterscheidbarkeit und Benennung erhält auch jeder Student eine verschiedene Nummer von 1 bis 26, mit der er aufgerufen wird.

Frage:

Auf wieviele Arten kann man die 26 Studenten auf die 26 Stühle setzen, d.h. wieviele Sitzordnungen gibt es?

Lösung:

- Der Student Nr. 1 kommt herein. Er findet 26 freie Stühle vor. Somit hat er für sich 26 Sitzmöglichkeiten.
- Der Student Nr. 2 kommt herein, Student Nr. 1 sitzt auf irgend einem Stuhl. Student Nr. 2 findet nur noch 25 freie Stühle vor. Somit hat er für sich nur noch 25 Sitzmöglichkeiten. Diese 25 Sitzmöglichkeiten hat er aber bei jeder Platzierung von Student Nr. 1, welcher sich auf 26 verschiedene Arten platzieren konnte. Zur ersten von Student Nr. 1 benutzten Möglichkeit hat Student Nr. 2 nun 25 Möglichkeiten, zur zweiten Möglichkeit von Student Nr. 1 hat Nr. 2 nun 25 Möglichkeiten, etc., zur letzten Möglichkeit von Student Nr. 1 hat Nr. 2 wiederum 25 Möglichkeiten. Zusammen haben beide also $26 \cdot 25$ Möglichkeiten. Die Anzahlen der Möglichkeiten multiplizieren sich!
- Der Student Nr. 3 kommt herein. Die Studenten Nr. 1 und Nr. 2 sitzen bereits. Student Nr. 3 findet nur noch 24 freie Stühle vor. Somit hat zu jeder der $26 \cdot 25$ Sitzmöglichkeiten der beiden ersten Studenten noch 24 Möglichkeiten. Zusammen haben sie also $26 \cdot 25 \cdot 24$ Möglichkeiten, da sich ja die Anzahlen der Möglichkeiten multiplizieren.
- So geht es dann weiter. Schliesslich kommt der Student Nr. 25 herein. Er hat bei jeder der Sitzmöglichkeiten der vorher hineingegangenen Studenten noch 2 freie Plätze zur Auswahl. Total haben also die 25 ersten Studenten $26 \cdot 25 \cdot 24 \cdot \dots \cdot 3 \cdot 2$ Sitzmöglichkeiten.

³Ein Paradigma ist ein Lehrbeispiel

- Endlich kommt der letzte Student mit der Nummer 26 herein. Er hat bei jeder der Sitzmöglichkeiten der andern Studenten noch einen freien Platz zur Auswahl. Total haben somit die 26 Studenten $26 \cdot 25 \cdot 24 \cdot \dots \cdot 3 \cdot 2 \cdot 1 = 26!$ Sitzmöglichkeiten.

Bemerkung: Falls in einer Menge von Individuen jedes Element (Individuum) einen unterscheidbaren Namen trägt, so kann man die Elemente auch den „Namen nach“, d.h. alphabetisch ordnen, so wie in einem Lexikon: A... kommt vor B... etc., Aa... vor Ab... etc.. In einem solchen Fall spricht man von einer *lexikographischen Anordnung*.

Zum Nachdenken:

Falls die Klasse in 10 Sekunden einen Platzwechsel schafft, so braucht sie also $10 \cdot 26!$ Sekunden für alle Platzwechsel. Das sind $\frac{10 \cdot 26!}{60 \cdot 60 \cdot 24 \cdot 365}$ Jahre $= 1.278310^{20}$ Jahre. Vergleich: Das Alter des Universums bei der Urknalltheorie wird gegenwärtig auf ca. 1 bis 2 mal 10^{10} Jahre geschätzt⁴. Um die Sitzordnungen alle ohne Pause zu realisieren, bräuchte es also etwa 10^{10} mal soviel Zeit, wie das Universum alt ist!

Verallgemeinerung des Problems:

Statt mit 26 Studenten kann man das Problem gleich allgemein mit n Studenten lösen. In der Argumentation ist dann 26 durch n , 25 durch $n - 1$ etc. zu ersetzen. Man erhält schliesslich so total ($n!$) Möglichkeiten, n Studenten auf n Plätze zu setzen.

Das abstrakte Problem

Gegeben sei eine Menge \mathcal{M}_n mit n Elementen, welche durchnummieriert sind mit den Nummern von 1 bis n . \mathcal{M}_n entspricht der Menge der Studenten im vorherigen Beispiel. Dadurch hat man eine bijektive Zuordnung der nummerierten Elemente n_k zur Teilmenge der natürlichen Zahlen $\mathbf{N}_n = \{1, 2, 3, \dots, n\}$. (Damit hat man eine bijektive Funktion). Da die Zuordnung eindeutig ist, können wir die n_k jeweils gerade durch k ersetzen, ohne das Problem zu verändern: $\mathcal{M} = \mathbf{N}_n = \{1, 2, 3, \dots, n\}$. Gesucht ist nun die Anzahl der Möglichkeiten, die Menge $\mathbf{N}_n = \{1, 2, 3, \dots, n\}$ auf sich selbst abzubilden, d.h. im obigen Problem die Menge der Nummern der Studenten \mathbf{N}_n der Menge der Nummern der Stühle \mathbf{N}_n zuzuordnen.

Sei $\sigma(k)$ bei einer solchen Zuordnung (im obigen Problem eine Sitzmöglichkeit) das Bild (oben die Stuhlnummer) von k (k entspricht oben der Nummer des Studenten). Dann wird also durch eine solche Zuordnung σ die Menge $\{1, 2, 3, \dots, n\}$ (oben die Menge der Studenten) der Menge $\{\sigma(1), \sigma(2), \sigma(3), \dots, \sigma(n)\}$ (oben die Stühle) zugeordnet. Schreibt man die Bilder $\sigma(k)$ unter die Urbilder k , so erscheint die durch σ ausgesonderte Relationsmenge in folgender Gestalt:

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \dots & \sigma(n) \end{pmatrix}$$

Damit ist also eine Teilmenge von $\mathbf{N}_n \times \mathbf{N}_n$ gegeben, für die die Relation „Funktion σ “ zutrifft. Durch das folgende Schema wird daher eine neue Anordnung $\sigma(1), \sigma(2), \sigma(3), \dots, \sigma(n)$ der Elemente $1, 2, 3, \dots, n$ definiert.

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \dots & \sigma(n) \end{pmatrix}$$

Wir sagen:

Definition 3.3 (Permutation:)

Die Anordnung $\sigma(1), \sigma(2), \sigma(3), \dots, \sigma(n)$ der Elemente aus \mathbf{N}_n heisst **Permutation \mathcal{P} der Anordnung $(1, 2, 3, \dots, n)$ dieser Elemente**.

⁴Die Fachleute streiten sich allerdings über diesen Wert. Je nach Wissensstand wird er laufend berichtigt.

Um eine Permutation zu geben, können wir auch schreiben:

$$P = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \dots & \sigma(n) \end{pmatrix}$$

Die Reihenfolge der Spalten kann beliebig sein.

Beispiel: Durch die folgende Anordnung ist eine solche Permutation gegeben:

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 5 & 3 & 2 \end{pmatrix}$$

1 wird auf 4, 2 auf 1 u.s.w.. abgebildet. Nun können wir unser Problem mit den Studenten und den Stühlen abstrakt und allgemein stellen:

Problem 3.2

Permutationen ohne Wiederholung:

Frage:

Wieviele Permutationen \mathcal{P} der Nummern $1, 2, \dots, n$ gibt es?

Oder anders gefragt: Wieviele Anordnungsmöglichkeiten der Zahlen $1, 2, \dots, n$ in einer Reihe gibt es?

Oder nochmals anders gefragt: Wieviele bijektive Funktionen $\mathbb{N}_n \rightarrow \mathbb{N}_n$ gibt es?

Symbole 1 : $P(n)$

Sei $P(n) =$ Anzahl Permutationen der Elemente von M_n der natürlichen Zahlen von 1 bis n .

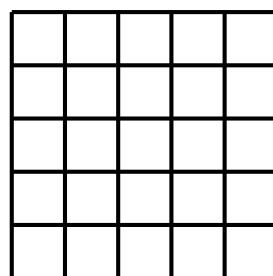
Nun wissen wir:

Satz 3.2

Permutationen ohne Wiederholung:

$$P(n) = n!$$

Abbildung 3.1: Teilflächen, verschieden zu färben ... • Surfaces partielles, à colorer de manière différente...



Beispiel: Wieviele Möglichkeiten gibt es, die in 3.1 gezeigten Teilflächen mit verschiedenen Farben zu färben? – Bei einer Färbung werden den 25 verschiedenen Flächen 25 verschiedene Farben zugeordnet. Statt Flächen und Farben kann man auch nur die Nummern 1 bis 25 betrachten. Man hat also eine bijektive Abbildung einer Menge \mathcal{M}_{25} oder von \mathbb{N}_{25} auf sich. Es wird also nach $P(25) =$ Anzahl Permutationen von $1, 2, 3, \dots, 25$ gefragt. Das gibt $25! \approx 1.55112 \cdot 10^{25}$. Wie lange hätte wohl einer, um alle Möglichkeiten auszuprobieren?

3.2.2 Permutationen mit Wiederholung

Paradigma

Problem 3.3

Vertauschungsmöglichkeiten von Briefen:

Situation: Ein Personalchef hat 20 verschiedene Briefe geschrieben. Davon sind 7 identische Kopien eines Informationsschreibens an eine Gruppe von Mitarbeitern, die andern 13 Briefe sind vertrauliche und persönliche Antworten in Lohnfragen anderer Mitarbeiter. Die zugehörigen 20 Couverts liegen ebenfalls bereit.

Frage: Wieviele Möglichkeiten hat die Sekretärin, die Briefe zu verschicken, sodass für sie Probleme entstehen könnten?

Lösung: Wenn alle Briefe verschieden wären, so hätte sie $20!$ Möglichkeiten, die Briefe in Couverts zu stecken. Da nur eine Möglichkeit akzeptiert werden kann, führen dann $(20! - 1)$ Möglichkeiten zu Problemen.

Wenn nun 7 Briefe gleich sind, können diese 7 Briefe unter sich vertauscht werden, ohne dass ein Problem entsteht. Man kann das auf $7!$ Arten tun. Wenn nun X die Anzahl der Platzierungsmöglichkeiten der 13 verschiedenen Briefe in den 20 Couverts ist, so können zu jeder der X Möglichkeiten der verschiedenen Briefe die restlichen, gleichen Briefe auf $Y = 7!$ Arten unter sich vertauscht werden, ohne dass etwas passiert. Da das bei jeder der X Möglichkeiten der Fall ist, multiplizieren sich die Anzahlen der Möglichkeiten zur Gesamtzahl der Möglichkeiten. Andere Vertauschungsmöglichkeiten als die hier vorkommenden hat man nicht. Somit gilt: $20! = X \cdot Y = X \cdot 7!$ und damit $X = \frac{20!}{7!}$.

Die Anzahl unerwünschter Möglichkeiten ist somit $X - 1 = \frac{20!}{7!} - 1 = 482718652416000 - 1 \approx 4.82719 \cdot 10^{14}$.

Verallgemeinerung des Problems:

Wir gehen wieder von 20 Briefen aus, 7 davon gleich, die wir zur *Klasse 1* zusammenfassen. Weiter sei jetzt nochmals ein spezieller Brief da, zu welchem sich noch zwei gleiche finden. Diese 3 seien in einer *Klasse 2* zusammengefasst. Dann finden wir nochmals 4 gleiche, die in einer *Klasse 3* zusammengefasst werden. Sei nun Y_i die Anzahl der Vertauschungsmöglichkeiten der Briefe in der *Klasse i* unter sich und X wie vorhin die Anzahl der Vertauschungsmöglichkeiten der restlichen ungleichen Briefe. Dann gilt aus demselben Grunde wie oben:

$$20! = X \cdot Y_1! \cdot Y_2! \cdot Y_3! = X \cdot 7! \cdot 3! \cdot 4!, \text{ also}$$

$$X = \frac{20!}{7! \cdot 3! \cdot 4!}$$

Das führt uns auf folgendes allgemeinere Problem:

Gegeben:

Total n Briefe, n Couverts, davon k Klassen je unter sich gleicher Briefe wie folgt:

Klasse₁: m_1 gleiche Briefe vom Typ 1,

Klasse₂: m_2 gleiche Briefe vom Typ 2,

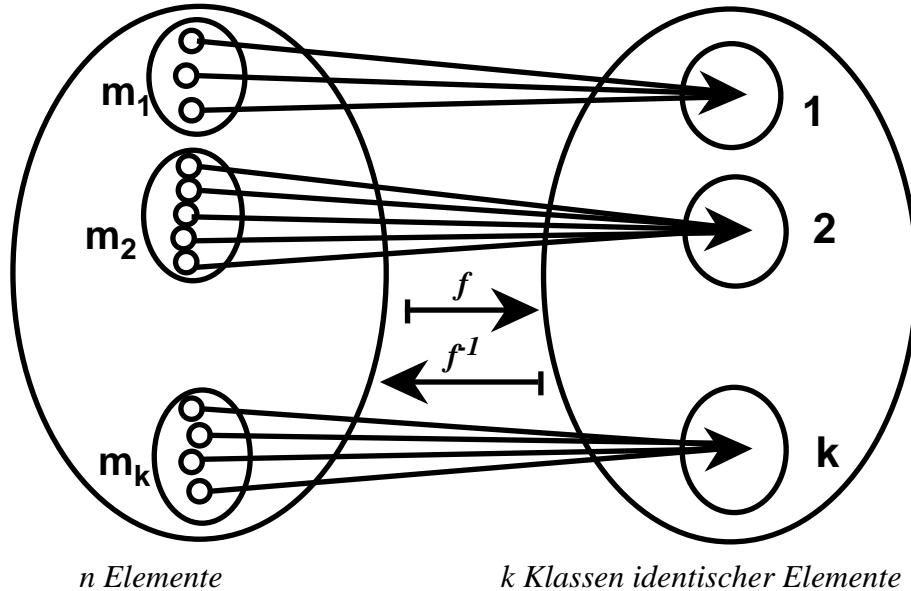
⋮ ⋮

Klasse_k: m_k gleiche Briefe vom Typ k

Gesucht: Anzahl Möglichkeiten $P_n(m_1, m_2, \dots, m_k)$, die Briefe in die Couverts zu platzieren.

Symbol 2 : $P_n(m_1, m_2, \dots, m_k)$

$P_n(m_1, m_2, \dots, m_k)$ = Anzahl Möglichkeiten, die eben beschriebenen n Objekte (hier Briefe, wobei k Klassen mit je n_j gleichen Objekten darunter sind) auf n Plätze (hier Couverts) zu platzieren.

Abbildung 3.2: Anzahl möglicher Umkehrabbildungen f^{-1} ? • Possibilités d'applications inverses f^{-1} ?

Definition 3.4 (Permutat. m. Wiederholung:) Gegeben sei eine Menge \mathcal{M}_n mit n Elementen. Darin seien k Klassen mit je n_i gleichen Elementen (pro Klasse i) enthalten. Bei der Nummerierung der Elemente erhalten alle Elemente einer Klasse dieselbe Nummer. Eine Permutation der Elemente von \mathcal{M}_n nennen wir Permutation mit Wiederholung.

Wir wissen jetzt:

Satz 3.3 (Permutationen mit Wiederholung) :

Anzahl der Permutationen mit Wiederholung:

$$P_n(m_1, m_2, \dots, m_k) = \frac{n!}{m_1! \cdot m_2! \cdot m_k!}$$

Das abstrakte Problem

Gegeben:

Eine Menge mit n Elementen, z.B. $\mathbf{R}_n = \{1, 2, 3, \dots, n\}$ sowie eine Menge mit k Elementen, z.B. $\mathbf{R}_k = \{1, 2, 3, \dots, k\}$, $n \geq k$. Man betrachte dann die möglichen Funktionen $f : \mathbf{R}_n \rightarrow \mathbf{R}_k$ (ein Beispiel ist in Abb. 3.2 dargestellt).

Gesucht:

Anzahl möglicher Umkehrabbildungen $f^{-1} : \mathbf{R}_k \rightarrow \mathbf{R}_n$. Dabei wird das erste Element (1 rechts im Bild) m_1 mal abgebildet, das zweite Element (2 rechts im Bild) m_1 mal u.s.w..

Es werden also die k Klassen gleicher Elemente (gleiche Briefe im Paradigma) auf die n verschiedenen Elemente (Couverts im Paradigma) abgebildet. Die gesuchte Anzahl ist dann $P_n(m_1, m_2, \dots, m_k)$.

Beispiel: Auf wieviele Arten lassen sich in einer Klasse mit 26 Studenten 5 Arbeitsgruppen mit 4, 5, 5, 6 und 6 Studenten bilden? Die Lösung ist:

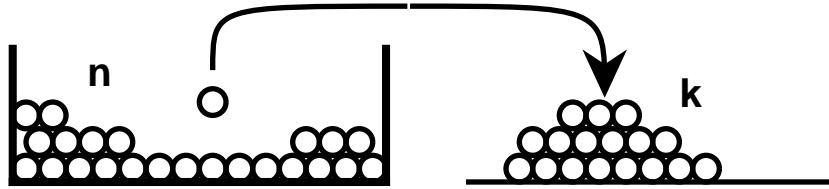
$$P_{26}(4, 5, 5, 6, 6) = \frac{26!}{4! \cdot 5!^2 \cdot 6!^2} = 2251024905729600 \approx 2.25102 \cdot 10^{15}$$

3.3 Auswahlprobleme mit und ohne Anordnung

3.3.1 Die Fragestellungen

Kombinationen

Abbildung 3.3: Auswahlproblem, Kombinationen • *Problème de sélection, combinaisons*



Problem 3.4 (Auswahlproblem:)

Gegeben:

Eine Kiste mit n wohlunterscheidbaren Objekten, z.B. verschiedenfarbigen Kugeln.
Aus der Kiste werden dann auf eine beliebige Art und Weise k Objekte herausgegriffen und nebenan aufgehäuft. (Vgl. Abb. 3.3.)

Frage:

Auf wieviele Arten sind solche Haufenbildungen nebenan möglich?

Wohlverstanden spielt bei der Haufenbildung die Anordnung der Objekte resp. der Kugeln keine Rolle. Dieses Problem lässt sich ohne viel Denkaufwand gleich abstrakt stellen. Die Kugeln in der Kiste bilden eine Menge \mathcal{M}_n , z.B. $\mathcal{M}_n = \mathbf{N}_n = \{1, 2, 3, \dots, n\}$. Herausgegriffen wird eine Teilmenge $\mathcal{M}_k \subseteq \mathcal{M}_n$, z.B. $\mathbf{N}_k = \{1, 2, 3, \dots, k\}$, $k \leq n$. Diese Teilmenge bildet den Haufen nebenan.

Definition 3.5

Kombination ohne Wiederholung

Eine solche Auswahl von k Elementen aus \mathcal{M}_n heisst **Kombination k -ter Ordnung ohne Wiederholung bei n Elementen**, kurz: **Kombination k -ter Ordnung**.

Symbole 3 (Anzahl Kombinationen:)

$C(k, n) = \text{Anzahl Kombinationen } k\text{-ter Ordnung bei } n \text{ Elementen.}$

Abstraktes Problem (Kombinationen ohne Wiederholung):

Gegeben: Eine Menge \mathcal{M}_n mit n Elementen.

Frage: $C(k, n) = ?$ D.h. wieviele Teilmengen mit genau k Elementen kann man bilden?

Variationen

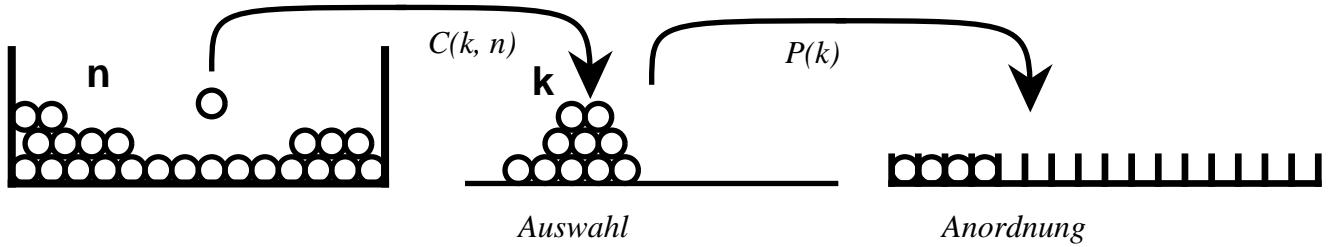
In Abb. 3.4 wird die Auswahl (Kombination) anschliessend noch angeordnet. Zwei solche Kombinationen mit denselben Elementen, aber verschiedener Anordnungen sind jetzt unterscheidbar. Man definiert daher:

Definition 3.6

Variation ohne Wiederholung:

Werden die aus \mathcal{M}_n ausgewählten Elemente (die Kombination also) noch angeordnet, so spricht man von einer **Variation k -ter Ordnung ohne Wiederholung bei n Elementen**. Kurz: **Variation k -ter Ordnung**.

Abbildung 3.4: Relationsmenge, Abbildung • Ensemble de relations et d'applications (choix, disposition)



Symbol 4 (Anzahl Variationen:)

$V(k, n) = \text{Anzahl Variationen } k\text{-ter Ordnung bei } n \text{ Elementen.}$

Beispiel:

Gegeben seien die Elemente a, b und c . Gesucht sind alle Kombinationen und Variationen 2-ter Ordnung.

Lösung:

Kombinationen : $a b \quad a c \quad b c: \quad 3 \text{ Stück.}$

Variationen: $a b \quad a c \quad b c$
 $b a \quad c a \quad c b: \quad 6 \text{ Stück.}$

Wiederholungen

Ersetzt man in der Vorratsmenge M_n jedes der Elemente e_i durch eine Menge E_i mit gleichen Elementen, die sich nur durch einen *internen Index* unterscheiden (z.B. $E_i = \{e_{i1}, e_{i2}, e_{i3}, \dots\}$), so wird es möglich, ein Element e_i mehrmals auszuwählen, wobei der interne Index nach der Auswahl wieder weggelassen werden kann⁵. Denselben Effekt erzielen wir, wenn wir nach der Auswahl eines Elementes eine identische Kopie dieses Elementes wieder zurücklegen. Wir stellen uns also vor, dass sich ein Element e_i bei seiner Auswahl dupliziert, sodass trotz Auswahl und Entfernung des Elements die Menge M_n unverändert bleibt. Ein Element wird also bei der Auswahl und Entfernung aus M_n sofort wieder in M_n nachgeliefert, etwa so wie bei einem bestimmten Artikel im Regal eines gut geführten Selbstbedienungsladens, wo die Regale immer sofort wieder aufgefüllt werden. Falls dieses Auffüllen, Duplizieren, Kopieren oder Zurücklegen beliebig oft möglich ist, so sagen wir, die Elemente in M_n seien *wiederholt auswählbar*. Wir definieren nun:

Definition 3.7

Kombination und Variation mit Wiederholung:

Sind bei der Bildung einer Kombination oder einer Variation die Elemente aus M_n wiederholt auswählbar, so spricht man von einer Kombination oder einer Variation mit Wiederholung.

Wir beginnen nun mit der Variation ohne Wiederholung:

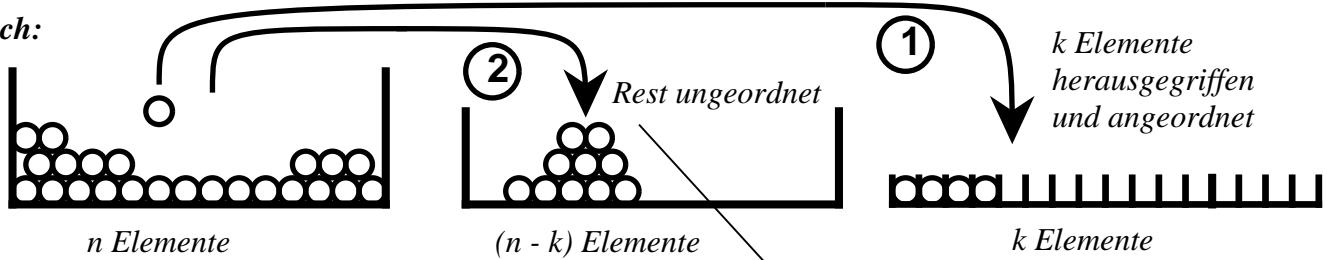
3.3.2 Variation ohne Wiederholung

Aus n Elementen werden k Elemente herausgegriffen und angeordnet, ohne Wiederholung, so wie in Abb. 3.5 dargestellt. Dort wird z.B. das Element e_1 auf den Platz 1, e_2 auf den Platz 2 u.s.w.. gelegt. Alle

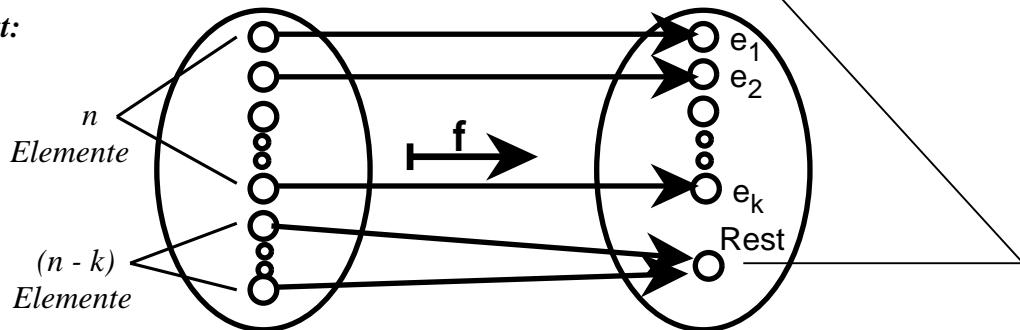
⁵Der interne Index wird nur zur Bildung der „Mengen gleicher Elemente E_i “ gebraucht, die notwendig sind, um eine wiederholte Auswahl desselben Elements möglich zu machen.

Abbildung 3.5: Variationen ohne Wiederholung • Arrangement sans répétition (image — abstrait, éléments, reste)

Bildlich:



Abstrakt:



$(n - k)$ nicht ausgewählten Elementen, der Rest also, kann man sich anschliessend in eine Kiste nebenan gelegt denken, auf einen Haufen also. Diese anschliessende Operation verändert die Anzahl Auswahl- und Anordnungsmöglichkeiten der ersten k Elemente nicht, denn diese Haufenbildung ist eine einzige, unabhängige Handlung, die nichts weiteres beiträgt. In dieser Restkiste nebenan spielt also die Anordnung der Elemente keine Rolle. Man unterscheidet diese Elemente demnach nicht, es ist egal, wie sie liegen. Daher bilden sie eine Klasse nicht unterschiedener, also gleicher Elemente, die auf nur eine einzige Art angeordnet werden können (da sie als nicht unterscheidbar gelten). Daher hat man folgendes Problem: Man hat n Elemente, k verschiedene und $n - k$ gleiche. Diese Elemente sind anzuordnen. Oder abstrakt: Man sucht die Anzahl der möglichen Umkehrfunktionen f^{-1} (vgl. Abb. 3.5). Das Problem haben wir aber bereits bei den Permutationen mit Wiederholung gelöst: Die Anzahl ist $P_n(n - k) = \frac{n!}{(n - k)!}$

Satz 3.4 (Variationen ohne Wiederholung:)

$$V(k, n) = P_n(n - k) = \frac{n!}{(n - k)!}$$

Beispiel:

Auf wieviele Arten kann man 20 verschiedene vorhandene Ferienjobs an 26 verschiedene Studenten verteilen, die alle einen solchen Job haben wollen, wenn diese Jobs nicht in Teiljobs aufteilbar sind?

Es handelt sich um die Auswahl 20 aus 26 mit anschliessender Zuordnung zu unterscheidbaren Studenten, d.h. Anordnung. Die Lösung ist somit:

$$V(20, 26) = \frac{26!}{(26 - 20)!} = \frac{26!}{(6)!} = 67215243521100939264000000 \approx 6.72152 \cdot 10^{25}$$

Ein Spezialfall: $V(n, n) = P_n(n - n) = P_n(0) = P(n)$

~ Permutation ohne Wiederholung!

3.3.3 Kombination ohne Wiederholung

Die Formel

Auf Seite 33 haben wir gesehen, dass sich bei Aussortierung einer Teilmenge gleicher Elemente die Anzahl der Möglichkeiten multiplikativ verhalten. Da war $20! = X \cdot Y = X \cdot 7!$. Die gleiche Situation finden wir beim Übergang von den Kombinationen zu den Variationen: Eine Variation (k Elemente aus n Elementen) entsteht aus einer Kombination durch Anordnung der k ausgewählten Elementen. Dazu hat man $P(k) = k!$ Möglichkeiten. Es gilt also:

Lemma 3.1 (Variationen und Kombination:)

$$V(k, n) = C(k, n) \cdot P(k), \text{ also } \frac{n!}{(n-k)!} = C(k, n) \cdot k!$$

Daraus folgt:

Satz 3.5 (Kombination ohne Wiederholung)

$$C(k, n) = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{1 \cdot 2 \cdots k}$$

Das Beispiel Zahlenlotto „6 aus 45“:

Auf wieviele Arten kann man 6 verschiedene Zahlen aus den 45 ersten natürlichen Zahlen auswählen? Hier handelt es sich um eine typische Frage nach der Anzahl Kombinationen $C(6, 45)$. Diese ist gleich:

$$\frac{45!}{6! \cdot (45-6)!} = \frac{45!}{6! \cdot (39)!} = 8145060 \approx 8.14506 \cdot 10^6$$

Binomialkoeffizienten

Multipliziert man das Binom $(a+b)^n = \overbrace{(a+b) \cdot (a+b) \cdot \dots \cdot (a+b)}^n$ nach den Regeln des Distributivgesetzes aus, so entstehen lauter Summanden der Form $m_k \cdot a^k \cdot b^{n-k}$ mit $0 \leq k \leq n$ und $m, k, n \in \mathbb{N}_0$.

Beim Ausmultiplizieren nimmt man der Reihe nach aus jedem Faktor $(a+b)$ einen der Summanden a oder b und multipliziert diese Faktoren zu einem Produkt $a^k \cdot b^{n-k}$. Falls man in jedem Summanden a und nie b nimmt, entsteht $a^n \cdot b^0$. Falls man in j Summanden a und folglich in $n-j$ Summanden b nimmt, entsteht $a^j \cdot b^{n-j}$. Dabei gibt es hier verschiedene Möglichkeiten, das a oder das b auszuwählen: Man kann z.B. im ersten Faktor a , im zweiten b , im dritten wieder a wählen etc., man kann aber auch zuerst b , dann a und dann wieder a wählen etc.. m_k ist die Anzahl der Möglichkeiten, a in genau k Faktoren $(a+b)$ und b in genau $n-k$ Faktoren zu wählen. Es ist dann:

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

Wie gross ist nun m_k ? — Hier handelt es sich um ein Auswahlproblem: Auf wieviele Arten kann man aus den n Faktoren $(a+b)$ k Faktoren auswählen und dort jeweils den Anteil a (und folglich in den restlichen $n-k$ Faktoren jeweils den Anteil b) nehmen? Diese Frage ist äquivalent mit der einfacheren Frage: Auf wieviele verschiedene Arten kann man aus n Elementen (Faktoren $(a+b)$) jetzt k Elemente auswählen? Die Antwort ist nun einfach: $m_k = C(k, n)$. m_k hat einen Namen:

Definition 3.8 (Binomialkoeffizient:) :

m_k heisst **Binomialkoeffizient**.

Symbol 5 (Binomialkoeffizient:) $m_k := \binom{n}{k}$

Wir wissen jetzt:

Satz 3.6 (Binomischer Lehrsatz)

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

Binomialkoeffizienten kann man bekanntlich im *Pascalschen Dreieck* ablesen:

Pascalsches Dreieck:

$$\begin{array}{llllll}
 n = 0 & \dots & & & & 1 \\
 n = 1 & \dots & & & 1 & 1 \\
 n = 2 & \dots & & 1 & 2 & 1 \\
 n = 3 & \dots & & 1 & 3 & 3 & 1 \\
 n = 4 & \dots & & 1 & 4 & 6 & 4 & 1 \\
 \text{etc.} & \dots & & & & \dots & \\
 \end{array}$$

Die vertikale Position ist n , die horizontale k . Die Numerierung beginnt jeweils mit 0. So liest man z.B. ab: $\binom{4}{1} = 4$ und $\binom{4}{2} = 6$.

Für die Binomialkoeffizienten kann man mit Hilfe von $\binom{n}{k} = C(k, n) = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{1 \cdot 2 \cdot \dots \cdot k}$ sowie mit dem Prinzip der vollständigen Induktion⁶ leicht die folgenden Gesetze beweisen:

Satz 3.7

Einige Eigenschaften der Binomialkoeffizienten:

$$\begin{array}{ll}
 1) \quad \binom{n}{k} = \binom{n}{n-k} & 2) \quad \binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k} \\
 3) \quad 2^n = \sum_{k=0}^n \binom{n}{k} & 4) \quad \sum_{k=0}^r \binom{p}{k} \cdot \binom{q}{r-k} = \binom{p+q}{r} \\
 5) \quad \sum_{s=0}^{n-1} \binom{k+s}{k} = \binom{n+k}{k+1} & 6) \quad \sum_{k=0}^p \binom{p}{k}^2 = \binom{2p}{p}
 \end{array}$$

Z.B. die Formel $2^n = \sum_{k=0}^n \binom{n}{k}$ ergibt sich aus $2^n = (1+1)^n = \sum_{k=0}^n \binom{n}{k} \cdot 1^k \cdot 1^{n-k} = \sum_{k=0}^n \binom{n}{k}$ mit Hilfe des binomischen Lehrsatzes.

3.3.4 Variation mit Wiederholung

Die Formel

Die *Variation mit Wiederholung* ist auf Seite 36 erklärt worden. Die Formel für die Anzahl Variationen mit Wiederholung hingegen müssen wir noch erarbeiten. Dazu verwenden wir folgendes Symbol:

Symbole 6 : $\bar{V}(k, n)$

$\bar{V}(k, n) = \text{Anzahl Variationen mit Wiederholung bei einer Auswahl von } k \text{ Elementen aus einem Vorrat mit } n \text{ verschiedenen Elementen, die alle wiederholbar sind.}$

Herleitung der Formel:

Wir betrachten die k numerierten Plätze, auf denen die auszuwählenden Elemente anzuordnen sind (vgl. Abb. 3.5 oben links im Bild). Da wir jedes der n Elemente im Vorrat auswählen können, hat man n Möglichkeiten, den 1. Platz zu besetzen. Bei der Auswahl für den 2. Platz hat man aber wieder n Elemente im Vorrat zur Auswahl, da wegen der Wiederholbarkeit wieder jedes Element vorhanden ist und gewählt werden kann: Zu jeder der n Möglichkeiten für den 1. Platz hat man n Möglichkeiten für den 2. Platz, total also jetzt $n \cdot n = n^2$ Möglichkeiten. Genauso geht es für den 3. Platz: Zu jeder der n^2 Möglichkeiten für die Plätze 1 und 2 hat man n Möglichkeiten für den 3. Platz, total also jetzt $n^2 \cdot n = n^3$ Möglichkeiten. So fährt man fort: Für die Besetzung der ersten 4 Plätze hat man n^4 Möglichkeiten, für die Besetzung der ersten 5 Plätze n^5 Möglichkeiten und schliesslich für die Besetzung der ersten k Plätze hat man n^k Möglichkeiten. Wir haben somit den Satz:

⁶Vgl. Zahlenlehre

Satz 3.8 (Variationen mit Wiederholung:)

$$\bar{V}(k, n) = n^k$$

Beispiel:

Auf wieviele Arten können 26 (unterscheidbare) Studenten sich in 12 verschiedene Kurse einschreiben, wenn jeder Kurs 26 Plätze offen hat, also keine Platzbeschränkung besteht?

Lösung:

Der erste Student hat 12 Möglichkeiten, sich in einen Kurs einzuschreiben. Zu jeder dieser Möglichkeiten des ersten Studenten hat der zweite auch 12 Möglichkeiten, sich in einen Kurs einzuschreiben. Beide zusammen haben also 12^2 Möglichkeiten. Für den dritten, vierten etc. Studenten geht das auch so: Jeder hat die 12 Möglichkeiten, und die Möglichkeiten multiplizieren sich. Es handelt sich um eine Variation mit Wiederholung. Total gibt es $\bar{V}(k, n) = \bar{V}(26, 12) = 12^{26} = 11447545997288281555215581184 \approx 1.14475 \cdot 10^{28}$ Möglichkeiten.

Merke: Aus diesem Beispiel ersieht man, dass $k > n$ sein kann.

Anwendung: Die Mächtigkeit der Potenzmenge

Die Potenzmenge ist bekanntlich die Menge aller Teilmengen.

Problem 3.5

Mächtigkeit der Potenzmenge:

Gegeben: Eine Menge \mathcal{M} mit n Elementen.

Frage: Wieviele Teilmengen hat \mathcal{M} ?

Lösung:

$\binom{n}{k} = C(k, n)$ ist bekanntlich die Anzahl Teilmengen mit k Elementen, denn hier handelt es sich ja um das typische Auswahlproblem. Nun kann man eine oder mehrere Teilmengen mit 0 (leere Menge), 1, 2, ..., n Elementen wählen. Total hat man also:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = \sum_{k=0}^n \binom{n}{k} = \sum_{k=0}^n \binom{n}{k} \cdot 1^k \cdot 1^{n-k} = (1+1)^n = 2^n.$$

Satz 3.9 :

Mächtigkeit der Potenzmenge:

Die Potenzmenge einer Menge mit n Elementen hat 2^n Elemente.

Eine Menge mit n Elementen hat also genau 2^n Teilmengen.

3.3.5 Kombination mit Wiederholung

Hier sollen aus einer Menge mit n Elementen k Elemente ausgewählt werden, wobei jedes ausgewählte Element bei der Auswahl in der Menge dupliziert wird resp. nachgeliefert wird, so dass die Menge trotz Auswahl immer aus denselben Elementen besteht. Wie gross ist die Anzahl Auswahlmöglichkeiten?

Für die Berechnung dieser Anzahl ist es unwesentlich, ob die Menge \mathcal{M}_n aus Kugeln, Losen oder Zahlen etc. besteht, d.h. welcher Natur die Elemente sind. Wir dürfen daher annehmen, es handle sich um die natürlichen Zahlen von 1 bis n : $\mathcal{M}_n = \{1, 2, 3, \dots, n\}$. Wenn wir jetzt k Elemente (d.h. Zahlen) auswählen, so wollen wir diese immer ihrer Grösse nach aufreihen, statt sie bloss „auf einen Haufen zu legen“. Wir reden hier von der *Standardanordnung*. Eine solche Auswahl $\{e_1, e_2, \dots, e_k\}$ wird also immer in der Anordnung $e_1 \leq e_2 \leq \dots \leq e_k$ präsentiert. Dadurch wird die Anzahl der Auswahlmöglichkeiten ja

nicht verändert.

Wie ist nun dem Problem der Wiederholungen beizukommen? Die Idee, aus $k \cdot n$ Elementen auszuwählen, führt zu keinem Resultat, da die Elemente einer Auswahlmenge dann auf verschiedene Weise gewonnen werden können, was fälschlicherweise die Anzahl Auswahlmöglichkeiten erhöht. So geht es also nicht. Um der Sache beizukommen, muss man etwas weiter ausholen:

Wir führen dazu $k - 1$ neue Elemente J_1, J_2, \dots, J_{k-1} ein und fügen diese der Menge \mathcal{M}_n an. So erhalten wir eine neue Menge $\mathcal{M}_n^{k-1} = \{1, 2, 3, \dots, n, J_1, J_2, \dots, J_{k-1}\}$ mit $n + k - 1$ Elementen. Die neu gültige Standardanordnung entspreche der hier gegebenen Aufzählung der Elemente: Die J_i werden hinten den Nummern nach angefügt. Dabei gelte für die Elemente J_i die folgende Interpretation: J_i ist eine Vorschrift oder Funktion, die auf jenen ausgewählten Standardanordnungen operiert, in denen sie selbst allenfalls vorkommt. Die durch J_i gegebene Vorschrift lautet: Ersetze das Symbol J_i in einer ausgewählten Standardanordnung durch das Element e_i derselben Auswahl, nachdem alle J_p mit $p < i$ schon ersetzt sind. Führt man alle diese Ersetzungen durch, so erhält man aus einer *primären Auswahl* die *Endstandardanordnung*. Da k Elemente auszuwählen sind, es aber nur $k - 1$ Elemente J_i gibt, kommt in einer Standardauswahl immer mindestens ein Element $e_j \in \mathcal{M}_n$ vor, in unserem Falle eine der gegebenen natürlichen Zahlen $1, 2, 3, \dots, n$. J_i bewirkt somit immer eine Ersetzung durch ein weiter vorne vorkommendes Element in der Standardanordnung, also eine Duplikation. Da so jedes Element einmal ausgewählt und dann noch durch die J_i maximal $k - 1$ mal dupliziert werden kann, besteht die Möglichkeit, dass jedes Element von \mathcal{M}_n dann k mal in der Endstandardanordnung vorkommen kann. Auf diese Art können alle Kombinationen mit Wiederholung gewonnen werden.

Beispiel:

Gegeben sei $\mathcal{M}_7 = \{1, 2, 3, \dots, 7\}$. Daraus sollen 5 Elemente mit Wiederholung ausgewählt werden. Es ist dann $\mathcal{M}_7^{5-1} = \mathcal{M}_7^4 = \{1, 2, 3, \dots, 7, J_1, J_2, J_3, J_4\}$.

Wählt man z.B. $(1, 5, 7, J_1, J_4)$ (in Standardanordnung), so wird wie folgt ersetzt: Zuerst $J_1 \mapsto 1$ (der Index 1 ist kleiner als der Index 4). Das ergibt $(1, 5, 7, 1, J_4)$ in Nicht-Standardanordnung und $(1, 1, 5, 7, J_4)$ in neuer Standardanordnung. Dann wird ersetzt $J_4 \mapsto 7$, was zur Standardanordnung $(1, 1, 5, 7, 7)$ führt.

Ähnlich führt die Auswahl $(4, J_1, J_2, J_3, J_4)$ nach allen Ersetzungen zur Standardanordnung $(4, 4, 4, 4, 4)$.

Bei der Auswahl von 6 Elementen aus \mathcal{M}_8 führt die primäre Auswahl $(2, 3, 7, 8, J_2, J_4)$ auf die Endstandardanordnung $(2, 3, 3, 7, 7, 8)$.

Diese Beispiele machen klar, dass eine primäre Auswahl eindeutig einer Endstandardanordnung entspricht. Die Anzahl der auswählbaren primären Anordnungen ist gleich der Anzahl der Endstandardanordnungen, in welchen alle Elemente bis zu k mal wiederholt vorkommen können. Um $\bar{C}(k, n)$ zu finden, muss man also die Anzahl der primär auswählbaren Standardanordnungen bestimmen. Dort werden k Elemente aus den $n + k - 1$ Elementen $1, 2, 3, \dots, n, J_1, J_2, \dots, J_{k-1}$ ausgewählt. Daher ist $\bar{C}(k, n) = C(k, n + k - 1)$. Somit hat man:

Satz 3.10

Kombinationen mit Wiederholung:

$$\bar{C}(k, n) = C(k, n + k - 1) = \binom{n + k - 1}{k}$$

Beispiel:

Ein Abteilungsleiter hat 19 Ingenieure unter sich, von denen jeder als Projektleiter in Frage kommt.

Es stehen 8 neue Projekte an, die wahrscheinlich nacheinander bearbeitet werden müssen. Wieviele Möglichkeiten bieten sich dem Abteilungsleiter, Projektleiter zu bestimmen, wenn auch in Betracht gezogen werden darf, dass im Extremfall derselbe Ingenieur allen 8 Projekten vorsteht?

Hier handelt es sich um eine Kombination mit Wiederholung. Aus 19 Ingenieuren werden 8 Projektleiter ausgewählt, wobei jeder mehrmals vorkommen darf. Es ist dann:

$$\bar{C}(8, 19) = \binom{19 + 8 - 1}{8} = \binom{26}{8} = \frac{26!}{8! \cdot (26 - 8)!} = \frac{26!}{8! \cdot 18!} = 1562275 \approx 1.56228 \cdot 10^6.$$

Neuere Darstellung siehe:

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

3.4 Übungen

Übungen finden sich in *DIYMU*, Wirz, Bibl. A18 sowie in der klassischen Schulbuchliteratur für die Gymnasialstufe — oder speziell auch in der Literatur zur Wahrscheinlichkeitsrechnung und Statistik.

Kapitel 4

Wahrscheinlichkeitsrechnung

4.1 Einleitung

4.1.1 Problemstellung

Gesucht ist ein Instrument, das Aussagen über betrachtete Massenereignisse erlaubt.

Beispiele von Massenereignissen:

1. Grosse Zahl von Individuen
2. Grosse Zahl von zeitlich gestaffelten Einzelerscheinungen
3. Grosse Zahl von räumlich gestaffelten Einzelerscheinungen
4. Grosse Zahl von Möglichkeiten
5. u.s.w.

Zudem suchen wir Begriffe, die bei der Vermessung solcher **Ereignismengen** nützlich sind.

4.1.2 Anwendung

Wissenschaften, in denen solche Massenereignisse Gegenstand der Betrachtung sind, sind z.B. die folgenden:

1. Statistik
2. Prognosenlehre
3. Technik der Optimierung von Verteilungen (z.B. Datenverteilung auf Datenträgern)
4. Modellierung in Ökonomie, Finanzmathematik, Versicherungslehre, Technik, Medizin u.s.w.. (z.B. Sterblichkeitsvorhersagen, Börsenverhalten, Therapiechancen u.s.w.)
5. u.s.w.

Das mathematische Mittel, auf denen die genannten Theorien fussen, ist die **Wahrscheinlichkeitsrechnung**.

Beispiel Statistik: :

Zur blossen Beschreibung von Massenerscheinungen dient die **deskriptive** oder **beschreibende Statistik**.

Will man hingegen Massenerscheinungen auch noch beurteilen, vergleichen oder ihr Verhalten modellieren, so benützt man die **mathematische Statistik** (beurteilende oder affirmative Statistik, Teststatistik, explorative Statistik, Tests von Modellen wie z.B. Regressionen u.s.w.). Grundlage der mathematischen Statistik ist die **Wahrscheinlichkeitsrechnung**.

Wichtige Idee: Gesucht sind mathematische Methoden, die bei einem vertretbaren Arbeitsaufwand einen verlässlichen Schluss vom Verhalten weniger Individuen (Stichprobe) auf die Grundgesamtheit (sehr viele Individuen) zulassen.

Statistik gehört auch zur Allgemeinbildung des Ingenieurs. Sie ist vor allem wichtig in der Qualitätskontrolle und im Bereich des Vertrauens in technische Komponenten (Grundlage des Vertrauens).

4.1.3 Personen

Personen in der Entstehungsgeschichte der Wahrscheinlichkeitsrechnung:

- Plaiste Pascal (1623 – 1662)
- Pierre Fermat (1601 – 1665)
- Christian Huygens (1629 – 1695)
- Jakob Bernoulli (1654 – 1705) Def. der Wahrscheinlichkeit
- Abraham Moivre (1667 – 1754) Def. der Wahrscheinlichkeit
- Pierre Simon Laplace (1749 – 1827) Def. der Wahrscheinlichkeit
- Carl Friederich Gauss (1777 – 1855)
- Simon-Denis Poisson (1781 – 1840)
- Pafnuti Lwowitsch Tschebyschow (1821 – 1894)
- Andrei Andrejewitsch Markow (1856 – 1922)
- Andrei Nikolajewitsch Kolmogorow (1903 – 1987) Def. der Wahrscheinlichkeit
- Alexander Jakowlewitsch Chintchin (1894 – 1959)

4.2 Zufallsexperiment, Ereignis

Begriff: Zufallsvorgang:

Zwei Interpretationen sind möglich:

1. **Relativer Zufall:** Vorgang, der einen Grund hat, dessen Grund jedoch unbekannt ist.
2. **Absoluter Zufall:** Vorgang, der aus sich selbst heraus existiert und der prinzipiell keinen weiteren Grund hat, also von nichts abhängig ist¹.

Das Problem der Existenz des absoluten Zufalls gehört in die Philosophie. Denn aus der Unkenntnis einer Sache lässt sich nicht ihre Nichtexistenz folgern.

Definition:

Zufallsexperiment oder Zufallsbeobachtung:

Beliebig oft wiederholbarer, nach festen Vorschriften ausgeführter Vorgang mit nicht vorherbestimmbarem Ergebnis (Ergebnismenge) und mehreren möglichen, sich gegenseitig ausschliessenden Resultaten. („Abhängig vom Zufall“.)

Bsp.:

Würfeln, werfen einer Münze, Kartenziehen, Grösse von Menschen messen, Milchmenge einer Kuh messen, farbige Kugeln aus einer Urne ziehen u.s.w..

¹Für viele Theologen ist dies die Eigenschaft Gottes.

Sprechweise:

Das Ereignis „hängt vom Zufall ab“ := Das Ergebnis ist nicht im voraus bestimmbar.

Bsp.: „Würfeln der Zahl 6“.

Die Verteilungsfunktion liefert hier in der Regel keine Handhabe oder kein Merkmal, nach dem man ein Ereignis, z.B. würfeln einer 6, als „vor anderen solchen ausgezeichnet“ bewerten könnte (z.B. würfeln einer 5). Mangels besseren Wissens, aus pragmatischen Gründen also, müsste man daher von „gleichen Chancen für solche Ereignisse“ sprechen.

Definition:**Elementarergebnis:**

Mögliches, nicht weiter zerlegbares Ergebnis (ω_i) eines Zufallsexperiments.

Bsp.: Würfeln mit einem Würfel \rightsquigarrow 6 Möglichkeiten: $R(1), R(2), \dots, R(6)$. Oder Würfeln mit drei Würfeln. Zerlegung: Würfeln mit einem Würfel \rightsquigarrow neues, **atomares Experiment**.

Wichtige Eigenschaft:

Ein Elementarergebnis gehorcht der zweiseitigen Logik: Es kann eintreffen oder nicht.

Definition:

Elementarergebnismenge: Menge aller Elementarergebnisse $\Omega = \{\omega_i \mid i = \dots\}$.

Bemerkung:

Statt Ergebnismenge sagen wir auch **Universalmenge** des Experiments oder **Fundamentalmenge**.

Definition:**Ereignis:**

Teilmenge der Ergebnismenge eines Zufallsexperiments.

Definition:

Elementarereignis zum Ergebnis ω_i : $\{\omega_i\}$

Ereignisse sind demnach Mengen von Elementarereignissen.

Definition:**Ereignisraum:**

Menge aller Ereignisse eines Zufallsexperiments. ($= \mathcal{P}(\Omega)$, $\Omega =$ Grundmenge, Ergebnismenge, $\mathcal{P} =$ Potenzmenge.)

Schreibweise:

Für Ereignisse verwenden wir Grossbuchstaben (A, B, C, ..., X, Y, ...) wie für Mengen.

1. Beispiel: $A = \{\text{würfeln einer 6}\} = \{\omega_6\}$.

Kurz: $A = \{R(6)\}$, $\Omega = \{R(1), R(2), R(3), R(4), R(5), R(6)\} = \{\omega_1, \dots, \omega_6\}$

2. Beispiel: Gegeben sei ein Stab der Länge 1, der zufällig in 3 Stücke gebrochen wird (Stellen x und y , $0 < x < y < 1$). Resultat: 3 Stabstücke mit den Längen x , $y - x$, $1 - y$. \rightsquigarrow Ein Resultat dieses Experiments kann somit als Punkt im Dreieck $\{(x, y) \in ((0, 1) \times (0, 1)) \mid x < y\}$ dargestellt werden.

Bemerkung:

Im ersten Beispiel oben ist der **Ereignisraum endlich**, im zweiten Beispiel ist er **unendlich**.

Ein mögliches Ergebnis kann bei einem Zufallsexperiment eintreffen oder nicht. Nach dem Experiment können wir die Menge der eingetroffenen Ergebnisse E (**statistische Ergebnisse**, **statistische(s) Ereignis(se)**) von der Menge der nicht eingetroffenen $\Omega \setminus E$ unterscheiden. $\rightsquigarrow \omega$ ist eingetroffen $\Leftrightarrow \omega \in E$, ω ist nicht eingetroffen $\Leftrightarrow \omega \notin E$

4.2.1 Zufallsprozesse, Häufigkeit

Zufallsprozesse

Z.B. das einmalige Würfeln mit einem Würfel ist ein atomares Zufallsexperiment, das nicht weiter zerlegbar ist. Neben solchen **atomaren Zufallsexperimenten** trifft man in der Praxis aber auch komplizierte Zufallsprozesse, die z.B. aus mehreren atomaren und nacheinander oder gleichzeitig ablaufenden Zufallsexperimenten zusammengesetzt sind. Man denke z.B. an eine Urne mit farbigen Kugeln, aus der nacheinander 3 mal je eine Kugel gezogen werden soll. In einer ersten Stufe ziehen wir eine Kugel aus 6 vorhandenen, in einer 2. Stufe ziehen wir eine aus 5 vorhandenen und in einer 3. Stufe eine aus 4 vorhandenen. Ein Elementarereignis des Zufallsexperiments besteht hier aus 3 sich folgenden atomaren Ergebnissen.

Sprechweise:

Bei zusammengesetzten Zufallsexperimenten nennen wir die Nummer von sich folgenden atomaren Zufallsexperimenten die **Stufe** und das zusammengesetzte Experiment daher **mehrstufig**.

Definition:

Ein mehrstufiges Zufallsexperiment nennen wir **Zufallsprozess**.

Bsp.:

Ziehen wir zuerst eine schwarze, dann eine weisse und dann wieder eine schwarze Kugel, so schreiben wir als Resultat $R(s_1, w_2, s_3)$. Dieses Zufallsexperiment ist dreistufig. Der Index bedeutet die Stufe. Die Information des Niveaus ist notwendig, da die in der Urne vorhandenen Kugeln (Voraussetzung des Experiments) mit der Stufe ändern. Ein Elementarereignis eines n -stufigen Experiments besteht demnach aus einem geordneten n -Tupel von atomaren Elementarereignissen.

Häufigkeit

Sei das statistische Ereignis E das Resultat eines Versuches (Versuchsreihe vom Umfang n , n Teilversuche, Prozess zum Niveau n). Das gewünschte Ereignis A treffe bei $|E| = n$ Teilversuchen k mal ein. ($\rightsquigarrow n$ mögliche und k günstige Fälle.)

Beispiel: 12 mal würfeln: $E =$ Menge der realisierten und nummerierten Teilresultate. 3 mal kommt die 1. $A =$ Teilmenge der realisierten und nummerierten 1. $\rightsquigarrow |E| = n = 12$, $k = |A| = 3$.

Definition:

Absolute Häufigkeit von A :

$$H(A) := k$$

Relative Häufigkeit von A :

$$h(A) := \frac{k}{n} = \frac{|A|}{|E|}$$

Bsp.:

$A =$ ziehen eines Königs aus einem gemischten Kartenspiel mit 36 Karten bei $n = 100$ Versuchen (100 mal ziehen). Sei $k = 9 \rightsquigarrow H(A) = 9$, $h(A) = 0.09$. Ein Ereignis bei diesem 100–stufigen Zufallsprozess ist demnach ein geordnetes 100–Tupel mit 100 atomaren Ereignissen, welche jeweils das Resultat einer einfachen Ziehung sind.

Folgerung: Absolute Häufigkeiten sind Mächtigkeiten von eingetroffenen Ereignissen

Bsp.:

Versuch: 12 mal würfeln. Resultat: $E = \{2_1, 4_2, 5_3, 1_4, 2_5, 6_6, 3_7, 1_8, 1_9, 5_{10}, \cdot_{11}, 3_{12}\}$. Teilmenge der Elemente „1 ist eingetroffen“: $A := \{x_i \in E \mid x_i = 1_i\} = \{1_4, 1_8, 1_9\}$. ($\cdot_{11} = \cdot =$ „missing value“.) $\rightsquigarrow H(A) = 3$, $h(A) = \frac{3}{11}$.

Sprechweise:

Die betrachtete Teilmenge A von E , der unser besonderes Interesse gilt, nennen wir hier auch **Zielmenge**.

Daher gilt:

Satz: **Vor.:**

Sei E ein eingetroffenes Ereignis eines Zufallsprozesses, A Zielmenge.

Beh.:

1. $H(A) = |A|$ (Absolute Häufigkeit = Mächtigkeit)
2. $h(A) = \frac{H(A)}{H(E)} = \frac{|A|}{|E|}$ (Relative Häufigkeit)
3. $h(E) = 1$

Unmittelbar erkennt man folgenden Sachverhalt:

Satz: $\forall_A : 0 \leq h(A) \leq 1$

4.3 Ereignisalgebra

4.3.1 Ereignisalgebra und Mengenalgebra

Da Ereignisse Mengen sind, ist Ereignisalgebra nichts anderes als Mengenalgebra.

Seien A und B Ereignisse (\rightsquigarrow Mengen), die bei einem geplanten Experiment eintreffen können.

Definition:

Summenereignis:

$A + B$ oder $A \cup B$:= Gesamtereignis, das genau dann eintrifft, wenn A oder B oder beide gemeinsam eintreffen. (\rightsquigarrow Vereinigung von A und B .)

Definition:

Produktereignis:

$A \cdot B$ oder $A \cap B$:= Gesamtereignis, das genau dann eintrifft, wenn A und B gemeinsam eintreffen. (\rightsquigarrow Durchschnitt von A und B .)

Sei dazu Ω die Menge aller Elementarergebnisse (Universalmenge, Summenereignis aller atomaren Ereignisse, trifft stets ein) und $\{\}$ die leere Menge (trifft nie ein).

Definition:

Ein stets eintreffendes Ereignis (Ω) heisst **sicheres Ereignis**. Ein Ereignis, das nie eintreffen kann ($\{\}$) heisst **unmögliches Ereignis**.

Definition:

Folgeereignis:

A heisst **Folgeereignis** von B , wenn das Eintreffen von B immer das von A zur Folge hat ($B \supseteq A$).

Definition:

Komplementäres Ereignis:

$\bar{A} = A^c$ heisst **Komplementäres Ereignis** von A , wenn gilt:
 $A \cup \bar{A} = \Omega \wedge A \cap \bar{A} = \{\}$.

Definition:

Äquivalente Ereignisse:

A und B heissen **äquivalent**, wenn das Eintreffen von A immer das von B zur Folge hat und umgekehrt. $\rightsquigarrow (B \supseteq A) \wedge (A \supseteq B)$.

Definition:

Ausschliessende Ereignisse:

A und B heissen **ausschliessend**, falls A und B nie gleichzeitig eintreffen können. ($A \cap B = \{\}$.)

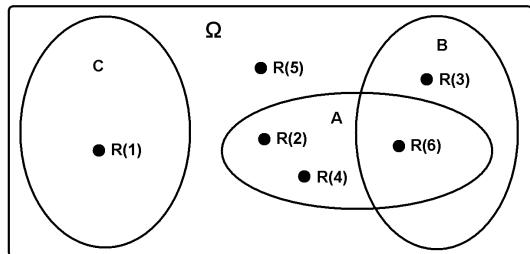
1. Beispiel:

Zufallsexperiment „ein Kind kommt zur Welt“: \rightsquigarrow Zwei Möglichkeiten: Ereignis $A = \{ \text{Es wird ein Junge} \}$. Ereignis $B = \{ \text{Es wird ein Mädchen} \}$.

$\rightsquigarrow A \cup B = \text{sicheres Ereignis}$, $A \cap B = \text{unmögliches Ereignis}$. A , B sind ausschliessend.

2. Beispiel:

$A = \{\text{würfeln einer geraden Zahl}\}$ (kurz $R(2 \vee 4 \vee 6)$), $B = \{\text{würfeln einer durch 3 teilbaren Zahl}\}$ (kurz $R(6 \vee 3)$), $C = \{\text{würfeln der Zahl 1}\}$ (kurz $R(1)$).



$$A \cup B = R(2 \vee 3 \vee 4 \vee 6)$$

$$A \cap B = R(6)$$

Bemerkung:

Eine Algebra ist eine Menge zusammen mit Operationen auf den Elementen. Die Potenzmenge von Ω bildet daher eine Algebra mit Operationen auf ihren Teilmengen, welche Mengen von Ereignissen sind. Die Operationen sind $\cup, \cap, \overline{\dots}, \subset$. Wir sprechen also hier von einer Ereignisalgebra.

4.3.2 Boolsche Algebren

Boolsche Algebren sind in Wirz, Bibl. A14 (Mathematikkurs für Ingenieure, Teil 4, Einführung in die Boolsche Algebra) besprochen. Wir resumieren hier nur die in diesem Kurs wichtige Anwendung bezüglich der Mengenalgebra.

Sei $\mathcal{A} \subset \mathcal{P}(\Omega)$, $\{\} \in \mathcal{A}$. ($\mathcal{P}(\Omega)$ = Potenzmenge)

Für unsere Bedürfnisse definieren wir:

Definition:

\mathcal{A} heisst **boolsche Algebra** (boolsche Mengenalgebra), wenn gilt:

1. $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$
2. $(A \in \mathcal{A}) \wedge (B \in \mathcal{A}) \Rightarrow (A \cup B) \in \mathcal{A}$

Es gilt (De Morgan!):

$$(A \in \mathcal{A}) \wedge (B \in \mathcal{A}) \Rightarrow (\overline{A} \in \mathcal{A}) \wedge (\overline{B} \in \mathcal{A}) \Rightarrow (\overline{A \cup B}) \in \mathcal{A} \Rightarrow \overline{\overline{A \cup B}} = (\overline{\overline{A}} \cap \overline{\overline{B}}) = A \cap B \in \mathcal{A}$$

Konsequenz: $(A \in \mathcal{A}) \wedge (B \in \mathcal{A}) \Rightarrow A \cap B \in \mathcal{A}$

Folgerungen:

1. $(A_i \in \mathcal{A}), i = 1, \dots, n \Rightarrow (\bigcup_{i=1}^n A_i) \in \mathcal{A}$
2. $(A_i \in \mathcal{A}), i = 1, \dots, n \Rightarrow (\bigcap_{i=1}^n A_i) \in \mathcal{A}$

Problem: Wenn Ω endlich ist, ist $\mathcal{P}(\Omega)$ ebenfalls endlich und daher problemlos zu bilden. $\mathcal{P}(\Omega)$ ist eine Menge von Teilmengen von Ω , also eine **Menge von Ereignissen**. Wie bildet man aber $\mathcal{P}(\Omega)$, wenn Ω unendlich ist? Und gilt dann $(\bigcup_{i=1}^{\infty} A_i) \in \mathcal{A}$ immer noch?

Wie oft in der Mathematik umgehen wir das Problem mit einer Definition:

Definition:

Eine boolsche Mengenalgebra heisst **σ -Algebra**, wenn gilt:
 $(A_i \in \mathcal{A}, i = 1, \dots, \infty) \Rightarrow (\bigcup_{i=1}^{\infty} A_i) \in \mathcal{A}$

Wegen De Morgan gilt dann wieder:

Folgerung:

$$(A_i \in \mathcal{A}, i = 1, \dots, \infty) \Rightarrow (\bigcap_{i=1}^{\infty} A_i) \in \mathcal{A}$$

Bemerkung:

Wenn Ω unendlich ist, ist \mathcal{A} im allgemeinen eine Teilmenge von $\mathcal{P}(\Omega)$.
 Denn die Ereignisse in \mathcal{A} sind normalerweise endliche Mengen, $\mathcal{P}(\Omega)$ besitzt hingegen in diesem Fall auch unendliche Teilmengen.

Konsequenz:

Zu einem Zufallsexperiment gehört von nun an immer ein Paar (Ω, \mathcal{A}) . (Menge der Ergebnisse Ω und Ereignisalgebra \mathcal{A} .)

4.3.3 Zur Mächtigkeit und Häufigkeit

Seien nun A und B als Zielmengen Teilmengen derselben Ergebnismenge E . Aus der Mengenalgebra können wir folgern:

Satz:

1. $H(A \cup B) = H(A) + H(B) - H(A \cap B)$
2. $h(A \cup B) = h(A) + h(B) - h(A \cap B)$
3. $A \cap B = \{\} \Rightarrow H(A \cup B) = H(A) + H(B)$ (ausschliessende Ereignisse)
4. $A \cap B = \{\} \Rightarrow h(A \cup B) = h(A) + h(B)$ (ausschliessende Ereignisse)
5. $\overline{A \cup B} = \overline{A} \cap \overline{B}$ (De Morgan)
6. $\overline{A \cap B} = \overline{A} \cup \overline{B}$ (De Morgan)
7. $h(\overline{A}) = 1 - h(A)$ (De Morgan)

Bsp.:

Versuch: 16 mal würfeln mit 2 Würfeln. \leadsto Ergebnis:

$$E = \{(2, 3)_1, (2, 5)_2, (1, 3)_3, (4, 6)_4, (1, 6)_5, (1, 6)_6, (2, 2)_7, (4, 6)_8, (4, 6)_9, (5, 6)_10, (1, 1)_11, (2, 6)_12, (2, 5)_13, (1, 1)_14, (1, 5)_15, (3, 4)_16\}$$

A : Mindestens eine Zahl ist durch 3 teilbar.

B : Beide Zahlen sind gerade.

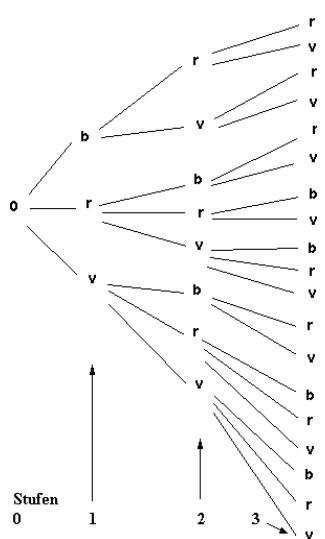
$$\leadsto H(E) = 16, \quad H(A) = 10, \quad H(B) = 5, \quad H(A \cap B) = 4, \quad H(A \cup B) = 11, \quad H(\overline{A}) = 6, \quad \dots$$

4.3.4 Ereignisbäume

Um die Übersicht über n -stufige Zufallsprozesse zu gewinnen, ist die Darstellung der Ereignisse durch **Ereignisbäume** günstig.

Bsp.:

Gegeben sei eine Urne, die eine blaue (b), zwei rote (r) und drei violette (v) Kugeln enthält. Es sollen in einem dreistufigen Zufallsprozess nacheinander drei Kugeln gezogen werden. Die Ereignisse sollen in einem Ereignisbaum dargestellt werden. Wieviele Möglichkeiten gibt es?

**Bsp.:**

$b \rightarrow r \rightarrow r$ ist eines der Ergebnisse.
Es gibt 19 verschiedene Ergebnisse (Variationen: Auswahl mit elementabhängiger Wiederholung und Anordnung).

Wichtig:

Da die Wiederholung der ausgewählten Elemente elementabhängig ist, kann keine der 6 klassischen Formeln der Kombinatorik direkt angewandt werden, um die Anzahl der Möglichkeiten zu berechnen.

4.4 Klassische Wahrscheinlichkeit nach Laplace

4.4.1 Laplace-Experiment, Gleichwahrscheinlichkeit

Im Folgenden studieren wir **Laplace-Experimente**. Das sind Experimente mit einer endlichen Ergebnismenge, bei denen man beobachtet, dass bei genügend grosser Anzahl Wiederholungen die Elementarereignisse alle ungefähr gleich grosse Häufigkeit aufweisen.

Z.B. beim Werfen einer Münze hat man beobachtet (Kreyszig, Bibl. A9, Statistische Methoden und ihre Anwendungen, p. 58):

| Experimentator | Anzahl Würfe | Eintreffen von Kopf | $h(R(Kopf))$ |
|----------------|--------------|---------------------|--------------|
| Buffon | 4040 | 2048 | 0.5069 |
| Pearson | 12000 | 6019 | 0.5016 |
| Pearson | 24000 | 12012 | 0.5005 |

Hier kann man demnach von einem Laplace-Experiment ausgehen. Man spricht auch von **statistischer Regelmässigkeit** oder von der **Stabilität der relativen Häufigkeit**. Ähnliche Erfahrungen macht man auch mit einem idealen Würfel.

Den Begriff des Laplace-Experiments können wir exakter fassen durch den Begriff der **Gleichwahrscheinlichkeit**:

Definition:

Ist bei einem Zufallsexperiment kein Ereignis vor einem andern ausgezeichnet, so heissen alle Ereignisse **gleichwahrscheinlich** oder **gleichmöglich**.

Folgerung:

Bei Laplace-Experimenten sind somit die Elementarereignisse **gleichwahrscheinlich**.

Bsp.: Laplace-Experimente: Kartenziehen, würfeln mit idealen Würfeln, ziehen von Kugeln aus einer Urne u.s.w..

4.4.2 Wahrscheinlichkeit als Mass für Gewinnchancen

Bsp.: Bei einem idealen Würfel sind alle 6 Ereignisse gleichwahrscheinlich (Laplace-Experiment).

Gedankengang von Laplace:

Problem: In einem Würfelspiel mit einem Würfel gewinnt ein Spieler, wenn er eine 3 oder eine 6 würfelt. Sonst verliert er. Was sind seine Gewinnchancen?

Beim genannten Spiel gibt es $m = 6$ mögliche Ereignisse (**mögliche Fälle**, Ereignismenge $\Omega = \{1, 2, 3, 4, 5, 6\}$, kurz geschrieben). Für den Gewinn günstig sind jedoch nur $g = 2$ Ereignisse (**günstige Fälle**, Ereignismenge $A = \{3, 6\}$). $m - g = 4$ Fälle sind demnach ungünstig für den Gewinn (Zahlen 1, 2, 4, 5).

$$\rightsquigarrow g = H(A) = m(A), \quad m = H(\Omega) = m(\Omega)$$

Da alle Ereignisse gleichwahrscheinlich sind, kann als Mass P für die Gewinnaussicht die relative Häufigkeit genommen werden:

$$P = \frac{g}{m} = \frac{H(A)}{H(\Omega)}$$

Dieses Mass für die Gewinnchance bei gleichwahrscheinlichen Elementarereignissen definiert man nun klassisch als **Wahrscheinlichkeit** für das Eintreffen von A

Definition:

Wahrscheinlichkeit (nach Laplace) $P(A)$ des Ereignisses A bei einem Experiment mit gleichwahrscheinlichen Elementarereignissen:

$$P(A) = \frac{g}{m} = \frac{H(A)}{H(\Omega)}$$

Bsp.: Zufallsprozess: Würfeln mit 3 Würfeln in 2 Versuchen (2-stufig). Gewonnen hat, wer mindestens einmal 2 6-er würfelt (Ereignis A). $\rightsquigarrow P(A) = ?$

Lösung:

Wir gliedern das Problem erst in Teilprobleme auf, deren Resultate wir dann anschliessend zusammenfügen.

- Das Ereignis A_1 sei definiert durch den Erfolg im 1. Versuch (Stufe 1), wobei das Resultat im 2. Versuch (Stufe 2) beliebig sein kann. Statt Würfel betrachten wir die Plätze P , auf die die Resultate notiert werden. Es gibt 3 Möglichkeiten, je mit zwei Würfeln eine 6 zu würfeln und mit dem jeweils 3. Würfel keine 6 zu würfeln ($E_{1,1}, E_{2,1}, E_{3,1}$). Dazu gibt es noch eine Möglichkeit, mit allen 3 Würfeln eine 6 zu würfeln ($E_{4,1}$).

$g_{j,k}$ = Anzahl günstige Möglichkeiten beim Ereignis $E_{j,1}$ am Platz Nummer k .

| Ereign. | Stufe 1 | | | Stufe 2 | | |
|-----------|------------------------|------------------------|------------------------|---------------|---------------|---------------|
| | $P = 1$ | $P = 2$ | $P = 3$ | $P = 4$ | $P = 5$ | $P = 6$ |
| $E_{1,1}$ | $g_{1,1} = 1$ | $g_{1,2} = 1$ | $\neg 6 : g_{1,3} = 5$ | $g_{1,4} = 6$ | $g_{1,5} = 6$ | $g_{1,6} = 6$ |
| $E_{2,1}$ | $g_{2,1} = 1$ | $\neg 6 : g_{2,2} = 5$ | $g_{2,3} = 1$ | $g_{2,4} = 6$ | $g_{2,5} = 6$ | $g_{2,6} = 6$ |
| $E_{3,1}$ | $\neg 6 : g_{3,1} = 5$ | $g_{3,2} = 1$ | $g_{3,3} = 1$ | $g_{3,4} = 6$ | $g_{3,5} = 6$ | $g_{3,6} = 6$ |
| $E_{4,1}$ | $g_{4,1} = 1$ | $g_{4,2} = 1$ | $g_{4,3} = 1$ | $g_{3,4} = 6$ | $g_{3,5} = 6$ | $g_{3,6} = 6$ |

A_1 liegt vor, wenn E_1 oder E_2 oder E_3 oder E_4 vorliegen.

$A_1 = E_1 \cup E_2 \cup E_3 \cup E_4 \Rightarrow m(A) = g = 5 \cdot 6 \cdot 6 \cdot 6 + 5 \cdot 6 \cdot 6 \cdot 6 + 5 \cdot 6 \cdot 6 \cdot 6 + 1 \cdot 6 \cdot 6 \cdot 6 = (3 \cdot 5 + 1) \cdot 6^3$.
Da auf jedem Platz 6 Zahlen möglich sind, gilt: $m = 6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 = 6^6$.

$$\Rightarrow P(A_1) = h(A_1) = \frac{m(A_1)}{E} = \frac{(3 \cdot 5 + 1) \cdot 6^3}{6^6} = \frac{16}{6^3} = \frac{8}{3 \cdot 6^2}$$

2. Das Ereignis A_2 sei definiert durch den Erfolg im 2. Versuch, wobei das Resultat im 1. Versuch beliebig sein kann. Aus Symmetriegründen erhalten wir dasselbe Resultat wie im 1. Versuch.

$$\Rightarrow P(A_2) = P(A_1) = h(A_2) = h(A_1) = \frac{8}{3 \cdot 6^2}$$

3. Sei $A_3 = A_1 \cap A_2$. A_3 bedeutet, dass der Erfolg sich im 1. und im 2. Versuch (Stufe) einstellen.

Für den Erfolg auf den Plätzen 1, 2, 3 (1. Stufe) gibt es $3 \cdot 5 + 1 = 16$ Möglichkeiten. Zu jeder dieser 16 Möglichkeiten gibt es wieder 16 Möglichkeiten für den Erfolg auf den Plätzen 4, 5, 6 (2. Stufe). Somit gibt es für den Erfolg auf beiden Stufen (A_3) $16 \cdot 16$ Möglichkeiten.

$$\Rightarrow P(A_3) = P(A_1 \cap A_2) = h(A_1 \cap A_2) = \frac{m(A_3)}{E} = \frac{16 \cdot 16}{6^6} = \frac{8 \cdot 8}{3^2 \cdot 6^4}$$

$$4. P(A) = P(A_1 \cup A_2) = h(A_1 \cup A_2) = h(A_1) + h(A_2) - h(A_1 \cap A_2) = \frac{8}{3 \cdot 6^2} + \frac{8}{3 \cdot 6^2} - \frac{8 \cdot 8}{3^2 \cdot 6^4} = \frac{2 \cdot 8}{3 \cdot 6^2} + \frac{8 \cdot 8}{3^2 \cdot 6^4} = \frac{47}{324} \approx 0.145062$$

Achtung:

Der Begriff der Gleichwahrscheinlichkeit versagt jedoch in der Praxis meistens: Würfel sind nicht exakt symmetrisch und homogen, die Psychologie des Werfers spielt hinein, und bei vielen statistischen Vorgängen liegen überhaupt keine Laplace-Experimente vor. Man denke z.B. nur an das Wetter! Daraus ist ein erweiterter Wahrscheinlichkeitsbegriff notwendig. Wir führen daher die **Wahrscheinlichkeit axiomatisch** ein nach **Kolmogoroff**.

4.5 Axiomatischer Wahrscheinlichkeitsbegriff

4.5.1 Begriff, Axiome, Folgerungen

Ausgangspunkt ist die Erfahrung: Ergebnisse von Zufallsexperimenten unterliegen auf die Dauer gewissen Gesetzmäßigkeiten. Nach Descartes ist es vernünftig, das einfachste und plausibelste Modell des beobachteten Vorgangs (Naturvorgangs) als Gesetzmäßigkeit zu akzeptieren, sofern nicht weitere Beobachtungen ein verändertes Modell erzwingen. Ein Modell in diesem Sinne ist jedoch keine kausale Naturerklärung sondern nur eine Naturbeschreibung.

Im Folgenden wollen wir uns auf Experimente oder Vorgänge einschränken, bei denen sich die relativen Häufigkeiten mit der grösser werdenden Anzahl Versuchen statistisch immer stabiler verhalten. Dabei muss jedoch keine Gleichwahrscheinlichkeit der Elementarereignisse vorausgesetzt werden. Zufällige Unregelmäßigkeiten sind nur für die Elementarereignisse, nicht aber für die von der Versuchsnummer abhängige relative Häufigkeit erlaubt, wenn die Versuchsnummer gross ist.

Grundannahme:

Auf der Grundlage der empirischen Beobachtungen lässt sich folgende **Annahme** vertreten: Sei n die Anzahl Versuche in einem Experiment unter gleichbleibenden Bedingungen. Dann ist $\lim_{n \rightarrow \infty} h_n(A) = h_\infty$ eine mit praktischer Gewissheit angebbare reelle Zahl $\in [0, 1]$.

Achtung! Da es in der Realität der Zeit unendlich viele Versuche nie geben kann, ist diese Annahme praktisch nicht überprüfbar. Sie ist rein theoretischer Natur. An die Stelle der praktischen Gewissheit tritt hier der praktische Erfolg mit der Annahme. Darauf beruhende Statistik ist demnach in der Grundlage keine deduktive, sondern eine induktive (experimentelle) Wissenschaft.

Bemerkung:

Um h_∞ zu gewinnen, ist man auf Daten, d.h. auf Erfahrung angewiesen.

Die nun verwendete Wahrscheinlichkeit beruht auf einer hypothetischen Interpretation, die wir in einer Definition fassen. Sie ist wieder als Mass für ein Ereignis aufzufassen.

Definition:

Wahrscheinlichkeit oder **Wahrscheinlichkeitsmass** (statistische Definition):

$$P : \mathcal{A} \mapsto [0, 1] \subset \mathbb{R}$$

$$A \mapsto P(A) := h_\infty = \lim_{n \rightarrow \infty} \frac{H(A)}{H(\Omega)}$$

Konsequenz: P ist eine Funktion auf \mathcal{A} .

Bemerkung:

Da es keine Methode geben kann um in der praktischen Realität h_∞ zu gewinnen, setzen wir $P(A) := h_\infty \approx h_n$, $n \gg 1$ (n sehr gross).

~ **Problem:** Stabilität von h_n mit n .

Da Ω in der Praxis oft nicht bekannt ist, schreiben wir nun S statt Ω .

Axiome:

1. $P(A) \in [0, 1]$ eindeutig bestimmt
2. (a) $A = S \Rightarrow P(A) = 1$ (sicheres Ereignis)
- (b) A und B äquivalente Ereignisse $\Rightarrow P(A) = P(B)$
- (c) A und B ausschliessende Ereignisse:
 $A \cap B = \{\} \Rightarrow P(A \cup B) = P(A) + P(B)$

Definition:

$\bar{A} = S \setminus A$ heisst **komplementäres Ereignis** von A .

Bemerkung:

1. $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$ wegen $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ (ausschliessend!)
 $\rightsquigarrow P(A \cup B) \leq P(A) + P(B)$
2. Sei $A \subseteq B \Rightarrow B = A \cup (B \setminus A)$
 $\Rightarrow P(B) = P(A) + P(B \setminus A) \Rightarrow P(B) \geq P(A)$
3. $A \cap \bar{A} = \{\} \Rightarrow 1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$
 $\Rightarrow P(A) = 1 - P(\bar{A})$
4. $P(S) = P(\bar{\bar{S}}) = 1 - P(\bar{S}) = 1 - P(\{\}) = 1 - 0 = 1$

Nachstehende Folgerungen sind nun unmittelbar einsichtig:

Folgerungen:

1. $P(A) = 1 - P(\bar{A})$
2. A_1, A_2, \dots, A_n ausschliessend
 $\Rightarrow P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$
3. A unmöglich ($A = \bar{S} = \{\}$) $\Rightarrow P(A) = 0$
4. A, B beliebig $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5. A, B beliebig $\Rightarrow P(A \cup B) \leq P(A) + P(B)$
6. $A \subseteq B \Rightarrow P(A) \leq P(B)$
7. $P(S) = 1$

Achtung:

1. $0 = P(A) = h_\infty(A) \approx h_n(A), n \gg 1$
 $\rightsquigarrow A$ trifft praktisch nie ein, was aber nicht bedeutet, dass A unmöglich ist!
 \rightsquigarrow
2. $1 = P(A) = h_\infty(A) \approx h_n(A), n \gg 1$
 $\rightsquigarrow A$ trifft praktisch immer ein, was aber nicht bedeutet, dass A absolut sicher ist!
 \rightsquigarrow

4.5.2 Der Begriff Wahrscheinlichkeitsraum

Wir betrachten eine diskrete Ereignismenge $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$ eines Zufallsexperiments. Für die Elementarereignisse ω_i sei $P(\omega_i) = p_i \geq 0$. Für $i \neq k$ muss $\omega_i \cap \omega_k = \{\}$ gelten, sonst wären die Elementarereignisse zerlegbar ($\omega_i = (\omega_i \cap \omega_k) \cup (\omega_i \setminus \omega_k)$). Daher folgt aus den Axiomen: $\sum_{i_1}^{\infty} p_i = 1$.

Da sich jedes Ereignis aus Elementarereignissen zusammensetzt ($A = \bigcup_{i_1}^{\infty} \omega_i$), gilt für die Wahrscheinlichkeit von A : $P(A) = \sum_{i_1}^{\infty} p_i$. Damit ist jedem Ereignis $A_k = \bigcup_{k_1}^{\infty} \omega_i$ eindeutig eine Wahrscheinlichkeit $P(A_k) = \sum_{k_1}^{\infty} p_k$ zugeordnet. Diese Zuordnung $A_k \mapsto P(A_k)$ ist eine Funktion. Diese Funktion macht den

Ereignisraum zu einem „**Wahrscheinlichkeitsraum**“.

Definition:

Ein **Wahrscheinlichkeitsraum** ist gegeben durch einen Ereignisraum Ω , eine Ereignisalgebra und ein Wahrscheinlichkeitsmass.

Man sieht sofort, dass im Falle eines endlichen Ereignisraumes $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ mit $P(\omega_1) = P(\omega_2) = \dots = P(\omega_n)$ die Laplace-Wahrscheinlichkeit herauskommt.

Sprechweise:

Im Falle der Laplace-Wahrscheinlichkeit reden wir von einem **Laplace-Wahrscheinlichkeitsraum**.

4.5.3 Bedingte Wahrscheinlichkeit

Der Begriff

Bsp.:

Seien A und B die folgenden Ereignisse:

$$A = \{R(\text{Kunde ist weiblich})\}$$

$$B = \{R(\text{Kunde kauft unser Produkt})\}$$

$$\leadsto B|A = \{R(\text{Weiblicher Kunde kauft unser Produkt})\}$$

Sprechweise:

$B|A = \leadsto$ Ereignis B unter der Voraussetzung von Ereignis A , B abhängig von A .

Bemerkung:

$B|A \leadsto$ Kunde ist weiblich **und** Kunde kauft \leadsto Schnittmenge!

Problem: Wir fragen nun nach der Wahrscheinlichkeit, dass der Kunde unser Produkt kauft, unter der Voraussetzung, dass der Kunde weiblich ist.

Definition:

$P(B|A) :=$ Wahrscheinlichkeit für das Eintreffen von B unter der Voraussetzung, dass A eingetroffen ist \leadsto **bedingte Wahrscheinlichkeit**.

Zur Berechnung von $P(B|A)$ betrachten wir zuerst den Fall einer endlichen Ereignismenge Ω , $A \neq \{\}$:

$$\text{Seien } P(A) := \frac{H(A)}{H(\Omega)} = \frac{a}{n}, \quad P(B) := \frac{H(B)}{H(\Omega)} = \frac{b}{n}, \quad P(A \cap B) := \frac{H(A \cap B)}{H(\Omega)} = \frac{s}{n}$$

Die Ergebnisse von $A \cap B$ treten in s Fällen ein. Dies geschieht unter der Bedingung, dass die Ergebnisse von A eingetroffen sind, was in a Fällen geschieht. A ist somit die Menge der sicheren Ereignisse bezüglich dieses Problems, d.h. Universalmenge für das Ereignis $A \cap B$. Daher gilt:

$$P(B|A) = \frac{H(A \cap B)}{H(A)} = \frac{s}{a} = \frac{\frac{s}{n}}{\frac{a}{n}} = \frac{P(A \cap B)}{P(A)}$$

Wegen $A \cap B = B \cap A$ ist obiges Gesetz symmetrisch: $P(A|B) = \frac{P(B \cap A)}{P(B)} = \frac{P(A \cap B)}{P(B)} \leadsto$

Satz:

$$\begin{aligned} 1. \quad P(B|A) &= \frac{P(A \cap B)}{P(A)} & P(A|B) &= \frac{P(B \cap A)}{P(B)} \\ 2. \quad H(A) = H(B) \Rightarrow P(A|B) &= P(B|A) \end{aligned}$$

Konsequenz: $P(A \cap B) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$
(Multiplikationssatz)

Z.B. für $A = \{\}$ ist der Sachverhalt auch richtig.
($\Rightarrow A \cap B = \{\} \Rightarrow P(A \cap B) = 0, P(A) = 0, P(B|A) = 0$)

$P(A \cap B) = P(B|A) \cdot P(A)$ lässt sich iterieren:

Korollar:

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

1. Beispiel:

Seien $|\Omega| = 100, |A \cup B| = 80, |A| = 50, |B| = 45$

Es gilt: $|A \cup B| = |A| + |B| - |A \cap B|$

$$\Rightarrow |A \cap B| = 50 + 45 - 80 = 15, P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{15/100}{50/100} = \frac{15}{50} = 0.3$$

2. Beispiel:

Gegeben: Schachtel mit 3 grünen und 7 roten Kugeln (\rightsquigarrow total 10). Wahrscheinlichkeit, dass bei 2 Zügen ohne Zurücklegen nur rote Kugeln kommen?

Sei r_1 = „rot im 1. Zug“, r_2 = „rot im 2. Zug“.

$$\rightsquigarrow P(\{r_1\}) = \frac{7}{10}, P(\{r_2\}|\{r_1\}) = \frac{6}{9} \Rightarrow P(\{r_1\} \cap \{r_2\}) = P(\{r_2\}|\{r_1\}) \cdot P(\{r_1\}) = \frac{6}{9} \cdot \frac{7}{10} = \frac{42}{90} \approx 0.47$$

3. Beispiel:

Lotto: Wahrscheinlichkeit für 6 richtige aus 49?

1. Methode: Kombinationen

$$P = \frac{1}{\binom{49}{6}} = \frac{(49-6)! \cdot 6!}{49!} = \frac{6!}{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44} = \frac{1}{13983816} \approx 7.15112 \cdot 10^{-8}$$

2. Methode: Abhängige Ereignisse

$$A_1 \rightsquigarrow 1. \text{ Zahl ziehen: } P(A_1) = \frac{6}{49}$$

$$A_2 \rightsquigarrow 2. \text{ Zahl ziehen unter der Voraussetzung, dass die 1. Zahl gezogen ist: } P(A_2|A_1) = \frac{5}{48}$$

$$A_3 \rightsquigarrow 3. \text{ Zahl ziehen unter der Voraussetzung, dass die 1. und die 2. Zahl gezogen sind:}$$

$$P(A_3|A_1 \cap A_2) = \frac{4}{47}$$

\vdots

$$\rightsquigarrow P(A_1 \cap A_2 \cap \dots \cap A_6) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \dots P(A_6|A_1 \cap A_2 \cap \dots \cap A_5) \\ = \frac{6}{49} \cdot \frac{5}{48} \cdot \frac{4}{47} \dots \frac{1}{44} = \frac{1}{13983816} \approx 7.15112 \cdot 10^{-8}$$

Unabhängige Ereignisse

Falls die Ergebnisse oder Elemente in der Menge $B|A$ dieselben sind wie diejenigen in der Menge B und somit $B|A = B$ gilt, hat A keinen Einfluss darauf, welche Elemente von $B|A$ bei der Auswahl aus B weggelassen werden müssen. A verändert also B nicht. Daher nennen wir B und A unabhängig. Dann wird $P(B|A) = P(B)$ und daher

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

Man kontrolliert sofort nach, dass das Ergebnis auch richtig ist für $A = \{\}$ oder $B = \{\}$.

Definition: B heisst von A **unabhängig** oder **stochastisch unabhängig**, wenn gilt: $B|A = B$.

Satz:

Vor.:

Sei B von A unabhängig.

Beh.:

1. $P(B|A) = P(B)$
2. $P(A \cap B) = P(A) \cdot P(B)$
3. A von B unabhängig.

1. Beispiel:

Gegeben: Kartenspiel mit 36 Karten. Ziehe 2 mal eine Karte und lege sie zurück. Wahrscheinlichkeit, dass 2 Asse kommen?

Lösung:

a_1 : As beim 1. Zug.

a_2 : As beim 2. Zug.

Die beiden Ereignisse $A_1 = \{a_1\}$ und $A_2 = \{a_2\}$ sind unabhängig. $\rightsquigarrow P(A_1) = P(A_2) \rightsquigarrow P(A_1) = P(A_2)$

$$\rightsquigarrow P(A_1) = \frac{4}{36}, \quad P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = \frac{4}{36} \cdot \frac{4}{36} = \frac{1}{81}$$

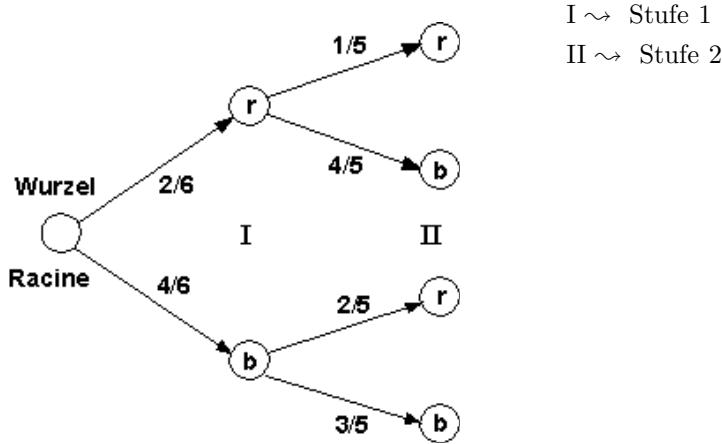
Benutzung des Ereignisbaums

2. Beispiel:

Urne mit 6 Kugeln: 2 rote (r), 4 blaue (b). Ziehe 2 mal eine Kugel ohne zurücklegen.

1. A: Wahrscheinlichkeit, 2 gleichfarbige Kugeln zu ziehen?
2. B: Wahrscheinlichkeit, 2 verschiedenfarbige Kugeln zu ziehen?

Lösung: Benutze einen Ereignisbaum:

**Wichtig: Methode**

Beachte: Den Pfeilen (Pfaden) entlang handelt es sich um abhängige Ereignisse, d.h. um Bedingte Wahrscheinlichkeit. An einem Verzweigungspunkt wird eine Bedingung gesetzt für die folgende Stufe. Daher multiplizieren sich die Wahrscheinlichkeiten auf den Pfaden. Hingegen sind die Ereignisse zu einem gegebenen Niveau ausschliessend. Die Wahrscheinlichkeiten addieren sich.

Lösung:

$$\rightsquigarrow P(A) = \frac{2}{6} \cdot \frac{1}{5} + \frac{4}{6} \cdot \frac{3}{5} = \frac{7}{15}, \quad P(B) = \frac{2}{6} \cdot \frac{4}{5} + \frac{4}{6} \cdot \frac{2}{5} = \frac{8}{15}, \quad P(A) + P(B) = 1$$

3. Beispiel:

Wie gross ist die Wahrscheinlichkeit, beim k -ten Wurf mit einer idealen Münze erstmals Zahl zu treffen? Was ist die Summe über k aller solcher Werte für die Wahrscheinlichkeit?

$$P(A_k) = \left(\frac{1}{2}\right)^k, \quad \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \left(\sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k\right) - 1 = \frac{1}{1 - \frac{1}{2}} - 1 = 2 - 1 = 1$$

4.5.4 Totale Wahrscheinlichkeit**Bsp.:**

Bei einem Zufallsexperiment ziehen wir aus einer von drei gegebenen Urnen U_i eine Kugel. U_1 enthält zwei rote und zwei blaue, U_2 zwei rote und eine blaue, U_3 eine rote und zwei blaue.

 \rightsquigarrow Fragen:

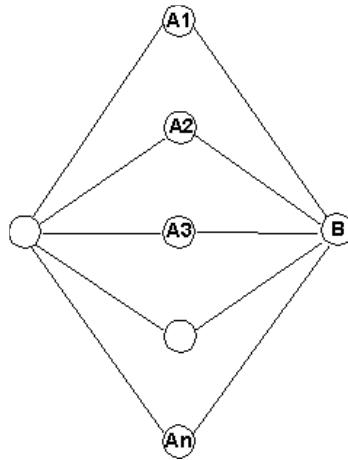
1. Wie gross ist die Wahrscheinlichkeit, eine rote Kugel zu ziehen?
2. Wie gross ist die Wahrscheinlichkeit, dass die gezogene rote Kugel aus U_2 stammt?

Lösung:

1. Niveau 1: $P(\{U_i\}) = \frac{1}{3} \forall_i$
2. Niveau 2: $P(\{r\}|\{U_1\}) = \frac{2}{4}, \quad P(\{r\}|\{U_2\}) = \frac{2}{3}, \quad P(\{r\}|\{U_3\}) = \frac{1}{3}$
3. Niveau 1 \rightarrow 2: $P(\{r \wedge U_1\}) = \frac{1}{3} \cdot \frac{2}{4} = \frac{1}{6}, \quad P(\{r \wedge U_2\}) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}, \quad P(\{r \wedge U_3\}) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$

$$4. \{r\}: P(\{r\}) = P(\{r \wedge U_1\}) + P(\{r \wedge U_2\}) + P(\{r \wedge U_3\}) = \sum_i P(\{U_i\}) \cdot P(\{r\}|U_i) \\ = \frac{1}{6} + \frac{2}{9} + \frac{1}{9} = \frac{1}{2} = 0.5$$

Selbstverständlich muss man bei diesem Experiment den Wert 0.5 erwarten, da total genau die Hälfte der Kugeln rot sind. Doch das Beispiel zeigt die Methode, die auch in andern Fällen funktioniert.



Setzen wir B an Stelle von $\{r\}$ und A_i an Stelle von U_i , so erhalten wir die folgende Formel:

Konsequenz: Seien $i \neq k$

$$\Rightarrow A_i \cap A_k = \emptyset \wedge A = (A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B), \quad P(B) = \sum_{i=1}^n P(A_i) \cdot P(B|A_i)$$

Definition:

$P(B)$ heisst hier die **totale Wahrscheinlichkeit** für das Eintreten des Ereignisses B .

Nun interessieren wir uns für die Wahrscheinlichkeit, dass der Pfad über A_k führt unter der Voraussetzung, dass B eingetreten ist. Die Wahrscheinlichkeit, dass der Pfad über A_k führt, ist dabei gleich der Wahrscheinlichkeit, dass A_k eintritt. Wir müssen demnach $P(A_k|B)$ berechnen. Gegeben sind aber $P(A_i)$ und $P(B|A_i)$.

Wir haben folgende Informationen zur Verfügung:

$$1. \quad P(B) = \sum_i P(A_i) \cdot P(B|A_i) \quad (\text{eben gewonnen})$$

$$2. \quad P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} \quad (\text{Multiplikationssatz})$$

$$3. \quad P(A_k \cap B) = P(A_k) \cdot P(B|A_k) \quad (\text{entlang einem Pfad})$$

Konsequenz: (Bayes'sche Formel)

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{\sum_i P(A_i) \cdot P(B|A_i)}$$

$$\rightsquigarrow P(\{U_2\}|\{r\}) = \frac{P(\{U_2\}) \cdot P(\{r\}|U_2)}{\sum_{i=1}^3 P(\{U_i\}) \cdot P(\{r\}|U_i)} = \frac{\frac{1}{3} \cdot \frac{2}{3}}{0.5} = \frac{2}{9 \cdot 0.5} = \frac{4}{9}$$

4.6 Wahrscheinlichkeitsverteilungen

4.6.1 Zufallsvariablen

Wir betrachten Zufallsexperimente. Wenn wir aus einer Urne farbige Kugeln ziehen, so müssen wir der Farben erst eine Skala zuordnen, um sie auf der Achse eines Diagramms eintragen zu können und nachher mit Wahrscheinlichkeitsfunktionen als Zahlenfunktionen arbeiten zu können. Wenn wir hingegen mit einem Würfel würfeln, so sind die möglichen Resultate bereits durch Zahlen gegeben $\in \{1, 2, \dots, 6\}$.

Reelle Variablen können wir wie folgt als Funktionen interpretieren:

$$x : z \longmapsto x(z) = z.$$

Die Funktion x ordnet der Zahl z den Wert $z = w(z)$ zu, was selbstverständlich immer zutrifft und daher nie so erwähnt wird.

In analoger Weise wird nach der Absicht im Experiment durch eine **Zufallsvariable** einem Ergebnis $\omega \in \Omega$ (oder einem Elementarereignis $\{\omega\} \in \mathcal{A}$) ein Wert $X(\omega) \in \mathbb{R}$ zugeordnet.

1. Beispiel:

In einem Zufallsexperiment wird mit zwei Würfeln gewürfelt. Die Ergebnismenge ist $\Omega = \{(m, n) \mid m, n \in \{1, 2, \dots, 6\}\}$. Die Zufallsvariable X sei nun wie folgt definiert: $X((m, n)) = m + n \in \mathbb{R}$. Es gilt dann: $X((m, n)) \in \{2, 3, \dots, 12\}$. Die Menge der Ereignisse mit $X((m, n)) < 4$ ist dann $\{(1, 1), (1, 2), (2, 1)\}$.

2. Beispiel:

In einem Zufallsexperiment wird mit einem Würfel gewürfelt. Die Ergebnismenge ist $\Omega = \{1, 2, \dots, 6\}$. Die Zufallsvariable X sei nun wie folgt definiert: $X(n) = n$. Es gilt dann: $X(\omega_k) = \omega_k \in \mathbb{R}$.

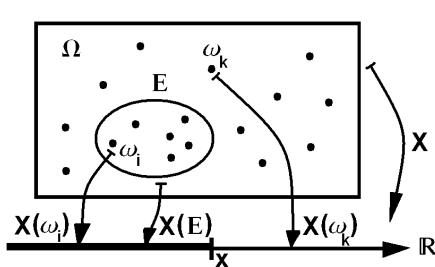
3. Beispiel:

In einem Zufallsexperiment werden Damen eines gegebenen Kollektivs nach dem Körperumfang ihres Traumpartners befragt. Für die Zufallsvariable gilt dann:

$$X : \text{Dame} \longmapsto X(\text{Dame}) = \text{Umfang Traumpartner} \in \mathbb{R}$$

Bemerkung:

Im 1. Beispiel ist $\{X(\dots)\} \subset \mathbb{R}$ diskret oder endlich, im 3. Beispiel ist $\{X(\dots)\} \subset \mathbb{R}$ theoretisch unendlich.



Für unseren weiteren Gebrauch definieren wir nun die Zufallsvariable etwas genauer. Gegeben sei ein Zufallsexperiment mit einer Fundamentalmenge Ω und einer Ereignisalgebra \mathcal{A} , die eine σ -Algebra ist.

Definition:

Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ mit $\omega \mapsto X(\omega)$ heisst **Zufallsvariable** oder **stochastische Variable**, **stochastische Funktion**, wenn gilt:

1. $\forall_{x \in \mathbb{R}} \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{A}$
2. Zu \mathcal{A} ist eine Wahrscheinlichkeitsfunktion P definiert, die den Wahrscheinlichkeitsaxiomen genügt.

Die erste Bedingung ist Messbarkeitsbedingung. Sie bedeutet, dass die Menge aller Ergebnisse, deren Realisation unterhalb eines bestimmten Wertes liegt, ein Ereignis bilden muss.

Bemerkung:

Da es sich bei $\{\omega\}$ und somit auch bei $\{X(\omega)\}$ um Zufallsereignisse handelt, sagen wir: Der Wert von X **hängt vom Zufall ab**.

Schreibweise: $X \leq x := \{\omega \in \Omega \mid X(\omega) \leq x\}$
 $\rightsquigarrow X \leq x, x < X$ u.s.w. sind Mengen!

Sprechweise: Z.B. für $a \leq X \leq b$ sagen wir: X nimmt einen Wert aus $[a, b]$ an.

Da Teilmengen von Ω Ereignisse sind, so folgt auf Grund der Definition von X und der Auswahl der Elemente sofort der Satz:

Satz:

1. $-\infty < X < \infty$ ist das sichere Ereignis.
2. Folgende Mengen sind immer Ereignisse: $a < X < b, a \leq X < b, a < X \leq b, a \leq X \leq b, a < X, a \leq X, X < a, x \leq a$
3. $P(X \leq a) + P(a < X) = P(-\infty < X < \infty) = 1$

Konsequenz: $P(a < X) = 1 - P(X \leq a)$

Bemerkung:

Wie wir oben in den Beispielen gesehen haben, müssen wir unterscheiden zwischen **diskreten** und **kontinuierlichen Zufallsvariablen**. Es ist zu erwarten dass bei uns im Falle der kontinuierlichen Zufallsvariablen die Wertemengen Intervalle sind. Aus Gründen der Mächtigkeit von Intervallen und der Messgenauigkeit macht in diesem Falle eine Aussage wie z.B. $P(X = a)$ praktisch nicht viel Sinn, im Gegensatz zu einer Aussage wie $P(X < a)$.

3. Beispiel: Würfeln mit einem Würfel, $X = \text{Zahl}$:

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

$$\Rightarrow P(1 < X < 2) = 0, P(1 \leq X < 6) = \frac{5}{6}, P\left(\frac{3}{2} < X < \frac{5}{2}\right) = \frac{1}{6}$$

4. Beispiel: Urne, 4 rote und 6 blaue Kugeln. 2 ziehen ohne zurücklegen. $X(\text{gezogene Kugeln}) = \text{Anzahl gezogene rote Kugeln}$.

$$\begin{aligned}
P(X = 0) &= \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3} \\
P(X = 1) &= ? \quad r_1 \rightsquigarrow P(Z_1) := P(X = 1|r_1) = \frac{4}{10} \cdot \frac{6}{9} = \frac{4}{15} \\
&\quad r_2 \rightsquigarrow P(Z_2) := P(X = 1|r_2) = \frac{6}{10} \cdot \frac{4}{9} = \frac{4}{15} \\
Z_1 \cap Z_2 &= \{\} \Rightarrow P(Z_1 \cup Z_2) = \frac{4}{15} + \frac{4}{15} = \frac{8}{15} \\
P(X = 2) &= \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15} \\
\rightsquigarrow P((X = 0) \vee (X = 1) \vee (X = 2)) &= \frac{1}{3} + \frac{8}{15} + \frac{2}{15} = 1 \\
P(1 < X < 2) &= 0 \\
P(X \leq 1) &= \frac{1}{3} + \frac{8}{15} = \frac{13}{15} \\
P(X \geq 1) &= \frac{8}{15} + \frac{2}{15} = \frac{2}{3} \\
P(0.5 < X < 10) &= P(x \geq 1) = \frac{2}{3}
\end{aligned}$$

4.6.2 Verteilungsfunktion

Anlässlich der Erklärung der Zufallsvariable ist auf den Begriff der Wahrscheinlichkeitsfunktion verwiesen worden: Zu \mathcal{A} ist eine Wahrscheinlichkeitsfunktion P definiert, die den Wahrscheinlichkeitsaxiomen genügt. Über das methodische Vorgehen zur Gewinnung von P ist aber nichts Allgemeines gesagt worden. Daher wollen wir hier zuerst den Begriff der Wahrscheinlichkeitsverteilung präzisieren:

Gegeben sei ein Zufallsexperiment und dazu eine Zufallsvariable X . P sei nun praktisch so gegeben, dass zu jedem $x \in \mathbb{R}$ die Wahrscheinlichkeit $P(X \leq x)$ exakt definiert ist. Wichtig ist dabei die Einsicht, dass $P(X \leq x)$ für diskrete und für kontinuierliche Variablen problemlos vernünftig definierbar ist, während z.B. $P(X = x)$ im kontinuierlichen Fall wenig praktischen Sinn macht.

Definition: $F(x) = P(X \leq x)$ heisst **Verteilungsfunktion** von X .

Bemerkung:

1. Nach Konstruktion ist $D_F = \mathbb{R}$.
2. Mit Hilfe der Verteilungsfunktion können wir die „Wahrscheinlichkeitsfunktion“ oder „Wahrscheinlichkeitsverteilung“ im diskreten und im kontinuierlichen Fall auf verschiedene Arten definieren.

Nach Definition ist die Verteilungsfunktion eine Wahrscheinlichkeit. Daher kann man ablesen:

1. $\forall_x : 0 \leq F(x) \leq 1$
2. $F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$
3. $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$
4. $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
5. $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

4.6.3 Verteilungstypen

Problem: Die Ausgangsfrage ist: Wie verteilen sich bei Zufallsexperimenten die verschiedenen Wahrscheinlichkeiten auf die verschiedenen Ereignisse? Vom Resultat her kennen wir schon diskrete und kontinuierliche Zufallsvariablen. Man kann aber auch die zugehörigen Experimente unterscheiden. Abgesehen von den Mischfällen unterscheiden wir zwei Klassen:

Experimentklassen:

1. **Zählexperimente** \rightsquigarrow diskrete Verteilung
2. **Messexperimente** \rightsquigarrow kontinuierliche Verteilung

1. Beispiel: Volkszählung:

Verteilung der Bewohner auf die Gemeinden $\rightsquigarrow H(X = a) = \text{Anzahl Bewohner der Gemeinde } a$.

Hier wird die Verteilung durch eine Wahrscheinlichkeitsfunktion $\frac{H(X = a)}{H(S)} = P(X = a)$ beschrieben.

2. Beispiel: Gewichtsfunktion:

Verteilung des Gewichts auf Balken gleichen Querschnitts $\rightsquigarrow H(X \leq x) = \int_a^x f(t) dt = \text{Anzahl Balken der Länge } x \in [a, b]$.

Hier wird die Verteilung durch eine Wahrscheinlichkeitsdichtefunktion $f(x)$ beschrieben:

$$\rightsquigarrow f(x) = \frac{d \left(\int_a^x f(t) dt \right)}{dx}. f \text{ liefert ein Modell für die Grundgesamtheit, die oft gar nicht genau bekannt ist.}$$

Zwischen Grundgesamtheit und Modell können wir folgenden Vergleich anstellen:

| Grundgesamtheit | Modell |
|--|--|
| 1. Wirklichkeit | 1. Theorie |
| 2. Praxis: Häufigkeitsverteilung oder Dichte | 2. Wahrscheinlichkeitsverteilung |
| 3. Oft nur Stichproben | 3. Problem der Bonität des Modells (Tests) |

4.6.4 Diskrete Verteilung

Wahrscheinlichkeitsfunktion

Für diskrete Variablen sind endliche Definitionsbereiche nicht zwingend. Wir definieren daher:

Definition:

Eine Zufallsvariable X heisst **diskret**, wenn gilt:

1. X kann nur endlich oder abzählbar unendlich viele Werte x_i mit $P(X = x_i) \neq 0$ annehmen.
2. $\{x_i \mid P(X = x_i) \neq 0\}$ hat keine Häufungspunkte.

Konsequenz:

In jedem endlichen Intervall $\subset \mathbb{R}$ liegen nur endlich viele Werte x_i mit $P(X = x_i) \neq 0$.

Bezeichnungen:

Statt $P(X = x_i)$ schreiben wir kurz $P(x_i)$. Die Werte x mit $P(x) \neq 0$ seien nummeriert: x_1, x_2, x_3, \dots . Die zugehörigen Wahrscheinlichkeiten schreiben wir ebenfalls kurz: $P(X = x_i) := P(x_i) := p_i$.

Definition:

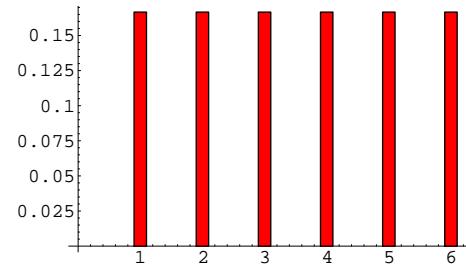
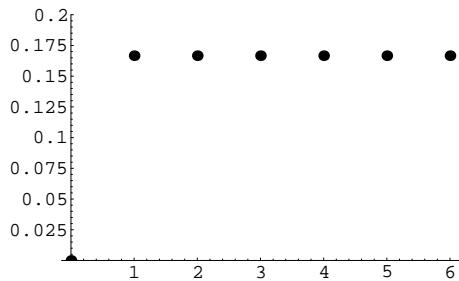
$$f(X) = \begin{cases} p_i & X = x_i \\ 0 & \text{sonst} \end{cases}$$

heisst **Wahrscheinlichkeitsfunktion** der Zufallsvariablen X

Bemerkung:

Der Graph einer solchen Wahrscheinlichkeitsfunktion besteht aus isolierten Punkten. Die graphische Darstellung wird daher oft als unübersichtlich empfunden und daher durch ein **Stabdiagramm** ersetzt.

Bsp.: Würfel:



Es ist unmittelbar einsichtig, dass durch die Wahrscheinlichkeitsfunktion die Wahrscheinlichkeitsverteilung oder die Verteilung der Zufallsvariablen X vollständig bestimmt ist.

Satz:

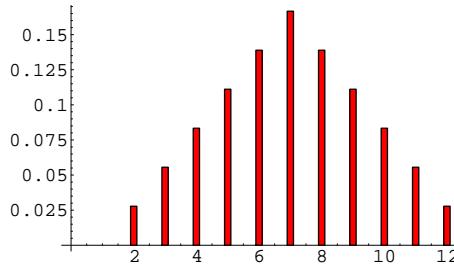
1. $\sum_i f(x_i) = 1$
2. $P(a \leq X \leq b) = \sum_{a \leq x_i \leq b} f(x_i) = \sum_{a \leq x_i \leq b} p_i$

Bsp.: Bestimme die Wahrscheinlichkeitsfunktion (Graph, Wertetabelle) beim Zufallsexperiment „würfeln mit zwei Würfeln“.

Wertetabelle:

| | | | | | | | | | | | | |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|
| x | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{7}{36}$ | $\frac{8}{36}$ | $\frac{9}{36}$ | $\frac{10}{36}$ | $\frac{11}{36}$ | $\frac{12}{36}$ |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|

Graph:



Verteilungsfunktion

Sei X eine Zufallsvariable aus einem Zufallsexperiment. Sei $P(X \leq x)$ damit bestimmt, $x \in \mathbb{R}$ beliebig. Sei $I = \{X \mid X \leq x\}$. Damit schreiben wir kurz: $P(X \leq x) = P(I)$. Von Seite 63 wissen wir, dass damit durch $F(x) := P(X \leq x)$ die Verteilungsfunktion der Zufallsvariablen X für alle $x \in \mathbb{R}$ definiert ist.

Bsp.: Wurf einer Münze, einmal. $X = X(\{\text{Anzahl Kopf}\})$.

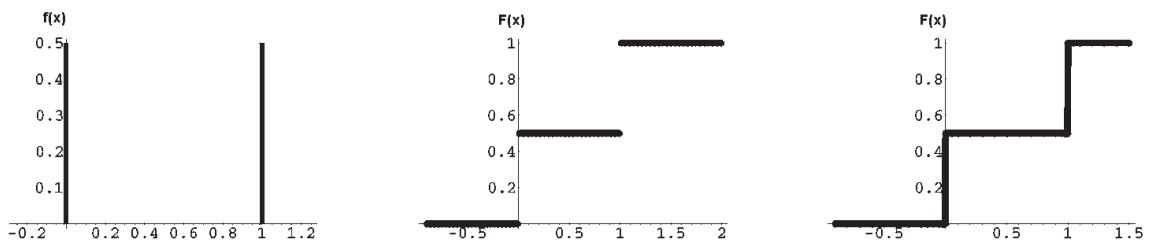
Es ist: $P(X = 0) = P(X = 1) = 0.5$.

1. $x < 0 \rightsquigarrow P(X < 0) = 0 \Rightarrow F(x) = 0, x < 0$
2. $x = 0 \rightsquigarrow P(X = 0) = 0.5 \Rightarrow F(0) = P(x \leq 0) = 0.5$
3. $x \in (0, 1) \rightsquigarrow \dots \Rightarrow F(x) = 0.5, x \in (0, 1)$
4. $x = 1 \rightsquigarrow P(X \leq 1) = P(X < 1) + P(X = 1) = P(X = 0) + P(X = 1) = 0.5 + 0.5 = 1$
5. $x > 1 \rightsquigarrow \dots \Rightarrow P(x) = 1, x > 1$

Resultat:

| x | $0 < x$ | $0 \leq x < 1$ | $1 \leq x$ |
|--------|---------|----------------|------------|
| $F(x)$ | 0 | 0.5 | 1 |

Diagramme:



Problem: **Geg.:** $X \in (a, b] \rightsquigarrow a < X \leq b$

Frage:

Wie lässt sich aus $F(x)$ die Wahrscheinlichkeit $P(a < X \leq b)$ berechnen?

Lösung:

$$\{X \mid X \leq a\} \cup \{X \mid a < X \leq b\} = \{X \mid X \leq b\} \Rightarrow P(X \leq a) + P(a < X \leq b) = P(X \leq b)$$

$$P(X \leq a) = F(a), \quad P(X \leq b) = F(b) \Rightarrow \{X \mid X \leq b\} \Rightarrow F(a) + P(a < X \leq b) = F(b)$$

Damit lässt sich bei der diskreten Verteilung die Wahrscheinlichkeitsfunktion aus der Verteilungsfunktion berechnen. Den folgenden Satz kennen wir schon von früher (Seite 63):

Satz:

Vor.:

Diskrete Verteilung

Beh.:

$$P(a < X \leq b) = F(b) - F(a)$$

Sei $x_1 \leq x_2 \leq \dots \leq x_n \leq x$

Von Seite 65 wissen wir:

$$P(a \leq X \leq b) = \sum_{a \leq x_i \leq b} f(x_i) = \sum_{a \leq x_i \leq b} p_i, \quad f(x_i) = p_i \rightsquigarrow F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i) = \sum_{x_i \leq x} p_i \rightsquigarrow$$

Satz:

$$F(x) = \sum_{x_i \leq x} f(x_i) = \sum_{x_i \leq x} p_i$$

Bsp.: $X = X(\{\text{Augenzahl beim Würfeln mit einem Würfel.}\})$

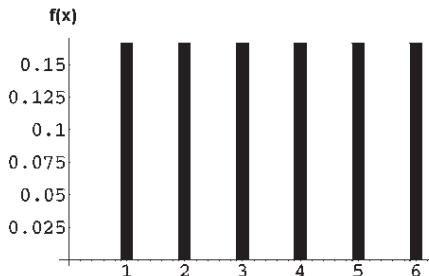
$F(x)$ ist weniger anschaulich als $f(x)$, besitzt jedoch Vorteile in der Handhabung.

Es ist: $P(X = 1) = P(X = 2) = \dots = P(X = 6) = \frac{1}{6}$.

rightsquigarrow **Resultat:**

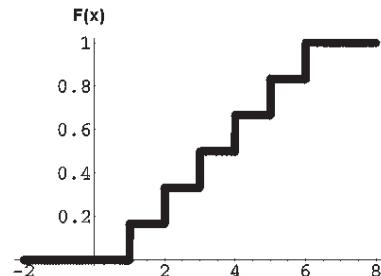
| x | $x < 1$ | $1 \leq x < 2$ | $2 \leq x < 3$ | $3 \leq x < 4$ | $4 \leq x < 5$ | $5 \leq x < 6$ | $6 \leq x$ |
|--------|---------|----------------|-----------------------------|-----------------------------|-----------------------------|----------------|-------------------|
| $F(x)$ | 0 | $\frac{1}{6}$ | $\frac{2}{6} = \frac{1}{3}$ | $\frac{3}{6} = \frac{1}{2}$ | $\frac{4}{6} = \frac{2}{3}$ | $\frac{5}{6}$ | $\frac{6}{6} = 1$ |

Diagramme:



rightsquigarrow Z.B.

$$P(X = 2 \dot{\vee} X = 3) = P(1 < X \leq 3) = F(3) - F(1) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$$



Bemerkung:

Eine Funktion wie $F(x)$ im obigen Beispiel nennen wir **Treppenfunktion**. Sie besitzt an diskreten Stellen (mögliche Werte x_i der Zufallsvariablen X) Sprünge, dazwischen verläuft sie konstant. Die Sprünge entstehen, weil dort $p(x) \neq 0$ ist. Treppenfunktionen kann man mit Hilfe der „Sprungfunktion“ $h(x)$ (vgl. Dirac'sche δ -Funktion, vgl. Analysis) konstruieren:

Satz:

$$F(x) = \sum_{k=1}^n h(x - x_k) \cdot p(x_k), \quad p(x_k) \neq 0, \quad k = 1, \dots, n, \quad h(x) = \operatorname{sgn}\left(\frac{\operatorname{sgn}(x) + 1}{2}\right)$$

4.6.5 Kontinuierliche Verteilung**Das Problem der Wahrscheinlichkeitsfunktion**

Wir definieren für unsere Bedürfnisse:

Definition:

Eine Zufallsvariable X heisst **stetig in einem Intervall I** , wenn alle Werte in I angenommen werden können.

Problem: Macht bei einer stetigen Zufallsvariablen $P(X = x_i)$ noch einen Sinn?

Das Problem ist ja, dass in diesem Fall unendlich viele Werte $x_i \in I$ angenommen werden können. Da aber im diskreten Fall die aufsummierte Wahrscheinlichkeit 1 ist, müsste daher auch hier konsequenterweise für die meisten x_i dann $P(X = x_i) = 0$ sein. Das besagt aber, dass diese angenommenen Werte dann doch nicht angenommen werden. Wir landen also in einem Widerspruch. Daher ist es in unserem Rahmen nicht sinnvoll, nach der Wahrscheinlichkeit $P(X = x_i)$ und daher nach einer Wahrscheinlichkeitsfunktion zu fragen. Wir fragen besser nach der Wahrscheinlichkeit für Intervalle, z.B. $P(a \leq X < b) = ?$ Das ist bei einer gegebenen Verteilungsfunktion kein Problem.

Situation beim diskreten Fall

Sei also: $F(x) = P(X \leq x)$

Im diskreten Fall ist:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i), \quad x_i : x_1 < x_2 < x_3 < \dots$$

Da die diskreten Werte x_i geordnet sind, besteht eine bijektive Beziehung zwischen dem Wert x_i und dem Index i :

$$\varphi : x_i \longmapsto \varphi(x_i) = i$$

Damit existiert auch $\varphi^{-1} = \psi$:

$$\psi(i) = x_i$$

$$\begin{aligned} \text{Sei } \varphi(x_j) = j, \quad \psi(j) = x_j \Rightarrow F(x_j) &= \sum_{x_i \leq x_j} f(x_i) = \sum_{\psi(i) \leq \psi(j)} f(\psi(i)) := \\ &= \sum_{\psi(i) \leq \psi(j)} \tilde{f}(i) = \sum_{\psi(i) \leq \psi(j)} \tilde{f}(i) \cdot 1 = \sum_{\psi(i) \leq \psi(j)} \tilde{f}(i) \cdot \underbrace{(i + 1 - i)}_{\Delta i} = \sum_{\psi(i) \leq \psi(j)} \tilde{f}(i) \cdot \Delta i \end{aligned}$$

$\tilde{f}(i) \cdot \Delta i$ hat die Bedeutung eines Rechteckinhalts. $F(x)$ hat somit die Bedeutung eines Flächeninhalts unter einer Treppenkurve. Die Bedeutung des Flächeninhalts hat aber auch im stetigen Fall noch Gültigkeit.

Wahrscheinlichkeitsdichte

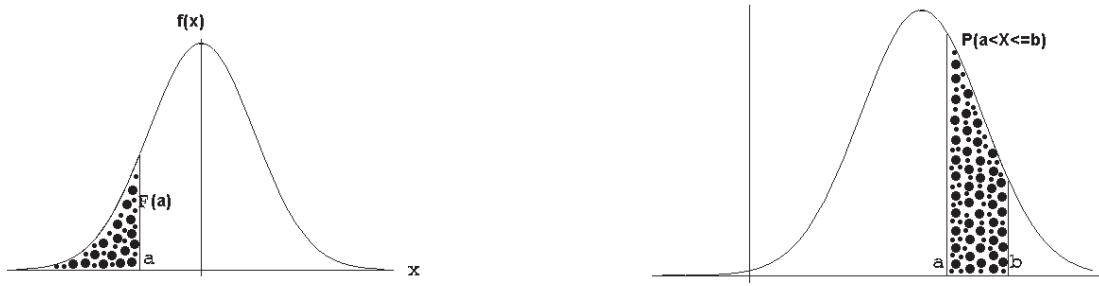
Im stetigen Fall liegt es daher nahe, $F(x) = P(X \leq x)$ ebenfalls als Flächeninhalt unter einer Kurve zu definieren.

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x^*) dx^* \Rightarrow f(x) = \frac{dF}{dx}$$

$f(x) = \frac{dF}{dx}$ kann hier nicht mehr die Bedeutung einer Wahrscheinlichkeit haben, denn alle Wahrscheinlichkeiten müssen sich ja zur Gesamtwahrscheinlichkeit 1 aufsummieren lassen können. Daher redet man hier von einer **Wahrscheinlichkeitsdichte**.

$$\rightsquigarrow P(a < x \leq b) = F(b) - F(a)$$

Diagramme:



Bsp.: Roulettepiel:

$$0 \leq X < 2\pi$$

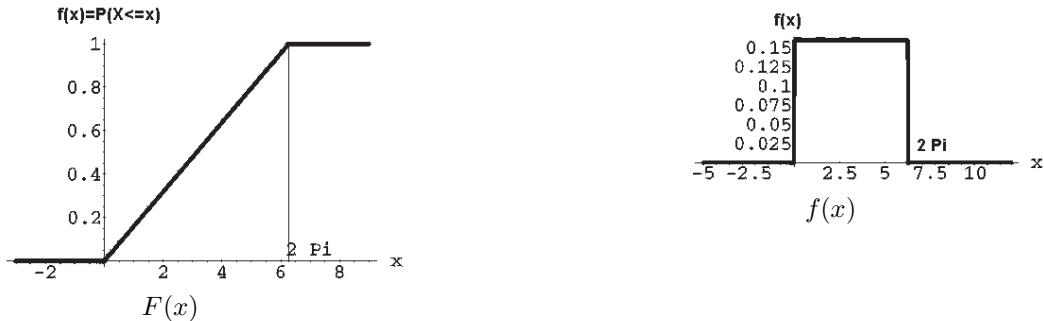
$X :=$ Winkel des Zeigers mit einer festen Richtung.

$$\text{Brauchbares Roulett} \rightsquigarrow P(X \leq x) = \frac{x}{2\pi}$$

$$\Rightarrow F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{2\pi} & 0 < x \leq 2\pi \\ 1 & 2\pi < x \end{cases}$$

$$f(x) = F'(x) \Rightarrow f(x) = \begin{cases} \frac{1}{2\pi} & x \in (0, 2\pi) \\ 0 & x \notin (0, 2\pi) \end{cases}$$

Diagramme:



Bemerkung:

Eine Verteilung wie im letzten Beispiel nennen wir **gleichförmige Verteilung** oder **Rechtecksverteilung**.

Hinweis:

Man beachte die Analogie zwischen Masseverteilung in der Physik und Wahrscheinlichkeitsverteilung. Punktmassen entsprechen dann Wahrscheinlichkeiten im diskreten Fall.

4.7 Mass- oder Kennzahlen einer Verteilung

4.7.1 Allgemeines

Problem: Wie lassen sich zwei Verteilungen rasch vergleichen?

1. Qualitativer Vergleich von $f(x)$, $F(x)$ oder Graphen (unendliche Mengen!).
2. Quantitativer Vergleich charakteristischer Masszahlen: Mittelwert μ , Varianz σ^2 , Standardabweichung σ , Schiefe γ u.s.w..

4.7.2 Mittelwert

1. Diskreter Fall:

Vergleich mit der Physik: Unter dem „Mittelpunkt einer Massenverteilung“ verstehen wir den Schwerpunkt. Betrachtet man an Stelle von Wahrscheinlichkeitsverteilungen nun statistische Häufigkeitsverteilungen, so tritt an die Stelle der Massenverteilung das arithmetische Mittel oder kurz der Mittelwert. Bei einer Wahrscheinlichkeitsverteilung tritt an die Stelle des Schwerpunktes somit auch Mittelwert. Es gilt:

$$\bar{x} \cdot m = \sum_{i=1}^n x_i \cdot m_i \Rightarrow \bar{x} = \sum_{i=1}^n \frac{x_i \cdot m_i}{m} = \sum_{i=1}^n \frac{m_i}{m} \cdot x_i \hat{=} \sum_{i=1}^n p_i \cdot x_i := \sum_{i=1}^n f(x_i) \cdot x_i$$

Dabei kann $n \in \mathbb{N}$ oder $n = \infty$ sein.

Daher definieren wir:

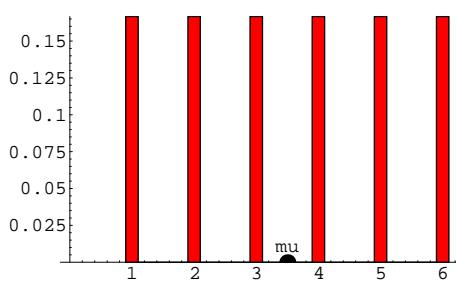
Definition:

Mittelwert: $\mu := \sum_i f(x_i) \cdot x_i \cdot 1$

Konsequenz:

$$(a) \ i \in \{1, 2, \dots, n\}, \ f(x_i) = p_i \Rightarrow \mu = \sum_{i=1}^n p_i \cdot x_i$$

$$(b) \ p_i = \frac{1}{n} \Rightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i$$



Bsp.:

X = Augenzahl beim Würfeln.

$$f(x_i) = \frac{1}{6} \quad x_i = 1, 2, 3, \dots, 6 \Rightarrow 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{i=1}^6 i = \frac{1}{6} \cdot 21 = 3.5.$$

2. Übertragung des Begriffs auf den stetigen Fall:

$$\text{Sei } x_i = i \Rightarrow \Delta x_i = (i+1) - i = 1 \Rightarrow \\ \mu = \sum_{i=n}^6 f(x_i) \cdot x_i = \sum_{i=n}^6 f(x_i) \cdot x_i \cdot 1 = \sum_{i=n}^6 f(x_i) \cdot x_i \cdot \Delta x_i \rightarrow \int_{-\infty}^{\infty} f(x) \cdot x \cdot dx \dots$$

Daher definieren wir analog zum Schwerpunkt der Physik:

Definition:

Mittelwert:

$$\mu := \int_{-\infty}^{\infty} f(x) \cdot x \cdot dx$$

4.7.3 Erwartungswert

Bsp.:

Zwei Spieler A und B spielen wie folgt:

Vor dem Spiel zahlt A an B den Betrag „durchschnittliche Gewinnerwartung beim Spiel“. A darf dann würfeln, während B an A Gewinne auszahlt nach dem folgenden Schema: Beim Resultat 1 oder 2 erhält A von B Fr. 10.-, beim Resultat 3 oder 4 erhält A von B Fr. 20.-, beim Resultat 5 erhält A von B Fr. 40.-, beim Resultat 6 erhält A von B Fr. 80.-.

Was ist die durchschnittliche Gewinnerwartung?

$$\rightsquigarrow \bar{G} = 10 \cdot \frac{1}{6} + 10 \cdot \frac{1}{6} + 20 \cdot \frac{1}{6} + 20 \cdot \frac{1}{6} + 40 \cdot \frac{1}{6} + 80 \cdot \frac{1}{6} = 30 \text{ oder } 6 \cdot \bar{G} = 10 + 10 + 20 + 20 + 40 + 80$$

In diesem Beispiel sehen wir eine Zuordnung folgender Art:

$$X \longleftrightarrow g(X) \text{ resp. } x_i \longleftrightarrow g(x_i)$$

| X | 1 | 2 | 3 | 4 | 5 | 6 |
|------|----|----|----|----|----|----|
| g(X) | 10 | 10 | 20 | 20 | 40 | 80 |

X ist unabhängige Zufallsvariable, $g(X)$ ist abhängige Zufallsvariable.

Konsequenz:

$g(X)$ ist ebenfalls Zufallsvariable

Um die durchschnittliche **Gewinnerwartung** zu berechnen, müssen wir daher x_i ersetzen durch $g(x_i)$, d.h. wir ersetzen $x_i \cdot f(x_i)$ durch $g(x_i) \cdot f(x_i)$. Dabei ist $f(x_i)$ die zu $g(x_i)$ gehörige Wahrscheinlichkeit.

Wir definieren daher analog zur durchschnittlichen Gewinnerwartung ganz allgemein die **Erwartung** resp. den **Erwartungswert** $E(g(X))$ zur **Erwartungsfunktion** $g(X)$. Analog zur Definition zu μ setzen wir fest:

Definition:

$$1. \text{ Diskreter Fall: } E(g(X)) := \sum_i g(x_i) \cdot f(x_i) \cdot 1 = \sum_i g(x_i) \cdot f(x_i) \cdot \Delta x$$

$$2. \text{ Stetiger Fall: } E(g(X)) := \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$$

Bemerkung:

Statt **Erwartungswert** sagen wir auch **mathematischer Erwartungswert**, **Erwartung** oder manchmal auch **kurz Mittelwert**.

Konsequenz: $g(x) = x \Rightarrow E(g(x)) = E(x) = \mu$

Speziell: (Diskreter Fall)

$$1. E(g(X)) = \sum_{i=1}^n g(x_i) \cdot f(x_i) = \left\langle \begin{pmatrix} g(x_1) \\ \vdots \\ g(x_n) \end{pmatrix}, \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} \right\rangle = \langle \vec{g}(x), \vec{f}(x) \rangle$$

$$2. g(x_i) = x_i \wedge f(x_i) = p_i \Rightarrow E(X) = \sum_{i_1}^n p_i \cdot x_i = \mu$$

$$3. g(x_i) = x_i \wedge f(x_i) = p_i = \frac{1}{n} \Rightarrow E(X) = \frac{1}{n} \sum_{i_1}^n x_i = \mu$$

Wegen der Linearität von Summe und Integral gilt:

Satz:

Vor.:

$$g(X) = a h(X) + b u(X)$$

Beh.:

$$E(g(X)) = a E(h(X)) + b E(u(X))$$

4.7.4 Symmetrische Verteilung

Definition:

Eine Verteilung heisst **symmetrisch** bezüglich c , wenn gilt:
 $\forall_{c \pm x \in D_f} : f(c+x) = f(c-x)$

Trivialerweise gilt für eine symmetrische Verteilung:

Satz:

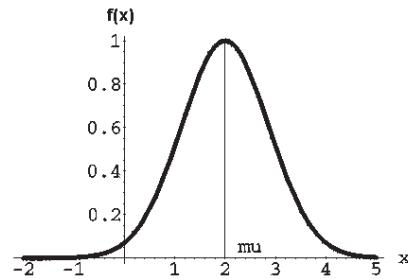
Vor.:

f symmetrisch bezüglich c

Beh.:

$\mu = c$

Bsp.:



4.7.5 Varianz, Standardabweichung

Statt nach dem Erwartungswert von x resp. x_i zu fragen, fragen wir nach dem Erwartungswert der quadratischen Abweichung von μ von x resp. x_i . (Achtung: Der Erwartungswert der linearen Abweichung gibt immer 0. Diese Grösse macht daher keinen Sinn.)

1. Diskreter Fall:

Definition:

Varianz

$$\sigma^2 = \sum_i (x_i - \mu)^2 \cdot f(x_i) \cdot 1$$

2. Stetiger Fall:

Wie beim Erwartungswert wird im stetigen Fall auch hier die Summe zu einem Integral.

Definition:

Varianz

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx$$

3. Definition:

$$\sigma = \sqrt{\sigma^2} \text{ heisst } \mathbf{Standardabweichung}$$

Bemerkung:

σ^2 ist wie erwähnt der Mittelwert der quadratischen Abweichung von μ . σ^2 gehört qualitativ als Mittelwert von Quadraten daher ebenfalls zur Klasse der Quadrate. σ ist somit ebenfalls ein Mass für die Grösse der mittleren Abweichung von μ , das qualitativ somit wieder zur Klasse der ursprünglichen nichtquadratierten Werte gehört. σ^2 oder auch σ werden manchmal als **Streuung** bezeichnet.

Konsequenz: $g(X) = (X - \mu)^2 \Rightarrow E(g(X)) = E((X - \mu)^2) = \sigma^2$

Speziell: (Diskreter Fall)

$$\begin{aligned} i \in \{1, 2, \dots, n\}, \quad g(x_i) &= x_i, \quad f(x_i) = p_i \\ \Rightarrow \sigma^2 &= E((X - \mu)^2) = E(X^2 - 2X\mu + \mu^2) = E(X^2) - E(2X\mu) + E(\mu^2) = E(X^2) - 2\mu E(X) + \mu^2 E(1) \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 \cdot \frac{1}{n} \sum_{i=1}^n 1 = E(X^2) - 2\mu^2 + \mu^2 \cdot \frac{1}{n} \cdot n = E(X^2) - \mu^2 = E(X^2) - (E(X))^2 \end{aligned}$$

Korollar: $i \in \{1, 2, \dots, n\}, \quad g(x_i) = x_i, \quad f(x_i) = p_i$

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

Bsp.:

$X = \text{„Anzahl Kopf“}$ beim einmaligen Wurf einer Münze.

$$\begin{aligned} X = 0 \rightsquigarrow P(X = 0) &= \frac{1}{2} & X = 1 \rightsquigarrow P(X = 1) &= \frac{1}{2} \\ \Rightarrow \mu &= 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2} \\ \Rightarrow \sigma^2 &= (0 - \frac{1}{2})^2 \cdot \frac{1}{2} + (1 - \frac{1}{2})^2 \cdot \frac{1}{2} = (\frac{1}{2})^2 \cdot \frac{1}{2} + (\frac{1}{2})^2 \cdot \frac{1}{2} = \frac{1}{4} \\ \Rightarrow \sigma &= \sqrt{\frac{1}{4}} = \frac{1}{2} \notin \{0, 1\} \end{aligned}$$

Konsequenz: $\sigma \in \{\dots, x_i, \dots\}$ ist nicht zwingend

4.7.6 Momente einer Verteilung

Größen wie Mittelwert und Varianz heissen allgemein **Momente** der Verteilung der Zufallsvariablen X mit der Wahrscheinlichkeitsfunktion $f(X)$.

Definition: Der folgende Erwartungswert heisst k -tes **Moment** der Verteilung:

1. Diskreter Fall:

$$\begin{aligned} m_k &= E(X^k) = \sum_i x_i^k \cdot f(x_i) \\ m_{k,g} &= E(g(X)^k) = \sum_i g(x_i)^k \cdot f(x_i) \end{aligned}$$

2. Stetiger Fall:

$$\begin{aligned} m_k &= E(X^k) = \int_{-\infty}^{\infty} x^k \cdot f(x) dx \\ m_{k,g} &= E(g(X)^k) = \int_{-\infty}^{\infty} g(x)^k \cdot f(x) dx \end{aligned}$$

Idee: $Y = X - \mu$, $f(x_i) = p = \frac{1}{n} \Rightarrow E(Y^k) = \sum_{i=1}^n \frac{(x_i - \mu)^k}{n} = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \mu)^k$
 \rightsquigarrow Mechanik!

Bsp.:

Mittelwert: $\mu = E(X) = E(X^1) = \sum_i x_i^1 \cdot f(x_i) = \sum_i x_i \cdot f(x_i)$

Varianz: $\sigma^2 = E((X - \mu)^2) = \sum_i (x_i - \mu)^2 \cdot f(x_i)$

Satz:

1. $E(X^2) = \sigma^2 + \mu^2 = E((X - E(X))^2) + (E(X))^2$
2. $E((X - \mu)^3) = E(X^3) - 3\mu E(X^2) + 2\mu^3$

Beweis:

1. $\sigma^2 = \sum_i (x_i - \mu)^2 \cdot f(x_i) = \sum_i x_i^2 \cdot f(x_i) - 2\mu \sum_i x_i \cdot f(x_i) + \mu^2 \cdot \sum_i f(x_i) = E(X^2) - 2\mu\mu + \mu^2 \cdot 1 = E(X^2) - \mu^2$

2. Übung.

Konsequenz: Daraus ergibt sich eine vereinfachte Berechnung der Varianz:

Formel: $\sigma^2 = E(X^2) - \mu^2 = E(X^2) - (E(X))^2$, $E(X^2) = \sum_i x_i^2 \cdot f(x_i)$

Schreibweise: Oft schreibt man auch:

$$\mu_n =: E(X^n) = \begin{cases} \sum_i x_i^n p_i & \text{diskr.} \\ \int_{-\infty}^{\infty} x^n f(x) dx & \text{stet.} \end{cases}$$

\rightsquigarrow **Formel:** $\sigma^2 = \mu_2 - \mu_1^2$

Aus der Linearität der Summe und des Integrals ergibt sich:

Satz: $E(a_n X^n + a_{n-1} X^{n-1} + \dots + a_1 X + a_0) = a_n E(X^n) + a_{n-1} E(X^{n-1}) + \dots + a_1 E(X) + a_0 = a_n \mu_n + a_{n-1} \mu_{n-1} + \dots + a_1 \mu_1 + a_0$

Daraus leiten wir ab:

$$\begin{aligned} E(aX + b) &= aE(X) + b \Rightarrow \mu(aX + b) = a\mu(X) + b \Rightarrow \\ (\sigma(aX + b))^2 &= E((aX + b - \mu(aX + b))^2) = E((aX + b)^2 - 2(aX + b)(\mu(aX + b)) + (\mu(aX + b))^2) = \\ E(a^2 X^2 + 2aXb + b^2 - 2(aX + b)(a\mu(X) + b) + (a\mu(X) + b)^2) &= \\ E(a^2 X^2 + 2aXb + b^2 - 2a^2 X\mu(X) - 2aXb - 2b\mu(X) - 2b^2 + a^2 \mu(X)^2 + 2a\mu(X)b + b^2) &= \\ E(a^2 X^2 - 2a^2 X\mu(X) + a^2 \mu(X)^2) &= E(a^2 (X - \mu(X))^2) = a^2 E((X - \mu(X))^2) = a^2 \sigma^2 \rightsquigarrow \end{aligned}$$

Satz: $(\sigma(aX + b))^2 = a^2 \sigma(X)^2$
oder $Var(aX + b) = a^2 Var(X)$

Der letzte Satz erlaubt manchmal eine Vereinfachung der Berechnung der Varianz.

Korollar: $(\sigma(b))^2 = Var(b) = 0$

Man rechnet sofort nach:

| | | |
|--------------|--------------|---|
| Satz: | Vor.: | Zu X gehöre der Mittelwert μ und die Varianz σ^2 . |
| | | $Z = \frac{X - \mu}{\sigma}$ |
| | Beh.: | Zu Z gehöre der Mittelwert 0 und die Varianz 1. |

Zum Beweis:

$$\begin{aligned} \mu(Z) = E(Z) &= \frac{1}{\sigma} \cdot (E(X) - E(\mu)) = \frac{1}{\sigma} \cdot (\mu - \mu \cdot E(1)) = \frac{1}{\sigma} \cdot (\mu - \mu \cdot 1) = 0 \\ \text{P. 75: } E(X^2) = \sigma^2 + \mu^2 \Rightarrow \sigma(Z)^2 &= E(Z^2) - \mu(Z^2) = E(Z^2) = E\left(\frac{1}{\sigma^2} \cdot (X^2 - 2X\mu + \mu^2)\right) \\ &= \frac{1}{\sigma^2} \cdot (E(X^2) - 2E(X)\mu + \mu^2) = \frac{1}{\sigma^2} \cdot (E(X^2) - \mu^2) = \frac{1}{\sigma^2} \cdot (\sigma^2 + \mu^2 - \mu^2) = 1 \end{aligned}$$

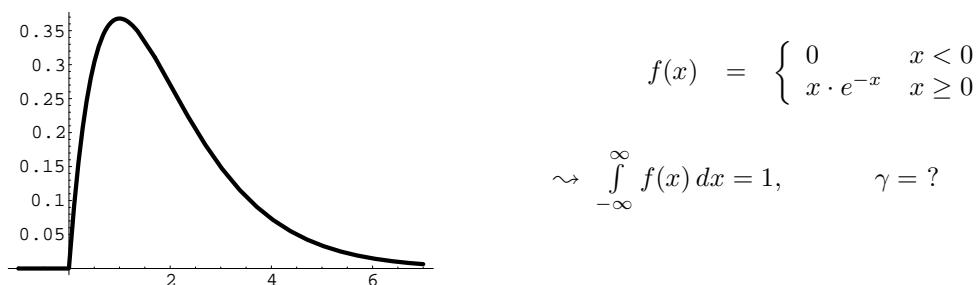
Infolge des letzten Satzes definieren wir:

Definition: $Z = \frac{X - \mu}{\sigma}$ heisst **Standardform** von X .

4.7.7 Schiefe einer Verteilung

Definition: Sei $\gamma = \frac{1}{\sigma^3} \cdot E((X - \mu)^3)$
 γ heisst **Schiefe** der Verteilung. .

Bemerkung: Die Schiefe der Verteilung ist ein Mass für die Assymetrie der Verteilung.



Mit partieller Integration erhalten wir:

$$\begin{aligned}\mu = E(X) &= \int_0^\infty x \cdot e^{-x} \cdot x \, dx = \int_0^\infty x^2 \cdot e^{-x} \, dx = 2, \quad E(X^2) = 6, \quad E(X^3) = 24 \\ \Rightarrow E((X - \mu)^3) &= 24 - 3 \cdot 2 \cdot 6 + 2 \cdot 2^3 = 4, \quad \sigma^2 = E(X^2) - (E(X))^2 = 6 - 2^2 = 2 \\ \Rightarrow \gamma &= \frac{1}{\sigma^3} \cdot E((X - \mu)^3) = \frac{1}{2^{(3)}} \cdot 4 = \sqrt{2} \approx 1.41421\end{aligned}$$

4.7.8 Weitere Kenngrößen

Variationskoeffizient

Der **Variationskoeffizient** oder **Variabilitätskoeffizient** ν einer Zufallsgröße X mit $\mu \neq 0$ ist wie folgt definiert:

Definition: $\nu = \frac{\sigma}{\mu}$

Median, Modus

Median und Modus sind Lagemasszahlen. Sie betreffen vor allem die beurteilende Statistik. (Vgl. 2.)

Definition:

Der **Median** einer Menge von Zahlen, die der Grösse nach geordnet sind, ist der Wert in der Mitte oder das arithmetische Mittel der beiden Werte in der Mitte der geordneten Datenfolge.

Falls die Daten in Klassen eingeteilt worden sind, muss die Definition angepasst werden (vgl. Lit.).

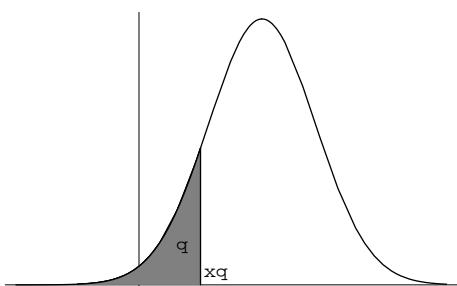
Quantile, Fraktile

Gegeben sei eine stetige Zufallsgröße X mit einer Dichtefunktion f und einer Verteilungsfunktion F . Sei $0 < q < 1$.

Definition:

$x_q \in \mathbb{R}$ heisst das **untere Quantil der Ordnung q** (q -Quantil), wenn gilt: $F(x_q) = P(X \leq x_q) = q$
Entsprechend für das **obere Quantil**.

Konsequenz: Das Quantil der Ordnung $q = 0.5$ ist gerade der Median.



Spezielle Quantile sind **Quartile** ($Q_1 = x_{0.25}$, $Q_3 = x_{0.75}$), **Quintile**, **Dezile** oder **Zentile** (P_{10}, P_{90}) u.s.w..

Bemerkung:

Der **Quartilabstand** $Q_3 - Q_1 = x_{0.75} - x_{0.25}$ ist ein Streuungsmaß. Ebenso der Zentilabstand $P_{90} - P_{10}$. Oft wird auch der **halbe Quartilabstand** $Q = \frac{Q_3 - Q_1}{2}$ benutzt.

Exzess, Wölbung

Z.B. der **Exzess (Kurtosis)** κ ist ein Mass für die Steilheit einer Verteilung.

Definition:

Exzess (Kurtosis):

$$\kappa = \frac{Q}{P_{90} - P_{10}} \quad (Q, P \text{ vgl. oben})$$

Momentenkoeffizient der Kurtosis:

$$a_4 = \frac{m_4}{m_2^2} \quad (m_k = E(X^k))$$

Andere Mittelwerte

Für andere Mittelwerte wie **harmonische** oder **geometrische** Mittelwerte sowie **empirische** Beziehungen vgl. Lit..

4.7.9 Momentenerzeugende charakteristische Funktion

Der Grund zu der folgenden Definition wird klar, wenn wir damit arbeiten:

Definition:

Die Erwartung $E(e^{tX})$ von e^{tX} heisst **momentenerzeugende Funktion** $G(t)$.

Wir berechnen $G(t)$ für beliebige Funktionen $f(x)$:

Diskreter Fall: $G(t) = E(e^{tX}) = \sum_i e^{tx_i} \cdot f(x_i)$

Für die k -te Ableitung nach t erhalten wir:

$$\frac{d^k G(t)}{dt^k} = G_t^{(k)}(t) = \sum_i x_i^k \cdot e^{tx_i} \cdot f(x_i)$$

Kontinuierlicher Fall: $G(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) dx$

$$\frac{d^k G(t)}{dt^k} = G_t^{(k)}(t) = \int_{-\infty}^{\infty} x^k \cdot e^{tx} \cdot f(x) dx$$

Satz:

\rightsquigarrow **Trick:**

$$t = 0 \Rightarrow G^{(k)}(0) = E(X^k) = \begin{cases} \sum_i x_i^k \cdot e^{0 \cdot x_i} \cdot f(x_i) & = \sum_i x_i^k \cdot f(x_i) \\ \int_{-\infty}^{\infty} x^k \cdot e^{0 \cdot x} \cdot f(x) dx & = \int_{-\infty}^{\infty} x^k \cdot f(x) dx \end{cases}$$

Speziell:

$$k = 1 \Rightarrow G'_t(0) = E(X^1) = E(X) = \mu \rightsquigarrow \text{Mittelwert}$$

$$\Rightarrow \sigma^2 = E(X^2) - \mu^2 = G''_t(0) - G'_t(0)^2 \rightsquigarrow \text{Varianz}$$

Bemerkung:

Die momenterzeugende Funktion wird uns im nächsten Kapitel bei der Bernoulli-Verteilung nützlich sein.

4.7.10 Laplace- und z-Transformation

Dem mit der Ingenieurmathematik vertrauten Leser fällt auf, dass die momenterzeugende Funktion $G(t)$ offensichtlich mit der Laplace- resp. der z -Transformation zu tun hat. Der Zusammenhang wird bei späterer Gelegenheit genauer erläutert, da er im Moment nicht unverzichtbar ist. Kenntnisse der Laplace- und der z Transformation erwirbt man in der höheren Analysis, vgl. Lit. Fortsetzung Mathematik, Wirz, Bibl. A16.

4.8 Spezielle diskrete Verteilungen

4.8.1 Bernoulliverteilung

Situation und Einzelversuch

Problem:

Wir studieren die Anzahl des Eintreffens eines bestimmten Ereignisses A bei der wiederholten Ausführung eines Zufallsexperiments, bei dem die Ereignisse unabhängig und gleichwahrscheinlich sind. Z.B. sei A das Eintreffen einer 2 beim Würfeln mit einem Würfel, wobei mehrmals gewürfelt wird. Das betrachtete Ereignis kann eintreffen oder nicht. Die dazugehörige Zufallsvariable X kann daher nur zwei Werte annehmen: 0 und 1.

Sei $X = 0$ falls A nicht eintrifft und $X = 1$ falls A eintrifft.

Sei $p =$ Wahrscheinlichkeit dass A eintrifft.

$q = 1 - p =$ Wahrscheinlichkeit dass A nicht eintrifft.

$X =$ Zufallsvariable: Anzahl des Eintreffens von A .

Folgerung:

Erfolgswahrscheinlichkeit beim Einzelversuch:

$$P(X = 0) = q, \quad f(0) = q$$

$$P(X = 1) = p, \quad f(1) = p$$

$$x \notin \{0, 1\} \Rightarrow f(x) = 0$$

Dieser Sachverhalt kann kurz in einer einzigen Formel wie folgt gefasst werden:

Konsequenz: Im Einzelversuch gilt:

$$f(x) = \begin{cases} p^x \cdot q^{1-x} & x \in \{0, 1\} \\ 0 & x \notin \{0, 1\} \end{cases}$$

Definition:

Den eben beschriebenen Versuch nennt man oft auch **Bernoulli–Versuch**.

Mehrfachversuch

Bei n –maliger Wiederholung des Experiments kann sich der Erfolg 0, 1, 2, … oder n mal einstellen. Erfolg heisst, dass A eintrifft. Daher kann also X die Werte 0, 1, 2, … oder n annehmen.

Bei n Versuchen treffe nun A x mal ein. Da die Versuche identisch und daher die Ereignisse unabhängig sind, ist die Wahrscheinlichkeit des Gesamtereignisses das Produkt der Wahrscheinlichkeiten der Einzelereignisse:

$$P(X = x) = \underbrace{p \cdot p \cdot \dots \cdot p}_{p^x} \cdot \underbrace{q \cdot q \cdot \dots \cdot q}_{q^{n-x}} = p^x \cdot q^{n-x}$$

Konsequenz: Die Formel für $n = 1$ ist hier eingeschlossen.

Der Erfolg kann sich bei gleicher Gesamtzahl auf verschiedene Arten auf die Einzelereignisse verteilen. Die Anzahl Möglichkeiten erhält man aus der Antwort auf die Frage, auf wieviele Arten man aus n Elementen x Elemente ohne Wiederholung auswählen kann. Das führt daher auf eine Kombination ohne Wiederholung:

$$\text{Möglichkeiten: } \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Bsp.: Bei 3 Versuchen soll ein Ereignis 2 mal eintreten. Wieviele Möglichkeiten gibt es? — Der Erfolg kann sich (a) im 1. und 2. Versuch oder (b) im 1. und 3. Versuch oder (c) im 2. und 3. Versuch einstellen. Es gibt also $|\{(a), (b), (c)\}| = 3 = \binom{3}{2}$ Möglichkeiten.

Satz:**Vor.:**

Bernoulli-Experiment mit n Wiederholungen

Beh.:

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

Definition:

Die durch $f(x) = \binom{n}{x} p^x q^{n-x}$ gegebene Verteilung mit den Parametern $n \in \mathbb{N}$ und $p \in (0, 1)$ heisst **Bernoulli–Verteilung** oder **Binomialverteilung** $Bi(n, p)$.

Korollar:

$$1 = \sum_{x=0}^n P(X = x) = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x}$$

Bemerkung:

Aus $p + q = 1$ bei n unabhängigen Wiederholungen deduziert man ebenfalls:

$$1 = 1^n = (p + q)^n = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x}$$

4.8.2 Gesetze für die Binomialverteilung

Die Verteilungsfunktion ergibt sich durch eine einfache Summation:

Korollar:
$$F(x) = \sum_{k \leq x} \binom{n}{k} p^k q^{n-k}, \quad x \geq 0$$

Weiter gilt:
$$\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!}$$

Daraus folgt:
$$\binom{n}{x+1} = \frac{n-x}{x+1} \cdot \binom{n}{x}$$

Damit erhalten wir eine Rekursionsformel:

Korollar:
$$f(x+1) = \left(\frac{n-x}{x+1}\right) \cdot \frac{p}{q} \cdot f(x)$$

Bsp.:

Es wird eine Sendung mit N Schrauben vorbereitet. M davon sind unbrauchbar (Ausschuss). Die Sendung wird vom Kunden geprüft, indem er n mal eine Schraube zieht und wieder zurücklegt. Wie gross ist die Chance, dass der Kunde x unbrauchbare Schrauben findet?

Offensichtlich handelt es sich um ein Bernoulliexperiment. Die Chance, bei einmaligem Ziehen Ausschuss zu finden ist $p = \frac{M}{N}$.

$$\rightsquigarrow f(x) = \binom{n}{x} \cdot \left(\frac{M}{N}\right)^x \cdot \left(1 - \frac{M}{N}\right)^{n-x}$$

Zur Berechnung von Mittelwert und Varianz benutzen wie die momentenerzeugende Funktion:

$$G(t) = E(e^{tX}) = \sum_i e^{tx_i} \cdot f(x_i) = \sum_{x=0}^n e^{tx} \cdot \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^n \binom{n}{x} \cdot (p \cdot e^t)^x \cdot q^{n-x} = (p \cdot e^t + q)^n$$

$$\Rightarrow G'(t) = n \cdot (p \cdot e^t + q)^{n-1} \cdot p \cdot e^t,$$

$$G''(t) = n \cdot (n-1) \cdot (p \cdot e^t + q)^{n-2} \cdot p^2 \cdot e^{2t} + n \cdot (p \cdot e^t + q)^{n-1} \cdot p \cdot e^t = n \cdot p \cdot e^t \cdot (p \cdot e^t + q)^{n-2} \cdot ((n-1) \cdot p \cdot e^t + (p \cdot e^t + q))$$

$$= e^t n p (e^t p + q)^{n-2} (e^t n p + q) \Rightarrow E(X^2) = G''(0) = n p \underbrace{(p+q)^{n-2}}_{=1} (n p + q) = n \cdot p (n p + q)$$

$$\Rightarrow \mu = G'(0) = n \cdot p, \quad \sigma^2 = G''(0) - \mu^2 = E(X^2) - (E(X))^2 = n^2 \cdot p^2 + n \cdot p \cdot q - n^2 \cdot p^2 = n \cdot p \cdot q$$

Satz:

Vor.:

Bernoulliverteilung

Beh.:

$$\mu = n \cdot p$$

$$\sigma^2 = n \cdot p \cdot q$$

Bsp.:

Kartenspiel, 32 Karten mit 4 Assen. Experiment: 6 mal eine Karte ziehen mit zurücklegen (Bernoulliexperiment). Ereignis A : mindestens 3 Asse kommen.

\rightsquigarrow Einzelexperiment: $p = \frac{4}{32}$

$\rightsquigarrow A$: 3, 4, 5 oder 6 mal ein As.

Kurz: $A = \{3, 4, 5, 6\} = \{3\} \cup \{4\} \cup \{5\} \cup \{6\}$

$$\Rightarrow P(A) = P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = \sum_{k=3}^6 P(\{k\}) = \sum_{k=3}^6 \binom{6}{k} \left(\frac{4}{32}\right)^k \left(1 - \frac{4}{32}\right)^{6-k} \approx 0.03.$$

Zum Vergleich: $\mu = \frac{3}{4} = 0.74$, $\sigma^2 = \frac{21}{32} \approx 0.656$, $\sigma \approx 0.81$

4.8.3 Poissonverteilung

Verteilungsfunktion

Problem: Gesucht ist eine Verteilung, die für grosse n und kleine p einerseits die Bernoulliverteilung gut annähert und andererseits mathematisch bequemer zu handhaben ist. Das Gewünschte leistet die **Poissonverteilung**.

Idee: Für die Bernoulliverteilung gilt: $\mu = n \cdot p$

$$\rightsquigarrow p = \frac{\mu}{n}, \quad p^x = \frac{\mu^x}{n^x}, \quad q^{n-x} = (1-p)^{n-x} = \left(1 - \frac{\mu}{n}\right)^{n-x} = \left(1 - \frac{\mu}{n}\right)^n \cdot \left(1 - \frac{\mu}{n}\right)^{-x}$$

$$\Rightarrow f(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{x!} \cdot \frac{\mu^x}{n^x} \cdot \left(1 - \frac{\mu}{n}\right)^n \cdot \left(1 - \frac{\mu}{n}\right)^{-x} =$$

$$= \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{n^x} \cdot \left(1 - \frac{\mu}{n}\right)^{-x} \cdot \frac{\mu^x}{x!} \cdot \left(1 - \frac{\mu}{n}\right)^n$$

Sei nun $n \rightarrow \infty$ und $p \rightarrow 0$ derart dass gilt $\mu = n \cdot p = \lim_{n \rightarrow \infty, p \rightarrow 0} n \cdot p$.

$$\rightsquigarrow \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{n^x} = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{n-x+1}{n} \xrightarrow{n \rightarrow \infty, x \ll n} 1$$

$$\text{Weiter gilt: } \left(1 - \frac{\mu}{n}\right)^n = \left(1 + \frac{-\mu}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\mu}, \quad \left(1 - \frac{\mu}{n}\right)^{-x} \xrightarrow{n \rightarrow \infty, x, \mu \ll n} 1$$

Damit erhalten wir zusammengefasst das folgende Resultat (**Grenzwertsatz von Poisson**):

Satz:

Vor.:

$$\mu = n \cdot p = \lim_{n \rightarrow \infty, p \rightarrow 0} n \cdot p$$

Beh.:

$$f(x) \xrightarrow{n \rightarrow \infty, x, \mu \ll n} \frac{\mu^x}{x!} \cdot e^{-\mu}$$

Definition:

$$f(x) = \frac{\mu^x}{x!} \cdot e^{-\mu} \text{ heisst } \textbf{Poissonverteilung } Po(\mu) \quad *^1$$

$$\textbf{Konsequenz: } x \geq 0 \Rightarrow F(x) = e^{-\mu} \cdot \sum_{s \leq x} \frac{\mu^s}{s!}$$

Bemerkung:

Da wegen $p \rightarrow 0$ die Wahrscheinlichkeiten der untersuchten Ereignisse sehr klein werden können, spricht man auch von der Verteilung der **seltenen Ereignissen**.

Momente

$$G(t) = \sum_i e^{tx_i} f(x_i) = \sum_i e^{tx_i} e^{-\mu} \cdot \frac{\mu^x}{x_i!} \Rightarrow G'(t) = e^{-\mu+e^t} \mu^{t+1} \mu \Rightarrow G'(0) = \mu \text{ O.k.!}$$

Damit ist aber für die Praxis nichts gewonnen. Um weiterzukommen erinnern wir uns, dass die Poissonverteilung eine Annäherung der Binomialverteilung für $x < n$ und $n \rightarrow \infty$ ist. Daher können wir in der Praxis für μ den Mittelwert der beobachteten relativen Häufigkeiten nehmen, wenn die genannten Bedingungen annähernd erfüllt sind und es sich um ein Bernoulliexperiment handelt.

$$\rightsquigarrow \mu \approx \frac{1}{n} \cdot \sum_{k=1}^j n_k \cdot x_k$$

$$G''(t) = e^{-\mu+e^t} \mu^{t+1} \mu (1 + e^t \mu) \Rightarrow G''(0) = \mu(1 + \mu) = \sigma^2 + \mu^2 \Rightarrow \sigma^2 = \mu^2, \sigma = \mu$$

$$\text{Ebenso berechnet man: } \gamma = \frac{1}{\sigma^3} \cdot E((X - \mu)^3) = \frac{1}{\sqrt{\mu}}$$

Satz:**Vor.:**

Poissonverteilung

Beh.:

$$\begin{aligned} \mu &\approx \frac{1}{n} \cdot \sum_{k=1}^j n_k \cdot x_k \\ \sigma^2 &= \mu^2, \sigma = \mu \\ \gamma &= \frac{1}{\sqrt{\mu}} \end{aligned}$$

Bsp.:

Wie gross ist die Wahrscheinlichkeit, dass an einer Schule mit $n = 1000$ Studierenden mindestens $k = 1$ Studierende am 1. Mai Geburtstag haben?

Bernoulliexperiment:

$$p = \frac{1}{365}, n = 1000, P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \left(\frac{365-1}{365}\right)^{1000} \approx 0.935654.$$

Poissonapproximation:

$$P(X = k) \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \lambda = n \cdot p = \frac{1000}{365}, P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) \approx 1 - \frac{\lambda^0}{0!} \cdot e^{-\lambda} \approx 0.935412.$$

4.8.4 Pascalverteilung

Das Experiment bestehe wieder aus unabhängigen Versuchen. Sei p im Gegensatz zum Bernoulliexperiment die gegebene Wahrscheinlichkeit für einen **Misserfolg** im Basisversuch. $1 - p$ ist daher hier die Wahrscheinlichkeit für den Erfolg. k sei der Parameter der Erfolge bis zum r -ten **Misserfolg (Schaden)**. X variiert auf $\{k_i\}$. An die Stelle von p^x bei der Bernoulliverteilung tritt jetzt $(1 - p)^k$, an die Stelle von $(1 - p)^{n-x}$ tritt p^r ($r = n - x = n - k$). Ausgewählt werden k Möglichkeiten aus einer Menge von $r + k - 1$ Möglichkeiten. (Die r Misserfolge beim Experiment ohne den letzten, der immer an letzter Stelle kommt und das Experiment abbricht, also die Anzahl der Auswahlmöglichkeiten nicht erhöht. k ist die Anzahl Erfolge.) Die Lösung des Auswahlproblems führt dann zu folgender Formel:

Formel:

$X = \text{Anzahl unabhängiger Versuche bis zum } r\text{-ten Misserfolg}, q = \text{Erfolgswahrscheinlichkeit (}p\text{: Misserfolg)}, n := k + r$.

$$nB_k := p_k = P(X = k) = \binom{r+k-1}{k} \cdot p^r \cdot (1-p)^k, \quad k = 0, 1, \dots, \quad P(X = n) = \binom{n-1}{r-1} \cdot p^r \cdot (1-p)^{n-r}$$

Da hier der Misserfolg p bestimmt, ist die folgende Definition plausibel:

Definition: $r > 0, \quad 0 < p < 1 \rightsquigarrow nB_k$ heisst **negative Binomialverteilung**.

Definition: Für $r \in \mathbb{N}$ heisst die Verteilung nB_k auch **Pascal-Verteilung** $Pa(r, p)$.

Anwendungen finden sich in der Schadenstheorie.

Die direkte Berechnung von μ und σ^2 von Hand ist etwas mühsam. Mit der bei der Binomialverteilung angewandten Methode und z. B. Mathematica ist man rasch am Ziel. Man findet:

Satz: $\mu = r \cdot \frac{1-p}{p}, \quad \sigma^2 = r \cdot \frac{1-p}{p^2} = \frac{\mu}{p}$

Bemerkung: $\mu < \sigma^2 \rightsquigarrow$ negative Binomialverteilung
 $\mu > \sigma^2 \rightsquigarrow$ Binomialverteilung
 $\mu = \sigma^2 \rightsquigarrow$ Poissonverteilung
Für $r = 1$ erhält man die geometrische Verteilung.

Konsequenz: Durch Schätzen von μ und σ^2 bei einer Stichprobe erhält man damit einen Hinweis auf den Typ der Verteilung.

Zur Berechnung der Wahrscheinlichkeiten kann man sofort eine Rekursionsformel ablesen:

Formel: $p_0 = p^r, \quad p_{k+1} = \frac{r+k}{k+1} (1-p) \cdot p_k$

Bsp.:

Gegeben: Maschine, die durchschnittlich alle 4 Wochen (20 Arbeitstage) einmal eingesetzt wird. Eine Revision nach einem Einsatz dauert 3 Tage. Wahrscheinlichkeit, dass sie genau am 1. Tag ($r = 1, \quad k = 0$) oder genau am 2. Tag ($r = 1, \quad k = 1$) oder genau am 3. Tag ($r = 1, \quad k = 2$) wieder gebraucht wird?

$$\rightsquigarrow p = \frac{1}{20}, \quad q = 1 - p, \quad P(X = k) = \binom{r+k-1}{r-1} \cdot p^r \cdot (1-p)^k, \quad r = 1$$

$$\rightsquigarrow P(X = 0) = 0.05, \quad P(X = 1) = 0.0475, \quad P(X = 2) = 0.0428688.$$

4.8.5 Geometrische Verteilung

Definition: Die **geometrische Verteilung** ist der Spezialfall der Pascalverteilung für $r = 1$.

Es gilt daher:

Satz:**Vor.:**

Geometrische Verteilung

Beh.:

$$p_k = P(X = k) = p \cdot (1 - p)^k = (1 - q) \cdot q^k, \quad k \in \mathbb{N}, \quad 0 < p < 1$$

$p = 1 - q$ ist die Wahrscheinlichkeit eines Misserfolgs, $q = 1 - p$ die Wahrscheinlichkeit eines Erfolgs und X gibt die zufällige Anzahl Erfolge bis zum 1. Misserfolg bei unabhängigen Versuchen. k sei der Parameter der Erfolge bis zum 1-ten **Misserfolg** ($k + 1$).

$$P(X \leq k) = P(X = 0) + \dots + P(X = k) = \sum_{j=0}^k (1 - q) \cdot q^j = (1 - q) \cdot \sum_{j=0}^k q^j = (1 - q) \cdot \frac{1 - q^{k+1}}{1 - q} = 1 - q^{k+1}$$

Satz:

$$P(X \leq k) = 1 - q^{k+1} \rightsquigarrow \text{Geometrische Folge!}$$

Aus der Pascalverteilung erhalten wir für $r = 1$ die Momente:

$$\text{Satz:} \quad \mu = \frac{1 - p}{p}, \quad \sigma^2 = \frac{1 - p}{p^2} = \frac{\mu}{p}$$

Bemerkung:

Die geometrische Verteilung (wegen der geometrischen Folge) ist das diskrete Analogon zur Exponentialverteilung.

Bsp.:

Gegeben: Maschine, die durchschnittlich alle 4 Wochen (20 Arbeitstage) einmal eingesetzt wird. Eben ist sie eingesetzt worden. Wahrscheinlichkeit, dass sie innerhalb der nächsten 10 Tagen gebraucht wird?

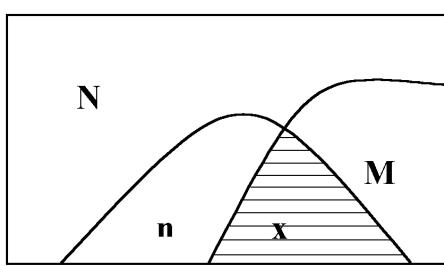
$$p = \frac{1}{20} = 0.05, \quad \mu = 20 - 1 = \frac{q}{1 - q} = \Rightarrow q = \frac{19}{20} = 0.95. \quad (p = 1 - q = \frac{1}{20} = 0.05).$$

$$P(X \leq 9) = 1 - P(X \geq 10) = 1 - \left(\frac{1}{20}\right)^{9+1} = 1 - 0.95^{10} \approx 0.401263.$$

4.8.6 Hypergeometrische Verteilung

Im Unterschied zur Bernoulliverteilung sind bei der hypergeometrischen Verteilung die Experimente nicht unabhängig. Im Modell der Bernoulliverteilung ziehen wir mit zurücklegen, im Modell der hypergeometrischen Verteilung hingegen ziehen wir **ohne zurücklegen**. p ändert daher hier von Versuch zu Versuch.

Modellversuch:



Modell:

Geg.: Gefäß mit N Elementen.

M Elemente sind defekt.

n Elemente werden gezogen ohne zurückzulegen.

x der gezogenen n Elemente sind defekt.

Problem:

Wahrscheinlichkeit, dass bei n mal ziehen x defekte Elemente kommen?

$$g = \text{günstige Fälle: Wähle } x \text{ aus } M \text{ und gleichzeitig } n - x \text{ aus } N - M. \rightsquigarrow g = \binom{M}{x} \cdot \binom{N - M}{n - x}$$

$$m = \text{mögliche Fälle: } n \text{ aus } N \text{ auswählen.} \rightsquigarrow m = \binom{N}{n}$$

$$\Rightarrow \frac{g}{m} = \frac{\binom{M}{x} \cdot \binom{N - M}{n - x}}{\binom{N}{n}}$$

Definition:

Beim beschriebenen Modellversuch ist die **hypergeometrische Verteilung** gegeben durch:

$$f(x) = \frac{\binom{M}{x} \cdot \binom{N - M}{n - x}}{\binom{N}{n}}$$

Bemerkung:

Der Name „hypergeometrische Verteilung“ kommt von der sogenannten „hypergeometrischen Funktion“.

Wie früher schon kann man auch hier $G(t)$ verwenden, um μ und σ^2 zu berechnen. Die Rechnung ergibt:

Satz:

Vor.:

Hypergeometrische Verteilung

Beh.:

$$\mu = n \cdot \frac{M}{N}$$

$$\sigma^2 = \frac{n \cdot M \cdot (N - M) \cdot (N - n)}{N^2 \cdot (N - 1)}$$

Bemerkung:

1. Sind $N, M, N - M$ gross und n klein, so ist die hypergeometrische Verteilung eine Approximation der Binomialverteilung. $\rightsquigarrow p = \frac{M}{N}$

2. Ist p klein, n gross und N gross gegenüber n , so ist die hypergeometrische Verteilung eine Approximation der Poissonverteilung. $\rightsquigarrow \mu = n \cdot p$

Bsp.:

Gegeben: Kiste mit 500 Kugellagern. Prüfung nach Vertrag: 50 (30) Lager entnehmen, untersuchen. Falls alle gut sind, wird die Lieferung akzeptiert. Sonst eben nicht. Nach Liefervertrag: Voraussichtlich maximal 2 % = 10 sind Ausschuss. Berechne: $P(k = 0)$.

$$N = 500, \ k = 0, \ p = \frac{2}{100} = 0.02 = \frac{M}{N} = \frac{M}{500} \Rightarrow M = 500 \cdot 0.02 = 10.$$

$$n = 50 \Rightarrow P(X = k = 0) = \frac{\binom{M}{k} \cdot \binom{N - M}{n - k}}{\binom{N}{n}} = \frac{\binom{10}{0} \cdot \binom{500 - 10}{50 - 0}}{\binom{500}{50}} = 0.345162.$$

$$n = 30 \Rightarrow P(X = k = 0) = \frac{\binom{M}{k} \cdot \binom{N - M}{n - k}}{\binom{N}{n}} = \frac{\binom{10}{0} \cdot \binom{500 - 10}{30 - 0}}{\binom{500}{30}} = 0.535489.$$

4.8.7 Beispiel

Bsp.:

Experiment: Eingespanntes Gewehr, 10 Schuss auf eine Scheibe, Trefferwahrscheinlichkeit pro Schuss = 0.1. Wie gross ist die Wahrscheinlichkeit, mindestens einen Treffer zu erzielen?

Trick: A : mindestens ein Treffer $\Rightarrow \bar{A}$: kein Treffer.

Sei A_0 : Treffer bei einem Schuss.

$$\rightsquigarrow P(A_0) = p = \frac{1}{10}, \ P(A) = 1 - P(\bar{A}), \ P(\bar{A}) = P(\bar{A}_0)^{10} \Rightarrow P(A) = 1 - P\left(\frac{9}{10}\right)^{10} \approx 0.651$$

\rightsquigarrow Verteilung?

4.9 Spezielle stetige Verteilungen

4.9.1 Allgemeines

Sei $f(x)$ eine stückweise stetige Wahrscheinlichkeitsdichte mit $D_f = \mathbb{R}$ und $F(x)$ die zugehörige Verteilungsfunktion. Im folgenden wollen wir nun spezielle Varianten von $f(x)$ diskutieren.

4.9.2 Rechtecksverteilung

Sei $f(x)$ die Wahrscheinlichkeitsdichte einer Zufallsvariablen X .

Die zu X gehörige Verteilung heisst **Rechtecksverteilung** oder **stetige gleichmässige Verteilung**, wenn gilt:

Definition:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \vee x > b \end{cases}$$

Man rechnet:

$$\begin{aligned} 1. \mu = m_1 = E(X^1) &= \int_{-\infty}^{\infty} x^1 \cdot f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{a+b}{2} \\ 2. \sigma^2 = m_2 - m_1^2 &= \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - \mu^2 = \int_a^b x^2 \cdot \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Satz:

Vor.:

Rechtecksverteilung

Beh.:

$$\mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}$$

Bsp.: Siehe Seite 69.

4.9.3 Normal- oder Gaussverteilung

Vorbemerkungen

Historisch gesehen ist die Normalverteilung schon von Gauss im Zusammenhang mit der Theorie der zufälligen (nicht systematische) Messfehler (Fehlerkurve) benutzt worden. Sie ist aus folgenden Gründen sehr wichtig:

1. In der Praxis sind viele Zufallsvariablen annähernd normalverteilt. Ein Problem bleibt aber die Tatsache, dass in der Praxis beliebig grosse Messgrößen wegen der physikalischen Beschränktheit nicht existieren.
2. Viele Messgrößen lassen sich einfach auf die Normalverteilung transformieren oder sind durch solche approximierbar.
3. Die Normalverteilung spielt auch bei statistischen Prüfverfahren eine Rolle.
4. Die Verteilung additiv überlagerter, zufälliger und unabhängiger Gesamtfehler hat als Grenzverteilung die Normalverteilung (**zentraler Grenzwertsatz**).

Verteilungsfunktion

Aus der Analysis wissen wir: $\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$

Hinweis zum Beweis:

$$\text{Sei } I(R) = \int_{-\infty}^R e^{-t^2} dt \Rightarrow (I(R))^2 = \int_{-\infty}^R e^{-x^2} dx \cdot \int_{-\infty}^R e^{-y^2} dy = \int_{-\infty}^R \left(\int_{-\infty}^R e^{-(x^2+y^2)} dx \right) dy$$

Statt über das Quadrat $[0, R] \times [0, R]$ zu integrieren, kann man auch über den innern und äussern Viertelskreis in Polarkoordinaten integrieren, die das Quadrat natürlich umschließen. Für $R \rightarrow \infty$ erhält man dann den Grenzwert $\sqrt{\pi}$.

Damit können wir folgern:

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi} \Rightarrow \int_{-\infty}^{\infty} e^{\frac{-t^2}{2}} dt = \sqrt{2\pi} \Rightarrow \int_{-\infty}^{\infty} e^{\frac{-t^2}{2 \cdot s^2}} dt = s \sqrt{2\pi} \Rightarrow \int_{-\infty}^{\infty} e^{\frac{-(t-m)^2}{2 \cdot s^2}} dt = s \sqrt{2\pi}$$

$$\Rightarrow \frac{1}{s \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{\frac{-(t-m)^2}{2 \cdot s^2}} dt = 1, \quad D_f = \mathbb{R}$$

Folgerung: $\frac{1}{s \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{\frac{-(t-m)^2}{2 \cdot s^2}} dt = 1 \Rightarrow f(x) := \frac{1}{s \sqrt{2\pi}} e^{\frac{-(t-m)^2}{2 \cdot s^2}}$ ist als stetige Verteilungsfunktion zulässig.

Definition:

$$\text{Sei } f(x) := \varphi(x; m, s^2) = \frac{1}{s \sqrt{2\pi}} e^{\frac{-(x-m)^2}{2 \cdot s^2}}$$

$$\text{Sei } \varphi(x) := \varphi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

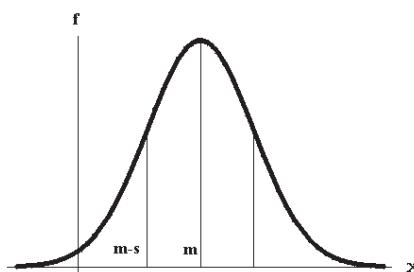
Die Dichtefunktion $f(x)$ definiert eine **Normalverteilung** oder **Gaussverteilung** mit den Parametern m und s zur Zufallsvariablen X .

Verteilungsfunktion:

$$F(x) := \Phi(x; m, s^2) = \frac{1}{s \sqrt{2\pi}} \cdot \int_{-\infty}^x e^{\frac{-(t-m)^2}{2 \cdot s^2}} dt$$

Symbol: $X \in N(m; s^2)$ ($\rightsquigarrow Z = \frac{-(X-m)^2}{2 \cdot s^2} \in N(0; 1)$)

Der Graph heisst auch **Gaußglocke**. Er ist symmetrisch zu $x = m$



Für den Erwartungswert gilt:

$$\mu = E(X) := \int_{-\infty}^{\infty} f(x) \cdot x dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{s \sqrt{2\pi}} e^{\frac{-(x-m)^2}{2 \cdot s^2}} dx = m$$

Berechnung der Varianz (z.B. Einsatz von Mathematica):

$$\sigma^2 = E(X^2) - \mu^2 = G''_t(0) - \mu^2 = \frac{d^2}{dt^2} \left(\int_{-\infty}^{\infty} e^{xt} f(x) dx \right) |_{t=0} - \mu^2 = s^2$$

Satz:Vor.:

$$X \in N(m; s^2)$$

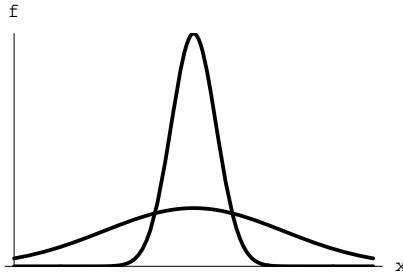
Beh.:

1. $\mu = m$
2. $\sigma^2 = s^2$

Definition:

Ist $\mu = 0$ und $\sigma^2 = 1$, so hat man die **standardisierte Normalverteilung**.

Flache Kurve: $\sigma = 1$, steile Kurve: $\sigma = 25$.

**Bemerkung:**

$$F(x) := \Phi(x; m, s^2) = \frac{1}{s \sqrt{2 \pi}} \cdot \int_{-\infty}^x e^{\frac{-(t-m)^2}{2 \cdot s^2}} dt$$

$$P(a < x \leq b) = F(b) - F(a)$$

$$\Phi(x; 0, 1) = \frac{1}{\sqrt{2 \pi}} \cdot \int_{-\infty}^x e^{\frac{-(t)^2}{2}} dt$$

$$\begin{aligned} F(x) &= \Phi(x; \mu, \sigma^2) = \Phi\left(\frac{x - \mu}{\sigma}; 0, 1\right) \\ &= \Phi(z; 0, 1) \end{aligned}$$

$$\rightsquigarrow X \in N(\mu; \sigma^2) \Leftrightarrow Z = \frac{X - \mu}{\sigma} \in N(0; 1)$$

Die Verteilungsfunktion $F(x)$ ist nicht elementar berechenbar. Man verwendet dazu Rechner (Error function, *Erf*) oder Tabellen. Oft ist $\Phi(z; 0, 1)$ tabelliert. Damit lässt sich auch $\Phi(x; \mu, \sigma^2)$ bestimmen.

Zwischen $\mu - \sigma$ und $\mu + \sigma$ liegt ca. 68% der Fläche unter der Kurve von $f(x)$, zwischen $\mu - 2\sigma$ und $\mu + 2\sigma$ liegt ca. 95.5% der Fläche und zwischen $\mu - 3\sigma$ und $\mu + 3\sigma$ liegt ca. 99.7% der Fläche.

Weitere Informationen siehe unter:

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

4.9.4 Grenzwertsätze von Moivre Laplace

Diese Sätze behandeln die Annäherung einer Bernoulliverteilung durch eine Normalverteilung.

4.9.5 Lokaler Grenzwertsatz

Geg.:

Bernoulliverteilung

$$f_n(x) = \binom{n}{x} p^x q^{n-x} \rightsquigarrow \mu = n \cdot p, \sigma^2 = n \cdot p \cdot q, q = 1 - p, 0 < p < 1$$

Normalverteilung

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2}}, \mu = n \cdot p, \sigma^2 = n \cdot p \cdot q, q = 1 - p, z = \frac{x - \mu}{\sigma} \in \mathbb{R}$$

Wir wollen zeigen, dass für „kleine“ x $f_n(x)$ gleichmässig gegen $f(x)$ konvergiert für alle x in einem beliebigen endlichen Intervall.

Satz:

Vor.:

Wie beschrieben

Beh.:

$$f_n(x) \sim f(x)$$

Symbol: „ \sim “: lies **asymptotisch gleich**

Beweis:

Wir greifen auf die Formel von Stirling zurück (Seite 30):

$$\rightsquigarrow k! \approx \sqrt{2\pi k} \left(\frac{k}{e}\right)^k, k! = \sqrt{2\pi k} \cdot \left(\frac{k}{e}\right)^k \cdot e^{\frac{\Phi}{12k}}, 0 < \Phi < 1$$

Der Beweis dieser Formel ist mit unseren Mitteln ein grösseres Unterfangen, vgl. z.B. Amann u. Escher, Analysis II (Bibl.: A2) oder van der Waerden Bibl. A13.

$$\begin{aligned} \rightsquigarrow f_n(x) &= \binom{n}{x} p^n q^{n-x} = \frac{n!}{x! \cdot (n-x)!} p^x q^{n-x} = \\ &\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{\Phi_n}{12n}} \cdot \frac{1}{\sqrt{2\pi x}} \cdot \left(\frac{x}{e}\right)^{-x} \cdot e^{-\frac{\Phi_x}{12x}} \cdot \frac{1}{\sqrt{2\pi(n-x)}} \cdot \left(\frac{(n-x)}{e}\right)^{-(n-x)} \cdot e^{-\frac{\Phi(n-x)}{12(n-x)}} \cdot p^x \cdot q^{n-x} = \\ &\frac{\sqrt{2\pi n}}{\sqrt{2\pi(n-x)} \cdot \sqrt{2\pi x}} \cdot \left(\frac{n}{e}\right)^n \cdot \left(\frac{e}{x}\right)^x \cdot \left(\frac{e}{(n-x)}\right)^{(n-x)} \cdot e^{\frac{\Phi_n}{12n}} \cdot e^{-\frac{\Phi_x}{12x}} \cdot e^{-\frac{\Phi(n-x)}{12(n-x)}} \cdot p^x \cdot q^{n-x} = \\ &\frac{1}{\sqrt{2\pi}} \cdot n^{n+\frac{1}{2}} \cdot p^x \cdot \frac{1}{x^{x+\frac{1}{2}}} \cdot \frac{1}{(n-x)^{(n-x)+\frac{1}{2}}} \cdot q^{n-x} \cdot e^{\frac{\Phi_n}{12n} - \frac{\Phi_x}{12x} - \frac{\Phi(n-x)}{12(n-x)}} = \\ &\frac{1}{\sqrt{2\pi} \cdot n \cdot p \cdot q} \cdot \left(\frac{n \cdot p}{x}\right)^{(x+\frac{1}{2})} \cdot \left(\frac{n \cdot q}{(n-x)}\right)^{(n-x+\frac{1}{2})} \cdot e^{\tau} = \frac{1}{\sqrt{2\pi} \cdot n \cdot p \cdot q} \cdot e^{-h} \cdot e^{\tau}, \\ \tau &= \frac{1}{12} \cdot \left(\frac{\Phi_n}{n} - \frac{\Phi_x}{x} - \frac{\Phi(n-x)}{(n-x)}\right), \Phi_i \in (0, 1), h = (x + \frac{1}{2}) \cdot \ln\left(\frac{n \cdot p}{x}\right) + (n - x + \frac{1}{2}) \cdot \ln\left(\frac{n \cdot q}{(n-x)}\right) \end{aligned}$$

Sei $z = \frac{x - \mu}{\sigma} = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot q}}$, $|z| < K = \text{const.}$, $p + q = 1$

$$\rightsquigarrow h(x) := w(z) = (n \cdot p + \frac{1}{2} + z \sqrt{n \cdot p \cdot q}) \cdot \ln(1 + z \sqrt{\frac{q}{n \cdot p}}) + (n \cdot q + \frac{1}{2} - z \sqrt{n \cdot p \cdot q}) \cdot \ln(1 - z \sqrt{\frac{p}{n \cdot q}})$$

n genügend gross $\rightsquigarrow u_1 = |z \sqrt{\frac{q}{n \cdot p}}| < 1$, $u_2 = |z \sqrt{\frac{p}{n \cdot q}}| < 1$

$\rightsquigarrow \ln(1 + u_1), \ln(1 - u_2)$ sind daher in Potenzreihen um das Zentrum 1 entwickelbar:

$$\ln(u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \frac{u^4}{4} + \dots$$

$$\rightsquigarrow \ln(1 + z \sqrt{\frac{q}{n \cdot p}}) = z \sqrt{\frac{q}{n \cdot p}} - \frac{z^2}{2} \frac{q}{n \cdot p} + \dots, \ln(1 - z \sqrt{\frac{p}{n \cdot q}}) = -z \sqrt{\frac{p}{n \cdot q}} - \frac{z^2}{2} \frac{p}{n \cdot q} + \dots$$

$$\rightsquigarrow w(z) = (n \cdot p + \frac{1}{2} + z \sqrt{n \cdot p \cdot q}) \cdot (z \sqrt{\frac{q}{n \cdot p}} - \frac{z^2}{2} \frac{q}{n \cdot p} + \dots) + (n \cdot q + \frac{1}{2} - z \sqrt{n \cdot p \cdot q}) \cdot (-z \sqrt{\frac{p}{n \cdot q}} - \frac{z^2}{2} \frac{p}{n \cdot q} + \dots)$$

Hier multiplizieren wir aus, ordnen nach Potenzen von z und vereinfachen ($q = 1 - p$, z.B. Mathematica verwenden), so erhalten wir:

$$w(z) = \left(\frac{q-p}{2\sqrt{npq}} \cdot z + z^2 \cdot \left(\frac{1}{2} - \frac{1}{2n(1-p)} - \frac{1}{4np(1-p)} - \frac{p}{2n(1-p)} \right) + \frac{1}{\sqrt{n}} \cdot R(n,p) \cdot z^3, \right. \\ \left. |R(n,p)| < \text{const.} (z \in (0,1)) \right)$$

$$\rightsquigarrow \lim_{n \rightarrow \infty} w(z) = \frac{1}{2} \cdot z^2, \quad z = \frac{x-\mu}{\sigma} = \frac{x-n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \in [\alpha, \beta] \subset \mathbb{R}$$

gleichmässige Konvergenz

$$\rightsquigarrow e^{-h(x)} = e^{-w(z)} \rightarrow e^{z^2/2} \quad \text{für } n \rightarrow \infty$$

$$\alpha \leq z = \frac{x-n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \leq \beta \Rightarrow x \geq np + \alpha \sqrt{npq} = np(1 + \alpha \sqrt{\frac{q}{np}}), \\ n - x \geq nq - \beta \sqrt{(npq)} = nq(1 - \beta \sqrt{\frac{p}{nq}})$$

$$\Rightarrow |\tau| = \left| \frac{1}{12} \cdot \left(\frac{\Phi_n}{n} - \frac{\Phi_x}{x} - \frac{\Phi(n-x)}{(n-x)} \right) \right| < \frac{1}{12} \cdot \left(\frac{1}{n} + \frac{1}{x} + \frac{1}{(n-x)} \right) \\ < \frac{1}{12} \cdot \left(\frac{1}{n} + \frac{1}{np(1+\alpha\sqrt{\frac{q}{np}})} + \frac{1}{nq(1-\beta\sqrt{\frac{p}{nq}})} \right) \rightarrow 0 \quad \text{für } n \rightarrow \infty$$

$$\rightsquigarrow e^\tau \rightarrow 1 \quad \text{für } n \rightarrow \infty$$

$$\rightsquigarrow f_n(x) = \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} \cdot e^{-h} \cdot e^\tau \rightarrow \underbrace{\frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}}}_{0} \cdot e^{z^2/2} \cdot \underbrace{e^0}_{1} = 0, \quad f_n(x) \sim \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} \cdot e^{z^2/2}$$

Satz:

Lokaler Grenzwertsatz

Vor.:

$$\alpha \leq z = \frac{x-n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \leq \beta, \quad (z \in [\alpha, \beta]), \\ n \rightarrow \infty$$

Beh.:

$$f_n(x) \sim \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} \cdot e^{z^2/2}$$

4.9.6 Grenzwertsatz von De Moivre/ Laplace

Der Grenzwertsatz von De Moivre/ Laplace ist ein Korollar des lokalen Grenzwertsatzes. Er ist ein Spezialfall des zentralen Grenzwertsatzes, der hier nicht behandelt wird (vgl. Lit.).

Sei $[\dots]$ die Gaussklammerfunktion (Floor, Int.).

Satz:**Grenzwertsatz von De Moivre/ Laplace**Vor.:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du, \quad a < b, \quad \alpha = \left[\frac{a - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right], \quad \beta = \left[\frac{b - n \cdot p + 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right],$$

$$\alpha \leq z = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \leq \beta$$

Beh.:

$$P(a \leq X \leq b) = \sum_{x=a}^b \binom{n}{x} p^x q^{n-x} \sim \Phi(\beta) - \Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2} du$$

4.9.7 Das Gesetz von Bernoulli der grossen Zahlen

Das **Gesetz der grossen Zahlen** liefert eine Wahrscheinlichkeitsaussage zur Vermutung, dass sich die relative Häufigkeit für grosse n verhält wie die Wahrscheinlichkeit.

Vor.:

Sei A : Ereignis mit der Wahrscheinlichkeit $p \in (0, 1)$ bei einem Zufallsexperiment.

Sei X = Anzahl des Eintreffens von A beim n -maligen Ausführen des Experiments.

Sei $\varepsilon \in \mathbb{R}^+$.

Satz:**Bernoulli**Beh.:

$$P\left(\left|\frac{x}{n} - p\right| < \varepsilon\right) \xrightarrow{n \rightarrow \infty} 1$$

Bemerkung:

Der Satz besagt also: Die Wahrscheinlichkeit oder Chance, dass die Abweichung der relativen Häufigkeit von der Wahrscheinlichkeit $p = P(X)$ kleiner als jedes beliebige positive ε wird, wenn nur n genügend gross ist, konvergiert gegen 1.

Achtung: Nur die Wahrscheinlichkeit konvergiert, nicht aber etwa die Differenz $\left|\frac{x}{n} - p\right|$. Eine grosse Wahrscheinlichkeit bedeutet nicht etwa Sicherheit!

Wir können die Behauptung aus dem Satz von De Moivre/ Laplace folgern:

Beweis:

$$\begin{aligned} \left|\frac{x}{n} - p\right| < \varepsilon &\Leftrightarrow -\varepsilon < \frac{x}{n} - p < \varepsilon \Leftrightarrow p - \varepsilon < \frac{x}{n} < p + \varepsilon \Leftrightarrow a := (p - \varepsilon) \cdot n < x < (p + \varepsilon) \cdot n := b \\ \Rightarrow P\left(\left|\frac{x}{n} - p\right| < \varepsilon\right) &= P((p - \varepsilon) \cdot n < x < (p + \varepsilon) \cdot n) = P(a < x < b) = P(a \leq x \leq b), \end{aligned}$$

$$\text{De Moivre/ Laplace} \rightsquigarrow P(a \leq x \leq b) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2} du \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{-u^2/2} du$$

Denn es gilt:

$$\alpha = \left[\frac{a - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{(p - \varepsilon) \cdot n - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{-\varepsilon \cdot n - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{-\varepsilon \cdot \sqrt{n}}{\sqrt{p \cdot q}} + \frac{0.5}{\sqrt{p \cdot q}} \right] \rightarrow -\infty,$$

$$\text{beta} = \left[\frac{b - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{(p + \varepsilon) \cdot n - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{+\varepsilon \cdot n - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{+\varepsilon \cdot \sqrt{n}}{\sqrt{p \cdot q}} + \frac{0.5}{\sqrt{p \cdot q}} \right] \rightarrow \infty$$

4.9.8 Bemerkung zum Zufall

Wie wir eingangs gesehen haben, wird beim absoluten Zufall ein Grund ausgeschlossen, beim relativen Zufall ist ein Grund jedoch bloss nicht wahrnehmbar, existiert jedoch. Nun macht es aber keinen Sinn über die Existenz einer nicht wahrnehmbaren Sache zu rätseln. Man kann da nur noch glauben, nicht glauben oder eben die Frage Frage sein lassen.

Wir können daher von folgendem Begriffsverständnis ausgehen: Für den beobachtenden Menschen ist es Zufall, wenn das Eintreten eines Ereignisses unvorhergesehen, unbeabsichtigt ist, ohne erkennbare Ursache oder Gesetzmässigkeit. Zufall ist es, wenn das Eintreten eines Ereignisses nicht notwendig aus einem gegebenen Gesamtereignis folgt, wenn alles auch hätte anders verlaufen können. Bloss eine Wirkung ist sichtbar, eine Ursache aber nicht.

Bei grösseren Ereignismengen werden aber nun dennoch Gesetzmässigkeiten sichtbar. Die relativen Häufigkeiten von Experimenten nähern sich den aus Modellen herleitbaren theoretischen Wahrscheinlichkeiten an (Gesetz der grossen Zahlen). Zufällige Ursachen haben hier also gesetzmässige, empirisch beobachtbare Wirkungen. Man kann daher trotzdem die sinnvolle Frage stellen, ob die Resultate nicht doch aus einem komplexen Wirkungszusammenhang gesetzmässig folgen. Die Vorhersage des Resultates könnte daher nur praktisch zu kompliziert, theoretisch aber möglich sein. (Z.B. Galton-Brett-Experiment!). Zufall wäre daher eine Folge der Komplexität. Die Sache kann aber auch umgekehrt sein: Relativ exakt bekannte Ursachen haben manchmal trotz bekannter Naturgesetze eine recht zufällige Wirkung. Z.B. bei der Wettervorhersage oder bei der Berechnung der Erdposition im Sonnensystem vor Jahrtausenden. Die unvermeidlichen Messfehler können sich bei Berechnungen derart fortpflanzen, dass der Endfehler den Wertebereich des Resultats übersteigt. Kleinste Veränderungen der Ursachen können grösste Abweichungen in der Wirkung zur Folge haben (Schmetterlingseffekt).

Man denke in diesem Zusammenhang an die Ausschweifungen des Determinismus (Laplace'scher Dämon, Negation des freien Willens) oder an die Unschärferelation (gesetzmässige Beschreibung des Unbekannten).

4.9.9 Tschebyscheffsche Ungleichung

Sei Y eine beliebige Zufallsgrösse, die nicht normalverteilt sein muss. $E(Y)$ sei der Erwartungswert, $Var(Y)$ sei die Varianz und ε sei eine beliebige positive Zahl. Dann gilt (ohne Beweis, vgl. Anhang):

Formel:

Tschebyscheff

$$P(|Y - E(Y)| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}$$

4.9.10 Logarithmische Normalverteilung

$$\text{Sei } f(t) := \varphi(t; \mu_L, \sigma_L^2) = \frac{1}{\sigma_L \sqrt{2\pi}} e^{\frac{-(t-\mu_L)^2}{2\sigma_L^2}}$$

= Dichtefunktion der Normalverteilung, Parameter μ_L und σ_L , Zufallsvariablen Y .

$$\rightsquigarrow \text{Verteilungsfunktion: } F(y) := \Phi(y; \mu_L, \sigma_L^2) = \frac{1}{\sigma_L \sqrt{2\pi}} \cdot \int_{-\infty}^y e^{\frac{-(t-\mu_L)^2}{2\sigma_L^2}} dt$$

$\rightsquigarrow Y$ ist normalverteilt.

Sei $X = \log(Y)$, $x = \log(y)$ (Substitution)

Falls in $f(y)$ einfach $y = \log_a(x)$, $a^y = x$ gesetzt wird, entspricht das Vorgehen nicht der bekannten Substitutionsregel. Wegen der Normierung muss gelten:

$$1 = \int_{-\infty}^{\infty} f(y) dy = \int_{x=a^{-\infty}}^{x=a^{\infty}} f(\log_a(x)) \frac{dy}{dx} dx = \int_0^{\infty} f(\log_a(x)) \frac{\log_a(e)}{x} dx$$

Definition: Gegeben sei die Dichtefunktion:

$$h(x) = \begin{cases} 0 & x \leq 0 \\ \frac{\log_a(e)}{x} \varphi(\log_a(x); \mu_L, \sigma_L^2) & x > 0 \end{cases}$$

Die durch $h(x)$ gegebene Verteilung heisst **logarithmische Normalverteilung** oder **Logonormalverteilung**.

Die Verteilungsfunktion erhält man aus der Substitution von $F(y) := \Phi(y; \mu_L, \sigma_L^2)$:

Formel:
$$H(x) = \frac{1}{\sigma_L \sqrt{2\pi}} \cdot \int_0^x e^{\frac{-(\log_a(z) - \mu_L)^2}{2\sigma_L^2}} \frac{\log_a(e)}{t} dt$$

Für $a = e$ erhält man durch Berechnung von Erwartungswert, Streuung und Median für die Logonormalverteilung:

Formel:
$$\begin{aligned} E(X) &= e^{\mu_L + \sigma_L^2/2} \\ Var(X) &= e^{2\mu_L + \sigma_L^2} (e^{\sigma_L^2} - 1) \\ x_{0.5} &= e^{\mu_L} \end{aligned}$$

Anwendung: Zeitstudien, Lebensdaueranalysen.

Normalverteilung: Additive Überlagerungen \rightsquigarrow Logonormalverteilung: Multiplikative Überlagerungen.

4.9.11 Exponentialverteilung

Es gilt:

$$\int_0^{\infty} \alpha e^{-\alpha x} dy = -\frac{\alpha}{\alpha} e^{-\alpha x} \Big|_0^{\infty} = -(e^{-(\infty)} - e^{-(0)}) = -(0 - 1) = 1$$

Definition:

Die stetige Zufallsvariable X genügt einer **Exponentialverteilung** mit dem Parameter α , wenn sie folgende Dichtefunktion besitzt:

$$f(x) = \begin{cases} 0 & x < 0 \\ \alpha e^{-\alpha x} & x \geq 0 \end{cases}$$

\rightsquigarrow Verteilungsfunktion:

$$\int_{-\infty}^x f(t) dt = \int_0^x \alpha e^{-\alpha t} dt = 1 - e^{-\alpha x}$$

Formel:

$$F(x) = \int_0^x f(t) dt = \begin{cases} 0 & x < 0 \\ 1 - e^{-\alpha x} & x \geq 0 \end{cases}$$

Erwartungswert, Streuung und Median:

$$\mu = E(X) = \int_0^\infty x \cdot \alpha e^{-\alpha x} dx = x \cdot \frac{\alpha}{-\alpha} e^{-\alpha x} \Big|_0^\infty - \int_0^\infty (-1) e^{-\alpha x} dx = 0 - 0 + \int_0^\infty e^{-\alpha x} dx = \frac{1}{-\alpha} e^{-\alpha x} \Big|_0^\infty = \frac{1}{\alpha}$$

$$\sigma = \text{Var}(X) = \int_0^\infty x^2 \cdot \alpha e^{-\alpha x} dx - \mu^2 = \dots = \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}$$

$$F(x_{0.5}) = 1 - e^{-\alpha x_{0.5}} = \frac{1}{2} \Rightarrow \dots \Rightarrow x_{0.5} = \frac{1}{\alpha} \ln(2)$$

Anwendung: Zeitmessungen (Atomzerfall, Arbeitszeiten, ...), Lebensdauer.

4.9.12 Weibullverteilung

Definition:

Eine stetige Zufallsvariable X besitzt eine dreiparametrische **Weibull-Verteilung** mit den Parametern $a > 0$, $b > 0$ und c , wenn ihre Dichtefunktion wie folgt gegeben ist:

$$f(x) = \begin{cases} 0 & x \leq x \\ \frac{b}{a} \left(\frac{x-c}{a}\right)^{b-1} e^{-\left(\frac{x-c}{a}\right)^b} & x > c \end{cases}$$

\rightsquigarrow Verteilungsfunktion:

Formel:

$$F(x) = \begin{cases} 0 & x \leq x \\ 1 - e^{-\left(\frac{x-c}{a}\right)^b} & x > c \end{cases}$$

$c = 0 \rightsquigarrow$ Zweiparametrische Verteilung

$a = 1 \rightsquigarrow$ Reduzierte Verteilung

Bekannt aus der Analysis:

Symbol: Sei $\Gamma(x)$ die **Gamma-Funktion**.

Definition: $\Gamma(x) := \int_0^\infty e^{-t} t^{x-1} dt, \quad x > 0$

Formel: $\Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma(x) = (x-1)\Gamma(x-1)$

Durch Berechnung erhält man:

Formel: $\mu = E(X) = c + a \cdot \Gamma\left(1 + \frac{1}{b}\right)$
 $\sigma^2 = Var(X) = a^2 \left(\Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right)\right)$
 $x_{0.5} = c + a (\ln(2))^{1/b}$

Anwendung: Lebensdauer- und Zuverlässigkeitssanalysen.

4.9.13 Gammaverteilung

Definition: Eine stetige Zufallsvariable X besitzt eine **Gammaverteilung** mit den Parametern $b > 0$ und $p > 0$, wenn ihre Dichtefunktion wie folgt gegeben ist:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} & x > 0 \end{cases}$$

Bemerkung:

Die Exponentialverteilung ergibt sich als Spezialfall der Gammaverteilung für $b = \alpha$ und $p = 1$. Für $p \in \mathbb{N}$ heisst die Verteilung auch **Erlangverteilung**.

Formel: $\mu = E(X) = \frac{p}{b}, \quad \sigma^2 = Var(X) = \frac{p}{b^2}$

Anwendung: Zuverlässigkeitssanalysen (Lebensdauerverteilung), Bedienungszeitenverteilung.

4.9.14 Ausblick

Um weitere wichtige Verteilungen behandeln zu können, müssen wir erst Kenntnisse der mehrdimensionalen Zufallsgrößen (Zufallsvektoren) erwerben. Wichtig sind die Prüfverteilungen für statistische Tests.

4.10 Zufallsvektoren und deren Verteilung

↪ Mehrdimensionale Wahrscheinlichkeit

4.10.1 Fragestellung, Begriffe

Zufallsvektor, Verteilungsfunktion

Bis jetzt haben wir **eindimensionale Zufallsgrößen** X betrachtet, die Werte $\in \mathbb{R}$ annehmen können. Die Verteilungsfunktion ist dann eine Funktion einer unabhängigen Variablen.

In einem Zufallsexperiment können jedoch verschiedene Ereignisse eintreten, die unabhängig oder abhängig sein können. Ebenso können zur Beschreibung eines Zufallsexperimentes mehrere Zufallsvariablen notwendig sein, die unabhängig oder abhängig sein können. Wir wollen hier den Modellfall von zwei Zufallsvariablen kurz studieren. ($N = 2$.)

Bsp.:

Bei der Qualitätskontrolle eines Loses von Zahnradwellen werden die Dicke X und gleichzeitig am selben Stück jeweils auch die Länge Y kontrolliert. Dicke und Länge werden jeweils in einem separaten Arbeitsgang gefertigt. Sie können daher als unabhängig angesehen werden. Da die Messungen jedoch paarweise vorgenommen werden, hängen Y und X dennoch irgendwie zusammen. Das führt zu Spezialfragen der mathematischen Statistik, auf die hier nicht eingegangen werden kann.

Seien X und Y zwei diskrete oder kontinuierliche Zufallsvariablen, die bei einem Zufallsexperiment auftreten. $\rightsquigarrow X$ kann Werte annehmen in $\{x_1, x_2 \dots x_k \dots\}$ oder in \mathbb{R} (allgemein $x \in \mathbb{R}$) und Y Werte in $\{y_1, y_2 \dots y_k \dots\}$ oder in \mathbb{R} (allgemein $y \in \mathbb{R}$).

Definition: Der Vektor $\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ heisst **zweidimensionaler Zufallsvektor**.
 Entsprechend ist $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ ein n -**dimensionaler Zufallsvektor**

Sei $A = \{\omega\}$ ein Ereignis, das bei einem Zufallsexperiment mit zwei Zufallsvariablen eintritt. Dann nimmt X den Wert $X(A) = x$ an und gleichzeitig nimmt Y den Wert $Y(A) = y$ an. D.h. $\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ nimmt den Wert $\begin{pmatrix} y \\ x \end{pmatrix}$ (oder kurz das Wertepaar (x, y)) an. $\begin{pmatrix} y \\ x \end{pmatrix}$ ist dann eine **Realisierung** von \vec{X} .

Bemerkung:

Statt von einem Paar (X, Y) von Zufallsvariablen zu sprechen, kann man auch von einem **Wahrscheinlichkeitsvektor** $\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ sprechen. Der Wahrscheinlichkeitszusammenhang des Paars (X, Y) (Wahrscheinlichkeitsfunktion) ist dann gegeben durch eine Doppelfolge $\langle (x_i, y_k, p_{i,k}) \rangle$.

(Für die Entwicklung der Theorie beschränken wir uns hier auf $N = 2$. Die Übertragung auf $N > 2$ ist problemlos.)

Für die zugehörigen Wahrscheinlichkeiten definieren wir:

Definition: $F(x, y) = P(X \leq x \wedge Y \leq y)$, $x \in \mathbb{R}$, $y \in \mathbb{R}$
 Im diskreten Fall ist speziell:
 $p_{ik} := P(X = x_i \wedge Y = y_k)$

$\rightsquigarrow F(x, y)$ ist eine Funktion $\mathbb{R}^2 \xrightarrow{F} [0, 1]$.

Eigenschaften:

1. F ist in X und in Y monoton wachsend sowie rechtsseitig stetig.
2. $F(-\infty, y) = 0, \ F(x, -\infty) = 0, \ F(\infty, \infty) = 1$

Diskreter Fall: $\sum_{i,k} p_{ik} = 1$

3. $x_1 < x_2 \wedge y_1 < y_2 \Rightarrow F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \leq 0$
4. $P(X \leq x) = P(X \leq x, Y \leq \infty) = F(x, \infty)$
 $P(Y \leq y) = P(X \leq \infty, Y \leq y) = F(\infty, y)$

Die ersten beiden Eigenschaften sind unmittelbar einsichtig. Bei der vierten Eigenschaft handelt es sich um eine **Spezialisierung** (Projektion oder Restriktion von $D_F = \mathbb{R}^N$ auf \mathbb{R}^{N-1}). Die dritte folgt wegen:

$$\begin{aligned}
 & F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \\
 &= (P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1)) - (P(X \leq x_1, Y \leq y_2) + P(X \leq x_1, Y \leq y_1)) \\
 &= \underbrace{(P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1))}_{=P(X \leq x_2, y_1 < Y \leq y_2)} - \underbrace{(P(X \leq x_1, Y \leq y_2) - P(X \leq x_1, Y \leq y_1))}_{=P(X \leq x_1, y_1 < Y \leq y_2)} \\
 &= P(X \leq x_2, y_1 < Y \leq y_2) - P(X \leq x_1, y_1 < Y \leq y_2) \\
 &= P(x_1 < X \leq x_2, y_1 < Y \leq y_2) \geq 0 \quad (P(\dots, \dots) \in [0, 1]!) \quad \text{☺}
 \end{aligned}$$

Randverteilungen, unabhängige Zufallsvariablen

Für die Restriktionen definieren wir:

Definition:

Randverteilungen der zweidimensionalen Zufallsvariablen \vec{X} :

$$\begin{aligned}
 F_X(x) &:= F(x, \infty) = P(X \leq x) = P(X \leq x, Y \leq \infty) \\
 F_Y(y) &:= F(\infty, y) = P(Y \leq y) = P(X \leq \infty, Y \leq y)
 \end{aligned}$$

Bemerkung:

$F_X(x)$ und $F_Y(y)$ sind die Verteilungsfunktionen der Komponenten X und Y .

Sei $A = \{X \leq x, Y \leq \infty\}, \ B = \{X \leq \infty, Y \leq y\} \Rightarrow A \cap B = \{X \leq x, Y \leq y\}$

Für unabhängige Ereignisse gilt nun:

$$\begin{aligned}
 P(A \cap B) &= P(A) \cdot P(B) \\
 \rightsquigarrow F(x, y) &= P(X \leq x, Y \leq y) = P(A \cap B) = P(A) \cdot P(B) \\
 &= P(X \leq x, Y \leq \infty) \cdot P(X \leq \infty, Y \leq y) = F_X(x) \cdot F_Y(y)
 \end{aligned}$$

Daher können wir den Begriff der Unabhängigkeit von den Ereignissen auf die Variablen übertragen:

Definition:

X und Y heißen **unabhängig**, wenn gilt:

$$\forall_{(x,y)} F(x, y) = F_X(x) \cdot F_Y(y)$$

Identisch verteilte Variablen

Sei $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$, $\vec{X}^T = (X_1, \dots, X_n) \rightsquigarrow F_{\vec{X}}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$

Definition: Falls alle Vektorkomponenten X_i , ($i = 1, \dots, n$) dieselbe Verteilungsfunktion $F_Z(Z)$, ($Z = X_1, \dots, X_n$) besitzen, so heißen die X_i **identisch verteilt**.

Definition: Die Variablen X_i , ($i = 1, \dots, n$) heißen **vollständig unabhängig**, wenn jede der Variablen von allen andern unabhängig ist.

Mittels der Definition für $n = 2$ kann man folgern:

Satz:

Vor.:

X_i , ($i = 1, \dots, n$) vollständig unabhängig und identisch verteilt

Beh.:

$$\forall_{(x_1, x_2, \dots, x_n)} F_{\vec{X}}(x_1, x_2, \dots, x_n) = F(x_1) \cdot F(x_2) \cdot \dots \cdot F(x_n)$$

Bemerkung:

F bestimmt in diesem Fall also $F_{\vec{X}}$.

4.10.2 Der diskrete Fall

Sei $p_{ik} := P(X = x_i, Y = y_k) := f(x_i, y_k)$, $in \in \mathbb{N}$

Allgemein sagen wir:

Definition: $\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ heisst **diskret**, wenn $\{x_i, y_k\}$ resp. $\{x_i, y_k, p_{ik}\}$ abzählbar ist.

Formel:

$$F(x, y) = \sum_{x_i \leq x, y_k \leq y} p_{ik}$$

Bsp.:

Bei der Qualitätskontrolle eines Loses von Zahnradwellen werden die Dicke d und gleichzeitig am selben Stück jeweils auch die Länge l kontrolliert. (Dicke und Länge werden jeweils in einem separaten Arbeitsgang gefertigt. Sie können daher als unabhängig angesehen werden.) Wir können die Werte von X und Y frei wie folgt festlegen:

| Kriterium: | X | Y |
|----------------------------|-----------|-----------|
| d innerhalb der Toleranz | $x_1 = 0$ | — |
| d ausserhalb Toleranz | $x_2 = 1$ | — |
| l innerhalb der Toleranz | — | $y_1 = 0$ |
| l ausserhalb Toleranz | — | $y_2 = 1$ |

Aus Erfahrung wissen wir:

| Kriterium: | Menge: |
|--------------------|--------|
| Ausschuss | 5% |
| d falsch | 1% |
| l falsch | 3% |
| d und l falsch | 1% |

Konsequenz:

| X | 0 | 1 | |
|-----|-----------------|-----------------|-----------------|
| Y | | | |
| 0 | $p_{11} = 0.95$ | $p_{21} = 0.01$ | $p_{.1} = 0.96$ |
| 1 | $p_{12} = 0.03$ | $p_{22} = 0.01$ | $p_{.2} = 0.04$ |
| | $p_{1.} = 0.98$ | $p_{2.} = 0.02$ | $p_{tot} = 1$ |

(Z.B. ist: $p_{11} + p_{21} = p_{.1}$, $p_{.1} = P(Y = 0) \dots$)

Definition: Wahrscheinlichkeiten wie $f_Y(Y = 1) := p_{.1}$ heissen **Randsummen**.

Übersicht über die Randsummen:

$$\begin{aligned} p_{.1} &= P(Y = 0) = p_{11} + p_{21} = 0.96 \\ p_{.2} &= P(Y = 1) = p_{12} + p_{22} = 0.04 \\ p_{1.} &= P(X = 0) = p_{11} + p_{12} = 0.98 \\ p_{2.} &= P(X = 1) = p_{21} + p_{22} = 0.02 \end{aligned}$$

Es gilt:

Satz:

$$\begin{aligned} f_X(X = x_i) &= p_{i.} = P(X = x_i) = \sum_k p_{ik} \\ f_Y(Y = y_i) &= p_{.k} = P(Y = y_k) = \sum_i p_{ik} \\ \sum_k p_{ik} &= p_{i.} p_{.k} = 1 \end{aligned}$$

Aus der allgemeinen Definition der Unabhängigkeit von Variablen folgt:

Korollar: Diskrete Variablen X und Y unabhängig

$$\forall_{(i,k)} p_{ik} = p_{i.} \cdot p_{.k}$$

Bsp.: $p_{1.} \cdot p_{.1} = 0.98 \cdot 0.96 \approx 0.94 \neq p_{11} = 0.95$

→ In diesem Beispiel sind die Variablen **nicht unabhängig!** → dans cet exemple les variables **ne sont pas indépendantes!**

4.10.3 Der stetige Fall

Wir betrachten wieder den zweidimensionalen Modellfall.

Definition:

Wir nennen \vec{X} stetig, wenn \vec{X} überabzählbar-unendlich viele reelle Wertpaare annehmen kann und zudem eine **Wahrscheinlichkeitsdichte** $f(x, y)$ gegeben ist mit einer (stückweise) stetigen Verteilungsfunktion $F(x, y)$ mit:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$$

Aus den allgemeinen Definitionen folgt:

Folgerung:

$$1. \quad f(x, y) \geq 0$$

$$2. \quad F(\infty, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) du dv = 1$$

Für die **Dichte der Randverteilungen** gilt:

Folgerung:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Aus der Unabhängigkeitsdefinition durch die Verteilungsfunktion ($F(x, y) = F_X(x) \cdot F_Y(y)$) kann man mit Hilfe des Mittelwertsatzes der Integralrechnung folgern:

Satz:Vor.:

X, Y unabhängig

Beh.:

$$f(x, y) = f(x) \cdot f(y)$$

4.11 Mehrdimensionale Erwartung

4.11.1 Erwartung, Mittelwert

Wir betrachten wieder den Modellfall $n = 2$. Sei $f(x, y)$ die Wahrscheinlichkeitsfunktion resp. die Dichte. Analog zum eindimensionalen Fall kann man auch jetzt Kenngrößen für X oder Y definieren:

Definition:

Erwartungswert $E(g(X, Y))$ einer gegebenen Funktion $g(X, Y)$:

$$E(g(X, Y)) : = \begin{cases} \sum_i \sum_k g(x_i, y_k) \cdot f(i, k) & dis. \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f(x, y) dx dy & cont. \end{cases}$$

Voraussetzung:

$$\sum_i \sum_k |g| \cdot f \quad \text{oder} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g| \cdot f dx dy$$

exist. resp. conv.

Wie im eindimensionalen Fall folgert man (Rechnung):

Satz: $E(a g(X, Y) + b h(X, Y)) = a E(g(X, Y)) + b E(h(X, Y))$

Speziell: $a = b = 1, g(X, Y) = X, h(X, Y) = Y \rightsquigarrow$

Korollar: $E(X + Y) = E(X) + E(Y)$

Verallgemeinerung:

Satz: **Additionssatz für Mittelwerte**

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

Allgemein ist $E(X^2) \neq E(X)^2$.

Bsp.: Würfeln $\rightsquigarrow E(X) = \frac{7}{2}$
 $E(X^2) = \sum_{k=1}^6 k^2 \cdot \frac{1}{6} = \frac{91}{6} \neq \left(\frac{7}{2}\right)^2 = E(X)^2$

Für unabhängige Variablen gilt jedoch infolge von $f(x, y) = f(x) \cdot f(y)$ resp. von $F(x, y) = F(x) \cdot F(y)$ z.B. im diskreten Fall:

$$E(X Y) = \sum_i \sum_k x_i \cdot y_k \cdot f(x_i, y_k) = \sum_i x_i \cdot f_X(x_i) \cdot \sum_k y_k \cdot f_Y(x_k) = E(X) \cdot E(Y)$$

Entsprechend im stetigen Fall. \rightsquigarrow

Konsequenz:

$$X, Y \text{ unabhängig} \Rightarrow E(X Y) = E(X) \cdot E(Y)$$

Verallgemeinerung:

Satz: **Multiplikationssatz für Mittelwerte**

Vor.: $i \neq k \Rightarrow X_i, X_k \text{ unabhängig}$

Beh.:

$$E(X_1 \cdot X_2 \cdot \dots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \dots \cdot E(X_n)$$

4.11.2 Varianz, Kovarianz, Korrelation

Bekannt vom Fall mit einer Variablen:

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

Wir definieren:

Definition:

Varianzen:

$$\begin{aligned}\sigma_X^2 &= \text{Var}(X) = E((X - \mu_X)^2) \\ \sigma_Y^2 &= \text{Var}(Y) = E((Y - \mu_Y)^2)\end{aligned}$$

Sei $Z = X + Y$

$$\begin{aligned}\sigma_{X+Y}^2 &= \sigma_Z^2 = E(Z^2) - (\mu_Z)^2 = E(Z^2) - (E(Z))^2 = E(X^2 + 2XY + Y^2) - (E(X + Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - (E(X) + E(Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\ &= (E(X^2) - E(X)^2) + 2(E(XY) - E(X)E(Y)) + (E(Y^2) - E(Y)^2) \\ &= \sigma_X^2 + 2(E(XY) - E(X)E(Y)) + \sigma_Y^2\end{aligned}$$

Definition:

$\sigma_{XY} := \text{cov}(XY) := (E(XY) - E(X)E(Y))$
heisst **Kovarianz** von X und Y

Es gilt:

$$\begin{aligned}E((X - \mu_X)(Y - \mu_Y)) &= E(X \cdot Y - X \cdot \mu_Y - \mu_X \cdot Y + \mu_X \cdot \mu_Y) \\ &= E(X \cdot Y) - \mu_Y \cdot E(X) - \mu_X \cdot E(Y) + \mu_X \cdot \mu_Y \cdot E(1) = E(X \cdot Y) - \mu_Y \cdot \mu_X - \mu_X \cdot \mu_Y + \mu_X \cdot \mu_Y \\ &= E(X \cdot Y) - \mu_X \cdot \mu_Y = E(X \cdot Y) - E(X)E(Y) \Rightarrow E((X - \mu_X)(Y - \mu_Y)) = \sigma_{XY}\end{aligned}$$

Definition:

$\rho_{XY} := \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{cov}(XY)}{\sigma_X \sigma_Y}$
heisst **Korrelationskoeffizient** von X und Y
($\sigma_X, \sigma_Y \neq 0$)

Man sieht unmittelbar:

Satz:

1. $\sigma_{X+Y}^2 = \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = \sigma_X^2 + \sigma_Y^2 + 2\text{cov}(XY)$
2. $\sigma_{XY} = \text{cov}(XY) = E(XY) - \mu_X \mu_Y$
 $= E(X \cdot Y) - E(X)E(Y) = E((X - \mu_X)(Y - \mu_Y))$
3. $\sigma_{XY} = \sigma_{YX}$, $\text{cov}(XY) = \text{cov}(YX)$
4. $X = Y \Rightarrow |\rho_{XY}| = 1$
5. $|\rho_{XY}| \leq 1$

Zur letzten Behauptung (die andern sind bewiesen) betrachten wir als Beispiel den stetigen Fall:

$$\begin{aligned}\rho_{XY} &:= \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{(E((X - \mu_X)^2)E((Y - \mu_Y)^2))^{\frac{1}{2}}} \\ &\Rightarrow \rho_{XY} \cdot (E((X - \mu_X)^2)E((Y - \mu_Y)^2))^{\frac{1}{2}} = E((X - \mu_X)(Y - \mu_Y)) \\ &\Rightarrow \rho_{XY} \cdot \underbrace{\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x, y) dx dy \right)^{\frac{1}{2}}}_{I_1} \cdot \underbrace{\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_Y)^2 \cdot f(x, y) dx dy \right)^{\frac{1}{2}}}_{I_2} = \rho_{XY} \cdot (I_1 \cdot I_2)^{\frac{1}{2}}\end{aligned}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy \Rightarrow \rho_{XY} \cdot (I_1 \cdot I_2)^{\frac{1}{2}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy$$

Hier wenden wir auf $I_1 \cdot I_2$ die Ungleichung von Cauchy–Schwarz an.

$$(\text{Cauchy–Schwarz: } \int_G |f g|)^2 \leq (\int_G f^2) \cdot (\int_G g^2), \quad \int_G f g \leq \int_G |f g| \Rightarrow (\int_G f g)^2 \leq (\int_G f^2) \cdot (\int_G g^2))$$

(Die Ungleichung von Cauchy–Schwarz ist bekannt als „Skalarprodukt–Ungleichung“. Sie gilt auch im diskreten Fall entsprechend.)

Daraus folgert man:

$$\rho_{XY} \cdot \left(\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy \right)^2 \right)^{\frac{1}{2}} \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy \Rightarrow |\rho_{XY}| \leq 1$$

(•)

Seien X und Y unabhängig $\rightsquigarrow P(X Y) = P(X) \cdot P(Y) \Rightarrow \text{cov}(X Y) = \sigma_{XY} = E(X Y) - E(X) E(Y) = E(X) E(Y) - E(X) E(Y) = 0 \Rightarrow \rho_{XY} = \frac{\text{cov}(X Y)}{\sigma_X \sigma_Y} = 0, (\sigma_X, \sigma_Y \neq 0)$

Satz:

Vor.:

X und Y unabhängig

Beh.:

$$\text{cov}(X Y) = \rho_{XY} = 0$$

Bemerkung:

Im Falle $\text{cov}(X Y) = E((X - \mu_X)(Y - \mu_Y)) = 0$ kann man $E((X - \mu_X)(Y - \mu_Y))$ als ein Skalarprodukt begreifen, dass 0 ist. Man hat es daher mit senkrecht stehenden Vektoren zu tun.

Im Falle $\rho_{XY} = 1$ wird die oben angewandte Cauchy–Schwarzsche Ungleichung zu einer Gleichung. In diesem Fall kann man den folgenden Satz folgern (Beweis vgl. Kreyszig, Bibl. A10):

Satz:

Zwischen zwei Zufallsvariablen X und Y mit positiver Varianz besteht dann und nur dann eine lineare Beziehung $Y = \alpha X + \beta$ resp. $X = \gamma Y + \delta$, wenn $|\rho_{XY}| = 1$ gilt.

Konsequenz:

Der Korrelationskoeffizient (vgl. auch Seite 180) ist somit ein **Mass für den linearen Zusammenhang zweier Zufallsvariablen**. Dieser Sachverhalt kann zur Prüfung eines linearen Zusammenhangs benutzt werden.

Achtung:

Ein formaler Zusammenhang beweist keinen kausalen Zusammenhang!

Bsp.: (Beispiel zu „formalem und kausalem Zusammenhang“.) Von Norden bis Süden nimmt die Körpergrösse h ab, jedoch die Anzahl der Katholiken k zu. Die Korrelation $h \sim \frac{1}{k}$ ist aber bekanntlich eine **Scheinkorrelation!**

Bemerkung:

Scheinkorrelation findet man häufig in interessengesetzlichen Argumenten. Für den Laien sind sie oft plausibel...

Sei $X + Y = Z$, X, Y unkorreliert \rightsquigarrow
 $cov(X, Y) = 0 = E(XY) - E(X)E(Y) \Rightarrow E(XY) = E(X)E(Y) \rightsquigarrow X, Y$ unabhängig

Oben haben wir hergeleitet:

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = \sigma_X^2 + \sigma_Y^2 + 2cov(XY) + \sigma_X^2 + \sigma_Y^2 + 0 \Rightarrow Var(X+Y) = Var(X) + Var(Y)$$

Satz:

Vor.:

X_1, \dots, X_n unkorreliert

Beh.:

1. X_1, \dots, X_n unabhängig
2. $Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$

4.11.3 Der diskrete Fall

Folgerung:

1. $E(g(X, Y)) = E(g(X))$
 $\Rightarrow E(g(X)) = \sum_i g(x_i) \sum_k f(x_i, y_k) = \sum_i g(x_i) \cdot f_X(x_i)$
 \rightsquigarrow Formel für eine einzige Variablen, o.k.!

2. **Erwartungswerte:**

$$\begin{aligned} \mu_X &:= E(X) := \sum_i x_i \cdot f_X(x_i) = \sum_i \sum_k x_i \cdot p_{ik} \\ \mu_Y &:= E(Y) := \sum_k y_k \cdot f_X(y_k) = \sum_k \sum_i y_k \cdot p_{ik} \end{aligned}$$

4.11.4 Der stetige Fall

Analog zum diskreten Fall kann man auch jetzt Kenngrößen für X oder Y definieren:

Definition:

$$\begin{aligned}
 1. \quad & E(g(X, Y)) = E(g(X)) \\
 \Rightarrow \quad & E(g(X)) = \int_{-\infty}^{\infty} g(x) \int_{-\infty}^{\infty} f(x, y) dy dx = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx \\
 \rightsquigarrow \quad & \text{Formel für eine einzige Variabel, o.k.!}
 \end{aligned}$$

2. **Erwartungswerte:**

$$\begin{aligned}
 \mu_X := E(X) &:= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) dx dy \\
 \mu_Y := E(Y) &:= \int_{-\infty}^{\infty} y \cdot f_X(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \cdot f(x, y) dx dy
 \end{aligned}$$

4.12 Mehrdimensionale Verteilungen

4.12.1 Zweidimensionale Normalverteilung

Im stetigen Fall spielt diese Verteilung eine besondere Rolle.

Definition:

Gegeben seien die Parameter $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho_{XY}$. $\binom{X}{Y}$ besitzt eine zweidimensionale Normalverteilung, wenn die Dichte durch folgende Funktion gegeben ist:

$$\begin{aligned}
 f(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \cdot e^{-\frac{h(x, y)}{2}} \\
 h(x, y) &= \frac{1}{(1-\rho_{XY}^2)} \left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\frac{\rho_{XY}(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right) \\
 x, y &\in \mathbb{R}
 \end{aligned}$$

Ohne Beweis folgender Satz:

Satz:

Vor.:

1. \vec{X} besitzt eine zweidimensionale Normalverteilung
2. $\text{conv}(XY) = 0$

Beh.:

1. X, Y unabhängig

$$2. \quad f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \cdot e^{-\frac{1}{2} \left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right)}$$

$$3. \quad \mu_X = \mu_Y = 0, \quad \sigma_X^2 = \sigma_Y^2 = 1 \quad \Rightarrow \quad f(x, y) = \frac{1}{2\pi} \cdot e^{-\frac{1}{2}(x^2+y^2)}$$

4.12.2 Stichprobenfunktion, Testverteilungen

Bemerkung:

Die in der Statistik wichtigen Verteilungen kann man in zwei Klassen einteilen: Verteilungen, die im Zusammenhang von **mathematischen Modellen** von Zufallsexperimenten auftreten und Verteilungen, **Testverteilungen** oder **Prüfverteilungen**, die die Grundlage statistischer Tests bilden. Testverteilungen sind hier Verteilungen von Stichprobenfunktionen.

Vor.:

Seien im Folgenden alle Komponenten X_i , $i = 1, \dots, n$ von \vec{X} unabhängig und identisch normalverteilt mit dem Erwartungswert μ und der Varianz σ^2 . (\sim Stichprobe vom Umfang n mit $N(\mu, \sigma^2)$. Die Grundgesamtheit sei normalverteilt.)

Mittelwert-Verteilung

Diese Verteilung ist die Verteilung der Stichprobenfunktion \bar{X} . Sie ist wichtig bei der Behandlung der Vertrauensintervalle (**Konfidenzintervalle**) für den Mittelwert einer Normalverteilung bei bekannter Varianz. Dabei gilt:

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

Satz: \bar{X} ist normalverteilt mit dem Erwartungswert μ und der Varianz $\frac{\sigma^2}{n} \sim N(\mu, \frac{\sigma^2}{n})$

Bemerkung:

Es ist wichtig zu bemerken, dass der Satz auch näherungsweise noch gilt, wenn die dazugehörige Grundgesamtheit mit denselben Parametern beliebig verteilt ist.

Sei $\bar{X} \xrightarrow{h} Z = h(\bar{X}) = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$

Die Koordinatentransformation h bewirkt, dass der Satz ein Korollar ist aus folgendem Lemma:

Lemma:

Vor.:

$$\begin{aligned} & X_i, \quad i = 1, \dots, n \text{ von } \vec{X} \\ & X_i \text{ unabhängig, identisch normalverteilt mit } N(\mu, \sigma^2) \\ & \bar{X} = \frac{1}{n} (X_1 + X_2 + \dots, X_n) \\ & Z = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \end{aligned}$$

Beh.:

Z genügt der standardisierten Normalverteilung nach $N(0, 1)$

Dieses Lemma wiederum ist ein Korollar (Spezialfall) eines allgemeineren Satzes, wie man unmittelbar einsehen kann:

Satz:**Summe unabhängiger normalverteilter Variablen****Vor.:**

Seien X_1, X_2, \dots, X_n unabhängig, normalverteilt mit
 Mittelwerte $\mu_1, \mu_2, \dots, \mu_n$
 Varianzen $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$

Beh.:

$$\sum_{i=1}^n X_i \text{ normalverteilt}$$

$$\text{Mittelwert } \mu = \sum_{i=1}^n \mu_i$$

$$\text{Varianz } \sigma^2 = \sum_{i=1}^n \sigma_i^2$$

Für den Beweis benötigen wir folgende Formel:

Lemma:**Vor.:**

$$Z = g(X, Y) = X + Y \quad X, Y \text{ unabh.}$$

$$\text{Dichten: } X \sim f_1(x), Y \sim f_2(Y)$$

Beh.:

$$F(Z) = \int_{-\infty}^{\infty} f_2(y) \left(\int_{-\infty}^{z-y} f_1(x) dx \right) dy$$

$$f_1 \in \text{stetig} \Rightarrow f(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-y) dx = \int_{-\infty}^{\infty} f_1(z-x) f_2(y) dy$$

Beweis Lemma:

$$\text{Von Seite 102: } f(x, y) = f_1(x) \cdot f_2(y)$$

$$F(z) = \int_{x+y \leq z} \int f(x, y) dx dy = \int_{x+y \leq z} \int f_1(x) \cdot f_2(y) dx dy = \int_{-\infty}^{\infty} (f_2(y) \int_{-\infty}^{x=z-y} f_1(x) \cdot dx) dy \quad \text{☺}$$

$$z = x + y, \quad \frac{d z}{d x} = 1 \Rightarrow F(z) = \int_{-\infty}^{\infty} (f_2(y) \int_{-\infty}^z f_1(z-y) \cdot dx) dz \Rightarrow f(z) := \frac{d}{d z} F(z) =$$

$$\frac{d}{d z} \int_{-\infty}^{\infty} (f_2(y) \int_{-\infty}^z f_1(z-y) dz) dy = \int_{-\infty}^{\infty} (f_2(y) \frac{d}{d z} \int_{-\infty}^z f_1(z-y) dz) dy = \int_{-\infty}^{\infty} (f_2(y) f_1(z-y)) dy$$

Zieht man bei der Berechnung von F statt f_2 die Funktion f_1 nach vorne, so folgt:
 $f(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-y) dy \quad \text{☺}$

Beweis Satz:

1. Induktion, Verankerung; $n = 1$: $X = X_1$ ist normalverteilt.

2. Induction, Vererbungsgesetz von $n = 1$ auf $n = 2$.

Sei zuerst: $X = X_1 + X_2$

$$\begin{aligned} \text{Nach Vor.: } f_1(x) &= \frac{1}{\sqrt{2\pi}\sigma_1} \cdot e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})}, \quad f_2(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \cdot e^{-\frac{1}{2}(\frac{x-\mu_2}{\sigma_2})} \\ \Rightarrow f(x) &= \int_{-\infty}^{\infty} f_1(x-y) f_2(y) dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \cdot e^{-\frac{1}{2}(\frac{x-y-\mu_1}{\sigma_1})} \frac{1}{\sqrt{2\pi}\sigma_2} \cdot e^{-\frac{1}{2}(\frac{y-\mu_2}{\sigma_2})} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\frac{x-y-\mu_1}{\sigma_1} + \frac{y-\mu_2}{\sigma_2})} dy \end{aligned}$$

$$\text{Sei } \mu := \mu_1 + \mu_2, \quad \sigma^2 := \sigma_1^2 + \sigma_2^2, \quad V := -\frac{1}{2} \left(\frac{x-y-\mu_1}{\sigma_1} + \frac{y-\mu_2}{\sigma_2} \right)$$

$$\text{Sei } V_1 := \frac{\sigma}{\sigma_1\sigma_2} \left(y - \frac{\sigma_1^2\mu_2 + \sigma_2^2(x-\mu_1)}{\sigma^2} \right), \quad V_2 := \frac{x-\mu}{\sigma}$$

Durch ausmultiplizieren, gleichnamig machen und umformen zeigt man, dass die folgende Identität gilt (viele Terme \rightsquigarrow z.B. mit Mathematica!):

$$V \equiv V_1^2 + V_2^2$$

V_2 ist unabhängig von y , $V_1 := \tau$ wird als neue Integrationsvariable verwendet:

$$\begin{aligned} \rightsquigarrow f(x) &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}(V_1^2+V_2^2)} dy = \frac{1}{2\pi\sigma_1\sigma_2} \cdot e^{-\frac{1}{2}V_2^2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}V_1^2} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot e^{-\frac{1}{2}V_2^2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\tau^2} dy \end{aligned}$$

$$\text{Es gilt: } \frac{dV_1}{dy} = \frac{d\tau}{dy} = \frac{\sigma}{\sigma_1\sigma_2} \Rightarrow dy = \frac{\sigma_1\sigma_2}{\sigma} d\tau, \quad \int_{-\infty}^{\infty} e^{-\frac{1}{2}\tau^2} d\tau = \sqrt{2\pi} \quad (\text{p. 89})$$

$$\Rightarrow f(x) = \frac{1}{2\pi\sigma_1\sigma_2} \cdot e^{-\frac{1}{2}V_2^2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\tau^2} \frac{\sigma_1\sigma_2}{\sigma} d\tau = \frac{1}{2\pi\sigma} \cdot e^{-\frac{1}{2}V_2^2} \cdot \sqrt{2\pi} = \frac{\sqrt{2\pi}}{\sigma} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

\rightsquigarrow Für $n = 2$ ist X damit normalverteilt.

3. Induktionsschluss von $n = m$ auf $n = m + 1$:

Vor.: $Y_1 = X_{m+1}$, $Y_2 = \sum_{i=1}^m X_i$ normalverteilt.

Gesetz: Y_1, Y_2 normalverteilt

\rightsquigarrow Wie bei X_1, X_2 auch Summe normalverteilt.

\rightsquigarrow Beh.: $Y_1 + Y_2 = \sum_{i=1}^m X_1 + X_{m+1} = \sum_{i=1}^{m+1} X_i$ normalverteilt. ☺

Durch Integralsubstitution findet man noch den folgenden Satz:

Satz:

Vor.:

X normalverteilt mit $N(\mu, \sigma^2)$
 $\tilde{X} = c_1 \cdot X + c_2, \quad c_1, c_2 = \text{const.}$ (Lineare Transf.)

Beh.:

\tilde{X} normalverteilt
 $\tilde{\mu} = c_1 \cdot \mu + c_2, \quad \tilde{\sigma}^2 = c_1^2 \cdot \sigma^2$

Die Formeln für $\tilde{\mu}$ und $\tilde{\sigma}^2$ sind schon auf Seite 76 in allgemeiner Form hergeleitet worden.

4.12.3 Chi-Quadrat-Verteilung

Definition, Verteilungsfunktion

Sei $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ (empirische Streuung)

Zur Erinnerung: $X_i \sim N(\mu, \sigma^2)$

Damit kann man die folgende Stichprobenfunktion bilden:

Definition: $\chi^2 = \frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$

Im Falle von $N(0, 1)$ ($\mu = 0, \sigma^2 = 1$) erhalten wir: Au cas de $N(0, 1)$ ($\mu = 0, \sigma^2 = 1$) nous obtenons:

$$\chi^2 = \sum_{i=1}^n X_i^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

χ^2 ist wieder eine Zufallsvariable. Die Herleitung der **Wahrscheinlichkeitsdichte** von χ^2 ist ein etwas grösseres Unterfangen (vgl. z.B. Lit. Kreyszig, Bibl. A10). Aus Gründen des Umfangs soll hier auf die Wiedergabe verzichtet werden (vgl. Anhang). Man erhält:

Satz: Wahrscheinlichkeitsdichte von χ^2

$$f(x) = \begin{cases} 0 & x \leq 0 \\ K_n x^{\frac{n-2}{2}} e^{-\frac{x}{2}} & x > 0 \end{cases}$$

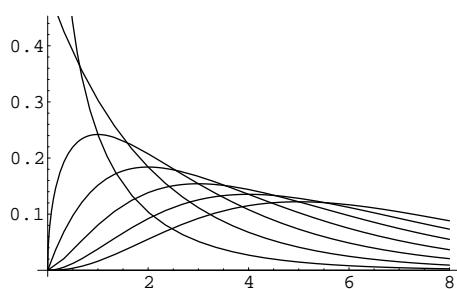
$$K_n := \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$$

Die Verteilungsfunktion erhält man durch Integration:

Formel: $F(x) = K_n \int_0^x u^{\frac{n-2}{2}} e^{-\frac{u}{2}} du$

(K_n ist auf Grund der Forderung $\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} f(x) dx = 1$ berechnet worden.)

Definition: n heisst die **Anzahl der Freiheitsgrade** der Verteilung. Γ ist die **Gammafunktion**.



Das Bild zeigt $f(x)$ die für die Freiheitsgrade $n = 0, 1, \dots, 7$

Gammafunktion, Gamma- und Betaverteilung

Für die Gammafunktion gilt:

1. $\Gamma(\alpha) := \int_0^\infty e^{-t} t^{\alpha-1} dt$
2. $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ (Partielle Integration)
3. $\Gamma(1) = \int_0^\infty e^{-t} t^{1-1} dt = \int_0^\infty e^{-t} dt = 1$
4. $\Gamma(2) = 1 \cdot \Gamma(1) = 1!, \quad \Gamma(3) = 2 \cdot \Gamma(2) = 2!, \quad \dots, \Gamma(n + 1) = n!$
5. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma\left(\frac{3}{2}\right) = \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$ u.s.w

Bemerkung:

Die Chi-Quadrat-Verteilung ist ein Sonderfall der **Gammaverteilung**.

Definition:

Sei $\alpha > 0$

Die **Gammaverteilung** ist durch die folgende Dichtefunktion gegeben:

$$f(x) = \begin{cases} 0 & x < 0 \\ e^{-x} \frac{x^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \end{cases}$$

Definition:

Sei $\alpha > 0, \beta > 0$

$B(\alpha, \beta)$ heisst **Beta-Funktion**:

$$B(\alpha, \beta) := \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Definition:

Sei $\alpha > 0, \beta > 0$

Die **Betaverteilung** ist durch die folgende Dichtefunktion gegeben:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} & x > 0 \end{cases}$$

4.12.4 Sätze zur Chi-Quadrat-Verteilung

Mit Hilfe der Formeln für die Momente berechnet man:

Satz:

Vor.:

Chi-Quadrat-Verteilung

Beh.:

$$1. \mu = n$$

$$2. \sigma^2 = 2n$$

Für grosse n kann man die Chi-Quadrat-Verteilung durch die Normalverteilung annähern. Es gilt:

Satz:

1. Die Zufallsvariable χ^2 ist asymptotisch normalverteilt mit $\mu = n$ und $\sigma^2 = 2n$. Für grosse n ist:

$$F(x) \approx \Phi\left(\frac{x - n}{\sqrt{2n}}\right)$$

2. Die Zufallsvariable $\sqrt{2\chi^2}$ ist asymptotisch normalverteilt mit $\mu = \sqrt{2n-1}$ und $\sigma^2 = 1$. Für grosse n ist:

$$F(x) \approx \Phi(\sqrt{2x}\sqrt{2n-1})$$

4.12.5 t-Verteilung von Student

Herleitung

Die **t-Verteilung** bildet die Grundlage für wichtige Tests. Sie wurde von W.S. Gosset unter dem Decknamen „Student“ publiziert.

Wir bilden zu den voneinander unabhängigen Stichprobenfunktionen X (normalverteilt) und Y (χ^2_n -verteilt) die neue Stichprobenfunktion t (oder T):

$$T = \frac{X}{\sqrt{Y}} \sqrt{n}, \quad t = \frac{X}{\sqrt{Y}/\sqrt{n}}$$

Zur Erinnerung: $X_i \in N(\dots), i = 1, 2, \dots, n \rightsquigarrow \bar{X} \in N(\dots)$

Z.B. für $\mu = 0$ und $\sigma^2 = 1$ ($N(0, 1)$) erhält man:

$$T = \frac{\bar{X}}{S} \sqrt{m} = \frac{\bar{X}}{\sqrt{Y/m}}, \quad Y = S^2, \quad m = n - 1 \quad (\text{Vgl. Vertrauensintervalle.})$$

Definition:

Die zu T resp. t gehörige Verteilung heisst **Student-Verteilung**.

Später werden wir zeigen: $Z = \bar{X}$ normalverteilt ($N(0, 1)$) und $Y = S^2$ ist $\chi^2_{m=n-1}$ -verteilt.

Definition: n resp. m heisst die **Anzahl der Freiheitsgrade** der Verteilung.

Die Herleitung der **Wahrscheinlichkeitsdichte** von T wollen wir auf später verschieben (vgl. z.B. Lit. Kreyszig, Bibl. A10 resp. Anhang). Man erhält ($T \rightsquigarrow z$):

Satz:

$$f(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{z^2}{n})^{(n+1)/2}}$$

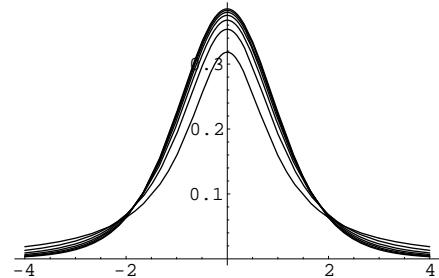
$$F(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \cdot \int_{-\infty}^z \frac{1}{(1 + \frac{u^2}{n})^{(n+1)/2}} du$$

Definition:Für $n = 1$ erhalten wir die **Chauchy–Verteilung**.**Sätze****Formel:**

1. $n = 1$: Die Chauchy–Verteilung besitzt keinen Mittelwert und keine Varianz.
2. $n = 2 \rightsquigarrow$ keine Varianz
3. $n \geq 2, n \in \mathbb{N} \rightsquigarrow \mu = 0$ (z^2 , Symmetrie!)
4. $n \geq 3 \Rightarrow \sigma^2 = \frac{n}{n-2}$

Das Bild zeigt $f(x)$ die für die Freiheitsgrade $n = 1, \dots, 7$

Es gilt:

**Satz:**

Für $n \rightarrow \infty$ strebt die Verteilungsfunktion $f(x)$ ($z \rightsquigarrow x$) der T –Verteilung gegen diejenige der standardisierten Normalverteilung ($N(0, 1)$).

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{x^2}{n})^{(n+1)/2}} \rightarrow \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

4.12.6 F–Verteilung von Fisher

Gegeben seien zwei voneinander unabhängige Stichproben mit den Umfängen n_1 und n_2 . Damit werden die Stichprobenfunktionen (Zufallsvariablen) $\bar{X}_1, \bar{X}_2, S_1^2, S_2^2$ gebildet: \rightsquigarrow

$$\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}, \quad S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2$$

$$\bar{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i}, \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2$$

S_1^2 und S_2^2 sind unabhängig.

Seien die zugehörigen Grundgesamtheiten normalverteilt mit den Parametern $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$. Zudem setzen wir voraus:

Vor.: $\sigma_1^2 = \sigma_2^2$

Nun bilden wir die folgende Stichprobenfunktion:

$$\mathcal{F} = \frac{S_1^2}{S_2^2} = \frac{\left(\frac{\chi_1^2 \sigma_1^2}{n_1 - 1}\right)}{\left(\frac{\chi_2^2 \sigma_2^2}{n_2 - 1}\right)} = \frac{\left(\frac{\chi_1^2}{n_1 - 1}\right)}{\left(\frac{\chi_2^2}{n_2 - 1}\right)} := \frac{\left(\frac{\chi_1^2}{m_1}\right)}{\left(\frac{\chi_2^2}{m_2}\right)}, \quad m_1 = n_1 - 1, \quad m_2 = n_2 - 1$$

Definition: Die zu \mathcal{F} gehörige Verteilung heisst **Fisher–Verteilung** oder **\mathcal{F} –Verteilung** mit (m_1, m_2) Freiheitsgraden.

Die Verteilungsfunktion und die Dichte berechnen sich wie folgt:

Satz: Vor.: \mathcal{F} –Verteilung

Beh.:

$$1. \quad F(x) = P(\mathcal{F} \leq x) = \frac{\Gamma(\frac{m_1 + m_2}{2})}{\Gamma(\frac{m_1}{2}) \Gamma(\frac{m_2}{2})} \cdot m_1^{\frac{m_1}{2}} \cdot m_2^{\frac{m_2}{2}} \cdot \int_0^x \frac{t^{\frac{m_1 - 2}{2}}}{(m_1 t + m_2)^{\frac{m_1 + m_2}{2}}} dt$$

$$2. \quad f(x) = \begin{cases} 0 & x \leq 0 \\ f(x) = E_{m_1, m_2} x^{\frac{m_1}{2} - 1} (m_2 + m_1 \cdot x)^{-\frac{m_1 + m_2}{2}} & x > 0 \end{cases}$$

(f aus F durch ableiten!)

4.13 Anhang I: Einige Beweise

4.13.1 Formel zur Gammafunktion

Im nächsten Abschnitt brauchen wir folgenden Satz:

Satz:Vor.:

$$a, b > 0$$

Beh.:

$$\frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)} = \int_0^1 u^{a-1} \cdot (1-u)^{b-1} du$$

Beweis:

$$\text{Es ist: } \Gamma(\alpha) = \int 0^\infty e^{-t} t^{\alpha-1} dt$$

$$\rightsquigarrow \Gamma(a) \cdot \Gamma(b) = \left(\int_0^\infty e^{-t} t^{a-1} dt \right) \cdot \left(\int_0^\infty e^{-v} v^{b-1} dv \right) = \int_0^\infty \int_0^\infty e^{-(t+v)} t^{a-1} v^{b-1} dt dv = \int_0^\infty \int_0^\infty h(t, v) dt dv$$

$$\text{Sei } t = r \cdot u, v = r(1-u) \Rightarrow \frac{\partial(t, v)}{\partial r, u} = \begin{vmatrix} \frac{\partial t}{\partial r} & \frac{\partial t}{\partial u} \\ \frac{\partial v}{\partial r} & \frac{\partial v}{\partial u} \end{vmatrix} = \begin{vmatrix} u & r \\ 1-u & -r \end{vmatrix} = -ur + r - (-ur) = -r$$
$$t = r \cdot u, v = r(1-u), t, r \in [0, \infty), u \in [0, 1] \text{ (Durch Betrachtung der Grenzfälle)}$$

$$\begin{aligned} \Rightarrow \Gamma(a) \cdot \Gamma(b) &= \int_0^\infty \int_0^\infty h(t(r, u), v(r, u)) \cdot \left| \frac{\partial(t, v)}{\partial r, u} \right| dr du = \int_0^1 \int_0^\infty e^{-(r)} (r u)^{a-1} (r(1-u))^{b-1} r dr du \\ &= \int_0^1 \int_0^\infty u^{a-1} (1-u)^{b-1} r dr du = \left(\int_0^1 e^{-(r)} (r)^{a+b-1} r dr \right) \left(\int_0^\infty e^{-(r)} (r)^{a+b-1} u^{a-1} (1-u)^{b-1} du \right) = \\ \Gamma(a+b) \Rightarrow \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)} &= \int_0^\infty e^{-(r)} (r)^{a+b-1} u^{a-1} (1-u)^{b-1} du \quad \text{QED} \end{aligned}$$

4.13.2 Dichte der Chi-Quadrat-Verteilung

 $X_i : \in N(0, 1) \rightsquigarrow \text{normalverteilt}, \mu = 0, \sigma^2 = 1. \quad \text{Es gilt:}$

1. $x < 0 \Rightarrow P(\chi^2 = X_1^2 + \dots + X_n^2 = x < 0) = 0 \Rightarrow (x < 0 \Rightarrow f(x) = 0)$
2. $(0 \leq x_i^2 \leq x \Leftrightarrow -\sqrt{x} \leq X_i \leq \sqrt{x}) \Rightarrow P(X_i^2 \leq x) = P(0 \leq X_i^2 \leq x) = P(-\sqrt{x} \leq X_i \leq \sqrt{x})$
3. $X_i : \in N(0, 1) \Rightarrow$

$$\forall_i F_i(x) = F_i(x) = P(X_i^2 \leq x) = P(-\sqrt{x} \leq X_i \leq \sqrt{x}) = 2 \cdot P(0 \leq X_i \leq \sqrt{x}) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{u^2}{2}} du$$

$$\text{Sei } v = u^2, u = v^{\frac{1}{2}} \Rightarrow \frac{du}{dv} = \frac{1}{2} v^{-\frac{1}{2}}, du = dv \frac{1}{2} v^{-\frac{1}{2}} = dv \frac{1}{2} \frac{1}{\sqrt{v}}$$

$$\Rightarrow F_1(x) = \frac{2}{\sqrt{2\pi}} \frac{1}{2} \int_0^{\sqrt{v}=u=\sqrt{x}} e^{-\frac{u^2}{2}} \cdot \frac{1}{\sqrt{v}} dv = \frac{1}{\sqrt{2\pi}} \int_0^{v=x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{v}} dv$$

$$\Rightarrow f_1(x) = \frac{dF_1(x)}{dx} = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{x}{2}}}{\sqrt{x}}, x > 0$$

$$4. n = 1 \Rightarrow \chi^2 = X_1^2, \rightsquigarrow f(x) = K_1 x^{\frac{1-2}{2}} e^{-\frac{x}{2}} = \frac{1}{2^{\frac{1}{2}} \Gamma(\frac{1}{2})} \cdot x^{\frac{1-2}{2}} e^{-\frac{x}{2}} = \Big|_{\Gamma(\frac{1}{2})=\sqrt{2}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} \cdot e^{-\frac{x}{2}} \quad \text{QED}$$

Damit ist die Formel für $n = m = 1$ bewiesen. Wir können somit einen Induktionsbeweis versuchen.

Sei $\chi^2 = X_1^2 + \dots + X_n^2 := \chi_n^2$, $f(x) := f_n(x)$

Sei die Formel richtig für $m=n-1$.

$$\rightsquigarrow f_{n-1}(x) = K_{n-1} x^{\frac{n-1-2}{2}} e^{-\frac{x}{2}} = \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} \cdot x^{\frac{n-1-2}{2}} e^{-\frac{x}{2}} = \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} \cdot x^{\frac{n-3}{2}} e^{-\frac{x}{2}}$$

Es gilt: $\chi_n^2 = \chi_{n-1}^2 + X_n^2 \quad \chi_{n-1} \leftrightarrow f_{n-1}(x) \wedge \chi_n \leftrightarrow f_n(x) = f_1(x)$

Nach Voraussetzung sind die X_i unabhängig und daher sind auch χ_{n-1} und X_n^2 unabhängig. Daher können wir die folgende Formel für die Dichte von Seite 109 benutzen:

$$f_n(x) = \int_{-\infty}^{\infty} f_n(x-y) \cdot f_{n-1}(y) dy = \int_0^x f_1(x-y) \cdot f_{n-1}(y) dy$$

$(\int_{-\infty}^{\infty} \dots = \int_0^x$ wegen $y < 0 \Rightarrow f_{n-1}(y) = 0$, $x > y \Rightarrow x-y < 0 \Rightarrow f_1(x-y) = 0$)

$$\begin{aligned} \rightsquigarrow f_n(x) &= \int_0^x f_1(x-y) \cdot f_{n-1}(y) dy = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2^{\frac{n-1}{2}} \cdot \Gamma(\frac{n-1}{2})} \cdot \int_0^x \frac{e^{-\frac{x-y}{2}}}{\sqrt{x-y}} \cdot y^{\frac{n-3}{2}} \cdot e^{-\frac{y}{2}} dy \\ &= \frac{1}{2^{\frac{n}{2}} \cdot \sqrt{\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot e^{-\frac{x}{2}} \cdot \int_0^x (x-y)^{-\frac{1}{2}} \cdot y^{\frac{n-3}{2}} dy := \frac{1}{2^{\frac{n}{2}} \cdot \sqrt{\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot e^{-\frac{x}{2}} \cdot I, \quad I = \int_0^x (x-y)^{-\frac{1}{2}} \cdot y^{\frac{n-3}{2}} dy \end{aligned}$$

Sei $y := ux$, $y \in [0, x] \Rightarrow u = \frac{y}{x} \in [0, 1]$, $x-y = x-u x = x(1-u)$, $\frac{dy}{du} = x$, $dy = du x$
 $\rightsquigarrow I = \int_0^x (x(1-u))^{-\frac{1}{2}} \cdot (u x)^{\frac{n-3}{2}} x du = (x^{-\frac{1}{2}} x^{\frac{n-3}{2}} x) \int_0^x (1-u)^{-\frac{1}{2}} \cdot (u)^{\frac{n-3}{2}} du = x^{\frac{n-2}{2}} \int_0^x (1-u)^{-\frac{1}{2}} \cdot u^{\frac{n-3}{2}} du$

Satz über die Gammafunktion von Seite 116

$$\rightsquigarrow I = x^{\frac{n-2}{2}} \int_0^x (1-u)^{\frac{1}{2}-1} \cdot u^{\frac{n-1}{2}-1} du = x^{\frac{n-2}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{1}{2})} \Gamma(\frac{n-1}{2} + \frac{1}{2}), \quad a = \frac{1}{2} > 0, \quad b = \frac{n-1}{2} > 0, \quad (n > 2)$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi} \Rightarrow f_n(x) = \frac{1}{2^{\frac{n}{2}} \cdot \sqrt{\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot e^{-\frac{x}{2}} \cdot x^{\frac{n-2}{2}} \frac{\Gamma(\frac{n-1}{2}) \cdot \sqrt{\pi}}{\Gamma(\frac{n-1}{2} + \frac{1}{2})} = \frac{1}{2^{\frac{n}{2}}} \cdot e^{-\frac{x}{2}} \cdot x^{\frac{n-2}{2}} \frac{1}{\Gamma(\frac{n}{2})} = f(x) \quad \text{QED}$$

Das ist die behauptete Formel.

4.13.3 Dichte der Student-Verteilung

Verteilungen nach Voraussetzung:

$$X \rightsquigarrow N(0, 1) \rightsquigarrow X \leftrightarrow f(x) := f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$Y \rightsquigarrow Y = \chi_n^2 \rightsquigarrow X \leftrightarrow f(y) := f_2(y) = \frac{1}{2^{\frac{n}{2}}} \cdot e^{-\frac{y}{2}} \cdot y^{\frac{n-2}{2}} \frac{1}{\Gamma(\frac{n}{2})}, \quad y > 0 \quad (y \leq 0 \Rightarrow f_2(y) = 0)$$

X, Y unabhängig $\rightsquigarrow f(x, y) = f_1(x) f_2(y)$

$$\begin{aligned}
F(z) &:= P(T \leq z) = P\left(\frac{X}{Y/n} \leq z\right) = P(X \leq z \cdot Y/n) \quad \rightsquigarrow F(z) = \int_{x \leq z \cdot Y/n} \int f(x, y) dx dy = \\
&= \int_{x \leq z \cdot \sqrt{y/n}, \quad y > 0} f_1(x) f_2(y) dx dy = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2^{\frac{n}{2}}} \cdot \frac{1}{\Gamma(\frac{n}{2})} \int_{x \leq z \cdot \sqrt{y/n}, \quad y > 0} \int_{x=z \cdot \sqrt{y/n}}^{\infty} e^{-\frac{x^2}{2}} \cdot e^{-\frac{y}{2}} \cdot y^{\frac{n-2}{2}} dx dy = \\
&= C_n \int_{x \leq z \cdot \sqrt{y/n}, \quad y > 0} e^{-\frac{x^2}{2}} \cdot e^{-\frac{y}{2}} \cdot y^{\frac{n-2}{2}} dx dy = C_n \int_0^{\infty} \left(\int_{-\infty}^{x=z \cdot \sqrt{y/n}} e^{-\frac{x^2}{2}} \cdot e^{-\frac{y}{2}} \cdot y^{\frac{n-2}{2}} dx \right) dy, \\
C_n &= \frac{1}{\sqrt{2\pi} \cdot 2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})} \quad \text{Sei } x := u \sqrt{\frac{y}{n}} \Rightarrow dx = du \sqrt{\frac{y}{n}}, \quad H := 1 + \frac{u^2}{n}, \quad \frac{H y}{2} = \frac{y}{2} + \frac{x^2}{n} \\
\rightsquigarrow F(z) &= \frac{C_n}{\sqrt{n}} \int_0^{\infty} \left(\int_{-\infty}^{u \sqrt{\frac{y}{n}} = z \cdot \sqrt{y/n}} e^{-\frac{(u \sqrt{\frac{y}{n}})^2}{2}} \cdot e^{-\frac{y}{2}} \cdot y^{\frac{n-2}{2}} du \right) \sqrt{y} dy = \frac{C_n}{\sqrt{n}} \int_0^{\infty} \left(\int_{-\infty}^{u=z} e^{-\frac{(x)^2 - y}{2}} \cdot y^{\frac{n-1}{2}} dy \right) du \\
&= \frac{C_n}{\sqrt{n}} \int_0^{\infty} \left(\int_{-\infty}^{u=z} e^{-\frac{H y}{2}} \cdot y^{\frac{n-1}{2}} dy \right) du \quad \text{Sei } H y := 2v, \quad y = \frac{2v}{H}, \quad dy = dv \frac{2}{H} \rightsquigarrow \\
\rightsquigarrow F(z) &= \frac{C_n}{\sqrt{n}} \int_{v=0}^{v=\infty} \left(\int_{u=-\infty}^{u=z} e^{-\frac{2v}{2}} \cdot \left(\frac{2v}{H}\right)^{\frac{n-1}{2}} \cdot \frac{2}{H} dv \right) du = \frac{C_n}{\sqrt{n}} \int_{v=0}^{v=\infty} \left(\frac{1}{H}\right)^{\frac{n-1}{2}} \cdot \frac{2}{H} \left(\int_{u=-\infty}^{u=z} e^{-\frac{2v}{2}} \cdot (2v)^{\frac{n-1}{2}} dv \right) du \\
&= \frac{C_n}{\sqrt{n}} \cdot 2^{\frac{n+1}{2}} \cdot \left(\int_{v=0}^{v=\infty} \left(\frac{1}{H}\right)^{\frac{n+1}{2}} du \right) \cdot \left(\int_{u=-\infty}^{u=z} e^{-\frac{2v}{2}} \cdot v^{\frac{n-1}{2}} dv \right) = \frac{C_n}{\sqrt{n}} \cdot 2^{\frac{n+1}{2}} \cdot \left(\int_{u=-\infty}^{u=z} \left(\frac{1}{H}\right)^{\frac{n+1}{2}} du \right) \cdot \Gamma\left(\frac{n+1}{2}\right) \\
&= \frac{1}{\sqrt{2\pi} \cdot 2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})} \cdot \frac{1}{\sqrt{n}} \cdot 2^{\frac{n+1}{2}} \cdot \left(\int_{u=-\infty}^{u=z} \left(\frac{1}{1 + \frac{u^2}{n}}\right)^{\frac{n+1}{2}} du \right) \cdot \Gamma\left(\frac{n+1}{2}\right) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \cdot \Gamma(\frac{n}{2})} \cdot \left(\int_{u=-\infty}^{u=z} \left(\frac{1}{1 + \frac{u^2}{n}}\right)^{\frac{n+1}{2}} du \right) \\
\rightsquigarrow F(z) &= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \cdot \Gamma(\frac{n}{2})} \cdot \left(\int_{u=-\infty}^{u=z} \left(\frac{1}{1 + \frac{u^2}{n}}\right)^{\frac{n+1}{2}} du \right), \quad f(z) = \frac{dF(z)}{dz} = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \cdot \Gamma(\frac{n}{2})} \cdot \left(\frac{1}{1 + \frac{z^2}{n}}\right)^{\frac{n+1}{2}} \quad \text{□}
\end{aligned}$$

4.13.4 Beweis Tschebyscheffsche Ungleichung

Sei Y eine beliebige Zufallsgrösse, die nicht normalverteilt sein muss. $E(Y) = \mu$ sei hier der Erwartungswert, $Var(Y) = \sigma^2$ sei hier die Varianz und ε sei eine beliebige positive Zahl. Dann gilt:

Formel:

Tschebyscheff

$$P(|Y - \mu| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}$$

Bemerkung:

$$P(|Y - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \Leftrightarrow P(|Y - \mu| < \varepsilon) > 1 - \frac{\sigma^2}{\varepsilon^2}$$

Definition:

Nichtnegative Zufallsvariable X : Nimmt mit Wahrscheinlichkeit 1 nur Werte aus \mathbb{R}_0^+ an.

Wir beweisen erst ein Lemma:

Lemma:**Vor.:** X nichtnegativ , $\delta > 0$ **Beh.:**

$$\frac{E(X)}{\delta} \leq P(X \geq \delta)$$

Beweis:

(Lemma)

1. Diskreter Fall:

$$E(X) = \sum_{k \in \{k\}} x_k p_k \geq \sum_{k \in \{k \mid x_k \geq \delta\} := H} x_k p_k \geq \sum_{k \in H} x_{\min} p_k \geq \sum_{k \in H} \delta p_k \geq \delta \sum_{k \in H} p_k = \delta P(X \geq \delta)$$

2. Stetiger Fall:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \geq \int_0^{\infty} x f(x) dx \geq \int_{\delta}^{\infty} x f(x) dx \geq \int_{\delta}^{\infty} x_{\min} f(x) dx \geq \int_{\delta}^{\infty} \delta f(x) dx \\ &= \delta \int_{\delta}^{\infty} f(x) dx = \delta P(X \geq \delta) \end{aligned}$$

Beweis:

(Satz)

Sei $\delta = \varepsilon^2$, $X = |Y - \mu|^2 \rightsquigarrow$ nichtnegativ

$$\begin{aligned} \sigma^2 &= E(|Y - \mu|^2) = E((Y - E(Y))^2) = E(X) \geq \delta P(X \geq \delta) = \varepsilon^2 P(|Y - E(Y)|^2 \geq \varepsilon^2) \\ &= \varepsilon^2 P(|Y - E(Y)| \geq \varepsilon) \Rightarrow P(|Y - E(Y)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \end{aligned}$$

4.14 Anhang II: Ergänzungen

4.14.1 Quadratsumme contra Betragssumme

1. **Frage:** Für welches c ist die Quadratsumme $\sum_i (x_i - c)^2$ minimal?

$$\begin{aligned} &\rightsquigarrow \frac{d}{dc} \sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (-1) \cdot 2(x_i - c) = -2 \sum_{i=1}^n (x_i - c) = 0 \Rightarrow \sum_{i=1}^n x_i = \sum_{i=1}^n c = n \cdot c \\ &\Rightarrow c = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \bar{x} \Rightarrow c = \bar{x} \rightsquigarrow c = \text{Mittelwert} \end{aligned}$$

2. **Frage:** Für welches c ist die Betragssumme $\sum_i |x_i - c|$ minimal?

$$\begin{aligned} &\rightsquigarrow \frac{d}{dc} \sum_{i=1}^n |x_i - c| = \frac{d}{dc} \sum_{i=1}^n \text{sgn}(x_i - c) \cdot (x_i - c) = \underset{|x_i \neq c, \text{ sgn} = \text{const}}{\sum_{i=1}^n} (-1) \cdot \text{sgn}(x_i - c) \\ &= -(\underbrace{-1 - 1 - \dots - 1}_{x_i < c}) + (\underbrace{+1 + 1 + \dots + 1}_{x_i > c}) = 0 \\ &\Rightarrow -(\underbrace{-1 - 1 - \dots - 1}_{x_i < c}) = (\underbrace{+1 + 1 + \dots + 1}_{x_i > c}) \\ &\rightsquigarrow (\text{Anzahl „Werte } < c“}) = (\text{Anzahl „Werte } > c“) \rightsquigarrow c = \text{Median} \end{aligned}$$

Konsequenz:

$\sum_i (x_i - c)^2$ wird minimal für den Mittelwert $c = \bar{x}$, $\sum_{i=1}^n |x_i - c|$ wird minimal für den Median $c = \tilde{x}$.

4.14.2 Verteilungstreue u.s.w.

Auf Seite 109 haben wir folgenden Satz gesehen:

Satz:

Summe unabhängiger normalverteilter Variablen

Vor.:

Seien X_1, X_2, \dots, X_n unabhängig, normalverteilt mit

Mittelwerte $\mu_1, \mu_2, \dots, \mu_n$
Varianzen $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$

Beh.:

$\sum_{i=1}^n X_i$ normalverteilt

Mittelwert $\mu = \sum_{i=1}^n \mu_i$

Varianz $\sigma^2 = \sum_{i=1}^n \sigma_i^2$

Summen unabhängiger normalverteilter Variablen sind somit wieder normalverteilt. Dieses Verhalten tritt aber auch bei anderen Verteilungsarten zutage:

Satz:

Vor.:

X, Y unabhängig

$X : \in Bi(n_1, p), Y : \in Bi(n_2, p)$

Beh.:

$X + Y : \in Bi(n_1 + n_2, p)$

Zum Beweis in einem Spezialfall:

Sei X = Anzahl des Eintretens des Ereignisses A mit $p(A) = p$ bei n_1 -maligem Wiederholen des zugehörigen Versuches und Y = Anzahl des Eintretens des Ereignisses A mit $p(A) = p$ bei n_2 -maligem Wiederholen des zugehörigen Versuches. Offensichtlich gehört dann $X + Y$ zum selben Versuch mit $p(A) = p$: $X + Y$ = Anzahl des Eintretens des Ereignisses A bei $n_1 + n_2$ -maligem Wiederholen des zugehörigen Versuches. Die Verteilung ist immer eine Binomialverteilung. (Zum Beweis vgl auch Lit. A4)

Da die Binomialverteilung sich bei $n p = \lambda = \text{const.}$ einer Poissonverteilung annähert, lässt sich der Satz auch auf die Poissonverteilung übertragen:

Satz:**Vor.:**

$$X, Y \text{ unabhängig}$$

$$X : \in Po(\lambda_1), \quad Y : \in Bi(\lambda_2)$$

Beh.:

$$X + Y : \in Bi(\lambda_1 + \lambda_2)$$

Für die Beweise der folgenden Sätze sei auf die Literatur verwiesen (falls nicht schon erledigt):

Satz:**Vor.:**

$$X, Y \text{ unabhängig}$$

$$X, Y \chi^2\text{-verteilt}$$

$$X : \in \chi^2_{m_1}, \quad Y : \in \chi^2_{m_2}$$

Beh.:

$$X + Y : \chi^2_{m_1 + m_2}$$

Satz:**Vor.:**

$$X_1, X_2, \dots, X_n \text{ unabhängig}$$

$$X_i \text{ } N(0, 1)\text{-verteilt}$$

Beh.:

$$Z = X_1^2 + X_2^2 + \dots + X_n^2 : \in \chi^2_n$$

Satz:**Vor.:**

$$X_1, X_2, \dots, X_n \text{ unabhängig}$$

$$X_i \text{ } N(\mu, \sigma^2)\text{-verteilt}$$

$$\bar{X} = \frac{1}{n} \sum_{i_1}^n X_i$$

Beh.:

$$\frac{1}{\sigma^2} \sum_{i_1}^n (X_i - \bar{X})^2 : \in \chi^2_{n-1}$$

Bezüglich der t -Verteilung und der χ^2 -Verteilung können wir noch rekapitulierend zusammenfassen oder folgern:

Satz:**Vor.:**

$$X, Y \text{ unabhängig}$$

$$X : \in N(0, 1), \quad Y : \in \chi^2_n$$

$$Z = \sqrt{n} \cdot \frac{X}{\sqrt{Y}}$$

Beh.:

$$Z : \in t_n$$

Satz:**Vor.:** X_1, X_2, \dots, X_n unabhängig $X_i: N(\mu, \sigma^2)$ -verteilt

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad Y = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2, \quad Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Beh.:

$$\bar{X} : \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right), \quad Z : \in N(0, 1), \quad Y : \in \chi^2_{n-1}$$

Satz:**Vor.:** X_1, X_2, \dots, X_n unabhängig $X_i: N(\mu, \sigma^2)$ -verteilt

$$H = \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}}$$

Beh.:

$$H : \in t_{n-1}$$

4.14.3 Zentraler Grenzwertsatz

Auf Seite 93 haben wir den Grenzwertsatz von Moivre/ Laplace kennengelernt. Bei n Zufallsvariablen sieht er wie folgt aus:

Satz:**Vor.:**Sei $H_n : \in Bi(n, p)$, $p \in (0, 1)$, $n \in \mathbb{N}$

$$Y_n := \frac{H_n - E(H_n)}{\sqrt{Var(H_n)}} = \frac{H_n - E(H_n)}{\sqrt{n p (1-p)}}$$

 $\leadsto Y_n$ standardisiert F_{Y_n} Verteilungsfunktion**Beh.:**

$$\lim_{n \rightarrow \infty} F_{Y_n} = \Phi(x) = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R}$$

Bemerkung:

Der Satz hat eine Bedeutung, wenn es bei grossen n um approximative Berechnungen von Wahrscheinlichkeiten im Zusammenhang mit Binomialverteilungen geht. Hier ist dann $\mu = np$, $\sigma = \sqrt{np(1-p)}$. Dann gilt:

Folgerung:

$$P(a \leq X \leq b) = P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right)$$

Φ ist leicht bestimbar (Rechner, Tabellen, ...). Gute Näherungen erreicht man schon, wenn die **Laplace-Bedingung** $np(1-p) > 9$ erfüllt ist.

Wir schreiben $A \in R_i$, wenn das Ereignis A bei der i -ten Wiederholung eines Versuches eintritt. Sonst ist $A \notin R_i$. Wir wählen nun H_n speziell:

$$H_n(A) = \sum_{i=1}^n X_i, \quad X_i = \begin{cases} 1 & A \in R_i \\ 0 & A \notin R_i \end{cases}$$

Man sieht sofort, dass jedes X_i für sich binomialverteilt ist mit den Parametern $n = 1$ und p , d.h. mit dem Erwartungswert $\mu = np = p$ und der Varianz $\sigma^2 = np(1-p) = p(1-p)$. Wie wir schon wissen, ist die Variable $\sum_{i=1}^n$ als Summe binomialverteilter Zufallsvariablen wieder binomialverteilt.

$$\text{Sei } Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sqrt{\sigma}/\sqrt{n}} \sim F_{Y_n} \rightarrow \Phi(n), \quad n \rightarrow \infty$$

Dieses Verhalten ist auch beweisbar, wenn $\langle X_n \rangle$ eine beliebige Folge unabhängiger und identisch verteilter Zufallsgrößen ist mit gemeinsamem Erwartungswert μ und gemeinsamer Varianz $\sigma^2 \in (0, k)$, $k = \text{const.} \in \mathbb{R}^+$.

Satz:

Zentraler Grenzwertsatz

Vor.:

Wie beschrieben

Beh.:

$$\lim_{n \rightarrow \infty} F_{Y_n} = \Phi(x) = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R}$$

Beweise finden sich in der Literatur, z.B. in Bibl. A3.

4.14.4 Lineare Transformationen

Auf Seite 72 haben wir den Satz erkannt:

Satz:

Vor.:

$$g(X) = a h(X) + b u(X)$$

Beh.:

$$E(g(X)) = a E(h(X)) + b E(u(X))$$

Und auf Seite 72 sahen wir:

Satz: $E(a_n X^n + a_{n-1} X^{n-1} + \dots + a_1 X + a_0) = a_n E(X^n) + a_{n-1} E(X^{n-1}) + \dots + a_1 E(X) + a_0 = a_n \mu_n + a_{n-1} \mu_{n-1} + \dots + a_1 \mu_1 + a_0$

\rightsquigarrow **Speziell:** $E(X) = \mu, X^* = c_1 X + c_2 = g(X)$

Korollar:

Vor.:

$$X^* = c_1 X + c_2$$

Beh.:

$$\mu^* = c_1 \mu + c_2$$

Weiter können wir folgern:

$$X^* - \mu^* = (c_1 X + c_2) - (c_1 \mu + c_2) = c_1 (X - \mu), E((X^* - \mu^*)^k) = E(c_1^k (X - \mu)^k) = c_1^k E((X - \mu)^k)$$

\rightsquigarrow **Speziell:** $k = 2 : \sigma_{X^*}^2 = E((X^* - \mu^*)^2) = c_1^2 E((X - \mu)^2) = c_1^2 \sigma^2$

Korollar:

Vor.:

$$X^* = c_1 X + c_2$$

Beh.:

$$\sigma_{X^*}^2 = c_1^2 \sigma^2$$

Kapitel 5

Mathematische Statistik

5.1 Qualitätskontrolle

5.1.1 Allgemeines, SQC

Qualitätskontrolle und speziell **statistische Qualitätskontrolle (SQC, statistical quality control)** umfasst Verfahren, die zur Qualitätssicherung, Qualitätsprüfung und Qualitätslenkung benutzt werden. Heute sind die Rahmenbedingungen oft Gegenstand von Normierungen (z.B. DIN 55350). Generell kann man zwischen qualitativen Kontrollen (z.B. gut, schlecht) und quantitativen Kontrollen (Messung von Variablen) unterscheiden.

Entsprechend der Anwendung unterscheidet man **zwei Hauptgebiete**:

1. Statistische Prozesskontrolle, Fertigungsüberwachung oder Qualitätsregulierung (**SPC, statistical process control**):

Hier geht es um die Überwachung des Fertigungsprozesses bei der Herstellung von Massenprodukten, um während der Produktion Qualitätsstörungen entdecken und dann unmittelbar eingreifen und steuern zu können.

2. Annahmekontrolle oder Annahmestichprobenprüfung (AC, acceptance sampling):

Hier geht es um die Eingangs-, Zwischen- und Endkontrolle eines Betriebes zur Überprüfung der Erzeugnisse ohne direkten Einfluss auf die Produktion. Damit wird aber der Umfang von produziertem Ausschuss beziffert. Die Annahmekontrolle dient auch zur Rückweisung von eingehender Ware. Sie beeinflusst daher indirekt die Produktion.

5.2 SQC1: Prozesskontrolle

5.2.1 Problemstellung

Bsp.:

Es sollen Sollwerte innerhalb gewisser Toleranzen eingehalten werden. Z.B. bei der Serienfertigung von Zahnrädern muss der Durchmesser der zentralen Bohrung einen bestimmten Sollwert μ_0 aufweisen.

Aus Erfahrung weiss man aber, dass auch bei aller Sorgfalt stets Abweichungen auftreten. Gründe sind menschliches Versagen, Abnutzungerscheinungen bei den Maschinen und Apparaten u.s.w. sowie Materialprobleme. Daher gilt es, die Produktion permanent mittels SPC zu überwachen und gegebenenfalls sofort einzutreten. Ziel ist es, den Ausschuss auf ein Minimum zu beschränken.

5.2.2 Beispiel Mittelwert

In der Praxis gibt es verschiedenste Problemstellungen, die wir hier nicht alle behandeln können. Wir greifen daher beispielhaft ein Problem heraus. Bei einer momentan mit einem Automaten produzierten Serie eines Artikels soll der Mittelwert μ_0 innerhalb eines nicht kritischen Toleranzbereiches konstant eingehalten werden.

Methodisches Vorgehen:

1. Annahme:

- (a) Aus Erfahrung wissen wir und nehmen daher auch jetzt an: $\bar{X} = \mu$ genügt einer Normalverteilung mit bekannter Standardabweichung σ . Diese Annahme ist jedoch nur in einem gewissen Bereich vernünftig, da der Definitionsbereich der Normalverteilung $(-\infty, \infty) = \mathbb{R}$ ist, der Bereich, in dem Messwerte x_i und damit Mittelwerte \bar{x} überhaupt feststellbar sind, ist $(0, M_{max})$. M_{max} hängt von den praktischen Möglichkeiten ab. Negative Messwerte existieren nicht. μ_0 ist vorgeschrrieben. Für σ können wir vorläufig einen Schätzwert benutzen, den wir aus schon akzeptierten Stichproben gewinnen können.
- (b) Die folgende Forderung oder Hypothese H_0 sei erfüllt: Der Sollwert μ_0 wird während der ganzen Produktion eingehalten, d.h. $H_0 : \mu = \mu_0$. Die Alternative dazu wäre $H_1 : \mu \neq \mu_0$.

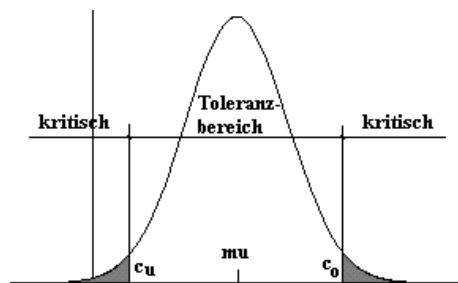
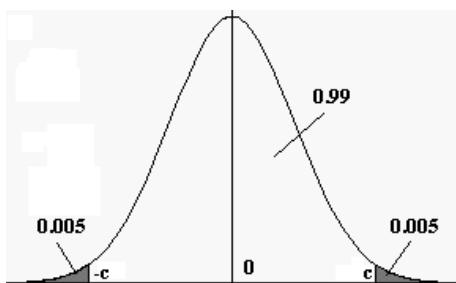
Definition:

H_0 heisst **Nullhypothese**, H_1 heisst **Alternativhypothese**.

- 2. In regelmässigen Zeitabständen sollen der Produktion Stichproben von gleichbleibendem Umfang n entnommen werden. Daraus wird der Stichprobenmittelwert \bar{x} bestimmt \rightsquigarrow **Messexperiment**.
- 3. Weiter hat uns die Erfahrung gelehrt, dass es vernünftig ist die Produktion zu korrigieren, wenn die Wahrscheinlichkeit, dass die Alternativhypothese $H_1 : \mu \neq \mu_0$ eintrifft, $\alpha = 0.01$ oder 1% übersteigt. (Wenn α erreicht ist, so bedeutet das etwas (fz. signifie!)! Man hat das Signifikanzniveau erreicht.) Dies ist so in Ordnung, denn bisher hat die Produktion sich etwa nach dem Gesetz der Normalverteilung verhalten, wobei 1% Ausschuss akzeptiert worden ist. Wir betrachten dazu die Standardnormalverteilung $N(0, 1)$ (links) im Vergleich zu $N(\mu, \sigma^2)$ (rechts):

Definition:

α heisst **Signifikanzniveau**. Die Zahl bezeichnet eine Grenze, an der meist nach Erfahrung etwas Bedeutendes geschieht.



Der Übergang von der normalverteilten Zufallsvariablen \bar{X} zur Standardisierten U ist durch die folgende Transformation gegeben:

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

In der standardisierten Form lautet die Bedingung mit α :

$$P(-c \leq U \leq c) \leq 1 - \alpha = 1 - 0.01 = 0.99, \quad \pm c = \sqrt{n} \cdot \frac{c_{o,u}^* - \mu_0}{\sigma}, \quad c_{o,u}^* = \mu + \frac{\pm c \cdot \sigma}{\sqrt{n}}$$

4. Es gilt: $P(-c \leq U \leq c) = \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{x^2}{2}} dx \leq 0.99 \rightsquigarrow c = ?$

Man kann auch rechnen:

$$P(-c \leq U \leq c) = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1 = 2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^c e^{-\frac{x^2}{2}} dx$$

Mit Mathematica (oder Tabelle) findet man:

```
FindRoot[1/(Sqrt[2 Pi]) Integrate[E^(-x^2/2), {x, -c, c}] == 0.99, {c, 1}]
```

$$\begin{aligned} \rightsquigarrow 2.57582 \leq c \leq 2.57583, \quad c \approx 2.5758 \Rightarrow c_o^* \approx \mu + \frac{2.5758 \cdot \sigma}{\sqrt{n}}, \\ [c_u^*, c_o^*] \approx [\mu - \frac{2.5758 \cdot \sigma}{\sqrt{n}}, \mu + \frac{2.5758 \cdot \sigma}{\sqrt{n}}] \end{aligned}$$

5. Sobald \bar{x} aus dem Bereich $[c_u^*, c_o^*]$ herausläuft, muss also eingeschritten werden! Die Abweichung von \bar{x} wird dann nicht mehr als zufallsbedingt eingestuft und toleriert, sondern als signifikant und nicht tolerierbar.
6. Oft wählt man zu den **kritischen Grenzen** $c_{\alpha,u,o}^* = c_{0.01,u,o}^*$ auch **Warngrenzen** $d_{\beta,u,o}^* = d_{0.05,u,o}^*$. Dabei macht man für d zu $\beta = 0.05$ die entsprechende Rechnung wie für c zu $\alpha = 0.01$.
- $$\rightsquigarrow [d_u^*, d_o^*] \approx [\mu - \frac{1.95996 \cdot \sigma}{\sqrt{n}}, \mu + \frac{1.95996 \cdot \sigma}{\sqrt{n}}]$$

7. Zahlenbeispiel:

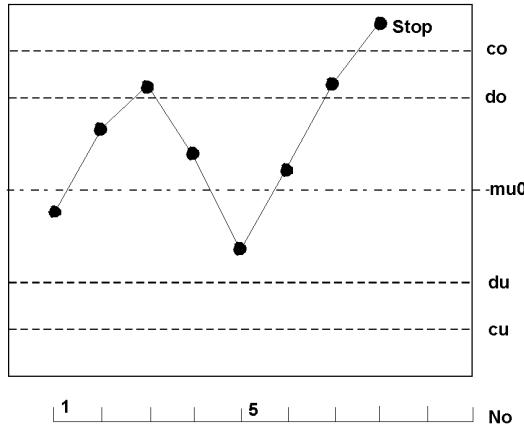
Sei $\mu = 120, \sigma = 0.01, n = \dots$

| $c_{0.01,u,o}$ | $n = 1$ | $n = 2$ | $n = 10$ | $n = 20$ | $n = 100$ |
|----------------|---------|---------|----------|----------|-----------|
| c_u | 119.974 | 119.982 | 119.992 | 119.994 | 119.997 |
| c_0 | 120.026 | 120.018 | 120.008 | 120.006 | 120.003 |
| $d_{0.05,u,o}$ | $n = 1$ | $n = 2$ | $n = 10$ | $n = 20$ | $n = 100$ |
| d_u | 119.980 | 119.986 | 119.994 | 119.996 | 119.998 |
| d_0 | 120.020 | 120.014 | 120.006 | 120.004 | 120.002 |

5.2.3 Überwachungstechnik mittels Kontrollkarten

Es ist oft üblich, die Resultate der Stichprobenerhebungen graphisch übersichtlich und anschaulich darzustellen. Dazu verwendet man in der Industrie **Kontrollkarten** diversester Art. Exemplarisch geben wir die Kontrollkarte zu einer Serie von Stichproben zu obigem Beispiel.

$$\mu = 120, \sigma = 0.01, n = 10, \alpha = 0.01, \beta = 0.05, c_{\alpha,u,o}^*, d_{\beta,u,o}^*$$



5.3 SQC2: Annahmekontrolle

Spez. franz..

Stichproben und Randomisierung

Problem:

Als Kunden erhalten wir eine grössere Lieferung einer Ware. Man denke z. B. an Zahnräder. Da wir auf die Einhaltung der vertraglich abgesicherten Toleranzen angewiesen sind um die Teile problemlos einbauen und die Funktion garantieren zu können, müssen wir die Einhaltung der Toleranzen der Teile prüfen. Wir führen daher eine **Annahmekontrolle** durch, um auf Grund der Verträge zu entscheiden, ob die Lieferung angenommen werden kann oder nicht. Falls es sich z.B. um die Prüfung der chemischen Zusammensetzung handelt (z.B. Prüfung der Bleifreiheit der Glasuren von Keramik–Essgeschirr), ist oft zur Prüfung eine Zerstörung des Stückes notwendig. Weil aber auch sonst eine Prüfung aller Teile zu hohe Kosten verursacht, beschränken wir uns auf Stichproben und mathematische Methoden, um daraus Rückschlüsse auf die Grundgesamtheit zu ziehen. Stichproben lassen sich exakter kontrollieren als grosse Stückzahlen, wo die Routine auf Kosten der Sorgfalt geht. Überlegt man sich die Konsequenzen eines falschen Entscheides betreffend Annahme und Weiterverarbeitung einer Lieferung, so merkt man, dass davon das überleben einer Firma abhängen kann.

Die Modalitäten einer **Annahmekontrolle** des Kunden müssen entsprechend den Abmachungen in einem Prüfplan fixiert sein. Die genau gleiche Kontrolle kann auch der Lieferant für sich durchführen. Hier heisst sie dann **Ausgangskontrolle, Endkontrolle** u.s.w..

Bei der Erhebung der Stichprobe spielt die Zufälligkeit der Stückauswahl eine wesentliche Rolle, um systematische Verfälschungen und Bevorzugungen auszuschliessen. Es ist daher sehr gefährlich, gedankenlos oder blind irgendwelche ersten Stücke auszuwählen. Die Auswahl muss nach einem **Randomisierungsverfahren** geschehen. Z.B. kann man den Stücken Nummern zuordnen und dann diejenigen Nummern wählen, die in einer Randomisierungsliste als Zufallszahlen stehen oder von einem

Pseudo-Zufallsgenerator generiert worden sind. Zum Problem der Generierung von Zufallszahlen muss des engen Rahmens wegen auf die Literatur verwiesen werden. (Möglichkeit: Nehme Ziffern aus der Dezimalbruchentwicklung von π .)

5.3.1 Beispiel hypergeometrische Verteilung, Urnenmodell

Oben haben wir gesehen, dass man generell unterscheiden kann zwischen qualitativen Kontrollen (z.B. in Ordnung, nicht in Ordnung) und quantitativen Kontrollen (Messung von Variablen).

Wir betrachten nun hier ein Beispiel aus der Klasse der qualitativen Annahmekontrollen. Gegeben sei eine Warenlieferung mit N Teilen. Davon seien M Teile defekt und daher $N - M$ Teile in Ordnung. Statt eine Warenlieferung zu betrachten, können wir auch an N Kugeln in einer Urne denken: M schwarze und $N - M$ weisse.

Aus der Lieferung resp. aus der Urne nehmen wir eine Zufallsstichprobe von n Stück ohne zurücklegen, $n \leq N$. Von früher wissen wir, dass die möglichen Resultate dieses Experiments ein Modell für die hypergeometrische Verteilung bilden. Sei $X = m$ die Anzahl der gezogenen defekten Stücke — oder beim Urnenmodell die Anzahl gezogener schwarzer Kugeln. Sei $m \in \{1, 2, \dots, n\}$. Dann wissen wir von der hypergeometrische Verteilung:

$$p_d = \frac{M}{N}, \quad M = p_d \cdot N, \quad p_m = P(X = m) = \frac{\binom{M}{m} \cdot \binom{N - M}{n - m}}{\binom{N}{n}}$$

Bsp.: $N = 120, \quad M = 4, \quad n = 3$
 $\rightsquigarrow p_0 = P(X = 0) = \frac{\binom{4}{0} \cdot \binom{120 - 4}{3 - 0}}{\binom{120}{3}} \approx 0.902507,$
 $p_1 = P(X = 1) \approx 0.0950007, \quad p_2 = P(X = 2) \approx 0.00247828, \quad p_3 = P(X = 3) \approx 0.000014243$
 $\Rightarrow P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$

5.3.2 Annahmevertrag und Prüfplan

Produzent und Konsument resp. Lieferant und Kunde müssen sich nun über die Annahmebedingung der Lieferung einig sein. Die Annahmebedingung muss in einem Vertrag (Lieferungsvertrag, Annahmevertrag) geregelt sein. Die Beschreibung der Annahmebedingungen mündet in einen Prüfungsplan (**Stichprobenprüfplan**), in dem die Details der Annahmeprüfung definiert sind.

Als Beispiel, das zur oben beschriebenen Situation passt, studieren wir einen einfachen Stichprobenprüfplan (**SSP, simple sampling plan**). Dann definieren wir:

Vor.:

Sei X die Anzahl defekter Teile resp. schwarzer Kugeln in der gezogenen Stichprobe ($0 \leq X = x \leq n$).

Dann definieren wir:

Definition:

Ein **einfacher Stichprobenprüfplan** ist gegeben durch:

1. Stichprobenumfang n
2. Eine Annahmezahl (**kritische Zahl**) $c \in \mathbb{R}_0^+$
3. Eine Entscheidungsregel:

| |
|-------------------------------------|
| $X \leq c \rightsquigarrow$ Annahme |
| $X > c \rightsquigarrow$ Ablehnung |

Problem:

Allgemein kennt man bei einer Lieferung $p_d = \frac{M}{N} \in [0, 1]$ nicht. p_d ist daher ein unbekannter Wert einer Zufallsvariablen Y . Es ist also notwendig, die Annahmewahrscheinlichkeit der Lieferung in Funktion von p_d in Modellen zu untersuchen, um die theoretisch gewonnenen Erkenntnisse praktisch dann kostensenkend verwenden zu können.

Im oben besprochenen Beispiel (hypergeometrische Verteilung, Urnenmodell) ist $X = m$ und $M = p_d \cdot N$. Daher gilt für $p_m = P(X = m)$:

$$p_m = P(X = m) = \frac{\binom{p_d \cdot N}{m} \cdot \binom{N - p_d \cdot N}{n - m}}{\binom{N}{n}} = \frac{\binom{p_d \cdot N}{m} \cdot \binom{N(1 - p_d)}{n - m}}{\binom{N}{n}}$$

$$\Rightarrow P(X = m \leq c) = \sum_{i, m_i \leq c} P(X = m_i) = \sum_{m=0}^{m \leq c} P(X = m) = \sum_{m=0}^{m \leq c} \frac{\binom{p_d \cdot N}{m} \cdot \binom{N(1 - p_d)}{n - m}}{\binom{N}{n}}$$

$P(X = m \leq c)$ ist die Wahrscheinlichkeit, maximal c defekte Stücke resp. schwarze Kugeln in der Stichprobe zu haben. Dabei ist c fix, $p_d = \frac{M}{N}$ jedoch eine unbekannte Realisation von Y . $P(X = m \leq c)$ ist somit eine Funktion von Y mit den bekannten Parametern N, n, c . $P(X = m \leq c)$ ist die **Annahmewahrscheinlichkeit** $L_{N,n,c}$ in Abhängigkeit von Y :

$$L_{N,n,c}(Y) := P(X = m \leq c) = \sum_{m=0}^{m \leq c} \frac{\binom{Y \cdot N}{m} \cdot \binom{N(1 - Y)}{n - m}}{\binom{N}{n}}, \quad Y \in [0, 1]$$

Generell gehört zu einem einfachen Versuchsplan eine Funktion $L_c(Y) := P(X \leq c)$, $Y = p_d \in [0, 1]$.

Definition:

$$L_c(Y) := P(X \leq c), \quad Y = p_d \in [0, 1]$$

heisst **Annahmewahrscheinlichkeitsfunktion** (der Lieferung oder des Loses) durch den Konsumenten

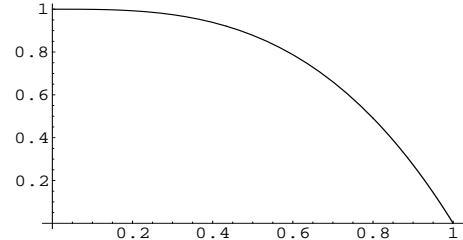
Der Graph von L_c heisst **Effizienzkurve** oder **Operationscharakteristik** des einfachen Versuchsplanes

Bemerkung:

\rightsquigarrow Engl. operating characteristic, **OC**-curve

In obigem Beispiel sei z.B. $c = 2$:

$$\begin{aligned}
L_{120,3,c=2}(Y) &:= P(X = m \leq 2) = \sum_{m=0}^{m \leq 2} \frac{\binom{Y \cdot 120}{m} \cdot \binom{120(1-Y)}{3-m}}{\binom{120}{3}} \\
&= \frac{\binom{Y \cdot 120}{0} \cdot \binom{120(1-Y)}{3}}{\binom{120}{3}} + \frac{\binom{Y \cdot 120}{1} \cdot \binom{120(1-Y)}{2}}{\binom{120}{3}} + \frac{\binom{Y \cdot 120}{2} \cdot \binom{120(1-Y)}{1}}{\binom{120}{3}}
\end{aligned}$$



Speziell:

$$L_{120,3,c=2}(0.9) \approx 0.273052$$

Eigenschaften von L_c :

$L_c(0) = P(X \leq c)|_{M=0}$ bedeutet, dass kein defektes Stück da ist und daher die Annahme sicher ist $\rightsquigarrow Y = p_d = 1$

Wenn die Anzahl der defekten Stücke steigt, so wird auch die Chance, dass bei einer Stichprobe defekte Stücke gefunden werden, grösser. Die Annahmewahrscheinlichkeit wird damit also kleiner. Somit muss $L_c(Y)$ mit Y monoton fallend sein.

$L_c(1) = P(X \leq c)|_{M=N}$ bedeutet, dass nur defekte Stücke da sind und daher die Annahme sicher verweigert wird $\rightsquigarrow Y = p_d = 0$

Eigenschaften: $L_c(0) = 1, L_c(1) = 0, D_L = [0, 1,]$
 $L_c(Y)$ monoton fallend

5.3.3 Das Beispiel Binomialmodell, Urnenmodell

Oben hatten wir eine Urne und die Parameter M, N, n, m . Das Experiment ist ohne zurücklegen ausgeführt worden. Grundlage war eine hypergeometrische Verteilung. Nun wollen wir das Experiment mit zurücklegen ausführen. Das führt uns auf eine Binomialverteilung

Sei $H = \{1, 2, \dots, M, M+1, \dots, N\}$ Wir können davon ausgehen, dass wir ein geordnetes n -Tupel aus der geordneten Menge H auswählen. Dann finden wir:

Total Anzahl Auswahlmöglichkeiten = N^n

Total Anzahl Möglichkeiten nur defekte Stücke zu wählen = M^m

Total Anzahl Möglichkeiten nur gute Stücke zu wählen = $(N - M)^{n-m}$

Anzahl Möglichkeiten in der Stichprobe Plätze an die defekten Stücke zu verteilen: = $\binom{n}{m}$

Die Anzahl Möglichkeiten, m defekte und zugleich $n - m$ gute Stücke zu wählen und dann noch Plätze auf die defekten zu verteilen, d.h. die defekten und die guten zu vermischen, ist gleich der Anzahl Stichproben mit m defekten Stücken bei zurücklegen. Dividiert man durch die Gesamtzahl N^n der Möglichkeiten, so erhält man die Laplace-Wahrscheinlichkeit $P(X = m)$:

$$\frac{M}{N} := p \Rightarrow P(X = m) = \binom{n}{m} \cdot \frac{M^m \cdot (N - M)^{n-m}}{N^n} = \binom{n}{m} \cdot \frac{M^m}{N^m} \cdot \frac{(N - M)^{n-m}}{N^{n-m}} = \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m}$$

Bekanntlich haben wir hier eine Binomialverteilung.

Folgender Zusammenhang rechtfertigt den Namen:

$$\sum_{m=1}^n p_m = \sum_{m=1}^n \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m} = (1 + (1-p))^n = 1^n = 1$$

Vor.:

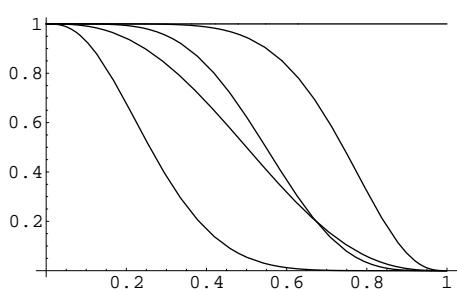
$$\text{Sei } n \ll N, m \in \{0, 1, \dots, n\}, p = \frac{M}{N}, N \rightarrow \infty$$

→ Dann kann man die hypergeometrische Verteilung durch die Binomialverteilung approximieren (praktisch z.B. für $10n \leq N$):

$$\lim_{N \rightarrow \infty} \frac{\binom{M}{m} \cdot \binom{N - M}{n - m}}{\binom{N}{n}} = \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m}$$

In diesem Fall gilt für die Annahmewahrscheinlichkeitsfunktion:

$$L_{n,c}(Y) = P(X \leq c) = \sum_{m=0}^{m \leq c} \binom{n}{m} \cdot Y^m \cdot (1-Y)^{n-m}$$



Links Kurven für:

$$(n, c) = (5, 2), (10, 2), (10, 5), (10, 7), (5, 5)$$

Beobachtung: Je grösser c , desto weiter rückt die Kurve nach rechts. Je grösser n , desto weiter rückt die Kurve nach unten.

5.3.4 Produzentenrisiko, Konsumentenrisiko

$L_{n,c}(Y) = L_{n,c}(p_d) = P(X \leq c)_{|Y=p_d}$ gibt bei den Parametern n und c die Wahrscheinlichkeit $P(X \leq c)$ in Abhängigkeit von $Y = p_d$, d.h. bei unbekanntem p_d . Da p_d unbekannt ist und trotzdem ein Entscheid gefällt werden muss ($X \leq c \rightsquigarrow$ Annahme, $X > c \rightsquigarrow$ Ablehnung), so kann es vorkommen, dass der Entscheid bei falsch angenommenem p_d zu ungünstigen des Produzenten oder auch zu ungünstigen des Konsumenten ausfällt.

Wird z.B. die **Ablehnungsgrenze** oder Schätzung $Y = p_\beta$ des effektiven Ausschussanteils p_d vom Konsumenten zu tief angesetzt ($Y = p_\beta < p_d$), so kann es vorkommen, dass die **Lieferung nach Prüfplan akzeptiert** wird, obwohl sie einen nicht **tolerierbaren Ausschussanteil** aufweist. Denn die Annahmewahrscheinlichkeit könnte dann zu gross werden.

Definition:

Die Annahme eines Loses durch den Konsumenten trotz zuviel Ausschuss heisst **Fehler 2. Art**. Die dazugehörige Annahmewahrscheinlichkeit $\beta = L_{n,c}(p_\beta) = P(X \leq c)$ heisst **Konsumentenrisiko**. p_β heisst **Ablehnungsgrenze**.

(Hit! Konsumentenrisiko wird von der Maschine mit "Le risque de canard de consommation" übersetzt.)

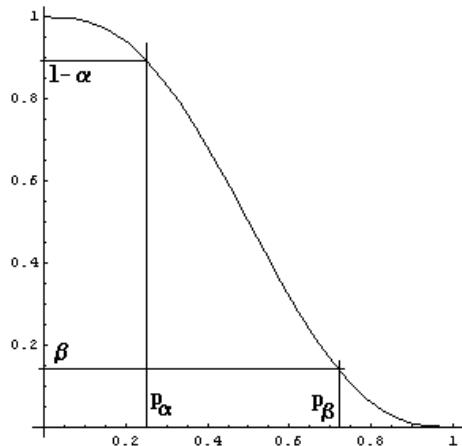
Entsprechend definieren wir:

Definition:

Die Ablehnung eines Loses obwohl es in Ordnung ist und der Ausschuss tolerierbar ist, heisst Fehler 1. Art. Die zugehörige Wahrscheinlichkeit $\alpha = P(X > c) = 1 - P(X \leq c) = 1 - L_{n,c}(p_\alpha)$ heisst **Produzentenrisiko**. p_α heisst **Annahmegrenze**.

Da α und β Fehlerwahrscheinlichkeiten sind, muss der Prüfplan so eingerichtet werden (Vertrag!), dass die beiden Größen möglichst klein werden. Aus dem Diagramm ist ersichtlich, dass das erreicht werden kann, wenn die Kurve möglichst steil gewählt wird (Modellierung!) oder wenn p_α möglichst rechts und p_β möglichst links liegt.

Man bedenke, dass grosse n mehr kosten!



Idee:

Sei $emp = \text{empirisch}$

Wähle

$$\left(\frac{M}{N}\right)_{emp} \leq p_\alpha \leq \left(\frac{M}{N}\right)_{emp} + \varepsilon$$

Wird hier M während der Produktion grösser, so sollte man das merken!

Bemerkung:

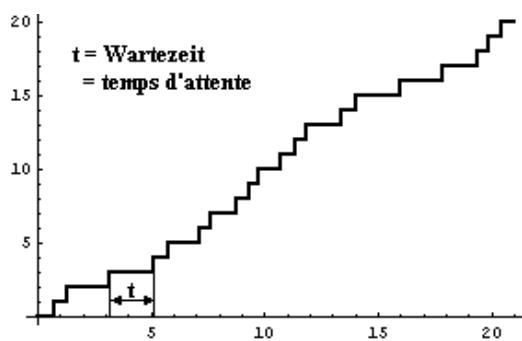
Stichprobenprüfung: $\leadsto p_\alpha$ heisst auch **annehmbare Qualitätslage, Gutlage** oder **AQL** (acceptable quality level). p_β heisst auch **rückweisende Qualitätsgrenzlage, Schlechtlage** oder **LQ, LQL, RQL** (limiting quality level, rejectable quality level).

Normierungen:

DIN ISO 2859, US Military Standard 105 D, DIN 40080, Philips-Standard, DIN ISO 3951, ISO 8423 etc.
 $1\% \leq \alpha\% \leq 12\%$, norm. $\beta\% \leq 10\%$, except. $\beta\% \leq 13\% \dots$ (Lit. Storm, Bibl. A12.)

5.4 Poisson–Prozess, Warteschlange

5.4.1 Poisson–Prozess



Sei $X(t) =$ Anzahl Zufallsereignisse von der Zeit 0 bis t
 \rightsquigarrow Die Zufallsvariable (stochastische Variable) ist zeitabhängig (\rightsquigarrow Prozess) sowie diskret ($X(t) \in \mathbb{N}$).

Definition: Erfüllt $X(t)$ obige Bedingungen, so reden wir von einem **stochastischen Prozess**.

Da $X(t)$ von Natur aus nicht abnehmen kann, gilt:

Lemma: $X(t)$ ist monoton wachsend.

Vor.:

Zur Vereinfachung gehen wir hier von der Situation mit der **Startbedingung** $X(0) = 0$ aus.

Sei $\Delta X(t_0, t) := X(t_0 + t) - X(t_0)$

\rightsquigarrow Der Zuwachs $\Delta X(t_0, t)$ von X im Intervall $[t_0, t_0 + t]$ hängt von t_0 und t ab.

Wir betrachten hier einen einfachen Fall:

Definition: Der Prozess $X(t)$ heisst **homogen**, falls $\Delta X(t_0, t)$ unabhängig ist von t_0 .

Dann gilt:

$$\forall_{t_i} : \Delta X(t_0, t) = \Delta X(t_1, t) \Rightarrow \Delta X(t_0, t) = \Delta X(0, t) = X(0 + t) - X(0) = X(t) - 0 = X(t)$$

Lemma:

Vor.:

$X(t)$ homogen

$$\Delta X(t_0, t) := X(t_0 + t) - X(t_0)$$

Beh.:

$$\Delta X(t_0, t) = \Delta X(0, t) = X(t)$$

Sei $p_m(t) := P(X(t) = m)$

$\rightsquigarrow p_m(t)$ ist die Wahrscheinlichkeit, dass in $[0, t]$ genau m Ereignisse liegen.

Definition:

Zwei Ereignisse heißen **statistisch unabhängig**, wenn das Auftreten irgend eines der beiden Ereignisse keinen Einfluss auf das Auftreten des andern Ereignis hat.

Wir betrachten folgende Situation:

$$t_0 < t_1 < t_2 < \dots < t_n, \quad \Delta X_r := X(t_r) - X(t_{r-1}), \quad r \in \{1, 2, 3, \dots, n\}$$

Die Intervalle $I_r = [t_r, t_{r-1}]$ haben nichts gemeinsam ausser den Grenzen, wenn es sich um Nachbarintervalle handelt. Daher wollen wir hier voraussetzen, dass die ΔX_r statistisch unabhängig sind.

$S = \sum_{m=0}^{\infty} p_m(\Delta t) =$ Wahrscheinlichkeit, dass in Δt 0 oder 1 oder 2 oder \dots oder unendlich viele Ereignisse auftreten. Jede Zahl von Ereignissen ist daher möglich $\rightsquigarrow S = 1$. $\rightsquigarrow \rightsquigarrow$

Lemma: $1 - p_0(\Delta t) - p_1(\Delta t) = \sum_{m=2}^{\infty} p_m(\Delta t) := S_2$

S_2 ist die Wahrscheinlichkeit, dass in Δt mehr als 2 Ereignisse auftreten. In der Praxis ist es oft vernünftig anzunehmen, dass für kleine Δt $S_2 \approx 0$ gilt. Das wollen wir hier auch tun:

Vor.:

$$S_2 = 1 - p_0(\Delta t) - p_1(\Delta t) = \sum_{m=2}^{\infty} p_m(\Delta t) \approx 0$$

Konsequenz: $\rightsquigarrow p_0(\Delta t) \approx 1 - p_1(\Delta t)$

Weiter ist es in der Praxis oft vernünftig anzunehmen, dass die Wahrscheinlichkeit für ein Ereignis $p_1(\Delta t)$ und für kleine Δt proportional zur Zeit ist. Auch das wollen wir hier tun:

Vor.:

$$p_1(\Delta t) \approx a \cdot \Delta t, \quad \Delta t < \text{const.}, \quad a > 0$$

Satz:**Vor.:**

1. $X(0) = 0$
2. $X(t)$ homogen
3. ΔX_r ($r = 1, 2, \dots$) statist. unabh.: $\in A$
4. $0 < \Delta t \ll \text{const.} \Rightarrow p_0(\Delta t) \approx 1 - p_1(\Delta t)$
5. $0 < \Delta t \ll \text{const.} \Rightarrow p_1(\Delta t) \approx a \cdot \Delta t, a > 0$
6. $S_2 = \sum_{m=2}^{\infty} p_m(\Delta t) = \varphi(\Delta t) \cdot (\Delta t)^2, |\varphi(\Delta t)| < \text{const.}$

Beh.: $X(t)$ ist poissonverteilt mit $\lambda = a \cdot t$:

$$p_m(t) = P(X(t) = m) = \frac{(at)^m}{m!} e^{-at} = \frac{\lambda^m}{m!} e^{-\lambda}$$

Bemerkung:

Modell für: $X(t)$ = Anzahl Kunden, die an einem Bedienungsschalter eintreffen oder $X(t)$ = Anzahl Ersetzungen von Bauteilen mit identischer Lebenserwartung und gleicher Betriebszeit.

Beweis:

1. $X(t)$ ist die monoton wachsende absolute Häufigkeit ≤ 0

$$\begin{aligned} &\rightsquigarrow X(t + \Delta t) = X(t) + X(\Delta t) \\ &\rightsquigarrow p_0(t + \Delta t) = P(X(t + \Delta t) = 0) = P(X(t) + X(\Delta t) = 0) = P(X(t) = 0 \wedge X(\Delta t) = 0) \\ &= \underset{| \in A}{=} P(X(t) = 0) \cdot P(X(\Delta t) = 0) = p_0(t) \cdot p_0(\Delta t) \approx p_0(t) \cdot (1 - p_1(\Delta t)) \approx p_0(t) \cdot (1 - a \cdot \Delta t) \\ &= p_0(t) - p_0(t) \cdot a \cdot \Delta t \Rightarrow \frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = \frac{p_0(t) - p_0(t) \cdot a \cdot \Delta t - p_0(t)}{\Delta t} = -p_0(t) \cdot a \\ &\Rightarrow \frac{dp_0(t)}{dt} = -p_0(t) \cdot a, \quad p_0(0) = P(\underbrace{X(0) = 0}_{O.k.}) = 0 \\ &\Rightarrow \text{AWP} \rightsquigarrow \text{bekannte Lösung: } p_0(t) = e^{-at} \end{aligned}$$
2. $p_m(t + \Delta t) = P(X(t + \Delta t) = m) = P(X(t) + X(\Delta t) = m) =$

$$\begin{aligned} &P((X(t) = m \wedge X(\Delta t) = 0) \vee (X(t) = m-1 \wedge X(\Delta t) = 1) \vee \dots \vee (X(t) = 0 \wedge X(\Delta t) = m)) \\ &= \underset{| \in A}{=} P(X(t) = m) \cdot P(X(\Delta t) = 0) + P(X(t) = m-1) \cdot P(X(\Delta t) = 1) + \dots + P(X(t) = 0) \cdot P(X(\Delta t) = m) \\ &= p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) + p_{m-2}(t) \cdot p_2(\Delta t) + \dots + p_0(t) \cdot p_m(\Delta t) \\ &\leq p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) + p_{m-2}(t) \cdot 1 + \dots + p_0(t) \cdot 1 \\ &= p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) + \sum_{k=1}^m p_k(t) \approx p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) + \varphi(\Delta t) \cdot (\Delta t)^2, \\ &\Delta t \rightarrow 0 \Rightarrow \varphi(\Delta t) \cdot (\Delta t)^2 \rightarrow 0 \\ &\rightsquigarrow p_m(t + \Delta t) \approx p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) = p_m(t) \cdot (1 - p_1(\Delta t)) + p_{m-1}(t) \cdot p_1(\Delta t) \\ &= p_m(t) \cdot (1 - a \cdot \Delta t) + p_{m-1}(t) \cdot a \cdot \Delta t = p_m(t) - p_m(t) \cdot a \cdot \Delta t + p_{m-1}(t) \cdot a \cdot \Delta t \Rightarrow \\ &\frac{p_m(t + \Delta t) - p_m(t)}{\Delta t} = \frac{p_m(t) - p_m(t) \cdot a \cdot \Delta t + p_{m-1}(t) \cdot a \cdot \Delta t - p_m(t)}{\Delta t} = -p_m(t) \cdot a + p_{m-1}(t) \cdot a \\ &\Rightarrow \frac{dp_m(t)}{dt} = -a \cdot p_m(t) + a \cdot p_{m-1}(t), \quad X(0) = 0 \rightsquigarrow P(X(0) = m > 0) = p_{m>0}(0) = 0 \\ &\Rightarrow \text{AWP} \rightsquigarrow \end{aligned}$$

- (a) $m = 1 \rightsquigarrow \frac{dp_1(t)}{dt} = -a \cdot p_1(t) + a \cdot p_0(t) = -a \cdot p_1(t) + a \cdot e^{-a t} = , \quad p_1(0) = 0$
 $\rightsquigarrow \mathbf{AWP} \rightsquigarrow$ bekannte Lösung: $p_1(t) = a \cdot t \cdot e^{-a t}$
- (b) $m = 2 \rightsquigarrow \frac{dp_2(t)}{dt} = -a \cdot p_2(t) + a \cdot p_1(t) = -a \cdot p_2(t) + a \cdot a \cdot t \cdot e^{-a t} = , \quad p_2(0) = 0$
 $\rightsquigarrow \mathbf{AWP} \rightsquigarrow$ bekannte Lösung: $p_2(t) = \frac{(a \cdot t)^2}{2} \cdot e^{-a t}$
- (c) Induktion $\rightsquigarrow p_m(t) = \frac{(a \cdot t)^m}{m!} \cdot e^{-a t}$

Konsequenz:

Die Wahrscheinlichkeit, dass im Zeitintervall $[t_0, t_0 + \Delta t]$ kein Ereignis eintritt, ist unabhängig von t_0 und somit $= p_0(\Delta t) = e^{-a \Delta t}$.

Sei T = Wartezeit zwischen zwei Ereignissen, z.B. zwischen zwei Ausfällen von Bauteilen oder Maschinen.

$\rightsquigarrow R(t) := P(T > t) = p_0(t) = e^{-a t}$ = Wahrscheinlichkeit, dass wegen $T > t$ bis t kein Bauteil ausfällt

$\rightsquigarrow F(t) := 1 - R(t)$ ist die **Überlebenswahrscheinlichkeit**

$\rightsquigarrow F(t) := 1 - R(t) = P(T \leq t) = 1 - e^{-a t} \rightsquigarrow$ Exponentialverteilung

Satz:

Vor.:

Wie im letzten Satz
 $R(t) := P(T > t)$

Beh.:

$$F(t) := P(T \leq t) = 1 - e^{-a t}$$

Konsequenz: Hier war die Häufigkeit $X(t)$ poissonverteilt. Daraus folgt somit hier, dass die Wartezeiten T nach $P(T \leq t) = 1 - e^{-a t}$ exponentialverteilt sind.

Bemerkung:

Von diesem Sachverhalt kann man auch die Umkehrung beweisen.

Satz:

Vor.:

$$F(t) = P(T \leq t) = 1 - e^{-a t}, \quad \lambda = a \cdot t$$

Beh.:

$$p_m(t) = \frac{\lambda \cdot t}{m!} \cdot e^{-\lambda \cdot t}$$

5.4.2 Warteschlangenmodelle

In den praktischen Anwendungen der Mathematik im Bereich der Planung spielt das Problem der Warteschlangen eine grössere Rolle (Kundenschalter, Drucker, Übermittlung u.s.w.). Betrachten wir z.B.

das Problem von wartenden Kunden an einer Kasse. Die Anzahl der Kunden, die sich pro Zeiteinheit vor der Kasse einfinden, ist meist Poissonverteilt. Die Zeit aber, die pro Kunde zur Abfertigung benötigt wird, ist meist exponentialverteilt. Beide Prozesse spielen hier also zusammen. Die Kombination bestimmt ein Warteschlangensystem. Die weiterreichende Behandlung des Problems ist wieder Gegenstand einer ausgedehnten Theorie, die den Rahmen dieses Kurses sprengt.

5.5 Schätzungen

5.5.1 Schätzer, Punktschätzer

Problem:

Gegeben sei eine Stichprobe von Werten $\{x_1, x_2, \dots, x_n\}$ aus einer weiter nicht bekannten Grundgesamtheit U . Die Werte werden von einer Variablen X angenommen, für deren Verteilungsfunktion ein Modell vorliegen kann. Für U sollen Parameter wie μ, σ^2 u.s.w. mit Hilfe von bekannten empirischen Parametern der Stichprobe geschätzt werden. Als Methode verwenden wir die **Momentenmethode**. Die zu schätzenden Parameter sollen durch die Momente der Verteilung $\mu = E(X)$, $\sigma = E((X - \mu)^2)$ u.s.w. ausgedrückt werden. Dazu verwenden wir **Schätzfunktionen**, **Punktschätzer** oder **Schätzer**. **Schätzwerte** sind zahlenmässige Realisierungen von Punktschätzern. Möchte man hingegen Aussagen gewinnen über die Genauigkeit oder die Sicherheit von Schätzwerten, so kann man dazu **Schätzintervalle** oder **Konfidenzintervalle** (**Vertrauensintervalle**) berechnen.

Definition:

Schätzfunktionen, **Punktschätzer** oder **Schätzer** sind Stichprobenfunktionen, durch die wir unbekannte theoretische Wahrscheinlichkeitsfunktionen oder Parameter einer Grundgesamtheit zufriedenstellend ersetzen wollen.

Trick: Für die Stichprobenwerte x_i wollen wir jetzt individuell neue Variablen X_i einführen:

Annahme: x_1 sei ein Wert für X_1 , x_2 ein Wert für X_2 u.s.w.. Die so postulierten Variablen haben alle dieselbe Verteilung wie X .

Bsp.:

$$\text{Schätzer für } \mu: \bar{X} = \bar{X}(n) = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\text{Schätzer für } \sigma^2: S^2 = S^2(n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ oder}$$

$$Q^2 = Q^2(n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1) S^2}{n}$$

Problem:

Wie gut oder tauglich sind solche Schätzer?

5.5.2 Erwartungstreue

Definition:

Ein Schätzer Y resp. Y_n für einen Parameter Υ der Grundgesamtheit heisst **erwartungstreu** oder **unverzerrt**, wenn der Erwartungswert von Y mit dem Erwartungswert von Υ übereinstimmt. D.h. wenn gilt:

$$E(Y) = E(\Upsilon)$$

Bemerkung:

Gegeben sei eine Stichprobe $\{x_1, x_2, \dots, x_n\}$. Seien die Variablen X_1, X_2, \dots, X_n wieder verteilt wie X . Speziell gilt dann $E(X_i) = E(X)$, $i = 1, \dots, n$, sowie $\sigma^2 = E((X - \mu)^2) = E((X_i - \mu)^2)$, $i = 1, \dots, n$. Der Wertevektor $(x_1, x_2, \dots, x_n)^T$ gibt dann eine mögliche Kombination von Werten, die vom Variablenvektor $(X_1, X_2, \dots, X_n)^T$ angenommen werden können (T bedeutet „transponiert“). Dann sind der Stichprobenmittelwert und die empirische Stichprobenvianz erwartungstreue Schätzer für μ und σ^2 .

Symbol: $X_i \sim X \rightsquigarrow X_i$ verteilt wie X

Satz: (Mittelwert erwartungstreu)

Vor.:

$$X_1, X_2, \dots, X_n \sim X$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \mu = E(X)$$

\bar{X} Schätzer für μ

Beh.:

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

Beweis:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n \cdot \mu = \mu = E(X)$$

☺

Satz: (Empirische Varianz erwartungstreu)

Vor.:

$$X_1, X_2, \dots, X_n \sim X$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \sigma^2 = E((X - \mu)^2)$$

S^2 Schätzer für σ^2

Beh.:

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

Beweis:

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

$$\begin{aligned}
&= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + \frac{2}{n^2} E((X_i - \mu)(n\mu - \sum_{j=1}^n X_j)) + \frac{1}{n^2} E((n\mu - \sum_{j=1}^n X_j)^2) \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + \frac{2}{n^2} E((X_i - \mu) \sum_{j=1}^n (\mu - X_j)) + \frac{1}{n^2} E((\sum_{j=1}^n (\mu - X_j))^2) \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + \frac{2}{n^2} \sum_{j=1}^n E((X_i - \mu)(\mu - X_j)) + \frac{1}{n^2} \sum_{j=1}^n E((\mu - X_j)^2) \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + \frac{2}{n^2} \sum_{j=1}^n E(\underbrace{-(X_i - \mu)(X_j - \mu)}_{i \neq j: \text{ unabh.} \Rightarrow =0}) + \frac{n}{n^2} \sigma^2 \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 - \frac{2}{n^2} n E((X_i - \mu)^2) + \frac{n}{n^2} \sigma^2 \right) = \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 - \frac{2}{n} \sigma^2 + \frac{1}{n} \sigma^2 \right) = \frac{\sigma^2}{n-1} (n - \frac{2}{n} + \frac{n}{n}) \\
&= \frac{\sigma^2}{n-1} (n - 2 + 1) = \frac{\sigma^2}{n-1} (n-1) = \sigma^2 \quad \text{..}
\end{aligned}$$

Es gilt:

$$Q^2 = Q^2(n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)S^2}{n} \Rightarrow E(Q^2) = \frac{(n-1)}{n} E(S^2) = \frac{(n-1)}{n} \sigma^2 \neq \sigma^2 \rightsquigarrow$$

Korollar: Q^2 nicht erwartungstreu.

Bemerkung: Weiter kann man zeigen, dass S selbst nicht allgemein erwartungstreu ist (vgl. Lit. Storm, Bibl. A12).

Bemerkung: Eine schwächere Bedingung als die Erwartungstreue ist die asymptotische Erwartungstreue, für die man fordert:

$$\lim_{n \rightarrow \infty} E(Y_n) = E(\Upsilon)$$

5.5.3 Konsistenz

Sei wieder $X_i \sim X$.

Definition: Ein Schätzer Y resp. Y_n für einen Parameter Υ der Grundgesamtheit heisst **konsistent** oder **schwach konsistent** oder **stichhaltig**, wenn gilt:

$$\forall \varepsilon > 0 \lim_{n \rightarrow \infty} P(|Y_n - \Upsilon| < \varepsilon) = 1$$

D.h. mit wachsendem Stichprobenumfang konvergiert die Wahrscheinlichkeit, dass sich der Schätzer beliebig wenig vom zu schätzenden Parameter unterscheidet, gegen 1.

Satz:**Vor.:**

$$\begin{aligned}\bar{X} &= \bar{X}(n) = \frac{1}{n} \sum_{i=1}^n X_i \\ S^2 &= S^2(n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ Q^2 &= Q^2(n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1) S^2}{n}\end{aligned}$$

Beh.:

1. \bar{X} ist ein konsistenter Schätzer für μ
2. $\lim_{n \rightarrow \infty} \text{Var}(S^2) \rightarrow 0$, $\lim_{n \rightarrow \infty} \text{Var}(Q^2) \rightarrow 0$ oder $X_i \in [-M, M] \Rightarrow S^2$ und Q^2 sind konsistente Schätzer für σ^2

Zum Beweis verwenden wir das Lemma:

Lemma:**Vor.:**

$$\begin{aligned}\forall_i X_i &\sim X \\ E(X_i) &= E(X) = \mu \\ \text{Var}(X_i) &= \text{Var}(X) = \sigma^2\end{aligned}$$

Beh.:

1. $E(\bar{X}) = \mu$
2. $\text{Var}(\bar{X}) = E((\bar{X} - \mu)^2) = E(\bar{X}^2) - \mu^2 = \frac{\sigma^2}{n}$

Beweis:

(Lemma)

$$\begin{aligned}1. \ E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{n}{n} \mu = \mu \\ 2. \ \text{Var}(\bar{X}) &= E((\bar{X} - \mu)^2) = E\left(\left(\frac{1}{n} \sum_{i=1}^n X_i\right) - \left(\frac{n}{n} \mu\right)\right)^2 = E\left(\frac{1}{n^2} \left(\left(\sum_{i=1}^n X_i\right) - (n \mu)\right)^2\right) \\ &= \frac{1}{n^2} E\left(\left(\sum_{i=1}^n (X_i - \mu)\right)^2\right) = \frac{1}{n^2} E\left(\sum_{i,j=1}^n (X_i - \mu)(X_j - \mu)\right) = \frac{1}{n^2} \sum_{i,j=1}^n E\left(\underbrace{(X_i - \mu)(X_j - \mu)}_{i \neq j: \text{unabh.}}\right) \Rightarrow = 0 \\ &= \frac{1}{n^2} \sum_{i=1}^n E((X_i - \mu)^2) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

Zum Beweis des Satzes:Verwende Tschebyscheff: $P(|Y - E(Y)| \geq \varepsilon) \leq \frac{\text{Var}(Y)}{\varepsilon^2}$

$$1. \ \lim_{n \rightarrow \infty} P(|Y_n - \Upsilon| < \varepsilon) = \lim_{n \rightarrow \infty} P(|\bar{X} - \mu| < \varepsilon) = \lim_{n \rightarrow \infty} (1 - P(|\bar{X} - \mu| \geq \varepsilon))$$

$$\begin{aligned}
&= 1 - \lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \varepsilon), \quad P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{Var(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n \varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty \\
&\Rightarrow \lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \varepsilon) = 0 \Rightarrow \lim_{n \rightarrow \infty} P(|\bar{X} - \mu| < \varepsilon) = 1 \\
2. \quad &\lim_{n \rightarrow \infty} P(|Y_n - \Upsilon| < \varepsilon) = \lim_{n \rightarrow \infty} P(|S^2 - \sigma^2| < \varepsilon) = 1 - \lim_{n \rightarrow \infty} P(|S^2 - \sigma^2| \geq \varepsilon), \\
&P(|S^2 - \sigma^2| \geq \varepsilon) = P(|S^2 - E(S^2)| \geq \varepsilon) \leq \frac{Var(S^2)}{\varepsilon^2} = \frac{1}{\varepsilon^2} Var\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\
&(X_i - \bar{X}), (X_j - \bar{X}) \text{ unabhängig für } i \neq j \\
&\text{(p. 106)} \rightsquigarrow \frac{1}{\varepsilon^2} Var\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{\varepsilon^2} \sum_{i=1}^n Var\left(\frac{1}{n-1} (X_i - \bar{X})^2\right) \\
&= \frac{1}{\varepsilon^2} \sum_{i=1}^n E\left(\left(\frac{1}{n-1} (X_i - \bar{X})^2 - \frac{1}{n-1} E((X_i - \bar{X})^2)\right)^2\right) \\
&= \frac{1}{\varepsilon^2 (n-1)^2} \sum_{i=1}^n \underbrace{E(((X_i - \bar{X})^2 - E((X_i - \bar{X})^2))^2)}_{\leq Const=M))} \leq \frac{M n}{\varepsilon^2 (n-1)^2} \rightarrow 0, \quad n \rightarrow \infty
\end{aligned}$$

Weitere Begriffe

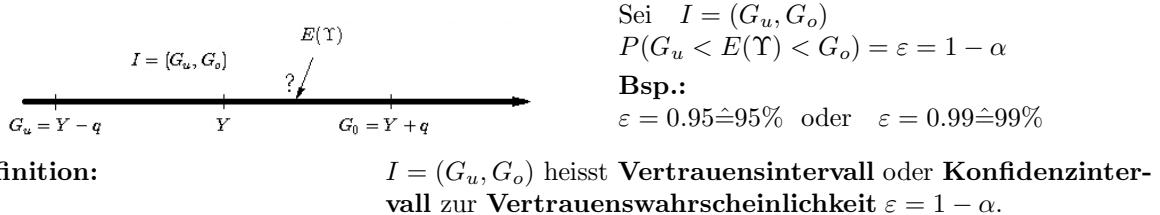
Wirksamkeit, Effizienz, Suffizienz, Wirkungsgrad etc. vgl. Lit..

5.5.4 Vertrauensintervalle I

Begriffe

Möchte man Aussagen gewinnen über die Genauigkeit oder die Sicherheit von Schätzwerten, so kann man dazu **Schätzintervalle** oder **Konfidenzintervalle** (**Vertrauensintervalle**) berechnen.

Sei Y resp. Y_n ein Schätzer für einen Parameter Υ der Grundgesamtheit.



Definition:

$I = (G_u, G_o)$ heisst **Vertrauensintervall** oder **Konfidenzintervall** zur **Vertrauenswahrscheinlichkeit** $\varepsilon = 1 - \alpha$.

Bemerkung:

Vertrauensintervalle interessieren in der Praxis!

Beispiel Mittelwert

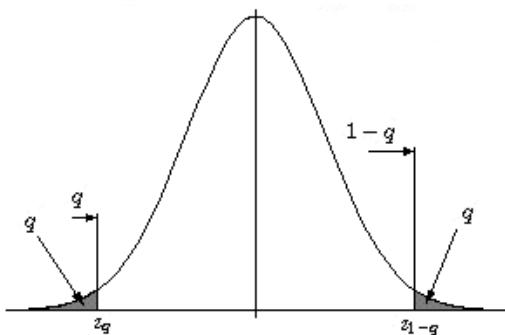
Bsp.:

Sei X gaussverteilt mit schon bekannter Varianz σ^2 . Gesucht ist das Konfidenzintervall für μ zu $\varepsilon = 1 - \alpha$.

Wir wissen: $X \in N(\mu, \sigma^2) \Rightarrow \bar{X} \in N\left(\mu, \frac{\sigma^2}{n}\right)$

$$X \in N(\mu, \sigma^2) \Leftrightarrow Z = \frac{X - \mu}{\sigma} \in N(0, 1)$$

$$\bar{X} \in N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \bar{Z} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \in N(0, 1)$$



Für das Weitere studieren wir die normierte Normalverteilung $N(0, 1)$. Wenn man mit Rechnern arbeitet, ist die Einschränkung auf die Standardsituation nicht notwendig. Wenn man wie früher mit Tabellen arbeitet aber schon.)

Bekannt:

$$1. F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = P(X \leq x) = \Phi(x; \mu, \sigma^2)$$

$$2. \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = P(Z \leq z) = \Phi(z; 0, 1)$$

$$3. Z = \frac{X - \mu}{\sigma} \in N(0; 1) \Leftrightarrow X \in N(\mu; \sigma^2)$$

$$\rightsquigarrow \text{Sei } \Phi(z) = \Phi(z; 0, 1) = P(X \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \rightsquigarrow \Phi(b) = P(X \leq b)$$

$$\rightsquigarrow \text{Sei } q := \Phi(z_q) = \Phi(z_q; 0, 1) \rightsquigarrow q\text{-Quantil (Ordnung } q)$$

Man sieht aus der Skizze:

Formel: $z_q = -z_{1-q}, \Phi(z_{1-q}) = 1 - q = 1 - \Phi(z_q)$

$$\text{Sei } z \sim N(0, 1), \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \text{ und } F(\bar{X}) = F(\bar{X}, \mu, \frac{\sigma^2}{n}) = \Phi\left(\frac{\bar{X} - \mu}{\sigma}; 0, 1\right) = q$$

$$\rightsquigarrow \text{Transformation: } Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \bar{X} = \mu + Z \cdot \frac{\sigma}{\sqrt{n}}$$

$$\text{Sei } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ (Stichprobenmittelwert)}$$

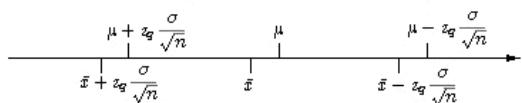
$$\text{Sei } z = z_q$$

$$\rightsquigarrow \text{Es gilt: } (z_q < 0 !!!)$$

$$\mu \in [\bar{x} + z_q \frac{\sigma}{\sqrt{n}}, \bar{x} - z_q \frac{\sigma}{\sqrt{n}}] \Leftrightarrow (\mu - \bar{x}) \in [z_q \frac{\sigma}{\sqrt{n}}, -z_q \frac{\sigma}{\sqrt{n}}] \Leftrightarrow (\bar{x} - \mu) \in [z_q \frac{\sigma}{\sqrt{n}}, -z_q \frac{\sigma}{\sqrt{n}}]$$

$$\Leftrightarrow |\bar{x} - \mu| \leq |z_q| \frac{\sigma}{\sqrt{n}} \Leftrightarrow \bar{x} \in [\mu + z_q \frac{\sigma}{\sqrt{n}}, \mu - z_q \frac{\sigma}{\sqrt{n}}]$$

Wichtig: $\rightsquigarrow \mu \in [\bar{x} + z_q \frac{\sigma}{\sqrt{n}}, \bar{x} - z_q \frac{\sigma}{\sqrt{n}}] \Leftrightarrow \bar{x} \in [\mu + z_q \frac{\sigma}{\sqrt{n}}, \mu - z_q \frac{\sigma}{\sqrt{n}}]$



$$\begin{aligned} \rightsquigarrow \varepsilon = 1 - \alpha &= P(\mu \in [\bar{X} + z_q \frac{\sigma}{\sqrt{n}}, \bar{X} - z_q \frac{\sigma}{\sqrt{n}}]) = P(\bar{X} \in [\mu + z_q \frac{\sigma}{\sqrt{n}}, \mu - z_q \frac{\sigma}{\sqrt{n}}]) \\ &= P(\mu + z_q \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu - z_q \frac{\sigma}{\sqrt{n}}) = F(\mu - z_q \frac{\sigma}{\sqrt{n}}) - F(\mu + z_q \frac{\sigma}{\sqrt{n}}) = (1 - q) - q = \Phi(z_{1-q}) - \Phi(z_q) \\ &= \Phi(z_{1-q}) - (1 - \Phi(z_{1-q})) = 2\Phi(z_{1-q}) - 1 \end{aligned}$$

$$\Rightarrow 2 - \alpha = 2\Phi(z_{1-q}), \quad \Phi(z_{1-q}) = 1 - \frac{\alpha}{2} = \int_{-\infty}^{z_{1-q}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Formel: $\Phi(z_{1-q}) = 1 - \frac{\alpha}{2} = \int_{-\infty}^{z_{1-q}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad \Phi(z_q) = \frac{\alpha}{2} = \int_{-\infty}^{z_q} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

z_q nennen wir daher **zweiteilige α -Schranke** der normierten Gaussverteilung Φ . z_q ist das $\frac{\alpha}{2}$ -Quantil.

Zahlenbeispiel: $\alpha = 0.05 \hat{=} 5\%$

$$\rightsquigarrow \Phi(z_{1-q}) = \int_{-\infty}^{z_{1-q}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1 - \frac{0.05}{2} = 0.975$$

Mathematica-Programm:

| FindRoot[Integrate[E^(-t^2/2)/Sqrt[2 Pi], {t, -Infinity, z}] == 0.975, {z, 1}]|

Output:

| {z -> 1.95996}|

\rightsquigarrow Wir finden daher $z_{1-q} \approx 1.960$.

Seien $\bar{x} = 156.38, \sigma^2 = 0.20, n = 16$.

$$\begin{aligned} \rightsquigarrow \varepsilon = 1 - \alpha = 0.95 &= P(\mu \in [\bar{X} + z_q \frac{\sigma}{\sqrt{n}}, \bar{X} - z_q \frac{\sigma}{\sqrt{n}}]) \\ &\approx P(\mu \in [156.38 - 1.960 \frac{\sqrt{0.20}}{\sqrt{16}}, 156.38 + 1.960 \frac{\sqrt{0.20}}{\sqrt{16}}]) \\ &\approx P(\mu \in [156.161, 156.599]) \rightsquigarrow I = [156.161, 156.599] = [a, b], \quad b - a \approx 0.438269 \end{aligned}$$

I ist das Vertrauensintervall zur Vertrauenswahrscheinlichkeit $\varepsilon = 1 - \alpha = 0.95$. Die Wahrscheinlichkeit, dass μ ausserhalb dieses Intervalls liegt, ist $\alpha = 0.05$.

5.5.5 Wichtige Testfunktionen

Oft kennt man μ und σ nicht. Das bedeutet, dass man mit Schätzungen für diese Grössen arbeiten muss. Es stellt sich dann die Frage nach den Typen der entsprechenden Verteilungsfunktionen, falls man μ und σ durch Schätzungen ersetzt. Da die Resultate eigentlich erstaunlich sind, müssen wir dazu einige Tatsachen herleiten.

$$\text{Sei } (\bar{X} = \sum_{i=1}^n X_i \Rightarrow \sum_{i=1}^n (X_i - \bar{X}) = (\sum_{i=1}^n X_i) - n \bar{X} = \sum_{i=1}^n X_i - n \frac{1}{n} \sum_{i=1}^n X_i = \sum_{i=1}^n X_i - \sum_{i=1}^n X_i = 0$$

$U_i = (X_i - \bar{X})$, $\sum_{i=1}^n U_i = 0 \rightsquigarrow$ Die Variablen $U_i = (X_i - \bar{X})$ sind abhängig.

$$\rightsquigarrow \text{Z.B. } U_1 = - \sum_{i=2}^n U_i$$

$$\rightsquigarrow 0 = 0^2 = (\sum_{i=1}^n U_i)^2 = \sum_{i=1}^n U_i^2 + \sum_{(i \neq k \wedge i, k=1)}^n U_i \cdot U_k$$

$$\Rightarrow U_1^2 = - \sum_{i=2}^n U_i^2 + \sum_{(i \neq k \wedge i, k=1)}^n U_i \cdot U_k = - \sum_{i=2}^n U_i^2 + \sum_{(i \neq k \wedge i, k=2)}^n U_i \cdot U_k + U_1 \cdot \sum_{k=2}^n U_k + U_1 \cdot \sum_{i=2}^n U_i$$

$$= - \sum_{i=2}^n U_i^2 + \sum_{(i \neq k \wedge i, k=2)}^n U_i \cdot U_k + (- \sum_{i=2}^n U_i) \cdot \sum_{k=2}^n U_k + (- \sum_{k=2}^n U_k) \cdot \sum_{i=2}^n U_i = h(U_2, U_3, \dots, U_n)$$

$$\rightsquigarrow U_1^2 = h(U_2, U_3, \dots, U_n)$$

\rightsquigarrow Die Variablen $U_i^2 = (X_i - \bar{X})^2$ sind abhängig.

Lemma: Die Variablen $U_i = (X_i - \bar{X})$ und auch die Variablen $U_i^2 = (X_i - \bar{X})^2$ sind abhängig

Weiter ist unmittelbar klar, dass eine Koordinatentransformation $X_i \mapsto X_i - \mu$ resp. $\bar{X}_i \mapsto \bar{X}_i - \mu$ keinen Einfluss auf die Unabhängigkeit oder Abhängigkeit hat. Wir verwenden folgendes Symbol:

Symbol: X_i unabhängig $\Leftrightarrow X_i \in \{\heartsuit\}$
 X_i abhängig $\Leftrightarrow X_i \in \{\spadesuit\}$

Lemma: Vor.: $\{X_i\} \in \{\heartsuit\}$ resp. $\{X_i\} \in \{\spadesuit\}$

Beh.: $\{X_i - \mu\} \in \{\heartsuit\}$ resp. $\{X_i - \mu\} \in \{\spadesuit\}$

$\rightsquigarrow X_i \in N(\mu, \sigma^2) \Rightarrow (X_i - \mu) \in N(0, \sigma^2)$
 $\rightsquigarrow X_i \in N(0, \sigma^2)$ bringt keine Einschränkung weiterer Bedingungen.

Sei $X_i \in N(0, \sigma^2)$, $Q_n = \sum_{i=1}^n (X_i - \bar{X})^2$

$\rightsquigarrow Q_n$ kann als Summe von $n - 1$ unabhängigen Quadraten geschrieben werden. Das wollen wir jetzt herleiten:

Lemma: Vor.:

$X_i \in N(0, \sigma^2)$, $Q_n = \sum_{i=2}^n (X_i - \bar{X})^2$

Beh.:

$\exists_{Y_1, Y_2, \dots, Y_{n-1} \in N(0, \sigma^2), \in \{\heartsuit\}} : Q_n = \sum_{i=2}^{n-1} Y_i^2$

Beweis:

(Lemma) Sei

$$\begin{aligned}
Y_1 &= \frac{1}{\sqrt{1 \cdot 2}} (X_1 - X_2) \\
Y_2 &= \frac{1}{\sqrt{2 \cdot 3}} (X_1 + X_2 - 2 X_3) \\
&\vdots = \vdots \\
Y_{n-1} &= \frac{1}{\sqrt{(n-1) \cdot n}} (X_1 + X_2 + \dots + X_{n-1} - (n-1) X_n) \\
Y_n &= \frac{1}{\sqrt{n}} (X_1 + X_2 + \dots + X_{n-1} + X_n) \\
\rightsquigarrow \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} &= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = M \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \\
M &= \begin{pmatrix} \frac{1}{\sqrt{1 \cdot 2}} & -\frac{1}{\sqrt{1 \cdot 2}} & 0 & \dots & 0 \\ \frac{1}{\sqrt{2 \cdot 3}} & \frac{1}{\sqrt{2 \cdot 3}} & -\frac{2}{\sqrt{2 \cdot 3}} & \ddots & \vdots \\ \vdots & & \vdots & & 0 \\ \frac{1}{\sqrt{(n-1) \cdot n}} & \dots & \frac{1}{\sqrt{(n-1) \cdot n}} & \frac{-(n-1)}{\sqrt{(n-1) \cdot n}} & \frac{1}{\sqrt{n}} \end{pmatrix} = (\vec{c}_1, \dots, \vec{c}_n) = \begin{pmatrix} \vec{z}_1^T \\ \vdots \\ \vec{z}_n^T \end{pmatrix} \\
\text{Z.B. } \vec{z}_j^T &= (\underbrace{\frac{1}{\sqrt{(j) \cdot (j+1)}}, \frac{1}{\sqrt{(j) \cdot (j+1)}}, \dots, \frac{1}{\sqrt{(j) \cdot (j+1)}}}_{j}, -\frac{j}{\sqrt{(j) \cdot (j+1)}}, 0, \dots, 0)
\end{aligned}$$

Sei $j < k < n$

$$\begin{aligned}
\rightsquigarrow \vec{z}_j^T \cdot \vec{z}_k &= \langle \vec{z}_j, \vec{z}_k \rangle = \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} + \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} + \dots \\
&+ \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} - \frac{j}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} + 0 = \frac{j-j}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} = 0
\end{aligned}$$

Sei $j = k < n \rightsquigarrow$

$$\begin{aligned}
\vec{z}_j^T \cdot \vec{z}_j &= \left(\frac{1}{\sqrt{(j) \cdot (j+1)}} \right)^2 + \left(\frac{1}{\sqrt{(j) \cdot (j+1)}} \right)^2 + \dots + \left(\frac{1}{\sqrt{(j) \cdot (j+1)}} \right)^2 + \left(-\frac{j}{\sqrt{(j) \cdot (j+1)}} \right)^2 = \\
j \cdot \frac{1}{(j) \cdot (j+1)} + \frac{j^2}{(j) \cdot (j+1)} &= \frac{j^2+1}{j^2+1} = 1
\end{aligned}$$

$$\begin{aligned}
\text{Sei } j < k = n \rightsquigarrow \vec{z}_j^T \cdot \vec{z}_n &= \langle \vec{z}_j, \vec{z}_n \rangle = \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} + \dots \\
&+ \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} - \frac{j}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} = \frac{j-j}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} = 0
\end{aligned}$$

Sei $j = k = n \rightsquigarrow$

$$\vec{z}_n^T \cdot \vec{z}_n = \left(\frac{1}{\sqrt{n}} \right)^2 + \left(\frac{1}{\sqrt{n}} \right)^2 + \dots + \left(\frac{1}{\sqrt{n}} \right)^2 = n \cdot \frac{1}{n} = 1 \rightsquigarrow$$

$$\vec{z}_j^T \cdot \vec{z}_k = \delta_{i,k}, \quad M \cdot M^T = E \Rightarrow M^T = M^{-1} \Rightarrow \frac{1}{\det M^{-1}} = \det M = \det M^T = \det M^{-1} \Rightarrow \det M = \pm 1$$

Konsequenz: $\rightsquigarrow M$ orthogonal:

$$\vec{Y} = M \cdot \vec{X} \Rightarrow r^2 = \vec{Y}^2 = \langle \vec{Y}, \vec{Y} \rangle = \langle M \cdot \vec{X}, M \cdot \vec{X} \rangle = (M \cdot \vec{X})^T \cdot M \cdot \vec{X} = (\vec{X}^T \cdot M^T) \cdot (M \cdot \vec{X}) = \vec{X}^T \cdot (M^T \cdot M) \cdot \vec{X} = \vec{X}^T \cdot E \cdot \vec{X} = \vec{X}^T \cdot \vec{X} = \langle \vec{X}, \vec{X} \rangle = \vec{X}^2 \Rightarrow |\vec{X}| = |\vec{Y}|, \sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2$$

Nach Konstruktion war:

$$\begin{aligned} Y_n &= \frac{1}{\sqrt{n}} (X_1 + X_2 + \dots + X_{n-1} + X_n) = \sqrt{n} \bar{X} \Rightarrow Y_n^2 = n \bar{X}^2 \rightsquigarrow \\ Y_n^2 &= \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2 = n \cdot \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 = n \cdot (\bar{X})^2 \Rightarrow Q_n = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i^2) - 2 X_i \bar{X} + \bar{X}^2 \\ &= \left(\sum_{i=1}^n (X_i^2) \right) - (2 \bar{X} \sum_{i=1}^n X_i) + \left(\sum_{i=1}^n \bar{X}^2 \right) = \left(\sum_{i=1}^n (X_i^2) \right) - (2 \bar{X} n \bar{X}) + (n \bar{X}^2) = \left(\sum_{i=1}^n (X_i^2) \right) - (n \bar{X}^2) \\ &= \left(\sum_{i=1}^n (Y_i^2) \right) - (Y_n^2) = \sum_{i=1}^n (Y_i^2) \end{aligned}$$

Dass die Y_i unabhängig sind, ersieht man aus der Definition der Y_i . Es ist $Y_{i-1} = Y_{i-1}(X_1, X_2, \dots, \cdot_i)$, $Y_i = Y_i(X_1, X_2, \dots, \cdot_{i+1})$. Daher ist es möglich, dass die eine Variable ändert, die andere aber nicht. $\rightsquigarrow \text{☺}$

Lemma:

$$\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2 = \sum_{i=1}^{n-1} Y_i^2 + Y_n = Q_n + n \bar{X}^2 = \sum_{i=2}^n (X_i - \bar{X})^2 + n \bar{X}^2 = \sum_{i=1}^{n-1} Y_i^2 + n \bar{X}^2, Y_n^2 = n \bar{X}^2$$

Weiter verwenden wir den folgenden Satz von Seite 109:

Satz:

Summe unabhängiger normalverteilter Variablen

Vor.:

Seien X_1, X_2, \dots, X_n unabhängig, normalverteilt mit

Mittelwerte $\mu_1, \mu_2, \dots, \mu_n$

Varianzen $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$

Beh.:

$\sum_{i=1}^n X_i$ normalverteilt

Mittelwert $\mu = \sum_{i=1}^n \mu_i$

Varianz $\sigma^2 = \sum_{i=1}^n \sigma_i^2$

Seien die X_i unabhängig und $\in N(\mu, \sigma^2)$, $k < n$. Dann sind auch die Summanden von Y_k , d.h. $X_1, X_2, \dots, -k X_{k+1}$ unabhängig und es gilt (vgl. Seite 124):

$$\begin{aligned} \mu(Y_k) &= \sum_{i=1}^n \mu_i = \frac{\mu}{\sqrt{k \cdot (k+1)}} + \frac{\mu}{\sqrt{k \cdot (k+1)}} + \dots - \frac{k \mu}{\sqrt{k \cdot (k+1)}} = \mu \left(k \frac{1}{\sqrt{k \cdot (k+1)}} - \frac{k}{\sqrt{k \cdot (k+1)}} \right) = 0 \\ \sigma^2(Y_k) &= \sum_{i=1}^n \sigma_i^2 = \frac{\sigma^2}{k \cdot (k+1)} + \frac{\sigma^2}{k \cdot (k+1)} + \dots + (-k)^2 \frac{\sigma^2}{k \cdot (k+1)} = \sigma^2 \frac{k+k^2}{k^2+k} = \sigma^2 \\ \Rightarrow \sigma^2 \left(\frac{Y_k}{\sigma} \right) &= \left(\frac{1}{\sigma} \right)^2 \cdot \sigma^2(Y_k) = \frac{\sigma^2}{\sigma^2} = 1 \rightsquigarrow \end{aligned}$$

Lemma:**Vor.:**

$$X_i \in N(\mu, \sigma^2), \quad X_i \in \{\heartsuit\}, \quad k < n,$$

$$Y_k = \frac{1}{\sqrt{k \cdot (k+1)}} (X_1 + X_2 + \dots + X_k - k \cdot X_{k+1})$$

Beh.:

$$\frac{Y_k}{\sigma} \in N(0, 1)$$

$$\text{Sei } Y := \frac{Q_n}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=2}^{n-1} Y_i^2 = \sum_{i=2}^{n-1} \frac{Y_i^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \rightsquigarrow$$

Y ist eine Summe von $n-1$ Quadraten von normalverteilten Variablen $\frac{Y_i}{\sigma} \in N(0, 1)$. Die Y_i sind unabhängig. Y ist somit χ^2 -verteilt mit $n-1$ Freiheitsgraden! Man hat somit den Satz:

Satz:**Vor.:**

$$X_1, \dots, X_n \in N(\mu, \sigma^2), \quad X_i \in \{\heartsuit\}$$

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

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

Beh.:

$$Y \in \chi_{n-1}^2$$

$$\text{Sei } Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \rightsquigarrow \text{Es gilt:}$$

$$Z = \frac{\sqrt{n}}{\sigma} \cdot \bar{X} + \frac{(-1) \cdot \sqrt{n} \cdot \mu}{\sigma} = c_1 \bar{X} + c_2 \Rightarrow (Z \in \{\heartsuit\} \Leftrightarrow \bar{X} \in \{\heartsuit\})$$

Weiter sieht man:

$$Q_n = \sum_{i=1}^n (X_i - \bar{X})^2 \Rightarrow \frac{\partial}{\partial \bar{X}} Q_n = \frac{\partial}{\partial \bar{X}} \sum_{i=1}^n (X_i - \bar{X})^2 = 2 \sum_{i=1}^n (X_i - \bar{X}) \cdot (-1) = -2 \left(\left(\sum_{i=1}^n X_i \right) - n \cdot \bar{X} \right)$$

$$= -2 \left((n \cdot \bar{X}) - n \cdot \bar{X} \right) = 0 \Rightarrow Q_n(\bar{X}) = \text{const.}(X_1, \dots, X_n)$$

Andererseits ist:

$k < n-1 \rightsquigarrow Y_k = Y_k(X_1, \dots, X_{k+1}), \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}(X_1, \dots, X_n) \rightsquigarrow$ Man kann Y_k verändern ohne \bar{X} zu verändern und umgekehrt.

$k = n-1 \rightsquigarrow Y_{n-1} = \lambda_1 \cdot \bar{X}(X_1, \dots, X_n) + \lambda_2 \cdot X_n \rightsquigarrow$ Auch hier kann man Y_k verändern ohne \bar{X} zu verändern und umgekehrt. \rightsquigarrow

Satz:**Vor.:**

$$X_1, \dots, X_n \in N(\mu, \sigma^2), \in \{\heartsuit\}$$

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

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

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

Beh.:

$$\{Y, Z\} \subset \{\heartsuit\}$$

Konsequenz:

Mit $Z \in N(0, 1)$ und $Y \in \chi^2_{n-1}$ lässt sich nun die Zufallsvariable $T = \frac{Z}{\sqrt{Y/(n-1)}}$ bilden, die somit einer t -Verteilung resp. einer Student-Verteilung genügt.

Daher folgt:

$$T = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \cdot \frac{1}{\sqrt{\frac{1}{\sigma^2} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} = \sqrt{n} \frac{\bar{X} - \mu}{\sqrt{\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2}} = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$$

Korollar:**Vor.:**

$$X_1, \dots, X_n \in N(\mu, \sigma^2), \in \{\heartsuit\}$$

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$$

Beh.:

$$T \in t_{n-1}$$

Bemerkung:

Man beachte, dass wir von $T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$ jetzt die Verteilungsfunktion kennen. In T steckt neben den Stichprobenwerten X_i aber nur μ . Man hat daher hier ein Mittel in der Hand anhand von Stichprobenwerten zu einer Aussage über μ zu gelangen!

5.5.6 Vertrauensintervalle II

Problem: Wegen der Konsistenz wissen wir, dass bei grossen Stichproben \bar{X} ein guter Schätzer für μ

ist. Wie ist es aber bei kleinen Stichproben?

Mittelwertschätzung bei unbekannter Varianz und kleinen Stichproben

Gegeben seien die Stichprobenwerte $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$, $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ und n . Damit bilden

wir die Variable $T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$, die nun einer t -Verteilung mit $n - 1$ Freiheitsgraden genügt. Wie kennen die zugehörigen Funktionen:

$$f(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{z^2}{n})^{(n+1)/2}}, \quad F(z) = P(T \leq z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \cdot \int_{-\infty}^z \frac{1}{(1 + \frac{u^2}{n})^{(n+1)/2}} du$$

Aus $f(z)$ ersieht man, dass die Verteilung symmetrisch ist zur y -Achse. Daher gilt:

$$F(-c) = 1 - F(c) \Rightarrow P(-c \leq T \leq c) = P(T \leq c) - P(T < -c) = F(c) - F(-c) = F(c) - (1 - F(c)) = 2F(c) - 1$$

Für eine vorgegebene Wahrscheinlichkeit $P(-c \leq T \leq c) = \gamma = 1 - \alpha$ kann man daher aus $\gamma = 2F(c) - 1$ das c bestimmen.

Mit Hilfe des nun bekannten c kann man jetzt auf ein Vertrauensintervall für μ schliessen:

$$\begin{aligned} -c \leq T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S} \leq c &\Rightarrow \frac{-c \cdot S}{\sqrt{n}} - \bar{X} \leq -\mu \leq \frac{c \cdot S}{\sqrt{n}} - \bar{X} \Rightarrow \bar{X} - \frac{c \cdot S}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{c \cdot S}{\sqrt{n}}, \\ I &= [\bar{X} - \frac{c \cdot S}{\sqrt{n}}, \bar{X} + \frac{c \cdot S}{\sqrt{n}}] \end{aligned}$$

Methode:

1. **Geg.:** $\{x_1, x_2, \dots, x_n\}$, $\gamma = 1 - \alpha$
2. $\rightsquigarrow \bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$, $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$, $T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$
3. $\rightsquigarrow \gamma = 2F(c) - 1 \rightsquigarrow c$
4. $\rightsquigarrow \mu \in I = [\bar{X} - \frac{c \cdot S}{\sqrt{n}}, \bar{X} + \frac{c \cdot S}{\sqrt{n}}]$ mit Wahrscheinlichkeit γ

Zahlenbeispiel

Sei $\gamma = 0.95$, $\alpha = 0.05$ Daten erzeugen:

Mathematica-Programm:

```

<< Graphics`Graphics`;
<< Statistics`DescriptiveStatistics`;
gamma = 0.95;
m = Table[243.79 + 2.9 Random[], {n, 1, 25}]

```

Output:

```
{246.21, 244.807, 244.481, 245.184, 246.195, 245.083, 244.915, 246.492,
243.98, 243.993, 246.646, 244.568,
245.833, 243.864, 246.659, 244.947,
246.081, 243.834, 245.631, 244.54, 246.166, 245.966, 245.905, 246.605,
246.645}
```

Mathematica-Programm:

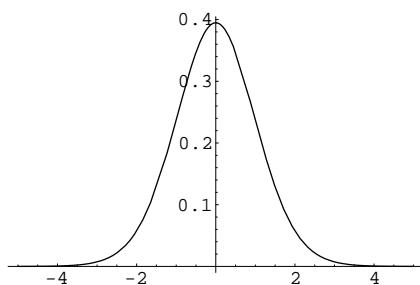
```
meanM = Mean /. LocationReport[m][[1]];
standD = StandardDeviation /. DispersionReport[m][[2]];
{LocationReport[m], DispersionReport[m], Length[m]}
```

Output:

```
 {{Mean -> 245.409, HarmonicMean -> 245.406, Median -> 245.631},
 {Variance -> 0.93796, StandardDeviation -> 0.968484, SampleRange -> 2.82473,
 MeanDeviation -> 0.8572, MedianDeviation -> 0.823889,
 QuartileDeviation -> 0.819016}, 25}
```

Mathematica-Programm:

```
f[z_, m_] :=
Gamma[(m + 1)/2]/(Sqrt[m Pi] Gamma[m/2]) / (1 + z^2/m)^((m + 1)/2);
f[u, Length[m] - 1];
Plot[f[u, Length[m] - 1], {u, -5, 5}];
```



Der Plot zeigt die Dichte der t -Verteilung $f(z, 24)$.

Nun berechnen wir c direkt mit Mathematica. Wer für eine solche Berechnung die Mittel nicht hat, muss auf **Tabellen** zurückgreifen!

Mathematica-Programm:

```
{"c", c = c /. FindRoot[gamma == 2 Evaluate[Integrate[f[u, 15],
{u, -Infinity, c}]] - 1, {c, 2}] // Chop,
"Interv", {xU = meanM - standD c/Sqrt[Length[m]],
x0 = meanM + standD c/Sqrt[Length[m]]}}
```

Output:

```
| {"c", 2.13145, "Interv", {244.996, 245.822}}
```

Mathematica-Programm:

```
| m1 = Select[m, (xU - 0.000001 < #1 < x0 + 0.000001) &]
```

Output:

```
| {245.184, 245.083, 245.631}
```

Mathematica-Programm:

```
| Complement[m, m1]
```

Output:

```
| {243.834, 243.864, 243.98, 243.993, 244.481, 244.54, 244.568, 244.807,
 244.915, 244.947, 245.833, 245.905, 245.966, 246.081, 246.166, 246.195,
 246.21, 246.492, 246.605, 246.645, 246.646, 246.659}
```

Lösung: Wie man sieht, liegt der Mittelwert μ mit einer Wahrscheinlichkeit von 95% im Intervall $I = [244.996, 245.822]$ ($\bar{X} \approx 245.409$). Die meisten der Daten liegen jedoch ausserhalb dieses Intervalls. Das zeigt, wie gut der Schätzer \bar{X} für μ ist.

5.5.7 Vertrauensintervalle III

Problem: Wegen der Konsistenz wissen wir, dass wir bei grossen Stichproben S^2 ein guter Schätzer für σ^2 ist. Wie ist es aber bei kleinen Stichproben?

Varianzschätzung bei unbekanntem Mittelwert und kleinen Stichproben

Auf Seite 148 haben wir festgestellt:

Satz:**Vor.:**

$$X_1, \dots, X_n \in N(\mu, \sigma^2), \in \{\heartsuit\}$$

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

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

Beh.:

$$Y \in \chi_{n-1}^2$$

Konsequenz:

$$\gamma = P(c \leq Y) = 1 - P(0 \leq Y < c) = 1 - (F_{\chi^2, n-1}(c) - F_{\chi^2, n-1}(0)) = 1 - F_{\chi^2, n-1}(c)$$

Bemerkung:

Da die Varianz stets positiv ist, begnügen wir uns hier mit einer einseitigen Abschätzung im Gegensatz zu der zweiseitigen Abschätzung beim Vertrauensintervall für μ .

Wir wissen:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ K_n x^{\frac{n-2}{2}} e^{-\frac{x}{2}} & x > 0 \end{cases} \quad K_n := \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \quad F(x) = K_n \int_0^{\infty} u^{\frac{n-2}{2}} e^{-\frac{u}{2}} du$$

Damit können wir c berechnen aus $\gamma = P(c \leq Y) = 1 - F_{\chi^2, n-1}(c)$.

$$\rightsquigarrow c \leq Y = \frac{n-1}{\sigma^2} S^2 \Rightarrow 0 \leq \sigma^2 \leq \frac{n-1}{c} S^2, \quad \sigma^2 \in [0, \frac{n-1}{c} S^2]$$

Methode:1. **Geg.:** $\{x_1, x_2, \dots, x_n\}$, $\gamma = 1 - \alpha$

$$2. \rightsquigarrow S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad Y = \frac{n-1}{\sigma^2} S^2$$

$$3. \rightsquigarrow \gamma = 1 - F_{\chi^2, n-1}(c) \rightsquigarrow c$$

$$4. \rightsquigarrow \sigma^2 \in [0, \frac{n-1}{c} S^2] \text{ mit Wahrscheinlichkeit } \textit{gamma}$$

ZahlenbeispielSei $\gamma = 0.95$, $\alpha = 0.05$ Daten erzeugen:**Mathematica-Programm:**

```

<< Graphics`Graphics`;
<< Statistics`DescriptiveStatistics`;
gamma = 0.95;
m = Table[243.79 + 2.9 Random[], {n, 1, 25}]

```

Output:

```
{244.584, 244.288, 245.634, 245.936, 244.285, 245.569, 245.374, 246.273,
243.942, 244.729, 244.258, 246.685, 246.468, 245.045, 245.432, 244.24,
245.035, 246.111, 246.083, 243.971, 245.717, 244.192, 246.688, 244.861, 244.923}
```

Mathematica-Programm:

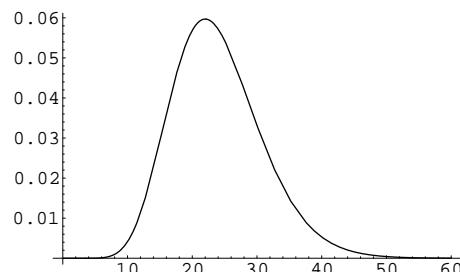
```
var = Variance /. DispersionReport[m][[1]];
{LocationReport[m], DispersionReport[m], Length[m]}
```

Output:

```
 {{Mean -> 245.213, HarmonicMean -> 245.21, Median -> 245.045},
 {Variance -> 0.768021, StandardDeviation -> 0.876368, SampleRange -> 2.74647,
 MeanDeviation -> 0.753255, MedianDeviation -> 0.760244,
 QuartileDeviation -> 0.842562}, 25}
```

Mathematica-Programm:

```
chi[x_, n_] := 1/(2^(n/2) Gamma[n/2]) x^((n - 2)/2) E^(-x/2);
Plot[chi[u, Length[m] - 1], {u, 0, 60}];
```



Der Plot zeigt die Dichte der χ^2 -Verteilung $\chi^2(24)$.

Nun berechnen wir c direkt mit Mathematica. Wer für eine solche Berechnung die Mittel nicht hat, muss auf **Tabellen** zurückgreifen!

Mathematica-Programm:

```
{"c", c = c /. FindRoot[gamma == 1 - Evaluate[Integrate[chi[u, 15],
{u, 0, c}]], {c, 40}] // Chop, "Interv", {xU = 0, x0 = ((Length[m]-1)/c) var }}
```

Output:

| |
|--|
| {"c", 7.26094, "Interv", {0, 2.53858}} |
|--|

Lösung: Wie man sieht, liegt die Varianz σ^2 mit einer Wahrscheinlichkeit von 95% im Intervall $I = [0, 2.53858]$ ($S^2 \approx 0.768021$). Weitere „Runs“ liefern: $\gamma = 0.99 \rightsquigarrow I = [0, 3.52486]$, $\gamma = 0.90 \rightsquigarrow I = [0, 2.15667]$

5.5.8 Vertrauensintervalle IV

Bemerkung zur Binomialverteilung

Wahrscheinlichkeitsfunktion:

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad q = 1 - p, \quad \mu = np, \quad \sigma^2 = npq$$

Problem: Gesucht ist ein Vertrauensintervall für p bei der Binomialverteilung.

Für grosse n gilt die Näherungsformel:

$$f(x) \approx f^*(x) = \frac{1}{\sqrt{w\pi\sigma}} \cdot e^{-\frac{z^2}{2}}, \quad Z = \frac{X - \mu}{\sigma}, \quad Z \in N(0, 1)$$

Seien X die Anzahl Erfolge beim n -maligen Ausführen des Experimentes. Dann kann man wieder wie folgt argumentieren:

$$\begin{aligned} \gamma &= P(-c \leq Z \leq c) \approx \int_{-c}^c f^*(x) dx \\ \Rightarrow c &= \dots \Rightarrow -c \leq Z = \frac{X - \mu}{\sigma} = \leq c \Rightarrow -c \cdot \sigma + \mu \leq X \leq +c \cdot \sigma + \mu \dots \end{aligned}$$

5.5.9 Automatische Bestimmung der Vertrauensbereiche

Dem geneigten Leser wird es wohl klar sein, dass heutzutage für alle Standardverfahren Computerprogramme vorhanden sind. Mit Mathematica (z.B. V.4.0) kann man Vertrauensbereiche wie folgt bestimmen:

Bsp.:

Vertrauensbereiche für Mittelwert und Varianz, Vertrauensniveau 0.9. Als Grundlage verwendete Verteilungen: t -Verteilung für den Mittelwert (mean) und χ^2 -Verteilung für die Varianz (var).

Mathematica-Programm:

| |
|--|
| <pre><< Statistics`ConfidenceIntervals`; data = {2.2, 1.3, 0.8, 1.0, 1.1, 3.1, 3.4, 3.3, 2.2, 0.5}; {"mean->", MeanCI[data, ConfidenceLevel -> 0.9], "var->", VarianceCI[data, ConfidenceLevel -> 0.9]}</pre> |
|--|

Output:

| |
|---|
| {"mean->", {1.25473, 2.52527}}, "var->", {0.638868, 3.25072}} |
|---|

5.6 Signifikanztests

5.6.1 Hypothesen

Idee: Einen Parameter testen heisst, zuerst über die Lage dieses Parameters **Hypothesen** aufzustellen und dann anschliessend die Wahrscheinlichkeit dieser Hypothesen zu untersuchen. Üblicherweise führt das dann zu einer Verwerfung oder im andern Fall zu einer Tolerierung resp. einer Akzeptanz der Hypothesen auf der Grundlage einer geforderten Wahrscheinlichkeit. Wie das genau funktioniert, wollen wir an charakteristischen Beispielen studieren.

Methode:

1. Wir beginnen mit der **Nullhypothese** H_0 : Z.B. postulieren wir, dass ein Parameter z einen Sollwert z_0 einhält: $H_0 = (z = z_0)$. Z.B. $H_0 = (\mu = \mu_0)$

Andere Beispiele:

$$H_0 = (z \neq z_0) \Leftrightarrow H_0 = (z \leq z_0), \quad H_0 = (z \not> z_0) \Leftrightarrow H_0 = (z \geq z_0)$$

2. Normalerweise wird der **Nullhypothese**, für die eine **hohe Wahrscheinlichkeit** zu erwarten ist, eine **Alternativhypothese** H_1 mit **kleiner Wahrscheinlichkeit** α gegenübergestellt, z.B. hier $H_1 = (z \neq z_0)$. Jetzt kann man die Wahrscheinlichkeit der Abweichung $z - z_0$ mittels einer geeigneten Testverteilung feststellen. (Mit z_0 fliesst somit auch die Nullhypothese als Voraussetzung in das Verfahren ein.) Man verwendet dazu normalerweise einen geeigneten Schätzer $\zeta(z)$ für z bei einem gegebenen Stichprobenumfang. Falls der real vorhandene Wert bei nur sehr kleiner verrechneten Wahrscheinlichkeit α trotzdem für die Alternativhypothese spricht, so hat man es mit einem seltenen Ereignis zu tun. Die Zufallsabweichung ist daher unwahrscheinlich. Der Alternativhypothese muss daher eine grössere als die angenommene Wahrscheinlichkeit zugebilligt werden, die Nullhypothese ist daher zu verwerfen. Im andern Fall kann man die kleine Wahrscheinlichkeit der Alternativhypothese nicht angreifen und somit die Nullhypothese nicht verwerfen. Man muss diese also tolerieren resp. man akzeptiert sie („auf dem Niveau α “).

5.6.2 Zweiseitige Alternative, t –Test

Bsp.: Geg.:

$X \in N(\mu, \sigma^2)$, $H_0 = (\mu = \mu_0) \rightsquigarrow H_1 = (\mu \neq \mu_0)$, $\zeta(\mu) = \bar{x} = \bar{x}_n$
 μ_0 ist ein abgemachter, durch Pläne vorgegebener und einzuhaltender Wert.

Sei $\bar{x} = 127.3$, $\mu_0 = 128$, $s = 1.2$, $n = 20$, $\alpha = 0.01$

Dabei kann man als **Testverteilung** die Stichprobenfunktion „normierte Differenz“ oder Abweichung von μ_0 verwenden. Hier wird als Voraussetzung die Nullhypothese als in Form des Terms $\bar{X} - \mu_0$ eingebaut. μ_0 wird bei der Normierung auf null abgebildet.

$$\rightsquigarrow T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$$

Wichtig: $\rightsquigarrow T$ genügt einer t -Verteilung mit $n - 1$ Freiheitsgraden.

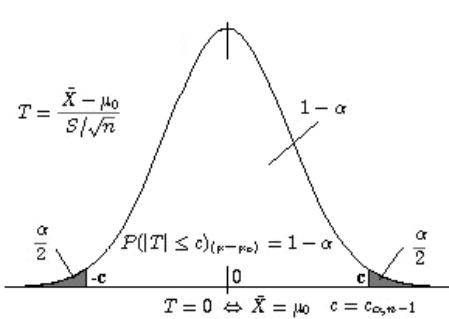
$$T \sim f(T) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{T^2}{n})^{\frac{n+1}{2}}}, \quad F(c) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \cdot \int_{-\infty}^c \frac{1}{(1 + \frac{u^2}{n})^{\frac{n+1}{2}}} du$$

\rightsquigarrow Studiere: $P(|T| > c) = \alpha$

Die Wahrscheinlichkeit, dass unter der Voraussetzung $\mu = \mu_0$ die normierte Differenz grösser ist als c soll hier gleich einer gegebenen Wahrscheinlichkeit α sein. α ist eine Zahl, die man vernünftig festsetzen muss.

Definition:

α heisst **Signifikanzniveau** (Bedeutungsniveau).



Es gilt:

$$\begin{aligned} P(|T| > c) = \alpha &\Leftrightarrow P(|T| \leq c) = 1 - \alpha, \\ P(|T| \leq c) &= P(-c \leq T \leq c) \\ &= F(c) - F(-c) = F(c) - (1 - F(c)) = 2F(c) - 1, \\ 1 - \alpha &= 2F(c) - 1 \Rightarrow 2 - 2F(c) = \alpha \\ 1 - F(c) &= \frac{\alpha}{2} \rightsquigarrow c = \dots \rightsquigarrow \\ P(|T| > c) = \alpha &\Rightarrow \\ (|T_{real}| > c)_{P=\alpha} &\Leftrightarrow (\neg(|T_{real}| \leq c))_{P=\alpha} \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow \neg((-c \leq T_{real} = \sqrt{n} \frac{\bar{X} - \mu_0}{S} \leq c))_{P=\alpha} \Leftrightarrow (\neg(-\frac{c \cdot s}{\sqrt{n}} + \mu_0 \leq \bar{X} \leq \frac{c \cdot s}{\sqrt{n}} + \mu_0))_{P=\alpha} \\ &\Leftrightarrow ((\bar{X} < -\frac{c \cdot s}{\sqrt{n}} + \mu_0) \vee (\frac{c \cdot s}{\sqrt{n}} + \mu_0 < \bar{X}))_{P=\alpha} \end{aligned}$$

Nun berechnen wir c direkt mit Mathematica. Wer für eine solche Berechnung die Mittel nicht hat, muss auf **Tabellen** zurückgreifen!

Mathematica-Programm:

```
alpha = 0.01;
f[z_, n_] :=
  Gamma[(n + 1)/2]/(Sqrt[n Pi] Gamma[n/2]) / (1 + z^2/n)^((n + 1)/2);
{"c", c = c /. FindRoot[
  alpha/2 == Evaluate[1 - Integrate[f[u, 20], {u, -Infinity, c}]], {c, 2}] // Chop, "Interv", {-c, +c}]}
```

Output:

```
{"c", 2.84534, "Interv", {-2.84534, 2.84534}}
```

$$\begin{aligned} &((\bar{X} < -\frac{c \cdot s}{\sqrt{n}} + \mu_0) \vee (\frac{c \cdot s}{\sqrt{n}} + \mu_0 < \bar{X}))_{P=\alpha} \\ &\rightsquigarrow ((127.3 < -\frac{2.84534 \cdot 1.2}{\sqrt{20}} + 128) \vee (\frac{2.84534 \cdot 1.2}{\sqrt{20}} + 128 < 127.3))_{P=0.01} \\ &\Leftrightarrow ((127.3 < 127.237) \vee (128.763 < 127.3))_{P=0.01} \Leftrightarrow (127.3 \notin [127.237, 128.763])_{P=0.01} \end{aligned}$$

Ergebnis: Mit einer Wahrscheinlichkeit von $\alpha = 0.01$ soll also 127.3 nicht im Intervall [127.237, 128.763] liegen. 127.3 liegt aber in diesem Intervall. Das bedeutet, dass das so ist, weil wirklich die Nullhypothese $H_0 = (\mu = \mu_0)$ richtig ist oder weil das zufällig so ist. **Man kann daher hier die Nullhypothese auf Grund des Tests bei einem Signifikanzniveau $\alpha = 0.01$ nicht ablehnen.** Würde jedoch 127.3 ausserhalb des Intervalls liegen, so wäre die Wahrscheinlichkeit dafür 0.01. Bei unserer Testphilosophie müsste man die Nullhypothese also ablehnen.

Achtung:

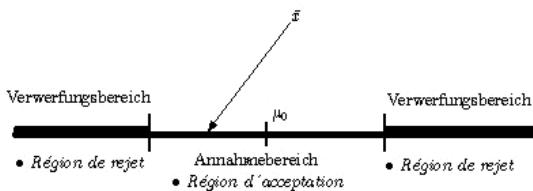
Kleine Wahrscheinlichkeit bedeutet nicht Unmöglichkeit!

Bemerkung:

Das hier verwendete Testverfahren nennen wir **t -Test**, da wir die t -Verteilung von Student als Testverteilung verwendet haben.

Annahmebereich, Verwerfungsbereich

Im eben besprochenen Beispiel ist die Alternativhypothese verworfen worden, weil der berechnete Wert des Schätzers im gefundenen Intervall und nicht ausserhalb lag. Wir reden hier von **Annahmebereich** und **Verwerfungsbereichen** (vgl. Skizze).



Achtung:

Die Verwerfungsbereiche müssen nicht in allen Fällen symmetrisch liegen. Sie können auch einseitig liegen!

5.6.3 Einseitige Alternative, t -Test

Wir betrachten hier den Fall $H_0 = (z = z_0)$ contra $H_1 = (z > z_0) \rightsquigarrow$ **einseitiger Alternativtest**.

Bsp.:

Jemand hat eine $N(U) = 10'000$ mal zufällig eine Münze geworfen. $N(E) = 5094$ mal ist Kopf resp. Wappen realisiert worden. Da $h(E) = \frac{N(E)}{N(U)} = 0.5094$ gilt, wird $H_0 = (p_E = 0.5)$ angezweifelt. H_0 soll daher hier gegen die vermutete einseitige Alternative $H_1 = (p_E > 0.5)$ auf dem Signifikanzniveau $\alpha = 0.05$ getestet werden.

Sei $X =$ Anzahl Kopf. Der Sollwert für $p = 0.5$ ist $X = 5'000$. Die Stichprobe ist verteilt nach der Binomialverteilung $Bi(10'000; 0.5) \hat{=} N(\mu, \sigma^2)$ (Die Binomialverteilung nähert sich der Normalverteilung an!). Dabei gilt:

$$\mu = np, \quad \sigma^2 = npq = np(1-p), \quad P(X \leq c) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^c e^{-\frac{(u-\mu)^2}{2\sigma^2}} du,$$

$$d = \frac{c-\mu}{\sigma}, \quad Z = \frac{X-\mu}{\sigma} \Rightarrow P(X \leq c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{u^2}{2}} du = P(Z \leq d)$$

$$H_1 = (p_E > 0.5) \Rightarrow P(X > c)_{H_0=(p_E=0.5)} = \alpha \Rightarrow P(X \leq c) = P(Z \leq d) = 1 - \alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{u^2}{2}} du$$

In Zahlen unter der Voraussetzung der Nullhypothese $p = 0.5$:

$$\mu = np = 10'000 \cdot 0.5 = 5'000, \sigma^2 = npq = 10'000 \cdot 0.5 \cdot (1 - 0.5) = 10'000 \cdot 0.25 = 2'500, \sigma = 50$$

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 5'000}{50} = 0.1X - 500, X = Z \cdot 50 + 5'000 = 50Z + 5'000$$

$$P(X \leq c) = P(Z \leq d) = 1 - \alpha = 0.95 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{u^2}{2}} du$$

Mathematica-Programm:

```
alpha=0.05;
root = FindRoot[
  Integrate[E^(-u^2/2)/Sqrt[2 Pi], {u, -Infinity, d}] == 1-alpha, {d, 1}];
{d1 = d /. root, 50 d1 + 5000}
```

Output:

```
{1.64485, 5082.24}
```

Wir finden daher als Resultat die Schranke $c = 5082.24$ auf der Grundlage $\alpha = 0.05$. Die Wahrscheinlichkeit der Alternativhypothese $H_1 = (p_E > 0.5)$ ist $P(X > c)_{H_0=(p_E=0.5)} = P(X > 5082.24) = 0.05$. Dagegen haben wir aber festgestellt: $H(E) = N(E) = 5094$ bei $P(X > 5082.24) = 0.05$. Mit $H(E) = 5094$ ist daher so gesehen ein seltenes Ereignis mit der Wahrscheinlichkeit < 0.05 eingetreten. Daher kann man die Alternativhypothese nicht für zufällig halten und nimmt sie auf der Grundlage von $\alpha = 0.05$ an. Damit ist die Nullhypothese verworfen. Die Münze wird daher auf dieser Grundlage nicht für homogen gehalten. Wie man sieht, ist hier der Ablehnungsbereich oder kritische Bereich einseitig. ($\{X \mid X > 5082.24\}$).

Direkte Rechnung

Eine andere, kürzere Methode ist es, $p_1 = P(X > 5094)$ direkt zu berechnen und mit der signifikanten Wahrscheinlichkeit $\alpha = 0.05 = P(X > c)_{H_0=(p_E=0.5)}$ der Alternativhypothese $H_1 = (p_E > 0.5)$ zu vergleichen. Ist $p_1 < \alpha = 0.05$, so ist die Alternativhypothese daher nicht zufällig und die Nullhypothese daher abzulehnen.

Mathematica-Programm:

```
sigma = 50; mu = 5000; cH = 5094;
p1 = 1/(Sqrt[2Pi] sigma) Integrate[
  E^(-(mu-u)^2/(2 sigma^2)), {u, a, b}]/. {a->cH, b->Infinity} // N
```

Output:

```
0.030054
```

Konsequenz: $p_1 = 0.030054 < \alpha = 0.5 \leadsto H_0$ ablehnen

5.6.4 Testrisiken

Aufgrund eines Tests könnte es möglich sein, dass man eine richtige Nullhypothese zufällig und fälschlicherweise ablehnt oder eine falsche Nullhypothese zufällig und fälschlicherweise nicht ablehnt. Etwas Ähnliches ist uns schon bei den Konfidenzintervallen begegnet. Wir übertragen daher die Begriffe von damals:

Definition: Lehnen wir aufgrund eines Tests eine wahre Nullhypothese zufällig ab, so begehen wir einen **Fehler erster Art**.

Einen Fehler 1. Art begehen wir, wenn wir die Alternativhypothese an Stelle der Nullhypothese annehmen. Die Alternativhypothese ist mit der Wahrscheinlichkeit α erfüllt (Signifikanzzahl).

Konsequenz:

Die Wahrscheinlichkeit eines Fehlers 1. Art ist gerade die Signifikanzzahl α .

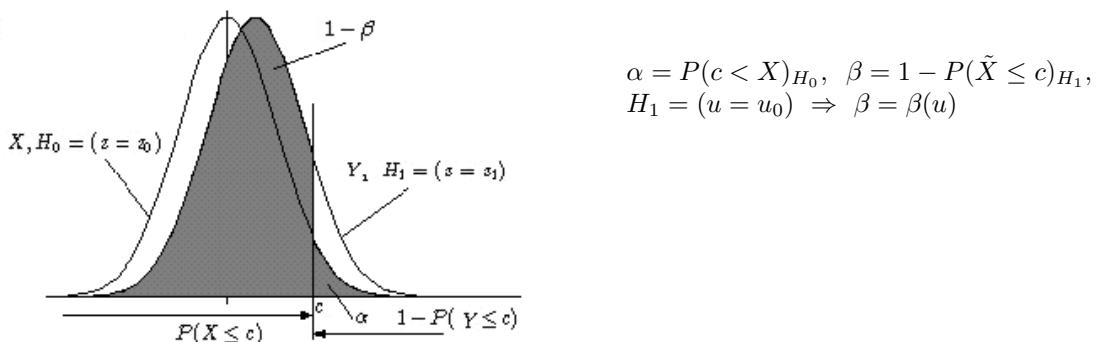
Definition: Lehnen wir aufgrund eines Tests eine falsche Nullhypothese zufällig nicht ab, so begehen wir einen **Fehler zweiter Art**.

Begehen wir einen Fehler zweiter Art, so wäre z.B. an Stelle der Nullhypothese $H_0 = (z = z_0)$ eine Alternativhypothese $H_1 = (z = z_1)$ richtig. Aufgrund von H_0 hat man eine Testfunktion konstruiert, die unrichtig ist. Die richtige Testfunktion müsste man aufgrund von H_1 konstruieren. Die Wahrscheinlichkeit für das Eintreffen der zu H_1 gehörigen Alternativhypothese bezüglich des aufgrund von H_0 berechneten c sei β . β ist somit gleich der Signifikanzzahl einer neuen Alternativhypothese bezüglich der neuen Nullhypothese H_1 . Somit ist $1 - \beta$ die Wahrscheinlichkeit die neue Nullhypothese H_1 nicht abzulehnen und daher die alte Nullhypothese H_0 abzulehnen, d.h. einen Fehler 2. Art zu vermeiden.

Definition: Die Wahrscheinlichkeit β , einen Fehler 2. Art zu vermeiden heisst **Macht des Tests**.

Die Wahrscheinlichkeit $1 - \beta$ einen Fehler 2. Art nicht zu vermeiden, heisst **Risiko 2. Art** oder **Produzentenrisiko**.

Die Signifikanzzahl oder Wahrscheinlichkeit eines Fehlers 1. Art heisst auch **Risiko 1. Art** oder **Konsumentenrisiko**.



Bemerkung:

Ein grösseres c bedeutet zwar ein kleineres α , jedoch auch ein grösseres $1 - \beta$. Mit grösseren n bekommt man aber in der Regel schlankere Kurven. Bei gegebenem α wird dadurch $1 - \beta$ kleiner. Der Test gewinnt an **Trennschärfe**.

5.6.5 Chi-quadrat-Test für die Varianz

Bsp.:

Sei $X_i \in N(\mu, \sigma^2)$, $\sigma_0 = 50$, $n = 20$, $s^2 = 61.3$, $H_0 = (\sigma^2 = \sigma_0^2)$, $H_1 = (\sigma^2 > \sigma_0^2)$

Bemerkung: $\sigma^2 > \sigma_0^2$ ist die einzige interessante Alternative.

$$\frac{S^2}{\sigma_0^2} = \frac{1}{n-1} \sum_{i_1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} = \frac{1}{n-1} \sum_{i_1}^n Z_i^2, \quad Y_i \in N(0, 1) \Rightarrow T = \frac{S^2}{\sigma_0^2} (n-1) \text{ hat eine } \chi_{n-1}^2 \text{-Verteilung}$$

Sei $\alpha = 0.05 \rightsquigarrow P(T > c)_{\sigma^2=50} = \alpha = 0.05 \Rightarrow P(T \leq c)_{\sigma^2=50} = 1 - \alpha = 0.95$

Mathematica-Programm:

```
chi[x_, n_] := 1/(2^(n/2) Gamma[n/2]) x^((n - 2)/2) E^(-x/2);
alpha = 0.05; sQ = 61.3; sigmQ = 50; n = 20;
{"c", c = c /. FindRoot[
  1 - alpha == Evaluate[Integrate[chi[u, n - 1], {u, 0, c}]], {c, 40}]
 // Chop, "Interv", {xU = 0, sQc = c sigmQ/(n - 1) }}
```

Output:

```
| {"c", 30.1435, "Interv", {0, 79.325}} |
```

Die Alternativhypothese besagt, dass ein Wert mit der Wahrscheinlichkeit $\alpha = 0.05$ ausserhalb des Intervalls $\{0, 79.325\}$ zu liegen kommt. $s^2 = 61.3$ liegt jedoch innerhalb des Intervalls. Es ist demnach kein Grund vorhanden die Alternativhypothese anzunehmen und damit die Nullhypothese zu verwerfen.

5.6.6 Automatisches testen

Bsp.:

Mit Mathematica, Zitat aus den Unterlagen: (Man beachte die vorausgesetzte Normalverteilung und die t -Verteilung!)

A test of a statistical hypothesis is a test of assumption about the distribution of a variable. Given sample data, you test whether the population from which the sample came has a certain characteristic. You can use functions in this package to test hypotheses concerning the mean, the variance, the difference in two population means, or the ratio of their variances.

The data that is given as an argument in a test function is assumed to be normally distributed. As a consequence of the Central Limit Theorem, you can disregard this normality assumption in the case of tests for the mean when the sample size, n , is large and the data is unimodal. The test functions accept as arguments the list of univariate data, a hypothesized parameter and relevant options.

Hypothesis tests for the mean are based on the normal distribution when the population variance is known, and the Student t distribution with $n-1$ degrees of freedom when the variance has to be estimated. If you know the standard deviation instead of the variance, you can also specify KnownStandardDeviation \rightarrow std.

The output of a hypothesis test is a pvalue, which is the probability of the sample estimate being as extreme as it is given that the hypothesized population parameter is true. A twosided test can be requested using TwoSided → True. For more detailed information about a test use FullReport → True. This causes the parameter estimate and the test statistic to be included in the output. You can also specify a significance level using SignificanceLevel → siglev, which yields a conclusion of the test, stating acceptance or rejection of the hypothesis.

Mathematica-Programm:

```
<< Statistics`HypothesisTests`;
data = {35, 33, 42, 32, 42, 43, 37, 44, 41, 39};
MeanTest[data, 34, KnownVariance -> 8, SignificanceLevel -> 0.9]
```

Output:

```
{OneSidedPValue -> 4.01256 * 10^-8,
"Reject null hypothesis at significance level" -> 0.9}
```

Mathematica-Programm:

```
| MeanTest[data1, 38, KnownVariance -> 10, SignificanceLevel -> 0.1]
```

Output:

```
{OneSidedPValue -> 0.211855,
"Fail to reject null hypothesis at significance level" -> 0.1}
```

Mathematica-Programm:

```
| MeanTest[data1, 38, KnownVariance -> 10]
```

Output:

```
{OneSidedPValue -> 0.211855,
OneSidedPValue -> 0.211855}
```

Mathematica-Programm:

```
| MeanTest[data1, 38]
```

Output:

```
| OneSidedPValue -> 0.286061
```

Mathematica-Programm:

```
| MeanTest[data1, 38, KnownVariance -> 10, FullReport -> True] // TeXForm
```

Output:

```
\{ {{\!\!\!\text{Mvariable}\!\!\!}\{FullReport\}}\!\!\!\rightarrow
  {\!\!\!\text{matrix}\!\!\!}\{ 38.8 \& 0.8 \& \!\!\!{\text{Muserfunction}\!\!\!}\{NormalDistribu
    tion\}()\ \!\!\!{\text{cr}\!\!\!}\} \!\!\!{\text{,}}\!\!\!},
  {{\!\!\!\text{Muserfunction}\!\!\!}\{OneSidedPValue\}}\!\!\!\rightarrow
  \{0.211855\}}\!\!\!
```

~

```
{FullReport -> Mean      TestStat      Distribution
 38.8          0.8          NormalDistribution[], OneSidedPValue -> 0.211855}
```

5.6.7 Testrezept im Überblick

1. Fasse das zu untersuchende Problem in eine Hypothese \sim Nullhypothese z.B. $H_0 = (a = a_0)$. Gegeben sind hier ein Wert a_0 und ein Schätzer $a = \tilde{a}_0$. Zum Wert a soll eine Verteilungsfunktion X existieren mit speziell $X = a$.
2. Über $H_0 = (a = a_0)$ ist oft direkt keine Wahrscheinlichkeitsaussage möglich. Formuliere daher zu H_0 eine Alternativhypothese H_1 , über die mittels einer geeigneten Verteilungsfunktion und einem Intervall eine Wahrscheinlichkeitsaussage gemacht werden kann. Ein Beispiel ist die einseitige Alternativhypothese $H_1 = (a > a_0)$.
3. Transformiere X bijektiv und stetig in eine andere Zufallsvariable $Z = g(X)$ (Testverteilung!), deren Verteilfunktion direkt berechenbar ist oder in Form von Tabellen vorliegt. $\sim X = g^{-1}(Z)$. Wenn g bijektiv und stetig ist, sind g und g^{-1} streng monoton. Ungleichungen bezüglich Z transformieren sich daher in Ungleichungen bezüglich X .
4. Zu $H_1 = (a > a_0) = H_1 = (a_0 < a)$ ist bei einem geeigneten a_1 aufgrund der Verteilung von X dann die Aussage " $a_0 < a \leq a_1$ " oder " $a_0 < X \leq a_1$ " wahrscheinlich und " $a_1 < a$ " oder " $a_1 < X$ " unwahrscheinlich. Transformiert haben diese Aussagen wegen der Monotonie z.B. die Form " $g(a_0) < g(X) = Z \leq g(a_1)$ " resp. z.B. zu " $g(a_1) = c < g(X) = Z$ ".
5. Zu einer vorgeschlagenen Wahrscheinlichkeit (Signifikanzzahl α) kann man nun die Ungleichung $P(c < Z) = \alpha$ resp. $P(Z \leq c) = 1 - \alpha$ mit Computern oder Tabellen nach c auflösen. $\sim c$
6. $c < Z$ entspricht z.B. $a_1 = g^{-1}(c) < X = g^{-1}(Z)$. Ist nun der Schätzer $\tilde{a}_0 \in (c, \infty)$, so hat man ein seltenes Ereignis der Alternativhypothese in einem Intervall mit kleiner Wahrscheinlichkeit α . Die Realität „fällt also in ein unwahrscheinliches Intervall“! Man kann daher die Alternativhypothese nicht verwerfen. Daher lehnt man in diesem Fall die Nullhypothese ab.

5.6.8 Vergleichstest für Mittelwerte I

Man kann auf den ersten Blick zwei verschiedenartige Stichproben unterscheiden: Erstens solche, die aus verschiedenen Individuen bestehen und zweitens solche, die aus verschieden gewonnenen Messwerten an denselben Individuen bestehen. Wir betrachten hier einmal den zweiten Fall.

Bsp.: Geg.:

Zwei Stichproben: Paarweise am selben Werkstück an zwei verschiedenen Orten A , B mit dem Verfahren V_A gemessene Daten und mit dem Verfahren V_B gemessene Daten:

| x_i | y_i | $d_i = x_i - y_i$ |
|-------|-------|-------------------|
| 15.46 | 14.37 | +0.09 |
| 16.35 | 16.36 | -0.01 |
| 14.11 | 14.03 | +0.08 |
| 19.21 | 19.23 | -0.02 |
| 17.89 | 17.82 | +0.07 |
| 15.66 | 15.62 | +0.04 |

Wir haben es hier mit zwei verbundenen Stichproben zu tun mit paarweise zusammengehörigen Werten (x_i, y_i) und den zugehörigen Variablen (X_i, Y_i) . Wir wollen voraussetzen, dass alle X_i resp. Y_i unabhängig von i sind und gleiche Varianzen σ_X^2 resp. σ_Y^2 haben.

Unter diesen Voraussetzungen wollen wir die Hypothesen $H_0 = (\mu_{X_i} = \mu_{Y_i})$ gegen die einseitigen Alternativen $H_1 = (\mu_{X_i} > \mu_{Y_i})$ testen.

Wir wissen, dass die Variablen Gaussverteilungen haben und daher zu den Differenzen $d_i = x_i - y_i$ eine gaussverteilte Variable D gehört mit $\mu_D = \mu_X - \mu_Y$ und $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2$. Nun können wir mit Hilfe des t -Tests die Nullhypothese $\mu_D = 0$ gegen die Alternative $\mu_D > 0$ testen.

$$\begin{aligned} \rightsquigarrow \bar{d} &= \frac{1}{n} \sum_{i_1}^n d_i = 0.0416667, \quad s_D = \sqrt{\frac{1}{n-1} \sum_{i_1}^n (d_i - \bar{d})^2} = 0.0470815, \quad \mu_0 = 0, \quad \alpha = 0.05, \quad n = 6 \\ Z = g(X) = T &= \frac{\bar{D} - \mu_0}{S/\sqrt{n}} = \frac{X - \mu_0}{S/\sqrt{n}} \rightsquigarrow X = \bar{D} = g^{-1}(Z) = \mu_0 + \frac{TS}{\sqrt{n}} = \mu_0 + \frac{ZS}{\sqrt{n}} \end{aligned}$$

$$P(c < T = Z) = \alpha = 0.05 \Rightarrow P(T \leq 0) = 1 - \alpha \rightsquigarrow c$$

$$c < Z \Leftrightarrow g^{-1}(c) < g^{-1}(Z) \Leftrightarrow \mu_0 + \frac{cS}{\sqrt{n}} = g^{-1}(c) < g^{-1}(Z) = X$$

$$\Leftrightarrow 0 + c \frac{0.0470815}{\sqrt{6}} = c \cdot 0.0192209 = g^{-1}(c) < g^{-1}(Z) = X$$

$$\rightsquigarrow \text{Frage: } c \cdot 0.0192209 \stackrel{?}{<} X = \bar{d} = 0.0416667 \rightsquigarrow$$

Mathematica-Programm:

```

alpha = 0.05;
f[z_, m_] :=
  Gamma[(m + 1)/2]/(Sqrt[m Pi] Gamma[m/2]) /((1 + z^2/m)^(m + 1)/2);
FindRoot[
  1 - alpha == Evaluate[Integrate[f[t, 5], {t, -Infinity, c}]], {c, 3}]
// Chop

```

Output:

```
{c -> 2.01505}
```

$$\rightsquigarrow c \cdot 0.0192209 = 2.01505 \cdot 0.0192209 = 0.0387311 < 0.0416667 = \bar{d}$$

\rightsquigarrow Seltenes Ereignis, Alternativhypothese nicht verwerfen, Nullhypothese verwerfen.

5.6.9 Vergleichstest für Mittelwerte II

Wir betrachten den Fall, dass zwei unabhängige Stichproben $\{x_1, \dots, x_n\}$ für $X \in N(\mu_X, \sigma_X^2)$ und $\{y_1, \dots, y_m\}$ für $Y \in N(\mu_Y, \sigma_Y^2)$ mit allgemein $n = n_X \neq m = n_Y$ und $\sigma_X = \sigma_Y$ vorliegen. Gegeben seien die Schätzungen \bar{x} für μ_X und entsprechend \bar{y} , s_X , s_Y . Getestet werden soll die Nullhypothese $H_0 = (\mu_X = \mu_Y)$ gegen die Alternative $H_1 = (\mu_X > \mu_Y)$. Man sieht rasch, dass man wegen $n_X \neq n_Y$ jetzt nicht mehr mit den Differenzen eine neue Variable konstruieren kann. Um trotzdem eine Testvariable zu konstruieren, braucht es hier also etwas mehr Theorie. Des Rahmens wegen verzichten wir hier auf eine weitere Herleitung und begnügen uns mit dem Resultat:

Formel:

Die folgende Variable genügt unter unseren Voraussetzungen einer t -Verteilung mit $n_X + n_Y - 2$ Freiheitsgraden:

$$T_{XY} = \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}}$$

Bsp.: $\alpha = 0.05$

| x_i | y_i | $d_i = x_i - y_i$ |
|-------|-------|-------------------|
| 15.46 | 14.37 | +0.09 |
| 16.35 | 16.36 | -0.01 |
| 14.11 | 14.03 | +0.08 |
| 19.21 | 19.23 | -0.02 |
| 17.89 | 17.82 | +0.07 |
| 15.66 | 15.62 | +0.04 |
| 15.48 | | |

Sei $H_0 = (\mu_X = \mu_Y)$, $H_1 = (\mu_X > \mu_Y)$

\rightsquigarrow Mit Mathematica:

```

a = {15.46, 16.35, 14.11, 19.21, 17.89, 15.66, 15.48};
b = {14.37, 16.36, 14.03, 19.23, 17.82, 15.62};

Length[a]

7

Length[b]

6

LocationReport[a]

{Mean -> 16.3086, HarmonicMean -> 16.1616, Median -> 15.66}

LocationReport[b]

{Mean -> 16.2383, HarmonicMean -> 16.0372, Median -> 15.99}

DispersionReport[a]

{Variance -> 2.93031, StandardDeviation -> 1.71182, SampleRange -> 5.1,
 MeanDeviation -> 1.29265, MedianDeviation -> 0.69,
 QuartileDeviation -> 1.02}

DispersionReport[b]

{Variance -> 4.04326, StandardDeviation -> 2.01079, SampleRange -> 5.2,
 MeanDeviation -> 1.565, MedianDeviation -> 1.725,
 QuartileDeviation -> 1.725}

alpha = 0.05; n = Length[a] + Length[b] - 2;
f[z_, m_] :=
  Gamma[(m + 1)/2]/(Sqrt[m Pi] Gamma[m/2]) / (1 + z^2/m)^((m + 1)/2);
c = c /.
  FindRoot[
    1 - alpha == Evaluate[Integrate[f[u, n], {u, -Infinity, c}]], {c, 3}]
// Chop

1.79588

nX = Length[a]; nY = Length[b];
mX = Mean /. LocationReport[a][[1]];
mY = Mean /. LocationReport[b][[1]];
sX = StandardDeviation /. DispersionReport[a][[2]];
sY = StandardDeviation /. DispersionReport[b][[2]];
{nX, nY, mX, mY, sX, sY}

{7, 6, 16.3086, 16.2383, 1.71182, 2.01079}

const = Sqrt[((nX - 1)sX^2 + (nY - 1)sY^2)(nX + nY)/(nX nY (nX + nY - 1))]

```

0.987397

c * const

1.77325

c * (mX - mY)

0.12614

$$\rightsquigarrow P(c < T_{XY}) = \alpha, \quad T = T_{XY} = \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}}$$

$$\rightsquigarrow (c < T_{XY})_{\alpha=0.05} \Leftrightarrow |_{\alpha=0.05} :$$

$$c \cdot \sqrt{\frac{((n_X - 1) S_X^2 + (n_Y - 1) S_Y^2) \cdot (n_X + n_Y)}{n_X n_Y (n_X + n_Y - 2)}} < (\bar{x} - \bar{y}) \cdot \sqrt{\frac{((n_X - 1) S_X^2 + (n_Y - 1) S_Y^2) \cdot (n_X + n_Y)}{n_X n_Y (n_X + n_Y - 2)}} \\ \rightsquigarrow 1.77325 < 0.12614$$

Die letzte Aussage trifft aber nicht zu, d.h. ($c < T_{XY}$) ist hier nicht erfüllt. Die Alternativhypothese kann man also nicht annehmen und die Nullhypothese daher nicht ablehnen.

5.7 Weitere Testverfahren

5.7.1 Eine Typeneinteilung von Tests

Grundsätzlich kann man sich beliebige Kriterien ausdenken, um Tests in Klassen einzuteilen. Für uns leisten die folgenden zwei Typologien ihre Dienste:

1. **Parametertests**, z.B. verteilungsabhängige Tests. Z.B. $H_0 = (\mu_X = \mu_Y)$ soll getestet werden.

2. **Parameterfreie Tests**, z.B. die **Rangtests** oder der **Vorzeichentest**.

1. **Verteilungsabhängige Tests**, z.B. Parametertests.

2. **Verteilungsunabhängige Tests**, z.B. Rangtests. Oder auch **Verteilungs- oder Anpassungstests**, mit denen geprüft wird, ob eine vermutete Verteilung zu den empirischen Werten passt.

5.7.2 Parameterfreier Test: Der Vorzeichentest

Bsp.: Geg.:

Tägliche Umsatzmengen von zwei gleichen Arbeitsteams A und B mit identischen Aufträgen während 10 Tagen. Notiert ist nur die Differenz $W_A - W_B$ in Verrechnungseinheiten. Man teste die Nullhypothese, dass beide Teams gleich gut sind gegen die Alternative, dass man mit Team A einen systematisch höheren

Umsatz erzieht als mit B . Sei $\alpha = 0.05$ (klein): Wenn die Realität für H_1 eine Wahrscheinlichkeit $P < \alpha$ ergibt, ist H_1 bemerkenswert, unwahrscheinlich oder „signifikant“.

| | | | | | | | | | | |
|-------------|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|
| $W_A - W_B$ | 0.9 | 2.1 | 1.1 | 0.0 | 1.1 | -0.4 | 1.4 | 0.6 | 0.5 | 0.1 |
| <i>sgn</i> | + | + | + | . | + | - | + | + | + | + |

Es gilt: $H_0 = (W_A = W_B)$ wahr $\rightsquigarrow P(\text{sgn}(W_A - W_B) = " + ") = P(\text{sgn}(W_B - W_A) = " + ")$

Der Wert 0.0 kann gestrichen werden, da er die Entscheidung nicht ändern kann. Somit haben wir 8 mal „+“ unter 9 Werten.

Idee: Sei $X = \text{Anzahl } "+ \text{ unter den } n \text{ Werten}$. Unter der Voraussetzung von H_0 ist X binomialverteilt mit $p = 0.5$ ($P(\text{sgn}(W_A - W_B) = " + ") = P(\text{sgn}(W_B - W_A) = " + ")$).

$$\begin{aligned} \rightsquigarrow P(X \geq 8) &= 1 - P(X \leq 7) = 1 - \sum_{k=0}^7 P(X = k) = 1 - \sum_{k=0}^7 \binom{9}{k} p^k q^{9-k} \\ &= 1 - \sum_{k=0}^7 \binom{9}{k} 0.5^k 0.5^{9-k} = 1 - 0.5^9 \sum_{k=0}^7 \binom{9}{k} = 1 - 0.5^9 (2^9 - \binom{9}{8} - \binom{9}{9}) \approx 0.0195313 \\ &\Rightarrow P(X \geq 8) \approx 0.0195313 \leq \alpha = 0.05 \end{aligned}$$

Man hat es also hier mit einem seltenen Ereignis mit $X = 8$ zu tun, das eingetroffen ist, obwohl $P(X \geq 8) \approx 0.0195313 \leq P(X \geq c) = \alpha = 0.05$ gilt. Aus denselben Überlegungen wie vorher mussten wir daher eine Alternativhypothese akzeptieren und die Nullhypothese daher verwerfen. Wäre hingegen $\alpha = 0.01$, so könnte man die Alternativhypothese nicht akzeptieren und die Nullhypothese daher nicht verwerfen.

5.7.3 Chi-Quadrat-Anpassungstest

Problem: Es ist mittels eines Tests zu prüfen, ob eine bestimmte Variable eine gegebene Verteilungsfunktion $F(t)$ besitzt. Wir wollen das an einem Beispiel studieren:

Bsp.: Geg.:

1. Disjunkte Ereignisse A_1, \dots, A_m als Resultate eines Zufallsexperiments.

$$A_i \cap A_k = \{\}, \quad i \neq k, \quad \bigcup_{i=1}^m = \Omega$$

2. $P(A_i) = p_i$ unbekannt $\rightsquigarrow H_0 = (P(A_i) = p_i), \quad i = 1, \dots, n$

3. Test von H_0 : Nehme Stichprobe vom Umfang n , $n_i = |A_i|$

$$4. \text{ Bilde: } \tau = \sum_{i=1}^m \frac{(n p_i - n_i)^2}{n p_i} = \sum_{i=1}^m \frac{(\nu_i - n_i)^2}{\nu_i} = \sum_{i=1}^m \nu_i \left(1 - \frac{n_i}{\nu_i}\right)^2$$

Es lässt sich zeigen, dass τ für grosse m annähernd χ^2 verteilt ist mit $m - 1$ Freiheitsgraden, vgl. z.B. Lit. Kreyszig, Bibl. A10 sowie Bibl. A7 (Cramer).

Entscheidungsfindung: Analoge Überlegungen wie schon bei den obigen Beispielen führen uns zum Entscheid H_0 abzulehnen, sobald der Wert von τ grösser ist als $c = \chi_{\alpha, m-1}^2$ (c = obere Grenze des Integrals mit dem Wert α).

Bsp.: (vgl. Lit Kreyszig, Bibl. A10.)

Gregor Mendel, Kreuzungsversuche mit Erbsen. p_i vermutet nach den „Mendelschen Gesetzen“.

Vermutung: $n_A : n_B : n_C : n_D = 9 : 3 : 3 : 1 \rightsquigarrow$ Resultat:

| | <i>A</i> | <i>B</i> | <i>C</i> | <i>D</i> | $m = 4$ |
|---------|----------------|----------------|----------------|----------------|---------|
| | rund, gelb | rund, grün | kantig, gelb | kantig, grün | Total |
| n_i | 315 | 108 | 101 | 32 | 556 |
| p_i | $\frac{9}{16}$ | $\frac{3}{16}$ | $\frac{3}{16}$ | $\frac{1}{16}$ | 1 |
| $n p_i$ | 312.75 | 104.25 | 104.25 | 34.75 | 556 |

Test der Vermutung mit $\alpha = 0.05$: $n_A : n_B : n_C : n_D = 9 : 3 : 3 : 1$

$$\tau = \frac{(312.75 - 315)^2}{312.75} + \frac{(104.25 - 108)^2}{104.25} + \frac{(104.25 - 101)^2}{104.25} + \frac{(34.75 - 32)^2}{34.75}$$

~ Mit Mathematica:

$$\frac{(312.75 - 315)^2/312.75 + (104.25 - 108)^2/104.25 + (104.25 - 101)^2/104.25 + (34.75 - 32)^2/34.75}{104.25 + (34.75 - 32)^2/34.75}$$

0.470024

```
chi[x_, n_] := 1/(2^(n/2) Gamma[n/2]) x^((n - 2)/2) E^(-x/2);
alpha = 0.05; m = 4; {"c",
  c = c /. FindRoot[
    1 - alpha == Evaluate[Integrate[chi[u, m - 1], {u, 0, c}]], {c,
    8}]] // Chop}

{"c", 7.81471}
```

~ Entscheid: $\tau = 0.470024 < c_{\alpha=0.5} = 7.81471 \rightsquigarrow H_0$ nicht abgelehnt! Problem: τ ist für grosse m annähernd χ^2 verteilt. Und bei uns war $m = 4 \dots$

Bemerkung:

Ein Ereignis A_i könnte es auch sein, dass X in ein gewisses Intervall $[x_i, x_{i+1}]$ fällt. Mit Intervallen ist auch die kontinuierliche Situation vertreten. So lässt sich z.B. **testen, ob eine Funktion normalverteilt ist**.

5.7.4 Zum Fishertest

Auf Seite 115 haben wir gesehen:

Satz:

Vor.:

\mathcal{F} -Verteilung

Beh.:

$$\begin{aligned}
 1. \quad F(x) &= P(\mathcal{F} \leq x) = \frac{\Gamma(\frac{m_1 + m_2}{2})}{\Gamma(\frac{m_1}{2}) \Gamma(\frac{m_2}{2})} \cdot m_1^{\frac{m_1}{2}} \cdot m_2^{\frac{m_2}{2}} \cdot \int_0^x \frac{t^{\frac{m_1-2}{2}}}{(m_1 t + m_2)^{\frac{m_1+m_2}{2}}} dt \\
 2. \quad f(x) &= \begin{cases} 0 & x \leq 0 \\ f(x) = E_{m_1, m_2} x^{\frac{m_1}{2}-1} (m_2 + m_1 \cdot x)^{-\frac{m_1+m_2}{2}} & x > 0 \end{cases}
 \end{aligned}$$

Nun gilt (Beweise vgl. Lit. Kreyszig, Bibl. A10):

Satz:

Vor.:

Zufallsvariablen $V_1, V_2 \in \{\heartsuit\}$ sowie $V_1 \in \chi^2_{m_1}$, $V_2 \in \chi^2_{m_2}$

Beh.:

$V = \frac{V_1/m_1}{V_2/m_2}$ besitzt eine \mathcal{F} -Verteilung

$$\begin{aligned}
 x \leq 0 \Rightarrow F(x) &= 0, \quad 0 < x \Rightarrow F(x) = P(V \leq x) = \\
 &= \frac{\Gamma(\frac{m_1 + m_2}{2})}{\Gamma(\frac{m_1}{2}) \Gamma(\frac{m_2}{2})} \cdot m_1^{\frac{m_1}{2}} \cdot m_2^{\frac{m_2}{2}} \cdot \int_0^x \frac{t^{\frac{m_1-2}{2}}}{(m_1 t + m_2)^{\frac{m_1+m_2}{2}}} dt
 \end{aligned}$$

Satz:

Vor.:

Zufallsvariablen

$X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2} \in \{\heartsuit\}$ sowie $X_i \in N(\mu_1, \sigma_1^2)$, $Y_i \in N(\mu_2, \sigma_2^2)$
 S_1^2, S_2^2 Schätzer für σ_1^2, σ_2^2 (Empirische Varianzen)

Beh.:

$V = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ besitzt eine \mathcal{F} -Verteilung mit $(n_1 - 1, n_2 - 1)$ Freiheitsgraden

Konsequenz:

Bei $H_0 = (\sigma_1^2 = \sigma_2^2)$ wird $V = \frac{S_1^2}{S_2^2}$. Damit hat man ein Mittel um die Gleichheit der Varianzen zu testen!

5.7.5 Kontingenztafeln

Mit dem χ^2 -Test kann man auch die **statistische Unabhängigkeit** zweier Zufallsvariablen X und Y testen. Wir betrachten dazu ein Beispiel einer Kontingenztafel. Darin sind Kunden, die entweder das Produkt A , B oder C kaufen, in Altersklassen eingeteilt. Aufgeführt sind die Häufigkeiten (empirische Werte):

| n_{ik} | $Y = A$ | $Y = B$ | $Y = C$ | $n_{i \cdot}$ |
|------------------|---------------------|----------------------|---------------------|----------------------|
| $X \in [16, 30)$ | $n_{11} = 132$ | $n_{12} = 778$ | $n_{13} = 592$ | $n_{1 \cdot} = 1502$ |
| $X \in [30, 55)$ | $n_{21} = 55$ | $n_{22} = 304$ | $n_{23} = 248$ | $n_{2 \cdot} = 607$ |
| $X \in [55, 85)$ | $n_{31} = 25$ | $n_{32} = 111$ | $n_{33} = 155$ | $n_{3 \cdot} = 291$ |
| $n_{\cdot k}$ | $n_{\cdot 1} = 212$ | $n_{\cdot 2} = 1193$ | $n_{\cdot 3} = 995$ | $n = 2400$ |

$n_{\cdot k}, n_{i \cdot} \rightsquigarrow$ Randhäufigkeiten

Problem: Sind X und Y statistisch unabhängig?

Idee:

Die empirischen Werte $n_{ik}, n_{i \cdot}, n_{\cdot k}$ betrachten wir als Realisationen von zu erwartenden theoretischen Besetzungszahlen $n \cdot p_{ik}, n \cdot p_{i \cdot}, n \cdot p_{\cdot k}$. Die gestellte Frage können wir nun in eine Nullhypothese fassen. Statistische Unabhängigkeit bedeutet, dass für die Wahrscheinlichkeiten gelten muss: $p_{ik} = p_{i \cdot} \cdot p_{\cdot k}$.

$\rightsquigarrow H_0 = (p_{ik} = p_{i \cdot} \cdot p_{\cdot k})$
 $\frac{n_{ik}}{n}$ ist ein Schätzwert für p_{ik} , $\frac{n_{i \cdot}}{n}$ ist ein Schätzwert für $p_{i \cdot}$, $\frac{n_{\cdot k}}{n}$ ist ein Schätzwert für $p_{\cdot k}$. Unabhängigkeit müsste also bedeuten:

$$\rightsquigarrow H_0 = (p_{ik} = p_{i \cdot} \cdot p_{\cdot k}) \approx H_0 = \left(\frac{n_{ik}}{n} = \frac{n_{i \cdot}}{n} \cdot \frac{n_{\cdot k}}{n} \right) = H_0 = \left(n_{ik} = \frac{n_{i \cdot} \cdot n_{\cdot k}}{n} \right)$$

Weiter kann man zeigen (vgl. Lit.):

Satz:

Die folgende Testvariable ist approximativ $\chi^2_{(r-1) \cdot (s-1)}$ -verteilt:

$$T = \sum_{i=1}^r \sum_{k=1}^s \frac{\left(\frac{n_{i \cdot} \cdot n_{\cdot k}}{n} - n_{ik} \right)^2}{\frac{n_{i \cdot} \cdot n_{\cdot k}}{n}}$$

Mittels der Randhäufigkeiten berechnen wir nun die geschätzten erwarteten Besetzungszahlen:

$$H_0 \rightsquigarrow n \cdot p_{ik} \approx \frac{n_{i \cdot} \cdot n_{\cdot k}}{n} = u_{ik}$$

| u_{ik} | $Y = A$ | $Y = B$ | $Y = C$ |
|------------------|---------|---------|---------|
| $X \in [16, 30)$ | 132.677 | 746.619 | 622.704 |
| $X \in [30, 55)$ | 53.6183 | 301.73 | 251.652 |
| $X \in [55, 85)$ | 25.705 | 144.651 | 120.644 |

$$\text{Es gilt: } T = \sum_{i=1}^r \sum_{k=1}^s \frac{\left(\frac{n_{i \cdot} \cdot n_{\cdot k}}{n} - n_{ik} \right)^2}{\frac{n_{i \cdot} \cdot n_{\cdot k}}{n}} = \sum_{i=1}^r \sum_{k=1}^s \frac{(u_{ik} - n_{ik})^2}{u_{ik}} \Rightarrow T = t = \dots$$

Sei $\alpha = 0.01$

\rightsquigarrow **Mit Mathematica:**

```

v1 = {{212, 1193, 995}}; v2 = {{1502, 607, 291}};
matrV = {{132, 778, 592}, {55, 304, 248}, {25, 111, 155}};
matrU = Transpose[v2].v1/2400 //
N; matrV = {{132, 778, 592}, {55, 304, 248}, {25, 111, 155}};
t={1, 1, 1}.({1, 1, 1}.((matrU - matrV)^2/2400)) // Evaluate

```

20.5737

```
chi[x_, n_] := 1/(2^(n/2) Gamma[n/2]) x^((n - 2)/2)E^(-x/2);
alpha = 0.01; r = 3; s = 3; m = (r - 1)(s - 1);
{"c", c = c /. FindRoot[
  1 - alpha == Evaluate[Integrate[chi[u, m], {u, 0, c}]], {c, 8}] //}
  Chop}

{"c", 13.2766}
```

~ Entscheid: $t = 20.5737 < ? c_{\alpha=0.1} = 13.2766$

~ H_0 abgelehnt! Problem: T ist für grosse r, s annähernd χ^2 verteilt. Und bei uns war $r = s = 4 \dots$

Kapitel 6

Fehlerrechnung, Regression, Korrelation — Calcul de l'erreur, régression, corrélation

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

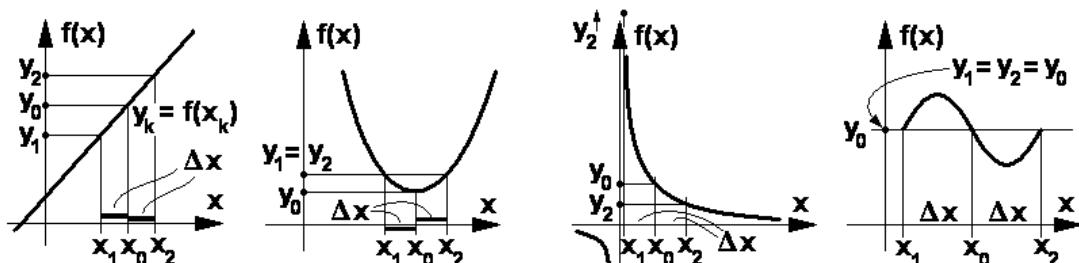
6.1 Fehlerrechnung (Abhängigkeit) — Calcul de l'erreur (dépendance)

Unter dem Begriff „Fehlerrechnung“ werden in der Literatur zwei verschiedene Problemtypen behandelt: Erstens das Problem der Verpflanzung von Messfehlern bei der Anwendung von Funktionen auf fehlerbehaftete Messgrößen. Und zweitens das Problem der Fehler von statistischen Kenngrößen (z.B. Mittelwert), welche meistens von einer grossen Datenmenge abhängen. Hier wollen wir den ersten Fehler- typ behandeln. Der zweite wird im getrennt herausgegebenen Anhang zu diesem Skript besprochen.

6.1.1 Das Problem der Verpflanzung — Le problème de la dépendance

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

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



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

Situation: (b): $y = f(x_{1_0}, x_{2_0}, \dots, x_{n_0})$

Gemessen werden die Werte $x_{1_0}, x_{2_0}, \dots, x_{n_0}$ der Variablen x_1, x_2, \dots, x_n . Bei kontinuierlichen Messwerten gibt es immer Ableseungenaugkeiten, die aber abschätzbar sind. Diese zugehörigen „Messfehler“ betragen $\Delta x_1, \Delta x_2, \dots, \Delta x_n$. Die „exakten Werte“ x_k^* , $k = 1, \dots, n$ liegen daher in den Intervallen

$[x_{k_0} - \Delta x_k, x_{k_0} + \Delta x_k]$. Zudem sei eine Funktion $f(x_1, x_2, \dots, x_n)$ gegeben, mit deren Hilfe eine weitere Grösse berechnet werden muss.

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

6.1.2 Verwendung des totalen Differentials — Appliquer la différentielle totale

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

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

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

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

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

Satz:

Vor.:

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

Beh.:

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

Konsequenz:

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

Definition:

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

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

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

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

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

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

5. Beispiel:

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

Bemerkung:

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

6. Beispiel:

Messwerte:

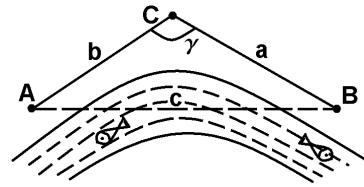
$$a = 364.76 \pm 0.05 \text{ m}$$

$$b = 402.35 \pm 0.05 \text{ m}$$

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

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

$$c = ?$$



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

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

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

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

6.2 Regression — Régression

6.2.1 Der Begriff — La notion

Geg.:

Stichprobe von Beobachtungen $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

x_k z.B. Maschineneinstellung,

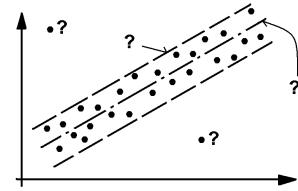
$y_k = f(x_k)$ Messung .

Problem: Interpretation des Verhaltens der Messwerte (Gesetz)?

Z.B. Vermutung: Die Messwerte liegen auf einer passenden Geraden oder sonstigen Kurve.

Problem:

Wie findet man die „beste“ Gerade (resp. Kurve)?



\rightsquigarrow **Probleme:**

1. Berechnung der Parameter der hypothetischen Kurve.

2. Beurteilung der Zuverlässigkeit der getroffenen Wahl.

Die so gefundene Kurve trägt den etwas unverständlichen Namen **Regressionskurve**.

Wieso dieser Name? Der Name stammt aus einer Beschreibung von Beobachtungen von F. Galton²¹ zum Größenwachstum von Menschen. Galton hat festgestellt:

²¹ Engl. Naturforscher, 1822 – 1911

1 Grössere Väter haben grössere Söhne.

2 Jedoch beobachtet man die Tendenz, dass grosse Väter grosse Söhne haben, die aber im Mittel kleiner sind als die genannten Väter selbst. Es ist also ein **Rückschritt** \rightsquigarrow **Regress** vorhanden.

Sei x = Grösse der Väter ,

y = Grösse der Söhne .

$y(x) = a x + b$ Gerade

\rightsquigarrow „Regressionsgerade“

Der Begriff *Regressionsgerade* ist also historisch verankert, hat aber inhaltlich nichts mit der Mathematik zu tun.

6.2.2 Methode der kleinsten Quadrate — Méthode des carrés minimaux

Das Problem — Le problème

Problem: Welche Kurve ist die „beste“ Kurve? – Wie ist das Kriterium „beste“ definitorisch am vernünftigsten zu fassen?

Da wir Menschen selbst entscheiden müssen, was unsere „Vernunft“ sein soll, können wir hier „vernünftig“ mit „ökonomisch“ und daher mit „pragmatisch“ und „einfach“ gleichsetzen. Wir wählen daher hier ein pragmatisches Vorgehen. (In der Literatur ist eine etwas fundiertere Begründung üblich, auf der Grundlage des „Maximum-likelihood-Prinzips“.)

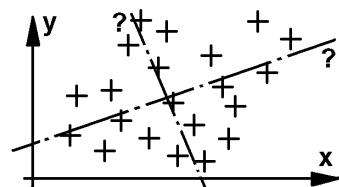
Wichtig: Dass die gesuchte Kurve eine Gerade oder sonst irgend eine Kurve ist, lässt sich theoretisch nicht exakt entscheiden. Man kann nur zu Aussagen kommen wie „die eine Kurve ist wahrscheinlicher als die andere“. Denn dass die gefundenen Messwerte überhaupt auf einer stetigen Kurve liegen, beruht auf einer Arbeitshypothese, die wir nach Descartes mit dem Argument der Erfahrung und der Arbeitsökonomie begründen.

Die Begründung der Methode — La déduction de la méthode

Problem: Durch eine „Punktwolke“ von Messwerten soll eine „beste“ Gerade gelegt werden. Wie ist vorzugehen?

Idee 1:

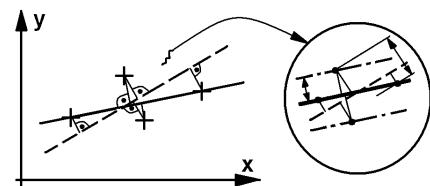
Versuche g so festzulegen, dass die Summe der Abstände zu g gleich 0 wird.



Problem: Diese Methode funktioniert nicht, da offensichtlich (Skizze) mehrere passende „beste“ Geraden möglich sind.

Idee 2:

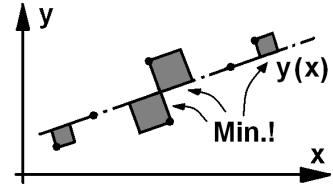
Lege die Gerade so, dass die Summe der Beträge der Abstände minimal wird.



Problem: Diese Methode funktioniert praktisch schlecht, da beim Berechnen der Abstände im \mathbb{R}^2 Wurzeln vorkommen, was die Rechnung sehr kompliziert.

Idee 3:

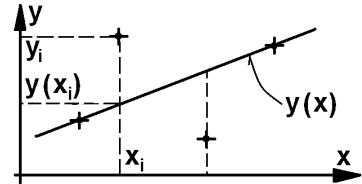
Nimm statt der Summe der Beträge der Abstände die Summe der Quadrate der Abstände zu g . Dadurch fallen bei der Rechnung die Quadratwurzeln weg.



Problem: Praktisch ist zu einem gegebenen Wert x_i der Messwert y_i gegeben $\rightsquigarrow P_i$. Die Gerade soll so bestimmt werden, dass für die nächstgelegenen Punkte $P_i^* \in g$ die Summe der Distanzquadrate $|\overline{P_i P_i^*}|^2$ minimal ist. Man hat das Problem der Minimalisierung unter der Bedingung, dass $\overline{P_i P_i^*} \perp g$ gilt, was die Rechnung wieder sehr kompliziert. Einfacher wird es, hier einen Fehler in Kauf zu nehmen und $x_i^* = x_i$ zu setzen.

Idee 4:

Nimm statt Summe der Quadrate der Abstände zu g nur die Summe der Δy_i .



Problem: Unter der Hypothese, dass die Kurve eine Gerade $y = g(x) = a \cdot x + b$ ist, gilt es nun, a und b zu berechnen. Analog geht man bei andern Kurven vor, z.B. $y = h(x) = a \sin(bx + c)$.

Lösung:

$$(1) \sum_{i=1}^n (\Delta y_i)^2 = \sum_{i=1}^n (y_i - g(x_i))^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2 := f(a, b) \rightarrow \text{Min}$$

$$(2) \sum_{i=1}^n (\Delta y_i)^2 = \sum_{i=1}^n (y_i - h(x_i))^2 = \sum_{i=1}^n (y_i - (a \sin(bx_i + c)))^2 := f(a, b) \rightarrow \text{Min}$$

$$\text{Bedingung: } \frac{\partial f}{\partial a} = \frac{\partial f}{\partial b} = 0 \quad \rightsquigarrow$$

$$\begin{aligned} \frac{\partial f}{\partial a} &= \sum_{i=1}^n 2(y_i - ax_i - b) \cdot (-x_i) = -2 \sum_{i=1}^n (y_i - ax_i - b) \cdot x_i = -2 \sum_{i=1}^n y_i \cdot x_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i \\ &= -2 \sum_{i=1}^n y_i \cdot x_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial f}{\partial b} &= \sum_{i=1}^n 2(y_i - ax_i - b) \cdot (-1) = -2 \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - b = 0 \\ &\Rightarrow \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - b = \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - nb = 0 \end{aligned}$$

$$\text{Sei } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ (Mitterwert der } x_i \text{), } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\Rightarrow \begin{aligned} \sum_{i=1}^n x_i y_i &= a \sum_{i=1}^n x_i^2 + b n \bar{x} \\ \Rightarrow \cancel{a} \bar{y} &= \cancel{a} \bar{x} + \cancel{b} \Rightarrow y(\bar{x}) = a \bar{x} + b = \bar{y} \end{aligned}$$

$$b = \bar{y} - a \bar{x} \text{ einsetzen: } (\Rightarrow \bar{y} = a \bar{x} + b)$$

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i^2 + (\bar{y} - a \bar{x}) n \bar{x} = a \left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) + n \bar{x} \bar{y}$$

Hinweis:

$$\text{Sei } \vec{d}_e := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} := \vec{1}, \quad \vec{x} := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \vec{y} := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\Rightarrow \sum_{i=1}^n x_i y_i = \langle \vec{x}, \vec{y} \rangle, \quad \sum_{i=1}^n x_i = \langle \vec{x}, \vec{d}_1 \rangle, \quad \sum_{i=1}^n y_i = \langle \vec{y}, \vec{d}_1 \rangle \quad \rightsquigarrow$$

Formel:

Vor.:

Sei

$y = a \cdot x + b$ Regressionsgerade g

Beh.:

$$a = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\langle \vec{x}, \vec{y} \rangle - n \bar{x} \bar{y}}{\langle \vec{x}, \vec{x} \rangle - n \bar{x}^2}$$

$$b = \bar{y} - \bar{x} \cdot \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \bar{y} - \bar{x} \cdot \frac{\langle \vec{x}, \vec{y} \rangle - n \bar{x} \bar{y}}{\langle \vec{x}, \vec{x} \rangle - n \bar{x}^2}$$

$$y(\bar{x}) = a \bar{x} + b = \bar{y} \Rightarrow (\bar{x}, \bar{y}) \in g$$

Schreibweise mit Varianz und Kovarianz — Façon d'écrire à l'aide de la variance et de la covariance

Betrachtung:

$$s_x^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \right) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - 2 \bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \right)$$

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

$$= \frac{1}{n-1} (\langle \vec{x}, \vec{x} \rangle - n \bar{x}^2)$$

In der mathematischen Statistik ist es üblich, die **Varianzen** s_x^2 , s_y^2 und die **Kovarianz** s_{xy} wie folgt zu definieren:

Definition:

$$s_x^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) = \frac{1}{n-1} (\langle \vec{x}, \vec{x} \rangle - n \bar{x}^2)$$

$$s_y^2 := \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \right) = \frac{1}{n-1} (\langle \vec{y}, \vec{y} \rangle - n \bar{y}^2)$$

$$s_{xy} := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{n-1} \left(\left(\sum_{i=1}^n x_i y_i \right) - n \bar{x} \bar{y} \right)$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right) \right) = \frac{1}{n-1} (\langle \vec{x}, \vec{y} \rangle - n \bar{x} \bar{y})$$

Es gilt: $\bar{x} = \frac{1}{n} \langle \vec{x}, \vec{1} \rangle$, $\bar{y} = \frac{1}{n} \langle \vec{y}, \vec{1} \rangle$.

Mit Hilfe der Varianz und der Kovarianz lässt sich die Regressionsgerade einfacher schreiben. Dabei benutzen wir:

Lemma: $a = \frac{s_{xy}}{s_x^2}$, $b = \bar{y} - \bar{x} \cdot \frac{s_{xy}}{s_x^2}$

(Durch ausmultiplizieren verifiziert man leicht: $s_{xy} = s_x^2 \cdot a$)

Korollar:

Vor.:

Sei

$y = a \cdot x + b$ Regressionsgerade

Beh.:

$$y = \frac{s_{xy}}{s_x^2} \cdot x + \bar{y} - \bar{x} \cdot \frac{s_{xy}}{s_x^2}$$

Bemerkung:

Für die Summe der quadratischen Abstände gilt:

$$\begin{aligned} \sum_{i=1}^n (\Delta y_i)^2 &= \sum_{i=1}^n (y_i - a x_i - b)^2 = \sum_{i=1}^n (y_i - \bar{y} - a (x_i - \bar{x}))^2 \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2 a \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) + a^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= (n-1) \cdot (s_y^2 - 2 a s_{xy} + a^2 s_x^2) = (n-1) \cdot (s_y^2 - 2 a a s_x^2 + a^2 s_x^2) = (n-1) \cdot (s_y^2 - a^2 s_x^2) \end{aligned}$$

$$\leadsto \sum_{i=1}^n (\Delta y_i)^2 = 0 \Leftrightarrow s_y^2 = a^2 s_x^2 = \left(\frac{s_{xy}}{s_x^2} \right)^2 s_x^2 = \frac{s_{xy}^2}{s_x^2} \Leftrightarrow (s_x \cdot s_y)^2 = s_{xy}^2$$

Satz:

$$\sum_{i=1}^n (\Delta y_i)^2 = 0 \Leftrightarrow (s_x \cdot s_y)^2 = s_{xy}^2$$

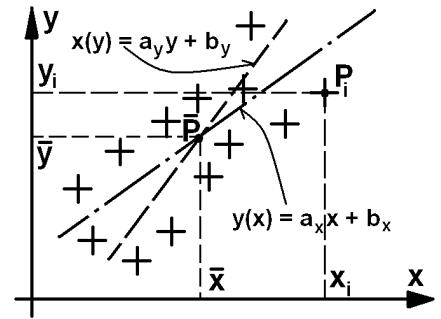
6.2.3 Korrelation — Corrélation

Vorhin haben wir die Gerade $g_x : y(x) = a \cdot x + b$ untersucht.

Da y die abhängige und x die unabhängige Variable war, schreiben wir präziser:

$$\begin{aligned} y(x) &= a_x \cdot x + b_x, \\ a_x &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{s_{xy}}{s_x^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

Wir vertauschen nun alle Wertepaare (x_i, y_i) . D.h. wir betrachten y als die abhängige und x als die unabhängige Variable. Dann lässt sich mit der Methode der kleinsten Quadrate wieder die „beste“ Gerade $g_y : x(y) = a_y \cdot y + b_y$ durch die Messpunkte finden.



a_y und b_y erhält man aus a_x und b_x durch Rollenvertauschung der x_i, \bar{x} und y_i, \bar{y} :

$$a_y = \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n y_i^2 - n \bar{y}^2} = \frac{s_{yx}}{s_y^2} = \frac{\sum_{i=1}^n (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

$$b_y = \bar{x} - \bar{y} \cdot \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n y_i^2 - n \bar{y}^2}$$

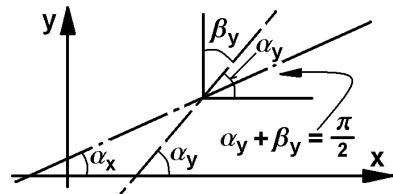
Wenn man g_x und g_y in dasselbe Koordinatensystem einzeichnet, ist zu erwarten, dass zwei verschiedene Geraden entstehen.

Ideal wäre allerdings, wenn g_x und g_y zusammenfallen würden. Dann hätte man:

$$a_x = \tan(\alpha_x),$$

$$a_y = \tan(\beta_y) = \tan\left(\frac{\pi}{2} - \alpha_x\right) = \frac{1}{\tan(\alpha_x)}$$

$$\Rightarrow a_x \cdot a_y = \tan(\alpha_x) \cdot \frac{1}{\tan(\alpha_x)} = 1$$

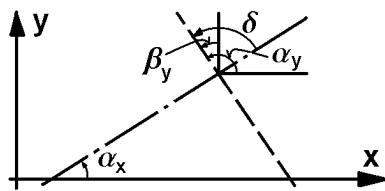


Andernfalls ist:

$$a_x \cdot a_y = \tan(\alpha_x) \cdot \tan(\beta_y)$$

$$= \frac{\tan(\alpha_x)}{\tan\left(\frac{\pi}{2} - \beta_y\right)} = \frac{\tan(\alpha_x)}{\tan(\alpha_y)}$$

$$= \frac{\tan(\alpha_x)}{\tan(\alpha_x + \delta)} \neq 1$$



$a_x \cdot a_y = \frac{\tan(\alpha_x)}{\tan(\alpha_x + \delta)}$ ist umso verschiedener von 1, je verschiedener δ von 0 ist. \rightsquigarrow
 $a_x \cdot a_y$ ist ein Mass für den Zusammenhang der beiden Geraden, d.h. für die **Korrelation**.

Es gilt: $(s_x, s_y \geq 0)$

$$a_x \cdot a_y = \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \cdot \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n y_i^2 - n \bar{y}^2} = \frac{\left(\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2\right) \cdot \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2\right)},$$

$$\sqrt{a_x \cdot a_y \cdot \text{sgn}(a_x \cdot a_y)} = \frac{|\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}|}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2) \cdot (\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} = \frac{|s_{xy}|}{\sqrt{s_x^2 \cdot s_y^2}} := |r_{xy}|$$

Definition: $r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}}$ heisst **Korrelationskoeffizient**

6.2.4 Korrelationskoeff.: Bedeutung — Coeff. de corrélation: Signification

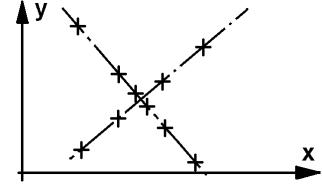
Untersuchung:

$$\begin{aligned} \rightsquigarrow r_{xy} = 0 \Leftrightarrow 0 = s_{xy} := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{n-1} \sum_{i=1}^n (\Delta x_i) \cdot (\Delta y_i) = \\ \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right) \Leftrightarrow \frac{1}{n} \cdot \sum_{i=1}^n x_i y_i = \overline{\langle \vec{x}, \vec{y} \rangle} = \bar{x} \cdot \bar{y}, \vec{x} = (x_1, \dots, x_n)^T, \vec{y} = (y_1, \dots, y_n)^T \end{aligned}$$

Konsequenz:

$$r_{xy} = 0 \Leftrightarrow \langle \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix}, \begin{pmatrix} \Delta y_1 \\ \vdots \\ \Delta y_n \end{pmatrix} \rangle = 0 \Leftrightarrow \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} \perp \begin{pmatrix} \Delta y_1 \\ \vdots \\ \Delta y_n \end{pmatrix} \Leftrightarrow \overline{\langle \vec{x}, \vec{y} \rangle} = \bar{x} \cdot \bar{y}$$

Diese Situation tritt auf, wenn die Punkte „wolkenartig“ verteilt sind, wie man schon mit vier Punkten sieht, die die Ecken eines achsenparallelen Rechtecks bilden.



Detailliertere Untersuchung:

Seien

$$\Delta \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \bar{x} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} = \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix},$$

Hier ist:

$$\varphi = \varphi(\Delta \vec{x}, \Delta \vec{y}).$$

$$\Delta \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \bar{y} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \Delta y_1 \\ \vdots \\ \Delta y_n \end{pmatrix}$$

Es gilt:

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{\frac{1}{n-1} \cdot \langle \Delta \vec{x}, \Delta \vec{y} \rangle}{\sqrt{\frac{1}{n-1} \cdot \langle \Delta \vec{x}, \Delta \vec{x} \rangle \cdot \frac{1}{n-1} \cdot \langle \Delta \vec{y}, \Delta \vec{y} \rangle}} = \frac{\frac{1}{n-1} \cdot |\Delta \vec{x}| \cdot |\Delta \vec{y}| \cdot \cos(\varphi)}{\frac{1}{n-1} \cdot |\Delta \vec{x}| \cdot |\Delta \vec{y}|} = \cos(\varphi)$$

$$r_{xy} = 0 \Leftrightarrow \langle \Delta \vec{x}, \Delta \vec{y} \rangle = 0 \Leftrightarrow \Delta \vec{x} \perp \Delta \vec{y} \Leftrightarrow \cos(\varphi) = 0 \Leftrightarrow \varphi = \frac{\pi}{2} + n \cdot \pi, n \in \mathbb{Z}$$

$$r_{xy} = 1 \Leftrightarrow \langle \Delta \vec{x}, \Delta \vec{y} \rangle = |\Delta \vec{x}| \cdot |\Delta \vec{y}| \Leftrightarrow \Delta \vec{x} \parallel \Delta \vec{y} \Leftrightarrow |\cos(\varphi)| = 1 \Leftrightarrow \varphi = n \cdot \pi, n \in \mathbb{Z}$$

Korollar:

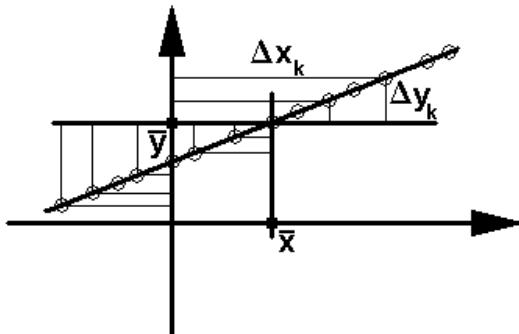
Vor.:

$$\varphi = \varphi(\Delta\vec{x}, \Delta\vec{y}), \quad \Delta\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \bar{x} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \Delta\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \bar{y} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

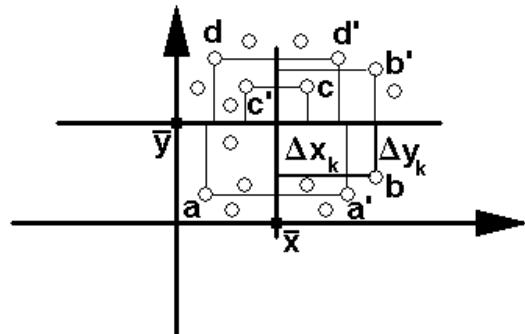
Beh.:

$$r_{xy} = 0 \Leftrightarrow \varphi = \frac{\pi}{2} + n \cdot \pi, \quad n \in \mathbb{Z}$$

$$|r_{xy}| = 1 \Leftrightarrow \varphi = n \cdot \pi, \quad n \in \mathbb{Z}$$



$$\varphi = n \cdot \pi, \quad n \in \mathbb{Z}$$



$$\varphi = \frac{\pi}{2} + n \cdot \pi, \quad n \in \mathbb{Z}$$

Bsp.:

1. $x = \{2, 3, 4, 6, 8, 12\}, \quad y = \{4, 6, 8, 12, 16, 24\} \quad (y = 2x) \Rightarrow r = 1$
2. $x = \{-2.1, -2.1, 0, 0, 2, 2.1\}, \quad y = \{4, -4, 3, -3, 4, -4\} \Rightarrow r = -0.0106422$
3. $x = \{-2.1, -2, 2, 2.1\}, \quad y = \{4, -4, 4, -4\} \Rightarrow r = -0.024383$

6.2.5 Rechnen mit Punktschätzern: Probleme — Calculer avec des estimateurs: Problèmes

Sei μ_X = Erwartungswert einer Messgrösse X , $Y := f(X)$.

Sei μ_X geschätzt durch eine erwartungstreue Punkt-Schätzung $\hat{\mu}_X$.

$$\begin{aligned} & \Rightarrow Y = \mu_Y + \Delta Y \approx f(\mu_X) + f'(\mu_X)(X - \mu_X) \quad (\text{Taylor}). \\ \rightsquigarrow \bar{y} & \approx \bar{\mu}_Y = \mu_Y \approx \bar{f}(\mu_X) + \overline{f'(\mu_X)(X - \mu_X)} = f(\mu_X) + \bar{f}'(\mu_X)(\bar{X} - \bar{\mu}_X) = f(\mu_X) + \underbrace{f'(\mu_X)(\bar{X} - \mu_X)}_{=0} \\ & \Rightarrow \bar{y} \approx \mu_Y \approx f(\mu_X) \end{aligned}$$

Satz:

$$Y = f(X) \Rightarrow \bar{y} \approx \mu_Y \approx f(\mu_X)$$

Verbesserung der Approximation:

$$Y = \mu_Y + \Delta Y \approx f(\mu_X) + f'(\mu_X)(X - \mu_X) + \frac{1}{2} f''(\mu_X)(X - \mu_X)^2$$

$$\longrightarrow \mu_Y \approx f(\mu_X) + f'(\mu_X) \cdot 0 + \frac{1}{2} f''(\mu_X) \sigma_X^2 = f(\mu_X) + \frac{1}{2} f''(\mu_X) \sigma_X^2$$

Satz: $\bar{x} \approx \mu_X, \bar{y} \approx \mu_Y, Y = f(X) \Rightarrow \bar{y} \approx f(\mu_X) + \frac{1}{2} f''(\mu_X) \sigma_X^2$

Anwendung für Vertrauensintervalle:

Sei $\mu_X \in [\bar{x} - \Delta x, \bar{x} + \Delta x], \Delta x \approx \sigma_X$ (kurz $\mu_X = \bar{x} \pm \Delta x$).

→ Problem: $\sigma_Y = ?$

$$Y = f(X), Y \approx f(\mu_X) + f'(\mu_X)(X - \mu_X) \Rightarrow Y - f(\mu_X) \approx f'(\mu_X)(X - \mu_X)$$

$$\Rightarrow (Y - f(\mu_X))^2 \approx (f'(\mu_X))^2 (X - \mu_X)^2 \Rightarrow \sigma_Y^2 \approx (f'(\mu_X))^2 \sigma_X^2.$$

Satz: $\sigma_Y^2 \approx (f'(\mu_X))^2 \sigma_X^2$

Die hier angefangenen Ausführungen sollen in den Übungen weiter verfolgt werden... Weitere Ausführungen finden sich unter dem nachstehend angegebenen Link:

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

6.3 Zum Schluss

Damit sind die ersten Grundlagen der Statistik besprochen. In der Praxis kann man damit allerdings noch nicht viel anfangen. Von Varianzanalyse, Zeitreihenanalyse und dergleichen haben wir noch nicht gesprochen. Viele Tests sind nicht erwähnt...

Problem:

Die Normalverteilung existiert in der Realität praktisch nie, da reale Intervalle praktisch nie unendlich gross und oft auch nur positiv sind. Viele Testvariablen sind nur Approximationen. Seriöse Arbeit benötigt ein grosses Datenmaterial... Da hilft die Fachliteratur weiter.

Literaturverzeichnis

- [1] Fachlexikon a b c. Verlag Harri Deutsch, Bibliographisches Institut Mannheim, Wien, Zürich. Dudenverlag (Bibl.: A1)
- [2] Herbert Amann, Joachim Escher, Analysis I, II, Birkhäuser (Bibl.: A2)
- [3] Heinz Bauer, Wahrscheinlichkeitstheorie und Grundzüge der Massstheorie, De Gruyter (Bibl.: A3)
- [4] Bosch K., Elementare Einführung in die Wahrscheinlichkeitsrechnung, Vieweg (Bibl.: A4)
- [5] Brenner, Lesky. Mathematik für Ingenieure und Naturwissenschaftler. AULA-Verlag Wiesbaden (Bibl.: A5)
- [6] Claus, Schwill. Schüler–Duden, Die Informatik. Bibliographisches Institut Mannheim, Wien, Zürich. Dudenverlag (Bibl.: A6)
- [7] Cramer, Mathematical Methods of Statistics, Princeton (Bibl.: A7)
- [8] Iyanaga, Kawada. Encyclopedic Dictionary of Mathematics. MIT Press, Cambridge Mass., London GB (Bibl.: A8)
- [9] Erwin Kreyszig, Statistische Methoden und ihre Anwendungen, Vandenhoeck & Ruprecht Göttingen 1965 / 72 (Bibl.: A9)
- [10] Erwin Kreyszig, Statistische Methoden und ihre Anwendungen, Vandenhoeck und Ruprecht, Neuauflage (Bibl.: A10)
- [11] Meschkowski. Mathematisches Begriffswörterbuch. BI Hochschultaschenbücher. Bibliographisches Institut Mannheim. (Bibl.: A11)
- [12] Regina Storm, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Statistische Qualitätskontrolle, Fachbuchverlag Leibzig, Köln (Bibl.: A12)
- [13] L.B. van der Waerden, Mathematische Statistik, Birkhäuser (Bibl.: A13)
- [14] Mathematikkurs für Ingenieure
Teil 4 ◇ Einführung in die Boolsche Algebra, Rolf Wirz, Biel 1999, (Bibl.: A14)

<http://rowicus.ch/Wir//Scripts/Teil4Bool.pdf>

- [15] Script ◇ Math ◇ Ing, ◇ Analysis ◇ Analyse ◇, kurz & bündig ◇ concis, Rolf Wirz, Biel 1999, (Bibl.: A15)

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

- [16] Script ◇ Math ◇ Ing, ◇ Fortsetzung Mathematik ◇ Suite mathématiques ◇, kurz & bündig ◇ concis, Rolf Wirz, Biel 2000/01,... (Bibl.: A16)

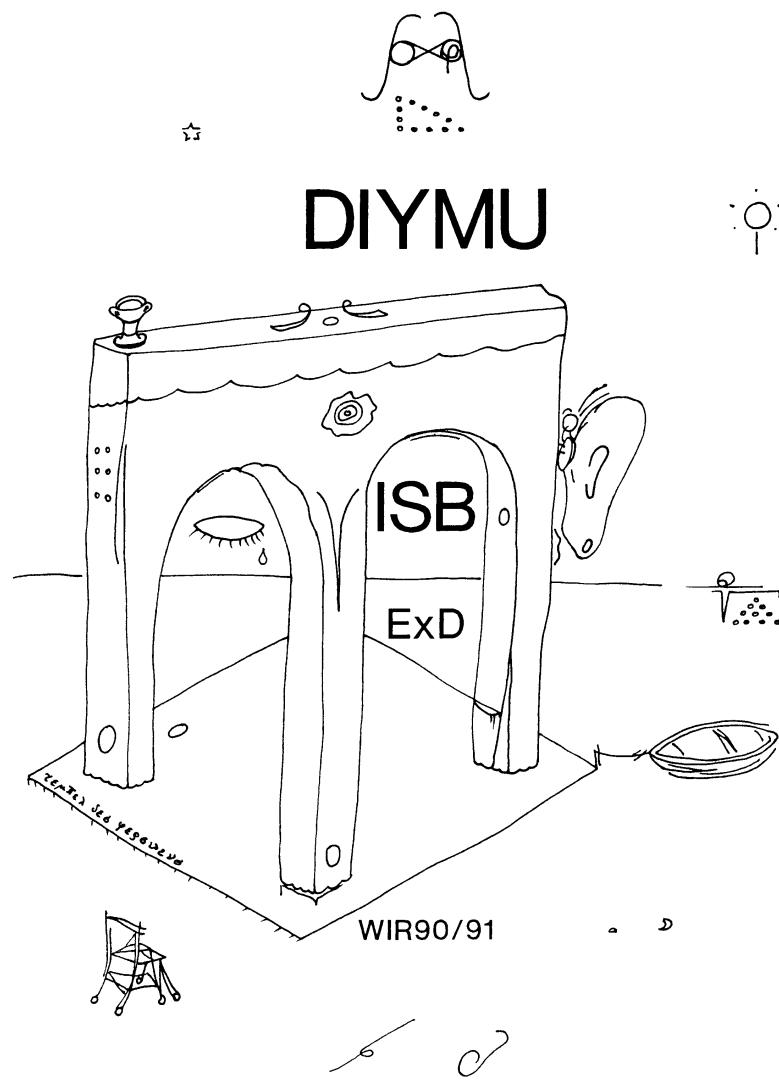
<http://rowicus.ch/Wir//Scripts/KursMathPubl.pdf>

- [17] Vom Autor. Mathematik für Ingenieure *Teile 1 ff* (Bibl.: A17)

- [18] Vom Autor. *DIYMU* (Do it yourself Mathematik Übungsbuch). Ingenieurschule Biel 1991 (Bibl.: A18)

Anhang A

Aus dem DIYMU



Übungen

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

Anhang B

Fehler von statistischen Kenngrössen und Standardfehler

Siehe auf

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

Anhang C

Monte-Carlo, Resampling und anderes

Siehe auf

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

Anhang D

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

Siehe auf

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

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

(Print: http://rowicus.ch/Wir/Scripts/KursWahrschStatistdf_Print.pdf)

Anhang E

Datensatzänderung

Siehe auf

<http://rowicus.ch/Wir/Scripts/Zusatz/AnhangStatistDatenkorrektur.pdf>

oder

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

Anhang F

Spezielle Wahrscheinlichkeitssituationen

Siehe auf

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

Anhang G

Hinweise zur Datenanalyse

Siehe auf

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

Anhang H

Crashkurs Kombinatorik, Wahrscheinlichkeit: Link

Neuere Darstellung siehe:

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

Ende *Fin*