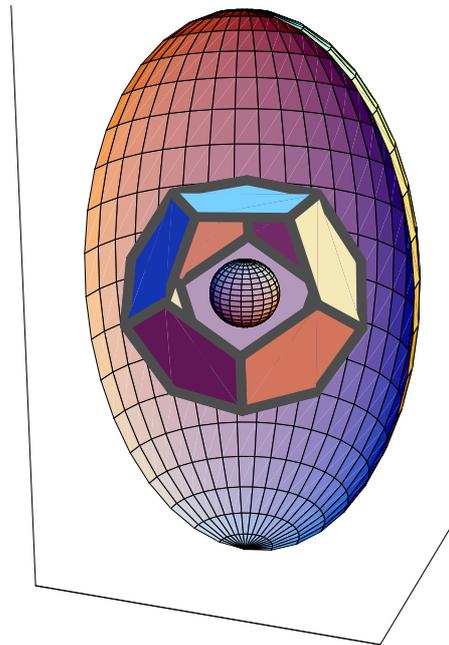


Script \diamond Math \diamond Ing
 \diamond Probabilité & Statistique \diamond
 \diamond concis



Scripta bilingua

von de

Rolf Wirz

Berner Fachhochschule BFH \diamond TI und et AHB

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

Contient partie 6 (analyse combinatoire) d'un cours de répétition et livret à l'accompagnement et complément de la leçon.
Produit avec LaTeX sur NeXT-Computer/ PCTeX WIN98.

Quelques graphismes ont été produits avec Mathematica. 1999, l'auteur a été violé d'un crash d'ordinateur. A la suite du changement de système provoqué par cela, quelques graphismes ont souffert tellement. Ils seront aménagés de nouveau dès qu'il y aura assez de temps à disposition pour cela.

Des fois le bonheur aidera, mais le travail aide toujours ...

Sagesse de l'Inde

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

En particulier les illustrations fait à la main sont pris des représentations publiques et plus précoces de l'auteur. Les copyrights pour cela appartient en privé à l'auteur.

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

La version LaTeX actuellement utilisée n'accepte pas un changement du titre du tableau des matières en français – c'est un programme ...

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	5
2.1.4 Arten von Massenerscheinungen — Types d'événements de masse	5
2.1.5 Stichproben — Echantillons	6
2.2 Arbeitsweise der math. Statistik — Façon der travailler d. l. stat. math.	7
2.3 Beschreibende Statistik — Statistique descriptive	8
2.3.1 Fragen — Questions	8
2.3.2 Häufigkeitsverteilungen — Répartition de fréquences	8
2.3.3 Häufigkeitsfunktion — Fonction de fréquences	10
2.3.4 Darstellungstechniken — Techniques de représentation	10
2.3.5 Zur Klassenbildung — Quant à la formation ce classes	14
2.4 Masszahlen einer Stichprobe — Mesures d'un échantillon	15
2.4.1 Mittelwert und empirische Varianz — Moyenne et variance empirique	16
2.4.2 Vereinfachungsmeth. bei Berechnungen — Méthodes de simplif. aux calculs	18
2.4.3 Berechnungen, Häufigkeitsfunktion — Calculs, fonction de fréquence	19
2.4.4 Häufigkeitsverteilung, Massenverteilung — Distr. de fréq. et de masse	20
2.4.5 Beispiel mit Mathematica — Exemple avec Mathematica	20
2.5 Auswertung: Beispiel — Exploitation: Exemple	22
2.5.1 Dateneingabe — Entrée de données	22
2.5.2 Kenngrößen — Caractéristiques	22
2.5.3 Darstellung mittels Kenngrößen — Représentat. par caract.: BoxWhiskerPlot	23
2.5.4 Andere statistische Plots — D'autres représentation statistiques	24
2.6 Weitere Kenngrößen — D'autres caractéristiques	25
2.6.1 Diverse Mittelwerte einer Verteilung — Certaines valeurs moyen. d'une distr.	25
2.6.2 Momente einer Verteilung — Moments d'une distribution	27
2.6.3 Die Schiefe einer Verteilung — Le biais d'une distribution	28
2.6.4 Kurtosis und Exzess — Kurtosis et excès	29
2.6.5 Sinn und Gefahr von Kenngrößen — Caractéristiques: Sens propre et danger	30
3 Kombinatorik — Analyse combinatoire	31
3.1 Einleitung — Introduction	31
3.1.1 Problemstellung — Problème	31
3.1.2 Fakultäten — Factorielles	31
3.2 Anordnungsprobleme — Problèmes d'arrangement	32
3.2.1 Permutationen ohne Wiederholung — Permutations sans répétition	32
3.2.2 Permutationen mit Wiederholung — Permutations avec répétition	35

3.3	Auswahlprobleme — Problèmes de choix	37
3.3.1	Die Fragestellungen — Les questions	37
3.3.2	Variation ohne Wiederholung — Arrangement sans répétition	39
3.3.3	Kombination ohne Wiederholung — Combinaison sans répétition	41
3.3.4	Variation mit Wiederholung — Arrangement avec répétition	42
3.3.5	Kombination mit Wiederholung — Combinaison avec répétition	44
3.4	Übungen — Exercices	45
4	Wahrscheinlichkeitsrechnung — Calcul des probabilités	47
4.1	Einleitung — Introduction	47
4.1.1	Problemstellung — Problème	47
4.1.2	Anwendung — Application	47
4.1.3	Personen — Personnages	48
4.2	Zufallsexperiment, Ereignis — Expérience de hasard, événement	48
4.2.1	Zufallsprozesse, Häufigkeit — Processus aléatoire, fréquence	51
4.3	Ereignisalgebra — Algèbre des événements	52
4.3.1	Ereignisalgebra und Mengenalg. — Alg. des événem. et alg. d. ensembles	52
4.3.2	Boolsche Algebren — Algèbres de Boole	54
4.3.3	Zur Mächtigkeit und Häufigkeit — Quant à la puissance et la fréquence	55
4.3.4	Ereignisbäume — Des arbres d'événements	56
4.4	Klass. Wahrscheinlichkeit n. Laplace — Probabilité class. d'a. Lapl.	56
4.4.1	Laplace-Experiment, Gleichwarsch. — Exp. de Laplace, probab. identique	56
4.4.2	Warsch. als Mass für Gewinnchancen — Prob. comme mesure p. gagner	57
4.5	Axiomatischer Warsch'begriff — Probabilité axiomatique	59
4.5.1	Begriff, Axiome, Folgerungen — Notion, axiomes, conclusions	59
4.5.2	Der Begriff Wahrscheinlichkeitsraum — La notion espace de probabilité	61
4.5.3	Bedingte Wahrscheinlichkeit — Probabilité conditionnelle	62
4.5.4	Totale Wahrscheinlichkeit — Probabilité totale	65
4.6	Wahrscheinlichkeitsverteilungen — Fonctions de répartition	67
4.6.1	Zufallsvariablen — Variables aléatoires	67
4.6.2	Verteilungsfunktion — Fonction de répartition	69
4.6.3	Verteilungstypen — Types de répartition	70
4.6.4	Diskrete Verteilung — Répartition discrète	71
4.6.5	Kontinuierliche Verteilung — Répartition continue	75
4.7	Mass- oder Kennzahlen einer Verteilung — Mesures de répartition	77
4.7.1	Allgemeines — Considérations générales	77
4.7.2	Mittelwert — Valeur moyenne	77
4.7.3	Erwartungswert — Valeur d'espérance	78
4.7.4	Symmetrische Verteilung — Distribution symétrique	79
4.7.5	Varianz, Standardabweichung — Variance, écart-type	80
4.7.6	Momente einer Verteilung — Moments d'une distribution	81
4.7.7	Schiefe einer Verteilung — Dissymétrie d'une distribution	84
4.7.8	Weitere Kenngrößen — Autres caractéristiques	84
4.7.9	Momentenerzeugende Funktion — Fonction char. génér. de moments	86
4.7.10	Laplace- und z-Transformation — Transf. de Laplace et en z	86
4.8	Spezielle diskrete Verteilungen — Distributions discrètes spéciales	87
4.8.1	Bernoulliverteilung — Distribution de Bernoulli	87
4.8.2	Gesetze für die Binomialverteilung — Lois pour la distribution de Bernoulli	88
4.8.3	Poissonverteilung — Distribution de Poisson	90
4.8.4	Pascalverteilung — Distribution de Pascal	91
4.8.5	Geometrische Verteilung — Distribution géométrique	93
4.8.6	Hypergeometrische Verteilung — Distribution hypergéométrique	94
4.8.7	Beispiel — Exemple	95

4.9	Spezielle stetige Verteilungen — Distributions continues spéciales	96
4.9.1	Allgemeines — Généralités	96
4.9.2	Rechtecksverteilung — Distribution rectangulaire	96
4.9.3	Normalverteilung — Distribution normale ou de Gauss	97
4.9.4	Grenzwertsätze von Moivre Laplace — Théorèmes limites de Moivre Laplace	99
4.9.5	Lokaler Grenzwertsatz — Théorème limite locale	99
4.9.6	Grenzwertsatz von De Moivre/ Laplace — Théorème limite de De Moivre/ Laplace	101
4.9.7	Das Gesetz der grossen Zahlen von Bernoulli — La loi des grands nombres	102
4.9.8	Bemerkung zum Zufall — Remarques quant au hasard	102
4.9.9	Tschebyscheffsche Ungleichung — Inéquation de Tschebyscheff	103
4.9.10	Logarithmische Normalverteilung — Distribution normale logarithmique	103
4.9.11	Exponentialverteilung — Distribution exponentielle	104
4.9.12	Weibullverteilung — Distribution Weibull	105
4.9.13	Gammaverteilung — Distribution gamma	106
4.9.14	Ausblick — Autres distributions	106
4.10	Zufallsvektoren — Vecteurs aléatoires	106
4.10.1	Fragestellung, Begriffe — Question, notions	107
4.10.2	Der diskrete Fall — Le cas discret	109
4.10.3	Der stetige Fall — Le cas continu	111
4.11	Mehrdimensionale Erwartung — Espérance multidimensionnelle	112
4.11.1	Erwartung, Mittelwert — Espérance, moyenne	112
4.11.2	Varianz, Kovarianz, Korrelation — Variance, covariance, corrélation	113
4.11.3	Der diskrete Fall — Le cas discret	116
4.11.4	Der stetige Fall — Le cas continu	116
4.12	Mehrdimensionale Verteilungen — Répartitions multidimensionnelles	117
4.12.1	Zweidimensionale Normalverteilung — Distribution normale bidimensionnelle	117
4.12.2	Stichprobenfunkt., Testverteilungen — Fonct. d'échant., distrib. de test	117
4.12.3	Chi-Quadrat-Verteilung — Distribution du Khi-deux	121
4.12.4	Sätze zur Chi-Quadrat-Verteilung — Théorèmes sur la distribution du Khi-deux	123
4.12.5	t-Verteilung von Student — Distribution de Student	123
4.12.6	F-Verteilung von Fisher — Distribution de Fisher	125
4.13	Anhang I: Einige Beweise — Annexe I: Certaines preuves	126
4.13.1	Formel zur Gammafunktion — Formule pour la fonction gamma	126
4.13.2	Dichte der Chi-Quadrat-Verteilung — Densité de la distribution Khi-deux	127
4.13.3	Dichte der Student-Verteilung — Densité de la distribution de Student	128
4.13.4	Beweis Tschebyscheffsche Ungleichung — Preuve d'inéquation de Tschebyscheff	129
4.14	Anhang II: Ergänzungen — Annexe II: Suppléments	130
4.14.1	Quadr'summe contra Betrags. — Somme d. carrés com. à la somme d. val. abs.	130
4.14.2	Verteilungstreue u.s.w. — Conformité du type de répartition etc.	130
4.14.3	Zentraler Grenzwertsatz — Théorème limite central	133
4.14.4	Lineare Transformationen — Transformations linéaires	134
5	Math. Statistik — Statist. mathématique	137
5.1	Qualitätskontrolle — Contrôle de qualité	137
5.1.1	Allgemeines, SQC — Généralités, SQC	137
5.2	SQC1: Prozesskontrolle — SQC1: Contrôle de processus	137
5.2.1	Problemstellung — Problème	137
5.2.2	Beispiel Mittelwert — Exemple moyenne	138
5.2.3	Überw'techn. m. Kontr'karten — Techn. de surv. à l'aide d. cartes de contr.	140
5.3	SQC2: Annahmekontrolle — SQC2: Contrôle d'acceptation	140
5.3.1	Hypergeom. Verteil., Urnenmodell — Répart. hypergéom., modèle d'urne	141
5.3.2	Annahmevertrag, Prüfplan — Contrat d'accept., plan d'échant.	142
5.3.3	Binomialmodell, Urnenmodell — Modèle binomial, modèle d'urne	144

5.3.4	Produzenten- u. Konsumentenrisiko — Risque du producteur e.d. consommateur	145
5.4	Poisson-Prozess, Warteschlange — Files d'attente, processus de Poisson	147
5.4.1	Poisson-Prozess — Processus de Poisson	147
5.4.2	Warteschlangemodelle — Modèles de files d'attente	151
5.5	Schätzungen — Estimations	151
5.5.1	Schätzer, Punktschätzer — Estimations ponctuelles	151
5.5.2	Erwartungstreue — Propriété d'être sans biais	152
5.5.3	Konsistenz — Consistance	154
5.5.4	Vertrauensintervalle I — Intervalles de confiance I	156
5.5.5	Wichtige Testfunktionen — Fonctions de test importantes	158
5.5.6	Vertrauensintervalle II — Intervalles de confiance II	164
5.5.7	Vertrauensintervalle III — Intervalles de confiance III	167
5.5.8	Vertrauensintervalle IV — Intervalles de confiance IV	170
5.5.9	Automat. Bestimm. der Vertrauensber. — Calcul automat. des domaines de conf.	170
5.6	Signifikanztests — Tests de signification	171
5.6.1	Hypothesen — Hypothèses	171
5.6.2	Zweiseitige Alternative, t -Test — Alternative bilatérale, test de Student	171
5.6.3	Einseitige Alternative, t -Test — Alternative unilatérale, test de Student	173
5.6.4	Testrisiken — Risques (aléas) aux tests	175
5.6.5	Chi-quadrat-Test für die Varianz — Test khi deux pour la variance	176
5.6.6	Automatisches testen — Tester automatiquement	177
5.6.7	Testrezept im Überblick — Vue d'ensemble d'une recette de test	179
5.6.8	Vergleichstest für Mittelwerte I — Test de compar. de moyennes I	179
5.6.9	Vergleichstest für Mittelwerte II — Test de compar. de moyennes II	181
5.7	Weitere Testverfahren — Autres méthodes de test	183
5.7.1	Eine Typeneinteilung von Tests — Une classification des tests	183
5.7.2	Parameterfreier Test: Vorz'test — Test libre de param.: Test du signe	183
5.7.3	Chi-Quadrat-Anpassungstest — Test d'adaptation de Khi deux	184
5.7.4	Zum Fishertest — Quant au test de Fisher	185
5.7.5	Kontingenztafeln — Tableaux de contingence	187
6	Fehlerech., Regr., Korr. — Calc. de l'err., régr., corr.	189
6.1	Fehlerrechnung (Abhängigkeit) — Calcul de l'erreur (dépendance)	189
6.1.1	Das Problem der Verpflanzung — Le problème de la dépendance	189
6.1.2	Verwendung des totalen Differentials — Appliquer la différentielle totale	190
6.2	Regression — Régression	191
6.2.1	Der Begriff — La notion	191
6.2.2	Methode der kleinsten Quadrate — Méthode des carrés minimaux	192
6.2.3	Korrelation — Corrélation	196
6.2.4	Korrelationskoeff.: Bedeutung — Coeff. de corrélation: Signification	198
6.2.5	Rechnen mit Punktschätzern: Probleme — Calculer avec des estimateurs: Probl.	199
6.3	Zum Schluss — Quant à la fin	200
A	Aus dem DIYMU: Link	203
B	Fehler von statistischen Kenngrößen und Standardfehler: Link	205
C	Monte-Carlo, Resampling und anderes: Link	207
D	Eine Bootstrap-Anwendung Schritt für Schritt: Link	209
E	Datensatzänderung: Link	211
F	Spezielle Wahrscheinlichkeitssituationen: Link	213

G Hinweise zur Datenanalyse: Link	215
H Crashkurs Kombinatorik, Wahrscheinlichkeit: Link	217

Préface à la partie analyse combinatoire

Problèmes aux nombres entiers

Chère lectrice, cher lecteur,

L'analyse combinatoire fait partie du programme du gymnase classique. Dans les écoles qui préparent à la maturité professionnelle, il devrait être traité également. Mais quoi, si un étudiant n'a jamais eu contact avec cette matière pour n'importe quelle raison — ou s'il l'a oubliée? Alors il faut l'élaborer ou répéter. Par conséquent ce texte est conçu comme cours de répétition et comme perfectionnement.

L'importance de l'analyse combinatoire est incontestée. Elle est un outil pour la solution de problèmes qui nous surprennent parfois. Elle est la base pour le "calcul des probabilités et la statistique".

Ce texte est écrit en forme de script. Ça signifie qu'il représente une forme très abrégée de la manière à apprendre. Pour des explications plus vastes et détaillées, exemples, preuves exactes et suppléments, on conseille l'étudiant de consulter plusieurs livres de cours. Etudier signifie en grande partie d'élargir soi-même son savoir à l'aide de la littérature et acquérir de la matière, de l'approfondir et de l'utiliser. Pour cela, un script est seulement un indicateur d'itinéraire et ne remplace jamais un livre de cours. Chacun est libre de choisir ses livres de cours. On trouve le sujet de l'analyse combinatoire pratiquement dans toutes les oeuvres de mathématiques pour le gymnase classique. Concernant le niveau des hautes écoles professionnelles le lecteur est renvoyé à des ouvrages tels que A5 (Brenner, Lesky, Mathematik für Ingenieure und Naturwissenschaftler, Band 1).

Dans l'été 1996

L'auteur

Kapitel 1

Quant à l'organisation

Vue générale

1. Organisation, cadre
2. Matière
3. But, chemin, méthodes, feedback, groupe
4. Exercices, études personnelles
5. Technique de travail et d'apprendre, selfmanagement
6. Droits et devoirs de l'étudiant et de l'école
7. Principes et positions
8. Calculatrices de poche, ordinateurs, software en mathes
9. Organisation du semestre en mathématiques (nombre de notes, règlement des examens, Plan des examens, conditions concernant les examens, exigences formelles, conditions formelles, philosophie des notes, évaluation des exercices et des projets, porte-feuille, chef de classe, professeur chargé de la classe, chef chargé des copies, heures de consultation)
10. Aides (bibliothèque, calculatrice de poche, software en mathématiques, littérature)
11. Plan du temps à disposition
12. Introduction

Kapitel 2

Questions fondamentales de la statistique

2.1 Introduction

2.1.1 Stochastique et statistique

Stochastique est considérée comme le terme générique. Ce nom est dérivé des mots grecs "stochastike techne". En latin cela signifie "ars coniectandi" ce qui signifie "art des taux", art de la conjecture. La stochastique est proprement une branche des mathématiques, bien qu'aujourd'hui on l'enseigne dans des branches d'étude arrangées exprès pour elle. A quelques universités on peut terminer ainsi les études comme "statisticien" ce qui est en réalité "stochasticien". Sous le terme général de "stochastique" on réunit les domaines scientifiques de la "théorie probabiliste" et de la "statistiques".

Par contre aujourd'hui on comprend sous "statistique" le résumé de méthodes pour l'analyse de données empiriques. Parfois les notions "stochastiques" et "statistiques" ne sont pas distinguées si exactement, car l'idée de la notion "stochastiques" n'est pas connue à chacun et par conséquent n'a pas de grande efficacité publicitaire. Beaucoup d'auteurs intitulent leurs livres avec "statistiques", mais ils pensent à "stochastiques". Ce mauvais usage persiste souvent dans le texte du livre.

Le mot "statistiques" se déduit probablement du mot latin "statisticum" qui signifie "ce qui concerne l'état". D'autre part dans la langue italienne il existe le mot "statista" qui signifie homme d'Etat resp. politicien. Au 18ème siècle dans la langue allemande on a commencé à utiliser le mot "statistique". Par ce mot on spécifiait des données sur l'état (théorie d'état). La signification du mot de statistique comme collectionner et interpréter des données date probablement du 19ème siècle.

2.1.2 La nature de la statistique

La grande importance de la statistique est inhérente à elle-même: La statistique sert à élaborer un jugement sur les **événements de masse**. Ces événements de masse correspondent à l'esprit de notre siècle. De nouvelles méthodes rendent les statistiques très efficaces. Les vieilles méthodes empiriques des artisans ne suffisent plus.

Actuellement on fait la différence entre les types suivants des statistiques:

1. La statistique **descriptive** ou **empirique**
2. La statistique **inductive** ou **mathématique**
3. La statistique **généralisant des hypothèses, explorative, extraction des données** ou **data mining**

La statistique descriptive est mathématiquement simple ou inoffensive. Il s'agit ici de la présentation ou représentation du matériel statistique.

Par la statistique mathématique (statistique analytique ou inductive), on veut parvenir aux résultats affirmatifs. Il s'agit de méthodes d'estimation, d'évaluation ou de méthodes de test pour des hypothèses. Comme résultats on obtient des expressions probabilistes. Ici le calcul de probabilité entre comme partie de la théorie. On cherche des rapports, des relations ou des différences dans les données. On cherche aussi la mesure de la sécurité des suppositions, des résultats hypothétiques des modèles. On estime un résultat comme statistiquement assuré seulement si un résultat peut être confirmé dans le cadre de planifications d'expériences prévoyantes et prospectives avec une probabilité très grande. Ici il faut toujours discuter la forme de la question posée par rapport aux alternatives ainsi que la probabilité acceptée.

Du point de vue des méthodes, la statistique explorative ou généralisant des hypothèses est une forme intermédiaire des deux méthodes ou des différentes statistiques mentionnées en haut. Comme le besoin d'application et l'importance de la statistique augmentent, nous sommes actuellement témoins (fin de siècle) d'un développement d'une forme d'application indépendante de cette science.

On distingue entre **statistique analytique** ou **statistique descriptive** qui est mathématiquement inoffensive et qui a comme but la représentation ou présentation de matériel statistique (données). Nous distinguons entre la **statistique mathématique (statistique analytique ou inductive)** qui mène aux résultats affirmatifs ou qui travaille de façon explorative.

Nous connaissons la statistique descriptive et empirique déjà du temps de l'antiquité égyptienne (statistique d'oignons à l'occasion du bâtiment des pyramides ...). Aujourd'hui il s'agit très souvent d'arriver, partant d'une situation décrite par des nombres, à un jugement qui est comparable aux autres jugements, à l'aide d'une "exploitation" mathématique. Les méthodes pour obtenir et exploiter le matériel statistique sont fournies par les **statistiques mathématiques**.

Des statistiques, on exige les cinq qualités suivants, ce qui est facilement compréhensible:

1. **Objectivité** (Indépendance du point de vue du producteur de la statistique)
2. **Reliabilité** (Fiabilité)
3. **Validité** (Une validité qui dépasse le contexte)
4. **Portée** (La mesure d'être juste, la mesure d'être essentiel ou fondamental)
5. **Importance** (La mesure d'être signifiant)

2.1.3 Domaines de problèmes de la statistique mathématique

Concernant la méthode de la statistique mathématique, nous pouvons distinguer les domaines de problèmes suivants:

1. Manière et méthodes de la collection des données
2. classement, structure, disposition, résumé
3. Analyse: Exploitation, conclusion, examen d'hypothèse
4. Estimation la la fiabilité
5. Evaluation
6. Présentation de régularités statistiques

Application: Partout où des événements de masse jouent un rôle. P. ex. statistiques de réparation, statistiques de vente, statistiques de stock ou de matériaux, statistiques d'accidents, de naissances-, de décès-, statistiques de population, statistiques de santé, statistiques de climat, statistiquess dans la fabrication industrielle en vue de contrôles de qualité u.s.w..

Le profit qu'on obtient des statistiques est la mise au point de nombres stratégiques comme base de décisions stratégiques. L'intention est souvent la diminution de risques, l'abaissement de frais, l'élévation du gain ou l'évitement de pertes etc..

2.1.4 Types d'événements de masse

On peut distinguer entre **deux sortes fondamentalement différentes d'événements de masse**:

1. Grand nombre d'individus (statistiques où on compte). Le matériel des données consiste en nombres ($\in \mathbb{N}$).
2. Grands nombres d'événements individuels ou généralement de mesurages physiques. Le matériel des données consiste en valeurs indiquées, c.-à.-d. généralement des nombres réels ($\in \mathbb{R}$).

Quant au matériau observé (individus, phénomènes uniques), il peut s'agir généralement de choses indépendantes et localement et temporellement ou n'importe comment de nature différente, que nous appellons des **unités d'observation** ou des **caractéristiques** qui se manifestent par **sortes ou empreintes différentes**. Nous voulons classifier les sortes de manifestations suivantes:

1. Les caractéristiques **qualitatives**, c.-à.-d. des caractéristiques non quantifiables.
Exemple: On qualifie quelque chose de "beau", "moins beau", "disgracieux" ou "horrible". Un transfert sur une échelle graduée est toujours problématique parce qu'il s'agit dans le cas présent de sentiments humains, qu'on ne peut pas bien délimiter l'un contre l'autre et qui ne sont pas

absolument comparables encore moins s'il s'agit de taxateurs différents. Ce problème émerge souvent à l'occasion de notes d'école. Il paraît absurde de "calculer" avec des manifestations pareilles. On peut se demander, si "beau" plus "horrible" devrait donc être "beau", "moins beau", "disgracieux" ou "horrible". Le même vaut pour une multiplication éventuelle au lieu de une addition.

2.

Des caractéristiques **quantitatives**. Ici, on peut distinguer deux groupes:

(a)

Des caractéristiques quantifiables de façon **discrète**, par exemple des nombres d'individus. On pense peut-être au nombre de gens dans un ascenseur. Ici, les places après le virgule deviennent absurdes, s'il ne s'agit pas de caractéristiques comme par exemple une moyenne. Nous ne trouverons pas une cabine d'ascenseur dans laquelle on observe par exemple "4.852 personnes".

(b)

Les caractéristiques quantifiables comme **continues**, par exemple des résultats de mesures qu'on a mesurées à l'aide d'une échelle graduée. Imaginons qu'on ait ainsi à mesurer la longueur d'un bâton. Celui-ci mesure 3.952 m à quoi la dernière place est inexacte ($\pm 0.001\text{ m}$). Quant à cela il faut tenir compte du fait que lorsqu'il s'agit d'observations d'après la nature, il n'est pas possible d'obtenir une exactitude infinie des résultats. Ici, il faut donc comprendre la notion "continue" dans un sens moins exact c.-à.-d., ne le comprendre pas comme l'exactitude mathématique.

C'est souvent la complexité de la taille du matériel qui contraint à une **division préalable** des nombres présents **en classes** et par conséquent à une représentation de **classes** de remplacement. Contrairement aux événements de masse, il y a des irrégularités accidentelles lors qu'il s'agit d'individus. Les "régularités statistiques" manquent.

2.1.5 Echantillons

Souvent il s'agit, lors d'une analyse d'événements de masse, d'événements qui partent d'un très grand nombre d'événements individuels qu'on ne peut absolument pas saisir dans leur totalité. Par conséquent on est forcé de réduire les événements de façon représentative et de faire un choix. Nous appelons un tel choix représentatif un **échantillon** (prise ou récolte d'échantillons). Le nombre de notions individuelles choisies s'appelle **taille d'échantillon**.

Les **valeurs d'échantillonnage** recueillies ou rassemblées sont saisies normalement dans une **liste initiale** (protocole des mesurages, questionnaires etc.) Elles sortent d'une **population (univers)** de manifestations, qui normalement, à cause des inscriptions, enregistrements ou mentions réels sont disponibles et connues et, s'il est sensé, sont expansibles de façon délimitée dans le cadre de l'enregistrement. Ainsi on pourrait, à l'occasion d'un enregistrement de la diversité biologique, découvrir par exemple une nouvelle sorte de plante qui n'était jusqu'ici pas connue dans aucune liste. Pour cela, la totalité de la population est élargie maintenant de façon limitée.

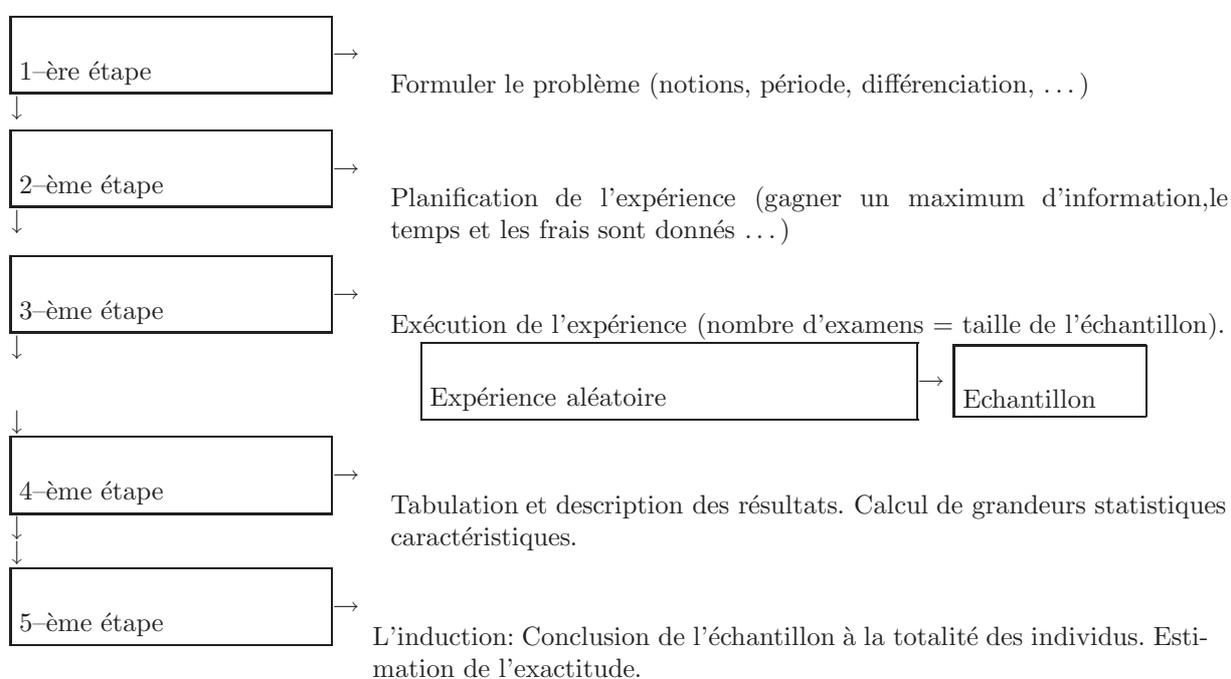
Evidemment, on peut aussi ordonner une liste initiale tout de suite. Comme ça, on obtient des listes initiales ordonnées, par exemple un classement. Mais lors de la construction d'un classement on perd déjà de l'information, parce que l'ordre de collectionner les données ou de collectionner les mesurages est transformé. Cet ordre peut avoir une influence systématique sur les données. Par conséquent le capital

est dans la liste initiale. On devrait la conserver toujours en tous cas comme la source la plus importante tant qu'on travaille encore au traitement statistique des données.

Un **échantillon** consiste en **unités d'observation** (par exemple des individus) qui portent des **caractéristiques** (variables). Celles-ci possèdent normalement une **empreinte** ou manifestation personnelle (valeur) pour chaque individu. Nous appelons la quantité des valeurs ramassées les **données d'échantillon**.

2.2 Façon der travailler dans la statistique mathématique

En pratique, les questions statistiques sont variées. Cependant les étapes d'élaboration et les méthodes utilisées sont souvent les mêmes:



Important:

1. Les conclusions de l'échantillon à la totalité des individus se fondent sur les mathématiques du calcul des probabilités. Des déductions de ce genre complètement sûres n'existent pas.
2. Les conclusions à la totalité des individus ne sont valables que s'il s'agit vraiment d'une **expérience de hasard** (voir 4.2).
3. Une expérience peut mener à la destruction de l'article. Dans ce cas, l'article est donc soustrait à son affectation.

2.3 Statistique descriptive

2.3.1 Questions

Les échantillons ne sont normalement raisonnables que s'ils consistent en nombres ou s'ils peuvent être représentés par des nombres.

Problème:

1. Comment est-ce que les données récoltées peuvent être représentées ingénieusement?
2. Comment est-ce qu'on peut exprimer les échantillons par peu de nombres pour pouvoir les comparer ainsi? (\rightsquigarrow Moyenne, variance ...)

Les idées importantes sont la **fréquence absolue** et **relative**, la **fonction de fréquence** $\tilde{f}(x)$, la **fréquence somme (intégrale)** – ou **fonction de répartition** $\tilde{F}(x)$ ainsi que les **caractéristiques** comme par exemple une moyenne. Les caractéristiques servent à mesurer, jauger ou repérer l'ensemble de données. Le but est par exemple la représentation des données par peu de nombres pour obtenir une comparabilité simple de paquets de données différents.

Remarque:

$$\begin{aligned}\tilde{f} &\rightsquigarrow \text{ engl. distribution function} \\ \tilde{F} &\rightsquigarrow \text{ engl. cumulative distribution function on the sample}\end{aligned}$$

2.3.2 Répartition de fréquences

Les résultats d'un échantillonnage (chompter, mesurer) doivent être retenus ou doivent être verbalisés. Ainsi on obtient une liste de base ou le procès-verbal. Si on mesure par exemple à chaque individu deux qualités, on obtient n paires de nombres et par conséquent on a la taille d'échantillon n .

Exemple:

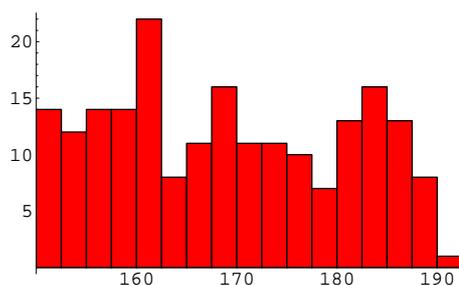
Liste de base: Grandeur de 200 participants d'un cours de pilotes, mesurer. Afin que les données soient mieux représentables, on a divisé les données directement en 40 classes (d'environ 150 à environ 190).

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

Présentation de l'échantillonnage: \rightsquigarrow

Valeurs ordonnées d'après la grandeur: **Tableau de fréquences** (Des fois aussi **liste avec des marques d'après la grandeur**, comme les joueurs de cartes). Les nombres cardinaux donnés s'appellent **fréquences absolues**. P.ex. la valeur 155.6 a la fréquence absolue 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)



Au lieu de faire une liste de fréquence ou une liste de marques, on peut inscrire les données tout de suite dans un histogramme ou dans un diagramme de points. \rightsquigarrow Un point par mesurage, points au lieu de bâtons.

Si on calcule la fréquence comme la fraction de la taille de l'échantillon plein, on parle de la **fréquence relative**. Par cette grandeur on peut comparer des échantillons d'une taille différente.

Définition:

Soit $H(a)$ la fréquence absolue de l'observation A dont la valeur soit $x = a$ et n la taille de l'échantillon.

\rightsquigarrow **Fréquence relative**

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

Il vaut clairement:

Théorème: $0 \leq h(a) \leq 1$

2.3.3 Fonction de fréquences

Donné:

m valeurs différentes $\{x_1, x_2, x_3, \dots, x_m\}$ avec les fréquences relatives $h(x_1), h(x_2), \dots, h(x_m)$, bref h_1, h_2, \dots, h_m . Les valeurs tombent dans m classes $\{[x]_1, \dots, [x]_m\}$. Le taille d'échantillon soit n .

Définition: **Fonction de fréquence:**

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

Définition:

Fonction de distribution: (Distribution discrète)

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

La fonction de fréquence de l'échantillon montre, comme les fréquences sont distribuées. Elle détermine ainsi la distribution de fréquence. Les relations suivant les sont directement compréhensibles:

Théorème:

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 Techniques de représentation

Exemple d'un ensemble de données

Donné:

Les résultats d'un comptage de composantes défectueuses de 40 appareils d'un type donné, qui ont été envoyés à la réparation. On a noté les nombres x_i de composantes.

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, 2, 1, ·

Remarque:

Le point (·) signifie **missing value**: La valeur n'a pas été enregistrée.
 Donc $n = 39$.

~> Tableau de fréquences:

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	—

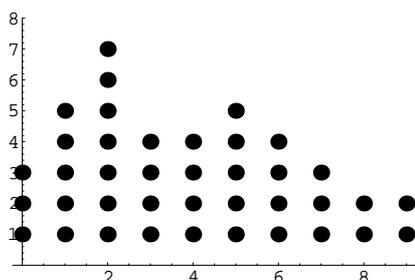
Diagramme de points

Diagramme de points (à la place d'une liste de fréquences)

Remarque:

Au lieu de travailler avec des listes de contrôle (avec des traits) on peut aussi travailler avec des **diagrammes de tronç-feuille**. Ici, pour un mesurage, les chiffres, sans le dernier, sont inscrits par exemple dans un classement (colonnes). Les derniers chiffres sont écrits dans la ligne affilié: Comme à une liste de traits on inscrit le chiffre au lieu d'un trait. Cela a un sens, si les mesurages se distinguent souvent seulement par le dernier chiffre.

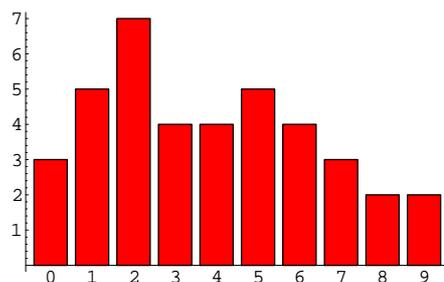
Diagramme en bâtons

Diagramme en bâtons

Diagramme en bâtons généralisé

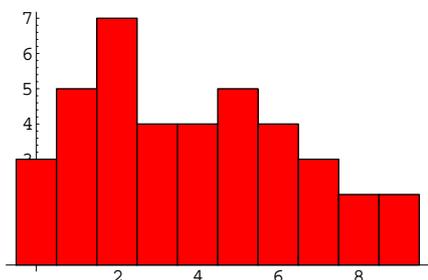
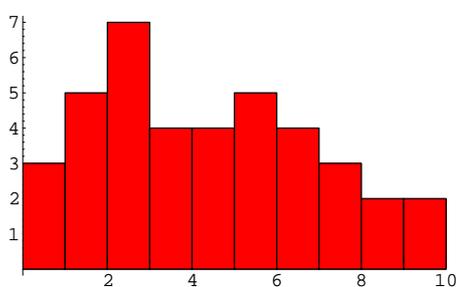


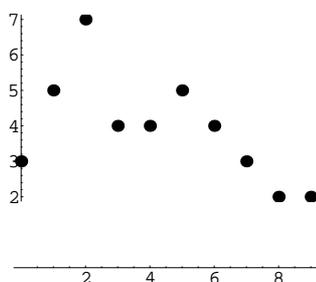
Diagramme en batons généralisé (ici la largeur des bâtons est plus grande)

Histogramme



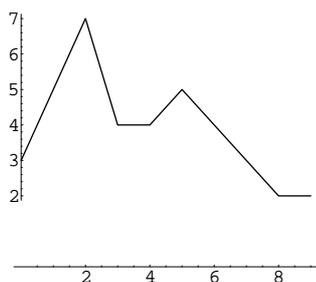
Histogramme (fonction en escaliers), normalement pour résumer des données dans des classes

ListPlot



Plot d'une liste

Polygone de fréquences



Polygone de fréquences

Diagramme à secteurs

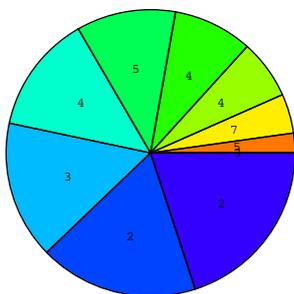
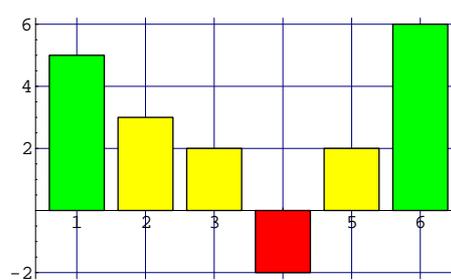
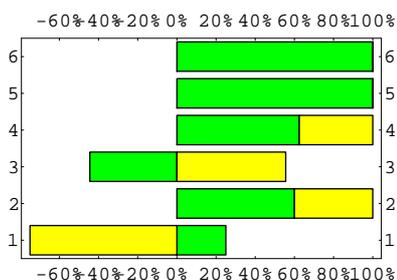
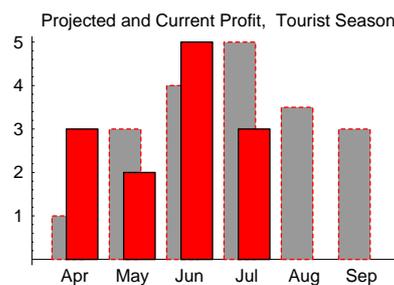
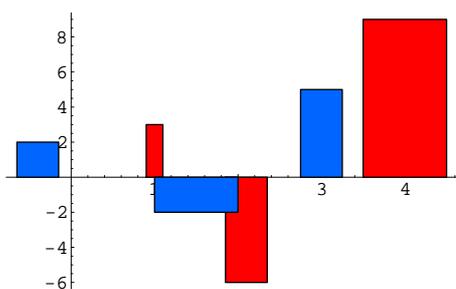
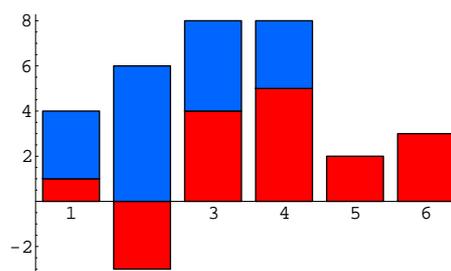
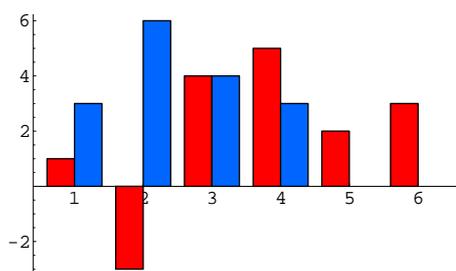
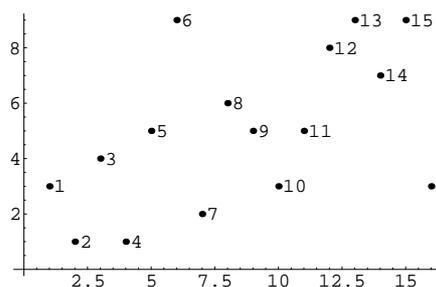
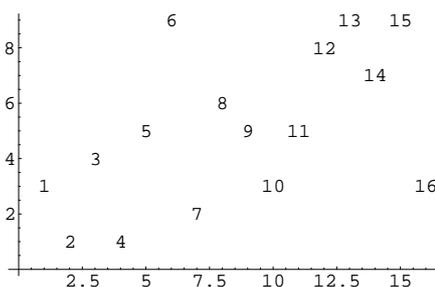
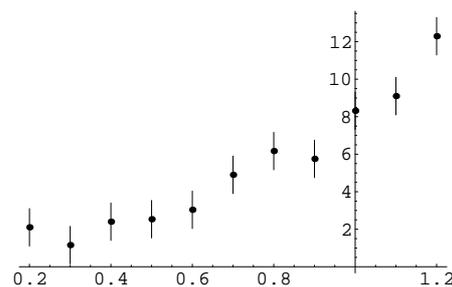
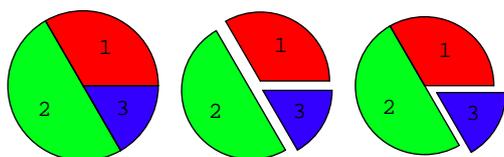


Diagramme à secteurs, diagramme en camembert, graphique sectoriel

Diagrammes structurés

Dans l'exemple suivant les quantités de données sont différentes de la quantité de données qu'on vient de traiter. La nature de la chose le rend nécessaire. Les exemples sont évidents.

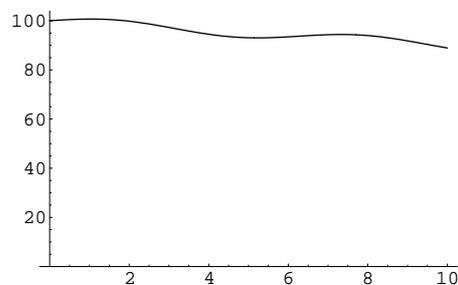
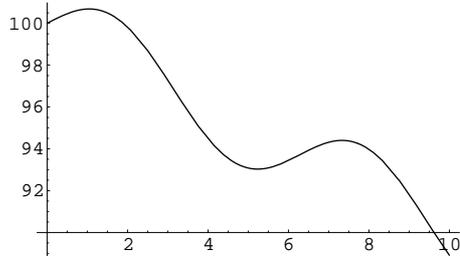




Dans les publications techniques et économiques, on trouve une quantité de types de représentations très vastes. P.ex. les **diagrammes de Sankey** qui rappellent des organigrammes ou des schémas de flux. En outre il faut mentionner les **diagrammes tridimensionnels** ou **représentations 3D**.

Attention escroquerie

L'exemple suivant explique tout:



2.3.5 Quant à la formation ce classes

Si on a une quantité de données assez grande, un tableau ou un graphisme deviennent vite peu clairs. Par groupement des données en classes (**formation de classes**), on peut de nouveau s'y retrouver et simplifier la chose. L'erreur faite est au maximum de la grandeur de la moitié de la largeur des classes.

L'intervalle, dans lequel se trouvent toutes les valeurs d'échantillon, soit subdivisé en **intervalles de classes** (intervalles partiels). Les centres de ces intervalles s'appellent **centres de classes**. Le nombre de valeurs qui sont situées dans une classe, s'appellent **fréquences absolues de la classe** ou **fréquences de la classe**. Quant à cela nous retenons:

Définition:

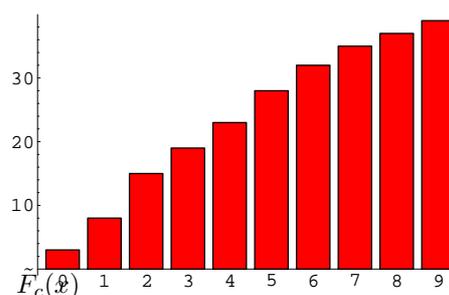
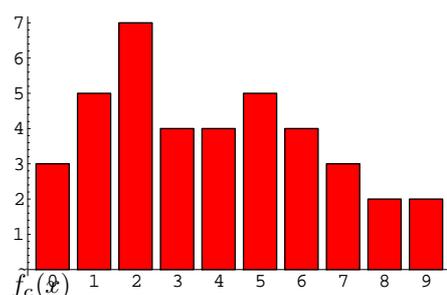
$$\text{Fréquence de classe relative} := \frac{\text{fréquence de classe absolue}}{\text{taille de l'échantillon}}$$

Remarque:

Maintenant nous supposons que toutes les valeurs d'une classe soient situées dans le centre de la classe. (Cette supposition est normalement faite de cette manière.) Ça nous produit une erreur qui est au maximum aussi grande que la demi-largeur de la classe.

La fréquence en dépendance du centre de la classe est maintenant décrite par la fonction de fréquence $\tilde{f}_c(x)$ de l'échantillon qui est maintenant divisé en classes. Liée avec $\tilde{f}_c(x)$ est la fonction de fréquence de somme ou fonction de distribution $\tilde{F}_c(x)$. $\tilde{F}_c(x)$ est une fonction en escalier qui croît monotonement.

L'exemple: Les diagrammes suivants montrent $n\tilde{f}_c(x)$ et $n\tilde{F}_c(x)$ représentés par des diagrammes de bâton. En effet cette représentation n'est pas tout à fait correcte, mais elle est aujourd'hui souvent produite facilement par les moyens de l'ordinateur.



Il est évident qu'il vaut:

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

Remarque:

1. $\tilde{f}_c(x)$ et $\tilde{F}_c(x)$ sont constantes sauf pour $x =$ limite de classe.
2. $\tilde{f}_c(x)$ et $\tilde{F}_c(x)$ ont des sauts à la limite des classes. La hauteur du saut de $\tilde{F}_c(x)$ est égale à $\tilde{f}_c(x)$.

2.4 Mesures d'un échantillon

Un échantillon peut être décrit d'une part par la fonction de fréquence ou par la fonction de distribution. D'autre part on peut aussi le caractériser par certains nombres ou certaines mesures. Ces mesures sont numériquement comparables aux mesures d'autres échantillons. C'est un avantage. Le désavantage est

que beaucoup d'informations qui étaient au début encore là, ne sont plus directement visibles par les mesures. Les mesures importantes sont la moyenne et la variance.

On divise de telles mesures en des catégories différentes. Ainsi la moyenne ou bien le minimum et le maximum d'un échantillon sont des **mesures de position**. Par contre la dispersion ou la variance sont des **mesures de dispersion**.

2.4.1 Moyenne et variance empirique

Moyenne

La moyenne montre la grandeur moyenne des valeurs d'échantillon ou des classes. C'est la moyenne arithmétique des valeurs d'échantillon ou des valeurs des classes:

Définition:

Moyenne

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

Problème:

Les échantillons 1, 2, 3, 4, 5 et 2.7, 3.0, 3.1, 3.2 ont la même moyenne, tandis qu'ils sont essentiellement différents. On cherche donc une mesure qui représente la différence des valeurs d'échantillon de la moyenne

Au lieu de la moyenne arithmétique ordinaire on utilise aussi d'autres moyennes. Il y en a beaucoup: Par exemple **la moyenne géométrique**, **la moyenne harmonique** ou **le moyen arithmétique pondéré** (correspondant à une somme de moments pondérés). Pour cela on consulte la littérature.

Variance empirique

Une possibilité serait de prendre la somme des différences $\sum(x_i - \bar{x})$. Mais ceci ne fonctionne pas comme on voit tout de suite:

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

$x_i - \bar{x}$ peut être négatif ou positif ou aussi zéro. Mais alors, les valeurs s'annulent dans la somme. Ceci pourrait être évité en prenant les valeurs absolues $|x_i - \bar{x}|$. Mais il est mathématiquement plus simple et plus profitable d'en prendre à cette place les carrés. En nous appuyant à la moyenne de l'écart carré des valeurs d'échantillon x_i de la moyenne d'échantillon \bar{x} ainsi imaginable, nous définissons:

Définition:

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

s'appelle **variance** de l'échantillon $\{x_1, \dots, x_n\}$.

$s = \sqrt{s^2}$ s'appelle **écart-type** (étalon, quadratique moyen, déviation normale)

(Problème de notions de langue allemande.)

Remarque:

Il est intéressant de remarquer que la variance empirique a dans le dénominateur $n - 1$ et non n , comme on s'attendrait pour la moyenne. D'une part ça n'a pratiquement aucune influence sur la valeur de s^2 si n est assez grand. D'autre part on atteint par cela que s^2 est une "appréciation de la diffusion σ^2 qui correspond à l'espérance mathématique" de l'ensemble de base de données. (Quant à cela voir les exposés dans Storm, Bibl. A12.)

La variance et l'écart standard ne sont pas robustes contre des valeurs aberrantes.

Exemple 1:

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

Exemple 2:

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

Définition:

$[\bar{x} - s, \bar{x} + s]$ s'appelle **intervalle standard**

Dans l'exemple dernier l'intervalle standard est environ $[2.78, 3.22]$.

Autres mesures

Définition:

Etendue de l'échantillon := valeur maximale moins valeur minimale de l'échantillon.

Médian (\tilde{x} , valeur centrale) de l'échantillon := valeur au centre si le nombre des valeurs de l'échantillon est impair, si non moyenne arithmétique de la valeur moyenne des deux valeurs au centre. (La valeur centrale (médiane) est robuste contre les valeurs aberrantes!)

Quartile, Quantile: Voir Box-Whisker-Plot, page [23](#).

Mode (D, moyenne de densité) de l'échantillon := valeur à la fréquence maximale. (Des fois, le mode est pris aussi de façon locale.)

Dissymétrie, kurtosis et des valeurs moyennes différentes, voir page [25](#).

2.4.2 Méthodes de simplification aux calculs

Il vaut:

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

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

Pour des nombres grands x_i , une transformation de coordonnées $x_i = c + x_i^*$ est favorable. (c choisi de manière que les x_i^* deviennent des nombres petits.)

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

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

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

Théorème:

Hyp.:

$$x_i = c + x_i^*$$

Thè.:

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

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

$$3. s = s^*$$

Des fois une **transformation linéaire** est utile:

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

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

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

Théorème:**Hyp.:**

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

Thè.:

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

2.4.3 Calculs au moyen de la fonction de fréquence

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

Soit n_k la fréquence dans la classe k . Pour les classes vaut donc:

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

Remarque:

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

Erreur: Au max. demi-largeur de classe.

Théorème:

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

Il vaut:

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

A la page 18 nous avons calculé:

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

Analogiquement nous déduisons aussi ici:

Théorème: Soit $s_K^2 := (\overline{[s]})^2$: $s^2 \approx s_K^2 = \frac{1}{n-1} \sum_{i=k}^j ([x]_k - \overline{[x]})^2 \cdot n \cdot \tilde{f}([x]_k)$

$$= \frac{1}{n-1} \cdot \sum_{i=k}^j ([x]_k \cdot n \cdot \tilde{f}([x]_k) - \frac{1}{n} \cdot (\sum_{i=k}^j x_k n \tilde{f}([x]_k))^2)$$

$$= \frac{1}{n-1} \cdot (\sum_{i=k}^j ([x]_k)^2) \cdot n \cdot \tilde{f}([x]_k) - n \cdot (\overline{[s]})^2$$

2.4.4 Distribution de fréquence et distribution de masse

Idée:

La masse 1 soit distribuée le long d'une droite.

Valeurs:	x_1	x_2	x_m
Fréquence:	$h_1 = \tilde{f}(x_1)$	$h_2 = \tilde{f}(x_2)$	$h_m = \tilde{f}(x_m)$

Analogiquement:

Point:	x_1	x_2	x_m
Masse:	h_1	h_2	h_m

$$\Rightarrow \tilde{F}(x_1) = \sum_{x_i \leq x} h_i = \sum \text{masses à gauche de } x$$

Conséquence:

Le **centre de gravité** x_S correspond à la **valeur moyenne** \bar{x} : $x_S \hat{=} \bar{x} \cdot 1 = \sum x_i \cdot h_i$

Le **moment d'inertie** Φ correspond, à l'exception d'un facteur, à la **variance** s^2 : $\Phi \hat{=} \frac{n}{n-1} \cdot s^2$

2.4.5 Exemple avec Mathematica

Charger les "packages" — **Programme en Mathematica:**

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

Données — **Programme en Mathematica:**

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

```
 {"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 Exploitation: Exemple

2.5.1 Entrée de données

Nous utilisons *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 Caractéristiques

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 Représentation par caract.: BoxWhiskerPlot

Le **Box-Whisker-Plot** est d'une valeur inestimable pour gagner rapidement une vue générale sur un paquet de données numériques. Ce paquet reçoit sa forme par un rectangle, qui est donné autour du médian par la distance entre les deux quartiles. Normalement ce sont les quartiles de 25 % et de 75 %. Normalement ce sont des "whiskers", c.-à.-d. des lignes comme les poils d'une moustache qui enferment un rectangle, qui marque l'élargissement du paquet des données soit entièrement, soit sans les valeurs aberrantes. Les points aberrants sont des points au-delà d'une distance de $3/2$ de la distance entre les quartiles mentionnées, mesurée depuis le rectangle. Puis les **grandes valeurs aberrantes** sont marquées par des points au-delà de 3 fois cette marge.

La **distance des quartiles (différence des quartiles)** est une mesure de la dispersion ou variance (hauteur du rectangle). En outre le médian est inscrit dans le rectangle et donne par sa situation dans le rectangle une impression du biais de la distribution des données. Ces quartiles ainsi que le médian résistent aux valeurs aberrantes: Elles ne changent pas énormément si on omet les valeurs aberrantes, contrairement à la moyenne arithmétique et à l'écart standard. Comme déjà le **médian** était une mesure de position **robuste**, ainsi par conséquent aussi la différence des quartiles est une mesure de l'écart robuste contre les valeurs aberrantes.

Extrait du fichier Help de *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 interquantile range from the edge of the box; far outliers are points beyond three times the interquantile range.“

Remarque:

Quant aux **quartiles** on utilise la définition suivante:

Définition: **Quartile q_j :**

Soit $n = 4k + r$, $k, r \in \mathbb{N}$, $x_k =$ valeur de mesure numéro k dans le classement des valeurs indiquées \rightsquigarrow

$$r = 0 \Rightarrow q_{0.25} = 0.25 x_k + 0.75 x_{k+1}$$

$$r = 1 \Rightarrow q_{0.25} = x_{k+1}$$

$$r = 2 \Rightarrow q_{0.25} = 0.75 x_{k+1} + 0.25 x_{k+2}$$

$$r = 3 \Rightarrow q_{0.25} = 0.5 x_{k+1} + 0.5 x_{k+2}$$

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

$$r = 0 \Rightarrow q_{0.75} = 0.75 x_{3k} + 0.25 x_{3k+1}$$

$$r = 1 \Rightarrow q_{0.75} = x_{3k+1}$$

$$r = 2 \Rightarrow q_{0.75} = 0.25 x_{3k+1} + 0.75 x_{3k+2}$$

$$r = 3 \Rightarrow q_{0.75} = 0.5 x_{3k+2} + 0.5 x_{3k+3}$$

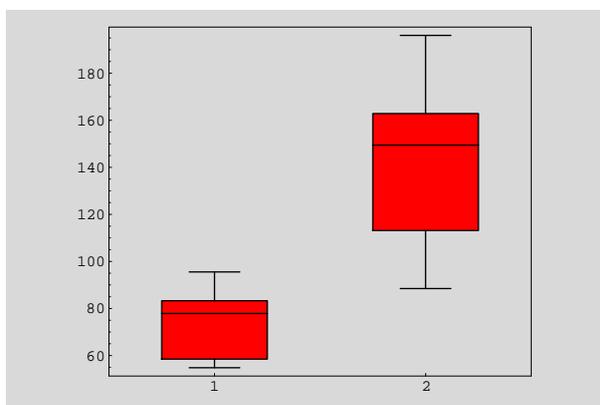
$Q_1 = q_{0.25} =$ **1er quartile**, $Q_2 = q_{0.5} =$ **2ème quartile** = médian, $Q_3 = q_{0.75} =$ **3ème quartile**, correspondant pour les **quintiles** etc.

La première quartile est donc la valeur environ à un quart des valeurs indiquées, la deuxième quartile est la valeur environ à la moitié des valeurs indiquées et la troisième quartile est la valeur à trois quart des valeurs indiquées.

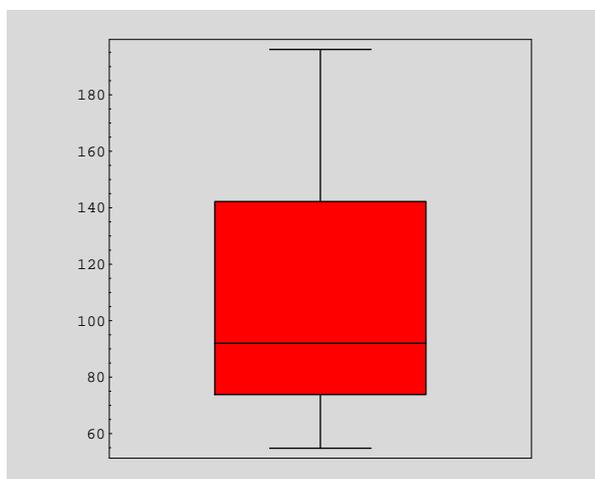
Voir aussi: <http://de.wikipedia.org/wiki/Boxplot>

In:

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



```
BoxWhiskerPlot [Transpose[{allData}]]];
```



2.5.4 D'autres représentation statistiques

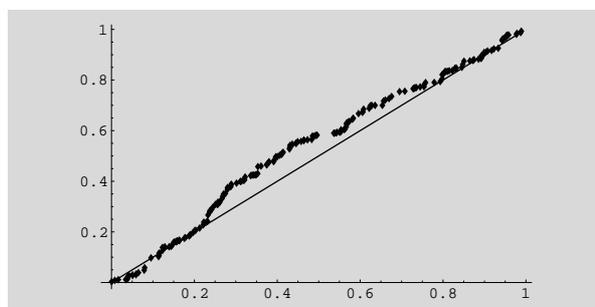
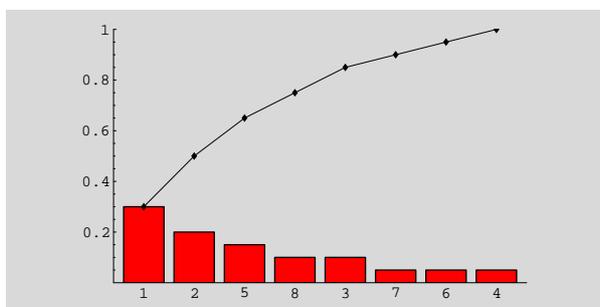
D'autres plots statistiques: Voir explications de *Mathematica*.

In:

```
data={1,1,1,2,2,3,2,4,1,5,5,6,3,1,1,7,2,8,8,5};
```

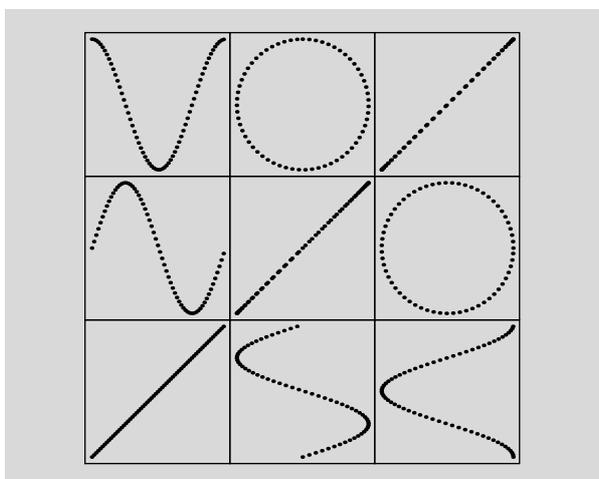
```
ParetoPlot [data];
```

```
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];
```

Voir Help dans *Mathematica*.

2.6 D'autres caractéristiques

2.6.1 Certaines valeurs moyennes d'une distribution

Définitions et conclusions:

1.

Moyenne arithmétique:

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

2.

Moyenne arithmétique pondérée, poids w_i à 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.

Moyenne géométrique:

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

4.

Moyenne géométrique pondérée:

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

5.

Moyenne harmonique:

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

Pour seulement deux valeurs x_1 et x_2 il vaut:

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

6.

Moyenne harmonique pondérée:

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

7.

Moyenne logarithmique de deux valeurs positives:

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

Alors il vaut:

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

8.

Moyenne de puissances avec des valeurs positives:

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

Remarque:

Pour $k = 1$ on a la moyenne arithmétique, pour $k = 2$ on a la moyenne quadratique \bar{x}_q , pour $k = -1$ on a la moyenne harmonique etc.

En outre avec le r -ième moment statistique (voir section suivante):

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

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

9.

Moyenne quasiment arithmétique: f soit continue de façon strictement monotone et continue,

$$w_j \in [0, 1], \quad \sum_{j=0}^n w_j = 1$$

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

10.

Des moyennes "winsorisées": Si les données sont contaminées par des valeurs aberrantes (quelques valeurs beaucoup trop hautes ou beaucoup trop basses) et si on ne veut pas considérer ces valeurs comme données valables, on "winsorise" les données. C.-à.-d. on assortit les valeurs d'observation d'après le rang ascendant ou descendant. Alors on trace un même nombre de valeurs au début et à la fin du classement si obtenu, de façon que les valeurs aberrantes disparaissent. Après on calcule la moyenne sur des valeurs restantes. Exemple: On produit une moyenne "winsorisée" du niveau de 20 % en traçant 10 % des nombres de toutes les valeurs à la fin inférieure et 10 % des valeurs à la fin supérieure du classement.

11.

D'autres moyennes intéressantes, cependant qu'on rencontre rarement dans la statistique descriptive, sont la **moyenne "a"** et des manifestations distinctes de **moyennes glissantes** ou de **moyennes arithmétiques-géométriques**. Concernant ces moyennes, le lecteur est prié de consulter la littérature spécialisée parce qu'elles excèdent le cadre de ce cours.

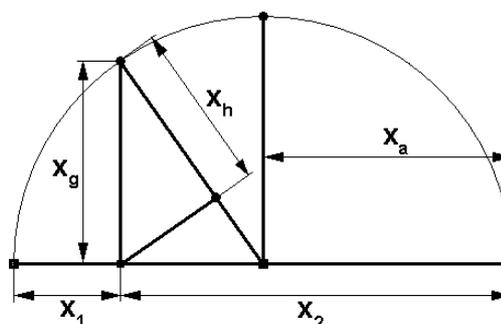
Indication:

A l'aide du théorème de la hauteur et

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

On voit dans la figure ci-contre:

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



2.6.2 Moments d'une distribution

Soyent x_1, x_2, \dots, x_n les valeurs indiquées, qu'on rencontre plusieurs fois. Nous supposons que, si nous faisons une liste avec chaque fois seulement un exemplaire de chaque valeur rencontrée, nous obtenons la liste $\{x_1, x_2, \dots, x_k\}$. Soient $f_1 = n_1, f_2 = n_2, \dots, f_k = n_k$ les fréquences ou fréquences absolues dues à la liste. Alors il vaut pour le nombre total des valeurs rencontrées $n = n_1 + n_1 + \dots + n_k = f_1 + f_1 + \dots + f_k$.

Analogiquement à la **moyenne** on définit maintenant $m_{r,a}$ comme **moment d'ordre r** de la série de mesure par référence à un point de départ ou valeur initiale "a". S'il vaut $a = \bar{x}$, nous écrivons brièvement m_r et nous parlons ici du moment d'ordre r sans mentionner le point de départ.

Définition:

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

Conclusions:

S'il vaut $a = \bar{x}$, on obtient:

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

S'il vaut en plus $f_1 = n_1 = f_2 = n_2 = \dots = f_k = n_k = 1$, on obtient:

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

S'il vaut $a = 0$, on obtient:

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

Par conséquent il vaut pour la moyenne \bar{x} et l'écart standard s :

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

En outre on obtient les relations suivantes:

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

Remarque:

Dans la littérature, on trouve une grande abondance de traitements des moments. Nous mentionnons "l'épreuve de Charlier" et "la correction de Sheppard" ainsi que la forme sans dimension de la définition des moments $M_r = \frac{m_r}{s^r}$. (Voir p.ex. Murray R. Spiegel, Statistique, McGraw-Hill, Schaum-Lectures.)

2.6.3 Le biais d'une distribution

Le biais devrait décrire la mesure du caractère unilatéral ou la déclivité d'une fonction de fréquence. Le biais est ainsi une mesure pour l'asymétrie. Dans certains domaines spécialisés, on définit une telle mesure d'après les besoins pratiques. Un accès physique évident au biais n'est pas directement visible. Le biais n'est normalement pas une grandeur à mesurer. Le biais indique, si (et avec quelle force) un distribution est penchée à droite (positivement) ou à gauche (négativement). Il est facilement compréhensible que les définitions suivantes sont raisonnables pour la statistique descriptive:

Définition:

Premier coefficient de biais de Pearson:

$$P_{S,1} := \frac{\text{valeur moyenne} - \text{modus}}{\text{écart-type}} = \frac{\bar{x} - x_{\text{modus}}}{s}$$

Deuxième coefficient de biais de Pearson:

$$P_{S,1} := \frac{3 \cdot (\text{valeur moyenne} - \text{médian})}{s} = 3 \cdot \frac{\bar{x} - x_{\text{median}}}{s}$$

Un coefficient de moment comme mesure de biais:

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

Indication: Le médian est disponible de façon plus simple que le mode, en particulier s'il s'agit d'une distribution multi-modale.

2.6.4 Kurtosis et excès

La "kurtosis" et "l'excès" décrivent le bombement ou le caractère pointu d'une distribution statistique. On appelle une distribution "leptokurtique" si elle est haute et mince. On appelle une distribution "platykurtique" si elle est inférieure et plate. Des distributions à peu près normales s'appellent "mesokurtiques". L'excès lié avec la kurtosis décrit l'écart sur le tracé de la distribution par rapport au tracé d'une distribution normale.

Définition:

Kurtosis des moments:

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

Kurtosis quartile-décile:

Analogiquement aux quartiles Q_1, Q_2, Q_3 on définit aussi les centiles P_1, P_2, \dots, P_{99} . Soit Q la distance des quartiles: $Q = \frac{1}{2}(Q_3 - Q_1)$. Alors on définit la kurtosis quartile-décile:

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

Excès:

$$\gamma_2 = \beta_2 - 3$$

2.6.5 Caractéristiques: Sens propre et danger

Un exemple peut illustrer le problème du sens des mesures statistiques. En pratique, on rencontre souvent la situation que les valeurs indiquées sont utilisées de manière irréfléchie dans des calculs de toutes sortes. Supposons qu'il s'agisse de traiter le bien-être de deux personnes, Jean et François. Jean est couché sur le lit. Dans sa chambre, on mesure la température de $30^{\circ}C$. La température est en ordre. Pour le bien-être de Jean, il n'y a pas de problème concernant la température de la chambre. La valeur moyenne de la température de ses deux mains est de $30^{\circ}C$.

Dans une autre maison dans une autre chambre, François est retenu par deux gangsters. Il est torturé. On lui a mis de façon serrée une main dans de l'eau bouillante. L'eau cuit à $100^{\circ}C$. À la surface de la main, où les neurones sont placés pour la sensation de la température, on constate par conséquent $100^{\circ}C$. On lui a gelé l'autre main dans un bloc de glace à $-40^{\circ}C$. À cette surface de la main on constate par conséquent $-40^{\circ}C$. La moyenne est $\frac{1}{2} \cdot (-40^{\circ}C + 100^{\circ}C) = 30^{\circ}C$. La température moyenne est donc la même chez François est chez Jean. Celui qui ne connaît pas la situation des deux et qui connaît seulement la moyenne de $30^{\circ}C$ et qui se fie à cette déclaration statistique, doit croire que les deux, Jean et François, concernant leur bien-être par rapport à la température, ne se distinguent pas. Ahurissant! On pense que pratiquement chaque jour on fait de la politique et de la publicité avec des données statistiques, sans que jamais un mot ne soit dit sur l'ensemble des données et leurs circonstances!

Kapitel 3

Analyse combinatoire

3.1 Introduction

3.1.1 Problème

Dans l'analyse combinatoire, nous traitons les 6 types de problèmes des nombres cardinaux classiques. Il s'agit des catégories suivantes de questions: Nous demandons le nombre des possibilités de pouvoir choisir des éléments dans un ensemble fini M d'après une prescription donnée, et éventuellement aussi de les ordonner ou de diviser l'ensemble en classes. Dans certains cas les éléments peuvent être répétés — ou bien non répétés. Comme le résultat y est chaque fois un nombre naturel, nous parlons de **fonctions dans les nombres cardinaux** $M \mapsto y$.

3.1.2 Factorielles

Dans l'analyse combinatoire, la notion des factorielles joue un grand rôle. On définit les factorielles¹ par la relation de récurrence suivante:

Définition 3.1 (Factorielle:)

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

Remarques:

1. Il vaut donc: $n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1 = \prod_{k=1}^n k$. (Voir⁽²⁾.)
Il vaut donc: $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$ etc..

2.

La notion de *récurrence* est très répandue aujourd'hui dans l'informatique. On entend par cette notion la définition d'une fonction ou une méthode par elle-même (voir aussi par exemple Bibl. A6 (Claus, Schwill)). Mais il ne faut pas confondre la notion de la *récurrence* qui cependant est utilisée ici dans un sens simple avec les notions *récurrence commune*, *récurrence primitive*, *fonction de récurrence* (dans la théorie des nombres) et *relation de récurrence* (dans des sens différents dans la logique et la théorie des ensembles) qui sont un peu difficiles et usuelle dans les mathématiques. Voir aussi Fachlexikon *a b c* (Bibl.: A1), Iyanaga, Kawada (Bibl. A8) et Meschkowski, Bibl. A11.

¹D'après le schéma de l'induction complète, voir le sujet des nombres naturels, axiome d'induction, un des axiomes du système de Peano.

² \prod signifie "produit".

Nous retenons:

Définition 3.2 (Récurrence (informatique))

Sous récurrence nous entendons la définition d'une fonction ou d'une méthode par elle-même.

De la définition de $n!$ on conclut: $f(n) = n \cdot f(n - 1)$. La fonction à la place n est donc définie par la fonction à la place $n - 1$ (ainsi par elle-même).

Les valeurs $f(n) = n!$ augmentent très vite. Par exemple il vaut:

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

Un programme simple sur un ordinateur peut donc très vite causer des problèmes. Voici une formule de *Stirling* très utile (sans la preuve):

Théorème 3.1 (Formule de Stirling:)

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

Ici e est le nombre de Euler: $e \approx 2.71828182845904523536028747135$ (à 30 places)

3.2 Problèmes d'arrangement

3.2.1 Permutations sans répétition

Paradigme (Exemple d'un problème pratique)³

Problème 3.1 (Possibilités de s'asseoir:)

Situation:

Dans une salle de classe il ne se trouve rien à l'exception de 26 chaises numérotées. Les numéros vont de 1 à 26. On vient d'enlever pupitres, bancs et tables. 26 étudiants attendent devant la porte. Pour pouvoir mieux distinguer les étudiants et pour mieux pouvoir les appeler, chaque étudiant reçoit un numéro différent de 1 à 26 avec lequel il est donc appelé.

Question:

De combien de manières différentes est-ce qu'on peut mettre les 26 étudiants sur les 26 chaises, c.-à.-d. combien est-ce que de répartitions des places existent?

Solution:

- L'étudiant no. 1 entre. Il trouve 26 chaises libres. Par conséquent il a 26 possibilités de s'asseoir.
- L'étudiant no. 2 entre. L'étudiant no. 1 est assis sur une chaise quelconque. L'étudiant Nr. 2 trouve encore 25 chaises libres. Par conséquent il a seulement 25 possibilités de s'asseoir. Mais il a ces 25 possibilités de s'asseoir pour chaque position de l'étudiant no. 1, qui pouvait se placer sur 26 sièges différents. A la première possibilité utilisée par l'étudiant no. 1, l'étudiant no. 2 a maintenant 25 possibilités, pour la deuxième possibilité de l'étudiant no. 1, l'étudiant no. 2 a aussi 25 possibilités, etc, pour la dernière possibilité de l'étudiant no. 1, l'étudiant no. 2 a de nouveau 25 possibilités. En tout les deux ont ainsi $26 \cdot 25$ possibilités. Les nombres des possibilités se multiplient!

³ Un paradigme est un exemple démonstratif

- L'étudiant no. 3 entre. Les étudiants no. 1 et no. 2 sont déjà assis. L'étudiant Nr. 3 ne trouve que 24 chaises libres. Il a par conséquent pour chacune des $26 \cdot 25$ possibilités des premiers deux étudiants encore 24 possibilités. En tout les trois ont ainsi $26 \cdot 25 \cdot 24$ possibilités, parce que les nombres des possibilités se multiplient.
- Ainsi on avance. Finalement l'étudiant no. 25 entre. Pour chaque façon de se placer des étudiants qui sont déjà là il a encore 2 sièges de libres et par conséquent 2 possibilités de se placer. Totalement les 25 premiers étudiants ont ainsi $26 \cdot 25 \cdot 24 \cdot \dots \cdot 3 \cdot 2$ possibilités de se placer.
- Enfin le dernier étudiant, numéro 26, entre. Pour chaque façon de se placer des autres étudiants il ne lui reste qu'une chaise de libre, il n'a donc qu'une seule possibilité de se placer. Totalement les 26 étudiants ont par conséquent $26 \cdot 25 \cdot 24 \cdot \dots \cdot 3 \cdot 2 \cdot 1 = 26!$ possibilités de se placer.

Remarque:

Si dans un ensemble chaque élément (individu) porte un nom distinctif, on peut ranger les éléments "d'après les noms", c.-à.-d. de façon alphabétique, comme dans un lexique: A... vient aevant B...etc, Aa... aevant Ab...etc.. Dans un tel cas on parle d'une disposition lexicographique.

A réfléchir:

Si la classe est capable d'exécuter un changement de place en 10 secondes, elle nécessite $10 \cdot 26!$ secondes pour tous les changements de place. Ça nous fait $\frac{10 \cdot 26!}{60 \cdot 60 \cdot 24 \cdot 365}$ ans = 1.278310^{20} ans. Comparaison: L'âge de l'univers d'après la théorie du big bang est actuellement estimée à env. 1 à 2 fois 10^{10} ans⁴. Pour réaliser toutes les répartitions des places sans pauses, il faudrait donc 10^{10} fois l'âge de l'univers!

Généralisation du problème:

Au lieu de résoudre le problème avec 26 étudiants on peut le résoudre généralement avec n étudiants. Dans l'argumentation il faut alors remplacer 26 par n , 25 par $n - 1$ etc.. Finalement on obtient totalement $(n!)$ possibilités de mettre n étudiants sur n places.

Le problème abstrait

Soit donné un ensemble \mathcal{M}_n avec n éléments, qui sont énumérotés par les numéros de 1 jusqu'à n . \mathcal{M}_n correspond à un ensemble d'étudiants dans l'exemple antérieur. On a ainsi un rapport bijectif des éléments numérotés n_k à un sous-ensemble $\mathbf{N}_n = \{1, 2, 3, \dots, n\}$ des nombres naturels. (On a ainsi une fonction bijective.) Comme le rapport est biunivoque, nous pouvons remplacer les n_k chaque fois par k , sans transformer le problème: $\mathcal{M} = \mathbf{N}_n = \{1, 2, 3, \dots, n\}$. Maintenant on cherche le nombre des possibilités d'appliquer l'ensemble $\mathbf{N}_n = \{1, 2, 3, \dots, n\}$ sur lui-même, c.-à.-d. dans le problème susdit appliquer l'ensemble des numéros des étudiants \mathbf{N}_n à l'ensemble des numéros des chaises \mathbf{N}_n .

Soit $\sigma(k)$ l'image (en haut le numéro des chaises) à une telle application (dans le problème susdit à une possibilité de s'asseoir) de k (en haut k correspond au numéro de l'étudiant). Alors par une telle application, on applique σ (en haut l'ensemble des étudiants) à l'ensemble $\{1, 2, 3, \dots, n\}$ (en haut les chaises). Si on écrit les images $\sigma(k)$ sous les originaux k , ainsi l'ensemble de relation défini par σ apparaît

⁴Les experts se disputent en effet au sujet de cette valeur. Elle est corrigée couramment d'après l'état des connaissances.

dans la forme suivante:

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

Ainsi un sous-ensemble de $\mathbf{N}_n \times \mathbf{N}_n$ est donné pour lequel le rapport de "fonction σ " est valable.

Par conséquent, par le schéma suivant, une nouvelle disposition $\sigma(1), \sigma(2), \sigma(3), \dots, \sigma(n)$ des éléments $1, 2, 3, \dots, n$ est définie.

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

Nous disons:

Définition 3.3 (Permutation:)

La disposition $\sigma(1), \sigma(2), \sigma(3), \dots, \sigma(n)$ des éléments de \mathbf{N}_n s'appelle **permutation** \mathcal{P} de la disposition $(1, 2, 3, \dots, n)$ de ces éléments.

Pour donner une permutation, nous pouvons aussi écrire:

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

Les colonnes se présentent dans un ordre quelconque.

Exemple Par la disposition suivante, une telle permutation est donnée:

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

1 est appliqué sur 4, 2 sur 1 etc.. Maintenant nous pouvons poser notre problème avec les étudiants et les chaises de façon abstraite et générale:

Problème 3.2

Permutations sans répétitions:

Question: *Combien de permutations \mathcal{P} des numéros $1, 2, \dots, n$ existent-ils?*

Autrement: Combien de possibilités de dispositions des nombres $1, 2, \dots, n$ dans un rang existent-elles?

Autrement encore: Combien de fonctions bijectives $\mathbf{N}_n \mapsto \mathbf{N}_n$ existent-elles?

Symboles 1 : $P(n)$

Soit $P(n) =$ nombre des permutations des éléments de M_n des nombres naturels de 1 jusqu'à n .

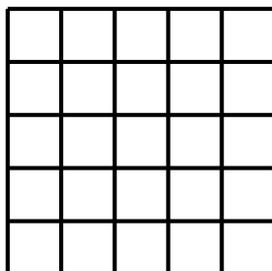
Nous savons:

Théorème 3.2

Les permutations sans répétition:

$$P(n) = n!$$

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



Exemple:

Combien de possibilités est-ce qu'il y a de colorer les différentes surfaces montrées dans 3.1 avec des couleurs différentes? – À une coloration, 25 couleurs différentes sont adjointes aux 25 surfaces différentes. Au lieu de surfaces et couleurs, on peut aussi considérer seulement les numéros de 1 jusqu'à 25. On a ainsi une application bijective d'un ensemble \mathcal{M}_{25} ou de \mathcal{M}_{25} sur soi-même. On cherche donc $P(25) =$ nombre de permutations de $1, 2, 3, \dots, 25$. Ça donne $25! \approx 1.55112 \cdot 10^{25}$. Combien de temps est-ce qu'il faudrait probablement pour exécuter toutes les possibilités?

3.2.2 Permutations avec répétition

Paradigme

Problème 3.3

Possibilités d'échange de lettres:

Situation: *Un chef de personnel a écrit 20 lettres différentes. Il y a 7 copies identiques d'une lettre d'information pour un groupe de collaborateurs et 13 lettres concernant des réponses adressées d'autres collaborateurs concernant des questions de salaire. Les 20 enveloppes sont aussi prêtes.*

Question:

Combien de possibilités d'envoyer les lettres est-ce que la secrétaire a, de façon que pour elle des problèmes pourraient se poser?

Solution:

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

Si toutes les lettres étaient différentes, elle auraient $20!$ possibilités de mettre les lettres dans les enveloppes. Comme une seule possibilité peut être acceptée, les $(20! - 1)$ autres possibilités causent des problèmes.

Si 7 lettres sont maintenant les mêmes, ces 7 lettres peuvent être échangées entre elles sans qu'il y ait de problèmes. Cela on peut le faire de $7!$ manières différentes. Si maintenant X est le nombre de possibilités de placer les 13 lettres différentes dans le 20 enveloppes, pour chacune des X possibilités des lettres différentes les lettres identiques qui restent peuvent être échangées entre elles de $Y = 7!$ manières différentes sans qu'il se passe quelque chose d'embêtant. Comme c'est le cas pour chacune des X possibilités, les nombres des possibilités se multiplient au nombre total des possibilités. On n'a pas d'autres possibilités d'échange que celle qui sont mentionnées ici. Par conséquent il vaut: $20! = X \cdot Y = X \cdot 7!$ et par conséquent $X = \frac{20!}{7!}$.

Le nombre de possibilités indésirables est par conséquent $X - 1 = \frac{20!}{7!} - 1 = 482718652416000 - 1 \approx$

4.82719 10¹⁴.

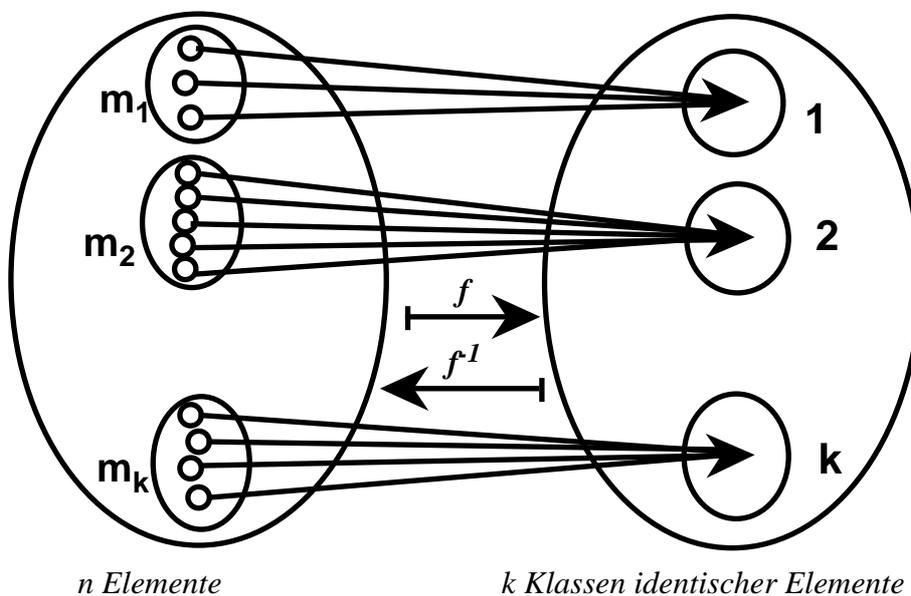
Généralisation du problème:

Nous partons encore de 20 lettres dont 7 sont les mêmes qui sont groupées dans une *classe* 1. En plus on a une lettre spéciale, pour laquelle on trouve deux autres identiques. Ces 3 soient réunies dans une *classe* 2. Nous trouvons encore 4 identiques qui sont réunies dans une *classe* 3. Soit maintenant Y_i le nombre des possibilités d'échange des lettres entre elles dans la *classe* i et X comme en haut le nombre des possibilités d'échange des lettres inégales et restantes. Alors il vaut par la même raison comme en haut:

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

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

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



Esquisse: n éléments \mapsto n classes d'éléments identiques

Ça nous mène au problème général:

Donné:

Totalement n lettres, n enveloppes dont on a k classes de lettres identiques entre elles:

*Classe*₁: m_1 lettres identiques du type 1,

*Classe*₂: m_2 lettres identiques du type 2,

⋮ ⋮

Classe _{k} : m_k lettres identiques du type k

Trouver:

Nombre de possibilités $P_n(m_1, m_2, \dots, m_k)$, de mettre les lettres dans les enveloppes.

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

$P_n(m_1, m_2, \dots, m_k)$ = nombre de possibilités, de placer les n objets qu'on vient de décrire (ici des lettres parmi lesquelles on trouve k classes avec n_j objets identiques entre eux) sur n places (ici des enveloppes).

Définition 3.4 (Permut. avec répétitions:)

Soit donné un ensemble \mathcal{M}_n avec n éléments. Dans cet ensemble on trouve k classes avec n_i éléments identiques (par classe i). A une énumération des éléments, tous les éléments d'une classe reçoivent le même numéro. Nous appelons une permutation des éléments de \mathcal{M}_n **une permutation avec répétitions**.

Maintenant nous savons:

Théorème 3.3 (Permut. avec répétition:) :

Nombre de permutations avec répétitions:

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

Le problème abstrait

Donné:

Un ensemble avec n éléments, par exemple $\mathbf{R}_n = \{1, 2, 3, \dots, n\}$ ainsi qu'un ensemble avec k éléments, z.B. $\mathbf{R}_k = \{1, 2, 3, \dots, k\}$. Puis on considère les fonctions possibles $f : \mathbf{R}_n \mapsto \mathbf{R}_k$ (un exemple est représenté dans image 3.2).

Trouver:

Nombre d'applications inverses possibles $f^{-1} : \mathbf{R}_k \mapsto \mathbf{R}_n$. Pour cela le premier élément (1 à droite dans l'image) est appliqué m_1 fois, le deuxième élément (2 à droite dans l'image) m_2 fois etc..

Les k classes d'éléments identiques (lettres identiques dans le paradigme) sont ainsi appliquées sur les n éléments différents (enveloppes dans le paradigme). Le nombre qu'on cherche est donc $P_n(m_1, m_2, \dots, m_k)$.

Exemple:

De combien de manières différentes est-ce qu'on peut former dans une classe de 26 étudiants 5 groupes de travail avec 4, 5, 5, 6 et 6 étudiants? La solution est:

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

3.3 Problèmes de sélection avec et sans rangement

3.3.1 Les questions

Combinaisons

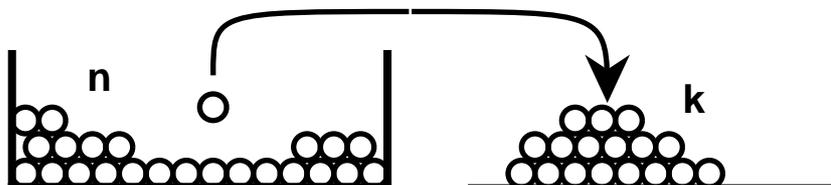
Problème 3.4 (Problème de sélection:)

Donné:

Une caisse qui contient n objets bien distincts, par exemple des boules de différents couleurs. Dans la caisse, on choisit k objets de manière quelconque et on les accumule à côté. (Voir fig. 3.3.)

Question:

De combien de manières différentes peut-on former le tas à côté?

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

La disposition des objets ne joue bien entendu aucun rôle à la disposition des objets resp. des boules. Il est possible de poser ce problème tout de suite abstraitement sans beaucoup de dépense de travail de cerveau. Les boules dans la caisse forment un ensemble \mathcal{M}_n , par exemple $\mathcal{M}_n = \mathbf{N}_n = \{1, 2, 3, \dots, n\}$. On choisit un sous-ensemble $\mathcal{M}_k \subseteq \mathcal{M}_n$, z.B. $\mathbf{N}_k = \{1, 2, 3, \dots, k\}$, $k \leq n$. Ce sous-ensemble forme le tas d'à côté.

Définition 3.5 Combinaison sans répétition

Un tel choix de k éléments dans \mathcal{M}_n s'appelle **combinaison d'ordre k sans répétition** pour n éléments, brièvement: *Combinaison d'ordre k* .

Symboles 3 (Nombre de combinaisons:)

$C(k, n) =$ nombre de combinaisons d'ordre k pour n éléments.

Problème abstrait (combinaisons sans répétition):

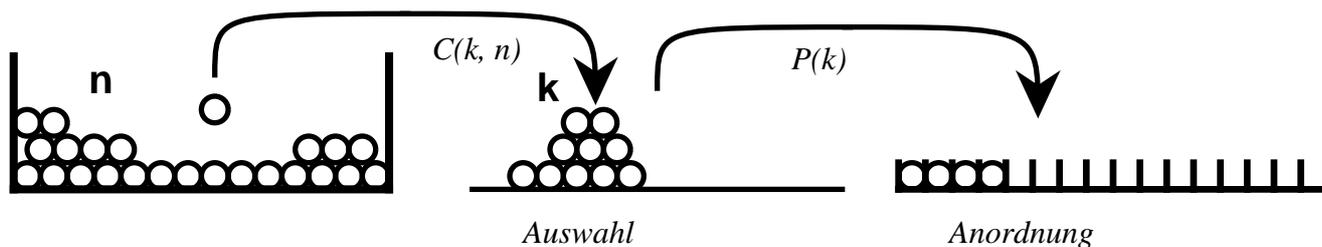
Donné:

Un ensemble \mathcal{M}_n à n éléments.

Question:

$C(k, n) = ?$ C.-à.-d. combien de sous-ensembles peut-on former avec exactement k éléments?

Arrangements

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

Dans la fig. 3.4 le choix (combinaison) est encore arrangé. On peut distinguer deux combinaisons de ce genre avec les mêmes éléments, mais de dispositions différentes. On définit par conséquent:

Définition 3.6

Arrangement sans répétition:

*Si les éléments choisis dans \mathcal{M}_n sont encore arrangés, on parle d'un **arrangement d'ordre k sans répétition** pour n éléments. Brièvement: Arrangement d'ordre k .*

Symboles 4 (Nombre d'arrangements:)

$V(k, n) =$ nombre d'arrangements d'ordre k pour n éléments.

Exemple

Soient donnés les éléments a, b et c . Trouver toutes les combinaisons et toutes les arrangements d'ordre 2.

Solution:

Combinaisons:	$a b$	$a c$	$b c$:	3 pièces
Arrangements:	$a b$	$a c$	$b c$	
	$b a$	$c a$	$c b$:	6 pièces

Répétitions

Si on remplace dans l'ensemble de réserve \mathcal{M}_n chacun des éléments e_i par un ensemble E_i avec les mêmes éléments, qui ne se distinguent que par un *indice interne* (par exemple $E_i = \{e_{i1}, e_{i2}, e_{i3}, \dots\}$), ainsi il devient possible de choisir un élément e_i plusieurs fois. L'indice interne peut être omis après le choix⁵. Nous obtenons le même effet si nous mettons en réserve une copie identique de cet élément après le choix de l'élément. Nous nous imaginons donc qu'un élément e_i se duplique lors de son choix, et, que malgré qu'on ait choisi et enlevé l'élément, l'ensemble \mathcal{M}_n reste inchangé. Un élément est donc fourni tout de suite dans \mathcal{M}_n lors qu'on l'a choisi et enlevé de \mathcal{M}_n . On peut comparer cela à la situation dans un supermarché où les marchandises sont remplacées sur les étagères au fur et à mesure qu'elles sont vendues. S'il est possible aussi souvent qu'on veut de remplir, copier ou remettre les éléments, nous disons que les éléments dans \mathcal{M}_n sont *répétitivement sélectionnables*. Nous définissons maintenant:

Définition 3.7

Combinaison et arrangement avec répétition:

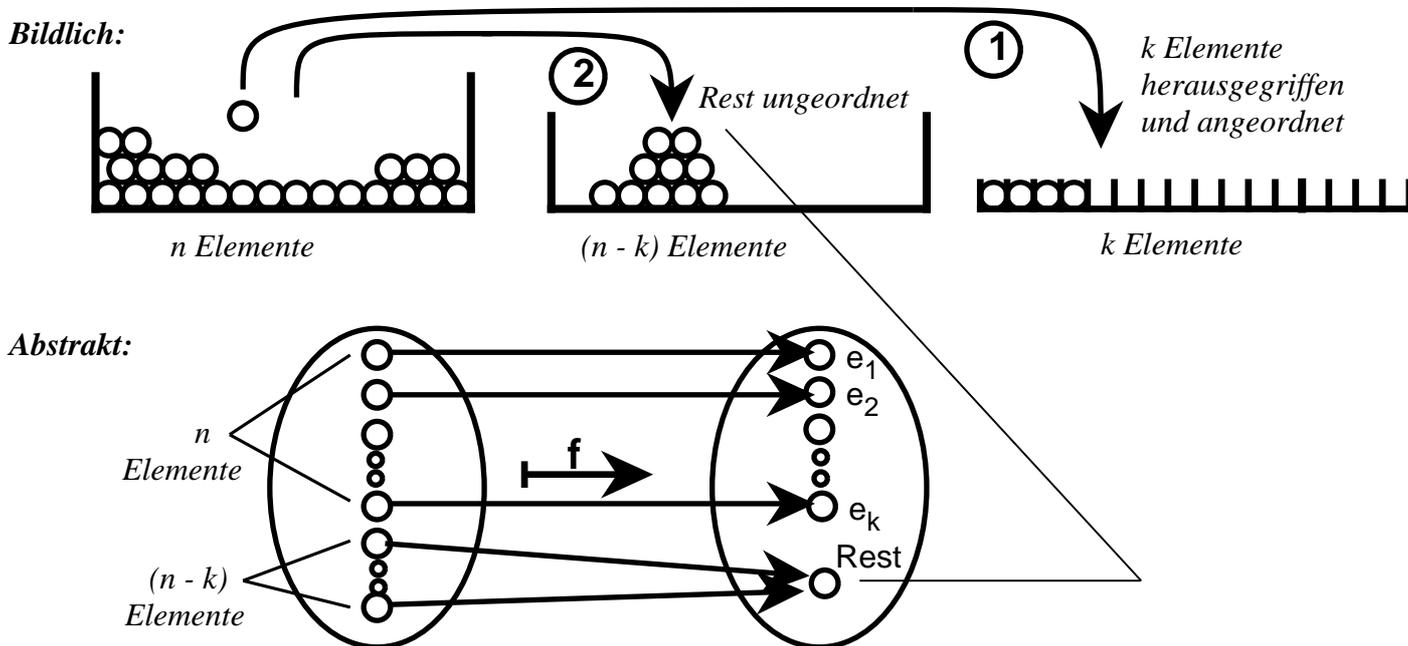
*Si à la formation d'une combinaison ou d'un arrangement les éléments de \mathcal{M}_n sont sélectionnables de façon répétitive, nous parlons d'une combinaison ou d'un arrangement **avec répétition**.*

Nous commençons maintenant avec l'arrangement sans répétition:

3.3.2 Arrangement sans répétition

Dans n éléments on choisit k éléments qui sont disposés immédiatement, sans répétition d'éléments, à

⁵L'indice interne n'est utilisé que pour former l'ensemble d'éléments identiques E_i , qui sont nécessaires pour rendre possible un choix répété d'éléments identiques.

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

l'instar de fig. 3.5. Là par exemple l'élément e_1 est mis sur la place 1, e_2 sur la place 2 etc.. Ensuite on s'imagine que tous les $(n - k)$ éléments non-choisis, donc le reste, sont mis dans une caisse à part resp. sur un tas. Cette opération ne change pas le nombre de possibilités de choix, le nombre de possibilités de disposition des premiers k éléments, car cette formation de tas est une action unique et indépendante qui ne contribue rien à l'opération. Dans cette caisse de restes à part, la disposition des éléments ne joue pas de rôle. On ne distingue donc pas ces éléments, c'est égal comme ils sont disposés. Par conséquent ils forment une classe d'éléments non-distingués et donc une classe d'éléments égaux qui sont disposés d'une seule manière (parce qu'ils comptent comme non-distinctifs). Par conséquent on a le problème suivant: On a n éléments, k sont distinctifs et $n - k$ sont égaux. Ces éléments sont à arranger. Ou bien abstraitement: On cherche le nombre des fonctions inverses possibles f^{-1} (voir fig. 3.5). Ce problème a été résolu déjà à l'occasion des permutations avec répétition: Le nombre est $P_n(n - k) = \frac{n!}{(n-k)!}$.

Théorème 3.4 (Arrangements sans répétition:)

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

Exemple:

De combien de manières est-ce qu'on peut distribuer 20 jobs de vacances différents et disponibles à 26 étudiants différents qui veulent avoir tous un semblable travail, si ces jobs ne sont pas divisibles dans des job partiels?

Il s'agit du choix de 20 sur 26 avec un classement des étudiants non distinctifs, c.-à.-d. un arrangement. La solution est par conséquent:

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

Un cas spécial: $V(n, n) = P_n(n - n) = P_n(0) = P(n)$
 \rightsquigarrow Permutation sans répétition!

3.3.3 Combinaison sans répétition

La formule

A la page ?? nous avons vu qu'à une sélection d'un sous-ensemble de mêmes éléments le nombre des possibilités se comporte de façon multipliquative. On y a trouvé: $20! = X \cdot Y = X \cdot 7!$. Nous trouvons la même situation au passage des combinaisons à l'arrangement: Un arrangement (k éléments de n éléments) peut être obtenu d'une combinaison par la disposition des k éléments choisis. On y a $P(k) = k!$ possibilités. Il vaut donc:

Lemme 3.1 (Arrangements et combinaison:)

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

Il en suit:

Théorème 3.5 (Combinaison sans répétition:)

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

L'exemple du jeu de loto "6 de 45":

De combien de manières différentes est-ce qu'on peut choisir 6 nombres différents dans les 45 premiers nombres naturels? Ici, il s'agit d'une question typique concernant le nombre des combinaisons $C(6, 45)$. Celle-ci est égale à:

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

Coefficients binomiaux

Si on multiplie le binôme $(a+b)^n = \overbrace{(a+b) \cdot (a+b) \cdot \dots \cdot (a+b)}^n$ d'après les règles de la loi distributive,

on n'obtient que des termes additionnels de la forme $m_k \cdot a^k \cdot b^{n-k}$ avec $0 \leq k \leq n$ et $m, k, n \in \mathbf{N}_0$. A la multiplication on prend selon le rang de chaque facteur $(a+b)$ un des termes additionnels a ou b et on les multiplie en obtenant un produit $a^k \cdot b^{n-k}$. Si on prend dans chaque terme additionnel a et jamais b , on obtient $a^n \cdot b^0$. Si on prend a dans j termes additionnels et b dans $n-j$ termes additionnels, on obtient $a^j \cdot b^{(n-j)}$. A cette occasion il existe plusieurs possibilités de choisir le a ou le b . Par exemple on peut choisir a dans le premier facteur, b dans le deuxième facteur, de nouveau a dans le troisième facteur etc., mais on peut aussi choisir d'abord b , après a et alors encore une fois a etc... m_k est le nombre des possibilités de choisir a dans exactement k facteurs $(a+b)$ et b dans exactement $n-k$ facteurs. Il vaut donc:

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

Quelle est maintenant la grandeur de m_k ?— Ici, il s'agit d'un problème de choix: De combien de manières différentes est-ce qu'on peut choisir entre les n facteurs $(a+b)$ k facteurs et y prendre la partie a (et par conséquent prendre dans les $n-k$ facteurs restants chaque fois la partie b)? Cette question est équivalente à la question plus simple: De combien de manières différentes est-ce qu'on peut maintenant choisir entre n éléments (facteurs $(a+b)$) k éléments? La réponse est maintenant très simple: $m_k = C(k, n)$. m_k porte un nom:

Définition 3.8 (Coefficient binomial:) :

m_k s'appelle **Coefficient binomial**.

Symboles 5 (Coefficient binomial:) $m_k := \binom{n}{k}$

Nous savons maintenant:

Théorème 3.6 (Théorème binomial:)

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

On peut lire les coefficients binomiaux dans le triangle de Pascal:

Triangle de Pascal:

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

La position verticale est n , la position horizontale k . Le numérotage commence chaque fois par 0. Ainsi on peut lire par exemple: $c\binom{4}{1} = 4$ und $\binom{4}{2} = 6$.

Pour les coefficients binomiaux, on peut prouver à l'aide de $\binom{n}{k} = C(k, n) = \frac{n!}{k! \cdot (n-k)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{1 \cdot 2 \cdot \dots \cdot k}$ ainsi qu'avec le principe de l'induction complète⁶ facilement les lois suivantes:

Théorème 3.7

Quelques qualités des coefficients binomiaux:

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

Par exemple la formule $2^n = \sum_{k=0}^n \binom{n}{k}$ est obtenue par $2^n = (1 + 1)^n = \sum_{k=0}^n \binom{n}{k} \cdot 1^k \cdot 1^{n-k} = \sum_{k=0}^n \binom{n}{k}$ à l'aide du théorème binomial.

3.3.4 Arrangement avec répétition

La formule

L'arrangement avec répétition a été expliqué à la page 39. La formule pour le nombre d'arrangements avec répétition par contre doit encore être élaborée. Pour cela nous utilisons le symbole suivant:

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

$\bar{V}(k, n) =$ nombre d'arrangements avec répétition pour un choix de k éléments dans une réserve avec n éléments différents, qui tous peuvent être répétés.

⁶Voir théorie des nombres

Déduction de la formule:

Nous considérons les k places numérotés sur lesquelles les éléments à choisir sont ordonnés (voir fig. 3.5 en haut à gauche dans l'image). Comme nous pouvons choisir chacun des n éléments dans la réserve, on a n possibilités d'occuper la 1ère place. Pour la 2ème place on a de nouveau n éléments dans la réserve à disposition pour le choix; à cause de la possibilité de répétition chaque élément existe toujours et peut être choisi: Pour chacune des n possibilités pour la 1ère place on a n possibilités pour la 2ème place, totalement donc $n \cdot n = n^2$ possibilités. Également pour la 3ème place: Pour chacun des n^2 possibilités pour les places 1 et 2 on a n possibilités pour la 3ème place, totalement donc $n^2 \cdot n = n^3$ possibilités. On continue ainsi: Pour l'occupation des premières 4 places on a n^4 possibilités, pour l'occupation des premières 5 places n^5 possibilités et finalement pour l'occupation des premières k places on a n^k possibilités. Par conséquent nous avons le théorème:

Théorème 3.8 (Arrangement avec répétitions:)

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

Exemple:

De combien de possibilités différentes est-ce que 26 étudiants (qu'on peut distinguer) peuvent s'inscrire dans 12 cours différents, si chaque cours offre 26 places, c.à.d. s'il n'y a pas de limites aux places?

Solution:

Le premier étudiant a 12 possibilités de s'inscrire dans un cours. Pour chacune de ces possibilités du premier étudiant le deuxième a aussi 12 possibilités de s'inscrire dans un cours. Les deux ensemble ont 12^2 possibilités. Pour le troisième, quatrième etc. étudiant ça fonctionne aussi d'après le même schéma: Chacun a les 12 possibilités, et les possibilités se multiplient. Il s'agit d'un arrangement avec répétitions. Totalement il y a $\bar{V}(k, n) = \bar{V}(26, 12) = 12^{26} = 11447545997288281555215581184 \approx 1.14475 \cdot 10^{28}$ possibilités.

A retenir: Par cet exemple, on voit que k peut être plus grand que n : $k > n$.

Application: Puissance de l'ensemble de parties

L'ensemble de parties est comme chacun sait l'ensemble de tous les sous-ensembles.

Problème 3.5**La puissance de l'ensemble de parties****Donné:**

Un ensemble \mathcal{M} à n éléments.

Question:

Combien d'ensembles partiels \mathcal{M} a-t-il?

Solution:

$\binom{n}{k} = C(k, n)$ est comme chacun sait le sous-ensembles avec k éléments, car ici, il s'agit d'un problème de choix typique. Maintenant on peut choisir un ou plusieurs sous-ensembles avec 0 (quantité vide), 1, 2, ..., n éléments. Totalement on a donc:

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

Théorème 3.9 :**Puissance de l'ensemble de parties**

L'ensemble de parties d'un ensemble qui contient n éléments possède 2^n éléments.

Donc un ensemble avec n éléments contient exactement 2^n sous-ensembles.

3.3.5 Combinaison avec répétition

Ici il faut choisir k éléments dans un ensemble avec n éléments. Chaque élément choisi se duplique dans l'ensemble de façon que l'ensemble reste toujours le même malgré le choix. Quel est le nombre des options quant au choix?

Pour le calcul de ce nombre, il n'est pas essentiel si l'ensemble \mathcal{M}_n consiste en boules, en billets de loterie ou en nombres, c.-à.-d. quelle est la nature des éléments. Nous pouvons supposer par conséquent qu'il s'agit de nombres naturels de 1 jusqu'à n : $\mathcal{M}_n = \{1, 2, 3, \dots, n\}$. Si nous choisissons maintenant k éléments (c.-à.-d. des nombres), nous voulons les ranger l'un à côté de l'autre au lieu de les "mettre sur un tas". Nous parlons ici du rangement standard (configuration). Nous présentons donc un tel choix $\{e_1, e_2, \dots, e_k\}$ toujours dans une disposition $e_1 \leq e_2 \leq \dots \leq e_k$. Par cela le nombre des options ne change pas.

Comment résoudre le problème des répétitions? L'idée de choisir parmi $k \cdot n$ éléments ne mène à aucun résultat parce que les éléments d'un ensemble de choix peuvent être obtenus de façon différente ce qui augmente faussement le nombre des possibilités de choix. Donc ça ne va pas de cette manière. Pour venir à bout de la chose on doit aller chercher plus loin:

A cette intention nous introduisons $k - 1$ nouveaux éléments J_1, J_2, \dots, J_{k-1} , les incluons dans l'ensemble \mathcal{M}_n . Ainsi nous recevons un nouveau ensemble $\mathcal{M}_n^{k-1} = \{1, 2, 3, \dots, n, J_1, J_2, \dots, J_{k-1}\}$ à $n + k - 1$ éléments. La disposition standard nouvellement valable correspond à l'énumération des éléments donnée ici: Les J_i sont ajoutés derrière d'après les numéros. Pour les éléments J_i l'interprétation suivante est valable: J_i est une prescription ou fonction qui opère sur les dispositions standard choisies, dans lesquelles on les trouve elles-mêmes. La prescription donnée par J_i dit: Remplacer le symbole J_i dans une disposition standard choisie par l'élément e_i du même choix après avoir remplacé tous les J_p par $p < i$. Si on effectue tous ces remplacements, on obtient d'un choix primaire la disposition finale standard. Comme il faut choisir k éléments et comme il n'existent que $k - 1$ éléments J_i , dans un choix standard on trouve toujours au moins un élément $e_j \in \mathcal{M}_n$, dans notre cas un des nombres naturels $1, 2, 3, \dots, n$. J_i effectue par conséquent toujours un remplacement par un élément qui est situé plus à l'avant dans la disposition standard, donc une duplicata. Comme chaque élément peut être choisi une fois ainsi et après peut être dupliqué par un J_i au maximum $k - 1$ fois, il existe la possibilité que chaque élément de \mathcal{M}_n peut se trouver donc k fois dans la disposition standard finale. De cette façon peuvent être obtenues toutes les combinaisons avec répétition.

Exemple:

Soit donné $\mathcal{M}_7 = \{1, 2, 3, \dots, 7\}$. Dans cet ensemble il faut choisir 5 éléments avec répétitions. Il vaut donc: $\mathcal{M}_7^{5-1} = \mathcal{M}_7^4 = \{1, 2, 3, \dots, 7, J_1, J_2, J_3, J_4\}$.

Si on choisit par exemple $(1, 5, 7, J_1, J_4)$ (dans la disposition standard), il faut remplacer comme suit: D'abord $J_1 \mapsto 1$ (l'indice 1 est plus petit que l'indice 4). Ça donne $(1, 5, 7, 1, J_4)$ dans la disposition non-standard et $(1, 1, 5, 7, J_4)$ dans la nouvelle disposition standard. Alors on remplace $J_4 \mapsto 7$ ce qui mène à la disposition standard $(1, 1, 5, 7, 7)$.

Semblablement le choix $(4, J_1, J_2, J_3, J_4)$ mène à la disposition standard $(4, 4, 4, 4, 4)$ après tous les remplacements.

Au choix de 6 éléments dans \mathcal{M}_8 le choix primaire mène à la disposition standard finale $(2, 3, 7, 8, J_2, J_4)$.

Ces exemples montrent qu'un choix primaire correspond clairement à une disposition standard finale. Le nombre des dispositions primaires et sélectionnables est égal au nombre des dispositions standard finales, dans lesquelles tous les éléments figurent répétés jusqu'à k fois. Pour trouver $\bar{C}(k, n)$, on doit donc trouver le nombre des dispositions standard primaires sélectionnables. Là, k éléments sont choisis entre $n+k-1$ éléments $1, 2, 3, \dots, n, J_1, J_2, \dots, J_{k-1}$. Par conséquent on trouve $\bar{C}(k, n) = C(k, n+k-1)$. On a donc:

Théorème 3.10

Combinaisons avec répétitions:

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

Exemple

Un chef de rayon dirige 19 ingénieurs desquels chacun est capable d'avoir la responsabilité pour un projet. 8 nouveaux projets sont à faire (en suspens), qui doivent être traités vraisemblablement l'un après l'autre. Combien de possibilités s'offrent au chef de rayon de nommer des responsables pour les projets, si on peut tirer en considération dans le cas extrême, que le même ingénieur assume (dirige) tous les 8 projets?

Ici, il s'agit d'une combinaison avec répétitions. Dans un ensemble de 19 ingénieurs, 8 responsables sont choisis de façon que chacun peut être nommé plusieurs fois. Il est donc:

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

Nouvel exposé allemand voir:

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

3.4 Exercices

On trouve des exercices dans *DIYMU*, Wirz, Bibl. A18, ainsi que dans la littérature scolaire classique pour le niveau gymnasial ou spécialement aussi dans les manuels du calcul des probabilités et statistiques.

Kapitel 4

Calcul des probabilités

4.1 Introduction

4.1.1 Problème

On cherche un instrument qui permet de faire des proposition concernant des événements de masse.

Exemples d'événements de masse:

1. Grand de nombre d'individus
2. Grand nombre d'apparences individuelles temporellement échelonnées
3. Grand nombre d'apparences individuelles échelonnées dans l'espace
4. Grand nombre de possibilités
5. etc.

En outre nous cherchons des notions qui sont utiles pour mesurer ces **ensembles d'événements**.

4.1.2 Application

Les sciences, dans lesquelles de tels événements de masse sont objet de la considération, sont par exemple les suivantes:

1. Les statistiques
2. La science de la prévision
3. La technique de l'optimisation de dispositions (par exemple la disposition de données sur les supports de données)

4. Créer des modèles en économie, mathématiques financière, théorie d'assurance, technique, médecine etc.. (par exemple prédictions de mortalité, comportement de bourse, chances de thérapie etc.)
5. etc.

Le moyen mathématique, sur lequel les théories mentionnées se fondent, est le **calcul des probabilités**.

Exemple de la statistique::

Quant à la description simple des événements de masse, on se sert de la **statistique descriptive**.

Si par contre on veut juger ou comparer des événements de masse ou bien étudier leur comportement à l'aide d'un modèle, on utilise les **statistiques mathématiques**, statistiques qui servent à juger ou affirmatives (statistiques de test, statistiques exploratives, tests de modèles, par exemple comme régressions etc.). Base des statistiques mathématiques est le **calcul des probabilités**.

Idée importante: On cherche des méthodes mathématiques qui admettent une conclusion sûre ou bien fiable du comportement de peu d'individus (échantillons prélevés) sur la totalité des individus (lots, beaucoup d'individus) sans dépenser trop de travail.

Aujourd'hui les méthodes statistiques font partie de la culture générale pour l'ingénieur. Elles sont importantes pour le contrôle de la qualité et aussi, quant aux composantes techniques, dans la théorie de la fiabilité.

4.1.3 Personnages

Les personnages de l'histoire du calcul des probabilités:

- Blaise Pascal (1623 – 1662)
- Pierre Fermat (1601 – 1665)
- Christian Huygens (1629 – 1695)
- Jakob Bernoulli (1654 – 1705) Déf. de la probabilité
- Abraham Moivre (1667 – 1754) Déf. de la probabilité
- Pierre Simon Laplace (1749 – 1827) Déf. de la probabilité
- 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) Déf. de la probabilité
- Alexander Jakowlewitsch Chintchin (1894 – 1959)

4.2 Expérience de hasard, événement

Notion: Événement au hasard:

Deux Interprétations sont possibles

1. **Hasard relatif:** Processus, qui a une raison mais dont la raison est inconnue.

2.

Hasard relatif absolu: Processus, qui n'existe que de lui-même et qui en principe n'a aucune raison, et qui ne dépend donc de rien¹.

Le problème de l'existence du hasard absolu est un problème de la philosophie. Car si une chose n'est pas connue, il n'est pas possible d'en conclure de son inexistence.

Définition:**Expérience de hasard ou observation de hasard:**

Processus qui est réalisé d'après des prescriptions fixes (ensemble d'événements), réitérable aussi souvent qu'on veut, avec un résultat non-définissable d'avance et qui a plusieurs résultats possibles qui s'excluent les uns les autres. ("Dépendant du hasard".)

Exemple:

Jouer aux dés, jeter une monnaie, tirage de cartes à jouer, mesurer la taille des gens, mesurer la quantité de lait au'une vache produit, tirer des boules colorées d'une urne etc..

Façon de dire:

L'événement "dépend du hasard" := Le résultat n'est pas prévisible.

Exemple:

Jouer au dés et faire un six.

La fonction de répartition ne fournit ici normalement aucun maniement ou aucune caractéristique, d'après lequel on pourrait distinguer un événement parmi d'autres semblables. Par exemple jouer aux dés un 6 et jouer un 5. Par manque de connaissances, donc par des raisons pragmatiques, on doit conséquemment parler de mêmes chances pour de tels événements.

Définition:**Résultat élémentaire:**

Résultat d'une expérience de hasard possible (ω_i) qui ne peut plus être partagé en des résultats partiels. (Mögliche Zerlegung verändert Experiment.)

Exemple: Jouer aux dés avec un dé \rightsquigarrow 6 possibilités: $R(1), R(2), \dots, R(6)$. Ou jouer aux dés avec trois dés. Décomposition: Jouer aux dés avec un dé \rightsquigarrow expérience nouvelle et **atomique**.

Qualité importante:

Un résultat élémentaire obéit à la logique bivalente: Il peut se réaliser ou ne pas se réaliser. (Décomp. change exp..)

Définition:

Ensemble des résultats élémentaires: Ensemble de tous les résultats élémentaires $\Omega = \{\omega_i \mid i = \dots\}$.

¹Pour beaucoup de théologiens c'est la qualité de Dieu.

Remarque:

Au lieu de dire ensemble des résultats, nous disons aussi **ensemble universel** de l'expérience ou **ensemble fondamental**.

Définition:**Événement:**

Sous-ensemble de l'ensemble des résultats d'une expérience de hasard.

Définition:

Événement élémentaire lié au résultat ω_i : $\{\omega_i\}$

Les événements sont donc des ensembles d'événements élémentaires.

Définition:**Espace d'événements:**

Ensemble de tous les événements ($\{\text{résultats}\}$) d'une expérience de hasard. (= $\mathcal{P}(\Omega)$, Ω = ensemble fondamental, ensemble des résultats, \mathcal{P} = ensemble de parties.)

Façon d'écrire:

Pour les événements, nous utilisons des lettres majuscules (A, B, C, ..., X, Y, ...) comme pour les ensembles.

Exemple 1: $A = \{\text{jouer un 6 aux dés}\} = \{\omega_6\}$.

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

Exemple 2:

Soit donné: Bâton de longueur 1 qui soit brisé au hasard en 3 parties (places x et y , $0 < x < y < 1$).
Résultat: 3 bouts de bâton de longueur x , $y - x$, $1 - y$. \rightsquigarrow Un résultat de cet expérience peut donc être représenté comme point dans le triangle $\{(x, y) \in ((0, 1) \times (0, 1)) \mid x < y\}$.

Remarque:

Dans le premier exemple l'espace des événements ou résultats est **fini**, dans le deuxième exemple il est **infini**.

Lors d'une expérience aléatoire, un résultat possible peut être réalisé ou ne pas être réalisé. Après l'expérience, nous pouvons distinguer l'ensemble des résultats E (**résultats statistiques, événement(s) statistique(s)**) qui ont été réalisés de l'ensemble des résultats $\Omega \setminus E$ qui n'ont pas été réalisés. \rightsquigarrow ω s'est réalisé $\Leftrightarrow \omega \in E$, ω ne s'est pas réalisé $\Leftrightarrow \omega \notin E$

4.2.1 Processus aléatoire, fréquence

Processus aléatoire

Par exemple jouer une seule fois aux dés avec un dé est une **expérience de hasard atomique** qui n'est plus décomposable. À côté de telles expériences de hasard atomiques, on trouve dans la pratique aussi des processus de hasard plus compliqués qui sont composés par exemple de plusieurs expériences de hasard atomiques qui se déroulent l'une après l'autre ou simultanément. Par exemple une urne remplie de boules colorées, de laquelle on tire 3 boules l'une après l'autre. A un premier niveau, nous tirons par exemple une boule sur 6 boules disponibles, à un 2. niveau on tire une boule sur 5 boules disponibles et à un 3. niveau une boule sur 4 boules disponibles. Ici un événement élémentaire de l'expérience de hasard consiste en 3 résultats atomiques qui se suivent.

Façon de dire:

Quant aux expériences de hasard composées, nous appelons le numéro des expériences de hasard atomiques dans une suite le **niveau**. Par conséquent l'expérience totale consiste en plusieurs niveaux.

Définition:

Une expériences de hasard à plusieurs niveaux s'appelle **processus aléatoire**.

Exemple:

Si nous tirons d'abord une boule noire, plus une blanche et ensuite encore une fois une boule noire, nous écrivons $R(n_1, b_2, n_3)$ pour le résultat. Cette expérience de hasard consiste en trois niveaux l'index indique le niveau. L'information du niveau est nécessaire parce que les boules disponibles dans l'urne (condition de l'expérience) changent avec le niveau. Un événement élémentaire d'une expérience au niveau n consiste donc en un multiplé- n arrangé d'événements élémentaires atomiques.

Fréquence

L'événement statistique E soit le résultat d'une expérience (série à n tentatives, n expériences partielles, processus du niveau n). L'événement A soit réalisé k fois à n tentatives partielles ou élémentaires. (n cas possibles et k cas favorables.)

Exemple: 12 fois jouer aux dés: E = ensemble des résultats partiels réalisés et numérotés. Le 1 vient 3 fois. A = sous-ensemble des 1 réalisés et numérotés. $\rightsquigarrow |E| = n = 12$, $k = |A| = 3$.

Définition:

Fréquence absolue de A :

$$H(A) := k$$

Fréquence relative de A :

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

Exemple:

A = tirer un roi d'un jeu de cartes bien mélangé avec 36 cartes et 100 tentatives (tirer 100 fois). Soit $k = 9 \rightsquigarrow H(A) = 9$, $h(A) = 0.09$. Un événement dans ce processus de hasard au niveau 100 est donc un

multiplet-100 arrangé avec 100 événements atomiques, qui sont chaque fois le résultat d'un tirage simple.

Conclusion: Les fréquences absolues sont des puissances d'événements **qui ont eu lieu**.

Exemple:

Expérience: jouer 12 fois au dé. Résultat: $E = \{2_1, 4_2, 5_3, 1_4, 2_5, 6_6, 3_7, 1_8, 1_9, 5_{10}, \cdot_{11}, 3_{12}\}$. Sous-ensemble "1 a eu lieu": $A := \{x_i \in E \mid x_i = 1_i\} = \{1_4, 1_8, 1_9\}$. ($\cdot_{11} = \cdot =$ "missing value".)
 $\rightsquigarrow H(A) = 3, h(A) = \frac{3}{11}$.

Façon de dire:

Le sous-ensemble A de E objet de notre intérêt particulier se dirige, s'appelle ici **ensemble de but**.

Il vaut donc:

Théorème: Hyp.:

Thè.:

1. $H(A) = |A|$
(Fréquence absolue = puissance)
2. $h(A) = \frac{H(A)}{H(E)} = \frac{|A|}{|E|}$
(Fréquence relative)
3. $h(E) = 1$

On reconnaît aussitôt:

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

4.3 Algèbre des événements

4.3.1 Algèbre des événements et algèbre des ensembles

Comme les événements sont des ensembles, l'algèbre des événements n'est rien d'autre que l'algèbre des ensembles.

Soient A et B des événements (\rightsquigarrow ensembles), qui peuvent se réaliser lors d'une expérience projetée.

Définition:**Événement somme:**

$A + B$ ou $A \cup B :=$ événement total, est réalisé exactement si A ou B ou les deux ensembles sont réalisés. (\rightsquigarrow Réunion de A et B .)

Définition:**Événement produit:**

$A \cdot B$ ou $A \cap B :=$ événement total, est réalisé exactement si A et B sont réalisés. (\rightsquigarrow Intersection de A et B .)

Soit Ω l'ensemble de tous les résultats élémentaires (ensemble universel, événement somme de tous les événements atomiques, est toujours réalisé) et $\{\}$ l'ensemble vide (n'est jamais réalisé).

Définition:

Un événement qui est toujours réalisé (Ω) s'appelle **événement sûr**. Un événement qui n'est jamais réalisé ($\{\}$) s'appelle **événement impossible**.

Définition:**Événement entraîné:**

B s'appelle **entraîné** par A , si la réalisation de A a toujours comme conséquence la réalisation de B ($B \supseteq A$).

Définition:**Événement complémentaire:**

$\bar{A} = A^c$ s'appelle **événement complémentaire** de A , si $A \cup \bar{A} = \Omega \wedge A \cap \bar{A} = \{\}$.

Définition:**Événements équivalents:**

A et B s'appellent **équivalents**, si la réalisation de A a toujours comme conséquence la réalisation de B et vice versa. $\rightsquigarrow (B \supseteq A) \wedge (A \supseteq B)$.

Définition:**Événements s'excluant mutuellement:**

A et B s'appellent **s'excluant mutuellement (incompatibles)**, si A et B ne sont jamais réalisés simultanément. ($A \cap B = \{\}$.)

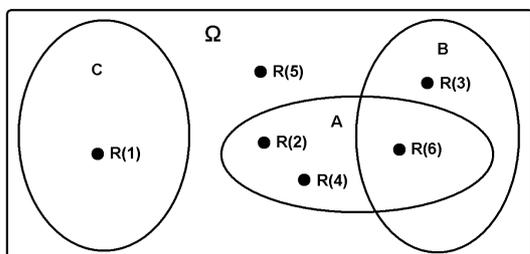
Exemple 1:

L'expérience de hasard "un enfant va naître": \rightsquigarrow Deux possibilités: événement $A = \{ \text{Ça va être un garçon} \}$. L'événement $B = \{ \text{Ça va être une fille} \}$.

$\rightsquigarrow A \cup B =$ événement sûr, $A \cap B =$ événement impossible. A, B sont s'excluent mutuellement.

Exemple 2:

$A = \{\text{jouer aux dés un nombre pair}\}$ (bref $R(2 \vee 4 \vee 6)$), $B = \{\text{jouer aux dés un nombre divisible par } 3\}$ (bref $R(6 \vee 3)$), $C = \{\text{jouer aux dés le nombre } 1\}$ (bref $R(1)$).



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

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

Remarque:

Une algèbre est un ensemble avec opérations sur les éléments. L'ensemble de parties de Ω forme par conséquent une algèbre avec des opérations sur leurs sous-ensembles qui sont des événements. Les opérations sont $\cup, \cap, \overline{\cdot}, \subset$. Nous parlons donc ici d'une **algèbre d'événements**.

4.3.2 Algèbres de Boole

L'algèbre de Boole est traitée dans Wirz, Bibl. A14 (Mathematikkurs für Ingenieure, Teil 4, Einführung in die Boolesche Algebra). Ici nous résumons seulement l'application concernant l'algèbre des ensembles importante dans ce cours.

Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$, $\{\cdot\} \in \mathcal{A}$. ($\mathcal{P}(\Omega)$ = ensemble de parties)

Nous définissons pour nos besoins:

Définition:

\mathcal{A} s'appelle **algèbre de Boole** (algèbre des ensembles de Boole), si l'on a:

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

Il vaut (De Morgan!):

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

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

Conclusions:

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

Problème:

Soit Ω fini. Par conséquent $\mathcal{P}(\Omega)$ est aussi fini et donc à former sans problèmes. $\mathcal{P}(\Omega)$ est un ensemble

de sousensembles de Ω et par conséquent un **ensemble d'événements**. Mais comment former $\mathcal{P}(\Omega)$, si Ω est infini? Et est-ce que $(\bigcup_{i=1}^{\infty} A_i) \in \mathcal{A}$ reste toujours valable?

Combien souvent dans les mathématiques nous éludons le problème avec une définition:

Définition:

Une algèbre des ensembles de Boole s'appelle σ -**algèbre**, s'il vaut:
 $(A_i \in \mathcal{A}, i = 1, \dots, \infty) \Rightarrow (\bigcup_{i=1}^{\infty} A_i) \in \mathcal{A}$

A cause de De Morgan il vaut donc:

Conclusion:

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

Remarque:

Si Ω est infini, \mathcal{A} est généralement un sous-ensemble de $\mathcal{P}(\Omega)$. Car les événements de \mathcal{A} sont normalement des ensembles finis, $\mathcal{P}(\Omega)$ possède par contre dans ce cas aussi des ensembles infinis.

Conséquence:

A une expérience aléatoire nous associons pour le moment toujours un couple (Ω, \mathcal{A}) . (Ensemble des résultats Ω et algèbre des événements (Ω, \mathcal{A}) .)

4.3.3 Quant à la puissance et à la fréquence

Soient maintenant A et B des sous-ensembles du même ensemble de résultats E . De l'algèbre des ensembles nous pouvons conclure:

Théorème:

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. $A \cap B = \{\} \Rightarrow H(A \cup B) = H(A) + H(B)$ (Événements incompatibles)
4. $A \cap B = \{\} \Rightarrow h(A \cup B) = h(A) + h(B)$ (Événements incompatibles)
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)

Exemple:

Expérience: jouer 16 fois aux dés avec 2 dés.

\rightsquigarrow Résultat:

$$E = \{(2, 3)_1, (2, 5)_2, (1, 3)_3, (4, 6)_4, (1, 6)_5, (1, 6)_6, (2, 2)_7, (4, 6)_8, (4, 6)_9, (5, 6)_{10}, (1, 1)_{11}, (2, 6)_{12}, (2, 5)_{13}, (1, 1)_{14}, (1, 5)_{15}, (3, 4)_{16}\}$$

A: Au moins un nombre est divisible par 3.

B: Les deux nombres sont pairs.

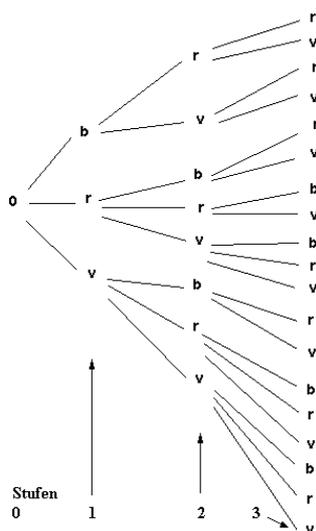
$$\leadsto H(E) = 16, H(A) = 10, H(B) = 5, H(A \cap B) = 4, H(A \cup B) = 11, H(\bar{A}) = 6, \dots$$

4.3.4 Des arbres d'événements

Pour obtenir une vue d'ensemble sur un processus aléatoire du niveau total n , la représentation des événements par un arbre d'événements est favorable.

Exemple:

Soit donnée une urne qui contient une boule bleu (b), deux boules rouges (r) et trois boules violetes (v). Par un processus aléatoire du niveau total trois, il faut tirer trois boules l'une après l'autre. Les événements doivent être représentés dans un arbre d'événements. Combien de possibilités est-ce qu'il y a?



Exemple:

$b - r - r$ Est un résultat.
Il y a 19 résultats différents (Arrangements: Combinaison avec répétition dépendant de l'élément et rangement).

Important:

Comme la répétition des éléments choisis est dépendante des éléments, aucune des 6 formules classiques de l'analyse combinatoire ne peut être appliquée directement pour calculer le nombre de possibilités.

4.4 Probabilité classique d'après Laplace

4.4.1 Expériences de Laplace, probabilité identique

Nous allons étudier les **expériences de Laplace**. Ce sont des expériences avec un ensemble de résultats fini où on peut observer qu'à un nombre suffisamment grand de répétitions tous les événements élémentaires ont à peu près la même fréquence.

P. ex. on a observé en jetant d'une monnaie (Kreyszig, Bibl. A9, Statistische Methoden und ihre Anwendungen, p. 58):

Expérimentateur	Nombre de jets	Face (tête) réalisée	$h(R(\text{tête}))$
Buffon	4040	2048	0.5069
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

Ici, on peut donc parler d'une expérience de Laplace. On parle aussi de la **régularité statistique** ou de la **stabilité de la fréquence relative**. On fait des expériences pareilles avec un dé idéal.

L'idée de l'expérience de Laplace peut être définie plus exactement par l'idée de la probabilité égale:

Définition:

Si à une expérience de hasard le comportement d'aucun événement n'est distingué du comportement des autres, on appelle tous les événements **de même probabilité** ou **de même possibilité**. (Probabilité identique.)

Conclusion:

Par conséquent quant à une expérience de Laplace les événements élémentaires sont de même probabilité.

Exemple:

Expériences de Laplace: Tirer des cartes dans un jeu de cartes, jouer aux dés avec des dés idéaux, tirer des boules d'une urne etc..

4.4.2 Probabilité comme mesure pour gagner

Exemple: Quant à un dé idéal, tous les 6 événements sont de la même probabilité (expérience de Laplace).

Raisonnement de Laplace:

Problème: Dans un jeu de dés avec un dé, un joueur gagne, s'il fait un 3 ou un 6. Sinon il perd. Quels sont ses chances de gagner?

Au jeu mentionné, il y a $m = 6$ événements possibles (**cas possibles**, ensemble d'événements $\Omega = \{1, 2, 3, 4, 5, 6\}$, écrit brièvement). Pour gagner cependant il n'y a que $g = 2$ événements (**cas favorables**). $m - g = 4$ cas sont donc défavorables pour gagner (nombres 1, 2, 4, 5), ensemble d'événements $A = \{3, 6\}$.

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

Comme tous les événements sont de la même probabilité, on peut prendre comme mesure P pour la chance de gagner la fréquence relative:

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

Classiquement on définit cette mesure pour la chance de gain (aux événements élémentaires de même probabilité) comme **probabilité** de réaliser l'événement A :

Définition:

Probabilité (d'après Laplace) $P(A)$ de l'événement A lors d'une expérience avec des événements élémentaires de même probabilité:

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

Exemple:

Processus aléatoire: Jouer aux dés avec 3 dés et 2 essais (niveau total 2). Gagner signifie: Faire au moins une fois (donc par niveau) 2 fois un 6 (événement A). $\rightsquigarrow P(A) = ?$

Solution:

D'abord nous décomposons le problème en problèmes partiels, dont nous joignons finalement les résultats.

1.

L'événement A_1 soit défini par le succès dans le 1-er essai (niveau 1), avec la restriction que le résultat du 2-ème essai (niveau 2) peut être quelconque. Au lieu de considérer les dés, nous considérons les places P sur lesquelles les résultats sont notés. Il y a 3 possibilités de faire un résultat égal 6 avec deux dés et inégal 6 avec le 3-ème dé restant ($E_{1,1}$, $E_{2,1}$, $E_{3,1}$). En plus il y a encore la possibilité de faire un 6 avec tous les 3 dés ($E_{4,1}$).

$g_{j,k}$ = Nombre de possibilités favorables à l'événement $E_{j,1}$ à la place numéro k .

Évén.	Niveau 1			Niveau 2		
	$P = 1$	$P = 2$	$P = 3$	$P = 4$	$P = 5$	$P = 6$
$E_{1,1}$	$g_{1,1} = 1$	$g_{1,2} = 1$	$-6 : g_{1,3} = 5$	$g_{1,4} = 6$	$g_{1,5} = 6$	$g_{1,6} = 6$
$E_{2,1}$	$g_{2,1} = 1$	$-6 : g_{2,2} = 5$	$g_{2,3} = 1$	$g_{2,4} = 6$	$g_{2,5} = 6$	$g_{2,6} = 6$
$E_{3,1}$	$-6 : g_{3,1} = 5$	$g_{3,2} = 1$	$g_{3,3} = 1$	$g_{3,4} = 6$	$g_{3,5} = 6$	$g_{3,6} = 6$
$E_{4,1}$	$g_{4,1} = 1$	$g_{4,2} = 1$	$g_{4,3} = 1$	$g_{3,4} = 6$	$g_{3,5} = 6$	$g_{3,6} = 6$

A_1 est réalisé, si E_1 ou E_2 ou E_3 ou E_4 est réalisé.

$A_1 = E_1 \cup E_2 \cup E_3 \cup E_4 \Rightarrow m(A) = g = 5 \cdot 6 \cdot 6 \cdot 6 + 5 \cdot 6 \cdot 6 \cdot 6 + 5 \cdot 6 \cdot 6 \cdot 6 + 1 \cdot 6 \cdot 6 \cdot 6 = (3 \cdot 5 + 1) \cdot 6^3$.

Comme sur chaque place 6 nombres sont possibles, il vaut: $m = 6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 = 6^6$.

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

2.

L'événement A_2 soit défini par le succès dans le 2-ème essai, avec la restriction que le résultat du 1-er essai peut être quelconque. Pour des raisons de symétrie, nous obtenons le même résultat comme dans le 1-er essai.

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

3. Soit $A_3 = A_1 \cap A_2$. A_3 signifie qu'on a du succès au 1-er et au 2-ème essai (niveau).

Pour le succès sur les places 1, 2, 3 (1-er niveau) il y a $3 \cdot 5 + 1 = 16$ possibilités. Pour chacune de

ces 16 possibilités il existe encore 16 possibilités pour le succès sur les places 4, 5, 6 (2^e-ème niveau). Par conséquent pour le succès sur deux niveaux (A_3) il y a $16 \cdot 16$ possibilités.

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

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

Attention:

L'idée de la probabilité identique des événements élémentaires manque normalement dans la pratique: Les dés ne sont pas exactement symétriques et homogènes, la psychologie du lanceur joue un rôle, et beaucoup de processus statistiques ne sont pas du tout des expériences de Laplace. Qu'on pense par exemple au temps! Par conséquent une notion de probabilité dans un sens plus large est nécessaire. Par conséquent nous introduisons la notion de la **probabilité axiomatique** d'après **Kolmogoroff**.

4.5 Probabilité axiomatique

4.5.1 Notion, axiomes, conclusions

Notre point de départ est l'expérience: Les résultats d'expériences de hasard suivent à la longue certaines lois. D'après Descartes, il est raisonnable d'accepter comme loi de la nature le modèle plus simple vraisemblable de l'événement observé (phénomène de nature), si d'autres observations ne nous forcent pas de changer le modèle. Un modèle dans ce sens n'est cependant pas une explication causale des phénomènes de la nature, il est seulement une description de la nature.

Dans ce qui suit, nous voulons nous restreindre à des expériences ou des processus qui montrent des fréquences relatives qui sont par le nombre croissant statistiquement stables, si le nombre des essais est assez grand. Pourtant il n'est pas nécessaire de présupposer que les événements élémentaires sont de même probabilité. Des irrégularités accidentelles sont permises seulement pour les événements élémentaires, mais non pas pour la fréquence relative dépendante du numéro de l'expérience, si le numéro de l'expérience est assez grand.

Hypothèse fondamentale:

Sur la base des observations empiriques, on peut supposer **l'hypothèse** suivante: Soit n le nombre d'essais dans une expérience à des conditions invariables. Donc $\lim_{n \rightarrow \infty} h_n(A) = h_\infty$ est un nombre réel $\in [0, 1]$ qui peut être indiqué avec une grande certitude pratique.

Attention: Comme dans la réalité du temps un nombre fini d'essais ne peut jamais exister, cette supposition ne peut jamais être contrôlée dans la pratique. Sa nature est purement théorique. Ici, le succès pratique prend la place de la certitude pratique de l'hypothèse. Les statistiques reposant sur ce fait ne sont donc, dans leurs bases, pas des sciences déductives, mais inductives (expérimentales).

Remarque:

Pour obtenir h_∞ , on a besoin de données, ç.v.d. de l'expérience.

La probabilité qui va être utilisée maintenant est fondée sur une interprétation hypothétique que nous résumons dans une définition. Elle est de nouveau une sorte de mesure d'un événement.

Définition: **Probabilité** ou **mesure de probabilité** (définition statistique):

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

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

Conséquence: P est une fonction sur \mathcal{A} .

Remarque:

Comme dans la réalité pratique il ne existe pas de méthode pour obtenir h_∞ , nous posons $P(A) := h_\infty \approx h_n$, $n \ll 1$ (n très grand).

\leadsto **Problème:** Stabilité de h_n avec n .

Comme Ω est souvent inconnu dans la pratique, nous écrivons S à la place de Ω .

Axiomes:

1. $P(A) \in [0, 1]$ univoque
2. (a) $A = S \Rightarrow P(A) = 1$ (événement sûr)
- (b) A et B événements équivalents $\Rightarrow P(A) = P(B)$
- (c) A et B événements incompatibles:
 $A \cap B = \{\}$ $\Rightarrow P(A \cup B) = P(A) + P(B)$

Définition: $\bar{A} = S \setminus A$
s'appelle **événement complémentaire** de A .

Remarque:

1. $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$ à cause de
 $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ (\emptyset)
(incompatibles)
 $\leadsto P(A \cup B) \leq P(A) + P(B)$
2. Soit $A \subseteq B \Rightarrow B = A \cup (B \setminus A)$
 $\Rightarrow P(B) = P(A) + P(B \setminus A) \Rightarrow P(B) \geq P(A)$
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. $P(S) = P(\bar{S}) = 1 - P(S) = 1 - P(\{\}) = 1 - 0 = 1$

Les conclusions suivantes sont maintenant directement compréhensibles:

Conclusions:

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

Attention:

1. $0 = P(A) = h_\infty(A) \approx h_n(A), n \gg 1$
 \rightsquigarrow
 $\rightsquigarrow A$ n'est pratiquement jamais réalisé ce qui ne signifie pas que A soit impossible!
2. $1 = P(A) = h_\infty(A) \approx h_n(A), n \gg 1$
 \rightsquigarrow
 $\rightsquigarrow A$ est pratiquement toujours réalisé ce qui ne signifie pas que A soit absolument sûr!

4.5.2 La notion espace de probabilité

Nous considérons un ensemble d'événements discret $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$ lié à une expérience de hasard. Soit $P(\omega_i) = p_i \geq 0$ pour les événements élémentaires ω_i . Pour $i \neq k$ il doit être $\omega_i \cap \omega_k = \{\}$, sinon les événements élémentaires seraient démontables ($\omega_i = (\omega_i \cap \omega_k) \cup (\omega_i \setminus \omega_k)$). Par conséquent il suit des axiomes: $\sum_{i=1}^{\infty} p_i = 1$. Comme chaque événement se compose d'événements élémentaires ($A = \bigcup_{i_1}^{\dots} \omega_i$), il vaut pour la probabilité de A : $P(A) = \sum_{i_1}^{\dots} p_i$. Avec cela, à chaque événement $A_k = \bigcup_{k_1}^{\dots}$ il est adjoint clairement une probabilité $P(A_k) = \sum_{k_1}^{\dots} p_k$. Cette correspondance (assignation) est une fonction $A_k \mapsto P(A_k)$. Cette fonction transforme l'espace d'événements en un **espace de probabilité**.

Définition:

Un **espace de probabilité** est donné par un espace d'événements Ω , une algèbre d'événements et une mesure de probabilité.

On voit tout de suite que dans le cas de $\Omega = \{\omega_1, \omega_s, \dots, \omega_n\}$ avec $P(\omega_1) = P(\omega_2) = \dots = P(\omega_n)$ on obtient la probabilité de Laplace.

Façon de dire:

4.5.3 Probabilité conditionnelle

La notion

Exemple:

Soient A et B les ensembles suivants:

$$A = \{R(\text{ Client est féminin})\}$$

$$B = \{R(\text{ Client achète notre produit})\}$$

$$\rightsquigarrow B|A = \{R(\text{ Client féminin achète notre produit})\}$$

Façon d'exprimer:

$$B|A = \rightsquigarrow$$

Événement B sous la condition de l'événement A , B dépendant de A .

Remarque:

$B|A \rightsquigarrow$ Client est féminin **et** client achète \rightsquigarrow intersection des ensembles!

Problème:

Nous cherchons maintenant la probabilité que le client achète notre produit, à la condition que le client soit féminin.

Définition:

$P(B|A) :=$ probabilité pour la réalisation de B à la condition que A ait été réalisé \rightsquigarrow **probabilité à condition.**

Comment calculer $P(B|A)$? Nous considérons d'abord un ensemble d'événements Ω fini, , $A \neq \{\}$:

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

Les résultats de $A \cap B$ sont réalisés dans s cas. Cela arrive à condition que les résultats de A soient réalisés ce qui arrive dans a cas. A est par conséquent l'ensemble des événements sûrs par rapport à ce problème, c.-à.-d. l'ensemble universel pour l'événement $A \cap B$. Par conséquent vaut:

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

A cause de $A \cap B = B \cap A$, la loi qu'on vient de déduire est symétrique: $P(A|B) = \frac{P(B \cap A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$
 \rightsquigarrow

Théorème:

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

Conséquence: $P(A \cap B) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$
(Théorème de la multiplication)

P.ex. pour $A = \{\}$ cette formule reste correcte.
 ($\Rightarrow A \cap B = \{\} \Rightarrow P(A \cap B) = 0, P(A) = 0, P(B|A) = 0$)

$P(A \cap B) = P(B|A) \cdot P(A)$ le laisse itérer:

Corollaire:

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

Exemple 1:

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

Il vaut: $|A \cup B| = |A| + |B| - |A \cap B|$

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

Exemple 2:

Donné: Boîte avec 3 boules vertes et 7 boules rouges (\rightsquigarrow en tout 10). Probabilité qu'on n'obtient que des boules rouges en tirant 2 fois sans remettre les boules tirés?

Soit $r_1 =$ "rouge au 1er tirage", $r_2 =$ "rouge au 2ème tirage".

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

Exemple 3:

Lotto: Probabilité pour 6 nombres corrects de 49?

Méthode 1: Combinaisons

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

Méthode 2: Événements dépendants

$A_1 \rightsquigarrow$ Tirage du 1er nombre: $P(A_1) = \frac{6}{49}$

$A_2 \rightsquigarrow$ Tirage du 2ème nombre à la condition que le 1er nombre soit tiré: $P(A_2|A_1) = \frac{5}{48}$

$A_3 \rightsquigarrow$ Tirage du 3ème nombre à la condition que le 1er et le 2ème nombre soient tirés:

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

\vdots

$$\rightsquigarrow P(A_1 \cap A_2 \cap \dots \cap A_6) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap A_2 \cap \dots \cap A_5)$$

$$= \frac{6}{49} \cdot \frac{5}{48} \cdot \frac{4}{47} \cdot \dots \cdot \frac{1}{44} = \frac{1}{13983816} \approx 7.15112 \cdot 10^{-8}$$

Événements indépendants

Si les résultats ou les éléments dans l'ensemble $B|A$ sont les mêmes que ceux dans l'ensemble B et par conséquent s'il vaut $B|A = B$, l'ensemble A n'influence pas quels éléments de $B|A$ doivent être omis au choix dans B . Donc A ne change pas B . Par conséquent nous appelons B et A indépendants. On obtient par conséquence $P(B|A) = P(B)$ et donc

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

On contrôle tout de suite que le résultat soit aussi correct pour $A = \{\}$ ou $B = \{\}$.

Définition:

B s'appelle **indépendant** de A ou **indépendant de façon stochastique (aléatoire)** s'il vaut: $B|A = B$.

Théorème:

Hyp.:

Soit B indépendant de A .

Thè.:

1. $P(B|A) = P(B)$
2. $P(A \cap B) = P(A) \cdot P(B)$
3. A indépendant de B .

Exemple 1:

Donné: Jeux de cartes avec 36 cartes. Tirer 2 fois une carte et les remettre dans le jeu. Probabilité qu'on tire 2 fois un as?

Solution:

a_1 : As au 1er tirage.

a_2 : As au 2er tirage.

$\rightsquigarrow P(A_1) = P(A_2)$

Les deux événements $A_1 = \{a_1\}$ et $A_2 = \{a_2\}$ sont indépendants. $\rightsquigarrow P(A_1) = P(A_2)$

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

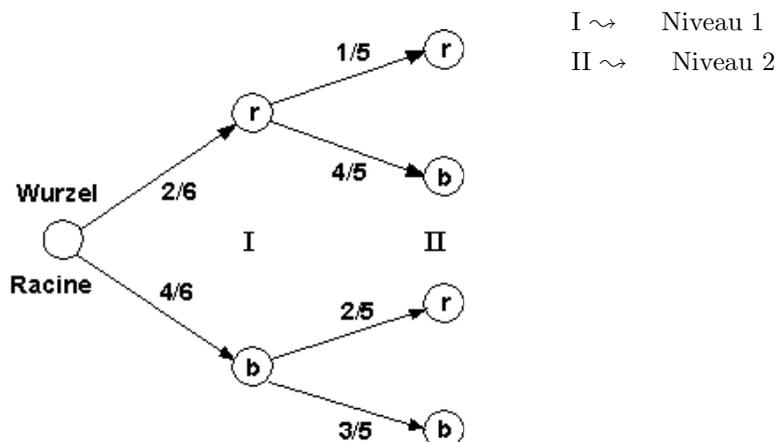
Utiliser l'arbre d'événements

Exemple 2:

Urne qui contient 6 boules: 2 rouges (r), 4 bleues (b). Tirer 2 boules sans les remettre.

1. A :
Probabilité qu'on tire 2 fois une boule avec la même couleur?
2. B :
Probabilité qu'on tire 2 fois une boule avec une couleur différente?

Solution: Utiliser un arbre d'événements:

**Important: Méthode**

Considérer: Le long des flèches (sentiers) il s'agit d'événements dépendants, c.-à.-d. de probabilité conditionnelle. À un point de bifurcation, une condition est posée pour le niveau suivant. Par conséquent les probabilités se multiplient le long des sentiers. Par contre les événements à un niveau donné sont exclusifs. Les probabilités s'additionnent.

Solution:

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

Exemple 3:

Quelle est la probabilité de réaliser la première fois pile au k -ème essai en lançant une monnaie idéale? Quelle est la somme sur k de toutes ces valeurs de probabilité?

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

4.5.4 Probabilité totale**Exemple:**

Lors d'une expérience aléatoire, nous tirons d'une de trois urnes données U_i une boule. U_1 contient deux boules rouges et deux bleues, U_2 deux rouges et une bleue, U_3 une rouge et deux bleues.

Questions:

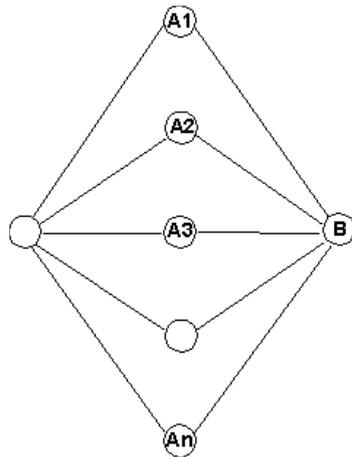
1. Quelle est la probabilité de tirer une boule rouge?
2. Quelle est la probabilité que la boule rouge a été tirée de U_2 ?

Solution:

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

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

Évidemment on s'attend la valeur 0.5 lors de cette expérience, parce que en tout les boules rouges font exactement la moitié des boules. Mais l'exemple montre la méthode qui fonctionne aussi dans d'autres cas.



Si nous mettons B à la place de $\{r\}$ et A_i à la place de U_i , nous obtenons la formule suivante:

Conséquence: Soient $i \neq k$

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

Définition: Ici $P(B)$ s'appelle la **probabilité totale** pour la réalisation de l'événement B .

Maintenant nous nous intéressons à la probabilité que le sentier traverse A_k à condition que B soit réalisé. La probabilité que le sentier traverse A_k correspond à la probabilité que A_k se réalise. Nous devons donc calculer $P(A_k|B)$. $P(A_i)$ et $P(B|A_i)$ sont donnés.

Nous avons à la disposition les informations suivantes:

1. $P(B) = \sum_i P(A_i) \cdot P(B|A_i)$ (dédit en haut)
2. $P(A_k|B) = \frac{P(A_k \cap B)}{P(B)}$ (théorème de la multiplication)
3. $P(A_k \cap B) = P(A_k) \cdot P(B|A_k)$ (le long d'un sentier)

Conséquence: **Formule de Bayes**

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

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

4.6 Fonctions de répartition

4.6.1 Variables aléatoires

Nous considérons des expériences de hasard. Si nous tirons des boules colorées d'une urne, nous devons d'abord attribuer une échelle graduée aux couleurs pour pouvoir les inscrire sur l'axe d'un diagramme à l'intention de travailler après avec les fonctions aléatoires comme fonctions sur les nombres. Si par contre nous jouons p. ex. aux dés avec un seul dé, les résultats possibles sont déjà les nombres $\in \{1, 2, \dots, 6\}$.

Nous pouvons interpréter les variable réelle comme fonctions comme il suit:

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

La fonction x applique le nombre z à la valeur $z = w(z)$, ce qui est toujours évident et par conséquent jamais mentionné.

De façon analogue on applique, d'après l'intention dans l'expérience, une valeur $X(\omega) \in \mathbb{R}$ par une variable aléatoire à un événement $\omega \in \Omega$ (ou un événement élémentaire $\{\omega\} \in \mathcal{A}$).

Exemple 1:

Dans une expérience de hasard, on joue aux dés avec deux dés. L'ensemble des résultats est $\Omega = \{(m, n) \mid m, n \in \{1, 2, \dots, 6\}\}$. Maintenant la variable aléatoire X soit définie comme suit: $X((m, n)) = m + n \in \mathbb{R}$. Par conséquent il vaut: $X((m, n)) \in \{2, 3, \dots, 12\}$. L'ensemble des événements avec $X((m, n)) < 4$ est donc $\{(1, 1), (1, 2), (2, 1)\}$.

Exemple 2:

Dans une expérience de hasard, on joue aux dés avec un dé. L'ensemble des résultats est $\Omega = \{1, 2, \dots, 6\}$. Maintenant la variable aléatoire X soit définie comme suit: $X(n) = n$. Par conséquent il vaut: $X(\omega_k) = \omega_k \in \mathbb{R}$.

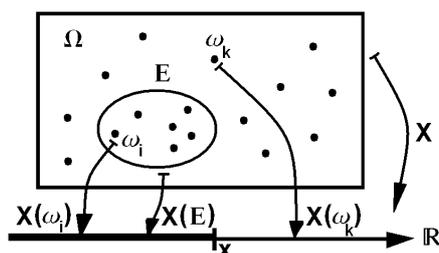
Exemple 3:

Dans une expérience de hasard, on questionne des dames d'un collectif donné sur le périmètre du corps de leur partenaire idéal. Pour la variable aléatoire il vaut donc:

$$X : \text{Dame} \mapsto X(\text{Dame}) = \text{périmètre partenaire idéal} \in \mathbb{R}$$

Remarque:

Dans le 1er exemple $\{X(\dots)\} \subset \mathbb{R}$ est discret ou fini, dans le 3ème exemple $\{X(\dots)\} \subset \mathbb{R}$ est théoriquement infini.



Pour notre emploi à venir, nous définissons maintenant la variable aléatoire plus exactement. Soit donnée une expérience aléatoire avec un ensemble fondamental Ω mené d'une algèbre d'événement \mathcal{A} qui est une σ -algèbre.

Définition:

Une fonction $X : \Omega \mapsto \mathbb{R}$ avec $\omega \mapsto X(\omega)$ s'appelle **variable aléatoire** ou **variable stochastique**, **fonction stochastique**, s'il vaut:

1. $\forall_{x \in \mathbb{R}} \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{A}$
2. Pour \mathcal{A} soit définie une fonction de probabilité P , qui suffit aux axiomes de la probabilité.

La première condition est la condition de la mesurabilité. Elle signifie que l'ensemble de tous les résultats dont la réalisation possède une valeur au-dessous d'une valeur x donnée, est considéré comme événement.

Remarque:

Comme à $\{\omega\}$ et donc aussi à $\{X(\omega)\}$ il s'agit d'événements de hasard, nous disons: La valeur de X **dépend du hasard**.

Façon d'écrire: $X \leq x := \{\omega \in \Omega \mid X(\omega) \leq x\}$
 $\leadsto X \leq x, x < X$ etc. sont des ensembles!

Façon de dire:

P. ex. pour $a \leq X \leq b$ nous disons: X prend une valeur dans $a \leq X \leq b$.

Comme les sous-ensembles de Ω sont des événements, on déduit s'appuyant sur la définition de X et la manière de choisir les éléments tout de suite le théorème:

Théorème:

1. $-\infty < X < \infty$ est l'événement sûr.
2. Les ensembles suivants sont toujours des événements: $a < X < b$, $a \leq X < b$, $a < X \leq b$, $a \leq X \leq b$, $a < X$, $a \leq X$, $X < a$, $x \leq a$
3. $P(X \leq a) + P(a < X) = P(-\infty < X < \infty) = 1$

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

Remarque:

Comme nous avons vu dans les exemples précédents, nous devons distinguer entre **variables aléatoires discrètes** et **continues**. On s'attend à ce que chez nous dans le cas des variables aléatoires continues les domaines de valeurs sont des intervalles. Pour les raisons de puissance d'intervalle et de l'exactitude des mesures, une proposition comme p. ex. $P(X = a)$ ne fait pratiquement pas beaucoup de sens, par contre une proposition comme $P(X < a)$

Exemple 3:

Jouer aux dés avec un dé, $X =$ pile (nombre):

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

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

Exemple 4: Urne, 4 boules rouges et 6 boules bleues. En tirer 2, sans les remettre. X (boules tirées) = nombre de boules rouges tirées.

$$P(X = 0) = \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}$$

$$P(X = 1) = ? \quad r_1 \rightsquigarrow P(Z_1) := P(X = 1|r_1) = \frac{4}{10} \cdot \frac{6}{9} = \frac{4}{15}$$

$$r_2 \rightsquigarrow P(Z_2) := P(X = 1|r_2) = \frac{6}{10} \cdot \frac{4}{9} = \frac{4}{15}$$

$$Z_1 \cap Z_2 = \{\} \Rightarrow P(Z_1 \cup Z_2) = \frac{4}{15} + \frac{4}{15} = \frac{8}{15}$$

$$P(X = 2) = \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15}$$

$$\rightsquigarrow P((X = 0) \vee (X = 1) \vee (X = 2)) = \frac{1}{3} + \frac{8}{15} + \frac{2}{15} = 1$$

$$P(1 < X < 2) = 0$$

$$P(X \leq 1) = \frac{1}{3} + \frac{8}{15} = \frac{13}{15}$$

$$P(X \geq 1) = \frac{8}{15} + \frac{2}{15} = \frac{2}{3}$$

$$P(0.5 < X < 10) = P(x \geq 1) = \frac{2}{3}$$

4.6.2 Fonction de répartition

Lors de l'explication de la variable aléatoire, on a mentionné l'idée de la fonction de probabilité: Pour \mathcal{A} soit définie une fonction de probabilité P , qui suffit aux axiomes de la probabilité. Mais la méthode d'obtenir P de façon précise n'a pas été traitée. C'est pourquoi nous voulons d'abord préciser l'idée de la fonction de répartition:

Soit donnée une expérience de hasard et liée à cela une variable de hasard X . Soit maintenant donné P de façon concrète et pratique qu'à chaque $x \in \mathbb{R}$ la probabilité $P(X \leq x)$ soit exactement définie. C'est important de comprendre que $P(X \leq x)$ peut être défini sans problème et raisonnablement pour des variables discrètes et continues tandis que par exemple dans le cas continu $P(X = x)$ ne donne pas de sens pratique.

Définition:

$F(x) = P(X \leq x)$
s'appelle **fonction de répartition** de X .

Remarque:

1. D'après la construction il est $D_F = \mathbb{R}$.
2. A l'aide de la fonction de répartition, nous pouvons définir la "fonction de probabilité" dans le cas discret et dans le cas continu de manière différente.

D'après la définition, la fonction de répartition est une probabilité. C'est pourquoi on peut déduire:

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 Types de répartition

Problème:

On est parti de la question: Comment est-ce que, lors d'expériences de hasard, les probabilités différentes sont distribuées aux événements différents? Partant du résultat, nous connaissons déjà des variables aléatoires discrètes et continues. Mais on peut aussi distinguer les expériences affiliés. Hormis les cas mixtes, nous distinguons deux classes:

Classes d'expériences:

1. **Expériences de compter** \rightsquigarrow répartition discrète
2. **Expériences de mesurer** \rightsquigarrow répartition continue

Exemple 1: Recensement de la population:

Répartition des habitants sur les communes $\rightsquigarrow H(X = a) =$ nombre d'habitants de la commune a .

Ici, la répartition est décrite par une fonction de probabilité $\frac{H(X = a)}{H(S)} = P(X = a)$.

Exemple 2: Fonction de poids:

Répartition du poids sur des poutres à la coupe transversale égale $\rightsquigarrow H(X \leq x) = \int_a^x f(t) dt =$ nombre de poutres de la longueur $x \in [a, b]$.

Ici, la répartition est décrite par une fonction de densité de probabilité $f(x)$:

$\rightsquigarrow f(x) = \frac{d(\int_a^x f(t) dt)}{dx}$. f livre un modèle pour l'ensemble de base, qui souvent n'est pas exactement

connu.

Nous pouvons comparer l'ensemble de base et le modèle comme il suit:

Ensemble de base	Modèle
1. Réalité	1. Théorie
2. Pratique: Répartition de fréquence ou densité	2. Répartition de probabilité
3. Souvent seulement échantillons	3. Problème de qualité du modèle (tests)

4.6.4 Répartition discrète

Fonction aléatoire

Pour les variables discrètes, les domaines de définition ne doivent pas nécessairement être finis. Nous définissons par conséquent:

Définition:

Une **variable aléatoire** X s'appelle **discrète**, s'il vaut:

1. X ne peut atteindre qu'un nombre fini ou infini de façon énumérable de valeurs x_i avec $P(X = x_i) \neq 0$.
2. $\{x_i \mid P(X = x_i) \neq 0\}$ n'a pas de points d'accumulation (limite).

Conséquence:

Dans chaque intervalle fini $\subset \mathbb{R}$ il n'y a qu'un nombre fini de x_i avec $P(X = x_i) \neq 0$.

Notations:

Au lieu d'écrire $P(X = x_i)$ nous notons brièvement $P(x_i)$. Les valeurs x avec $P(x) \neq 0$ soient numérotées: x_1, x_2, x_3, \dots . Pour les probabilités appartenantes à x_1, x_2, x_3, \dots nous notons aussi brièvement: $P(X = x_i) := P(x_i) := p_i$.

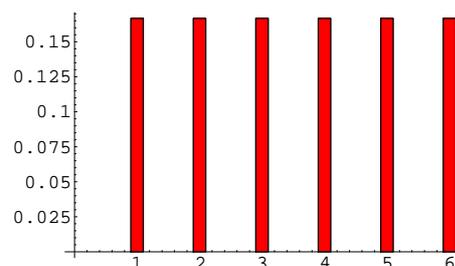
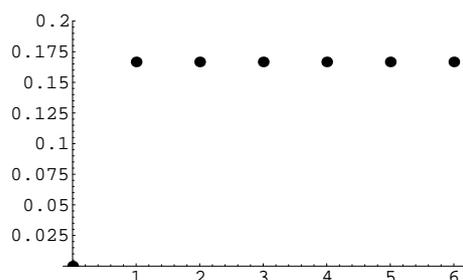
Définition:

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

s'appelle **fonction de probabilité** de la variable aléatoire X

Remarque:

Le graphe d'une telle fonction de probabilité consiste en points isolés. Souvent cette représentation graphique donne une impression peu claire et par conséquent on la remplace par un diagramme de bâton.

Exemple: Dé:

Il est bien clair que la fonction aléatoire détermine complètement la fonction de répartition ou la répartition de la variable aléatoire X .

Théorème:

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

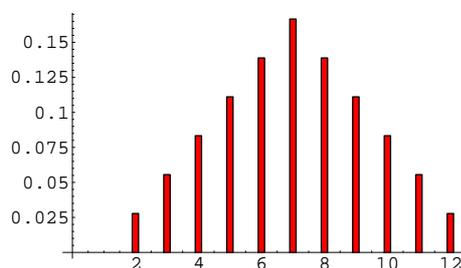
Exemple:

Déterminer la fonction aléatoire (graphe, tableau des valeurs) pour l'expérience de hasard "jouer aux dés avec deux dés".

Tableau des valeurs:

x	$\frac{2}{1}$	$\frac{3}{2}$	$\frac{4}{3}$	$\frac{5}{4}$	$\frac{6}{5}$	$\frac{7}{6}$	$\frac{8}{5}$	$\frac{9}{4}$	$\frac{10}{3}$	$\frac{11}{2}$	$\frac{12}{1}$
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Graphe:

**Fonction de répartition**

Soit X une variable aléatoire d'une expérience de hasard. Soit $P(X \leq x)$ donc défini, $x \in \mathbb{R}$ quelconque. Soit $I = \{X \mid X \leq x\}$. Ainsi nous écrivons brièvement: $P(X \leq x) = P(I)$. De la page ?? on sait qu'ainsi

la fonction aléatoire de la variable aléatoire X est définie par $F(x) := P(X \leq x)$ pour tous les $x \in \mathbb{R}$.

Exemple: Jeter une fois une monnaie. $X = X(\{\text{nombre de faces (têtes) réalisées}\})$.

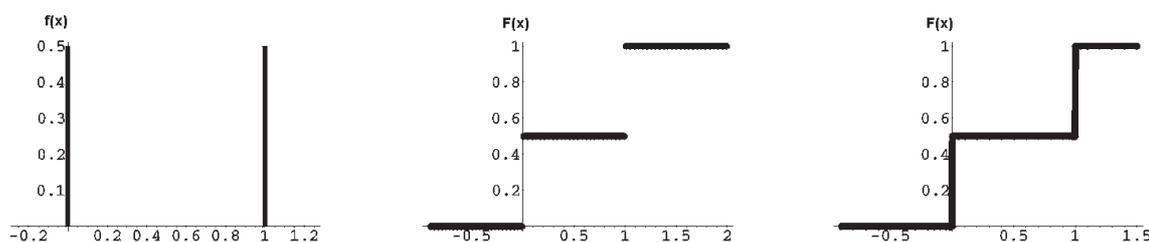
On trouve: $P(X = 0) = P(X = 1) = 0.5$.

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. $x > 1 \rightsquigarrow \dots \Rightarrow P(x) = 1, x > 1$

Résultat:

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

Diagrammes:



Problème: Donné: $X \in (a, b] \rightsquigarrow a < X \leq b$

Question:

Comment est-ce qu'on peut calculer la probabilité $P(a < X \leq b)$ de $F(x)$?

Solution:

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

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

Donc si la répartition est discrète, on peut calculer la fonction aléatoire de la fonction de répartition. Nous connaissons le théorème suivant depuis en haut (page 70):

Théorème:**Hyp.:**

Répartition discrète

Thè.:

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

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

De la page 72 nous savons:

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

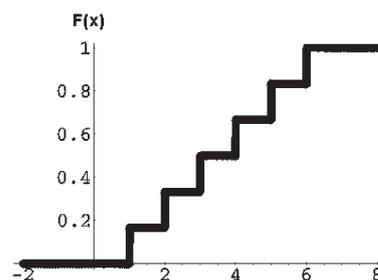
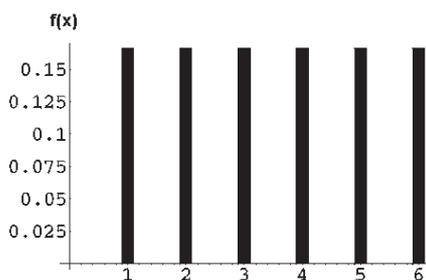
Théorème:

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

Exemple: $X = X(\{\text{Nombre de points en jouant avec un dé.}\})$ $F(x)$ est moins compréhensible que $f(x)$, mais F possède des avantages pour l'application.On trouve: $P(X = 1) = P(X = 2) = \dots = P(X = 6) = \frac{1}{6}$. \rightsquigarrow **Résultat:**

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	$\frac{1}{6}$	$\frac{2}{6} = \frac{1}{3}$	$\frac{3}{6} = \frac{1}{2}$	$\frac{4}{6} = \frac{2}{3}$	$\frac{5}{6}$	$\frac{6}{6} = 1$

Diagrammes:


 \rightsquigarrow P.ex. $P(X = 2 \vee X = 3) = P(1 < X \leq 3) = F(3) - F(1) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$
Remarque:

Une fonction comme $F(x)$ dans l'exemple susdit s'appelle **fonction en escalier (à palier)**. Elle a des sauts à places discrètes (valeurs x_i possibles de la variable aléatoire X), entre ces places elle est constante. Les sauts résultent de la valeur $p(x) \neq 0$. On peut construire la fonction en escalier à l'aide de la "fonction de saut" $h(x)$ (voir fonction δ de Dirac, voir analyse):

Théorème:

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

4.6.5 Répartition continue

Le problème de la fonction aléatoire

Pour nos besoins nous définissons:

Définition:

Une variable aléatoire X s'appelle **continue dans un intervalle** I , si toutes les valeurs de I peuvent être réalisées.

Problème:

Est-ce que pour une variable aléatoire continue $P(X = x_i)$ a un sens?

Le problème est que dans ce cas il y a un nombre infini de valeurs $x_i \in I$ qui peuvent être réalisées. Comme dans le cas discret la somme de la probabilité est 1, logiquement et par conséquent aussi dans ce cas pour la plupart des x_i il vaut $P(X = x_i) = 0$. Ça signifie que ces valeurs réalisées par contre ne sont pas réalisées. Ainsi nous arrivons à une contradiction. Par conséquent, dans ce cadre il n'est pas raisonnable, de demander la probabilité $P(X = x_i)$ ni par conséquent une fonction de probabilité. Nous demandons donc plutôt la probabilité pour les intervalles, par exemple $P(a \leq X < b) = ?$ Ceci n'est pas un problème si une fonction de probabilité est donnée.

Situation au cas discret

Soit donc: $F(x) = P(X \leq x)$

Dans le cas discret il est:

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

Comme les valeurs discrètes x_i sont ordonnées, il existe une liaison bijective entre la valeur x_i et l'index i :

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

Il existe donc aussi $\varphi^{-1} = \psi$:

$$\psi(i) = x_i$$

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

$\tilde{f}(i) \cdot \Delta i$ a la signification d'une aire d'un rectangle. $F(x)$ a donc la signification d'une aire sous une courbe en escalier. Cette signification d'une aire garde sa valabilité aussi pour le cas continu.

Densité de probabilité

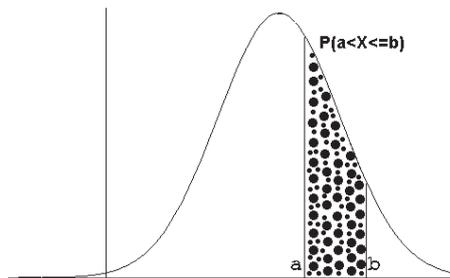
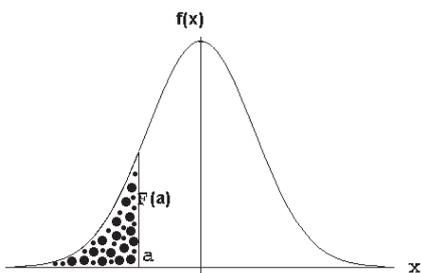
Il est donc bien clair que dans le cas continu, on peut définir $F(x) = P(X \leq x)$ comme mesure de surface sous une courbe.

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

Ici, $f(x) = \frac{dF}{dx}$ ne peut plus avoir la signification d'une probabilité, car toutes les probabilités doivent donner la somme 1. C'est pourquoi on parle ici d'une **densité de probabilité**.

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

Diagrammes:



Exemple: Jouer à la roulette:

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

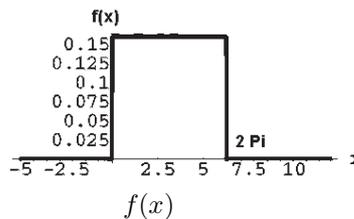
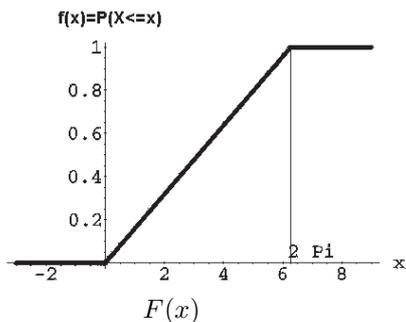
X := angle entre l'aiguille et une direction fixe.

Roulette utilisable $\leadsto P(X \leq x) = \frac{x}{2\pi}$

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

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

Diagrammes:



Remarque: Nous appelons une distribution comme on la voit dans le dernier exemple **distribution rectangulaire** ou **distribution uniforme**.

Indication:

Il faut bien considérer l'analogie entre la distribution de la masse dans la physique et la distribution aléatoire. Les masses ponctuelles correspondent à la probabilité dans le cas discret.

4.7 Mesures de répartition

4.7.1 Considérations générales

Problème: Comment peut-on vite comparer deux distributions?

1. Comparaison qualitative de $f(x)$, $F(x)$ ou des graphiques (quantités infinies!).
2. Comparaison quantitative de mesures caractéristiques: La moyenne μ , variance σ^2 , écart type (quadratique moyen) (déviatoin standard) σ , dissymétrie γ etc..

4.7.2 Valeur moyenne

1. Cas discret:

Comparaison avec la physique: Par le "centre d'une distribution de masses", nous comprenons le centre de gravité. Si l'on considère des distributions de fréquence statistiques à la place de distributions aléatoires, la moyenne arithmétique (ou brièvement la moyenne) prend la place de la distribution de masse. Quant à une distribution aléatoire, c'est donc aussi la valeur moyenne qui remplace le centre de gravité. Il vaut:

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

On peut trouver le cas $n \in \mathbb{N}$ ou $n = \infty$.

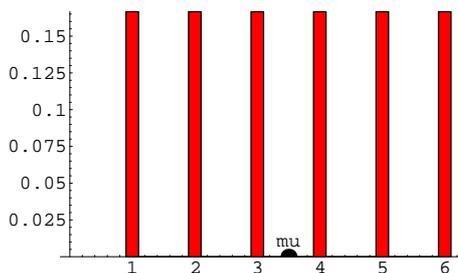
C'est pourquoi nous définissons:

Définition: **Valeur moyenne:** $\mu := \sum_i f(x_i) \cdot x_i \cdot 1$

Conséquence:

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

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



Exemple:

X = nombre de points en jouant avec un dé.

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

2. Transcription de l'idée pour le cas discontinu:

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

C'est pourquoi nous définissons analogiquement au centre de gravité de la physique:

Définition:

Valeur moyenne:

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

4.7.3 Valeur d'espérance

Exemple:

Les deux joueurs A et B jouent comme il suit: Avant le jeu A paie à B le montant = profit moyen attendu au jeu. Alors, A peut jouer aux dés pendant que B paie le gain après le schéma suivant: Au résultat 1 ou 2, A reçoit de B sfr. 10. -, au résultat 3 ou 4, A reçoit de B sfr. 20. -, au résultat 5, A reçoit de B sfr. 40. -, au résultat 6, A reçoit de B sfr. 80. Quelle est le profit moyen?

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

Dans cet exemple, nous constatons consécutivement un classement de la sorte suivante:

$$X \longleftrightarrow g(X) \quad \text{resp. } x_i \longleftrightarrow g(x_i) \quad \begin{array}{c|cccccc} X & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline g(X) & 10 & 10 & 20 & 20 & 40 & 80 \end{array}$$

X est variable aléatoire indépendante, $g(X)$ est variable aléatoire dépendante.

Conséquence:

$g(X)$ est aussi variable aléatoire.

Pour calculer le **gain moyen attendu (l'espérance mathématique)**, nous devons remplacer par conséquent x_i par $g(x_i)$, c.-à.-d. nous remplaçons $x_i \cdot f(x_i)$ par $g(x_i) \cdot f(x_i)$. Là, $f(x_i)$ est la probabilité appartenant à $g(x_i)$.

Nous définissons donc tout à fait analogiquement à l'espérance mathématique moyenne **l'espérance (valeur d'espérance, valeur attendue) $E(g(X))$** pour la **fonction d'espérance $g(X)$** . Analogiquement à la définition de μ nous fixons:

Définition:

1. **Cas discret:** $E(g(X)) := \sum_i g(x_i) \cdot f(x_i) \cdot 1 = \sum_i g(x_i) \cdot f(x_i) \cdot \Delta x$

2. **Cas continu:** $E(g(X)) := \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$

Remarque:

Au lieu de dire **valeur d'espérance**, nous disons souvent aussi **valeur d'espérance mathématique** ou bien brièvement **valeur moyenne**.

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

Spécialement: (Cas discret)

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

A cause de la linéarité de la somme et de l'intégrale il vaut:

Théorème:

Hyp.: $g(X) = a h(X) + b u(X)$

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

4.7.4 Distribution symétrique

Définition:

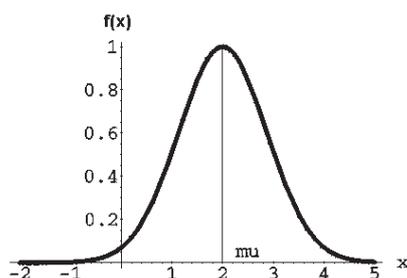
Une distribution s'appelle **symétrique** par rapport à c , s'il est:
 $\forall_{c \pm x \in D_f} : f(c + x) = f(c - x)$

Pour des raisons bien visibles, il vaut pour une distribution symétrique:

Théorème: Hyp.: f symétrique par rapport à c

Thè.: $\mu = c$

Exemple:



4.7.5 Variance, écart-type

Au lieu de s'intéresser à la valeur d'espérance de x resp. de x_i , nous considérons la valeur d'espérance de l'écart carré de μ de x resp. x_i . (Attention: La valeur d'espérance de l'écart linéaire donne toujours 0. Cette quantité par conséquent n'a pas de sens.)

1. Cas discret:

Définition: **Variance (dispersion)**

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

2. Cas continu:

Comme à la valeur d'espérance aussi dans le cas continu la somme devient une intégrale.

Définition: **Variance (dispersion)**

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

3. Définition:

$\sigma = \sqrt{\sigma^2}$ s'appelle **écart type (étalon, quadratique moyen), écart-type** ou **déviati**on normale

Remarque:

σ^2 est, comme mentionné, la moyenne de l'écart carré de μ . Comme la moyenne de carrés, σ^2 appartient par conséquent aussi qualitativement à la classe des carrés. σ est donc par conséquent aussi une mesure pour le montant de l'écart moyen de μ , qui appartient par conséquent qualitativement encore à la classe des valeurs non-carrées et originales. σ^2 ou bien aussi σ sont parfois qualifiés de **dispersion**.

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

Spécialement: (Cas discret)

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

Corollaire:

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

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

Exemple:

$X =$ "Nombre de faces" en jettant une monnaie une fois.

$$X = 0 \rightsquigarrow P(X = 0) = \frac{1}{2} \qquad X = 1 \rightsquigarrow P(X = 1) = \frac{1}{2}$$

$$\Rightarrow \mu = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

$$\Rightarrow \sigma^2 = (0 - \frac{1}{2})^2 \cdot \frac{1}{2} + (1 - \frac{1}{2})^2 \cdot \frac{1}{2} = (\frac{1}{2})^2 \cdot \frac{1}{2} + (\frac{1}{2})^2 \cdot \frac{1}{2} = \frac{1}{4}$$

$$\Rightarrow \sigma = \sqrt{\frac{1}{4}} = \frac{1}{2} \notin \{0, 1\}$$

Conséquence:

$\sigma \in \{\dots, x_i, \dots\}$ n'est pas toujours le cas.

4.7.6 Moments d'une distribution

Les grandeurs comme la moyenne et la variance s'appellent généralement des **moments** de la distribution de la variable aléatoire X avec la fonction aléatoire $f(X)$.

Définition:

La valeur d'espérance suivante s'appelle k -ème **moment** de la distribution:

1. **Cas discret:**

$$m_k = E(X^k) = \sum_i x_i^k \cdot f(x_i)$$

$$m_{k,g} = E(g(X)^k) = \sum_i g(x_i)^k \cdot f(x_i)$$

2. **Cas continu:**

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

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

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

~> Mécanique!

Exemple:

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

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

Théorème:

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

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

Preuve:

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

2. Exercice.

Conséquence:

Il en résulte une façon simple de calculer la variance:

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

Façon d'écrire: Souvent on trouve aussi:

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

↪ **Formule:** $\sigma^2 = \mu_2 - \mu^2$

De la linéarité de la somme et l'intégrale on déduit:

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

A l'aide de cette formule on déduit:

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

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

Le théorème ci-dessus permet parfois une simplification du calcul de la variance.

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

On peut vérifier tout de suite par le calcul:

Théorème: Hyp.:

Pour X on a la moyenne μ et la variance σ^2 .

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

Thè.:

Pour Z on a la moyenne 0 et la variance 1.

Quant à la preuve:

$$\mu(Z) = E(Z) = \frac{1}{\sigma} \cdot (E(X) - E(\mu)) = \frac{1}{\sigma} \cdot (\mu - \mu \cdot E(1)) = \frac{1}{\sigma} \cdot (\mu - \mu \cdot 1) = 0$$

$$\begin{aligned} \text{P. 82: } E(X^2) = \sigma^2 + \mu^2 \Rightarrow \sigma(Z)^2 &= E(Z^2) - \mu(Z)^2 = E(Z^2) = E\left(\frac{1}{\sigma^2} \cdot (X^2 - 2X\mu + \mu^2)\right) \\ &= \frac{1}{\sigma^2} \cdot (E(X^2) - 2E(X)\mu + \mu^2) = \frac{1}{\sigma^2} \cdot (E(X^2) - \mu^2) = \frac{1}{\sigma^2} \cdot (\sigma^2 + \mu^2 - \mu^2) = 1 \end{aligned}$$

Suivant le théorème ci-dessus nous définissons:

Définition:

$$Z = \frac{X - \mu}{\sigma} \text{ s'appelle } \mathbf{forme\ standard} \text{ de } X.$$

4.7.7 Dissymétrie d'une distribution

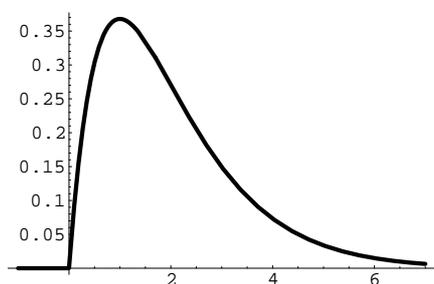
Définition:

$$\text{Soit } \gamma = \frac{1}{\sigma^3} \cdot E((X - \mu)^3)$$

γ s'appelle **dissymétrie** de la distribution.

Remarque:

La dissymétrie est une mesure pour le manque de symétrie de la distribution.



$$f(x) = \begin{cases} 0 & x < 0 \\ x \cdot e^{-x} & x \geq 0 \end{cases}$$

$$\leadsto \int_{-\infty}^{\infty} f(x) dx = 1, \quad \gamma = ?$$

A l'aide de l'intégration partielle nous recevons:

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

4.7.8 Autres caractéristiques

Coefficient de variation

Le **coefficient de variation** ou **coefficient de variabilité** ν d'une variable aléatoire X avec $\mu \neq 0$ est défini comme il suit:

Définition:

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

Médian, mode

Médian et mode sont des mesures pour la position du graphe d'une fonction. Ils concernent en particulier la statistique descriptive. (Voir 2.)

Définition:

Le **médian** d'un ensemble de nombres qui sont ordonnés selon la grandeur, est la valeur du milieu ou la moyenne arithmétique des deux valeurs du milieu de la suite rangée des données.

Si les données ont été groupées en classes, la définition doit être adaptée à cette situation, voir. lit.

Quantiles, fractiles

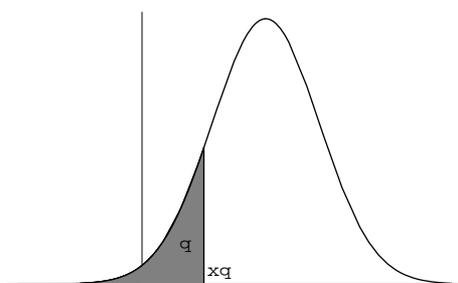
Soit donné une variable aléatoire continue X avec une fonction de densité f et une fonction de distribution F . Soit $0 < q < 1$.

Définition:

$x_q \in \mathbb{R}$ s'appelle **quantile inférieur de l'ordre q (quantile q)**, s'il vaut: $F(x_q) = P(X \leq x_q) = q$
Pareillement pour le **quantile supérieur**.

Conséquence:

Le quantile de l'ordre $q = 0.5$ est exactement le médian.



Voici des quantiles spéciaux: Les **quartiles** ($Q_1 = x_{0.25}$, $Q_3 = x_{0.75}$), les **quintiles**, les **deciles** ou les **centiles** (P_{10} , P_{90}) etc..

Remarque:

La **distance des quartiles** $Q_3 - Q_1 = x_{0.75} - x_{0.25}$ est une mesure de dispersion. De même la distance des centiles $P_{90} - P_{10}$. Souvent on utilise aussi la **demi-distance des quartiles**

$$Q = \frac{Q_3 - Q_1}{2}.$$

Excès, kurtosis, courbure, convexité

P.ex. le **kurtosis** κ est une mesure pour la pente d'une distribution.

Définition:

kurtosis:
 $\kappa = \frac{Q}{P_{90} - P_{10}}$ (Q, P voir en haut)

Coefficient de moments du kurtosis:
 $a_4 = \frac{m_4}{m_2^2}$ ($m_k = E(X^k)$)

Autres moyennes

Pour les autres moyennes comme les moyennes **harmoniques** ou **géométriques** ainsi que relations **empiriques** voir lit..

4.7.9 Fonction caractéristique génératrice de moments

En travaillant avec la définition suivante, on comprend la raison pour laquelle on l'utilise:

Définition:

La valeur espérée $E(e^{tx})$ de e^{tx} s'appelle **fonction caractéristique génératrice de moments** $G(t)$.

Nous calculons $G(t)$ pour des fonctions quelconques $f(x)$:

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

Pour la k -ème dérivée d'après t nous obtenons:

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

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

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

Théorème:

\rightsquigarrow **Truc:**

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

Spécialement:

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

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

Remarque:

La fonction génératrice de moments nous sera utile dans le chapitre prochain au sujet de la distribution de Bernoulli.

4.7.10 Transf. de Laplace et en z

Le lecteur qui connaît les mathématiques d'ingénieur remarque que la fonction génératrice de moments

$G(t)$ évidemment à faire avec les transformations de Laplace resp. les transformations en z . Le rapport est expliqué plus exactement plus tard parce qu'il n'est pas indispensable en ce moment. On acquiert les connaissances de la transformation de Laplace et de la transformation en z dans l'analyse supérieure, voir lit. Suite mathématiques, Wirz, Bibl. A16.

4.8 Distributions discrètes spéciales

4.8.1 Distribution de Bernoulli

Situation et expérience indépendante

Problème:

Nous étudions le nombre de réalisations d'un certain événement A qui soit exécuté répétitivement pendant une expérience aléatoire dont les événements sont indépendants et ont la même probabilité. Soit A p.ex. la réalisation d'un 2 en jouant aux dés avec un dé, le dé sera lancé plusieurs fois. L'événement considéré peut être réalisé ou ne pas être réalisé. La variable aléatoire X nécessaire peut donc atteindre seulement deux valeurs: 0 et 1.

Soit $X = 0$ si A n'est pas réalisé et $X = 1$ si A est réalisé.

Soit $p =$ probabilité que A soit réalisé.

$q = 1 - p =$ probabilité que A ne soit pas réalisé.

$X =$ Variable aléatoire: Nombre de réalisations de A .

Conclusion:

Probabilité de succès à l'expérience sans répétition:

$$\begin{aligned} P(X = 0) &= q, & f(0) &= q \\ P(X = 1) &= p, & f(1) &= p \\ x \notin \{0, 1\} &\Rightarrow & f(x) &= 0 \end{aligned}$$

Ce fait peut être écrit brièvement dans une seule formule:

Conséquence: Lors d'une expérience unique il vaut:

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

Définition:

L'expérience qu'on vient de décrire s'appelle souvent **expérience de Bernoulli**.

Expérience exécutée multiplement

Si l'expérience est répétée n -fois le succès peut arriver 0, 1, 2, ... ou n fois. Succès veut dire que A est réalisé. X peut donc prendre les valeurs 0, 1, 2, ... ou n .

L'événement A soit réalisé x fois à n répétitions de l'expérience. Comme les expériences partielles sont identiques et par conséquent indépendantes, la probabilité de l'événement total est le produit des probabilités des événements partiels:

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

Conséquence:

La formule pour $n = 1$ est donc encore valable.

Au même nombre total, le succès peut être réalisé de manières différentes par les événements partiels. On obtient le nombre de possibilités de la réponse à la question suivante: De combien de manières différentes est-ce qu'on peut choisir x éléments parmi n éléments sans répétition. Ça nous mène par conséquent à une combinaison sans répétition:

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

Exemple:

Lors de 3 expériences un événement doit se réaliser 2 fois. Combien de possibilités est-ce qu'il y a? — Le succès peut se réaliser (a) lors des 1. et 2. expériences ou (b) lors des 1. et 3. expériences ou (c) lors des 2. et 3. expériences. Il y a donc $|\{(a), (b), (c)\}| = 3 = \binom{3}{2}$ possibilités.

Théorème:

Hyp.:

Expérience de Bernoulli avec n répétitions

Thè.:

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

Définition:

La distribution donnée par $f(x) = \binom{n}{x} p^x q^{n-x}$ avec les paramètres $n \in \mathbb{N}$ et $p \in (0, 1)$ s'appelle **distribution de Bernoulli** oder **distribution binomiale** $Bi(n, p)$.

Corollaire:

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

Remarque:

De $p + q = 1$ à n répétitions indépendantes on déduit aussi:

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

4.8.2 Lois pour la distribution de Bernoulli

La fonction de distribution est obtenue de façon simple en prenant la somme:

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

En outre il vaut:
$$\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!}$$

On en déduit:
$$\binom{n}{x+1} = \frac{n-x}{x+1} \cdot \binom{n}{x}$$

Ça nous donne une formule de récurrence:

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

Exemple:

On prépare une livraison avec N vis. M pièces sont inutilisables (rebut). La livraison est examinée par le client. Il en tire n fois une vis et la remet. Quelle est la chance que le client trouve x vis inutilisables?

Il est bien visible qu'il s'agit d'une expérience de Bernoulli. La chance de trouver du rebut en tirant une seule fois est $p = \frac{M}{N}$.

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

Pour calculer la moyenne et la variance nous utilisons la fonction génératrice de moments:

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

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

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

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

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

Théorème: Hyp.:

Distribution de Bernoulli

 Thè.:

$$\mu = n \cdot p$$

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

Exemple:

Jeu de cartes, 32 cartes avec 4 as. L'expérience: Tirer 6 fois une carte et la remettre (expérience de Bernoulli). Événement A : Tirer au moins 3 as.

↪ Expérience indépendante: $p = \frac{4}{32}$

↪ A : 3, 4, 5 ou 6 fois un as.

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

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

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

4.8.3 Distribution de Poisson

Fonction de distribution

Problème:

On cherche une distribution qui d'une part approche relativement bien la distribution de Bernoulli pour les grands n et des petits p et qui d'autre part est commode pour des manipulations mathématiques. Ce qu'on veut est effectué par la **distribution de Poisson**.

Idée: Für die Bernoulliverteilung gilt: Pour la distribution de Bernoulli il vaut: $\mu = n \cdot p$

$$\begin{aligned} \sim p &= \frac{\mu}{n}, \quad p^x = \frac{\mu^x}{n^x}, \quad q^{n-x} = (1-p)^{n-x} = \left(1 - \frac{\mu}{n}\right)^{n-x} = \left(1 - \frac{\mu}{n}\right)^n \cdot \left(1 - \frac{\mu}{n}\right)^{-x} \\ \Rightarrow f(x) &= \binom{n}{x} \cdot p^x \cdot q^{n-x} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{x!} \cdot \frac{\mu^x}{n^x} \cdot \left(1 - \frac{\mu}{n}\right)^n \cdot \left(1 - \frac{\mu}{n}\right)^{-x} = \\ &= \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{n^x} \cdot \left(1 - \frac{\mu}{n}\right)^{-x} \cdot \frac{\mu^x}{x!} \cdot \left(1 - \frac{\mu}{n}\right)^n \end{aligned}$$

Soit maintenant $n \rightarrow \infty$ et $p \rightarrow 0$ de façon qu'il vaut $\mu = n \cdot p = \lim_{n \rightarrow \infty, p \rightarrow 0} n \cdot p$.

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

En outre il vaut: $\left(1 - \frac{\mu}{n}\right)^n = \left(1 + \frac{-\mu}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\mu}$, $\left(1 - \frac{\mu}{n}\right)^{-x} \xrightarrow{n \rightarrow \infty, x, \mu \ll n} 1$

Nous obtenons donc en somme le résultat suivant (**Théorème limite de Poisson**):

Théorème:

Hyp.:

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

Thè.:

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

Définition:

$$f(x) = \frac{\mu^x}{x!} \cdot e^{-\mu} \quad *^1$$

s'appelle **distribution de Poisson** $Po(\mu)$

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

Remarque:

Comme par conséquent de $p \rightarrow 0$ les probabilités des évènements examinés peuvent être très petites, on parle aussi **d'évènements rares**.

Moments

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

Par ce calcul on n'a rien gagné pour la pratique. Pour avancer, nous nous souvenons que la distribution de Poisson est une approximation de la distribution binomiale pour $x \ll n$ et $n \rightarrow \infty$. Par conséquent nous pouvons prendre la moyenne des fréquences relatives et observées dans la pratique pour la moyenne μ si les conditions données sont satisfaites et et s'il s'agit d'une expérience de Bernoulli.

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

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

Analogiquement on calcule: $\gamma = \frac{1}{\sigma^3} \cdot E((X - \mu)^3) = \frac{1}{\sqrt{\mu}}$

Théorème:

Hyp.:

Distribution de Poisson

Thè.:

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

Exemple:

Quelle est la probabilité qu'à une école avec $n = 1000$ étudiants il y a au moins $k = 1$ étudiants qui ont leur anniversaire le 1er mai?

Expérience de Bernoulli:

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

Approximation de Poisson:

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

4.8.4 Distribution de Pascal

L'expérience consiste de nouveau en expériences partielles et indépendantes. Soit p par contre à l'expérience de Bernoulli la probabilité donnée d'un **échec** dans l'expérience de base. $1 - p$ est donc par

conséquent la probabilité pour le succès. k soit le paramètre des succès jusqu'au r -ème **échec (ne pas réussir)**. X varie sur $\{k_i\}$. $(1-p)^k$ prend maintenant la place de p^x dans la distribution de Bernoulli, p^r prend la place de $(1-p)^{n-x}$ ($r = n - x = n - k$). (On choisit k possibilités dans un ensemble de $r + k - 1$ possibilités, parce que lors de l'expérience on a r échecs sans le dernier qui vient toujours à la dernière place et termine donc l'expérience, ce qui n'augmente pas les possibilités du choix. k est le nombre de succès.) La solution du problème de choix mène donc à la formule suivante:

Formule: $X =$ nombre d'expérience indépendantes jusqu' à ne pas réussir pour le r -ème fois, $q =$ probabilité de succès (p : échec), $n := k + r$.

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

Comme ici l'échec détermine p , la définition suivante devient compréhensible:

Définition:

$r > 0$, $0 < p < 1 \rightsquigarrow$ ${}_n B_k$ s'appelle **distribution binomiale négative**.

Définition:

Pour $r \in \mathbb{N}$ la distribution ${}_n B_k$ s'appelle aussi **distribution de Pascal** $Pa(r, p)$.

On trouve des applications dans la théorie du dégât.

Calculer directement μ et σ^2 à la main est un peu fatigant. Par la méthode appliqué lors de la distribution binomiale et p.ex. Mathematica on trouve vite le résultat désiré. On trouve:

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

Remarque:

$\mu < \sigma^2 \rightsquigarrow$ distribution binomiale négative
 $\mu > \sigma^2 \rightsquigarrow$ distribution binomiale
 $\mu = \sigma^2 \rightsquigarrow$ distribution de Poisson
 Pour $r = 1$ on obtient la distribution géométrique.

Conséquence:

L'estimation de μ et σ^2 d'un échantillon, on obtient une indication pour déterminer le type de la distribution.

Pour calculer les probabilités, on peut tout de suite déduire une formule de récurrence:

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

Exemple:

Donnée: Une machine qui est utilisée en moyenne tous les 4 semaines (20 journées de travail). Après une utilisation, on fait une révision du moteur qui dure 3 jours. Probabilité qu'elle sera utilisée exactement le 1. jour ($r = 1$, $k = 0$) ou exactement au 2. jour ($r = 1$, $k = 1$) ou exactement au 3. jour ($r = 1$, $k = 2$) pendant la révision?

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

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

4.8.5 Distribution géométrique

Définition:

La **distribution géométrique** est le cas spécial de la distribution de Pascal pour $r = 1$.

Il vaut donc:

Théorème:

Hyp.:

Distribution géométrique

Thè.:

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

$p = 1 - q$ est la probabilité d'un échec, $q = 1 - p$ la probabilité d'un succès et X donne le nombre aléatoire de succès jusqu'au premier échec. Les expériences partielles sont indépendantes. k soit le paramètre des succès jusqu'au 1-er **échec** ($k + 1$).

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

Théorème:

$$P(X \leq k) = 1 - q^{k+1} \leadsto \text{Suite géométrique!}$$

On obtient les moments à l'aide de la distribution de Pascal pour $r = 1$:

Théorème:

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

Remarque:

La distribution géométrique (à cause de la suite géométrique) prend la place de la distribution exponentielle aux distributions discrètes.

Exemple:

Donnée: Une machine qui est utilisée en moyenne tous les 4 semaines (20 journées de travail). Maintenant

on vient de l'utiliser. Probabilité qu'elle sera utilisée dans les prochains 10 jours?

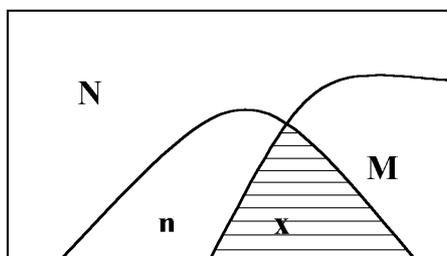
$$p = \frac{1}{20} = 0.05, \quad \mu = 20 - 1 = \frac{q}{1 - q} \Rightarrow q = \frac{19}{20} = 0.95. \quad (p = 1 - q = \frac{1}{20} = 0.05).$$

$$P(X \leq 9) = 1 - P(X \geq 10) = 1 - \left(\frac{1}{20}\right)^{9+1} = 1 - 0.95^{10} \approx 0.401263.$$

4.8.6 Distribution hypergéométrique

Par contre à la distribution de Bernoulli, à la distribution hypergéométrique les expériences ne sont pas indépendantes. Dans le modèle de la distribution de Bernoulli, nous tirons un élément qui tout de suite après est remis à sa place originale. Dans le modèle de la distribution hypergéométrique par contre nous tirons **sans remettre l'élément**. Par conséquent p change ici d'une expérience à l'autre.

Expérience modèle:



Modèle:

Donné: Récipient qui contient N éléments.

M éléments sont défectueux.

On tire n éléments sans les remettre.

x parmi les n éléments sont défectueux.

Problème:

Probabilité d'obtenir x éléments défectueux si on tire n éléments?

$$g = \text{cas favorables: Choisir } x \text{ parmi } M \text{ et au même temps } n - x \text{ parmi } N - M. \rightsquigarrow g = \binom{M}{x} \cdot \binom{N - M}{n - x}$$

$$m = \text{cas possibles: Choisir } n \text{ parmi } N. \rightsquigarrow m = \binom{N}{n}$$

$$\Rightarrow \frac{g}{m} = \frac{\binom{M}{x} \cdot \binom{N - M}{n - x}}{\binom{N}{n}}$$

Définition:

A l'expérience modèle décrit, la **distribution hypergéométrique** est donnée par:

$$f(x) = \frac{\binom{M}{x} \cdot \binom{N - M}{n - x}}{\binom{N}{n}}$$

Remarque:

La notion „distribution hypergéométrique“ vient de la „fonction hypergéométrique“.

Comme on l'a déjà exécuté dans d'autres cas, on peut utiliser aussi dans ce cas $G(t)$ pour calculer μ et σ^2 . Le calcul livre:

Théorème:

Hyp.:

Distribution hypergéométrique

Thè.:

$$\mu = n \cdot \frac{M}{N}$$

$$\sigma^2 = \frac{n \cdot M \cdot (N - M) \cdot (N - n)}{N^2 \cdot (N - 1)}$$

Remarque:

1. Si $N, M, N - M$ sont grands et n est petit, la distribution hypergéométrique est une approximation de la distribution binomiale. $\leadsto p = \frac{M}{N}$
2. Si p est petit, n grand et N grand par rapport à n , la distribution hypergéométrique est une approximation de la distribution de Poisson. $\leadsto \mu = n \cdot p$

Exemple:

Donné: Une caisse qui contient 500 roulements à billes. L'examen selon le contrat: Retirer 50 (30) roulements, examiner. Si tous sont en ordre, la livraison est acceptée. Sinon, justement on ne les accepte pas. D'après le contract de livraison: Probablement maximal 2% $\hat{=} 10$ sont de mauvaise qualité. Calculer: $P(k = 0)$.

$$N = 500, \quad k = 0, \quad p = \frac{2}{100} = 0.02 = \frac{M}{N} = \frac{M}{500} \Rightarrow M = 500 \cdot 0.02 = 10.$$

$$n = 50 \Rightarrow P(X = k = 0) = \frac{\binom{M}{k} \cdot \binom{N - M}{n - k}}{\binom{N}{n}} = \frac{\binom{10}{0} \cdot \binom{500 - 10}{50 - 0}}{\binom{500}{50}} = 0.345162.$$

$$n = 30 \Rightarrow P(X = k = 0) = \frac{\binom{M}{k} \cdot \binom{N - M}{n - k}}{\binom{N}{n}} = \frac{\binom{10}{0} \cdot \binom{500 - 10}{30 - 0}}{\binom{500}{30}} = 0.535489.$$

4.8.7 Exemple

Exemple:

Expérience: Fusil abloqué, 10 coups sur une cible, probabilité par coup de toucher le but = 0.1. Quelle

est la probabilité de toucher au moins une fois?

Truc: A : toucher au moins une fois $\Rightarrow \bar{A}$: ne pas toucher.

Soit A_0 : Toucher avec un seul coup.

$$\leadsto P(A_0) = p = \frac{1}{10}, \quad P(A) = 1 - P(\bar{A}), \quad P(\bar{A}) = P(\bar{A}_0)^{10} \Rightarrow P(A) = 1 - P\left(\frac{9}{10}\right)^{10} \approx 0.651$$

\leadsto Distribution?

4.9 Distributions continues spéciales

4.9.1 Généralités

$f(x)$ soit une densité de probabilité pièce par pièce incessante avec $D_f = \mathbb{R}$ et $F(x)$ la fonction de distribution affiliée. Nous allons maintenant discuter des variantes spéciales de $f(x)$.

4.9.2 Distribution rectangulaire

Soit $f(x)$ la fonction de densité d'une variable aléatoire X .

La distribution donnée par X s'appelle **distribution rectangulaire**, s'il vaut:

Définition:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \vee x > b \end{cases}$$

On calcule:

$$1. \mu = m_1 = E(X^1) = \int_{-\infty}^{\infty} x^1 \cdot f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{a+b}{2}$$

$$2. \sigma^2 = m_2 - m_1^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - \mu^2 = \int_a^b x^2 \cdot \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}$$

Théorème:

Hyp.:

Distribution rectangulaire

Thè.:

$$\mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}$$

Exemple: Voir page 76.

4.9.3 Distribution normale ou de Gauss

Remarques préliminaires

Historiquement la distribution normale a déjà été utilisée par Gauss dans le contexte de la théorie des erreurs de mesure aléatoires, non-systématiques ou non-méthodiques (courbe des erreurs). Les raisons suivantes montrent son importance:

1. Dans la pratique, beaucoup de variables aléatoires sont à peu près distribuées de façon normale. Reste le problème que, dans la pratique, les grandeurs infinies mesurées n'existent pas pour des raisons physiques.
2. Beaucoup de grandeurs mesurées peuvent être transformées de façon simple en une distribution normale ou elles peuvent être approximées par une telle distribution.
3. La distribution normale joue aussi un rôle aux méthodes de contrôle statistiques.
4. La distribution totale de la superposition d'erreurs aléatoires et indépendantes a comme distribution limite la distribution normale **théorème limite central**.

Fonction de distribution

De l'analyse on sait: $\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$

Indication quant à la preuve:

$$\text{Soit } I(R) = \int_{-\infty}^R e^{-t^2} dt \Rightarrow (I(R))^2 = \int_{-\infty}^R e^{-x^2} dx \cdot \int_{-\infty}^R e^{-y^2} dy = \int_{-\infty}^R \left(\int_{-\infty}^R e^{-(x^2+y^2)} dx \right) dy$$

Au lieu d'intégrer sur le carré $[0, R] \times [0, R]$, on peut aussi intégrer en coordonnées polaires sur les cercles intérieurs et extérieurs qui incluent naturellement le carré. Pour $R \rightarrow \infty$, on obtient donc la valeur limite $\sqrt{\pi}$.

Nous pouvons en déduire:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi} &\Rightarrow \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi} \Rightarrow \int_{-\infty}^{\infty} e^{-\frac{t^2}{2 \cdot s^2}} dt = s\sqrt{2\pi} \Rightarrow \int_{-\infty}^{\infty} e^{-\frac{(t-m)^2}{2 \cdot s^2}} dt = s\sqrt{2\pi} \\ &\Rightarrow \frac{1}{s\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{(t-m)^2}{2 \cdot s^2}} dt = 1, \quad D_f = \mathbb{R} \end{aligned}$$

Conclusion:

$$\frac{1}{s\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{(t-m)^2}{2 \cdot s^2}} dt = 1 \Rightarrow f(x) := \frac{1}{s\sqrt{2\pi}} \cdot e^{-\frac{(t-m)^2}{2 \cdot s^2}} \text{ est}$$

admissible comme fonction de distribution continue.

Définition:

$$\text{Soit } f(x) := \varphi(x; m, s^2) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2 \cdot s^2}}$$

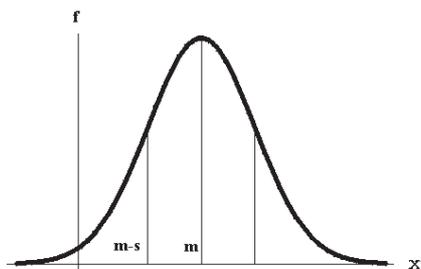
$$\text{Soit } \varphi(x) := \varphi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

La fonction de densité $f(x)$ définit une **distribution normale** ou **distribution de Gauss** avec les paramètres m et s pour la variable aléatoire X .

Fonction de distribution:

$$F(x) := \Phi(x; m, s^2) = \frac{1}{s\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{(t-m)^2}{2 \cdot s^2}} dt$$

Symbole: $X \in N(m; s^2)$ ($\rightsquigarrow Z = \frac{-(X-m)^2}{2 \cdot s^2} \in N(0; 1)$)



On appelle le graphe aussi **cloche de Gauss**.
Il est symétrique à $x = m$

Pour la valeur moyenne il vaut:

$$\mu = E(X) := \int_{-\infty}^{\infty} f(x) \cdot x dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2 \cdot s^2}} dx = m$$

Calculer la variance (p.ex. employer Mathematica):

$$\sigma^2 = E(X^2) - \mu^2 = G_t''(0) - \mu^2 = \frac{d^2}{dt^2} \left(\int_{-\infty}^{\infty} e^{xt} f(x) dx \right) \Big|_{t=0} - \mu^2 = s^2$$

Théorème:

Hyp.:

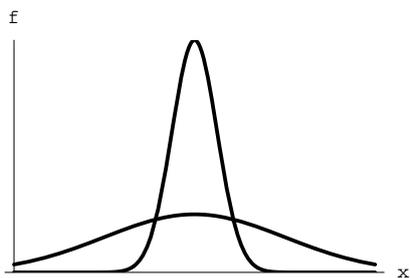
$$X \in N(m; s^2)$$

Thè.:

1. $\mu = m$
2. $\sigma^2 = s^2$

Définition:

S' $\mu = 0$ et $\sigma^2 = 1$, on a la **distribution normale standar-disée**.



Courbe plate: $\sigma = 1$, courbe raide: $\sigma = 25$.

Remarque:

$$F(x) := \Phi(x; m, s^2) = \frac{1}{s\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{(t-m)^2}{2 \cdot s^2}} dt$$

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

$$\Phi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

$$F(x) = \Phi(x; \mu, \sigma^2) = \Phi\left(\frac{x - \mu}{\sigma}; 0, 1\right) = \Phi(z; 0, 1)$$

$$\leadsto X \in N(\mu; \sigma^2) \Leftrightarrow Z = \frac{X - \mu}{\sigma} \in N(0; 1)$$

On ne peut pas calculer la fonction de répartition $F(x)$ de façon élémentaire. Pour arriver quand-même à un résultat, on utilise des calculateurs (error function, *Erf*) ou des tableaux. Souvent on trouve $\Phi(z; 0, 1)$ dans des tableaux. A l'aide de ces tableaux on peut trouver $\Phi(x; \mu, \sigma^2)$.

Entre $\mu - \sigma$ et $\mu + \sigma$ on trouve environ 68% de la surface sous la courbe de $f(x)$, entre $\mu - 2\sigma$ et $\mu + 2\sigma$ on trouve environ 95.5% de la surface et entre $\mu - 3\sigma$ et $\mu + 3\sigma$ on trouve environ 99.7% de la surface.

D'autres informations voir:

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

4.9.4 Théorèmes limites de Moivre Laplace

Ces théorèmes traitent l'approximation de la répartition de Bernoulli par une distribution normale.

4.9.5 Théorème limite locale

Donné:

Répartition de Bernoulli

$$f_n(x) = \binom{n}{x} p^x q^{n-x} \leadsto \mu = n \cdot p, \quad \sigma^2 = n \cdot p \cdot q, \quad q = 1 - p, \quad 0 < p < 1$$

Distribution normale

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2}}, \quad \mu = n \cdot p, \quad \sigma^2 = n \cdot p \cdot q, \quad q = 1 - p, \quad z = \frac{x - \mu}{\sigma} \in \mathbb{R}$$

Nous voulons prouver, que pour des x "petits" $f_n(x)$ converge de façon uniforme vers $f(x)$ pour tous les x dans un intervalle quelconque et fini.

Théorème:**Hyp.:**

Comme décrit

Thè.:

$$f_n(x) \sim f(x)$$

Symbole: " \sim ": lire égale de façon asymptotique**Preuve:**

Nous utilisons la formule de Stirling (page 32):

$$\rightsquigarrow k! \approx \sqrt{2\pi k} \left(\frac{k}{e}\right)^k, \quad k! = \sqrt{2\pi k} \cdot \left(\frac{k}{e}\right)^k \cdot e^{\frac{\Phi}{12k}}, \quad 0 < \Phi < 1$$

La preuve de cette formule est une affaire étendue, voir p.ex. Amann u. Escher, Analysis II (Bibl.: A2) ou van der Waerden Bibl. A13.

$$\begin{aligned} \rightsquigarrow f_n(x) &= \binom{n}{x} p^n q^{n-x} = \frac{n!}{x! \cdot (n-x)!} p^x q^{n-x} = \\ &= \frac{\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{\Phi_n}{12n}} \cdot \frac{1}{\sqrt{2\pi x}} \cdot \left(\frac{x}{e}\right)^{-x} \cdot e^{-\frac{\Phi_x}{12x}} \cdot \frac{1}{\sqrt{2\pi(n-x)}} \cdot \left(\frac{n-x}{e}\right)^{-(n-x)} \cdot e^{-\frac{\Phi_{n-x}}{12(n-x)}} \cdot p^x \cdot q^{n-x}}{\frac{\sqrt{2\pi n}}{\sqrt{2\pi(n-x)} \cdot \sqrt{2\pi x}} \cdot \left(\frac{n}{e}\right)^n \cdot \left(\frac{x}{e}\right)^x \cdot \left(\frac{n-x}{e}\right)^{(n-x)} \cdot e^{\frac{\Phi_n}{12n}} \cdot e^{-\frac{\Phi_x}{12x}} \cdot e^{-\frac{\Phi_{n-x}}{12(n-x)}} \cdot p^x \cdot q^{n-x}} = \\ &= \frac{1}{\sqrt{2\pi}} \cdot n^{n+\frac{1}{2}} \cdot p^x \cdot \frac{1}{x^{x+\frac{1}{2}}} \cdot \frac{1}{(n-x)^{(n-x)+\frac{1}{2}}} \cdot q^{n-x} \cdot e^{\frac{\Phi_n}{12n} - \frac{\Phi_x}{12x} - \frac{\Phi_{n-x}}{12(n-x)}} = \\ &= \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} \cdot \left(\frac{n \cdot p}{x}\right)^{(x+\frac{1}{2})} \cdot \left(\frac{n \cdot q}{n-x}\right)^{(n-x+\frac{1}{2})} \cdot e^\tau = \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} \cdot e^{-h} \cdot e^\tau, \\ \tau &= \frac{1}{12} \cdot \left(\frac{\Phi_n}{n} - \frac{\Phi_x}{x} - \frac{\Phi_{n-x}}{n-x}\right), \quad \Phi_i \in (0, 1), \quad h = \left(x + \frac{1}{2}\right) \cdot \ln\left(\frac{n \cdot p}{x}\right) + \left(n-x + \frac{1}{2}\right) \cdot \ln\left(\frac{n \cdot q}{n-x}\right) \end{aligned}$$

$$\text{Soit } z = \frac{x-\mu}{\sigma} = \frac{x-n \cdot p}{\sqrt{n \cdot p \cdot q}}, \quad |z| < K = \text{const.}, \quad p+q=1$$

$$\rightsquigarrow h(x) := w(z) = \left(np + \frac{1}{2} + z\sqrt{npq}\right) \cdot \ln\left(1 + z\sqrt{\frac{q}{np}}\right) + \left(nq + \frac{1}{2} - z\sqrt{npq}\right) \cdot \ln\left(1 - z\sqrt{\frac{p}{nq}}\right)$$

$$n \text{ assez grand } \rightsquigarrow u_1 = |z\sqrt{\frac{q}{np}}| < 1, \quad u_2 = |z\sqrt{\frac{p}{nq}}| < 1$$

 \rightsquigarrow On peut développer $\ln(1+u_1)$, $\ln(1-u_2)$ dans des séries de puissances autour du centre 1:

$$\ln(u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \frac{u^4}{4} + \dots$$

$$\rightsquigarrow \ln\left(1 + z\sqrt{\frac{q}{np}}\right) = z\sqrt{\frac{q}{np}} - \frac{z^2}{2} \frac{q}{np} + \dots, \quad \ln\left(1 - z\sqrt{\frac{p}{nq}}\right) = -z\sqrt{\frac{p}{nq}} - \frac{z^2}{2} \frac{p}{nq} + \dots$$

$$\rightsquigarrow w(z) = \left(np + \frac{1}{2} + z\sqrt{npq}\right) \cdot \left(z\sqrt{\frac{q}{np}} - \frac{z^2}{2} \frac{q}{np} + \dots\right) + \left(nq + \frac{1}{2} - z\sqrt{npq}\right) \cdot \left(-z\sqrt{\frac{p}{nq}} - \frac{z^2}{2} \frac{p}{nq} + \dots\right)$$

Ici, nous multiplions les termes, après nous les ordonnons d'après les puissances de z et nous simplifions ($q = 1 - p$, par exemple appliquer Mathematica). Ainsi nous obtenons:

$$w(z) = \left(\frac{q-p}{2\sqrt{npq}} \cdot z + z^2 \cdot \left(\frac{1}{2} - \frac{1}{2n(1-p)} - \frac{1}{4np(1-p)} - \frac{p}{2n(1-p)}\right)\right) + \frac{1}{\sqrt{n}} \cdot R(n, p) \cdot z^3, \\ |R(n, p)| < \text{const.} (z \in (0, 1))$$

$$\rightsquigarrow \lim_{n \rightarrow \infty} w(z) = \frac{1}{2} \cdot z^2, \quad z = \frac{x-\mu}{\sigma} = \frac{x-n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \in [\alpha, \beta] \subset \mathbb{R}$$

\leadsto convergence uniforme

$\leadsto e^{-h(x)} = e^{-w(z)} \rightarrow e^{z^2/2}$ pour $n \rightarrow \infty$

$$\alpha \leq z = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \leq \beta \Rightarrow x \geq np + \alpha \sqrt{npq} = np(1 + \alpha \sqrt{\frac{q}{np}}),$$

$$n - x \geq nq - \beta \sqrt{npq} = nq(1 - \beta \sqrt{\frac{p}{nq}})$$

$$\Rightarrow |\tau| = \left| \frac{1}{12} \cdot \left(\frac{\Phi_n}{n} - \frac{\Phi_x}{x} - \frac{\Phi(n-x)}{(n-x)} \right) \right| < \frac{1}{12} \cdot \left(\frac{1}{n} + \frac{1}{x} + \frac{1}{(n-x)} \right)$$

$$< \frac{1}{12} \cdot \left(\frac{1}{n} + \frac{1}{np(1 + \alpha \sqrt{\frac{q}{np}})} + \frac{1}{nq(1 - \beta \sqrt{\frac{p}{nq}})} \right) \rightarrow 0 \quad \text{pour } n \rightarrow \infty$$

$\leadsto e^\tau \rightarrow 1$ pour $n \rightarrow \infty$

$$\leadsto f_n(x) = \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} \cdot e^{-h} \cdot e^\tau \rightarrow \underbrace{\frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}}}_0 \cdot \underbrace{e^0}_1 = 0, \quad f_n(x) \sim \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} \cdot e^{z^2/2}$$

Théorème:

Théorème limite locale

Hyp.:

$$\alpha \leq z = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \leq \beta, \quad (z \in [\alpha, \beta]),$$

$$n \rightarrow \infty$$

Thè.:

$$f_n(x) \sim \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} \cdot e^{z^2/2}$$

4.9.6 Théorème limite de De Moivre/ Laplace

Le théorème de limite de De Moivre/ Laplace est un corollaire du théorème de limite locale. C'est un cas spécial du théorème limite centrale qui n'est pas traité ici (voir lit.).

Soit [...] la fonction de parenthèses de Gauss (Floor, Int).

Théorème:

Théorème de limite de De Moivre/ Laplace

Hyp.:

$$n \rightarrow \infty,$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du, \quad a < b, \quad \alpha = \left[\frac{a - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right], \quad \beta = \left[\frac{b - n \cdot p + 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right],$$

$$\alpha \leq z = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \leq \beta$$

Thè.:

$$P(a \leq X \leq b) = \sum_{x=a}^b \binom{n}{x} p^x q^{n-x} \sim \Phi(\beta) - \Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2} du$$

4.9.7 La loi des grands nombres de Bernoulli

La **loi des grands nombres** fournit une explication propositionnelle pour la supposition que la fréquence suit la probabilité pour les grand n .

Hyp.:

Soit A : événement avec la probabilité $p \in (0, 1)$ à une expérience de hasard.

Soit X =: nombre de réalisations de A si on exécute l'expérience n fois.

Soit $\varepsilon \in \mathbb{R}^+$.

Théorème:

Bernoulli

Thè.:

$$P\left(\left|\frac{x}{n} - p\right| < \varepsilon\right) \xrightarrow{n \rightarrow \infty} 1$$

Remarque:

Le théorème affirme donc: La probabilité ou la chance que l'écart de la fréquence relative de la probabilité $p = P(X)$ soit plus petit que n'importe quel nombre ε positif quelconque si n est suffisamment grand, converge vers 1. Attention: Seulement la probabilité converge, mais ne pas la différence $\left|\frac{x}{n} - p\right|$. Une grande probabilité non signifie pas sécurité!

Nous pouvons déduire l'assertion du théorème de De Moivre/ Laplace:

Preuve:

$$\begin{aligned} \left|\frac{x}{n} - p\right| < \varepsilon &\Leftrightarrow -\varepsilon < \frac{x}{n} - p < \varepsilon \Leftrightarrow p - \varepsilon < \frac{x}{n} < p + \varepsilon \Leftrightarrow a := (p - \varepsilon) \cdot n < x < (p + \varepsilon) \cdot n := b \\ \Rightarrow P\left(\left|\frac{x}{n} - p\right| < \varepsilon\right) &= P((p - \varepsilon) \cdot n < x < (p + \varepsilon) \cdot n) = P(a < x < b) = P(a \leq x \leq b), \end{aligned}$$

$$\text{De Moivre/ Laplace} \rightsquigarrow P(a \leq x \leq b) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2} du \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du$$

Car il vaut:

$$\begin{aligned} \alpha &= \left[\frac{a - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{(p - \varepsilon) \cdot n - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{-\varepsilon \cdot n - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{-\varepsilon \cdot \sqrt{n}}{\sqrt{p \cdot q}} + \frac{0.5}{\sqrt{p \cdot q}} \right] \rightarrow -\infty, \\ \text{beta} &= \left[\frac{b - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{(p + \varepsilon) \cdot n - n \cdot p - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{+\varepsilon \cdot n - 0.5}{\sqrt{n \cdot p \cdot (1-p)}} \right] = \left[\frac{+\varepsilon \cdot \sqrt{n}}{\sqrt{p \cdot q}} + \frac{0.5}{\sqrt{p \cdot q}} \right] \rightarrow \infty \end{aligned}$$

4.9.8 Remarques quant au hasard

Comme nous avons vu au début, au hasard absolu une raison est exclue, par contre au hasard relatif, une raison n'est cependant pas perceptible, mais elle existe. Or, cela n'a aucun sens de se poser des questions au sujet de l'existence d'une chose qui n'est pas perceptible. On peut croire, ne pas croire ou laisser la question ouverte.

Par conséquent nous pouvons commencer avec la compréhension de l'idée suivante: Pour l'être humain qui observe, c'est un hasard, si la réalisation d'un événement est imprévue, pas voulue, sans raison reconnaissable ni régularité. C'est un hasard si l'événement ne résulte pas nécessairement d'un ensemble d'événement donnés et si tout aurait pu se dérouler différemment. Seulement l'effet est visible, mais non pas une cause .

Pourtant, aux ensembles d'événements plus grands, des lois deviennent maintenant visibles. Les fréquences relatives d'expériences s'approchent aux probabilités déduisibles de façon théorique de modèles (loi des grands nombres). Les observations empiriques et accidentelles ont ici des effets qui suivent des lois. Par conséquent on peut poser quand même la question raisonnable si les résultats ne résultent pas suivant des lois dans un contexte d'effets complexe. Par conséquent la prédiction du résultat pourrait être pratiquement trop compliquée, tandis qu'elle est théoriquement possible. (P.ex. expérience avec la planche de Galton). Le hasard serait donc une suite de la complexité. Mais la chose peut être aussi inversée: Des causes connues relativement exactement ont parfois un effet bien accidentel malgré la loi de la nature connue. P.ex. à la prévision météorologique ou au calcul de la position que la terre avait dans le système solaire il y a des millénaires. Les erreurs de mesure inévitables peuvent tellement s'amplifier durant les calculs que l'erreur finale surmonte le domaine de valeur du résultat. Par conséquent les changements les plus petits des causes (constellations originales) peuvent avoir des écarts énormes dans l'effet (effet de papillon).

Qu'on pense dans ce contexte aux débauches du déterminisme (démon de Laplace, négation de la volonté libre) ou au rapport du flou (description déterminée de l'inconnu).

4.9.9 Inéquation de Tschebyscheff

Soit Y une variable aléatoire quelconque, qui ne doit pas être distribuée de façon normale. Soit $E(Y)$ la valeur d'espérance, $Var(Y)$ soit la variance et ε soit un nombre positif et quelconque. Alors il vaut (sans preuve, voir annexe):

Formule: **Tschebyscheff**

$$P(|Y - E(Y)| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}$$

4.9.10 Distribution normale logarithmique

Soit $f(t) := \varphi(t; \mu_L, \sigma_L^2) = \frac{1}{\sigma_L \sqrt{2\pi}} e^{-\frac{(t-\mu_L)^2}{2\sigma_L^2}}$

= fonction de densité $f(y)$ de la distribution normale, paramètres μ_L et σ_L , variable aléatoire Y .

~> Fonction de distribution: $F(y) := \Phi(y; \mu_L, \sigma_L^2) = \frac{1}{\sigma_L \sqrt{2\pi}} \cdot \int_{-\infty}^y e^{-\frac{(t-\mu_L)^2}{2\sigma_L^2}} dt$

~> Y est distribué de façon normale.

Soit $X = \log(Y)$, $x = \log(y)$ (Substitution)

Si on remplace $y = \log_a(x)$, $a^y = x$ dans $f(y)$, ça ne correspond pas à la règle de la substitution. (On n'a pas tenu compte de la dérivée intérieure). À cause de la normalisation il doit valoir:

$$1 = \int_{-\infty}^{\infty} f(y) dy = \int_{x=a^{-\infty}}^{x=a^{\infty}} f(\log_a(x)) \frac{dy}{dx} dx = \int_0^{\infty} f(\log_a(x)) \frac{\log_a(e)}{x} dx$$

Définition:

Soit donnée la fonction de densité:

$$h(x) = \begin{cases} 0 & x \leq 0 \\ \frac{\log_a(e)}{x} \varphi(\log_a(x); \mu_L, \sigma_L^2) & x > 0 \end{cases}$$

La distribution donnée par $h(x)$ s'appelle **distribution normale logarithmique**.

Nous obtenons la fonction de distribution par la substitution de $F(y) := \Phi(y; \mu_L, \sigma_L^2)$:

Formule:
$$H(x) = \frac{1}{\sigma_L \sqrt{2\pi}} \cdot \int_0^x e^{-\frac{(\log_a(z) - \mu_L)^2}{2 \cdot \sigma_L^2}} \frac{\log_a(e)}{t} dt$$

Pour $a = e$ on obtient par calcul de la valeur d'espérance, de la variance et du médian pour la distribution normale logarithmique:

Formule:
$$\begin{aligned} E(X) &= e^{\mu_L + \sigma_L^2/2} \\ Var(X) &= e^{2\mu_L + \sigma_L^2} (e^{\sigma_L^2} - 1) \\ x_{0,5} &= e^{\mu_L} \end{aligned}$$

Application: études des temps, analyse de longévité

Distribution normale: Superpositions additives \rightsquigarrow distribution normale logarithmique: superpositions multiplicatives.

4.9.11 Distribution exponentielle

Il vaut:

$$\int_0^\infty \alpha e^{-\alpha x} dy = -\frac{\alpha}{\alpha} e^{-\alpha x} \Big|_0^{x=\infty} = -(e^{-\infty} - e^{-\alpha \cdot 0}) = -(0 - 1) = 1$$

Définition:

La variable aléatoire X satisfait une **distribution exponentielle** avec le paramètre α , si elle a la fonction de densité suivante:

$$f(x) = \begin{cases} 0 & x < 0 \\ \alpha e^{-\alpha x} & x \geq 0 \end{cases}$$

\rightsquigarrow Fonction de distribution:

$$\int_{-\infty}^x f(t) dt = \int_0^x \alpha e^{-\alpha t} dt = 1 - e^{-\alpha x}$$

Formule:

$$F(x) = \int_0^x f(t) dt = \begin{cases} 0 & x < 0 \\ 1 - e^{-\alpha x} & x \geq 0 \end{cases}$$

Valeur d'espérance, variance et médian:

$$\mu = E(X) = \int_0^{\infty} x \cdot \alpha e^{-\alpha x} dx = x \cdot \frac{\alpha}{-\alpha} e^{-\alpha x} \Big|_0^{\infty} - \int_0^{\infty} (-1) e^{-\alpha x} dx = 0 - 0 + \int_0^{\infty} e^{-\alpha x} dx = \frac{1}{-\alpha} e^{-\alpha x} \Big|_0^{\infty} = \frac{1}{\alpha}$$

$$\sigma = \text{Var}(X) = \int_0^{\infty} x^2 \cdot \alpha e^{-\alpha x} dx - \mu^2 = \dots = \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}$$

$$F(x_{0.5}) = 1 - e^{-\alpha x_{0.5}} = \frac{1}{2} \Rightarrow \dots \Rightarrow x_{0.5} = \frac{1}{\alpha} \ln(2)$$

Application: mesure du temps (désintégration de l'atome, heures de travail, ...), durée de vie.

4.9.12 Distribution Weibull

Définition:

Une variable aléatoire continue X a une **distribution de Weibull** aux trois paramètres $a > 0$, $b > 0$ et c , si la fonction de densité est donnée comme il suit:

$$f(x) = \begin{cases} 0 & x \leq c \\ \frac{b}{a} \left(\frac{x-c}{a}\right)^{b-1} e^{-\left(\frac{x-c}{a}\right)^b} & x > c \end{cases}$$

 \rightsquigarrow Fonction de distribution:**Formule:**

$$F(x) = \begin{cases} 0 & x \leq c \\ 1 - e^{-\left(\frac{x-c}{a}\right)^b} & x > c \end{cases}$$

 $c = 0 \rightsquigarrow$ Distribution à deux paramètres $a = 1 \rightsquigarrow$ Distribution réduite

Connue de l'analyse:

Symbole: Soit $\Gamma(x)$ la **fonction Gamma**.**Définition:** $\Gamma(x) := \int_0^{\infty} e^{-t} t^{x-1} dt, \quad x > 0$

Formule: $\Gamma(1) = 1, \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \Gamma(x) = (x-1)\Gamma(x-1)$

On obtient en calculant:

Formule:

$$\mu = E(X) = c + a \cdot \Gamma\left(1 + \frac{1}{b}\right)$$

$$\sigma^2 = Var(X) = a^2 \left(\Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right) \right)$$

$$x_{0.5} = c + a (\ln(2))^{1/b}$$

Application: analyses de durée de vie et de fiabilité.

4.9.13 Distribution gamma

Définition:

Une variable aléatoire continue X a une **distribution gamma** aux paramètres $a > 0$, $b > 0$ et c , si la fonction de densité est donnée comme il suit:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} & x > 0 \end{cases}$$

Remarque:

On obtient la distribution exponentielle comme cas spécial de la distribution gamma pour $b = \alpha$ et $p = 1$. Pour $p \in \mathbb{N}$ la distribution s'appelle aussi **distribution de Erlang**.

Formule: $\mu = E(X) = \frac{p}{b}, \sigma^2 = Var(X) = \frac{p}{b^2}$

Application: analyses de fiabilité (distribution de durée de vie) distribution des temps de service.

4.9.14 Autres distributions

Pour pouvoir traiter d'autres distributions importantes, nous devons acquérir des connaissances sur les variables aléatoires aux dimensions supérieures (vecteurs aléatoires). Les lois usuelles pour les tests sont importantes quant aux tests statistiques.

4.10 Vecteurs aléatoires et leurs distributions

↪ Probabilité multidimensionnelle

4.10.1 Question, notions

Vecteur aléatoire, fonction de répartition

Jusqu'à maintenant nous avons étudié les **variables aléatoires unidimensionnelles** X , qui atteignent des valeurs $\in \mathbb{R}$. La fonction de répartition est donc une fonction d'une variable indépendante.

Par contre, dans une expérience aléatoire des événements différents peuvent être réalisés. Ces événements peuvent être indépendants ou dépendants. Également dans une expérience aléatoire plusieurs variables aléatoires peuvent être nécessaires à la description. Ces variables peuvent être indépendantes ou dépendantes. Ici nous allons étudier brièvement le cas modèle de deux variables aléatoires. ($N = 2$.)

Exemple:

Au contrôle de qualité d'un lot d'arbres de roues dentées on contrôle l'épaisseur X et simultanément à la même pièce chaque fois aussi la longueur Y . L'épaisseur et la longueur sont fabriquées chaque fois pendant une phase de travail séparée. Par conséquent on peut les considérer comme indépendantes. Comme les mesurages cependant sont faits en paires, Y et X ont quand-même un rapport. Ça mène à des questions spéciales de la statistique mathématique que nous ne pouvons pas traiter ici.

Soyent X et Y deux variables aléatoires discrètes ou continues, qui apparaissent à une expérience de hasard. \rightsquigarrow X peut prendre des valeurs dans $\{x_1, x_2 \dots x_k \dots\}$ ou dans \mathbb{R} (généralement $x \in \mathbb{R}$) et Y peut prendre des valeurs dans $\{y_1, y_2 \dots y_k \dots\}$ ou dans \mathbb{R} (généralement $y \in \mathbb{R}$).

Définition:

Le vecteur $\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ s'appelle **vecteur aléatoire bidimensionnel**. Correspondamment, $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ s'appelle **vecteur aléatoire de dimension n** .

Soit $A = \{\omega\}$ un événement, qui soit réalisé lors d'une expérience de hasard avec deux variables aléatoires. X y prend la valeur $X(A) = x$ et en même temps c'est Y qui prend la valeur $Y(A) = y$. Ça veut dire que $\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ prend la valeur $\begin{pmatrix} y \\ x \end{pmatrix}$ (ou brièvement la paire (x, y)). $\begin{pmatrix} y \\ x \end{pmatrix}$ est donc une **réalisation** de \vec{X} .

Remarque:

Au lieu de considérer une paire (X, Y) de variables aléatoires, on peut parler aussi d'un vecteur aléatoire $\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$. La loi de probabilité bi-dimensionnelle du couple (X, Y) (fonction aléatoire) est donc donnée par la double suite $\langle\langle x_i, y_k, p_{i,k} \rangle\rangle$

(Pour le développement de la théorie nous nous limitons ici à $N = 2$. Le transfert à $N > 2$ est sans problèmes.)

Pour les probabilités affiliés nous définissons:

Définition: $F(x, y) = P(X \leq x \wedge Y \leq y)$, $x \in \mathbb{R}$, $y \in \mathbb{R}$

Dans le cas discret il est spécialement:
 $p_{ik} := P(X = x_i \wedge Y = y_k)$

$\leadsto F(x, y)$ est une fonction $\mathbb{R}^2 \xrightarrow{F} [0, 1]$.

Qualités:

1. F est croissant de façon monotone dans X et dans Y et en plus continue depuis la droite.
2. $F(-\infty, y) = 0$, $F(x, -\infty) = 0$, $F(\infty, \infty) = 1$

Cas discret: $\sum_{i,k} p_{ik} = 1$

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. $P(X \leq x) = P(X \leq x, Y \leq \infty) = F(x, \infty)$
 $P(Y \leq y) = P(X \leq \infty, Y \leq y) = F(\infty, y)$

Les deux premiers propriétés sont évidentes. Quant à la quatrième propriété il s'agit d'une **spécialisation** (projetion ou restriction de $D_F = \mathbb{R}^N$ à \mathbb{R}^{N-1}). La troisième propriété se voit à cause de:

$$\begin{aligned}
 & F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \\
 &= (P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1)) - (P(X \leq x_1, Y \leq y_2) + P(X \leq x_1, Y \leq y_1)) \\
 &= \underbrace{(P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1))}_{=P(X \leq x_2, y_1 < Y \leq y_2)} - \underbrace{(P(X \leq x_1, Y \leq y_2) - P(X \leq x_1, Y \leq y_1))}_{=P(X \leq x_1, y_1 < Y \leq y_2)} \\
 &= P(X \leq x_2, y_1 < Y \leq y_2) - P(X \leq x_1, y_1 < Y \leq y_2) \\
 &= P(x_1 < X \leq x_2, y_1 < Y \leq y_2) \geq 0 \quad (P(\dots, \dots) \in [0, 1]!) \quad \text{☺}
 \end{aligned}$$

Répartitions marginales, variables aléatoires indépendantes

Quant aux restrictions nous définissons:

Définition:

distributions (répartitions) marginales de la variable aléatoire bidimensionnelle \vec{X} :

$$F_X(x) := F(x, \infty) = P(X \leq x) = P(X \leq x, Y \leq \infty)$$

$$F_Y(y) := F(\infty, y) = P(Y \leq y) = P(X \leq \infty, Y \leq y)$$

Remarque:

$F_X(x)$ et $F_Y(y)$ sont les fonctions des composantes X et Y .

Soit $A = \{X \leq x, Y \leq \infty\}$, $B = \{X \leq \infty, Y \leq y\} \Rightarrow A \cap B = \{X \leq x, Y \leq y\}$

Pour des événements indépendants il vaut:

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

$$\begin{aligned} \rightsquigarrow F(x, y) &= P(X \leq x, Y \leq y) = P(A \cap B) = P(A) \cdot P(B) \\ &= P(X \leq x, Y \leq \infty) \cdot P(X \leq \infty, Y \leq y) = F_X(x) \cdot F_Y(y) \end{aligned}$$

Par conséquent nous pouvons transférer la notion d'indépendance des événements aux variables:

Définition:

X et Y s'appellent **indépendantes**, s'il vaut:

$$\forall_{(x,y)} F(x, y) = F_X(x) \cdot F_Y(y)$$

Variables réparties de façon identique

Soit $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$, $\vec{X}^T = (X_1, \dots, X_n) \rightsquigarrow$

Définition:

Si toutes les composantes X_i , ($i = 1, \dots, n$) ont la même fonction de répartition $F_Z(Z)$, ($Z = X_1, \dots, X_n$), les X_i s'appellent **répartis de façon identique**.

Définition:

Les variables X_i , ($i = 1, \dots, n$) s'appellent **complètement indépendantes**, si chacune des variables est indépendante de toutes les autres.

A l'aide de la définition pour $n = 2$ on peut déduire:

Théorème:

Hyp.:

X_i , ($i = 1, \dots, n$) compl. indépendants et réparties de façon identique

Thè.:

$$\forall_{(x_1, x_2, \dots, x_n)} F_{\vec{X}}(x_1, x_2, \dots, x_n) = F(x_1) \cdot F(x_2) \cdot \dots \cdot F(x_n)$$

Remarque:

4.10.2 Le cas discret

Soit $p_{ik} := P(X = x_i, Y = y_k) := f(x_i, y_k)$, $in \in \mathbb{N}$

Généralement nous disons:

Définition:

$\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ s'appelle **discret**, si $\{x_i, y_k\}$ resp. $\{x_i, y_k, p_{ik}\}$ est dénombrable.

Formule:

$$F(x, y) = \sum_{x_i \leq x, y_k \leq y} p_{ik}$$

Exemple:

Au contrôle de qualité d'un lot d'arbres de roues dentées on contrôle l'épaisseur d et simultanément à la même pièce chaque fois aussi la longueur l . (L'épaisseur et la longueur sont fabriquées chaque fois pendant une phase de travail séparée. Par conséquent on les peut considérer comme indépendents.) Nous pouvons définir librement les valeurs de X et de Y comme il suit:

Critère:	X	Y
d dans la tolérance	$x_1 = 0$	—
d hors de la tolérance	$x_2 = 1$	—
l dans la tolérance	—	$y_1 = 0$
l hors de la tolérance	—	$y_2 = 1$

Nous savons par l'expérience:

Critère:	Montant:
Rebut	5%
d fautif	1%
l fautif	3%
d et l fautifs	1%

Conséquence:

X	0	1	
Y			
0	$p_{11} = 0.95$	$p_{21} = 0.01$	$p_{.1} = 0.96$
1	$p_{12} = 0.03$	$p_{22} = 0.01$	$p_{.2} = 0.04$
	$p_{1.} = 0.98$	$p_{2.} = 0.02$	$p_{tot} = 1$

(P.ex. on a: $p_{11} + p_{21} = p_{.1}$, $p_{.1} = P(Y = 0) \dots$)

Définition:

Probabilités comme $p_{.1}$ s'appellent **sommes marginales**.

Vue d'ensemble des sommes marginales:

$$\begin{aligned} p_{.1} &= P(Y = 0) = p_{11} + p_{21} = 0.96 \\ p_{.2} &= P(Y = 1) = p_{12} + p_{22} = 0.04 \\ p_{1.} &= P(X = 0) = p_{11} + p_{12} = 0.98 \\ p_{2.} &= P(X = 1) = p_{21} + p_{22} = 0.02 \end{aligned}$$

Il vaut:

Théorème:

$$f_X(X = x_i) = p_{i\cdot} = P(X = x_i) = \sum_k p_{i\ k}$$

$$f_Y(Y = y_k) = p_{\cdot k} = P(Y = y_k) = \sum_i p_{i\ k}$$

$$\sum_k p_{\cdot k} = p_{\cdot i} p_{i\cdot} = 1$$

De la définition générale de l'indépendance de variables il suit:

Corollaire:

Variables discrètes X et Y indépendantes

$$\forall_{(i,k)} p_{i\ k} = p_{i\cdot} \cdot p_{\cdot k}$$

Exemple: $p_{1\cdot} \cdot p_{\cdot 1} = 0.98 \cdot 0.96 \approx 0.94 \neq p_{1\ 1} = 0.95$

↪ ↪

4.10.3 Le cas continu

Nous considérons de nouveau le cas modèle bidimensionnel.

Définition:

On déduit des définitions générales:

Conclusion:

1. $f(x, y) \geq 0$
2. $F(\infty, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) du dv = 1$

Pour la densité des **répartitions marginales** il vaut:

Conclusion:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

De la définition de l'indépendance par la fonction de répartition ($F(x, y) = F_X(x) \cdot F_Y(y)$) on peut conclure à l'aide du théorème de la moyenne du calcul intégral:

Théorème:**Hyp.:** X, Y indépendantes**Thè.:**

$$f(x, y) = f(x) \cdot f(y)$$

4.11 Espérance multidimensionnelle

4.11.1 Espérance, moyenne

Nous considérons de nouveau le cas modèle $n = 2$. Soit $f(x, y)$ la fonction aléatoire resp. la densité. Analogiquement au cas unidimensionnel on peut maintenant aussi définir des grandeurs caractéristiques pour X ou Y :

Définition:**Valeur d'espérance** $E(g(X, Y))$ d'une fonction donnée $g(X, Y)$:

$$E(g(X, Y)) : = \begin{cases} \sum_i \sum_k g(x_i, y_k) \cdot f(i, k) & \text{dis.} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f(x, y) dx dy & \text{cont.} \end{cases}$$

Prémisse ou hypothèse:

$$\sum_i \sum_k |g| \cdot f \quad \text{oder} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g| \cdot f dx dy$$

exist. resp. conv.

Analogiquement au cas unidimensionnel on déduit (calcul):

Théorème:

$$E(a g(X, Y) + b h(X, Y)) = a E(g(X, Y)) + b E(h(X, Y))$$

Spécialement: $a = b = 1$, $g(X, Y) = X$, $h(X, Y) = Y \rightsquigarrow$ **Corollaire:**

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

Généralisation:

Théorème:**Théorème d'addition pour les moyennes**

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

Généralement il est $E(X^2) \neq E(X)^2$.

Exemple: Jouer au dés $\rightsquigarrow E(X) = \frac{7}{2}$

$$E(X^2) = \sum_{k=1}^6 k^2 \cdot \frac{1}{6} = \frac{91}{6} \neq \left(\frac{7}{2}\right)^2 = E(X)^2$$

Pour des variables indépendantes il vaut, à cause de la formule $f(x, y) = f(x) \cdot f(y)$ resp. à cause de la formule $F(x, y) = F(x) \cdot F(y)$, p.ex. dans le cas discret:

$$E(XY) = \sum_i \sum_k x_i \cdot y_k \cdot f(x_i, y_k) = \sum_i x_i \cdot f_X(x_i) \cdot \sum_k y_k \cdot f_Y(y_k) = E(X) \cdot E(Y)$$

Correspondamment au cas continu. \rightsquigarrow

Conséquence:

$$X, Y \text{ indépendantes} \Rightarrow E(XY) = E(X) \cdot E(Y)$$

Généralisation:

Théorème:

Théorème de multiplications pour les moyennes

Hyp.: $i \neq k \Rightarrow X_i, X_k \text{ indépendantes}$

Thè.:

$$E(X_1 \cdot X_2 \cdot \dots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \dots \cdot E(X_n)$$

4.11.2 Variance, covariance, corrélation

Connu du cas avec une variable:

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

Nous définissons:

Définition:

Variances:

$$\sigma_X^2 = \text{Var}(X) = E((X - \mu_X)^2)$$

$$\sigma_Y^2 = \text{Var}(Y) = E((Y - \mu_Y)^2)$$

Soit $Z = X + Y$

$$\begin{aligned} \rightsquigarrow \sigma_{X+Y}^2 &= \sigma_Z^2 = E(Z^2) - (\mu_Z)^2 = E(Z^2) - (E(Z))^2 = E(X^2 + 2XY + Y^2) - (E(X + Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - (E(X) + E(Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\ &= (E(X^2) - E(X)^2) + 2(E(XY) - E(X)E(Y)) + (E(Y^2) - E(Y)^2) \\ &= \sigma_X^2 + 2(E(XY) - E(X)E(Y)) + \sigma_Y^2 \end{aligned}$$

Définition:

$$\sigma_{XY} := \text{cov}(XY) := (E(XY) - E(X)E(Y))$$

s'appelle **covariance** de X et Y

Il vaut:

$$\begin{aligned} E((X - \mu_X)(Y - \mu_Y)) &= E(X \cdot Y - X \cdot \mu_Y - \mu_X \cdot Y + \mu_X \cdot \mu_Y) \\ &= E(X \cdot Y) - \mu_Y \cdot E(X) - \mu_X \cdot E(Y) + \mu_X \cdot \mu_Y \cdot E(1) = E(X \cdot Y) - \mu_Y \cdot \mu_X - \mu_X \cdot \mu_Y + \mu_X \cdot \mu_Y \\ &= E(X \cdot Y) - \mu_X \cdot \mu_Y = E(X \cdot Y) - E(X) E(Y) \Rightarrow E((X - \mu_X)(Y - \mu_Y)) = \sigma_{XY} \end{aligned}$$

Définition:
$$\rho_{XY} := \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{cov}(XY)}{\sigma_X \sigma_Y}$$

s'appelle **coefficient de corrélation** de X et Y
($\sigma_X, \sigma_Y \neq 0$)

Il est évident:

Théorème:

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. $|\rho_{XY}| \leq 1$

Quant au dernier théorème (les autres ont été vérifiés) nous considérons comme exemple le cas continu:

$$\begin{aligned} \rho_{XY} &:= \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{(E((X - \mu_X)^2) E((Y - \mu_Y)^2))^{\frac{1}{2}}} \\ &\Rightarrow \rho_{XY} \cdot (E((X - \mu_X)^2) E((Y - \mu_Y)^2))^{\frac{1}{2}} = E((X - \mu_X)(Y - \mu_Y)) \\ &\Rightarrow \rho_{XY} \cdot \underbrace{\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x, y) dx dy \right)^{\frac{1}{2}}}_{I_1} \cdot \underbrace{\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_Y)^2 \cdot f(x, y) dx dy \right)^{\frac{1}{2}}}_{I_2} = \rho_{XY} \cdot (I_1 \cdot I_2)^{\frac{1}{2}} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy \Rightarrow \rho_{XY} \cdot (I_1 \cdot I_2)^{\frac{1}{2}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy \end{aligned}$$

Ici nous appliquons l'inégalité de Cauchy-Schwarz à $I_1 \cdot I_2$.

$$\text{(Cauchy-Schwarz: } \int_G |fg|^2 \leq \left(\int_G f^2 \right) \cdot \left(\int_G g^2 \right), \quad \int_G fg \leq \int_G |fg| \Rightarrow \left(\int_G fg \right)^2 \leq \left(\int_G f^2 \right) \cdot \left(\int_G g^2 \right))$$

(L'inégalité de Cauchy-Schwarz est connue comme "inégalité du produit scalaire". Elle est correspondamment aussi valable dans le cas discret.)

On en déduit:

En haut on a déduit:

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = \sigma_X^2 + \sigma_Y^2 + 2\text{cov}(XY) + \sigma_X^2 + \sigma_Y^2 + 0 \Rightarrow \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

Théorème:

Hyp.:

X_1, \dots, X_n non corrélées

Thè.:

1. X_1, \dots, X_n indépendantes
2. $\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$

4.11.3 Le cas discret

Conclusion:

1. $E(g(X, Y)) = E(g(X))$
 $\Rightarrow E(g(X)) = \sum_i g(x_i) \sum_k f(x_i, y_k) = \sum_i g(x_i) \cdot f_X(x_k)$
 \leadsto
 Formule pour une seule variable, o.k.!

2. **Valeurs d'espérance.**

$$\mu_X := E(X) := \sum_i x_i \cdot f_X(x_i) = \sum_i \sum_k x_i \cdot p_{ik}$$

$$\mu_Y := E(Y) := \sum_k y_k \cdot f_X(y_k) = \sum_k \sum_i y_k \cdot p_{ik}$$

4.11.4 Le cas continu

Analogiquement au cas discret on peut maintenant définir des grandeurs caractéristiques pour X ou Y :

Définition:

1. $E(g(X, Y)) = E(g(X))$
 $\Rightarrow E(g(X)) = \int_{-\infty}^{\infty} g(x) \int_{-\infty}^{\infty} f(x, y) dy dx = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx$
 \leadsto
 Formule pour une seule variable, o.k.!

2. **Valeurs d'espérance.**

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

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

4.12 Répartitions multidimensionnelles

4.12.1 Distribution normale bidimensionnelle

Dans le cas continu cette distribution joue un rôle important.

Définition:

Soient donnés les paramètres $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho_{XY}$. $\begin{pmatrix} X \\ Y \end{pmatrix}$ possède une répartition normale bidimensionnelle si la densité est donnée par la fonction suivante:

$$f(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1 - \rho_{XY}^2}} \cdot e^{-\frac{h(x, y)}{2}}$$

$$h(x, y) = \frac{1}{(1 - \rho_{XY}^2)} \left(\frac{(x - \mu_X)^2}{\sigma_X^2} - 2 \frac{\rho_{XY} (x - \mu_X)(y - \mu_Y)}{\sigma_X \sigma_Y} + \frac{(y - \mu_Y)^2}{\sigma_Y^2} \right)$$

$x, y, \in \mathbb{R}$

Le théorème suivant sans preuve:

Théorème:

Hyp.:

1. \vec{X} a une distribution normale bidimensionnelle
2. $\text{cov}(X Y) = 0$

Thè.:

1. X, Y indépendantes
2. $f(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y} \cdot e^{-\frac{1}{2} \left(\frac{(x - \mu_X)^2}{\sigma_X^2} + \frac{(y - \mu_Y)^2}{\sigma_Y^2} \right)}$
3. $\mu_X = \mu_Y = 0, \sigma_X^2 = \sigma_Y^2 = 1 \Rightarrow f(x, y) = \frac{1}{2\pi} \cdot e^{-\frac{1}{2}(x^2 + y^2)}$

4.12.2 Fonctions d'échantillonnage, distributions de test

Remarque:

On peut diviser les distributions importants des statistiques en deux classes: Les répartitions qui apparaissent en rapport avec les **modèles mathématiques** d'expériences aléatoires et répartitions, qui sont des **distributions de test** ou des **lois usuelles pour les tests**, ce qui forme la base de tests statistiques. Les distributions de test sont ici des distributions de fonctions d'échantillonnage.

Hyp.:

Dans ce qui suit toutes les composantes X_i , $i = 1, \dots, n$ de \vec{X} soient indépendamment et identiquement distribuées de façon normale avec la valeur d'espérance μ et la variance σ^2 . (\leadsto Echantillon de la taille n avec $N(\mu, \sigma^2)$. L'ensemble de base soit distribué de façon normale.)

Distribution de la moyenne

Cette répartition est une distribution de la fonction d'échantillonnage \bar{X} . Elle est importante lors du traitement des **intervalles de confiance** pour la moyenne d'une distribution normale à la variance connue. Il vaut:

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

Théorème:

\bar{X} est distribué de façon normale avec la valeur d'espérance μ et la variance $\frac{\sigma^2}{n}$ $\leadsto N(\mu, \frac{\sigma^2}{n})$

Remarque:

Il est important de remarquer que le théorème vaut encore de façon approximative si la base des données est distribuée de façon quelconque avec les mêmes paramètres.

Soit $\bar{X} \xrightarrow{h} Z = h(\bar{X}) = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$

La transformation de coordonnées h cause que ce théorème est un corollaire du Lemme suivant:

Lemme:

Hyp.:

X_i , $i = 1, \dots, n$ de \vec{X}

X_i indépendantes, distribuées de façon normale avec $N(\mu, \sigma^2)$

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

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

Thè.:

Z satisfait la distribution normale standardisée d'après $N(0, 1)$

Ce lemme est de nouveau un corollaire d'un théorème plus général comme on voit tout de suite:

Théorème:

Somme de variables indépendantes et distribuées de façon normale

Hyp.:

Soient X_1, X_2, \dots, X_n indépendantes, distribuées de façon normale avec
Moyennes $\mu_1, \mu_2, \dots, \mu_n$
Variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$

Thè.:

$\sum_{i=1}^n X_i$ distribuées de façon normale

Moyenne $\mu = \sum_{i=1}^n \mu_i$

Variance $\sigma^2 = \sum_{i=1}^n \sigma_i^2$

Pour la preuve nous avons besoin de la formule suivante:

Lemme:

Hyp.:

$Z = g(X, Y) = X + Y \quad X, Y \text{ indep.}$

Densités: $X \rightsquigarrow f_1(x), Y \rightsquigarrow f_2(Y)$

Thè.:

$$F(Z) = \int_{-\infty}^{\infty} f_2(y) \left(\int_{-\infty}^{z-y} f_1(x) dx \right) dy$$

$$f_1 \in \text{cont} \Rightarrow f(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-y) dx = \int_{-\infty}^{\infty} f_1(z-x) f_2(y) dy$$

Preuve Lemme:

De la page 112: $f(x, y) = f_1(x) \cdot f_2(y)$

$$F(z) = \int_{x+y \leq z} \int f(x, y) dx dy = \int_{x+y \leq z} \int f_1(x) \cdot f_2(y) dx dy = \int_{-\infty}^{\infty} (f_2(y) \int_{-\infty}^{x=z-y} f_1(x) \cdot dx) dy \quad \text{☺}$$

$$z = x + y, \quad \frac{dz}{dx} = 1 \Rightarrow F(z) = \int_{-\infty}^{\infty} (f_2(y) \int_{-\infty}^z f_1(z-y) \cdot dx) dz \Rightarrow f(z) := \frac{d}{dz} F(z) =$$

$$\frac{d}{dz} \int_{-\infty}^{\infty} (f_2(y) \int_{-\infty}^z f_1(z-y) dz) dy = \int_{-\infty}^{\infty} (f_2(y) \frac{d}{dz} \int_{-\infty}^z f_1(z-y) dz) dy = \int_{-\infty}^{\infty} (f_2(y) f_1(z-y)) dy$$

Si en calculant F on met f_2 au lieu de f_1 en devant, on obtient: $f(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-y) dy \quad \text{☺}$

Preuve Théorème:

1.

Induction, "ancrer": $n = 1$: $X = X_1$ distribuée de façon normale.

2. Induktion, transmission de $n = 1$ à $n = 2$.

Soit d'abord: $X = X_1 + X_2$

$$\text{D'après hyp.: } f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}, \quad f_2(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}$$

$$\begin{aligned} \Rightarrow f(x) &= \int_{-\infty}^{\infty} f_1(x-y) f_2(y) dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \cdot e^{-\frac{1}{2}\left(\frac{x-y-\mu_1}{\sigma_1}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_2} \cdot e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-y-\mu_1}{\sigma_1} + \frac{y-\mu_2}{\sigma_2}\right)^2} dy \end{aligned}$$

$$\text{Soit } \mu := \mu_1 + \mu_2, \quad \sigma^2 := \sigma_1^2 + \sigma_2^2, \quad V := -\frac{1}{2} \left(\frac{x-y-\mu_1}{\sigma_1} + \frac{y-\mu_2}{\sigma_2} \right)$$

$$\text{Soit } V_1 := \frac{\sigma}{\sigma_1\sigma_2} \left(y - \frac{\sigma_1^2\mu_2 + \sigma_2^2(x-\mu_1)}{\sigma^2} \right), \quad V_2 := \frac{x-\mu}{\sigma}$$

Par multiplication des termes, mettre au même dénominateur et remodeler on démontre que l'identité suivante est valable (un tas de termes \rightsquigarrow p.ex. avec Mathematica!):

$$V \equiv V_1^2 + V_2^2$$

V_2 est indépendante de y , $V_1 := \tau$ est utilisée comme nouvelle variable d'intégration:

$$\begin{aligned} \rightsquigarrow f(x) &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}(V_1^2+V_2^2)} dy = \frac{1}{2\pi\sigma_1\sigma_2} \cdot e^{-\frac{1}{2}V_2^2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}V_1^2} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot e^{-\frac{1}{2}V_2^2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\tau^2} d\tau \end{aligned}$$

$$\text{Il vaut: } \frac{dV_1}{dy} = \frac{d\tau}{dy} = \frac{\sigma}{\sigma_1\sigma_2} \Rightarrow dy = \frac{\sigma_1\sigma_2}{\sigma} d\tau, \quad \int_{-\infty}^{\infty} e^{-\frac{1}{2}\tau^2} d\tau = \sqrt{2\pi} \quad (\text{p. 97})$$

$$\Rightarrow f(x) = \frac{1}{2\pi\sigma_1\sigma_2} \cdot e^{-\frac{1}{2}V_2^2} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\tau^2} \frac{\sigma_1\sigma_2}{\sigma} d\tau = \frac{1}{2\pi\sigma} \cdot e^{-\frac{1}{2}V_2^2} \cdot \sqrt{2\pi} = \frac{\sqrt{2\pi}}{\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

\rightsquigarrow Pour $n = 2$, X est donc distribuée de façon normale.

3. Conclusion de l'induction de $n = m$ à $n = m + 1$:

Hyp.: $Y_1 = X_{m+1}$, $Y_2 = \sum_{i=1}^m X_i$ distribuées de façon normale.

Loi: Y_1, Y_2 distribuées de façon normale

\rightsquigarrow Comme à X_1, X_2 : la somme est aussi distribuée de cette façon.

\rightsquigarrow Thèse.: $Y_1 + Y_2 = \sum_{i=1}^m X_1 + X_{m+1} = \sum_{i=1}^{m+1} X_i$ distribuées de façon normale. 

Par substitution sous l'intégrale on trouve le théorème suivant:

Théorème:**Hyp.:**

X distribuée de façon normale avec $N(\mu, \sigma^2)$
 $\tilde{X} = c_1 \cdot X + c_2, \quad c_1, c_2 = \text{const.}$ (Transf. lin.)

Thè.:

\tilde{X} distribuée de façon normale
 $\tilde{\mu} = c_1 \cdot \mu + c_2, \quad \tilde{\sigma}^2 = c_1^2 \cdot \sigma^2$

Les formules pour $\tilde{\mu}$ et $\tilde{\sigma}^2$ ont été déduites à la page 83 en forme plus générale.

4.12.3 Distribution du Khi-deux

Définition, fonction de répartition

Soit $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ (variance empirique)

Se rappeler: $X_i \rightsquigarrow N(\mu, \sigma^2)$

A partir de ceci, on peut former la fonction d'échantillon suivante:

Définition:
$$\chi^2 = \frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\chi^2 = \sum_{i=1}^n X_i^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

χ^2 est encore une variable de hasard. La déduction de la **densité de probabilité** de χ^2 demande beaucoup de temps et de place (voir, par exemple lit. Kreyszig, Bibl. A10). Pour des raisons de volume nous devons renoncer de reproduire la déduction ici (voir annexe). On obtient:

Théorème:

Densité de probabilité de χ^2

$$f(x) = \begin{cases} 0 & x \leq 0 \\ K_n x^{\frac{n-2}{2}} e^{-\frac{x}{2}} & x > 0 \end{cases}$$

$$K_n := \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$$

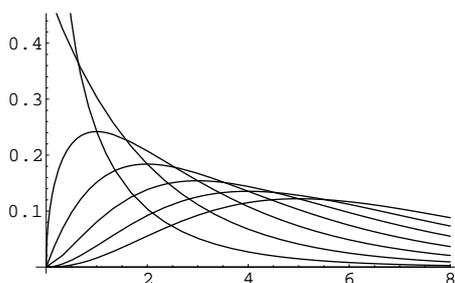
On obtient la fonction de distribution par intégration:

Formule:
$$F(x) = K_n \int_0^x u^{\frac{n-2}{2}} e^{-\frac{u}{2}} du$$

(K_n a été calculé sur la base de l'exigence $\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} f(x) dx = 1$.)

Définition:

n s'appelle **nombre de degrés de liberté** de la distribution. Γ est la **fonction gamma**.



L'image montre $f(x)$ pour les degrés de liberté $n = 0, 1, \dots, 7$

Fonction gamma, répartition gamma et bêta

Pour la fonction gamma il vaut:

1. $\Gamma(\alpha) := \int_0^{\infty} e^{-t} t^{\alpha-1} dt$
2. $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ (Intégration partielle)
3. $\Gamma(1) = \int_0^{\infty} e^{-t} t^{1-1} dt = \int_0^{\infty} e^{-t} dt = 1$
4. $\Gamma(2) = 1 \cdot \Gamma(1) = 1!$, $\Gamma(3) = 2 \cdot \Gamma(2) = 2!$, \dots , $\Gamma(n + 1) = n!$
5. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(\frac{3}{2}) = \frac{1}{2} \Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$ etc.

Remarque:

La distribution du Khi-deux est un cas spécial de la **distribution gamma**.

Définition:

Soit $\alpha > 0$

La **distribution gamma** est donnée par la fonction de densité suivante:

$$f(x) = \begin{cases} 0 & x < 0 \\ e^{-x} \frac{x^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \end{cases}$$

Définition:

Soit $\alpha > 0$, $\beta > 0$

$B(\alpha, \beta)$ s'appelle **fonction bêta**:

$$B(\alpha, \beta) := \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Définition:

Soit $\alpha > 0$, $\beta > 0$

La **distribution bêta** est donnée par la fonction de densité suivante:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} & x > 0 \end{cases}$$

4.12.4 Théorèmes sur la distribution du Khi-deux

A l'aide de la formule pour les moments on calcule:

Théorème:

Hyp.:

Distribution du Khi-deux

Thè.:

1. $\mu = n$
2. $\sigma^2 = 2n$

Pour des n grands on peut approximer la distribution du Khi-deux par la distribution normale. Il vaut:

Théorème:

1.

La variable aléatoire χ^2 est répartie de façon normale asymptotique avec $\mu = n$ et $\sigma^2 = 2n$. Pour des grands n il vaut:

$$F(x) \approx \Phi\left(\frac{x-n}{\sqrt{2n}}\right)$$

2.

La variable aléatoire $\sqrt{2}\chi^2$ est répartie de façon normale asymptotique avec $\mu = \sqrt{2n-1}$ et $\sigma^2 = 1$. Pour des grands n il vaut:

$$F(x) \approx \Phi(\sqrt{2}x\sqrt{2n-1})$$

4.12.5 Distribution de Student

Déduction

La **distribution de Student** est la base de tests importants. Elle a été publiée de W.S. Gosset sous le pseudonyme "Student".

Quant aux fonctions d'échantillonnage indépendantes X (distribué de façon normale) et Y (distribué de façon χ_n^2) nous formons une nouvelle fonction d'échantillonnage t (ou T):

$$T = \frac{X}{\sqrt{Y}} \sqrt{n}, \quad t = \frac{X}{\sqrt{Y}/\sqrt{n}}$$

Se rappeler: $X_i \in N(\dots), i = 1, 2, \dots, n \rightsquigarrow \bar{X} \in N(\dots)$

P.ex. pour $\mu = 0$ et $\sigma^2 = 1$ ($N(0, 1)$) on obtient:

$$T = \frac{\bar{X}}{S} \sqrt{m} = \frac{\bar{X}}{\sqrt{Y/m}}, \quad Y = S^2, \quad m = n - 1 \quad (\text{ Voir interv. de confiance.})$$

Définition:

La distribution liée à T resp. t s'appelle **distribution de Student**.

On va prouver plus tard: $Z = \bar{X}$ est distribuée de façon normale ($N(0, 1)$) et $Y = S^2$ est distribuée de façon $\chi_{m=n-1}^2$.

Définition:

n resp. m s'appelle **nombre de degrés de liberté** de la distribution.

Nous faisons la déduction de la **densité de probabilité** de T plus tard (voir, par exemple lit. Kreyszig, Bibl. A10 ou annexe). On obtient ($T \rightsquigarrow z$):

Théorème:

$$f(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \pi \Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{z^2}{n})^{(n+1)/2}}$$

$$F(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \pi \Gamma(\frac{n}{2})} \cdot \int_{-\infty}^z \frac{1}{(1 + \frac{u^2}{n})^{(n+1)/2}} du$$

Définition:

Pour $n = 1$ on obtient la **distribution de Cauchy**.

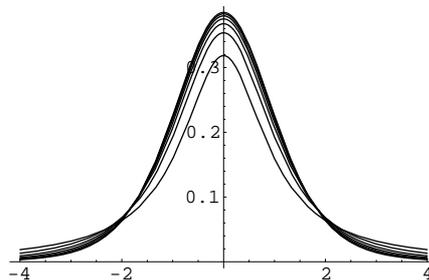
Théorèmes

Formule:

1. $n = 1$:
 \rightsquigarrow La distribution de Cauchy n'a ni de moyenne ni de variance.
2. $n = 2 \rightsquigarrow$ pas de variance
3. $n \geq 2, n \in \mathbb{N} \rightsquigarrow \mu = 0$ (z^2 , symétrie!)
4. $n \geq 3 \Rightarrow \sigma^2 = \frac{n}{n-2}$

L'image montre $f(x)$ pour les degrés de liberté
 $n = 1, \dots, 7$

Il vaut:



Théorème:

Pour $n \rightarrow \infty$ la fonction de distribution $f(x)$ ($z \rightsquigarrow x$) de la distribution de Student approche la fonction de la distribution normale standardisée ($N(0, 1)$).

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \pi \Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{x^2}{n})^{(n+1)/2}} \rightarrow \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

4.12.6 Distribution de Fisher

Soient données deux échantillons indépendants l'un de l'autre dont les tailles sont n_1 et n_2 . On les utilise pour former les fonctions d'échantillon (variables aléatoires) $\bar{X}_1, \bar{X}_2, S_1^2, S_2^2$: \rightsquigarrow

$$\begin{aligned} \bar{X}_1 &= \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}, & S_1^2 &= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 \\ \bar{X}_2 &= \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i}, & S_2^2 &= \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2 \end{aligned}$$

S_1^2 et S_2^2 sont indépendantes.

Soient les données de bases liées distribuées de façon normale aux paramètres $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$. En plus nous supposons:

Hyp.: $\sigma_1^2 = \sigma_2^2$

Maintenant nous formons la fonction d'échantillon suivante:

$$\mathcal{F} = \frac{S_1^2}{S_2^2} = \frac{\left(\frac{\chi_1^2 \sigma_1^2}{n_1 - 1}\right)}{\left(\frac{\chi_2^2 \sigma_2^2}{n_2 - 1}\right)} = \frac{\left(\frac{\chi_1^2}{n_1 - 1}\right)}{\left(\frac{\chi_2^2}{n_2 - 1}\right)} := \frac{\left(\frac{\chi_1^2}{m_1}\right)}{\left(\frac{\chi_2^2}{m_2}\right)}, \quad m_1 = n_1 - 1, \quad m_2 = n_2 - 1$$

Définition:

La distribution liée à \mathcal{F} s'appelle **distributeur de Fisher** ou **distribution \mathcal{F}** avec (m_1, m_2) degrés de liberté.

On calcule la fonction de distribution et la densité comme il suit:

4.13.2 Densité de la distribution Khi-deux

$X_i : \in N(0, 1) \rightsquigarrow$ distribution normale, $\mu = 0$, $\sigma^2 = 1$. Il vaut:

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. $X_i : \in N(0, 1) \Rightarrow$

$$\forall_i F_1(x) = F_i(x) = P(X_i^2 \leq x) = P(-\sqrt{x} \leq X_i \leq \sqrt{x}) = 2 \cdot P(0 \leq X_i \leq \sqrt{x}) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{u^2}{2}} du$$

$$\text{Soit } v = u^2, \quad u = v^{\frac{1}{2}} \Rightarrow \frac{du}{dv} = \frac{1}{2} v^{-\frac{1}{2}}, \quad du = dv \frac{1}{2} v^{-\frac{1}{2}} = dv \frac{1}{2} \frac{1}{\sqrt{v}}$$

$$\Rightarrow F_1(x) = \frac{2}{\sqrt{2\pi}} \frac{1}{2} \int_0^{\sqrt{v}=u=\sqrt{x}} e^{-\frac{v}{2}} \cdot \frac{1}{\sqrt{v}} dv = \frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{v}=u=\sqrt{x}} \frac{e^{-\frac{v}{2}}}{\sqrt{v}} dv$$

$$\Rightarrow f_1(x) = \frac{dF_1(x)}{dx} = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{x}{2}}}{\sqrt{x}}, \quad x > 0$$

4. $n = 1 \Rightarrow \chi^2 = X_1^2, \rightsquigarrow f(x) = K_1 x^{\frac{1-2}{2}} e^{-\frac{x}{2}} = \frac{1}{2^{\frac{1}{2}} \Gamma(\frac{1}{2})} \cdot x^{\frac{1-2}{2}} e^{-\frac{x}{2}} = \Big|_{\Gamma(\frac{1}{2})=\sqrt{\pi}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} \cdot e^{-\frac{x}{2}} \quad \text{☺}$

On a donc prouvée la formule pour $n = m = 1$. Maintenant nous pouvons essayer de prouver la formule générale par induction.

$$\text{Soit } \chi^2 = X_1^2 + \dots + X_n^2 := \chi_n^2, \quad f(x) := f_n(x)$$

La formule soit juste pour $m = n - 1$.

$$\rightsquigarrow f_{n-1}(x) = K_{n-1} x^{\frac{n-1-2}{2}} e^{-\frac{x}{2}} = \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} \cdot x^{\frac{n-1-2}{2}} e^{-\frac{x}{2}} = \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} \cdot x^{\frac{n-3}{2}} e^{-\frac{x}{2}}$$

$$\text{Il vaut: } \chi_n^2 = \chi_{n-1}^2 + X_n^2 \quad \chi_{n-1} \leftrightarrow f_{n-1}(x) \wedge \chi_n \leftrightarrow f_n(x) = f_1(x)$$

D'après l'hypothèse les X_i sont indépendantes. Par conséquent, χ_{n-1} et X_n^2 sont aussi indépendantes. Nous pouvons donc utiliser la formule suivante pour la densité de la page 119:

$$f_n(x) = \int_{-\infty}^{\infty} f_n(x-y) \cdot f_{n-1}(y) dy = \int_0^x f_1(x-y) \cdot f_{n-1}(y) dy$$

$$\left(\int_{-\infty}^{\infty} \dots = \int_0^x \text{ à cause de } y < 0 \Rightarrow f_{n-1}(y) = 0, \quad x > y \Rightarrow x - y < 0 \Rightarrow f_1(x - y) = 0 \right)$$

$$\rightsquigarrow f_n(x) = \int_0^x f_1(x-y) \cdot f_{n-1}(y) dy = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2^{\frac{n-1}{2}} \cdot \Gamma(\frac{n-1}{2})} \cdot \int_0^x \frac{e^{-\frac{x-y}{2}}}{\sqrt{x-y}} \cdot y^{\frac{n-3}{2}} \cdot e^{-\frac{y}{2}} dy$$

$$= \frac{1}{2^{\frac{n}{2}} \cdot \sqrt{\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot e^{-\frac{x}{2}} \cdot \int_0^x (x-y)^{-\frac{1}{2}} \cdot y^{\frac{n-3}{2}} dy := \frac{1}{2^{\frac{n}{2}} \cdot \sqrt{\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot e^{-\frac{x}{2}} \cdot I, \quad I = \int_0^x (x-y)^{-\frac{1}{2}} \cdot y^{\frac{n-3}{2}} dy$$

$$\begin{aligned} \text{Soit } y &:= ux, \quad y \in [0, x] \Rightarrow u = \frac{y}{x} \in [0, 1], \quad x - y = x - ux = x(1 - u), \quad \frac{dy}{du} = x, \quad dy = du x \\ &\leadsto I = \int_0^x (x(1-u))^{-\frac{1}{2}} \cdot (ux)^{\frac{n-3}{2}} x du = (x^{-\frac{1}{2}} x^{\frac{n-3}{2}} x) \int_0^x (1-u)^{-\frac{1}{2}} \cdot (u)^{\frac{n-3}{2}} du = x^{\frac{n-2}{2}} \int_0^x (1-u)^{-\frac{1}{2}} \cdot u^{\frac{n-3}{2}} du \end{aligned}$$

Théorème sur la fonction gamma de la page 126

$$\begin{aligned} &\leadsto I = x^{\frac{n-2}{2}} \int_0^x (1-u)^{\frac{1}{2}-1} \cdot u^{\frac{n-1}{2}-1} du = x^{\frac{n-2}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{1}{2})} \Gamma(\frac{n-1}{2} + \frac{1}{2}), \quad a = \frac{1}{2} > 0, \quad b = \frac{n-1}{2} > 0, \quad (n > 2) \\ \Gamma(\frac{1}{2}) &= \sqrt{\pi} \Rightarrow f_n(x) = \frac{1}{2^{\frac{n}{2}} \cdot \sqrt{\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot e^{-\frac{x}{2}} \cdot x^{\frac{n-2}{2}} \frac{\Gamma(\frac{n-1}{2}) \cdot \sqrt{\pi}}{\Gamma(\frac{n-1}{2} + \frac{1}{2})} = \frac{1}{2^{\frac{n}{2}}} \cdot e^{-\frac{x}{2}} \cdot x^{\frac{n-2}{2}} \frac{1}{\Gamma(\frac{n}{2})} = f(x) \quad \text{☺} \end{aligned}$$

C'est la formule qu'on a dû démontrer.

4.13.3 Densité de la distribution de Student

Distributions selon hypothèse:

$$\begin{aligned} X &\leadsto N(0, 1) \leadsto X \leftrightarrow f(x) := f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\ Y &\leadsto Y = \chi_n^2 \leadsto X \leftrightarrow f(y) := f_2(y) = \frac{1}{2^{\frac{n}{2}}} \cdot e^{-\frac{y}{2}} \cdot y^{\frac{n-2}{2}} \frac{1}{\Gamma(\frac{n}{2})}, \quad y > 0 \quad (y \leq 0 \Rightarrow f_2(y) = 0) \end{aligned}$$

X, Y indépendantes $\leadsto f(x, y) = f_1(x) f_2(y)$

$$\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 \leadsto F(z) = \int \int_{x \leq z \cdot y/n} f(x, y) dx dy = \\ &= \int_{x \leq z \cdot \sqrt{y/n}, y > 0} \int 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}, y > 0} 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}, 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{Soit } x := u \sqrt{\frac{y}{n}} \Rightarrow dx = du \sqrt{\frac{y}{n}}, \quad H := 1 + \frac{u^2}{n}, \quad \frac{Hy}{2} = \frac{y}{2} + \frac{x^2}{2} \\ &\leadsto F(z) = \frac{C_n}{\sqrt{n}} \int_0^\infty \left(\int_{-\infty}^{u \sqrt{\frac{y}{n}} = x = 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^{u=z} \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^{u=z} \left(\int_{-\infty}^{u=z} e^{-\frac{Hy}{2}} \cdot y^{\frac{n-1}{2}} dy \right) du \quad \text{Soit } Hy := 2v, \quad y = \frac{2v}{H}, \quad dy = dv \frac{2}{H} \leadsto \\ &\leadsto 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\left(\frac{n+1}{2}\right)}{\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) \end{aligned}$$

$$\leadsto F(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \cdot \Gamma(\frac{n}{2})} \cdot \left(\int_{u=-\infty}^{u=z} \left(\frac{1}{1+\frac{u^2}{n}}\right)^{\frac{n+1}{2}} du \right), \quad f(z) = \frac{dF(z)}{dz} = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \cdot \Gamma(\frac{n}{2})} \cdot \left(\frac{1}{1+\frac{z^2}{n}}\right)^{\frac{n+1}{2}} \quad \text{😊}$$

4.13.4 Preuve d'inéquation de Tschebyscheff

Soit Y une variable aléatoire quelconque, qui ne doit pas être distribuée de façon normale. Soit ici $E(Y) = \mu$ la valeur d'espérance, $Var(Y) = \sigma^2$ soit ici la variance et ε soit un nombre positif et quelconque. Alors il vaut:

Formule: **Tschebyscheff**

$$P(|Y - \mu| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}$$

Remarque:
$$P(|Y - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \Leftrightarrow P(|Y - \mu| < \varepsilon) > 1 - \frac{\sigma^2}{\varepsilon^2}$$

Définition: **Variable aléatoire non-négative X :** N'atteint que des valeurs \mathbb{R}_0^+ avec la probabilité 1.

Nous prouvons d'abord un lemme:

Lemme: **Hyp.:** X non-négative, $\delta > 0$

Thè.:
$$\frac{E(X)}{\delta} \leq P(X \geq \delta)$$

Preuve: (Lemme)

1. Cas discret:

$$E(X) = \sum_{k \in \{k\}} x_k p_k \geq \sum_{k \in \{k \mid x_k \geq \delta\} := H} x_k p_k \geq \sum_{k \in H} x_{min} p_k \geq \sum_{k \in H} \delta p_k \geq \delta \sum_{k \in H} p_k = \delta P(X \geq \delta)$$

2. Cas continu:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \geq \int_0^{\infty} x f(x) dx \geq \int_{\delta}^{\infty} x f(x) dx \geq \int_{\delta}^{\infty} x_{min} f(x) dx \geq \int_{\delta}^{\infty} \delta f(x) dx \\ &= \delta \int_{\delta}^{\infty} f(x) dx = \delta P(X \geq \delta) \end{aligned}$$

Preuve: (Théorème)

Soit $\delta = \varepsilon^2$, $X = |Y - \mu|^2 \leadsto$ non-négative

$$\begin{aligned} \sigma^2 &= E(|Y - \mu|^2) = E((Y - E(Y))^2) = E(X) \geq \delta P(X \geq \delta) = \varepsilon^2 P((|Y - E(Y)|)^2 \geq \varepsilon^2) \\ &= \varepsilon^2 P(|Y - E(Y)| \geq \varepsilon) \Rightarrow P(|Y - E(Y)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \end{aligned}$$

4.14 Annexe II: Suppléments

4.14.1 La somme des carrés contre somme de valeur absolue

1.

Question: Pour quel c la somme des carrés $\sum_i (x_i - c)^2$ devient-elle minimale?

$$\begin{aligned} \leadsto \frac{d}{dc} \sum_{i=1}^n (x_i - c)^2 &= \sum_{i=1}^n (-1) \cdot 2(x_i - c) = -2 \sum_{i=1}^n (x_i - c) = 0 \Rightarrow \sum_{i=1}^n x_i = \sum_{i=1}^n c = n \cdot c \\ \Rightarrow c &= \frac{1}{n} \cdot \sum_{i=1}^n x_i = \bar{x} \Rightarrow c = \bar{x} \leadsto c = \text{moyenne} \end{aligned}$$

2.

Question: Pour quel c la somme des valeurs absolue $\sum_i |x_i - c|$ devient-elle minimale?

$$\begin{aligned} \leadsto \frac{d}{dc} \sum_{i=1}^n |x_i - c| &= \frac{d}{dc} \sum_{i=1}^n \text{sgn}(x_i - c) \cdot (x_i - c) =_{|x_i \neq c, \text{sgn}=\text{const}} \sum_{i=1}^n (-1) \cdot \text{sgn}(x_i - c) \\ &= -(\underbrace{-1 - 1 - \dots - 1}_{x_i < c}) + (\underbrace{+1 + 1 + \dots + 1}_{x_i > c}) = 0 \\ \Rightarrow -(\underbrace{-1 - 1 - \dots - 1}_{x_i < c}) &= (\underbrace{+1 + 1 + \dots + 1}_{x_i > c}) \\ \leadsto (\text{Nombre de "valeurs } < c\text{"}) &= (\text{nombre de "valeurs } > c\text{"}) \leadsto c = \text{médian} \end{aligned}$$

Conséquence:

$\sum_i (x_i - c)^2$ devient minimal pour la moyenne $c = \bar{x}$, $\sum_i |x_i - c|$ est minimale pour le médian $c = \tilde{x}$.

4.14.2 Conformité du type de répartition etc.

A la page 119 on a vu le théorème suivant:

Théorème:

Somme de variables indépendantes et distribuées de façon normale

Hyp.:

Soient X_1, X_2, \dots, X_n indépendantes, distribuées de façon normale avec
Moyennes $\mu_1, \mu_2, \dots, \mu_n$
Variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$

Thè.:

$\sum_{i=1}^n X_i$ distribuées de façon normale

Moyenne $\mu = \sum_{i=1}^n \mu_i$

Variance $\sigma^2 = \sum_{i=1}^n \sigma_i^2$

Les sommes de variables distribuées de façon normale sont de nouveau distribuées de façon normale. On observe ce comportement aussi aux autres types de répartition:

Théorème:	<u>Hyp.:</u>	X, Y indépendantes $X : \in Bi(n_1, p), Y : \in Bi(n_2, p)$
	<u>Thè.:</u>	$X + Y : \in Bi(n_1 + n_2, p)$

Quant à la preuve dans un cas spécial:

Soit X = nombre de réalisations de l'événement A avec $p(A) = p$ si l'expérience affilée est répétée n_1 fois et Y = nombre de réalisations de l'événement A avec $p(A) = p$ si l'expérience affilée est répétée n_2 fois. Évidemment $X + Y$ est liée à la même expérience avec $p(A) = p$: $X + Y$ = nombre de réalisations de l'événement A si l'expérience affilée est répétée $n_1 + n_2$ fois. La répartition est toujours une répartition de Bernoulli (binomiale). (Quant à la preuve voir aussi lit. A4)

Comme la répartition binomiale sous la condition $np = \lambda = \text{const.}$ s'approche à une répartition de Poisson, on peut adapter le théorème aussi à une répartition de Poisson:

Théorème:	<u>Hyp.:</u>	X, Y indépendantes $X : \in Po(\lambda_1), Y : \in Bi(\lambda_2)$
	<u>Thè.:</u>	$X + Y : \in Bi(\lambda_1 + \lambda_2)$

Quant aux preuves des théorèmes suivants le lecteur est renvoyé à la littérature (si ce n'est pas encore terminé):

Théorème:	<u>Hyp.:</u>	X, Y indépendantes X, Y réparties de façon χ^2 $X : \in \chi_{m_1}^2, Y : \in \chi_{m_2}^2$
	<u>Thè.:</u>	$X + Y : \chi_{m_1+m_2}^2$

Théorème:	<u>Hyp.:</u>	X_1, X_2, \dots, X_n indépendantes X_i : réparties de façon $N(\mu, \sigma^2)$ $H = \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}}$
	<u>Thè.:</u>	$H : \in t_{n-1}$

4.14.3 Théorème limite central

A la page 101 on a vu le théorème de Moivre/ Laplace. Avec n variables aléatoires il a la forme suivante:

Théorème:	<u>Hyp.:</u>	Soit $H_n : \in Bi(n, p)$, $p \in (0, 1)$, $n \in \mathbb{N}$ $Y_n := \frac{H_n - E(H_n)}{\sqrt{Var(H_n)}} = \frac{H_n - E(H_n)}{\sqrt{np(1-p)}}$ $\rightsquigarrow Y_n$ standardisée F_{Y_n} fonction de distribution
	<u>Thè.:</u>	$\lim_{n \rightarrow \infty} F_{Y_n} = \Phi(x) = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R}$

Remarque:

Le théorème est important si pour des n grands, il s'agit de calculer approximativement des probabilités dans le contexte de distributions binomiales. Il est donc $\mu = np$, $\sigma = \sqrt{np(1-p)}$. Alors il vaut:

Conclusion:

$$P(a \leq X \leq b) = P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right)$$

Il est facile d'évaluer Φ (ordinateur, tableaux, ...). L'approximation est déjà bonne si la condition $np(1-p) > 9$ est satisfaite **condition de Laplace**.

Nous écrivons $A \in R_i$, si l'événement A est réalisé lors de la i -ième répétition d'une expérience. Autrement nous écrivons $A \notin R_i$. Soit H_n la variable comme il suit:

$$H_n(A) = \sum_{i=1}^n X_i, \quad X_i = \begin{cases} 1 & A \in R_i \\ 0 & A \notin R_i \end{cases}$$

On voit tout de suite que chaque X_i pour lui-même est distribué de façon binomiale avec les paramètres $n = 1$ et p , ç.v.d. avec la valeur d'espérance $\mu = np = p$ et la variance $\sigma^2 = np(1-p) = p(1-p)$. Comme nous savons, la variable $\sum_{i=1}^n$ est distribuée de façon binomiale comme somme de variables distribuées de façon binomiale.

$$\text{Soit} \quad Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sqrt{\sigma}/\sqrt{n}} \rightsquigarrow F_{Y_n} \rightarrow \Phi(n), \quad n \rightarrow \infty$$

Ce comportement est aussi prouvable si $\langle X_n \rangle$ est une suite quelconque de variables aléatoires indépendantes et identiquement distribuées avec une valeur d'espérance μ commune et une variance σ^2 commune, $\sigma^2 \in (0, k)$, $k = \text{const.} \in \mathbb{R}^+$.

Théorème:

Théorème limite central

Hyp.:

Comme décrit

Thè.:

$$\lim_{n \rightarrow \infty} F_{Y_n} = \Phi(x) = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R}$$

On trouve des preuves dans la littérature, p.ex. dans Bibl. A3.

4.14.4 Transformations linéaires

A la page 79 nous avons vu le théorème:

Théorème:

Hyp.:

$$g(X) = ah(X) + bu(X)$$

Thè.:

$$E(g(X)) = aE(h(X)) + bE(u(X))$$

E à la page 79 nous avons vu:

Théorème:

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

\leadsto **Spécialement:** $E(X) = \mu$, $X^* = c_1 X + c_2 = g(X)$

Corollaire:

$$\underline{\text{Hyp.:}} \quad X^* = c_1 X + c_2$$

$$\underline{\text{Thè.:}} \quad \mu^* = c_1 \mu + c_2$$

En outre nous déduisons:

$$X^* - \mu^* = (c_1 X + c_2) - (c_1 \mu + c_2) = c_1 (X - \mu), \quad E((X^* - \mu^*)^k) = E(c_1^k (X - \mu)^k) = c_1^k E((X - \mu)^k)$$

\leadsto **Spécialement:** $k = 2$: $\sigma_{X^*}^2 = E((X^* - \mu^*)^2) = c_1^2 E((X - \mu)^2) = c_1^2 \sigma^2$

Corollaire:

$$\underline{\text{Hyp.:}} \quad X^* = c_1 X + c_2$$

$$\underline{\text{Thè.:}} \quad \sigma_{X^*}^2 = c_1^2 \sigma^2$$

Kapitel 5

Statistique mathématique

5.1 Contrôle de qualité

5.1.1 Généralités, SQC

Sous contrôle de qualité et spécialement contrôle de qualité statistique (SQC, statistical quality control) on entend des méthodes qui sont utilisées pour l'assurance de qualité, l'examen de la qualité et la gestion de qualité. Aujourd'hui, les conditions sont souvent objet de normalisations, par exemple D.I.N. 55350. Généralement on peut distinguer entre les contrôles qualitatifs ou contrôles par attributs (p.ex. bon, mauvais) et les contrôles quantitatifs ou contrôles par variables (mesurage de variables).

Selon l'utilisation on distingue **deux domaines principaux**:

1. Contrôle statistique de processus, surveillance de fabrication ou réglage de qualité (**SPC**, statistical process control):

Ici, il s'agit de la surveillance d'un processus de fabrication pendant la production de produits de masse, pour découvrir des différences de qualité et pour pouvoir intervenir et conduire directement.

2. Contrôle de réception ou examen d'échantillon de réception (AC, acceptance sampling):

Ici, il s'agit du contrôle d'entrée, d'un contrôle pendant la production et d'un contrôle final des produits dans une entreprise (ou usine) sans influence directe sur la production. Ainsi le montant de rebut produit est mesuré. Le contrôle initial sert aussi à refuser la marchandise arrivante. Elle n'influence par conséquent la production que de manière indirecte.

5.2 SQC1: Contrôle de processus

5.2.1 Problème

Exemple:

Il faut observer des valeurs prescrites par rapport à certaines tolérances. P.ex. lors de la fabrication en série de roues dentées le diamètre du forage central doit montrer une valeur prescrite μ_0 .

Par expérience on sait qu'il y a des écarts qui apparaissent malgré tout le soin. Les raisons sont la défaillance humaine, usures ou abrasions aux machines et appareils etc. ainsi que des problèmes de matériel. Par conséquent il est essentiel de surveiller la production de façon permanente au moyen de SPC et d'intervenir immédiatement s'il est nécessaire. Le but est de limiter le rebut à un minimum.

5.2.2 Exemple moyenne

Dans la pratique, il y a des problèmes différents que nous ne pouvons pas tous traiter ici. Par conséquent nous choisissons un problème comme exemplaire. Lors de la production automatique d'une série d'un article il faudrait constamment observer la moyenne μ_0 pour qu'elle reste dans une certaine marge de tolérance non critique.

Procédé méthodique (manière d'agir):

1. Hypothese:

(a)

Nous savons par expérience et par conséquent nous l'acceptons aussi maintenant: $\bar{X} = \mu$ satisfait une distribution normale avec l'écart-type σ connu. Cette supposition est cependant seulement raisonnable dans un certain domaine parce que le domaine de définition de la répartition normale est $(-\infty, \infty) = \mathbb{R}$ tandis que le domaine, dans lequel on peut pratiquement trouver des valeurs de mesures x_i et par conséquent les moyennes \bar{x} , est $(0, M_{max})$. M_{max} dépend des possibilités pratiques. Des valeurs de mesure négatives n'existent pas. μ_0 est prescrit, pour σ nous pouvons utiliser provisoirement une valeur estimative que nous pouvons obtenir d'échantillons déjà acceptés.

(b)

L'hypothèse suivante H_0 soit satisfaite: La valeur prescrite μ_0 est respectée strictement pendant toute la production, ç.v.d. $H_0 : \mu = \mu_0$. L'alternative à cette hypothèse serait $H_1 : \mu \neq \mu_0$.

Définition:

H_0 s'appelle **hypothèse nulle**, H_1 s'appelle **hypothèse alternative**.

2.

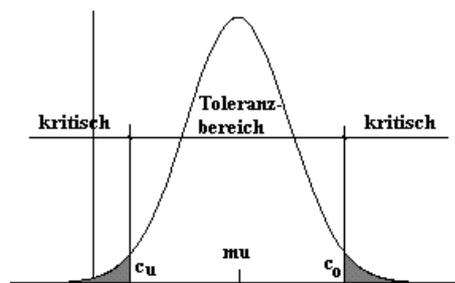
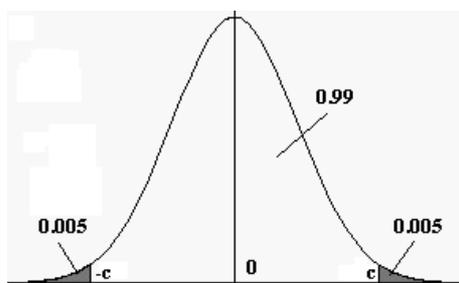
Selon des intervalles de temps réguliers, on prélève à la production des échantillons de la taille invariable n . Après on en calcule la moyenne d'échantillon $\bar{x} \rightsquigarrow$ **expérience de mesure**.

3.

En outre nous avons appris de l'expérience qu'il est raisonnable de corriger la production, si la probabilité que l'hypothèse alternative $H_1 : \mu \neq \mu_0$ soit réalisée dépasse $\alpha = 0.01$ ou 1%. (Si α est atteint, ça signifie donc quelque chose! On a atteint le niveau de signification.) Ainsi c'est en ordre, car jusqu'à présent la production a montré environ la loi de la répartition normale et on a accepté 1% de rebut. Pour comprendre mieux nous considérons la distribution normale $N(0, 1)$ (à gauche) comparée avec $N(\mu, \sigma^2)$ (à droite):

Définition:

α s'appelle **niveau de signification**. Le nombre marque une limite à laquelle il se passe quelque chose d'important, dans la plupart des cas d'après l'expérience.



On passe de la variable aléatoire distribuée de façon normale \bar{X} à la variable aléatoire standardisée U par la transformation suivante:

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

Utilisant α , la condition est la suivante dans la forme standardisée:

$$P(-c \leq U \leq c) \leq 1 - \alpha = 1 - 0.01 = 0.99, \quad \pm c = \sqrt{n} \cdot \frac{c_{o,u}^* - \mu_0}{\sigma}, \quad c_{o,u}^* = \mu + \frac{\pm c \cdot \sigma}{\sqrt{n}}$$

4. il vaut:
$$P(-c \leq U \leq c) = \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{x^2}{2}} dx \leq 0.99 \rightsquigarrow c = ?$$

On peut aussi déduire:

$$P(-c \leq U \leq c) = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1 = 2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^c e^{-\frac{x^2}{2}} dx$$

on trouve à l'aide de Mathematica (ou dans un tableau):

```
FindRoot[1/(Sqrt[2 Pi]) Integrate[E^(-x^2/2), {x, -c, c}] == 0.99, {c, 1}]
```

$$\rightsquigarrow 2.57582 \leq c \leq 2.57583, \quad c \approx 2.5758 \Rightarrow c_o^* \approx \mu + \frac{2.5758 \cdot \sigma}{\sqrt{n}},$$

$$[c_u^*, c_o^*] \approx \left[\mu - \frac{2.5758 \cdot \sigma}{\sqrt{n}}, \mu + \frac{2.5758 \cdot \sigma}{\sqrt{n}} \right]$$

5.

Dès que \bar{x} sort de l'intervalle $[c_u^*, c_o^*]$, il faut réagir! L'écart de \bar{x} n'est plus classé comme aléatoire et donc n'est plus toléré. Elle est regardée comme significative et intolérable.

6.

Souvent on choisit pour les **limites critiques** $c_{\alpha,u,o}^* = c_{0.01,u,o}^*$ des **limites d'avertissement** $d_{\beta,u,o}^* = d_{0.05,u,o}^*$. Le calcul pour d avec $\beta = 0.05$ correspond au calcul pour c avec $\alpha = 0.01$.

$$\leadsto [d_u^*, d_o^*] \approx [\mu - \frac{1.95996 \cdot \sigma}{\sqrt{n}}, \mu + \frac{1.95996 \cdot \sigma}{\sqrt{n}}]$$

7. **Exemple avec des nombres:**

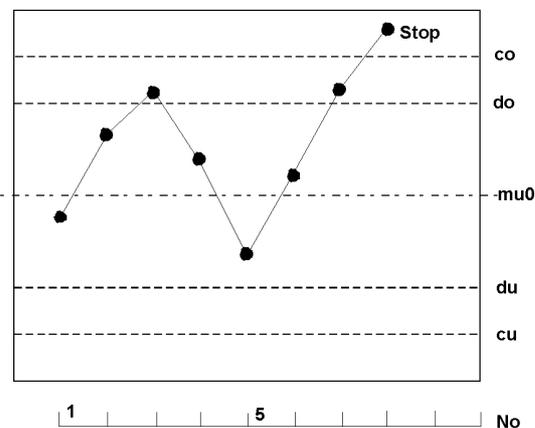
Soit $\mu = 120$, $\sigma = 0.01$, $n = \dots$

$c_{0.01,u,o}$	$n = 1$	$n = 2$	$n = 10$	$n = 20$	$n = 100$
c_u	119.974	119.982	119.992	119.994	119.997
c_o	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_o	120.020	120.014	120.006	120.004	120.002

5.2.3 Technique de surveillance à l'aide de cartes de contrôle

Souvent il est usuel de représenter les résultats des échantillonnages graphiquement et clairement. Dans l'industrie on utilise des différentes **cartes de contrôle**. Nous donnons exemplairement une carte de contrôle concernant une série d'échantillons concernant l'exemple susdit.

$$\mu = 120, \sigma = 0.01, n = 10, \alpha = 0.01, \beta = 0.05, c_{\alpha,u,o}^*, d_{\beta,u,o}^*$$



5.3 SQC2: Contrôle d'acceptation

Contrôle d'acceptation \leadsto de réception etc..

Echantillons et randomisation

Problème:

Etant clients nous recevons une grande livraison d'une marchandise. On pense par exemple à des roues dentées. Comme nous sommes obligés de respecter les tolérances assurées par contrat pour pouvoir

monter les pièces sans problèmes et de pouvoir garantir le fonctionnement, nous devons examiner les tolérances des pièces. Nous effectuons par conséquent un **contrôle d'acceptation** pour pouvoir décider sur la base des contrats si la livraison peut être acceptée ou non. S'il s'agit par exemple d'examiner de composition chimique (p.ex. examiner si les vernis (glaçure, couverte) des services de table contiennent du plomb), une destruction de la pièce lors de l'examen est souvent nécessaire. Comme d'autre part un examen de toutes les parties cause des frais très hauts, nous nous limitons aux échantillons et les méthodes mathématiques pour tirer des conclusions valables pour la totalité de l'ensemble fondamental. On peut contrôler les échantillons plus exactement qu'un grand nombre de pièces, où la routine remplace le soin. Si on considère les conséquences d'une fausse décision à propos de l'acceptation et de l'utilisation d'une livraison dans la production, on s'aperçoit la survie d'une entreprise en peut dépendre.

Les modalités d'un **contrôle d'acceptation** du client doivent être fixées selon les conventions dans le plan de contrôle. Le fournisseur peut aussi effectuer exactement le même contrôle pour lui-même. Ici, elle s'appelle le **contrôle de sortie**, **contrôle final** etc..

Au tirage de l'échantillon, le hasard du choix des pièces joue un rôle considérable pour exclure des falsifications systématiques et des préférences. Par conséquent il est très dangereux de choisir aveuglément quelques premières pièces. Le choix doit être fait d'après un système de **randomiser** le choix. P.ex. on peut attribuer un numéro à chaque pièce. Alors on choisit les numéros qu'on trouve dans une liste randomisée ou qui ont été générés par un pseudo-générateur de nombres aléatoires. Quant au problème de générer des nombres aléatoires, le lecteur est renvoyé à la littérature à cause de notre cadre étroit. (Possibilité: Prendre des chiffres du nombre π développé en fraction décimale.)

5.3.1 Exemple répartition hypergéométrique, modèle d'urne

En haut on a vu qu'on peut généralement distinguer entre les contrôles qualitatifs ou contrôles par attributs (p.ex. en ordre, pas en ordre) et les contrôles quantitatifs ou contrôles par variables (mesurage de variables).

Maintenant nous considérons un exemple de la classe des contrôles d'acceptation qualitatifs. Soit donné une livraison de marchandises à N parties. M parties en sont défectueuses et par conséquent $N - M$ parties sont en ordre. Au lieu de considérer une livraison de marchandises, nous pouvons aussi penser à N boules dans une urne: M sont noires et $N - M$ sont blanches.

De la livraison resp. de l'urne, nous tirons un échantillon aléatoire consistant en n pièces et sans les remettre, $n \leq N$. De la théorie nous savons que les résultats possibles de cette expérience forment un modèle pour la répartition hypergéométrique. Soit $X = m$ le nombre des pièces défectueuses qu'on vient de tirer — ou lors du modèle d'urne le nombre de boules noires tirées. Sois $m \in \{1, 2, \dots, n\}$. Alors nous connaissons de la répartition hypergéométrique:

$$p_d = \frac{M}{N}, \quad M = p_d \cdot N, \quad p_m = P(X = m) = \frac{\binom{M}{m} \cdot \binom{N - M}{n - m}}{\binom{N}{n}}$$

Exemple: $N = 120$, $M = 4$, $n = 3$

$$\begin{aligned} \leadsto p_0 = P(X = 0) &= \frac{\binom{4}{0} \cdot \binom{120-4}{3-0}}{\binom{120}{3}} \approx 0.902507, \\ p_1 = P(X = 1) &\approx 0.0950007, \quad p_2 = P(X = 2) \approx 0.00247828, \quad p_3 = P(X = 3) \approx 0.000014243 \\ &\Rightarrow P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1 \end{aligned}$$

5.3.2 Contrat d'acceptation et plan d'échantillonnage

Le producteur et le consommateur resp. le fournisseur et le client doivent être d'accord en ce qui concerne les conditions d'acceptation de la livraison. Les conditions d'acceptation doivent être réglées dans un contrat, le contrat de livraison ou le contrat d'acceptation. La description des conditions d'acceptation débouchent dans un plan d'examen, le **plan d'échantillonnage** dans lequel les détails de l'examen d'acceptation sont définis.

Nous étudions comme exemple, qui va avec la situation décrite en haut, un plan d'échantillonnage simple (**SSP, simple sampling plan**).

Hyp.:

Soit X le nombre de pièces défectueuses resp. de boules noires dans l'échantillon tiré ($0 \leq X = x \leq n$).

Alors nous définissons:

Définition:

Un **plan d'échantillonnage simple** est donné par:

1. La taille de l'échantillon n
2. Un nombre d'acceptation (**nombre critique**) $c \in \mathbb{R}_0^+$
3. Une règle de décision:

$X \leq c \leadsto$	acceptation
$X > c \leadsto$	lot refusé

Problème:

Généralement on ne connaît pas $p_d = \frac{M}{N} \in [0, 1]$ lors d'une livraison. Par conséquent p_d est une valeur inconnue d'une variable aléatoire Y . Ainsi il est nécessaire d'examiner la probabilité d'acceptation de la livraison en fonction de p_d par des modèles, pour pouvoir appliquer pratiquement les connaissances acquises de façon théorique à l'intention de diminuer les frais.

Dans l'exemple discuté en haut (répartition hypergéométrique, modèle d'urne) il était $X = m$ et $M = p_d \cdot N$. Il vaut donc pour $p_m = P(X = m)$:

$$p_m = P(X = m) = \frac{\binom{p_d \cdot N}{m} \cdot \binom{N - p_d \cdot N}{n - m}}{\binom{N}{n}} = \frac{\binom{p_d \cdot N}{m} \cdot \binom{N(1 - p_d)}{n - m}}{\binom{N}{n}}$$

$$\Rightarrow P(X = m \leq c) = \sum_{i, m_i \leq c} P(X = m_i) = \sum_{m=0}^{m \leq c} P(X = m) = \sum_{m=0}^{m \leq c} \frac{\binom{p_d \cdot N}{m} \cdot \binom{N(1 - p_d)}{n - m}}{\binom{N}{n}}$$

$P(X = m \leq c)$ est la probabilité de trouver au maximum c pièces défectueuses resp. c boules noires dans l'échantillon. c est fixé par contrat, mais par contre $p_d = \frac{M}{N}$ est une réalisation inconnue de Y . Donc $P(X = m \leq c)$ est une fonction de Y aux paramètres connus N, n, c . $P(X = m \leq c)$ est la **probabilité d'acceptation** $L_{N,n,c}$ qui dépend de Y :

$$L_{N,n,c}(Y) := P(X = m \leq c) = \sum_{m=0}^{m \leq c} \frac{\binom{Y \cdot N}{m} \cdot \binom{N(1 - Y)}{n - m}}{\binom{N}{n}}, \quad Y \in [0, 1]$$

Généralement on a toujours une fonction $L_c(Y) := P(X \leq c)$, $Y = p_d \in [0, 1]$ liée à un plan d'échantillonnage.

Définition:

$$L_c(Y) := P(X \leq c), \quad Y = p_d \in [0, 1]$$

s'appelle **fonction de probabilité d'acceptation** (du lot par le consommateur)

Le graphe de L_c s'appelle **courbe d'efficacité** du plan d'échantillonnage simple

Remarque:

↪ Angl. operating characteristic, **OC**-curve

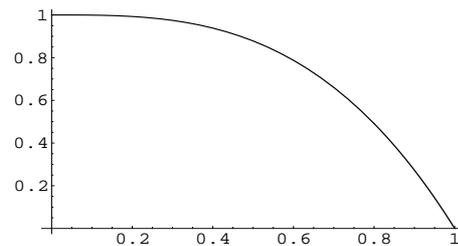
Dans l'exemple en haut c soit p.ex. $c = 2$:

$$L_{120,3,c=2}(Y) := P(X = m \leq 2) = \sum_{m=0}^{m \leq 2} \frac{\binom{Y \cdot 120}{m} \cdot \binom{120(1 - Y)}{3 - m}}{\binom{120}{3}}$$

$$= \frac{\binom{Y \cdot 120}{0} \cdot \binom{120(1 - Y)}{3}}{\binom{120}{3}} + \frac{\binom{Y \cdot 120}{1} \cdot \binom{120(1 - Y)}{2}}{\binom{120}{3}} + \frac{\binom{Y \cdot 120}{2} \cdot \binom{120(1 - Y)}{1}}{\binom{120}{3}}$$

Spécialement:

$$L_{120,3,c=2}(0.9) \approx 0.273052$$



Qualités de L_c :

$L_c(0) = P(X \leq c)_{|M=0}$ signifie qu'il n'y a pas de pièces défectueuses et par conséquent que l'acceptation est sûre $\leadsto Y = p_d = 1$

Si le nombre des pièces défectueuses augmente, la chance de trouver des pièces défectueuses lors d'un échantillonnage augmente aussi. La probabilité d'acceptation diminue donc. Par conséquent $L_c(Y)$ doit décroître de façon monotone avec Y .

$L_c(1) = P(X \leq c)_{|M=N}$ signifie qu'il n'y a que des pièces défectueuses et que l'acceptation sera donc refusée $\leadsto Y = p_d = 0$

Qualités: $L_c(0) = 1, L_c(1) = 0, D_L = [0, 1,]$
 $L_c(Y)$ décroissant de façon monotone

5.3.3 L'exemple modèle binomial, modèle d'urne

En haut nous avons eu une urne et les paramètres M, N, n, m . Nous avons fait l'expérience sans remettre les pièces. La base était donc une répartition hypergéométrique. Maintenant nous voulons exécuter l'expérience en remettant la pièce chaque fois. Cela nous mène à une distribution binomiale.

Soit $H = \{1, 2, \dots, M, M+1, \dots, N\}$

Nous pouvons considérer le fait que nous choisissons un système ordonné à n éléments de l'ensemble H . Alors nous trouvons:

Nombre total de possibilités de choisir = N^n

Nombre total de possibilités de choisir que des pièces défectueuses = M^m

Nombre total de possibilités de choisir que des pièces bonnes = $(N - M)^{n-m}$

Nombre de possibilités dans l'échantillonnage de distribuer des places aux pièces défectueuses = $\binom{n}{m}$

Le nombre de possibilités de choisir m pièces défectueuses et au même temps $n - m$ pièces bonnes et en plus de distribuer les places aux pièces défectueuses, c.v.d. de mélanger les pièces défectueuses et les autres, ceci nous donne le nombre d'échantillonnages à m pièces défectueuses en remettant la pièce chaque fois après l'avoir tirée. Si on divise encore par le nombre total de possibilités de choisir N^n , on obtient la probabilité de Laplace $P(X = m)$:

$$\frac{M}{N} := p \Rightarrow P(X = m) = \binom{n}{m} \cdot \frac{M^m \cdot (N - M)^{n-m}}{N^n} = \binom{n}{m} \cdot \frac{M^m}{N^m} \cdot \frac{(N - M)^{n-m}}{N^{n-m}} = \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m}$$

Comme nous savons, nous avons ici une répartition binomiale.

La loi suivante justifie le nom:

$$\sum_{m=1}^n p_m = \sum_{m=1}^n \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m} = (1 + (1-p))^n = 1^n = 1$$

Hyp.:

Soit $n \ll N, m \in \{0, 1, \dots, n\}, p = \frac{M}{N}, N \rightarrow \infty$

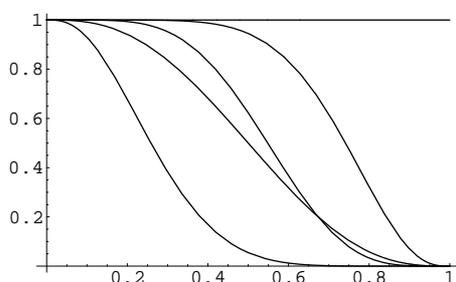
↪

Alors on peut approximer la répartition hypergéométrique par la répartition binomiale (en pratique pour $10n \leq N$):

$$\lim_{N \rightarrow \infty} \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}} = \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m}$$

Dans ce cas il vaut pour la fonction de probabilité d'acceptation:

$$L_{n,c}(Y) = P(X \leq c) = \sum_{m=0}^{m \leq c} \binom{n}{m} \cdot Y^m \cdot (1-Y)^{n-m}$$



A gauche courbes pour:

$(n, c) = (5, 2), (10, 2), (10, 5), (10, 7), (5, 5)$

Observation: Un c plus grand pousse la courbe à droite. Un n plus grand pousse la courbe vers le bas.

5.3.4 Risque du producteur, risque du consommateur

Utilisant les paramètres n et c , $L_{n,c}(Y) = L_{n,c}(p_d) = P(X \leq c)_{|Y=p_d}$ donne la probabilité $P(X \leq c)$ dépendante de $Y = p_d$ si p_d est inconnu. Comme p_d est inconnu et lorsqu'on doit décider quand-même ($X \leq c \rightsquigarrow$ acceptation, $X > c \rightsquigarrow$ lot refusé), il peut arriver que la décision est au désavantage du producteur ou du consommateur, parce que p_d a été mal estimé.

Si la **limite de refus** ou l'estimation $Y = p_\beta$ de la portion de rebut effectif p_d est fixée trop bas par le consommateur ($Y = p_\beta < p_d$), il peut arriver, que la **livraison est acceptée d'après le plan d'acceptation**, bien qu'elle contienne une portion de **rebut intolérable**. Car la probabilité d'acceptation pourrait devenir trop grande.

Définition:

L'acceptation d'un lot par le consommateur malgré trop de rebut s'appelle **erreur de la 2-ième sorte**. La probabilité d'acceptation nécessaire $\beta = L_{n,c}(p_\beta) = P(X \leq c)$ s'appelle le **risque de consommation**. p_β s'appelle **limite de refus**.

(Hit! A réfléchir ...)

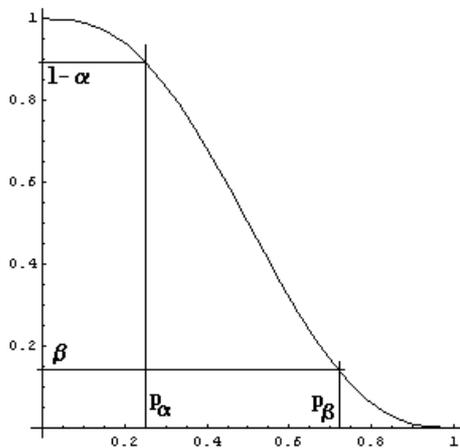
Au fur et à mesure nous définissons:

Définition:

Le refus d'un lot bien qu'il soit en ordre et la portion de rebut soit tolérable s'appelle **erreur de la 1-ère sorte**. La probabilité liée $\alpha = P(X > c) = 1 - P(X \leq c) = 1 - L_{n,c}(p_\alpha)$ s'appelle **risque du producteur**. p_α s'appelle **limite d'acceptation**.

Comme α et β sont des probabilités d'erreurs, il faut construire le plan d'échantillonnage (contrat!) de façon que les deux grandeurs deviennent aussi petites que possible. Comme on voit dans le diagramme, il est évident qu'on atteint ce fait si on choisit une courbe très raide (modélisation!) et si p_α est situé le plus que possible à gauche et p_β à droite.

Il faut y penser que les n plus grands coûtent davantage.

**Idée:**

Soit emp = empirique

Choisir

$$\left(\frac{M}{N}\right)_{emp} \leq p_\alpha \leq \left(\frac{M}{N}\right)_{emp} + \varepsilon$$

Si M devient plus grand pendant la production, on devrait maintenant le constater!

Remarque:

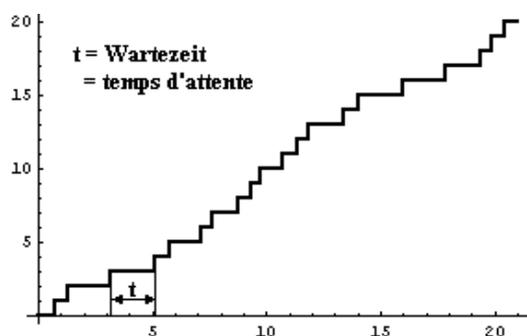
Echantillonnage: \leadsto p_α s'appelle aussi **la limite de qualité acceptable** ou **AQL** (acceptable quality level). p_β s'appelle aussi **la limite de qualité tolérée** ou **LQ, LQL, RQL** (limiting quality level, rejectable quality level).

Normalisations:

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 Files d'attente, processus de Poisson

5.4.1 Processus de Poisson



Soit $X(t)$ = nombre d'événements aléatoires du temps 0 jusqu'à t

\leadsto La variable aléatoire (stochastique) dépend du temps (\leadsto processus) et en plus elle est discrète ($X(t) \in \mathbb{N}$).

Définition:

Si $X(t)$ satisfait les conditions susdites, nous parlons d'un **processus stochastique**.

Comme par sa nature $X(t)$ ne peut pas diminuer, il vaut:

Lemme: $X(t)$ s'accroît de façon monotone.

Hyp.:

Pour simplifier le problème, nous partons ici de la situation de la condition initiale $X(0) = 0$.

Soit $\Delta X(t_0, t) := X(t_0 + t) - X(t_0)$

\leadsto

L'augmentation $\Delta X(t_0, t)$ de X dans l'intervalle $[t_0, t_0 + t]$ dépend de t_0 et t .

Ici nous considérons un cas simple:

Définition:

Le processus s'appelle **homogène**, si $\Delta X(t_0, t)$ est indépendant de t_0 .

Alors il vaut:

$$\forall t_i : \Delta X(t_0, t) = \Delta X(t_1, t) \Rightarrow \Delta X(t_0, t) = \Delta X(0, t) = X(0 + t) - X(0) = X(t) - 0 = X(t)$$

Lemme:**Hyp.:** $X(t)$ homogène

$$\Delta X(t_0, t) := X(t_0 + t) - X(t_0)$$

Thè.:

$$\Delta X(t_0, t) = \Delta X(0, t) = X(t)$$

Soit $p_m(t) := P(X(t) = m)$ $\rightsquigarrow p_m(t)$ est la probabilité que dans $[0, t]$ il y a exactement m événements.**Définition:**

Deux événements s'appellent **statistiquement indépendants**, si l'apparition de n'importe quel des deux événements n'a aucune influence sur l'apparition de l'autre des deux événements.

Nous considérons la situation suivante:

$$t_0 < t_1 < t_2 < \dots < t_n, \quad \Delta X_r := X(t_r) - X(t_{r-1}), \quad r \in \{1, 2, 3, \dots, n\}$$

Les intervalles $I_r = [t_r, t_{r-1}]$ n'ont rien en commun sauf les bords s'il s'agit d'intervalles voisins. C'est pourquoi nous voulons présupposer ici que les ΔX_r soient statistiquement indépendants.

 \rightsquigarrow

$$S = \sum_{m=0}^{\infty} p_m(\Delta t) = \text{probabilité qu'il y a dans } \Delta t \text{ 0 ou 1 ou 2 ou } \dots \text{ ou un nombre infini d'événements.}$$

Chaque nombre d'événements est donc possible $\rightsquigarrow S = 1$. \rightsquigarrow **Lemme:**

$$1 - p_0(\Delta t) - p_1(\Delta t) = \sum_{m=2}^{\infty} p_m(\Delta t) := S_2$$

S_2 est la probabilité que dans Δt il y a plus de 2 événements qui se sont réalisés. En pratique, il est souvent raisonnable de présupposer que pour des Δt petits il vaut $S_2 \approx 0$. C'est ce que nous voulons faire aussi ici:

Hyp.:

$$S_2 = 1 - p_0(\Delta t) - p_1(\Delta t) = \sum_{m=2}^{\infty} p_m(\Delta t) \approx 0$$

Conséquence: $\rightsquigarrow p_0(\Delta t) \approx 1 - p_1(\Delta t)$

En pratique il est en outre raisonnable de présupposer que la probabilité pour un événement $p_1(\Delta t)$ et pour des temps Δt petits soit proportionnel au temps. C'est ce que nous voulons faire ici:

Hyp.:

$$p_1(\Delta t) \approx a \cdot \Delta t, \quad \Delta t \ll \text{const.}, \quad a > 0$$

Théorème:

Hyp.:

1. $X(0) = 0$
2. $X(t)$ homogène
3. ΔX_r ($r = 1, 2, \dots$) statist. indép.: $\in A$
4. $0 < \Delta t \ll \text{const.} \Rightarrow p_0(\Delta t) \approx 1 - p_1(\Delta t)$
5. $0 < \Delta t \ll \text{const.} \Rightarrow p_1(\Delta t) \approx a \cdot \Delta t, \quad a > 0$
6. $S_2 = \sum_{m=2}^{\infty} p_m(\Delta t) = \varphi(\Delta t) \cdot (\Delta t)^2, \quad |\varphi(\Delta t)| < \text{const.}$

Thè.:

$X(t)$ est réparti d'après Poisson avec $\lambda = a \cdot t$:

$$p_m(t) = P(X(t) = m) = \frac{(at)^m}{m!} e^{-at} = \frac{\lambda^m}{m!} e^{-\lambda}$$

Remarque:

Modèle pour: $X(t)$ = nombre de clients qui arrivent à un bouton de service ou $X(t)$ = nombre de remplacements de composants avec l'espérance de vie identique et la même durée de fonctionnement.

Preuve:

1.

$X(t)$ est la fréquence absolue qui s'accroît de façon monotone ≤ 0

$$\rightsquigarrow X(t + \Delta t) = X(t) + X(\Delta t)$$

$$\rightsquigarrow p_0(t + \Delta t) = P(X(t + \Delta t) = 0) = P(X(t) + X(\Delta t) = 0) = P(X(t) = 0 \wedge X(\Delta t) = 0)$$

$$=_{|\in A} P(X(t) = 0) \cdot P(X(\Delta t) = 0) = p_0(t) \cdot p_0(\Delta t) \approx p_0(t) \cdot (1 - p_1(\Delta t)) \approx p_0(t) \cdot (1 - a \cdot \Delta t)$$

$$= p_0(t) - p_0(t) \cdot a \cdot \Delta t \Rightarrow \frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = \frac{p_0(t) - p_0(t) \cdot a \cdot \Delta t - p_0(t)}{\Delta t} = -p_0(t) \cdot a$$

$$\Rightarrow \frac{dp_0(t)}{dt} = -p_0(t) \cdot a, \quad p_0(0) = \underbrace{P(X(0) = 0)}_{O.k.} = 0$$

$$\Rightarrow \text{PVI} \rightsquigarrow \text{Solution connue: } p_0(t) = e^{-at}$$

2.

$$p_m(t + \Delta t) = P(X(t + \Delta t) = m) = P(X(t) + X(\Delta t) = m) =$$

$$P((X(t) = m \wedge X(\Delta t) = 0) \vee (X(t) = m-1 \wedge X(\Delta t) = 1) \vee \dots \vee (X(t) = 0 \wedge X(\Delta t) = m))$$

$$=_{|\in A} P(X(t) = m) \cdot P(X(\Delta t) = 0) + P(X(t) = m-1) \cdot P(X(\Delta t) = 1) + \dots + P(X(t) = 0) \cdot P(X(\Delta t) = m)$$

$$= p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) + p_{m-2}(t) \cdot p_2(\Delta t) + \dots + p_0(t) \cdot p_m(\Delta t)$$

$$\leq p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) + p_{m-2}(t) \cdot 1 + \dots + p_0(t) \cdot 1$$

$$= p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) + \sum_{k=1}^m p_k(t) \approx p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) + \varphi(\Delta t) \cdot (\Delta t)^2,$$

$$\Delta t \rightarrow 0 \Rightarrow \varphi(\Delta t) \cdot (\Delta t)^2 \rightarrow 0$$

$$\rightsquigarrow p_m(t + \Delta t) \approx p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) = p_m(t) \cdot (1 - p_1(\Delta t)) + p_{m-1}(t) \cdot p_1(\Delta t)$$

$$= p_m(t) \cdot (1 - a \cdot \Delta t) + p_{m-1}(t) \cdot a \cdot \Delta t = p_m(t) - p_m(t) \cdot a \cdot \Delta t + p_{m-1}(t) \cdot a \cdot \Delta t \Rightarrow$$

$$\frac{p_m(t + \Delta t) - p_m(t)}{\Delta t} = \frac{p_m(t) - p_m(t) \cdot a \cdot \Delta t + p_{m-1}(t) \cdot a \cdot \Delta t - p_m(t)}{\Delta t} = -p_m(t) \cdot a + p_{m-1}(t) \cdot a$$

$$\Rightarrow \frac{dp_m(t)}{dt} = -a \cdot p_m(t) + a \cdot p_{m-1}(t), \quad X(0) = 0 \rightsquigarrow P(X(0) = m > 0) = p_{m>0}(0) = 0$$

$$\Rightarrow \quad \mathbf{PVI} \rightsquigarrow$$

$$(a) \quad m = 1 \rightsquigarrow \frac{dp_1(t)}{dt} = -a \cdot p_1(t) + a \cdot p_0(t) = -a \cdot p_1(t) + a \cdot e^{-at} = , \quad p_1(0) = 0$$

$$\rightsquigarrow \quad \mathbf{PVI} \rightsquigarrow \quad \text{Solution connue: } p_1(t) = a \cdot t \cdot e^{-at}$$

$$(b) \quad m = 2 \rightsquigarrow \frac{dp_2(t)}{dt} = -a \cdot p_2(t) + a \cdot p_1(t) = -a \cdot p_2(t) + a \cdot a \cdot t \cdot e^{-at} = , \quad p_2(0) = 0$$

$$\rightsquigarrow \quad \mathbf{PVI} \rightsquigarrow \quad \text{Solution connue: } p_2(t) = \frac{(a \cdot t)^2}{2} \cdot e^{-at}$$

$$(c) \quad \text{Induction} \rightsquigarrow \quad p_m(t) = \frac{(a \cdot t)^m}{m!} \cdot e^{-at}$$

Conséquence:

La probabilité que dans l'intervalle temporel $[t_0, t_0 + \Delta t]$ aucun événement n'est réalisé est indépendante de t_0 et par conséquent $= p_0(\Delta t) = e^{-a \Delta t}$.

Soit T = temps'attente entre deux événements, par exemple entre deux pannes (défaillances) de composants ou machines.

$\rightsquigarrow R(t) := P(T > t) = p_0(t) = e^{-at}$ = probabilité qu'à cause de $T > t$ jusqu'à t , aucun composant tombe en panne

$\rightsquigarrow F(t) := 1 - R(t)$ est la **probabilité de survivre**

$\rightsquigarrow F(t) := 1 - R(t) = P(T \leq t) = 1 - e^{-at} \rightsquigarrow$ Répartition exponentielle

Théorème:**Hyp.:**

Comme au théorème dernier
 $R(t) := P(T > t)$

Thè.:

$F(t) := P(T \leq t) = 1 - e^{-at}$

Conséquence:

Ici, la fréquence $X(t)$ était distribuée selon la loi de Poisson. Par conséquent on peut donc conclure que les attentes T sont réparties selon la loi exponentielle $P(T \leq t) = 1 - e^{-at}$.

Remarque:

On peut aussi prouver le théorème inverse.

Théorème:**Hyp.:**

$$F(t) = P(T \leq t) = 1 - e^{-a \cdot t}, \quad \lambda = a \cdot t$$

Thè.:

$$p_m(t) = \frac{\lambda \cdot t}{m!} \cdot e^{-\lambda \cdot t}$$

5.4.2 Modèles de files d'attente

Lors de l'usage pratique des mathématiques dans le domaine de la planification, le problème des queues ou files d'attente joue un grand rôle (quiches, imprimante, transmission, etc.). Considérons par ex. le problème de clients qui font la queue devant une caisse. Le nombre de clients qui se trouve devant la caisse par unité temporelle est distribué d'après Poisson. Le temps que le client emploie pour être servi est pour la plupart distribué exponentiellement. Les deux processus entrent donc en relation. Leur combinaison détermine un système des files d'attente. Pour approfondir ce problème intensivement, il nous faudrait une ample théorie, ce qui déborderait le cadre de ce cours.

5.5 Estimations

5.5.1 Estimations ponctuelles

Problème:

Soit donné un échantillon de valeurs $\{x_1, x_2, \dots, x_n\}$ d'un ensemble de valeurs fondamental U qui soit inconnu dans sa totalité. Les valeurs sont atteintes une variable X dont on peut avoir un modèle pour la fonction de répartition. Pour U il faudrait estimer des paramètres comme μ, σ^2 etc. à l'aide de paramètres empiriques et connus de l'échantillon. Comme méthode, nous utilisons la méthode des moments. Les paramètres à estimer devraient être exprimés par les moments de la répartition $\mu = E(X)$, $\sigma = E((X - \mu)^2)$ etc.. Dans cette intention nous utilisons des **fonctions d'estimation** ou des **estimations ponctuelles**. Les **valeurs estimées** sont des réalisations d'estimations ponctuelles en nombres. Si on veut par contre obtenir des connaissances sur l'exactitude ou la sécurité de valeurs d'estimation, on peut calculer des **intervalles estimés** ou des **intervalles de confiance**.

Définition:

Des **fonctions d'estimation** ou des **valeurs estimées** sont des fonctions de l'échantillon par lesquelles nous voulons remplacer de façon satisfaisante des fonctions aléatoires théoriques ou des paramètres de l'ensemble fondamental.

Truc: Pour les valeurs x_i de l'échantillon nous voulons maintenant introduire individuellement de nouvelles variables X_i :

Hypothèse: x_1 soit une valeur pour X_1 , x_2 une valeur pour X_2 etc.. Les variables ainsi postulées ont toutes la même distribution que X .

Exemple:

Valeur estimée pour μ : $\bar{X} = \bar{X}(n) = \frac{1}{n} \sum_{i=1}^n X_i$

Valeur estimée pour σ^2 : $S^2 = S^2(n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ oder ou

$$Q^2 = Q^2(n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)S^2}{n}$$

Problème:

Quelle est la qualité de tels estimations ponctuelles?

5.5.2 Propriété d'être sans biais

Définition:

Une valeur estimée Y resp. Y_n pour un paramètre Υ de l'ensemble de base s'appelle **non biaisé**, si la valeur probable de Y est la même que la valeur probable de Υ . Ç.v.d. s'il vaut:

$$E(Y) = E(\Upsilon)$$

Remarque:

Soit donné un échantillonnage $\{x_1, x_2, \dots, x_n\}$. Les variables X_1, X_2, \dots, X_n soient encore réparties comme X . Il vaut donc spécialement $E(X_i) = E(X)$, $i = 1, \dots, n$ ainsi que $\sigma^2 = E((X - \mu)^2) = E((X_i - \mu)^2)$, $i = 1, \dots, n$. Alors le vecteur de valeurs $(x_1, x_2, \dots, x_n)^T$ donne une combinaison possible de valeurs qui sont atteintes par le vecteur de variables $(X_1, X_2, \dots, X_n)^T$ (T signifie „transposé“). Alors la moyenne de l'échantillonnage et la variance empirique sont des valeurs estimées pour μ et σ^2 non biaisées.

Symbole: $X_i \sim X \rightsquigarrow X_i$ répartit comme X

Théorème:

(Moyenne fidèle à l'espérance mathématique)

Hyp.:

$$X_1, X_2, \dots, X_n \sim X$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \mu = E(X)$$

\bar{X} valeur estimée pour μ

Thè.:

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

Preuve:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n \cdot \mu = \mu = E(X)$$



Théorème:

(Variance empirique fidèle à l'espérance mathématique)

Hyp.:

$$X_1, X_2, \dots, X_n \sim X$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \sigma^2 = E((X - \mu)^2)$$

 S^2 valeur estimée pour σ^2 **Thè.:**

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

Preuve:

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} \sum_{i=1}^n E((X_i - \bar{X})^2) = \frac{1}{n-1} \sum_{i=1}^n E(((X_i - \mu) + (\mu - \bar{X}))^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n E((X_i - \mu)^2 + 2(X_i - \mu)(\mu - \bar{X}) + (\mu - \bar{X})^2) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E((X_i - \mu)^2) + E(2(X_i - \mu)(\mu - \bar{X})) + E((\mu - \bar{X})^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + 2E((X_i - \mu) \left(\frac{n}{n}\mu - \frac{1}{n} \sum_{j=1}^n X_j\right)) + E\left(\left(\frac{n}{n}\mu - \frac{1}{n} \sum_{j=1}^n X_j\right)^2\right) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + \frac{2}{n^2} E((X_i - \mu) (n\mu - \sum_{j=1}^n X_j)) + \frac{1}{n^2} E((n\mu - \sum_{j=1}^n X_j)^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + \frac{2}{n^2} E((X_i - \mu) \sum_{j=1}^n (\mu - X_j)) + \frac{1}{n^2} E\left(\left(\sum_{j=1}^n (\mu - X_j)\right)^2\right) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + \frac{2}{n^2} \sum_{j=1}^n E((X_i - \mu) (\mu - X_j)) + \frac{1}{n^2} \sum_{j=1}^n E((\mu - X_j)^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 + \frac{2}{n^2} \sum_{j=1}^n \underbrace{E(-(X_i - \mu)(X_j - \mu))}_{\substack{i \neq j: \\ \text{indep.} \Rightarrow = 0}} + \frac{n}{n^2} \sigma^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 - \frac{2}{n^2} n E((X_i - \mu)^2) + \frac{n}{n^2} \sigma^2 \right) = \frac{1}{n-1} \left(\sum_{i=1}^n \sigma^2 - \frac{2}{n} \sigma^2 + \frac{1}{n} \sigma^2 \right) = \frac{\sigma^2}{n-1} \left(n - \frac{2n}{n} + \frac{n}{n} \right) \\ &= \frac{\sigma^2}{n-1} (n - 2 + 1) = \frac{\sigma^2}{n-1} (n - 1) = \sigma^2 \quad \text{☺} \end{aligned}$$

Il vaut:

$$Q^2 = Q^2(n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)S^2}{n} \Rightarrow E(Q^2) = \frac{(n-1)}{n} E(S^2) = \frac{(n-1)}{n} \sigma^2 \neq \sigma^2 \sim$$

Corollaire: Q^2 non biaisé.**Remarque:**En outre on peut démontrer que S n'est pas non biaisé (voir lit.Storm, Bibl. A12).

Remarque:

Une condition plus faible que la fidélité à l'espérance est la fidélité à l'espérance asymptotique:

$$\lim_{n \rightarrow \infty} E(Y_n) = E(\Upsilon)$$

5.5.3 Consistance

Soit de nouveau $X_i \sim X$.

Définition:

Une valeur estimée Y resp. Y_n pour un paramètre Υ de l'ensemble de base s'appelle **consistant** ou **consistant de façon faible**, s'il vaut:

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|Y_n - \Upsilon| < \varepsilon) = 1$$

C.-à.-d. avec la taille d'échantillon grandissante la probabilité que l'estimation ponctuelle se distingue aussi peu que possible du paramètre à estimer, converge vers 1.

Théorème:**Hyp.:**

$$\begin{aligned} \bar{X} &= \bar{X}(n) = \frac{1}{n} \sum_{i=1}^n X_i \\ S^2 &= S^2(n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ Q^2 &= Q^2(n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)S^2}{n} \end{aligned}$$

Thè.:

1. \bar{X} est une estimation consistante pour μ
2. $\lim_{n \rightarrow \infty} \text{Var}(S^2) \rightarrow 0$, $\lim_{n \rightarrow \infty} \text{Var}(Q^2) \rightarrow 0$ oder \Rightarrow ou $X_i \in [-M, M]$
 S^2 et Q^2 sont des estimations consistantes pour σ^2

Pour la preuve nous utilisons le lemme:

Lemme:**Hyp.:**

$$\begin{aligned} \forall_i X_i &\sim X \\ E(X_i) &= E(X) = \mu \\ \text{Var}(X_i) &= \text{Var}(X) = \sigma^2 \end{aligned}$$

Thè.:

1. $E(\bar{X}) = \mu$
2. $\text{Var}(\bar{X}) = E((\bar{X} - \mu)^2) = E(\bar{X}^2) - \mu^2 = \frac{\sigma^2}{n}$

Preuve:

(Lemme)

$$\begin{aligned} 1. E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{n}{n} \mu = \mu \\ 2. \text{Var}(\bar{X}) &= E((\bar{X} - \mu)^2) = E\left(\left(\frac{1}{n} \sum_{i=1}^n X_i - \frac{n}{n} \mu\right)^2\right) = E\left(\frac{1}{n^2} \left(\sum_{i=1}^n X_i - (n\mu)\right)^2\right) \\ &= \frac{1}{n^2} E\left(\left(\sum_{i=1}^n (X_i - \mu)\right)^2\right) = \frac{1}{n^2} E\left(\sum_{i,j=1}^n (X_i - \mu)(X_j - \mu)\right) = \frac{1}{n^2} \sum_{i,j=1}^n E(\underbrace{(X_i - \mu)(X_j - \mu)}_{i \neq j: \text{indép.} \Rightarrow =0}) \\ &= \frac{1}{n^2} \sum_{i=1}^n E((X_i - \mu)^2) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

Quant à la preuve du théorème:Utiliser Tschebyscheff: $P(|Y - E(Y)| \geq \varepsilon) \leq \frac{\text{Var}(Y)}{\varepsilon^2}$

$$\begin{aligned} 1. \lim_{n \rightarrow \infty} P(|Y_n - \Upsilon| < \varepsilon) &= \lim_{n \rightarrow \infty} P(|\bar{X} - \mu| < \varepsilon) = \lim_{n \rightarrow \infty} (1 - P(|\bar{X} - \mu| \geq \varepsilon)) \\ &= 1 - \lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \varepsilon), \quad P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n \varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty \\ &\Rightarrow \lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \varepsilon) = 0 \Rightarrow \lim_{n \rightarrow \infty} P(|\bar{X} - \mu| < \varepsilon) = 1 \\ 2. \lim_{n \rightarrow \infty} P(|Y_n - \Upsilon| < \varepsilon) &= \lim_{n \rightarrow \infty} P(|S^2 - \sigma^2| < \varepsilon) = 1 - \lim_{n \rightarrow \infty} P(|S^2 - \sigma^2| \geq \varepsilon), \\ P(|S^2 - \sigma^2| \geq \varepsilon) &= P(|S^2 - E(S^2)| \geq \varepsilon) \leq \frac{\text{Var}(S^2)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \text{Var}\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\ (X_i - \bar{X}), (X_j - \bar{X}) &\text{ indépendants pour } i \neq j \\ \text{(p. 116)} \rightsquigarrow \frac{1}{\varepsilon^2} \text{Var}\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) &= \frac{1}{\varepsilon^2} \sum_{i=1}^n \text{Var}\left(\frac{1}{n-1} (X_i - \bar{X})^2\right) \\ &= \frac{1}{\varepsilon^2} \sum_{i=1}^n E\left(\left(\frac{1}{n-1} (X_i - \bar{X})^2 - \frac{1}{n-1} E((X_i - \bar{X})^2)\right)^2\right) \\ &= \frac{1}{\varepsilon^2 (n-1)^2} \sum_{i=1}^n \underbrace{E(((X_i - \bar{X})^2 - E((X_i - \bar{X})^2))^2)}_{\leq \text{Const} = M} \leq \frac{M n}{\varepsilon^2 (n-1)^2} \rightarrow 0, \quad n \rightarrow \infty \end{aligned}$$

Autres notions

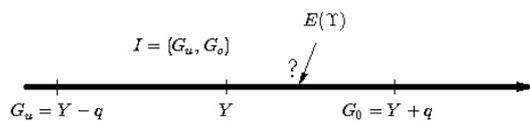
Efficacité, exhaustivité, rendement etc, voir lit.

5.5.4 Intervalles de confiance I

Notions

Si on veut gagner des connaissances sur l'exactitude ou la sécurité de valeurs d'estimation, on peut calculer des **intervalles estimés** ou des **intervalles de confiance**.

Soit Y resp. Y_n une estimation ponctuelle pour un paramètre Υ de l'ensemble de base.



Soit $I = (G_u, G_o)$
 $P(G_u < E(\Upsilon) < G_o) = \varepsilon = 1 - \alpha$

Exemple:

$\varepsilon = 0.95 \hat{=} 95\%$ oder ou $\varepsilon = 0.99 \hat{=} 99\%$

Définition:

$I = (G_u, G_o)$ s'appelle **intervalle de confiance** à la **probabilité de confiance** $\varepsilon = 1 - \alpha$.

Remarque:

Les intervalles de confiance intéressent en pratique!

Exemple moyenne

Exemple:

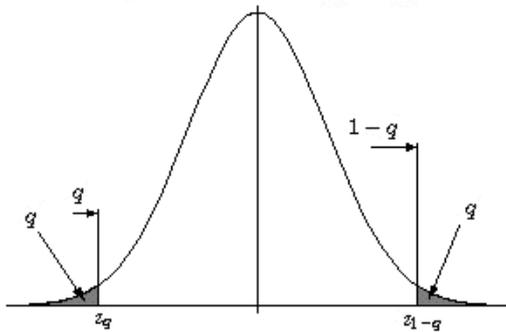
Soit X réparti d'après Gauss avec la variance σ^2 qu'on connaît. On cherche l'intervalle de confiance μ pour $\varepsilon = 1 - \alpha$.

Nous savons:

$$X \in N(\mu, \sigma^2) \Rightarrow \bar{X} \in N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$X \in N(\mu, \sigma^2) \Leftrightarrow Z = \frac{X - \mu}{\sigma} \in N(0, 1)$$

$$\bar{X} \in N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \bar{Z} = \frac{\bar{X} - \mu}{(\sigma/\sqrt{n})} \in N(0, 1)$$



Pour ce qu'on va faire, nous étudions la répartition normale $N(0, 1)$. Si on travaille en utilisant des calculateurs, la restriction à la situation standard n'est pas nécessaire. Si on travaille comme jadis avec les tableaux, il faut la restriction.

Nous savons:

$$1. F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = P(X \leq x) = \Phi(x; \mu, \sigma^2)$$

$$2. \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = P(Z \leq z) = \Phi(z; 0, 1)$$

$$3. Z = \frac{X - \mu}{\sigma} \in N(0; 1) \Leftrightarrow X \in N(\mu; \sigma^2)$$

$$\rightsquigarrow \text{ Soit } \Phi(z) = \Phi(z; 0, 1) = P(X \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \rightsquigarrow \Phi(b) = P(X \leq b)$$

$$\rightsquigarrow \text{ Soit } q := \Phi(z_q) = \Phi(z_q; 0, 1) \rightsquigarrow \text{ quantile } q \text{ (Ordre } q)$$

On voit dans l'esquisse:

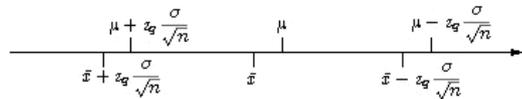
Formule: $z_q = -z_{1-q}, \Phi(z_{1-q}) = 1 - q = 1 - \Phi(z_q)$

$$\text{ Soit } z \rightsquigarrow N(0, 1), \bar{X} \rightsquigarrow N(\mu, \frac{\sigma^2}{n}) \quad \text{et} \quad F(\bar{X}) = F(\bar{X}, \mu, \frac{\sigma^2}{n}) = \Phi(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}; 0, 1) = q$$

$$\rightsquigarrow \text{ Transformation: } Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \bar{X} = \mu + Z \cdot \frac{\sigma}{\sqrt{n}}$$

$$\text{ Soit } \bar{x} = \frac{1}{n} \sum_{i_1}^n x_i \quad (\text{Valeur moyenne de l'échantillonnage})$$

$$\text{ Soit } z = z_q$$



\rightsquigarrow Il vaut: ($z_q < 0$!!!)

$$\begin{aligned} \mu \in [\bar{x} + z_q \frac{\sigma}{\sqrt{n}}, \bar{x} - z_q \frac{\sigma}{\sqrt{n}}] &\Leftrightarrow (\mu - \bar{x}) \in [z_q \frac{\sigma}{\sqrt{n}}, -z_q \frac{\sigma}{\sqrt{n}}] \Leftrightarrow (\bar{x} - \mu) \in [z_q \frac{\sigma}{\sqrt{n}}, -z_q \frac{\sigma}{\sqrt{n}}] \\ &\Leftrightarrow |\bar{x} - \mu| \leq |z_q| \frac{\sigma}{\sqrt{n}} \Leftrightarrow \bar{x} \in [\mu + z_q \frac{\sigma}{\sqrt{n}}, \mu - z_q \frac{\sigma}{\sqrt{n}}] \end{aligned}$$

Important: $\rightsquigarrow \mu \in [\bar{x} + z_q \frac{\sigma}{\sqrt{n}}, \bar{x} - z_q \frac{\sigma}{\sqrt{n}}] \Leftrightarrow \bar{x} \in [\mu + z_q \frac{\sigma}{\sqrt{n}}, \mu - z_q \frac{\sigma}{\sqrt{n}}]$

$$\rightsquigarrow \varepsilon = 1 - \alpha = P(\mu \in [\bar{X} + z_q \frac{\sigma}{\sqrt{n}}, \bar{X} - z_q \frac{\sigma}{\sqrt{n}}]) = P(\bar{X} \in [\mu + z_q \frac{\sigma}{\sqrt{n}}, \mu - z_q \frac{\sigma}{\sqrt{n}}])$$

$$= P(\mu + z_q \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu - z_q \frac{\sigma}{\sqrt{n}}) = F(\mu - z_q \frac{\sigma}{\sqrt{n}}) - F(\mu + z_q \frac{\sigma}{\sqrt{n}}) = (1 - q) - q = \Phi(z_{1-q}) - \Phi(z_q)$$

$$= \Phi(z_{1-q}) - (1 - \Phi(z_{1-q})) = 2\Phi(z_{1-q}) - 1$$

$$\Rightarrow 2 - \alpha = 2\Phi(z_{1-q}), \quad \Phi(z_{1-q}) = 1 - \frac{\alpha}{2} = \int_{-\infty}^{z_{1-q}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Formule:
$$\Phi(z_{1-q}) = 1 - \frac{\alpha}{2} = \int_{-\infty}^{z_{1-q}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad \Phi(z_q) = \frac{\alpha}{2} = \int_{-\infty}^{z_q} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Par conséquent nous appelons z_q **borne α bipartite** de la répartition de Gauss normalisée Φ . z_q est le quantile de la valeur $\frac{\alpha}{2}$.

Exemple avec des nombres: $\alpha = 0.05 \hat{=} 5\%$

$$\leadsto \Phi(z_{1-q}) = \int_{-\infty}^{z_{1-q}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1 - \frac{0.05}{2} = 0.975$$

Programme en Mathematica:

```
| FindRoot[Integrate[E^(-t^2/2)/Sqrt[2 Pi], {t, -Infinity, z}] == 0.975, {z, 1}]
```

Output:

```
| {z -> 1.95996}
```

\leadsto Nous trouvons donc $z_{1-q} \approx 1.960$.

Soient $\bar{x} = 156.38$, $\sigma^2 = 0.20$, $n = 16$.

$$\begin{aligned} \leadsto \varepsilon = 1 - \alpha = 0.95 &= P(\mu \in [\bar{X} + z_q \frac{\sigma}{\sqrt{n}}, \bar{X} - z_q \frac{\sigma}{\sqrt{n}}]) \\ &\approx P(\mu \in [156.38 - 1.960 \frac{\sqrt{0.20}}{\sqrt{16}}, 156.38 + 1.960 \frac{\sqrt{0.20}}{\sqrt{16}}]) \\ &\approx P(\mu \in [156.161, 156.599]) \leadsto I = [156.161, 156.599] = [a, b], \quad b - a \approx 0.438269 \end{aligned}$$

I est l'intervalle de confiance pour la probabilité de confiance $\varepsilon = 1 - \alpha = 0.95$. La probabilité que μ est situé à l'extérieur de cet intervalle, est $\alpha = 0.05$.

5.5.5 Fonctions de test importantes

Souvent on ne connaît ni μ ni σ . Ça signifie qu'on doit travailler avec des estimations. Alors il se pose la question d'après les types des fonctions de répartition correspondantes, si on doit remplacer μ et σ par des estimations. Comme les résultats sont proprement étonnants, nous devons maintenant déduire quelques faits.

$$\text{Soit } (\bar{X} = \sum_{i=1}^n X_i \Rightarrow \sum_{i=1}^n (X_i - \bar{X}) = (\sum_{i=1}^n X_i) - n\bar{X} = \sum_{i=1}^n X_i - n \frac{1}{n} \sum_{i=1}^n X_i = \sum_{i=1}^n X_i - \sum_{i=1}^n X_i = 0$$

$$U_i = (X_i - \bar{X}), \quad \sum_{i=1}^n U_i = 0 \rightsquigarrow$$

Les variables $U_i = (X_i - \bar{X})$ sont dépendantes.

$$\rightsquigarrow \text{ P.ex. } U_1 = - \sum_{i=2}^n U_i$$

$$\rightsquigarrow 0 = 0^2 = (\sum_{i=1}^n U_i)^2 = \sum_{i=1}^n U_i^2 + \sum_{(i \neq k \wedge i, k=1)}^n U_i \cdot U_k$$

$$\Rightarrow U_1^2 = - \sum_{i=2}^n U_i^2 + \sum_{(i \neq k \wedge i, k=1)}^n U_i \cdot U_k = - \sum_{i=2}^n U_i^2 + \sum_{(i \neq k \wedge i, k=2)}^n U_i \cdot U_k + U_1 \cdot \sum_{k=2}^n U_k + U_1 \cdot \sum_{i=2}^n U_i$$

$$= - \sum_{i=2}^n U_i^2 + \sum_{(i \neq k \wedge i, k=2)}^n U_i \cdot U_k + (- \sum_{i=2}^n U_i) \cdot \sum_{k=2}^n U_k + (- \sum_{k=2}^n U_k) \cdot \sum_{i=2}^n U_i = h(U_2, U_3, \dots, U_n) \\ \rightsquigarrow U_1^2 = h(U_2, U_3, \dots, U_n)$$

\rightsquigarrow Les variables $U_i^2 = (X_i - \bar{X})^2$ sont dépendantes.

Lemme:

Les variables $U_i = (X_i - \bar{X})$ et aussi les variables $U_i^2 = (X_i - \bar{X})^2$ sont dépendantes.

En outre il est évident qu'une transformation de coordonnées $X_i \mapsto X_i - \mu$ resp. $\bar{X}_i \mapsto \bar{X}_i - \mu$ n'a pas d'influence sur la dépendance ou indépendance. Nous utilisons le symbole suivant:

Symbole: X_i indépendant $\Leftrightarrow X_i \in \{\heartsuit\}$
 X_i dépendant $\Leftrightarrow X_i \in \{\spadesuit\}$

Lemme:

Hyp.: $\{X_i\} \in \{\heartsuit\}$ resp. $\{X_i\} \in \{\spadesuit\}$

Thè.: $\{X_i - \mu\} \in \{\heartsuit\}$ resp. $\{X_i - \mu\} \in \{\spadesuit\}$

\rightsquigarrow

\rightsquigarrow

$X_i \in N(0, \sigma^2)$ ne restreint pas d'autres conditions.

$$\text{Soit } X_i \in N(0, \sigma^2), \quad Q_n = \sum_{i=1}^n (X_i - \bar{X})^2$$

\rightsquigarrow

Il est possible d'écrire Q_n comme somme de $n - 1$ variables au carré qui sont indépendantes. C'est ce que nous allons déduire:

Lemme:

Hyp.:

$$X_i \in N(0, \sigma^2), \quad Q_n = \sum_{i=2}^n (X_i - \bar{X})^2$$

Thè.:

$$\exists_{Y_1, Y_2, \dots, Y_{n-1} \in N(0, \sigma^2), \in \{\heartsuit\}} : Q_n = \sum_{i=2}^{n-1} Y_i^2$$

Preuve:

(Lemme) Sei

$$\begin{aligned} Y_1 &= \frac{1}{\sqrt{1 \cdot 2}} (X_1 - X_2) \\ Y_2 &= \frac{1}{\sqrt{2 \cdot 3}} (X_1 + X_2 - 2X_3) \\ &\vdots \\ Y_{n-1} &= \frac{1}{\sqrt{(n-1) \cdot n}} (X_1 + X_2 + \dots + X_{n-1} - (n-1)X_n) \\ Y_n &= \frac{1}{\sqrt{n}} (X_1 + X_2 + \dots + X_{n-1} + X_n) \end{aligned}$$

$$\begin{aligned} \rightsquigarrow \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} &= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = M \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \\ M &= \begin{pmatrix} \frac{1}{\sqrt{1 \cdot 2}} & -\frac{1}{\sqrt{1 \cdot 2}} & 0 & \dots & 0 \\ \frac{1}{\sqrt{2 \cdot 3}} & \frac{1}{\sqrt{2 \cdot 3}} & -\frac{2}{\sqrt{2 \cdot 3}} & \ddots & \vdots \\ \vdots & & & \vdots & 0 \\ \frac{1}{\sqrt{(n-1) \cdot n}} & \dots & & \frac{1}{\sqrt{(n-1) \cdot n}} & -\frac{(n-1)}{\sqrt{(n-1) \cdot n}} \\ \frac{1}{\sqrt{n}} & & & & \frac{1}{\sqrt{n}} \end{pmatrix} = (\vec{c}_1, \dots, \vec{c}_n) = \begin{pmatrix} \vec{z}_1^T \\ \vdots \\ \vec{z}_n^T \end{pmatrix} \\ \text{P.ex. } \vec{z}_j^T &= \underbrace{\left(\frac{1}{\sqrt{(j) \cdot (j+1)}}, \frac{1}{\sqrt{(j) \cdot (j+1)}}, \dots, \frac{1}{\sqrt{(j) \cdot (j+1)}} \right)}_j, -\frac{j}{\sqrt{(j) \cdot (j+1)}}, 0, \dots, 0 \end{aligned}$$

Soit $j < k < n$

$$\begin{aligned} \rightsquigarrow \vec{z}_j^T \cdot \vec{z}_k &= \langle \vec{z}_j, \vec{z}_k \rangle = \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} + \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} + \dots \\ &+ \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} - \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} + 0 = \frac{j-j}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{(k) \cdot (k+1)}} = 0 \end{aligned}$$

Soit $j = k < n \rightsquigarrow$

$$\begin{aligned} \vec{z}_j^T \cdot \vec{z}_j &= \left(\frac{1}{\sqrt{(j) \cdot (j+1)}} \right)^2 + \left(\frac{1}{\sqrt{(j) \cdot (j+1)}} \right)^2 + \dots + \left(\frac{1}{\sqrt{(j) \cdot (j+1)}} \right)^2 + \left(-\frac{j}{\sqrt{(j) \cdot (j+1)}} \right)^2 = \\ j \cdot \frac{1}{(j) \cdot (j+1)} + \frac{j^2}{(j) \cdot (j+1)} &= \frac{j^2 + 1}{j^2 + 1} = 1 \end{aligned}$$

$$\begin{aligned} \text{Soit } j < k = n \rightsquigarrow \bar{z}_j^T \cdot \bar{z}_n &= \langle \bar{z}_j, \bar{z}_n \rangle = \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} + \dots \\ &+ \frac{1}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} - \frac{j}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} = \frac{j-j}{\sqrt{(j) \cdot (j+1)}} \cdot \frac{1}{\sqrt{n}} = 0 \end{aligned}$$

$$\begin{aligned} \text{Soit } j = k = n \rightsquigarrow \\ \bar{z}_n^T \cdot \bar{z}_n &= \left(\frac{1}{\sqrt{n}}\right)^2 + \left(\frac{1}{\sqrt{n}}\right)^2 + \dots + \left(\frac{1}{\sqrt{n}}\right)^2 = n \cdot \frac{1}{n} = 1 \rightsquigarrow \end{aligned}$$

$$\bar{z}_j^T \cdot \bar{z}_k = \delta_{i,k}, \quad M \cdot M^T = E \Rightarrow M^T = M^{-1} \Rightarrow \frac{1}{\det M^{-1}} = \det M = \det M^T = \det M^{-1} \Rightarrow \det M = \pm 1$$

Conséquence: \rightsquigarrow M orthogonale:

$$\begin{aligned} \bar{Y} = M \cdot \bar{X} \Rightarrow r^2 = \bar{Y}^2 &= \langle \bar{Y}, \bar{Y} \rangle = \langle M \cdot \bar{X}, M \cdot \bar{X} \rangle = (M \cdot \bar{X})^T \cdot M \cdot \bar{X} = (\bar{X}^T \cdot M^T) \cdot (M \cdot \bar{X}) = \\ \bar{X}^T \cdot (M^T \cdot M) \cdot \bar{X} &= \bar{X}^T \cdot E \cdot \bar{X} = \bar{X}^T \cdot \bar{X} = \langle \bar{X}, \bar{X} \rangle = \bar{X}^2 \Rightarrow |\bar{X}| = |\bar{Y}|, \quad \sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2 \end{aligned}$$

D'après la construction on avait:

$$\begin{aligned} Y_n &= \frac{1}{\sqrt{n}} (X_1 + X_2 + \dots + X_{n-1} + X_n) = \sqrt{n} \bar{X} \Rightarrow Y_n^2 = n \bar{X}^2 \rightsquigarrow \\ Y_n^2 &= \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2 = n \cdot \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 = n \cdot (\bar{X})^2 \Rightarrow Q_n = \sum_{i=i}^n (X_i - \bar{X})^2 = \sum_{i=i}^n (X_i^2) - 2 X_i \bar{X} + \bar{X}^2 \\ &= \left(\sum_{i=i}^n (X_i^2) \right) - (2 \bar{X} \sum_{i=i}^n X_i) + \left(\sum_{i=i}^n \bar{X}^2 \right) = \left(\sum_{i=i}^n (X_i^2) \right) - (2 \bar{X} n \bar{X}) + (n \bar{X}^2) = \left(\sum_{i=i}^n (X_i^2) \right) - (n \bar{X}^2) \\ &= \left(\sum_{i=i}^n (Y_i^2) \right) - (Y_n^2) = \sum_{i=i}^{n-1} (Y_i^2) \end{aligned}$$

Les Y_i sont indépendantes comme on voit par la définition des Y_i . Il est $Y_{i-1} = Y_{i-1}(X_1, X_2, \dots, \cdot_i)$, $Y_i = Y_i(X_1, X_2, \dots, \cdot_{i+1})$. Il est donc possible de changer la valeur d'une variable sans toucher les valeurs des autres variables. $\rightsquigarrow \odot$

Lemme:

$$\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2 = \sum_{i=1}^{n-1} Y_i^2 + Y_n = Q_n + n \bar{X}^2 = \sum_{i=2}^n (X_i - \bar{X})^2 + n \bar{X}^2 = \sum_{i=1}^{n-1} Y_i^2 + n \bar{X}^2, \quad Y_n^2 = n \bar{X}^2$$

On outre nous utilisons le théorème suivant de la page 119:

Théorème:

Somme de variables indépendantes et distribuées de façon normale

Hyp.:

Soient X_1, X_2, \dots, X_n indépendantes, distribuées de façon normale avec
Moyennes $\mu_1, \mu_2, \dots, \mu_n$
Variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$

Thè.:

$\sum_{i=1}^n X_i$ distribuées de façon normale

Moyenne $\mu = \sum_{i=1}^n \mu_i$

Variance $\sigma^2 = \sum_{i=1}^n \sigma_i^2$

Les X_i soient indépendantes et $\in N(\mu, \sigma^2)$, $k < n$. Par conséquent les termes de la somme dans Y_k , c.v.d. $\frac{X_1}{\sqrt{k \cdot (k+1)}}, \frac{X_2}{\sqrt{k \cdot (k+1)}}, \dots, -\frac{k X_{k+1}}{\sqrt{k \cdot (k+1)}}$ sont aussi indépendantes et il vaut (voir page 135):

$$\mu(Y_k) = \sum_{i=1}^n \mu_i = \frac{\mu}{\sqrt{k \cdot (k+1)}} + \frac{\mu}{\sqrt{k \cdot (k+1)}} + \dots - \frac{k \mu}{\sqrt{k \cdot (k+1)}} = \mu \left(k \frac{1}{\sqrt{k \cdot (k+1)}} - \frac{k}{\sqrt{k \cdot (k+1)}} \right) = 0$$

$$\sigma^2(Y_k) = \sum_{i=1}^n \sigma_i^2 = \frac{\sigma^2}{k \cdot (k+1)} + \frac{\sigma^2}{k \cdot (k+1)} + \dots + (-k)^2 \frac{\sigma^2}{k \cdot (k+1)} = \sigma^2 \frac{k + k^2}{k^2 + k} = \sigma^2$$

$$\Rightarrow \sigma^2\left(\frac{Y_k}{\sigma}\right) = \left(\frac{1}{\sigma}\right)^2 \cdot \sigma^2(Y_k) = \frac{\sigma^2}{\sigma^2} = 1 \rightsquigarrow$$

Lemme:

Hyp.:

$X_i \in N(\mu, \sigma^2)$, $X_i \in \{\heartsuit\}$, $k < n$,

$Y_k = \frac{1}{\sqrt{k \cdot (k+1)}} (X_1 + X_2 + \dots + X_k - k X_{k+1})$

Thè.:

$\frac{Y_k}{\sigma} \in N(0, 1)$

$$\text{Soit } Y := \frac{Q_n}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=2}^{n-1} Y_i^2 = \sum_{i=2}^{n-1} \frac{Y_i^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \rightsquigarrow$$

Y est une Somme de $n - 1$ variables au carré $\frac{Y_i}{\sigma} \in N(0, 1)$ qui sont distribuées de façon normale. Les Y_i sont indépendantes. Y est donc réparti selon la loi χ^2 avec $n - 1$ degrés de liberté! On a donc le théorème:

Théorème:**Hyp.:**

$$\begin{aligned}
& X_1, \dots, X_n \in N(\mu, \sigma^2), \in \{\heartsuit\} \\
& \bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i \\
& Y = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2
\end{aligned}$$

Thè.:

$$Y \in \chi_{n-1}^2$$

Soit $Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \rightsquigarrow$ Il vaut:

$$Z = \frac{\sqrt{n}}{\sigma} \cdot \bar{X} + \frac{(-1) \cdot \sqrt{n} \cdot \mu}{\sigma} = c_1 \bar{X} + c_2 \Rightarrow (Z \in \{\heartsuit\} \Leftrightarrow \bar{X} \in \{\heartsuit\})$$

En plus on a:

$$\begin{aligned}
Q_n &= \sum_{i=1}^n (X_i - \bar{X})^2 \Rightarrow \frac{\partial}{\partial \bar{X}} Q_n = \frac{\partial}{\partial \bar{X}} \sum_{i=1}^n (X_i - \bar{X})^2 = 2 \sum_{i=1}^n (X_i - \bar{X}) \cdot (-1) = -2 \left(\sum_{i=1}^n X_i - n \cdot \bar{X} \right) \\
&= -2 \left((n \cdot \bar{X}) - n \cdot \bar{X} \right) = 0 \Rightarrow Q_n(\bar{X}) = \text{const.}(X_1, \dots, X_n)
\end{aligned}$$

D'autre part on voit:

$$k < n - 1 \rightsquigarrow Y_k = Y_k(X_1, \dots, X_{k+1}), \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}(X_1, \dots, X_n) \rightsquigarrow \text{On peut changer le } Y_k \text{ sans changer le } \bar{X} \text{ et vice versa.}$$

$$k = n - 1 \rightsquigarrow Y_{n-1} = \lambda_1 \cdot \bar{X}(X_1, \dots, X_n) + \lambda_2 \cdot X_n \rightsquigarrow \text{On peut aussi changer le } Y_k \text{ sans changer le } \bar{X} \text{ et vice versa.} \rightsquigarrow$$

Théorème:**Hyp.:**

$$\begin{aligned}
& X_1, \dots, X_n \in N(\mu, \sigma^2), \in \{\heartsuit\} \\
& \bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i \\
& Y = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \\
& Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}
\end{aligned}$$

Thè.:

$$\{Y, Z\} \subset \{\heartsuit\}$$

Conséquence:

Avec $Z \in N(0, 1)$ et $Y \in \chi_{n-1}^2$ on peut construire la variable aléatoire $T = \frac{Z}{\sqrt{Y/(n-1)}}$ qui suit donc la loi de la répartition de Student.

On déduit donc:

$$T = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \cdot \frac{1}{\sqrt{\frac{1}{\sigma^2} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} = \sqrt{n} \frac{\bar{X} - \mu}{1} \cdot \frac{1}{\sqrt{S^2}} = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$$

Corollaire:**Hyp.:**

$$X_1, \dots, X_n \in N(\mu, \sigma^2), \in \{\heartsuit\}$$

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$$

Thè.:

$$T \in t_{n-1}$$

Remarque:

Considérons que nous connaissons maintenant la fonction de répartition de $T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$. Dans T il y a à côté des valeurs d'échantillon X_i seulement le μ . On a donc ici un moyen en main qui peut servir à obtenir de l'information sur le μ à partir des valeurs de l'échantillon!

5.5.6 Intervalles de confiance II**Problème:**

À cause de la consistance, nous savons que pour les grandes échantillons, \bar{X} est une bonne estimation de μ . Mais comment est-ce que se présente la chose pour les échantillons petits?

Estimation de la moyenne à la variance inconnue et des petits échantillons

Soient données les valeurs d'un échantillon $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$, $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ et n . Nous en

construisons la variable $T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$ qui suit la loi de Student à $n - 1$ degrés de liberté. Nous connaissons les fonctions liées:

$$f(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \pi \Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{z^2}{n})^{(n+1)/2}}, \quad F(z) = P(T \leq z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \pi \Gamma(\frac{n}{2})} \cdot \int_{-\infty}^z \frac{1}{(1 + \frac{u^2}{n})^{(n+1)/2}} du$$

On voit de $f(z)$ que la fonction de distribution est symétrique à l'axe y . Il vaut donc:

$$F(-c) = 1 - F(x) \Rightarrow P(-c \leq T \leq c) = P(T \leq c) - P(T < -c) = F(c) - F(-c) = F(x) - 1 + F(c) = 2F(x) - 1$$

Pour une probabilité donnée $P(-c \leq T \leq c) = \gamma = 1 - \alpha$ on peut donc calculer le c de l'équation $\gamma = 2F(c) - 1$.

A l'aide d'une valeur c connue on peut maintenant en conclure l'extension de l'intervalle de confiance:

$$-c \leq T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S} \leq c \Rightarrow \frac{-c \cdot S}{\sqrt{n}} - \bar{X} \leq -\mu \leq \frac{c \cdot S}{\sqrt{n}} - \bar{X} \Rightarrow \bar{X} - \frac{c \cdot S}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{c \cdot S}{\sqrt{n}},$$

$$I = \left[\bar{X} - \frac{c \cdot S}{\sqrt{n}}, \bar{X} + \frac{c \cdot S}{\sqrt{n}} \right]$$

Méthode:

1. **Donné:** $\{x_1, x_2, \dots, x_n\}$, $\gamma = 1 - \alpha$
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. $\rightsquigarrow \mu \in I = \left[\bar{X} - \frac{c \cdot S}{\sqrt{n}}, \bar{X} + \frac{c \cdot S}{\sqrt{n}} \right]$ avec la probabilité γ

Exemple avec des nombres

Soit $\gamma = 0.95$, $\alpha = 0.05$ Créer les données:

Programme en Mathematica:

```
<< 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}
```

Programme en Mathematica:

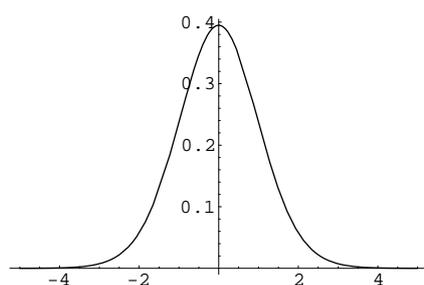
```
meanM = Mean /. LocationReport[m][[1]];
standD = StandardDeviation /. DispersionReport[m][[2]];
{LocationReport[m], DispersionReport[m], Length[m]}
```

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}
```

Programme en Mathematica:

```
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};
```



On voit dans l'esquisse la densité de la loi de Student $f(z, 24)$.

Maintenant nous calculons c directement avec Mathematica. Ceux qui n'ont pas les moyens pour un tel calcul, doivent avoir recours tableaux!

Programme en Mathematica:

```
{"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}}
```

Programme en Mathematica:

```
m1 = Select[m, (xU - 0.000001 < #1 < xD + 0.000001) &]
```

Output:

```
{245.184, 245.083, 245.631}
```

Programme en Mathematica:

```
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}
```

Solution:

Comme on peut remarquer, la moyenne μ se trouve avec une probabilité de 95% dans l'intervalle $I = [244.996, 245.822]$ ($\bar{X} \approx 245.409$). Par contre la plupart des données se trouve à l'extérieur de cet intervalle. Ça nous montre la qualité de l'estimation \bar{X} de μ .

5.5.7 Intervalles de confiance III

Problème:

À cause de la consistance, nous savons que pour les grandes échantillons, S^2 est une bonne estimation de σ^2 . Mais comment est-ce que se présente la chose pour les petits échantillons?

Estimation de la variance à la moyenne inconnue et les petits échantillons

A la page 163 on a constaté:

Théorème:**Hyp.:**

$$X_1, \dots, X_n \in N(\mu, \sigma^2), \in \{\heartsuit\}$$

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

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

Thè.:

$$Y \in \chi_{n-1}^2$$

Conséquence:

$$\gamma = P(c \leq Y) = 1 - P(0 \leq Y < c) = 1 - (F_{\chi^2, n-1}(c) - F_{\chi^2, n-1}(0)) = 1 - F_{\chi^2, n-1}(c)$$

Remarque:

Comme la variance est toujours positive, nous nous contentons ici d'une estimation seulement dans une direction, par contre à l'estimation dans deux directions pour l'intervalle de confiance pour μ .

Nous savons:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ K_n x^{\frac{n-2}{2}} e^{-\frac{x}{2}} & x > 0 \end{cases} \quad \begin{aligned} K_n &:= \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \\ F(x) &= K_n \int_0^x u^{\frac{n-2}{2}} e^{-\frac{u}{2}} du \end{aligned}$$

Nous pouvons donc calculer c de $\gamma = P(c \leq Y) = 1 - F_{\chi^2, n-1}(c)$.

$$\leadsto c \leq Y = \frac{n-1}{\sigma^2} S^2 \Rightarrow 0 \leq \sigma^2 \leq \frac{n-1}{c} S^2, \sigma^2 \in [0, \frac{n-1}{c} S^2]$$

Méthode:

1. **Donné:** $\{x_1, x_2, \dots, x_n\}$, $\gamma = 1 - \alpha$
2. $\leadsto S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, Y = \frac{n-1}{\sigma^2} S^2$
3. $\leadsto \gamma = 1 - F_{\chi^2, n-1}(c) \leadsto c$
4. $\leadsto \sigma^2 \in [0, \frac{n-1}{c} S^2]$ avec la probabilité γ

Exemple avec des nombres

Soit $\gamma = 0.95$, $\alpha = 0.05$ Créer les données:

Programme en Mathematica:

```
<< 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}
```

Programme en Mathematica:

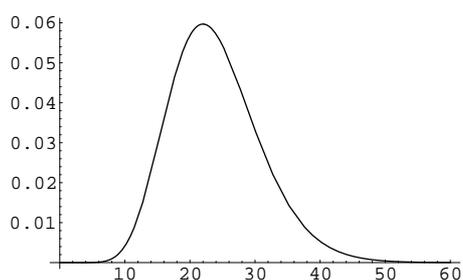
```
var = Variance /. DispersionReport[m][[1]];
{LocationReport[m], DispersionReport[m], Length[m]}
```

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}
```

Programme en Mathematica:

```
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}];
```



On voit dans l'esquisse la densité de la loi de khi deux $chi(x, 24)$.

Maintenant nous calculons c directement avec Mathematica. Ceux qui n'ont pas les moyens pour un tel calcul, doivent avoir recours aux tableaux!

Programme en Mathematica:

```
{"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}}
```

Solution:

Comme on peut remarquer, la variance σ^2 se trouve avec une probabilité de 95% dans l'intervalle $I = [0, 2.53858]$ ($S^2 \approx 0.768021$). Par d'autres "runs" on obtient: $\gamma = 0.99 \rightsquigarrow I = [0, 3.52486]$, $\gamma = 0.90$

$\rightsquigarrow I = [0, 2.15667]$

5.5.8 Intervalles de confiance IV

Remarque concernant la loi binomiale

Fonction de probabilité:

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad q = 1 - p, \quad \mu = np, \quad \sigma^2 = npq$$

Problème:

On cherche un intervalle de confiance pour p de la répartition binomiale.

Pour des n grands la formule d'approximation est valable:

$$f(x) \approx f^*(x) = \frac{1}{\sqrt{w\pi\sigma}} \cdot e^{-\frac{z^2}{2}}, \quad Z = \frac{X - \mu}{\sigma}, \quad Z \in N(0, 1)$$

Soit X le nombre de succès à l'occasion de n répétitions de l'expérience. Puis on peut argumenter comme il suit:

$$\gamma = P(-c \leq Z \leq c) \approx \int_{-c}^c f^*(x) dx$$

$$\Rightarrow c = \dots \Rightarrow -c \leq Z = \frac{X - \mu}{\sigma} \leq c \Rightarrow -c \cdot \sigma + \mu \leq X \leq +c \cdot \sigma + \mu \dots$$

5.5.9 Calcul automatique des domaines de confiance

Pour le lecteur moderne, il est certainement clair qu'aujourd'hui on a des programmes d'ordinateur à disposition pour toutes les méthodes standard. Avec Mathematica (par exemple V.4.0) on calcule des domaines de confiance comme il suit:

Exemple:

Intervalles de confiance pour la moyenne et la variance, niveau de confiance 0.9. Lois utilisées comme base: Loi de Student pour la moyenne (mean) et loi de χ^2 pour la variance (var).

Programme en Mathematica:

```
<< 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]}
```

Output:

```
{"mean->", {1.25473, 2.52527}, "var->", {0.638868, 3.25072}}
```

5.6 Tests de signification

5.6.1 Hypothèses

Idée:

Tester un paramètre signifie de postuler d'abord des **hypothèses** concernant la position du paramètre en question et ensuite d'étudier la probabilité de l'hypothèse. Normalement ce processus mène à un refus ou dans l'autre cas à la tolérance resp. à accepter l'hypothèse sur la base d'une probabilité demandée. Nous voulons étudier au moyen d'exemples caractéristiques, comment ça fonctionne exactement.

Méthode:

1.

Nous commençons par **l'hypothèse nulle** H_0 : P.ex. nous postulons qu'un paramètre z respecte une valeur de consigne z_0 : $H_0 = (z = z_0)$. P.ex. $H_0 = (\mu = \mu_0)$

Autres exemples:

$$H_0 = (z \neq z_0) \Leftrightarrow H_0 = (z \leq z_0), \quad H_0 = (z \neq z_0) \Leftrightarrow H_0 = (z \geq z_0)$$

2.

Normalement on oppose une **hypothèse alternative** H_1 avec une **petite probabilité** α à **l'hypothèse nulle** avec une **grande probabilité**, p.ex. ici $H_1 = (z \neq z_0)$. Maintenant on peut étudier la probabilité de la différence $z - z_0$ par le moyen d'une répartition de test convenable. (Avec z_0 , l'hypothèse nulle entre ici comme condition aussi dans la procédure.) Normalement on utilise une estimation convenable $\zeta(z)$ pour z et on a une taille d'échantillon donnée. Si sous la condition d'une probabilité α petite qu'on a utilisée pour calculer, la valeur qui est actuellement présente favorise quand-même l'hypothèse alternative, on a à faire à un événement rare. Une différence aléatoire est donc improbable. A l'hypothèse alternative il faut donc attribuer une probabilité qui est plus grande que celle (α) qu'on vient d'utiliser. Par conséquent on repousse l'hypothèse nulle. Dans l'autre cas, on ne peut pas attaquer la petite probabilité de l'hypothèse alternative. Par conséquent on ne peut pas repousser l'hypothèse nulle. On doit la tolérer ainsi resp. on l'accepte ("au niveau α ").

5.6.2 Alternative bilatérale, test de Student

Exemple: Donnée:

$$X \in N(\mu, \sigma^2), \quad H_0 = (\mu = \mu_0) \rightsquigarrow H_1 = (\mu \neq \mu_0), \quad \zeta(\mu) = \bar{x} = \bar{x}_n$$

μ_0 est une valeur qu'il faut respecter selon l'accord et qui est donnée par les plans.

Soit $\bar{x} = 127.3$, $\mu_0 = 128$, $s = 1.2$, $n = 20$, $\alpha = 0.01$

On peut utiliser comme **distribution de test** la fonction de l'échantillon "différence normée" ou l'écart de μ_0 . Ici on insère comme base aussi l'hypothèse nulle en forme du terme $\bar{X} - \mu_0$. μ_0 est appliqué à zéro lors de la normalisation.

$$\rightsquigarrow T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$$

Important:

$\rightsquigarrow T$ satisfait une répartition t avec $n - 1$ degrés de liberté.

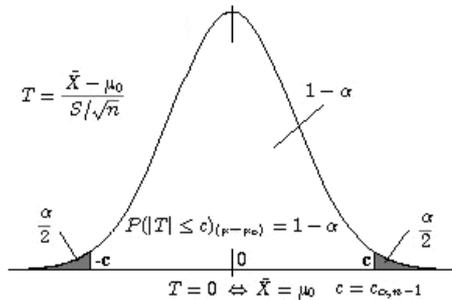
$$T \rightsquigarrow f(T) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \pi \Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{T^2}{n})^{\frac{n+1}{2}}}, \quad F(c) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \pi \Gamma(\frac{n}{2})} \cdot \int_{-\infty}^c \frac{1}{(1 + \frac{u^2}{n})^{\frac{n+1}{2}}} du$$

↪ Etudier: $P(|T| > c) = \alpha$

La probabilité que sous la condition $\mu = \mu_0$ la différence normalisée soit plus grande que c , est ici égale à une probabilité donnée α . α est donc un nombre qu'il faut fixer de façon raisonnable.

Définition:

α s'appelle **niveau de signification**.



Il vaut:

$$\begin{aligned} P(|T| > c) = \alpha &\Leftrightarrow P(|T| \leq c) = 1 - \alpha, \\ P(|T| \leq c) &= P(-c \leq T \leq c) \\ &= F(c) - F(-c) = F(c) - (1 - F(c)) = 2F(c) - 1, \\ 1 - \alpha &= 2F(c) - 1 \Rightarrow 2 - 2F(c) = \alpha \\ 1 - F(c) &= \frac{\alpha}{2} \rightsquigarrow c = \dots \rightsquigarrow \\ P(|T| > c) = \alpha &\Rightarrow \\ (|T_{real}| > c)_{P=\alpha} &\Leftrightarrow (\neg(|T_{real}| \leq c))_{P=\alpha} \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow \neg((-c \leq T_{real} = \sqrt{n} \frac{\bar{X} - \mu_0}{S} \leq c))_{P=\alpha} \Leftrightarrow \neg(-\frac{c \cdot s}{\sqrt{n}} + \mu_0 \leq \bar{X} \leq \frac{c \cdot s}{\sqrt{n}} + \mu_0)_{P=\alpha} \\ &\Leftrightarrow ((\bar{X} < -\frac{c \cdot s}{\sqrt{n}} + \mu_0) \vee (\frac{c \cdot s}{\sqrt{n}} + \mu_0 < \bar{X}))_{P=\alpha} \end{aligned}$$

Maintenant nous calculons c directement avec Mathematica. Ceux qui n'ont pas les moyens pour un tel calcul, doivent avoir recours à des tableaux!

Programme en Mathematica:

```
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}|
```

$$\begin{aligned} &((\bar{X} < -\frac{c \cdot s}{\sqrt{n}} + \mu_0) \vee (\frac{c \cdot s}{\sqrt{n}} + \mu_0 < \bar{X}))_{P=\alpha} \\ \rightsquigarrow &((127.3 < -\frac{2.84534 \cdot 1.2}{\sqrt{20}} + 128) \vee (\frac{2.84534 \cdot 1.2}{\sqrt{20}} + 128 < 127.3))_{P=0.01} \\ \Leftrightarrow &((127.3 < 127.237) \vee (128.763 < 127.3))_{P=0.01} \Leftrightarrow (127.3 \notin [127.237, 128.763])_{P=0.01} \end{aligned}$$

Résultat: Ainsi avec une probabilité de $\alpha = 0.01$, 127.3 ne devrait pas être trouvé dans l'intervalle [127.237, 128.763]. Mais 127.3 se trouve justement dans cet intervalle. Ça signifie que c'est comme ça parce que l'hypothèse nulle $H_0 = (\mu = \mu_0)$ est correcte ou parce que c'est comme ça par hasard. **Ici on ne peut donc pas refuser l'hypothèse nulle sur la base de ce test et le niveau de signification $\alpha = 0.01$.** Cependant si 127.3 était à l'extérieur de l'intervalle, la probabilité serait 0.01. Sur la base de notre philosophie de test il faudrait donc refuser l'hypothèse nulle.

Attention:

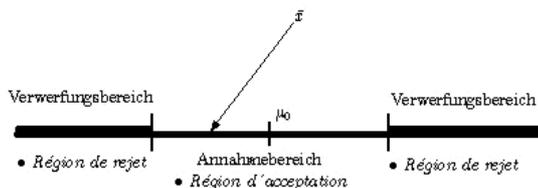
Petite probabilité ne signifie pas impossibilité

Remarque:

La méthode de test que nous venons d'utiliser ici s'appelle **test de Student** parce que nous avons utilisé la loi de Student comme fonction de distribution.

région d'acceptation, domaine de rejet

Dans l'exemple qu'on vient de discuter, l'hypothèse alternative a été repoussée parce que la valeur calculée de l'estimation se trouvait dans l'intervalle trouvé et pas à l'extérieur de l'intervalle. Nous parlons ici du **domaine d'acceptation** et des **domaines de refus** (voir esquisse).



Attention:

Ce n'est pas la règle que les domaines de rejet soient symétriques. Ils peuvent se trouver aussi seulement d'un côté.

5.6.3 Alternative unilatérale, test de Student

Nous considérons ici le cas $H_0 = (z = z_0)$ contra $H_1 = (z > z_0) \rightsquigarrow$ **test alternatif unilatéral.**

Exemple:

Quelqu'un a jeté au hasard une monnaie $N(U) = 10'000$ fois. Il a réalisé $N(E) = 5094$ fois la face resp. l'armoirie. Comme il vaut $h(E) = \frac{N(E)}{N(U)} = 0.5094$, on doute de l'hypothèse nulle $H_0 = (p_E = 0.5)$. Par conséquent il faut donc tester H_0 contre l'alternative unilatérale et supposée $H_1 = (p_E > 0.5)$ au niveau de signification $\alpha = 0.05$.

Soit $X =$ nombre de faces. La valeur de consigne pour $p = 0.5$ est $X = 5'000$ (la distribution binomiale s'approche de la distribution normale). L'échantillon est distribué d'après la loi binomiale $Bi(10'000; 0.5) \doteq N(\mu, \sigma^2)$. Donc il vaut:

$$\mu = np, \quad \sigma^2 = npq = np(1-p), \quad P(X \leq c) = \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^c e^{-\frac{(u-\mu)^2}{2\sigma^2}} du,$$

$$d = \frac{c - \mu}{\sigma}, \quad Z = \frac{X - \mu}{\sigma} \Rightarrow P(X \leq c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{u^2}{2}} du = P(Z \leq d)$$

$$H_1 = (p_E > 0.5) \Rightarrow P(X > c)_{H_0=(p_E=0.5)} = \alpha \Rightarrow P(X \leq c) = P(Z \leq d) = 1 - \alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{u^2}{2}} du$$

Avec des nombres sur la base de l'hypothèse nulle $p = 0.5$:

$$\mu = np = 10'000 \cdot 0.5 = 5'000, \quad \sigma^2 = npq = 10'000 \cdot 0.5 \cdot (1 - 0.5) = 10'000 \cdot 0.25 = 2'500, \quad \sigma = 50$$

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 5'000}{50} = 0.1 X - 500, \quad X = Z \cdot 50 + 5'000 = 50 Z + 5'000$$

$$P(X \leq c) = P(Z \leq d) = 1 - \alpha = 0.95 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{u^2}{2}} du$$

Programme en Mathematica:

```
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}
```

Par conséquent nous trouvons comme le résultat la barrière à $c = 5082.24$ sur la base $\alpha = 0.05$. La probabilité de l'hypothèse alternative $H_1 = (p_E > 0.5)$ est $P(X > c)_{H_0=(p_E=0.5)} = P(X > 5082.24) = 0.05$. Mais par contre à cela nous avons constaté: $H(E) = N(E) = 5094$ avec $P(X > 5082.24) = 0.05$. Avec $H(E) = 5094$ il s'est donc réalisé un événement classifié comme rare avec la probabilité < 0.05 . Par conséquent on ne peut pas considérer l'hypothèse alternative comme accidentelle et on l'accepte donc sur la base de $\alpha = 0.05$. L'hypothèse nulle est donc repoussée. Par conséquent sur cette base la monnaie n'est pas considérée comme homogène. Comme on voit, le domaine de rejet ou **domaine critique** est ici unilatéral. ($\{X \mid X > 5082.24\}$).

Calcul direct

Une autre méthode plus courte consiste à calculer directement $p_1 = P(X > 5094)$ et de comparer ce nombre avec la probabilité significative $\alpha = 0.05 = P(X > c)_{H_0=(p_E=0.5)}$ de l'hypothèse alternative $H_1 = (p_E > 0.5)$. Si on trouve $p_1 < \alpha = 0.05$, l'hypothèse alternative n'est pas aléatoire et il faut donc rejeter l'hypothèse nulle.

Programme en Mathematica:

```

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 |

```

Conséquence: $p_1 = 0.030054 < \alpha = 0.5 \rightsquigarrow$ refuser H_0

5.6.4 Risques (aléas) aux tests

En raison d'un test, il peut être possible qu'on refuse accidentellement et faussement une hypothèse nulle correcte ou qu'on ne refuse pas accidentellement et faussement une fausse hypothèse nulle. Nous avons déjà rencontré une chose pareille aux intervalles de confiance. Par conséquent nous reprenons les notations:

Définition:

Si nous refusons par hasard une hypothèse nulle vraie en raison d'un test, nous commettons une **erreur de première sorte**.

Nous faisons une erreur de 1ère sorte si nous acceptons l'hypothèse alternative à la place de l'hypothèse nulle. L'hypothèse alternative est réalisée par la probabilité α (nombre significatif).

Conséquence:

La probabilité d'une erreur de 1ère sorte est exactement le nombre significatif α .

Définition:

Si par hasard nous ne refusons pas une hypothèse nulle fausse en raison d'un test, nous commettons une **erreur de deuxième sorte**.

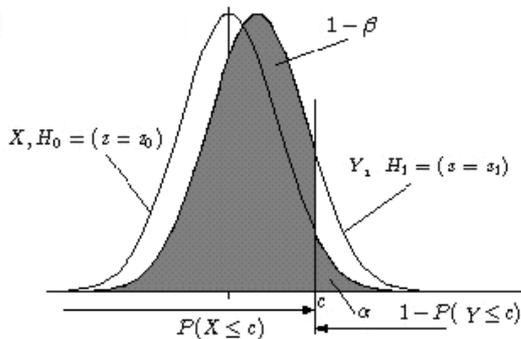
Si nous commettons une erreur de la deuxième sorte, à la place de l'hypothèse nulle, p.ex. $H_0 = (z = z_0)$, une hypothèse alternative $H_1 = (z = z_1)$ serait correcte. En raison de H_0 on a construit une fonction de test qui est fausse. On devrait construire la fonction de test correcte en raison de H_1 . β soit la probabilité pour la réalisation de l'hypothèse alternative liée à H_1 par rapport à c qui était calculé à la base de H_0 . Par conséquent β est égal au nombre significatif d'une nouvelle hypothèse alternative par rapport à la nouvelle hypothèse nulle H_1 . Par conséquent $1 - \beta$ est la probabilité de ne pas refuser la nouvelle hypothèse nulle H_1 et par conséquent de refuser la vieille hypothèse nulle H_0 , c.-à.-d. d'éviter une erreur de la 2ème sorte.

Définition:

La probabilité β d'éviter une erreur de la 2ème sorte s'appelle le **pouvoir du test**.

La probabilité $1 - \beta$ de ne pas éviter une erreur la 2ème sorte s'appelle le **risque de la 2ème sorte** ou le **risque du producteur**.

Le nombre significatif ou la probabilité d'une erreur de la 1ère sorte s'appelle aussi le **risque de la 1ère sorte** ou le **risque du consommateur**.



$$\alpha = P(c < X)_{H_0}, \quad \beta = 1 - P(\tilde{X} \leq c)_{H_1}, \\ H_1 = (u = u_0) \Rightarrow \beta = \beta(u)$$

Remarque:

Un c plus grand signifie en effet un α plus petit, cependant ça signifie aussi un $1 - \beta$ plus grand. Avec un n plus grand on obtient normalement des courbes plus fines. α étant donné, $1 - \beta$ devient plus petit. Le **degré de séparation** du test augmente.

5.6.5 Test khi deux pour la variance**Exemple:**

Soit $X_i \in N(\mu, \sigma^2)$, $\sigma_0 = 50$, $n = 20$, $s^2 = 61.3$, $H_0 = (\sigma^2 = \sigma_0^2)$, $H_1 = (\sigma^2 > \sigma_0^2)$

Remarque:

$\sigma^2 > \sigma_0^2$ est la seule alternative intéressante.

$$\frac{S^2}{\sigma_0^2} = \frac{1}{n-1} \sum_{i_1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} = \frac{1}{n-1} \sum_{i_1}^n Z_i^2, \quad Y_i \in N(0, 1) \Rightarrow T = \frac{S^2}{\sigma_0^2} (n-1)$$

$T = \frac{S^2}{\sigma_0^2} (n-1)$ est réparti d'après la loi χ_{n-1}^2 .

Soit $\alpha = 0.05 \rightsquigarrow P(T > c)_{\sigma^2=50} = \alpha = 0.05 \Rightarrow P(T \leq c)_{\sigma^2=50} = 1 - \alpha = 0.95$

Programme en Mathematica:

```
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}} |
```

L'hypothèse alternative signifie qu'une valeur quelconque soit située à l'extérieur de l'intervalle $\{0, 79.325\}$ avec une probabilité $\alpha = 0.05$. $s^2 = 61.3$ cependant est situé à l'intérieur de l'intervalle. Il n'existe donc pas de raison d'accepter l'hypothèse alternative et de repousser l'hypothèse nulle.

5.6.6 Tester automatiquement**Exemple:**

Avec Mathematica, citation de la littérature officielle: (Il faut considérer la base des lois normales et de Student!)

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 a 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` \rightarrow `True`. For more detailed information about a test use `FullReport` \rightarrow `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` \rightarrow `siglev`, which yields a conclusion of the test, stating acceptance or rejection of the hypothesis.

Programme en Mathematica:

```
<< 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}
```

Programme en Mathematica:

```
MeanTest[data1, 38, KnownVariance -> 10, SignificanceLevel -> 0.1]
```

Output:

```
{OneSidedPValue -> 0.211855,
 "Fail to reject null hypothesis at significance level" -> 0.1}
```

Programme en Mathematica:

```
MeanTest[data1, 38, KnownVariance -> 10]
```

Output:

```
{OneSidedPValue -> 0.211855,
 OneSidedPValue -> 0.211855}
```

Programme en Mathematica:

```
MeanTest[data1, 38]
```

Output:

```
OneSidedPValue -> 0.286061
```

Programme en Mathematica:

```
MeanTest[data1, 38, KnownVariance -> 10, FullReport -> True] // TeXForm
```

Output:

```
\{ {\Mvariable{FullReport}}\rightarrow
  {\matrix{ 38.8 & 0.8 & \Muserfunction{NormalDistribu
            tion}() \cr } },
 {\Muserfunction{OneSidedPValue}}\rightarrow
  {0.211855}}\}
```

↪

```
{FullReport → Mean      TestStat  Distribution
                38.8     0.8        NormalDistribution[], OneSidedPValue → 0.211855}
```

5.6.7 Vue d'ensemble d'une recette de test

1. Résumer le problème en question par une hypothèse \rightsquigarrow hypothèse nulle p.ex. $H_0 = (a = a_0)$. Ici on a donné une valeur a_0 et une estimation $a = \tilde{a}_0$. A la valeur a il doit exister une fonction de répartition X avec spécialement $X = a$
2. Souvent il n'est pas directement possible de donner un énoncé (assertion) de probabilité concernant $H_0 = (a = a_0)$. Par rapport à H_0 on formule donc une hypothèse alternative H_1 dont il est possible de trouver une assertion de probabilité à l'aide d'une fonction de répartition et d'un intervalle convenable. Un exemple est l'hypothèse alternative unilatérale $H_1 = (a > a_0)$.
3. Transformer X de façon bijective et monotone en une autre variable aléatoire $Z = g(X)$ (répartition de tet!), dont la fonction de répartition est directement calculable ou dont on a donné les valeurs dans un tableau. $\rightsquigarrow X = g^{-1}(Z)$. Si g est bijective et continue, g et g^{-1} sont strictement monotones. Des inégalités par rapport à Z se transforment en inégalités par rapport à X .
4. Par rapport à $H_1 = (a > a_0) = H_1 = (a_0 < a)$ et en considération de la répartition de X , la proportion " $a_0 < a \leq a_1$ " ou " $a_0 < X \leq a_1$ " a une probabilité grande si a_1 est une valeur convenable que la proportion " $a_1 < a$ " ou " $a_1 < X$ " soit improbable. A cause de la monotonie, ces proportions obtiennent après la transformation p.ex. la forme " $g(a_0) < g(X) = Z \leq g(a_1)$ " resp. p.ex. la forme " $g(a_1) = c < g(X) = Z$ ".
5. Soit donné une probabilité proposée (nombre significatif α); on peut résoudre l'inéquation $P(c < Z) = \alpha$ resp. $P(Z \leq c) = 1 - \alpha$ d'après c à l'aide d'ordinateurs ou de tableaux. $\rightsquigarrow c$
6. $c < Z$ correspond p.ex. à $a_1 = g^{-1}(c) < X = g^{-1}(Z)$. S'il vaut maintenant pour l'estimation $\tilde{a}_0 \in (c, \infty)$, un a un événement rare pour l'hypothèse alternative dans un intervalle où la probabilité α est petite. La réalité "tombe dans un intervalle improbable"! On ne peut donc pas rejeter l'hypothèse alternative. C'est pourquoi on rejette ici l'hypothèse nulle.

5.6.8 Test de compar. de moyennes I

On voit toute suite qu'on peut distinguer deux échantillons divers: Premièrement ceux qui consistent en individus différents et deuxièmement ceux qui consistent en valeurs obtenues de façon différentes, mais obtenues aux mêmes individus. Nous considérons ici le deuxième cas.

Exemple: Donné:

Deux échantillons: On a mesuré à deux endroits différents A , B à la même pièce des données par la méthode V_A et par la méthode V_B :

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

Nous avons ici à faire à deux échantillons conjoints qui contiennent des valeurs en paires (x_i, y_i) et les variables affiliées (X_i, Y_i) . Nous voulons présupposer que tous les X_i resp. Y_i soient indépendants de i et aient des variances σ_X^2 resp. σ_Y^2 égales.

A ces conditions, nous voulons tester les hypothèses $H_0 = (\mu_{X_i} = \mu_{Y_i})$ contre les alternatives unilatérales $H_1 = (\mu_{X_i} > \mu_{Y_i})$.

Nous savons que les variables sont distribuées selon la loi de Gauss et que par conséquent les différences $d_i = x_i - y_i$ ont une variable D qui est distribuée selon la loi de Gauss avec $\mu_D = \mu_X - \mu_Y$ et $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2$. Maintenant nous pouvons tester à l'aide du test de Student l'hypothèse nulle $\mu_D = 0$ contre l'alternative $\mu_D > 0$.

$$\leadsto \bar{d} = \frac{1}{n} \sum_{i_1}^n d_i = 0.0416667, \quad s_D = \sqrt{\frac{1}{n-1} \sum_{i_1}^n (d_i - \bar{d})^2} = 0.0470815, \quad \mu_0 = 0, \quad \alpha = 0.05, \quad n = 6$$

$$Z = g(X) = T = \frac{\bar{D} - \mu_0}{S/\sqrt{n}} = \frac{X - \mu_0}{S/\sqrt{n}} \leadsto X = \bar{D} = g^{-1}(Z) = \mu_0 + \frac{T S}{\sqrt{n}} = \mu_0 + \frac{Z S}{\sqrt{n}}$$

$$P(c < T = Z) = \alpha = 0.05 \Rightarrow P(T \leq 0) = 1 - \alpha \leadsto c$$

$$c < Z \Leftrightarrow g^{-1}(c) < g^{-1}(Z) \Leftrightarrow \mu_0 + \frac{c S}{\sqrt{n}} = g^{-1}(c) < g^{-1}(Z) = X$$

$$\Leftrightarrow 0 + c \frac{0.0470815}{\sqrt{6}} = c \cdot 0.0192209 = g^{-1}(c) < g^{-1}(Z) = X$$

\leadsto Question: $c \cdot 0.0192209 \stackrel{?}{<} X = \bar{d} = 0.0416667 \leadsto$

Programme en Mathematica:

```
alpha = 0.05;
f[z_, m_] :=
  Gamma[(m + 1)/2]/(Sqrt[m Pi] Gamma[m/2]) / (1 + z^2/m)^(m/2);
FindRoot[
  1 - alpha == Evaluate[Integrate[f[t, 5], {t, -Infinity, c}]], {c, 3}]
// Chop
```

Output:

```
{c -> 2.01505}
```

$$\leadsto c \cdot 0.0192209 = 2.01505 \cdot 0.0192209 = 0.0387311 < 0.0416667 = \bar{d}$$

~>

Événement rare, ne pas rejeter l'hypothèse alternative, rejeter l'hypothèse nulle.

5.6.9 Test de compar. de moyennes II

Nous considérons le cas de deux échantillons indépendants $\{x_1, \dots, x_n\}$ pour $X \in N(\mu_X, \sigma_X^2)$ et $\{y_1, \dots, y_m\}$ pour $Y \in N(\mu_Y, \sigma_Y^2)$ avec en générale $n = n_X \neq m = n_Y$ et $\sigma_X = \sigma_Y$. Soient données les estimations \bar{x} pour μ_X et correspondamment \bar{y} , s_X , s_Y . Il faut tester l'hypothèse nulle $H_0 = (\mu_X = \mu_Y)$ par contre à l'alternative $H_1 = (\mu_X > \mu_Y)$. On voit très vite que maintenant à cause de $n_X \neq n_Y$ on ne peut plus construire avec les différences une nouvelle variable. Pour construire quand même une variable de test, on a donc besoin d'un peu plus de théorie. À cause du cadre, nous renonçons ici à une dérivation plus vaste et nous nous contentons du le résultat:

Formule:

A nos conditions, la variable suivante satisfait la loi de Student avec $n_X + n_Y - 2$ degrés de liberté :

$$T_{XY} = \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}}$$

Exemple: $\alpha = 0.05$

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		

Soit $H_0 = (\mu_X = \mu_Y)$, $H_1 = (\mu_X > \mu_Y)$

~> **A l'aide de 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

$$\begin{aligned} &\leadsto P(c < T_{XY}) = \alpha, \quad T = T_{XY} = \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}} \\ &\leadsto (c < T_{XY})_{\alpha=0.05} \Leftrightarrow |_{\alpha=0.05} : \\ &c \cdot \sqrt{\frac{((n_X - 1) S_X^2 + (n_Y - 1) S_Y^2) \cdot (n_X + n_Y)}{n_X n_Y (n_X + n_Y - 2)}} < (\bar{x} - \bar{y}) \cdot \sqrt{\frac{((n_X - 1) S_X^2 + (n_Y - 1) S_Y^2) \cdot (n_X + n_Y)}{n_X n_Y (n_X + n_Y - 2)}} \\ &\quad \leadsto 1.77325 < 0.12614 \end{aligned}$$

La dernière proposition n'est pas vraie, c.-à-d. $(c < T_{XY})$ n'est pas satisfait ici. On ne peut donc pas accepter l'hypothèse alternative et par conséquent on ne peut pas refuser l'hypothèse nulle.

5.7 Autres méthodes de test

5.7.1 Une classification des tests

En principe on peut inventer des critères quelconques pour classifier les tests. Pour nous les deux typologies suivantes nous rendent service:

1. Des **tests de paramètres**, p.ex. il faut tester $H_0 = (\mu_X = \mu_Y)$.
 2. Des **tests libres de paramètres**, p.ex. les **tests de rang** ou le **test des signes** (signum).
-

1. Les **tests qui dépendent d'une loi de répartition**, p.ex. des tests de paramètres.
 2. Les **tests qui ne dépendent pas d'une loi de répartition**, p.ex. des tests de rang. Ou en outre des **tests de la qualité de la loi de distribution** où on examine si une répartition est compatible avec les valeurs empiriques.
-

5.7.2 Test libre de paramètres: Le test du signe

Exemple: Donné:

Quantités de vente quotidiennes de deux équipes de travail comparables *A* et *B* avec des tâches identiques pendant 10 jours. On a noté seulement la différence $W_A - W_B$ en unités de calcul. On teste l'hypothèse nulle où les deux équipes ont les mêmes capacités contre l'alternative où l'équipe *A* réalise une vente systématiquement plus haute que *B*. Soit $\alpha = 0.05$ (petit): Si de la réalité on obtient comme résultat une probabilité $P < \alpha$, H_1 devient remarquable, improbable ou "significative".

$W_A - W_B$	0.9	2.1	1.1	0.0	1.1	-0.4	1.4	0.6	0.5	0.1
<i>sgnum</i>	+	+	+	.	+	-	+	+	+	+

Il vaut: $H_0 = (W_A = W_B)$ vrai $\rightsquigarrow P(\text{sgn}(W_A - W_B) = '' +'') = P(\text{sgn}(W_B - W_A) = '' +'')$

La valeur 0.0 peut être tracée, car la décision ne peut pas changer. Ainsi nous avons 8 fois " + " parmi 9 valeurs.

Idée:

Soit $X =$ nombre de " + " entre les n valeurs. Sous la condition de H_0 X est réparti selon la loi binomiale avec $p = 0.5$ ($P(\text{sgn}(W_A - W_B) = '' +'') = P(\text{sgn}(W_B - W_A) = '' +'')$).

$$\begin{aligned} \rightsquigarrow P(X \geq 8) &= 1 - P(X \leq 7) = 1 - \sum_{k=0}^7 P(X = k) = 1 - \sum_{k=0}^7 \binom{9}{k} p^k q^{9-k} \\ &= 1 - \sum_{k=0}^7 \binom{9}{k} 0.5^k 0.5^{9-k} = 1 - 0.5^9 \sum_{k=0}^7 \binom{9}{k} = 1 - 0.5^9 (2^9 - \binom{9}{8} - \binom{9}{9}) \approx 0.0195313 \\ &\Rightarrow P(X \geq 8) \approx 0.0195313 \leq \alpha = 0.05 \end{aligned}$$

On l'a donc à faire ici à un événement rare avec $X = 8$ qui a été réalisé, bien qu'il vaille $P(X \geq 8) \approx 0.0195313 \leq P(X \geq c) = \alpha = 0.05$. Pour les mêmes réflexions qu'avant, il faut donc accepter une hypothèse alternative et rejeter par conséquent l'hypothèse nulle. Mais si l'on avait $\alpha = 0.01$, on ne pourrait pas accepter d'hypothèse alternative et on ne pourrait donc plus rejeter l'hypothèse nulle.

5.7.3 Test d'adaptation de Khi deux

Problème:

Il faut examiner au moyen d'un test, si une variable donnée suit la loi d'une fonction de distribution $F(t)$ donnée. C'est ce que nous voulons étudier par l'exemple suivant:

Exemple: Donné:

1. Événements disjoints A_1, \dots, A_m qui sont des résultats d'une expérience aléatoire.

$$A_i \cap A_k = \{\}, \quad i \neq k, \quad \bigcup_{i=1}^m = \Omega$$

2. $P(A_i) = p_i$ inconnu $\rightsquigarrow H_0 = (P(A_i) = p_i), \quad i = 1, \dots, n$

3.

Test de H_0 : Prendre un échantillon de la taille n , $n_i = |A_i|$

4. Formule: $\tau = \sum_{i=1}^m \frac{(n p_i - n_i)^2}{n p_i} = \sum_{i=1}^m \frac{(\nu_i - n_i)^2}{\nu_i} = \sum_{i=1}^m \nu_i \left(1 - \frac{n_i}{\nu_i}\right)^2$

on peut démontrer que τ est une approximation de la loi de χ^2 avec $m - 1$ degrés de liberté, si m est assez grand, voir. p.ex. lit. Kreyszig, Bibl. A10 et Bibl. A7 (Cramer).

Trouver la décision: Des réflexions analogiques aux exemples qu'on vient de traiter nous mènent à la décision de refuser H_0 dès que la valeur de τ surmonte $c = \chi_{\alpha, m-1}^2$ ($c =$ borne supérieure de l'intégrale à la valeur α).

Exemple: (vgl. Lit Kreyszig, Bibl. A10.)

Gregor Mendel, expériences avec des pois, hybrides. p_i supposé d'après les "lois de Mendel".
Supposition: $n_A : n_B : n_C : n_D = 9 : 3 : 3 : 1 \rightsquigarrow$ Résultat:

	A	B	C	D	$m = 4$
	rond, jaune	rond, vert	arêts, jaune	arêts, vert	total
n_i	315	108	101	32	556
p_i	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	1
np_i	312.75	104.25	104.25	34.75	556

Test de la supposition avec $\alpha = 0.05$: $n_A : n_B : n_C : n_D = 9 : 3 : 3 : 1$

$$\tau = \frac{(312.75 - 315)^2}{312.75} + \frac{(104.25 - 108)^2}{104.25} + \frac{(104.25 - 101)^2}{104.25} + \frac{(34.75 - 32)^2}{34.75}$$

\rightsquigarrow **A l'aide de Mathematica:**

```
(312.75 - 315)^2/312.75 + (104.25 - 108)^2/104.25 + (104.25 - 101)^2/
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}
```

$\rightsquigarrow \rightsquigarrow$ Décision: $\tau = 0.470024 < c_{\alpha=0.5} = 7.81471 \rightsquigarrow H_0$ n'est pas refusée! Problème: τ est une approximation de la loi de χ^2 , si m est assez grand. Et chez nous on a $m = 4 \dots$

Remarque:

Un événement A_i pourrait aussi être que X tombe dans un certain intervalle $[x_i, x_{i+1})$. Avec les intervalles, la situation continue est aussi représentée. Ainsi on peut **tester si une fonction est distribuée de façon normale**.

5.7.4 Quant au test de Fisher

A la page 125 vous avons vu:

Théorème: **Hyp.:** Distribution \mathcal{F}

Thè.:

$$1. F(x) = P(\mathcal{F} \leq x) = \frac{\Gamma(\frac{m_1 + m_2}{2})}{\Gamma(\frac{m_1}{2})\Gamma(\frac{m_2}{2})} \cdot m_1^{\frac{m_1}{2}} \cdot m_2^{\frac{m_2}{2}} \cdot \int_0^x \frac{t^{\frac{m_1-2}{2}}}{(m_1 t + m_2)^{\frac{m_1+m_2}{2}}} dt$$

$$2. f(x) = \begin{cases} 0 & x \leq 0 \\ f(x) = E_{m_1, m_2} x^{\frac{m_1}{2}-1} (m_2 + m_1 \cdot x)^{-\frac{m_1+m_2}{2}} & x > 0 \end{cases}$$

Maintenant il vaut (preuves voir lit. Kreyszig, Bibl. A10):

Théorème: **Hyp.:**

Variables aléatoires $V_1, V_2 \in \{\heartsuit\}$ ainsi que $V_1 \in \chi_{m_1}^2$, $V_2 \in \chi_{m_2}^2$

Thè.:

$V = \frac{V_1/m_1}{V_2/m_2}$ est réparti selon la loi \mathcal{F}

$$\begin{aligned} x \leq 0 &\Rightarrow F(x) = 0, & 0 < x &\Rightarrow F(x) = P(V \leq x) = \\ &= \frac{\Gamma(\frac{m_1 + m_2}{2})}{\Gamma(\frac{m_1}{2})\Gamma(\frac{m_2}{2})} \cdot m_1^{\frac{m_1}{2}} \cdot m_2^{\frac{m_2}{2}} \cdot \int_0^x \frac{t^{\frac{m_1-2}{2}}}{(m_1 t + m_2)^{\frac{m_1+m_2}{2}}} dt \end{aligned}$$

Théorème: **Hyp.:**

Variables aléatoires

$X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2} \in \{\heartsuit\}$ 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 Estimations de σ_1^2, σ_2^2

Variances empiriques

Thè.:

$V = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ est réparti selon la loi \mathcal{F} avec $(n_1 - 1, n_2 - 1)$ degrés de liberté

Conséquence:

Pour $H_0 = (\sigma_1^2 = \sigma_2^2)$ on obtient $V = \frac{S_1^2}{S_2^2}$. On a donc un moyen pour tester l'égalité des variances!

5.7.5 Tableaux de contingence

A l'aide du test de χ^2 on peut aussi tester l'**indépendance statistique** de deux variables aléatoires X et Y . Nous considérons un exemple de tableaux de contingence. Les clients qui achètent le produit A , B ou C y sont répartis dans des classes d'après leur âge. On y voit les fréquences (valeurs empiriques):

n_{ik}	$Y = A$	$Y = B$	$Y = C$	$n_{i\cdot}$
$X \in [16, 30)$	$n_{11} = 132$	$n_{12} = 778$	$n_{13} = 592$	$n_{1\cdot} = 1502$
$X \in [30, 55)$	$n_{21} = 55$	$n_{22} = 304$	$n_{23} = 248$	$n_{2\cdot} = 607$
$X \in [55, 85)$	$n_{31} = 25$	$n_{32} = 111$	$n_{33} = 155$	$n_{3\cdot} = 291$
$n_{\cdot k}$	$n_{\cdot 1} = 212$	$n_{\cdot 2} = 1193$	$n_{\cdot 3} = 995$	$n = 2400$

$n_{\cdot k}, n_{i\cdot} \rightsquigarrow$ fréquence marginales

Problème: Est-ce que X et Y sont statistiquement indépendants?

Idée:

Nous considérons les valeurs empiriques $n_{ik}, n_{i\cdot}, n_{\cdot k}$ comme réalisations de nombres d'occupation théoriques et attendues $n \cdot p_{ik}, n \cdot p_{i\cdot}, n \cdot p_{\cdot k}$. Nous pouvons maintenant poser la question comme hypothèse nulle. L'indépendance statistique signifie que pour la probabilité il vaut: $p_{ik} = p_{i\cdot} \cdot p_{\cdot k}$.

$\rightsquigarrow H_0 = (p_{ik} = p_{i\cdot} \cdot p_{\cdot k})$

$\frac{n_{ik}}{n}$ est une valeur estimée pour p_{ik} , $\frac{n_{i\cdot}}{n}$ est une valeur estimée pour $p_{i\cdot}$, $\frac{n_{\cdot k}}{n}$ est une valeur estimée pour $p_{\cdot k}$. L'indépendance doit donc signifier:

$\rightsquigarrow H_0 = (p_{ik} = p_{i\cdot} \cdot p_{\cdot k}) \approx H_0 = \left(\frac{n_{ik}}{n} = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot k}}{n}\right) = H_0 = (n_{ik} = \frac{n_{i\cdot} \cdot n_{\cdot k}}{n})$

En outre on peut démontrer:

Théorème:

La variable aléatoire suivante est distribuée approximativement selon la loi $\chi_{(r-1) \cdot (s-1)}^2$:

$$T = \sum_{i=1}^r \sum_{k=1}^s \frac{\left(\frac{n_{i\cdot} \cdot n_{\cdot k}}{n} - n_{ik}\right)^2}{\frac{n_{i\cdot} \cdot n_{\cdot k}}{n}}$$

Au moyen des fréquences aux limites nous calculons les valeurs d'occupation estimées attendues:

$$H_0 \rightsquigarrow n \cdot p_{ik} \approx \frac{n_{i\cdot} \cdot n_{\cdot k}}{n} = u_{ik}$$

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

Il vaut: $T = \sum_{i=1}^r \sum_{k=1}^s \frac{\left(\frac{n_{i\cdot} \cdot n_{\cdot k}}{n} - n_{ik}\right)^2}{\frac{n_{i\cdot} \cdot n_{\cdot k}}{n}} = \sum_{i=1}^r \sum_{k=1}^s \frac{(u_{ik} - n_{ik})^2}{u_{ik}} \Rightarrow T = t = \dots$

Soit $\alpha = 0.01$

→ **A l'aide de 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/matrU // 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}

→ Décision: $t = 20.5737 \stackrel{?}{<} c_{\alpha=0.1} = 13.2766$

→

H_0 refusé! Problème: T est une approximation de la loi de χ^2 , si r, s sont assez grands. Et chez nous on a $r = s = 4 \dots$

Kapitel 6

Fehlerrechnung, Regression, Korrelation — Calcul de l'erreur, régression, corrélation

Extrait de: Wirz, Bibl. A15, Script ◊ Math ◊ Ing, ◊ Analysis ◊ Analyse ◊, kurz & bündig. Avec supplément

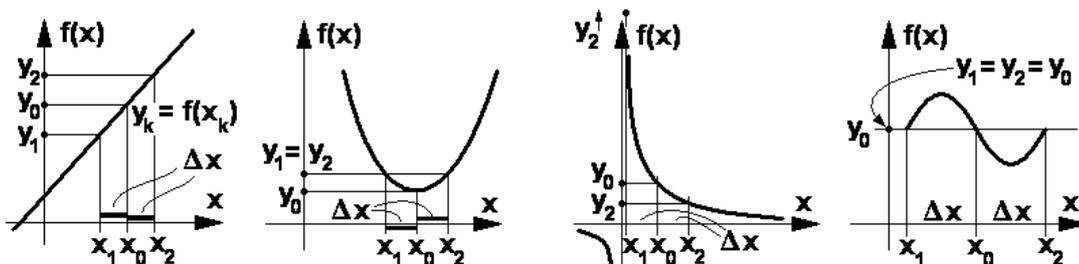
6.1 Fehlerrechnung (Abhängigkeit) — Calcul de l'erreur (dépendance)

Sous la notion de calcul d'erreur, deux types de problèmes différents sont traités dans la littérature: Premièrement le problème de la transposition d'erreurs de mesure à l'application de fonctions lors de grandeurs mesurées inexactes. Et deuxièmement le problème des erreurs de grandeurs qui marquent des données statistiques (p.ex. la moyenne) qui dépend d'ordinaire d'une grande quantité de données. Ici, nous voulons traiter le premier type d'erreur. Le deuxième est traité dans l'appendice (édition séparée) à ce script.

6.1.1 Das Problem der Verpflanzung — Le problème de la dépendance

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

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



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

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 Mess-

werten gibt es immer Ableseungenauigkeiten, die aber abschätzbar sind. Diese zugehörigen „Messfehler“ betragen $\Delta x_1, \Delta x_2, \dots, \Delta x_n$. Die „exakten Werte“ x_k^* , $k = 1, \dots, n$ liegen daher in den Intervallen $[x_{k_0} - \Delta x_k, x_{k_0} + \Delta x_k]$. Zudem sei eine Funktion $f(x_1, x_2, \dots, x_n)$ gegeben, mit deren Hilfe eine weitere Grösse berechnet werden muss. On mesure les valeurs $x_{1_0}, x_{2_0}, \dots, x_{n_0}$ des variables x_1, x_2, \dots, x_n . Pour les valeurs de mesures non-discrètes il y a toujours des inexactitudes quand on lit les échelles, mais elles sont estimables. Les erreurs de mesure qui en résultent sont $\Delta x_1, \Delta x_2, \dots, \Delta x_n$. Les „valeurs exactes“ x_k^* , $k = 1, \dots, n$ sont donc situées dans les intervalles $[x_{k_0} - \Delta x_k, x_{k_0} + \Delta x_k]$. En plus on a donné une fonction $f(x_1, x_2, \dots, x_n)$ à l'aide de laquelle on doit calculer une valeur en plus.

Problème: In welchem Intervall liegt der „wahre“ Wert $f(x_1^*, x_2^*, \dots, x_n^*)$? Dans quel intervalle la valeur „exacte“ $f(x_1^*, x_2^*, \dots, x_n^*)$ est-elle située ?

6.1.2 Verwendung des totalen Differentials — Appliquer la différentielle totale

Soit $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\vec{x}_0 = \begin{pmatrix} x_{1_0} \\ \vdots \\ x_{n_0} \end{pmatrix}$, $f(\vec{x}) := f(x_1, x_2, \dots, x_n)$, $D_k \geq |\Delta x_k|$ (D_k ist eine bezifferbare Schranke. D_k est une borne connue.)

Aus der Theorie des **totalen Differentials** weiss man: De la théorie de la **différentielle totale** on sait:

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

(O: Glieder höherer Ordnung Termes d'ordre supérieur)

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

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

Théorème:

Hyp.:

Messsituation wie oben beschrieben Situation de mesure comme décrit en haut,
 $f \in \mathcal{D}^{(1)}$

Thè.:

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

Conséquence:

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

Définition:

$|\Delta f|$ heisst **absoluter Fehler** s'appelle **erreur absolue**,
 $\left| \frac{\Delta f}{f(\vec{x}_0)} \right|$ heisst **relativer Fehler** s'appelle **erreur relative**.

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

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

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

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

Exemple 5:

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

Remarque:

Diese Beispiele zeigen, dass die oft geäußerte Meinung, es genüge mit den extremen Werten zu rechnen, wohl äusserst falsch sein muss. Ces exemples démontrent que l'opinion souvent communiquée, qu'il suffit de calculer avec les valeurs extrêmes, doit être complètement fausse.

Exemple 6:

Messwerte: Valeurs de mesure:

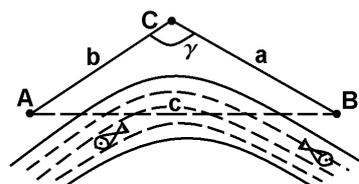
$$a = 364.76 \pm 0.05m$$

$$b = 402.35 \pm 0.05m$$

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

$$\leadsto \gamma \approx 1.1909 \pm 0.002$$

$$c = ?$$



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

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

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

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

6.2 Regression — Régression

6.2.1 Der Begriff — La notion

Donné:

Stichprobe von Beobachtungen Epreuve au hasard d'observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

x_k z.B. Maschineneinstellung p.ex. réglage (positionnement) de la machine,

$y_k = f(x_k)$ Messung measurement.

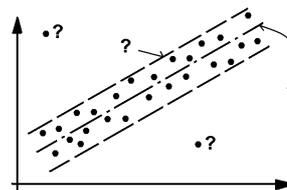
Problème: Interpretation des Verhaltens der Messwerte (Gesetz)?

Intérpretation du comportement des mesures (lois)?

Z.B. Vermutung: Die Messwerte liegen auf einer passenden Geraden oder sonstigen Kurve.
 P.ex. supposition: Les points mesures sont situés sur une droite qui convient ou sur une autre courbe.

Problème:

Wie findet man die „beste“ Gerade (resp. Kurve)? Comment trouver la „meilleure“ droite (resp. courbe)?



~> **Problèmes:**

1. Berechnung der Parameter der hypothetischen Kurve. Calculer les paramètres de la courbe hypothétique.
2. Beurteilung der Zuverlässigkeit der getroffenen Wahl. Juger l'authenticité du choix qu'on a fait.

Die so gefundene Kurve trägt den etwas unverständlichen Namen **Regressionskurve**.
 La courbe ainsi trouvée porte le nom un peu incompréhensible de **courbe de régression**.

Wieso dieser Name? Der Name stammt aus einer Beschreibung von Beobachtungen von F. Galton²¹ zum Größenwachstum von Menschen. Galton hat festgestellt: Pourquoi ce nom? Le nom vient d'une description d'observations de F. Galton²¹ quant à la croissance de l'être humain. Galton a observé:

- 1 Grössere Väter haben grössere Söhne. Les pères plus grands ont des fils plus grands.
- 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** ~> **Regress** vorhanden. Mais par contre on observe la tendance que les pères plus grands ont des fils grands mais qui sont en moyenne plus petits que les pères dont on parle. Il s'agit donc d'un **régression**.

Soit x = Grösse der Väter grandeur des pères,
 y = Grösse der Söhne grandeur des fils.
 $y(x) = ax + b$ Gerade droite
 ~> „**Regressionsgerade**“ „**droite de régression**“

Der Begriff *Regressionsgerade* ist also historisch verankert, hat aber inhaltlich nichts mit der Mathematik zu tun. La notion de droite de régression est donc fondée par l'histoire. La signification de cette notion n'a rien à faire avec les mathématiques.

6.2.2 Methode der kleinsten Quadrate — Méthode des carrés minimaux

Das Problem — Le problème

Problème: Welche Kurve ist die „beste“ Kurve? – Wie ist das Kriterium „beste“ defintorisch am vernünftigsten zu fassen? Quelle est la „meilleure“ des courbes? – Comment est-ce qu'il faut définir le critère „meilleur“ dans ce cadre de façon raisonnable?

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.) Comme nous, êtres humains, devons décider nous-mêmes ce qui est notre „raison“, nous pouvons remplacer ici „raisonnable“ par „économique“, donc par „pragmatique“ et „simple“. C'est pourquoi nous choisissons ici un chemin pragmatique. (Il est de coutume dans la littérature de présenter une explication un peu plus étoffée, basée sur le principe de „maximum-likelihood“.)

²¹Engl. Naturforscher Naturaliste anglais, 1822 – 1911

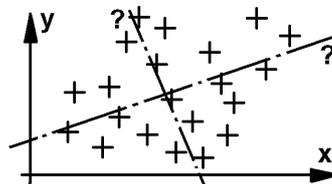
Important: Dass die gesuchte Kurve eine Gerade oder sonst irgend eine Kurve ist, lässt sich theoretisch nicht exakt entscheiden. Man kann nur zu Aussagen kommen wie „die eine Kurve ist wahrscheinlicher als die andere“. Denn dass die gefundenen Messwerte überhaupt auf einer stetigen Kurve liegen, beruht auf einer Arbeitshypothese, die wir nach Descartes mit dem Argument der Erfahrung und der Arbeitsökonomie begründen. On ne peut pas décider de façon théorique que la courbe cherchée est une droite ou n'importe quelle autre courbe. On peut seulement affirmer que "l'une des courbes est plus propable que l'autre". C'est une hypothèse que les valeurs mesurées sont situées sur une courbe continue, hypothèse que nous faisons d'après Descartes, nous basant sur l'expérience et l'économie du travail.

Die Begründung der Methode — La déduction de la méthode

Problème: Durch eine „Punktwolke“ von Messwerten soll eine „beste“ Gerade gelegt werden. Wie ist vorzugehen? Il faut tracer la "meilleure" droite à travers un "nuage de points mesures". Comment faut-il procéder?

Idee 1: Idée 1:

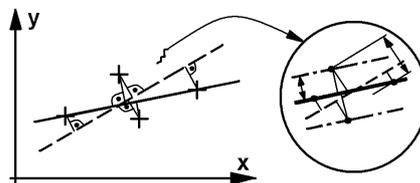
Versuche g so festzulegen, dass die Summe der Abstände zu g gleich 0 wird. Essaie de déterminer g de façon que la somme des distances à g devienne 0.



Problème: Diese Methode funktioniert nicht, da offensichtlich (Skizze) mehrere passende „beste“ Geraden möglich sind. Cette méthode ne fonctionne pas car visiblement (esquisse) plusieurs droites du type "meilleure" sont possibles.

Idee 2: Idée 2:

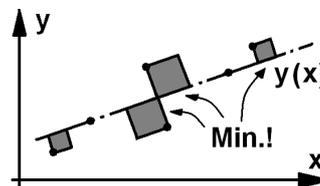
Lege die Gerade so, dass die Summe der Beträge der Abstände minimal wird. Choisis la droite de façon que la somme des valeurs absolues des distances à g soit minimale.



Problème: Diese Methode funktioniert praktisch schlecht, da beim Berechnen der Abstände im \mathbb{R}^2 Wurzeln vorkommen, was die Rechnung sehr kompliziert. Cette méthode fonctionne mal en pratique, car en calculant les distances dans le \mathbb{R}^2 on obtient des racines carrées ce qui complique passablement le calcul.

Idee 3: Idée 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. Choisis au lieu de la somme des valeurs absolues des distances à g la somme des carrés des valeurs absolues des distances. Par conséquent on élimine de cette manière les racines carrées.



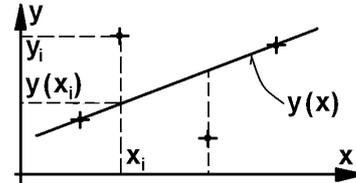
Problème: Praktisch ist zu einem gegebenen Wert x_i der Messwert y_i gegeben $\leadsto P_i$. Die Gerade soll so bestimmt werden, dass für die nächstgelegenen Punkte $P_i^* \in g$ die Summe der Distanzquadrate $|P_i P_i^*|^2$ minimal ist. Man hat das Problem der Minimalisierung unter der Bedingung, dass $P_i P_i^* \perp g$ gilt,

was die Rechnung wieder sehr kompliziert. Einfacher wird es, hier einen Fehler in Kauf zu nehmen und $x_i^* = x_i$ zu setzen. En pratique on a pour une valeur x_i donnée une valeur de mesure y_i donnée $\leadsto P_i$. La droite doit être déterminée de manière que pour les points les plus proches $P_i^* \in g$ la somme des carrés de distance $|\overline{P_i P_i^*}|$ soit minimale. Le problème de la minimalisation est lié à la condition suivante: $\overline{P_i P_i^*} \perp g$. Cela complique le calcul. Il est plus simple d'accepter une erreur et de mettre $x_i^* = x_i$.

Idee 4: Idée 4:

Nimm statt Summe der Quadrate der Abstände zu g nur die Summe der Δy_i .

Choisis, au lieu de la somme des carrés des valeurs absolues des distances à g , seulement la somme des Δy_i .



Problème: 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. $y = h(x) = a \sin(bx + c)$.

Sous l'hypothèse que la courbe soit une droite $y = g(x) = a \cdot x + b$, il faut maintenant calculer a et b . On procède de façon analogue pour d'autres courbes, p.ex. $y = h(x) = a \sin(bx + c)$ etc..

Solution:

$$(1) \sum_{i=1}^n (\Delta y_i)^2 = \sum_{i=1}^n (y_i - g(x))^2 = \sum_{i=1}^n (y_i - (ax + b))^2 := f(a, b) \rightarrow \text{Min}$$

$$(2) \sum_{i=1}^n (\Delta y_i)^2 = \sum_{i=1}^n (y_i - h(x))^2 = \sum_{i=1}^n (y_i - (a \sin(bx + c)))^2 := f(a, b) \rightarrow \text{Min}$$

Bedingung: Condition: $\frac{\partial f}{\partial a} = \frac{\partial f}{\partial b} = 0 \quad \leadsto$

$$\begin{aligned} \frac{\partial f}{\partial a} &= \sum_{i=1}^n 2(y_i - ax_i - b) \cdot (-x_i) = -2 \sum_{i=1}^n (y_i - ax_i - b) \cdot x_i = -2 \sum_{i=1}^n y_i \cdot x_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i \\ &= -2 \sum_{i=1}^n y_i \cdot x_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial f}{\partial b} &= \sum_{i=1}^n 2(y_i - ax_i - b) \cdot (-1) = -2 \sum_{i=1}^n (y_i - ax_i - b) = 0 \\ &\Rightarrow \sum_{i=1}^n y_i - \sum_{i=1}^n ax_i - \sum_{i=1}^n b = \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - nb = 0 \end{aligned}$$

Soit $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ (Mitterwert der x_i valeur moyenne des x_i), $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

$$\begin{aligned} \Rightarrow & \triangleright \sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i^2 + bn\bar{x} \\ & \triangleright \cancel{h} \bar{y} = \cancel{h} a \bar{x} + \cancel{h} b \Rightarrow y(\bar{x}) = a\bar{x} + b = \bar{y} \end{aligned}$$

$b = \bar{y} - a\bar{x}$ einsetzen: substituer: ($\Rightarrow \bar{y} = a\bar{x} + b$)

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i^2 + (\bar{y} - a\bar{x})n\bar{x} = a \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) + n\bar{x}\bar{y}$$

Indication:

$$\text{Soit } \vec{d}_e := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} := \vec{1}, \quad \vec{x} := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \vec{y} := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\Rightarrow \sum_{i=1}^n x_i y_i = \langle \vec{x}, \vec{y} \rangle, \quad \sum_{i=1}^n x_i = \langle \vec{x}, \vec{d}_1 \rangle, \quad \sum_{i=1}^n y_i = \langle \vec{y}, \vec{d}_1 \rangle \quad \rightsquigarrow$$

Formule:**Hyp.:**

Soit

 $y = a \cdot x + b$ Regressionsgerade g Droite de régression g **Thè.:**

$$a = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\langle \vec{x}, \vec{y} \rangle - n \bar{x} \bar{y}}{\langle \vec{x}, \vec{x} \rangle - n \bar{x}^2}$$

$$b = \bar{y} - \bar{x} \cdot \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \bar{y} - \bar{x} \cdot \frac{\langle \vec{x}, \vec{y} \rangle - n \bar{x} \bar{y}}{\langle \vec{x}, \vec{x} \rangle - n \bar{x}^2}$$

$$y(\bar{x}) = a \bar{x} + b = \bar{y} \Rightarrow (\bar{x}, \bar{y}) \in g$$

Schreibweise mit Varianz und Kovarianz — Façon d'écrire à l'aide de la variance et de la covariance

Considération:

$$\begin{aligned} s_x^2 &:= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \right) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - 2\bar{x} n \bar{x} + n \bar{x}^2 \right) = \frac{1}{n-1} \left(\left(\sum_{i=1}^n x_i^2 \right) - n \bar{x}^2 \right) = \frac{1}{n-1} \left(\left(\sum_{i=1}^n x_i^2 \right) - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) \\ &= \frac{1}{n-1} \left(\langle \vec{x}, \vec{x} \rangle - n \bar{x}^2 \right) \end{aligned}$$

In der mathematischen Statistik ist es üblich, die **Varianzen** s_x^2 , s_y^2 und die **Kovarianz** s_{xy} wie folgt zu definieren: Dans la statistique mathématique on a la coutume de définir les **variances** s_x^2 , s_y^2 et la **covariance** s_{xy} comme il suit:

Définition:

$$s_x^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) = \frac{1}{n-1} \left(\langle \vec{x}, \vec{x} \rangle - n \bar{x}^2 \right)$$

$$s_y^2 := \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \right) = \frac{1}{n-1} \left(\langle \vec{y}, \vec{y} \rangle - n \bar{y}^2 \right)$$

$$\begin{aligned} s_{xy} &:= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{n-1} \left(\left(\sum_{i=1}^n x_i y_i \right) - n \bar{x} \bar{y} \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right) \right) = \frac{1}{n-1} \left(\langle \vec{x}, \vec{y} \rangle - n \bar{x} \bar{y} \right) \end{aligned}$$

$$\text{Il vaut: } \bar{x} = \frac{1}{n} \langle \vec{x}, \vec{1} \rangle, \quad \bar{y} = \frac{1}{n} \langle \vec{y}, \vec{1} \rangle.$$

Mit Hilfe der Varianz und der Kovarianz lässt sich die Regressionsgerade einfacher schreiben. Dabei benutzen wir: A l'aide de la variance et de la covariance on peut simplifier la formule pour la droite de

régression. Pour celà nous utilisons:

Lemme:
$$a = \frac{s_{xy}}{s_x^2}, \quad b = \bar{y} - \bar{x} \cdot \frac{s_{xy}}{s_x^2}$$

(Durch ausmultiplizieren verifiziert man leicht: En multipliant les termes des deux côtés on vérifie facilement: $s_{xy} = s_x^2 \cdot a$)

Corollaire:

Hyp.:

Soit

$y = a \cdot x + b$ Regressionsgerade Droite de régression

Thè.:

$$y = \frac{s_{xy}}{s_x^2} \cdot x + \bar{y} - \bar{x} \cdot \frac{s_{xy}}{s_x^2}$$

Remarque:

Für die Summe der quadratischen Abstände gilt: Pour la somme des distances au carré il vaut:

$$\begin{aligned} \sum_{i=1}^n (\Delta y_i)^2 &= \sum_{i=1}^n (y_i - a x_i - b)^2 = \sum_{i=1}^n (y_i - \bar{y} - a(x_i - \bar{x}))^2 \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2a \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) + a^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= (n-1) \cdot (s_y^2 - 2a s_{xy} + a^2 s_x^2) = (n-1) \cdot (s_y^2 - 2a a s_x^2 + a^2 s_x^2) = (n-1) \cdot (s_y^2 - a^2 s_x^2) \\ \leadsto \sum_{i=1}^n (\Delta y_i)^2 = 0 &\Leftrightarrow s_y^2 = a^2 s_x^2 = \left(\frac{s_{xy}}{s_x^2}\right)^2 s_x^2 = \frac{s_{xy}^2}{s_x^2} \Leftrightarrow (s_x \cdot s_y)^2 = s_{xy}^2 \end{aligned}$$

Théorème:
$$\sum_{i=1}^n (\Delta y_i)^2 = 0 \Leftrightarrow (s_x \cdot s_y)^2 = s_{xy}^2$$

6.2.3 Korrelation — Corrélation

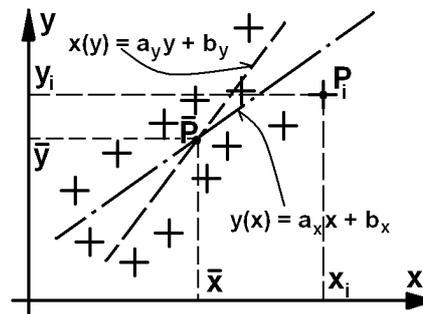
Vorhin haben wir die Gerade $g_x : y(x) = a \cdot x + b$ untersucht. Nous venons d'examiner la droite $g_x : y(x) = a \cdot x + b$.

Da y die abhängige und x die unabhängige Variable war, schreiben wir präziser: Comme y était la variable dépendante et x la variable indépendante, nous écrivons de manière plus précise:

$$y(x) = a_x \cdot x + b_x,$$

$$a_x = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{s_{xy}}{s_x^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

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. Nous échangeons maintenant les paires de valeurs (x_i, y_i) . Ç.v.d. nous considérons y comme variable indépendante et x comme variable dépendante. Par conséquent on peut de nouveau trouver la droite la "meilleure" $g_y : x(y) = a_y \cdot y + b_y$ par les points de mesure à l'aide de la méthode des carrés minimaux.



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} :
On obtient a_y et b_y de a_x et b_x par échangeement des rôles de x_i, \bar{x} et y_i, \bar{y} :

$$a_y = \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n y_i^2 - n \bar{y}^2} = \frac{s_{yx}}{s_y^2} = \frac{\sum_{i=1}^n (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

$$b_y = \bar{x} - \bar{y} \cdot \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n y_i^2 - n \bar{y}^2}$$

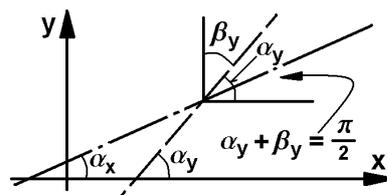
Wenn man g_x und g_y in dasselbe Koordinatensystem einzeichnet, ist zu erwarten, dass zwei verschiedene Geraden entstehen. Si on dessine g_x et g_y dans le même système de coordonnées, on peut s'attendre à deux droites différentes.

Ideal wäre allerdings, wenn g_x und g_y zusammenfallen würden. Dann hätte man: Il serait idéal si les deux droites étaient les mêmes. Alors on aurait:

$$a_x = \tan(\alpha_x),$$

$$a_y = \tan(\beta_y) = \tan\left(\frac{\pi}{2} - \alpha_x\right) = \frac{1}{\tan(\alpha_x)}$$

$$\Rightarrow a_x \cdot a_y = \tan(\alpha_x) \cdot \frac{1}{\tan(\alpha_x)} = 1$$

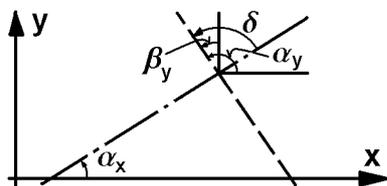


Andernfalls ist: Autrement on obtient:

$$a_x \cdot a_y = \tan(\alpha_x) \cdot \tan(\beta_y)$$

$$= \frac{\tan(\alpha_x)}{\tan(\frac{\pi}{2} - \beta_y)} = \frac{\tan(\alpha_x)}{\tan(\alpha_y)}$$

$$= \frac{\tan(\alpha_x)}{\tan(\alpha_x + \delta)} \neq 1$$



$a_x \cdot a_y = \frac{\tan(\alpha_x)}{\tan(\alpha_x + \delta)}$ ist umso verschiedener von 1, je verschiedener δ von 0 ist est plus différent de 1, plus δ est différent de 0. \leadsto

$a_x \cdot a_y$ ist ein Mass für den Zusammenhang der beiden Geraden, d.h. für die **Korrelation**.

$a_x \cdot a_y$ est une mesure pour le rapport entre les deux droites, ç.v.d. pour la **corrélacion**.

Es gilt: On constate: $(s_x, s_y \geq 0)$

$$a_x \cdot a_y = \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \cdot \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n y_i^2 - n \bar{y}^2} = \frac{(\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x})^2}{(\sum_{i=1}^n x_i^2 - n \bar{x}^2) \cdot (\sum_{i=1}^n y_i^2 - n \bar{y}^2)},$$

$$\sqrt{a_x \cdot a_y \cdot \operatorname{sgn}(a_x \cdot a_y)} = \frac{|\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}|}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2) \cdot (\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} = \frac{|s_{xy}|}{\sqrt{s_x^2 \cdot s_y^2}} := |r_{xy}|$$

Définition:

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} \text{ heisst } \mathbf{Korrelationskoeffizient}$$

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} \text{ s'appelle } \mathbf{coefficient de corrélation}$$

6.2.4 Korrelationskoeff.: Bedeutung — Coeff. de corrélation: Signification

Etude:

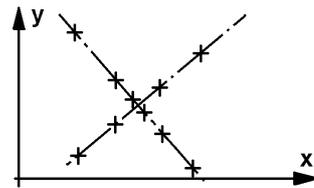
$$\leadsto r_{xy} = 0 \Leftrightarrow 0 = s_{xy} := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{n-1} \sum_{i=1}^n (\Delta x_i) \cdot (\Delta y_i) =$$

$$\frac{1}{n-1} ((\sum_{i=1}^n x_i y_i) - n \bar{x} \bar{y}) \Leftrightarrow \frac{1}{n} \cdot \sum_{i=1}^n x_i y_i = \langle \vec{x}, \vec{y} \rangle = \bar{x} \cdot \bar{y}, \quad \vec{x} = (x_1, \dots, x_n)^T, \quad \vec{y} = (y_1, \dots, y_n)^T$$

Conséquence:

$$r_{xy} = 0 \Leftrightarrow \left\langle \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix}, \begin{pmatrix} \Delta y_1 \\ \vdots \\ \Delta y_n \end{pmatrix} \right\rangle = 0 \Leftrightarrow \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} \perp \begin{pmatrix} \Delta y_1 \\ \vdots \\ \Delta y_n \end{pmatrix} \Leftrightarrow \overline{\langle \vec{x}, \vec{y} \rangle} = \bar{x} \cdot \bar{y}$$

Diese Situation tritt auf, wenn die Punkte „wolkenartig“ verteilt sind, wie man schon mit vier Punkten sieht, die die Ecken eines achsenparallelen Rechtecks bilden. On a cette situation si les points sont distribués en forme de nuage, comme on voit déjà avec quatre points qui sont situés dans les sommets d'un rectangle.



Etude détaillée:

Soient

$$\Delta \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \bar{x} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} = \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix},$$

Ici, on utilise:

$$\varphi = \chi(\Delta \vec{x}, \Delta \vec{y}).$$

$$\Delta \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \bar{y} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \Delta y_1 \\ \vdots \\ \Delta y_n \end{pmatrix}$$

Il vaut:

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{\frac{1}{n-1} \cdot \langle \Delta \vec{x}, \Delta \vec{y} \rangle}{\sqrt{\frac{1}{n-1} \cdot \langle \Delta \vec{x}, \Delta \vec{x} \rangle \cdot \frac{1}{n-1} \cdot \langle \Delta \vec{y}, \Delta \vec{y} \rangle}} = \frac{\frac{1}{n-1} \cdot |\Delta \vec{x}| \cdot |\Delta \vec{y}| \cdot \cos(\varphi)}{\frac{1}{n-1} \cdot |\Delta \vec{x}| \cdot |\Delta \vec{y}|} = \cos(\varphi)$$

$$r_{xy} = 0 \Leftrightarrow \langle \Delta \vec{x}, \Delta \vec{y} \rangle = 0 \Leftrightarrow \Delta \vec{x} \perp \Delta \vec{y} \Leftrightarrow \cos(\varphi) = 0 \Leftrightarrow \varphi = \frac{\pi}{2} + n \cdot \pi, \quad n \in \mathbb{Z}$$

$$r_{xy} = 1 \Leftrightarrow \langle \Delta \vec{x}, \Delta \vec{y} \rangle = |\Delta \vec{x}| \cdot |\Delta \vec{y}| \Leftrightarrow \Delta \vec{x} \parallel \Delta \vec{y} \Leftrightarrow |\cos(\varphi)| = 1 \Leftrightarrow \varphi = n \cdot \pi, \quad n \in \mathbb{Z}$$

Corollaire:

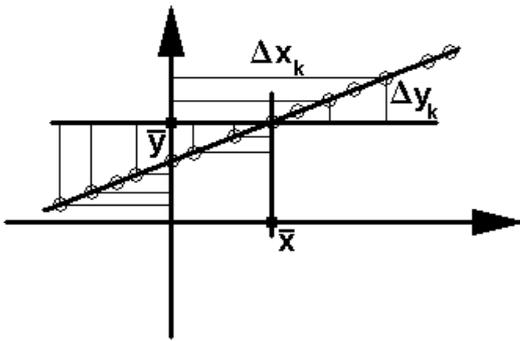
Hyp.:

$$\varphi = \angle(\Delta \vec{x}, \Delta \vec{y}), \quad \Delta \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \bar{x} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \Delta \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \bar{y} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

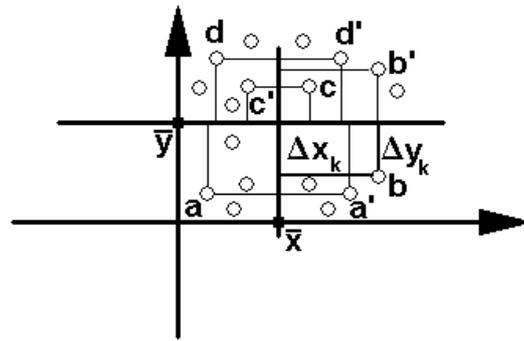
Thè.:

$$r_{xy} = 0 \Leftrightarrow \varphi = \frac{\pi}{2} + n \cdot \pi, \quad n \in \mathbb{Z}$$

$$|r_{xy}| = 1 \Leftrightarrow \varphi = n \cdot \pi, \quad n \in \mathbb{Z}$$



$$\varphi = n \cdot \pi, \quad n \in \mathbb{Z}$$



$$\varphi = \frac{\pi}{2} + n \cdot \pi, \quad n \in \mathbb{Z}$$

Exemple:

1. $x = \{2, 3, 4, 6, 8, 12\}$, $y = \{4, 6, 8, 12, 16, 24\}$ ($y = 2x$) $\Rightarrow r = 1$
2. $x = \{-2.1, -2.1, 0, 0, 2, 2.1\}$, $y = \{4, -4, 3, -3, 4, -4\}$ $\Rightarrow r = -0.0106422$
3. $x = \{-2.1, -2, 2, 2.1\}$, $y = \{4, -4, 4, -4\}$ $\Rightarrow r = -0.024383$

6.2.5 Rechner mit Punktschätzern: Probleme — Calculer avec des estimateurs: Problèmes

Soit μ_X = valeur attendue d'une grandeur mesurée X , $Y := f(X)$.

Soit μ_X estimé par un estimateur de vraisemblance sans biais $\hat{\mu}_X$.

$$\Rightarrow Y = \mu_Y + \Delta Y \approx f(\mu_X) + f'(\mu_X)(X - \mu_X) \text{ (Taylor).}$$

$$\begin{aligned} \leadsto \bar{y} &\approx \bar{\mu}_Y = \mu_Y \approx \bar{f}(\mu_X) + \overline{f'(\mu_X)(X - \mu_X)} = f(\mu_X) + \bar{f}'(\mu_X)(\bar{X} - \bar{\mu}_X) = f(\mu_X) + f'(\mu_X) \underbrace{(\bar{X} - \mu_X)}_{=0} \\ &\Rightarrow \bar{y} \approx \mu_Y \approx f(\mu_X) \end{aligned}$$

Théorème: $Y = f(X) \Rightarrow \bar{y} \approx \mu_Y \approx f(\mu_X)$

Amélioration de l'approximation:

$$\begin{aligned} Y = \mu_Y + \Delta Y &\approx f(\mu_X) + f'(\mu_X)(X - \mu_X) + \frac{1}{2} f''(\mu_X)(X - \mu_X)^2 \\ &\longrightarrow \mu_Y \approx f(\mu_X) + f'(\mu_X) \cdot 0 + \frac{1}{2} f''(\mu_X) \sigma_X^2 = f(\mu_X) + \frac{1}{2} f''(\mu_X) \sigma_X^2 \end{aligned}$$

Théorème: $\bar{x} \approx \mu_X, \bar{y} \approx \mu_Y, Y = f(X) \Rightarrow \bar{y} \approx f(\mu_X) + \frac{1}{2} f''(\mu_X) \sigma_X^2$

Application pour des intervalles de confiance:

Soit $\mu_X \in [\bar{x} - \Delta x, \bar{x} + \Delta x]$, $\Delta x \approx \sigma_X$ (bref $\mu_X = \bar{x} \pm \Delta x$).

\leadsto Problème: $\sigma_Y = ?$

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

Théorème: $\sigma_Y^2 \approx (f'(\mu_X))^2 \sigma_X^2$

Nous allons continuer de développer ces idées dans les exercices... On trouve des explications auxiliaires sous le lien qui suit:

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

6.3 Quant à la fin

On a ainsi discuté les premières bases des statistiques. Mais en pratique, on ne peut pas faire de grands sauts avec cela. On n'a pas encore parlé de l'analyse de variance, de l'analyse des séries chronologiques et de telles choses. Beaucoup de tests ne sont pas mentionnés...

Problème:

La distribution normale n'existe presque jamais dans la réalité, parce que les intervalles réels ne sont pratiquement jamais infiniment grands et souvent ils ne sont que positifs. Beaucoup de variables de test sont seulement des approximations. Pour un travail sérieux il faut un grand matériel de données... Là la littérature spécialisée va nous aider.

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 Masstheorie, 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, Neuausgabe (Bibl.: A10)
- [11] Meschkowski. Mathematisches Begriffswörterbuch. BI Hochschultaschenbücher. Bibliographisches Institut Mannheim. (Bibl.: A11)
- [12] Regina Storm, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Statistische Qualitätskontrolle, Fachbuchverlag Leipzig, Köln (Bibl.: A12)
- [13] L.B. van der Waerden, Mathematische Statistik, Birkhäuser (Bibl.: A13)
- [14] Mathematikkurs für Ingenieure
Teil 4 \diamond Einführung in die Boolesche Algebra, Rolf Wirz, Biel 1999, (Bibl.: A14)

<http://rowicus.ch/Wir//Scripts/Teil4Bool.pdf>
- [15] Script \diamond Math \diamond Ing, \diamond Analysis \diamond Analyse \diamond , kurz & bündig \diamond concis, Rolf Wirz, Biel 1999, (Bibl.: A15)

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

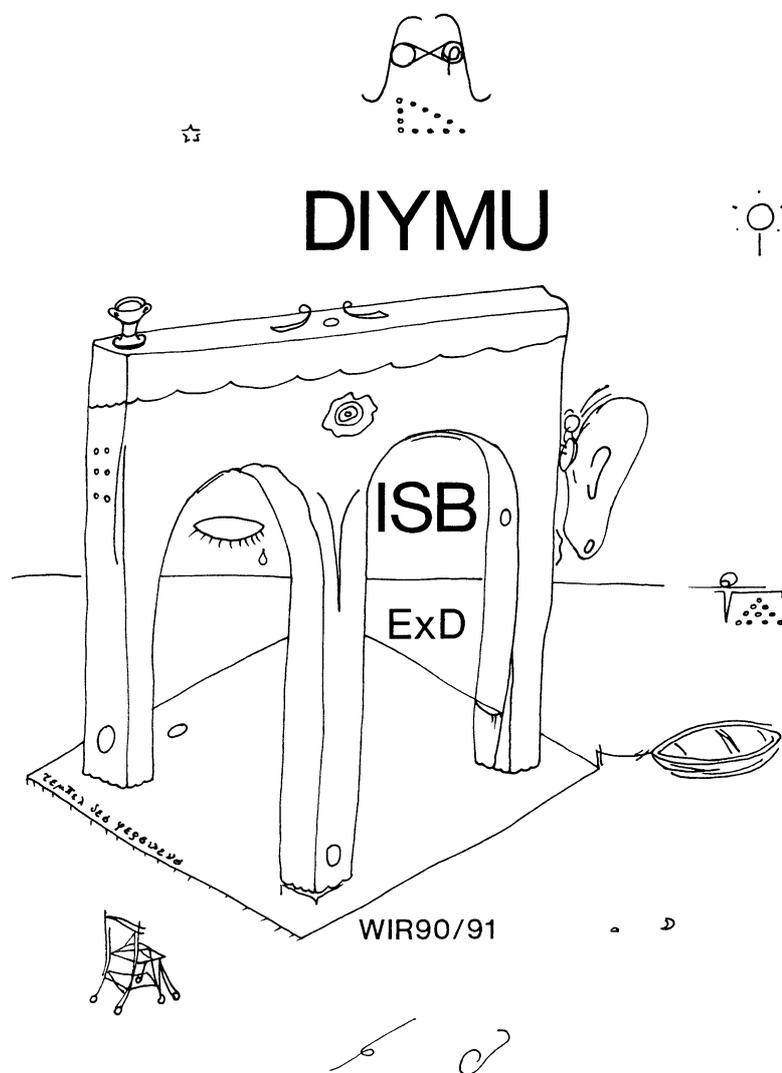
- [16] Script \diamond Math \diamond Ing, \diamond Fortsetzung Mathematik \diamond Suite mathématiques \diamond , kurz & bündig \diamond concis, Rolf Wirz, Biel 2000/01,... (Bibl.: A16)

<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



Excercises

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

Anhang B

Fehler von statistischen Kenngrößen und Standardfehler

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

Siehe auf

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

Anhang C

Monte-Carlo, Resampling und anderes

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

Siehe auf

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

Anhang D

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

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

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

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

Siehe auf

<http://rowicus.ch/Wir/Scripts/Zusatz/AnhangStatistDatenkorrektur.pdf>

oder

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

Anhang F

Spezielle Wahrscheinlichkeitssituationen

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

Siehe auf

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

Anhang G

Hinweise zur Datenanalyse

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

Siehe auf

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

Anhang H

Crashkurs Kombinatorik, Wahrscheinlichkeit: Link

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

Nouvel exposé allemand voir:

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

Ende *Fin*