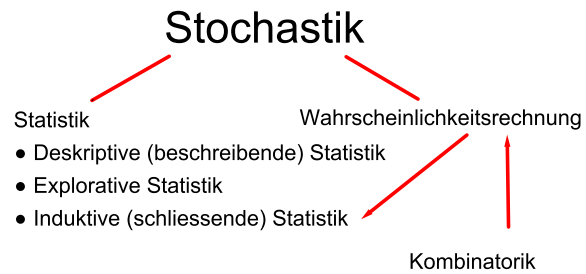


Wahrscheinlichkeitsrechnung und Statistik(WS)

Zusammenfassung (Stand FS2021)

“Die Wahrscheinlichkeitsrechnung lehrt, wie man mit Wahrscheinlichkeiten rechnet, die Statistik lehrt, wie man sie bestimmt.”



Inhaltsverzeichnis

1 Beschreibende Statistik	5
1.1 Statistische Merkmale und Variablen	5
1.1.1 Statistische Einheiten und Grundgesamtheiten	5
1.1.2 Merkmale und Merkmalsausprägungen	6
1.1.3 Niveau der Messbarkeit bzw. Skalenniveau	7
1.1.4 Stichproben und Teilgesamtheiten	8
1.2 Statistische Verteilung	9
1.2.1 Absolute Häufigkeit	10
1.2.2 Relative Häufigkeit	11
1.3 Empirische Häufigkeitsfunktion und empirische Verteilungsfunktion	12
1.4 Häufigkeitsdichte und Histogramm	15
2 Kenngrössen einer Stichprobe	18
2.1 Lagekennwerte	18
2.1.1 Arithmetisches Mittel	18
2.1.2 Geometrisches Mittel	20
2.1.3 Harmonisches Mittel	20
2.1.4 Modus und Median	22
2.2 Quantile	23
2.3 Streuungskenngrössen	26
2.3.1 Spannweite	26
2.3.2 Empirische Varianz	26
2.3.3 Interquartilsabstand	27
3 Lineare Korrelation	27
3.1 Streudiagramm	27
3.2 Kennzahl für das Ausmass des linearen Zusammenhangs	28
4 Lineare Regression	31

5	Wahrscheinlichkeitsrechnung	33
5.1	Kombinatorik	33
5.1.1	Fakultäten und Binomialkoeffizienten	34
5.1.2	Fundamental- oder Multiplikationsprinzip der Kombinatorik	35
5.1.3	Permutationen	37
5.1.4	Anzahl von Auswahlmöglichkeiten k aus n mit Berücksichtigung der Reihenfolge (Variationen)	38
5.1.5	Anzahl von Auswahlmöglichkeiten k aus n ohne Berücksichtigung der Reihenfolge (Kombinationen)	39
5.2	Grundlagen der Wahrscheinlichkeitsrechnung	41
5.2.1	Ereignisraum und Ereignismenge	41
5.2.2	Zuordnung von Wahrscheinlichkeiten	44
5.2.3	Wichtige Regeln für Wahrscheinlichkeiten	47
5.3	Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit	50
5.3.1	Totale Wahrscheinlichkeit und Bayes' Theorem	53
6	Zufallsvariablen	57
6.1	Zufallsvariablen und ihre Verteilungsfunktionen	57
6.2	Diskrete Zufallsvariablen	60
6.3	Stetige Zufallsvariablen	61
6.4	Der Erwartungswert einer Zufallsvariablen	63
6.4.1	Der Erwartungswert einer diskreten Zufallsvariablen	64
6.4.2	Der Erwartungswert einer stetigen Zufallsvariablen	64
6.4.3	Eigenschaften des Erwartungswerts	64
6.5	Die Varianz einer Zufallsvariablen	65
6.5.1	Die Varianz einer diskreten Zufallsvariablen	65
6.5.2	Die Varianz einer stetigen Zufallsvariablen	65
6.5.3	Eigenschaften der Varianz	66
6.6	Standardisieren von Zufallsvariablen	66
6.7	Quantil, Median und Modus (Kein Modulstoff!)	66

7	Spezielle Verteilungen	69
7.1	Diskrete Verteilungen	69
7.1.1	Gleichförmige Verteilung oder Gleichverteilung (diskret)	69
7.1.2	Bernoulli-Verteilung	70
7.1.3	Binomialverteilung	71
7.1.4	Poisson-Verteilung	73
7.1.5	Approximation der Binomialverteilung durch Poisson-Verteilung (Kein Modulstoff!)	75
7.2	Stetige Verteilungen	76
7.2.1	Stetige Gleichverteilung (oder Rechteckverteilung)	76
7.2.2	Exponentialverteilung (Kein Modulstoff!)	78
7.2.3	Normalverteilung	80
8	Grundlegende Problemstellung der schliessenden Statistik	84
9	Grenzwertsätze	85
9.1	Das Schwache Gesetz der grossen Zahlen	85
9.1.1	Tschebychev-Ungleichung	86
9.2	Der Zentrale Grenzwertsatz (ZGWS)	89
9.2.1	Stichprobenverteilungen	90
9.2.2	Approximation der Binomialverteilung durch die Normalverteilung	93
10	Punktschätzungen (Parameterschätzungen)	95
10.1	Punktschätzung für den Mittelwert	95
10.1.1	Eigenschaften der Punktschätzung für den Mittelwert	96
10.1.2	Schätzung des Anteilswerts als Spezialfall	96
10.2	Punktschätzung für die Varianz	97
10.3	Ziehen mit und ohne Zurücklegen: Korrekturfaktor	99
11	Intervallschätzungen	100
11.1	Intervallschätzungen für Mittelwerte	100

11.1.1 Fall 1: Stichprobenvariablen X_1, \dots, X_n unabh. normalverteilt mit Parametern μ und bekanntem σ^2	101
11.1.2 Fall 2: Stichprobenvariablen X_1, \dots, X_n unabh. normalverteilt mit Parametern μ und unbekanntem σ^2	102
11.1.3 Fall 3: X_1, \dots, X_n unabh. beliebig identisch verteilte ZV und $n \geq 30$	102
11.1.4 Intervallschätzungen für Mittelwerte mit grossen Stichproben (also $n \geq 30$), d.h. Fall 3)	103
11.1.5 Intervallschätzungen für den Mittelwert eines normalverteilten Merkmals bei kleinen Stichproben (Verletzung von $n \geq 30$), d.h. Fall 2	105
11.1.6 Intervallschätzungen für den Mittelwert bei Verletzung von $\frac{N-n}{N-1} \geq 0.9$	105
11.1.7 Intervallschätzungen für den Anteilswert als Spezialfall von Fall 3	106
12 Zwei-dimensionale Verteilungen	107
12.1 Streudiagramme und Kontingenztabellen	107
12.2 Randverteilungen	109
13 Testen von Hypothesen: Binomialtest, t-Tests, χ^2-Tests	110
13.1 Binomialtest	111
13.2 Einstichproben t -Test	114
13.2.1 Zweiseitiges Testniveau	114
13.2.2 Einseitiges Testniveau	116
13.3 t -Tests über Anteilswerte	118
13.4 χ^2 -Anpassungstest	119
13.5 χ^2 -Unabhängigkeitstest	122

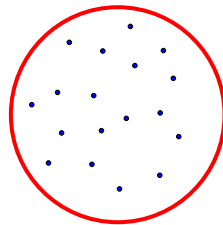
1 Beschreibende Statistik

Die beschreibende (deskriptive) Statistik hat zum Ziel, empirische Daten darzustellen (z.B. durch Tabellen und Graphiken), zu ordnen und zu komprimieren sowie zu analysieren und zu interpretieren. Durch geeignete grundlegende Kenngrößen wie z.B. Lage- und Streuungsparameter und Konzentrationsmasse versucht man darüber hinaus, Daten zahlenmässig zu beschreiben.

Von der induktiven Statistik (folgt später im Modul) unterscheidet sich die beschreibende Statistik dadurch, dass sie keine Aussagen zu einer über die untersuchten Fälle hinausgehende Grundgesamtheit macht und keine Überprüfung von Hypothesen ermöglicht.

1.1 Statistische Merkmale und Variablen

1.1.1 Statistische Einheiten und Grundgesamtheiten



• - statistische Einheit (Träger von Merkmalen oder Informationen)

Hauptuntersuchungsobjekte der beschreibenden Statistik sind *statistische Einheiten* (auch *Unit*, *Untersuchungs-* oder *Erhebungseinheit* genannt) als Träger von Merkmalen oder Informationen. Dabei stehen jedoch nicht einzelne statistische Einheiten im Mittelpunkt des Interesses, sondern Massenphänomene in einer bestimmten Menge von im wesentlichen gleichartigen statistischen Einheiten. Die Abgrenzung dieser Menge erfolgt durch *Identifikationskriterien*, die sachlicher, räumlicher und zeitlicher Natur sein können.

Beispiele 1.1: Identifikationskriterien

- i) Die Studierenden (sachlich), die sich gerade (zeitlich) in diesem Vorlesungssaal (räumlich) befinden.
- ii) Die Kraftfahrzeuge (sachlich), die im Jahr 2020 (zeitlich) im Kanton Zürich (räumlich) zugelassen waren.

Formal kann diese Menge, die sogenannte *Grundgesamtheit* G , wie folgt charakterisiert werden:

$$G = \{e \mid e \text{ erfüllt Identifikationskriterien}\}$$

(Die Grundgesamtheit G besteht aus allen statistischen Einheiten e , die die Identifikationskriterien erfüllen.)

Andere Namen für Grundgesamtheit sind *statistische Masse*, *Population* und *Kollektiv*.

Mit $n(G)$ bezeichnen wir die Anzahl oder den Umfang der statistischen Einheiten in der Grundgesamtheit. *Reale* Grundgesamtheiten enthalten stets nur endlich viele Elemente. Nur *fiktive* oder *hypothetische* Grundgesamtheiten können auch unendlich viele Einheiten enthalten.

1.1.2 Merkmale und Merkmalsausprägungen

Definition 1.2: Merkmal

Ein Merkmal ist eine erkennbare Eigenschaft, die eine Person, eine Sache oder einen abstrakten Zusammenhang von anderen unterscheidet.

Unterscheidbare Erscheinungsformen eines Merkmals heissen *Merkmalsausprägungen*, *Modalitäten* oder (vor allem bei quantitativen Merkmalen) *Merkmalswerte*. Z.B. sind Rot, Blau, Grün etc. Ausprägungen für das Merkmal "Farbe".

Merkmale können nach verschiedenen Gesichtspunkten typisiert werden:

- *Qualitative* Merkmale variieren artmässig (Farbe, Geschlecht, Religionszugehörigkeit, Staatsbürgerschaft, Dienstgrade etc.)
- *Quantitative* Merkmale variieren zahlenmässig (z.B. alles, was gemessen werden kann)
- *Diskrete* Merkmale haben nur endlich viele oder abzählbar unendlich viele Ausprägungen, d.h. Ausprägungen sind "isolierte" Punkte (z.B. Anzahlen, Platzierungen, Geschlecht etc.)
- *Stetige* Merkmale können jeden beliebigen Merkmalswert in einem ganzen Bereich (Intervall) annehmen. (z.B. Temperatur, Gewicht, etc.)

Bemerkung 1.3

In der Praxis ist die Unterscheidung zwischen stetigen und diskreten Merkmalen vielfach willkürlich. Die Körpergrösse einer Person ist sicher ein stetiges Merkmal. Sie kann aber infolge der begrenzten Messgenauigkeit nur in diskreten Sprüngen erfasst werden (z.B. auf 1cm oder 0.5cm etc. genau). Andererseits lassen sich "feinabgestufte" diskrete Merkmale (z.B. Gehälter

auf einen Rappen genau) bei der statistischen Auswertung genau so wie stetige behandeln.

1.1.3 Niveau der Messbarkeit bzw. Skalenniveau

Das *Skalenniveau* (bzw. *Messniveau*) ist eine wichtige Eigenschaft von Merkmalen. Je nach der Art des Merkmals bzw. je nachdem, welche Vorschriften bei seiner Messung eingehalten werden können, lassen sich verschiedene Stufen der Skalierbarkeit unterscheiden. Das Skalenniveau bestimmt

- die (mathematischen) Operationen, die mit einer entsprechend skalierten Variable zulässig sind.
- welche Transformationen mit entsprechend skalierten Variablen durchgeführt werden können, ohne Information zu verlieren bzw. zu verändern.
- welche Information das entsprechende Merkmal liefert, welche Interpretationen Ausprägungen des entsprechenden Merkmals zulassen.

Es wird (meistens) zwischen den folgenden Skalenniveaus unterschieden:

- i) *Nominal skalierte (oder nominal messbare) Merkmale (Variablen)*: Lediglich Gleichheit oder Andersartigkeit verschiedener Ausprägungen kann festgestellt werden; keinerlei Bewertung oder Quantifizierung beabsichtigt; stets qualitativ. Ausprägungen sind Namen oder Kategorien, die den Einheiten zugeordnet werden.

Beispiele 1.4: Nominal skalierte Merkmale

Farben, Religion, Gehaltsklasse, Geschlecht,... . Ausprägungen werden häufig durch Zahlen *signiert*.

- ii) *Ordinal skalierte (oder ordinal messbare) Merkmale (Variablen)*: Ausprägungen können geordnet werden, aber ihre Abstände sind nicht interpretierbar.

Beispiele 1.5: Ordinal skalierte Merkmale

Schulnoten (1,2,3,4,5,6); Kleidergrössen (S,M,L,XL), Gewichtsklassen, Angenehmheit von Düften, Härte von Mineralien,... .

- iii) *Intervallskalierte Merkmale (Variablen)*: Ausprägungen haben Rangfolge und quantitative Unterschiede zwischen ihnen können interpretiert werden (Addition und Subtraktion von Ausprägungen machen Sinn!). Ausprägungen sind immer numerisch gegeben. Quotienten zwischen Ausprägungen machen jedoch keinen Sinn. Intervallskala hält Intervalle zwischen den Skalenwerten (Ausprägungen) für empirisch bedeutsam, nicht aber die absoluten Werte. In einer Intervallskala ist deshalb sowohl der Ausgangswert (= Nullpunkt) als auch die numerische Grösse der Einheit frei wählbar.

Beispiele 1.6: Intervallskalierte Merkmale

Temperaturskalen (Celsius, Fahrenheit, **nicht** Kelvin): Gleiche Differenzen auf dieser Skala repräsentieren gleiche Unterschiede in der Temperatur, aber eine Temperatur von 30 Grad Fahrenheit/Celsius ist nicht doppelt so warm wie 15 Grad. Kalender-Daten.

- iv) *Verhältnisskalierte Merkmale (Variablen)* Über die für intervallskalierte Merkmale erwähnten Eigenschaften hinaus, haben verhältnisskalierte Merkmale einen absoluten Nullpunkt (Einheit ist willkürlich, aber Nullpunkt ist fest); Quotienten machen Sinn (“...doppelt so gross...”, “... halb so schwer...”,...)

Beispiele 1.7: Verhältnisskalierte Merkmale

Länge, Gewicht, Dichte,...

Bemerkungen 1.8

- i) Intervall- und Verhältnisskala werden unter dem Überbegriff der *metrischen* (oder *kardinalen*) Skala zusammengefasst.
- ii) Die Einteilung in Skalentypen ist nicht unumstritten. Es ist eine Vielzahl anderer Einteilungen denkbar.

1.1.4 Stichproben und Teilgesamtheiten

Ist G eine Grundgesamtheit und werden die Merkmalsausprägungen $M(e)$ eines Merkmals M für jede statistische Einheit $e \in G$ erhoben, spricht man von einer *Voll-* oder *Totalerhebung*. Solche Erhebungen sind jedoch oft weder sinnvoll noch möglich. Einige Gründe hierfür können sein:

- Kosten oder Zeitgründe (Grundgesamtheit steht prinzipiell zur Verfügung): Volksbefragungen, Wahlhochrechnungen, Qualitätskontrollen
- Grundgesamtheit steht prinzipiell nicht zur Verfügung: Zerstörende Materialprüfung, Kreditwürdigkeit von Schuldern, Bestimmung des Risikos bei Versicherungsabschlüssen, etc.

Merkmalsausprägungen werden daher häufig für eine Teilmenge $G^* \subset G$ von G erhoben. G^* wird dann als *Teilgesamtheit* und die auf ihr basierende Erhebung als *Teilerhebung* bezeichnet. Eine Teilgesamtheit wird als *Stichprobe* bezeichnet (Teilerhebungen und Stichproben haben **immer** endlichen Umfang), falls bei ihrer Auswahl der Zufall wesentlich beteiligt war. Man unterscheidet zwischen

- *Reine Stichprobe*: Jedes $e \in G$ hat gleiche “Chance” in G^* aufgenommen zu werden.

- *Repräsentative Stichprobe*: Teilgesamtheit ist repräsentativ für Grundgesamtheit, d.h. Verteilung wesentlicher Merkmale in G^* und G ist zumindest ähnlich.

Bemerkung 1.9

Hassloch in der Pfalz (Deutschland) ist in ökonomisch/sozialer Hinsicht ein “repräsentatives Abbild” Deutschlands, Deutschland im Kleinforma sozuagen. Deshalb führt die Gesellschaft für Konsumforschung Marktforschungsprojekte häufig in dieser Kleinstadt durch.^a

^ahttp://www.focus.de/finanzen/news/tid-12695/gesellschaft-klein-deutschland-mitten-in-der-pfalz_aid_351714.html

1.2 Statistische Verteilung

Eine *Urliste* entsteht durch Untersuchung einer Grundgesamtheit, Teilgesamtheit oder Stichprobe bzgl. eines Merkmals X .

Statistische Einheit	e_1	e_2	\cdots	e_i	\cdots	e_n
Merkmalswert $X(e_i)$	x_1	x_2	\cdots	x_i	\cdots	x_n

Beispiel 1.10: Urliste

Individuum e_1 wird zugeordnet x_1 = Körpergrösse von e_1 , Individuum e_2 wird zugeordnet x_2 = Körpergrösse von e_2 , etc.

x_1, x_2, \dots, x_n heisst *Beobachtungsreihe*.

Ein Beispiel hierfür ist die folgende Reihe. (Bitte beachten Sie, dass die Natur der Messdaten nicht relevant ist.)

$$x_1 = 5.0, x_2 = 1.6, 3.0, 4.1, 1.6, 3.0, 4.1, 4.1, 5.0, 5.0, 4.1, 5.0, 3.0, 4.1, 3.0, 5.0, 4.1, 5.0, 4.1, x_{20} = 4.1$$

Wir haben es hier mit einer Beobachtungsreihe vom Umfang $n = 20$ zu tun.

Bemerkung 1.11

Die Auswertung grosser Datenmengen kann nicht per Hand erledigt werden. Dafür gibt es Statistikprogramme wie z.B. R, SPSS oder SAS. Für den “Privatgebrauch” kann man mit der Datenanalyse jedoch schon sehr weit mit EXCEL unter WINDOWS kommen. Wir werden R benutzen.

Häufig werden Beobachtungsreihen der Grösse nach geordnet (Erinnern Sie sich: das Ordnen von Daten (wie elementar auch immer) ist eine der Aufgaben der (beschreibenden) Statistik!): Ordnen

aufsteigend der Grösse nach führt in unserem konkreten Beispiel zu:

$$x_{[1]} = 1.6, x_{[2]} = 1.6, 3.0, 3.0, 3.0, 3.0, 4.1, 4.1, 4.1, 4.1, 4.1, 4.1, 4.1, 4.1, 5.0, 5.0, 5.0, 5.0, 5.0, x_{[20]} = 5.0$$

Bemerkung 1.12

Bitte beachten Sie, dass Bärtil bei geordneten Beobachtungsreihen immer in eckige Klammern eingeschlossene Indizes benutzt! Es gilt also stets $x_{[1]} \leq x_{[2]} \leq x_{[3]} \leq \dots$.

Eines unserer Ziele wird sein, diese Daten (und jegliche andere Daten) zu beschreiben bzgl. ihrer Verteilung, d.h. der Häufigkeit ihres Auftretens, bzgl. ihrer Lage, Streuung u.s.w..

Die in einer Beobachtungsreihe enthaltene Information kann auf verschiedene Arten (ohne Verlust) gespeichert werden.

1.2.1 Absolute Häufigkeit

Anstatt die Beobachtungsreihe vollständig aufzulisten, könnte man geradeso gut sagen: Ausprägung 1.6 ist 2-mal aufgetreten, Ausprägung 3.0 ist 4-mal aufgetreten etc. Man kann das betrachten als Übergang von der Auflistung aller Elemente der Beobachtungsreihe, d.h. aller Einzelbeobachtungen, zur Auflistung der aufgetretenen Merkmalsausprägungen zusammen mit der (absoluten) Häufigkeit ihres Auftretens. (Wenn Bärtil nur noch die unterschiedlichen aufgetretenen Merkmalsausprägungen betrachtet, verwendet er dafür die Symbole a_i !)

Das wird folgendermassen notiert: (Die absolute Häufigkeit, dass das Merkmal X eine Ausprägung a_i annimmt): $\text{absH}(X = a_i) = n_i$. In unserem konkreten Fall also:

$$\text{absH}(X = a_1 = 1.6) = n_1 = 2, \dots, \text{absH}(X = a_4 = 5.0) = n_4 = 6$$

Tabellarisch sieht das wie folgt aus:

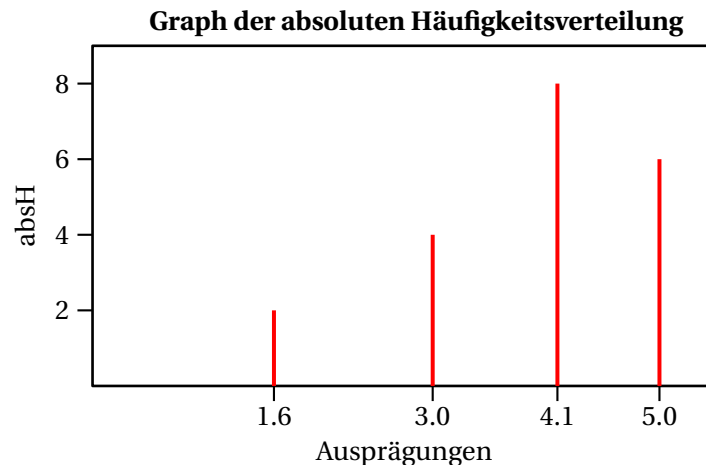
Ausprägung a_i	$a_1 = 1.6$	$a_2 = 3.0$	$a_3 = 4.1$	$a_4 = 5.0$
absolute Häufigkeit $\text{absH}(X = a_i)$	$n_1 = 2$	$n_2 = 4$	$n_3 = 8$	$n_4 = 6$

Die Zuordnung $a_i \mapsto \text{absH}(X = a_i) = n_i$ bzw. die Menge der Paare $(a_1, n_1), \dots, (a_k, n_k)$ heisst *absolute Häufigkeitsverteilung* des Merkmals X .

Für unser konkretes Beispiel haben wir also:

$$\begin{array}{cccc} 1.6 & 3.0 & 4.1 & 5.0 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 2 & 4 & 8 & 6 \end{array}$$

Diese Zuordnung kann auch graphisch dargestellt werden.



1.2.2 Relative Häufigkeit

Eine andere Möglichkeit der Aufzeichnung der Beobachtungsreihe ohne Informationsverlust besteht darin, anstatt der absoluten Häufigkeiten die *relativen Häufigkeiten* des Auftretens der verschiedenen Ausprägungen zu verwenden. Diese erhält man durch Teilen der absoluten Häufigkeiten durch den Umfang der Beobachtungsreihe. Die relative Häufigkeit, dass das Merkmal X eine Ausprägung a_i annimmt, wird folgendermassen notiert: $\text{relH}(X = a_i) = n_i^* (= \frac{n_i}{n})$. In unserem konkreten Fall also:

$$\text{relH}(X = a_1 = 1.6) = n_1^* = 0.1, \dots, \text{relH}(X = a_4 = 5.0) = n_4^* = 0.3$$

Tabellarisch:

Ausprägung a_i	$a_1 = 1.6$	$a_2 = 3.0$	$a_3 = 4.1$	$a_4 = 5.0$
relative Häufigkeit $\text{relH}(X = a_i)$	$n_1^* = 0.1$	$n_2^* = 0.2$	$n_3^* = 0.4$	$n_4^* = 0.3$

Beachten Sie bitte, dass die Summe aller relativen Häufigkeiten immer 1 ergibt.

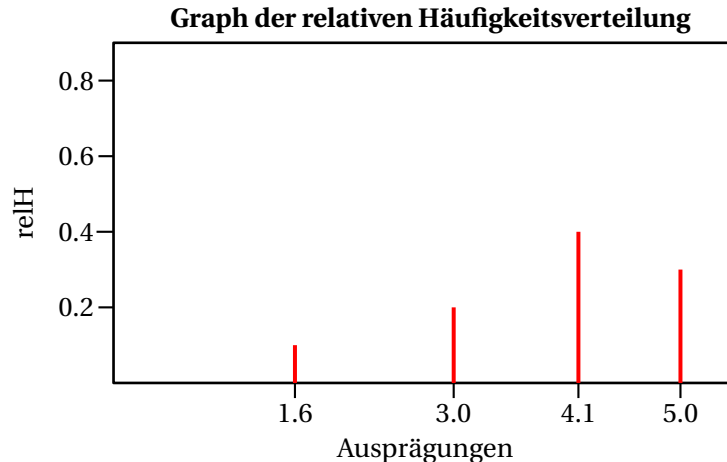
Die Zuordnung $a_i \mapsto \text{relH}(X = a_i) = n_i^*$ bzw. die Menge der Paare $(a_1, n_1^*), \dots, (a_k, n_k^*)$ heisst *relative Häufigkeitsverteilung* des Merkmals X .

Für unser konkretes Beispiel haben wir also:

$$\begin{array}{cccc}
 1.6 & 3.0 & 4.1 & 5.0 \\
 \downarrow & \downarrow & \downarrow & \downarrow \\
 0.1 & 0.2 & 0.4 & 0.3
 \end{array}$$

Auch diese Zuordnung kann wieder graphisch dargestellt werden.

Das Resultat:



Stabdiagramme dieser Art eignen sich vor allem zur Darstellungen von Häufigkeiten bei Merkmalen mit wenigen Ausprägungen. In der Praxis hat man es aber meist mit Merkmalen mit vielen möglichen Ausprägungen zu tun (Denken Sie nur an stetige Merkmale!), die jeweils vergleichsweise selten (oder nie) auftreten. Eine Abhilfe kann hier das Klassieren, d.h. das Zusammenfassen von Daten zu Gruppen, schaffen. Wir kommen darauf zurück.

1.3 Empirische Häufigkeitsfunktion und empirische Verteilungsfunktion

Die absolute und relative Häufigkeitsverteilung sind Funktionen, die nur in den aufgetretenen Ausprägungen eines Merkmals definiert sind (siehe letzter Abschnitt). In diesem Abschnitt definieren wir Funktionen, deren Definitionsgebiet die Menge der reellen Zahlen \mathbb{R} , d.h. die ganze Zahlengerade ist. Dazu betrachten wir wieder ein Merkmal X . Die verschiedenen aufgetretenen Ausprägungen bezeichnen wir wie üblich mit a_1, a_2, \dots, a_k .

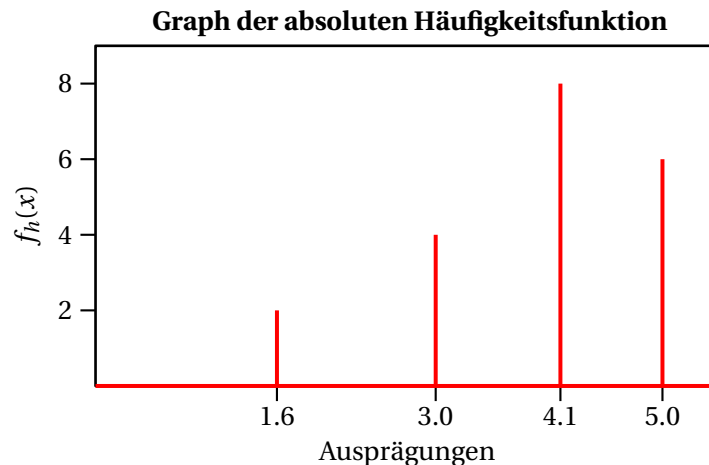
Die Funktion $f_h : \mathbb{R} \rightarrow [0, 1]$, definiert durch

$$f_h(x) = \begin{cases} n_i & \text{für } x = a_i \\ 0 & \text{sonst} \end{cases}$$

heisst *absolute (empirische) Häufigkeitsfunktion* von X .

Bemerkung 1.13

Nochmal der Unterschied zwischen absoluter Häufigkeitsverteilung und Häufigkeitsfunktion: erstere ist **nur** in a_1, a_2, \dots, a_k definiert. Die Funktionswerte sind die jeweiligen absoluten Häufigkeiten. Auch für die absolute Häufigkeitsfunktion gilt $f_h(a_i) = n_i$. **Darüber hinaus** sind aber für die Häufigkeitsfunktion “zwischen” den Ausprägungen die Funktionswerte gleich 0. Vergleichen Sie den Graph der folgenden Funktion mit der absoluten Häufigkeitsverteilung.



Ersetzt man in $f_h(x)$ die absoluten Häufigkeiten mit den relativen Häufigkeiten n_i^* , erhält man die *relative (empirische) Häufigkeitsfunktion* $f_h^*(x)$.

Wir kommen nun zur wichtigsten Funktion zur Beschreibung von Merkmalen, der sogenannten *relativen (empirischen) Verteilungsfunktion* $F_h^*(x)$.

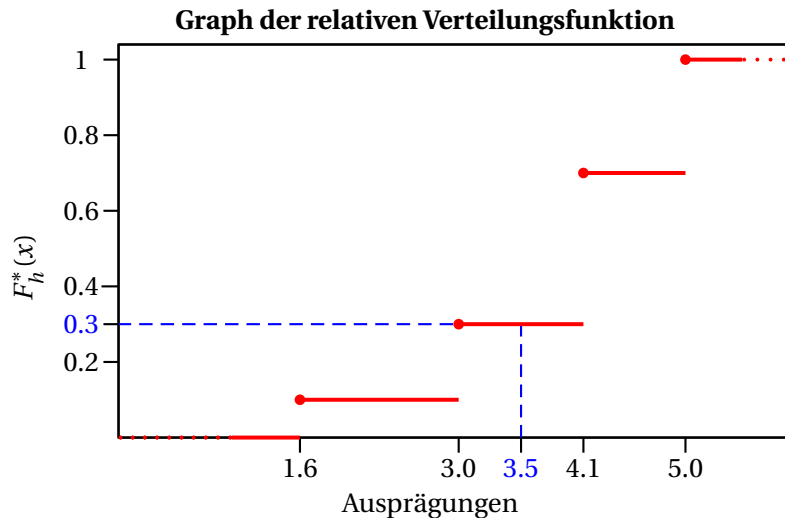
Dazu ordnen wir die aufgetretenen Ausprägungen aufsteigend: $a_{[1]} < a_{[2]} < \dots < a_{[k]}$. Die entsprechenden relativen Häufigkeiten bezeichnen wir wieder mit $n_1^*, n_2^*, \dots, n_k^*$.

Die empirische Verteilungsfunktion hat als Definitionsbereich die Menge der reellen Zahlen \mathbb{R} und als Wertebereich das Intervall $[0, 1]$. Sie wird definiert durch:

$$F_h^*(x) = \begin{cases} 0 & \text{für } x < a_{[1]} \\ \sum_{i: a_{[i]} \leq x} n_i^* & \text{für } a_{[1]} \leq x \leq a_{[k]} \\ 1 & \text{für } x > a_{[k]} \end{cases}$$

In Worten: F_h^* beginnt in “ $-\infty$ ” bei 0, springt in $a_{[1]}$ **um** n_1^* **auf** n_1^* , springt in $a_{[2]}$ **um** n_2^* **auf** $n_1^* + n_2^*, \dots$, springt in $a_{[k]}$ **um** n_k^* **auf** $n_1^* + n_2^* + \dots + n_k^* = 1$ und bleibt auf 1 bis “ $+\infty$ ”.

Die empirische Verteilungsfunktion hat also die Gestalt einer Treppenfunktion. Für unser Beispiel sieht das folgendermassen aus:



Bemerkungen 1.14

- i) Die Punkte in der obigen Graphik bezeichnen Funktionswerte. Es gilt also z.B. $F_h^*(1.6) = 0.1$. Das hat mit der Rechtsstetigkeit von Verteilungsfunktionen zu tun. Siehe die Eigenschaften von Verteilungsfunktionen im Anschluss.
- ii) Im Bärtil gibt es auch absolute (empirische) Verteilungsfunktionen. Bei diesen springt man in den Ausprägungen jeweils um die absoluten Häufigkeiten. Ich rate davon ab, diese zu benutzen. Das ist sehr ungewöhnlich!

Eine Charakterisierung der Verteilungsfunktion, die für uns von Bedeutung ist, ist die folgende:

$$F_h^*(x) = \text{relH}(X \leq x) = \text{Rel. Häufigkeit, dass Merkmal } X \text{ Ausprägungen } \leq x \text{ annimmt.}$$

Wir lesen also z.B. aus obiger Verteilungsfunktion ab, dass die rel. Häufigkeit, mit der unser Merkmal Werte ≤ 3.5 angenommen hat, 0.3 (oder 30%) beträgt. Mit rel. Häufigkeit von 0.7 (oder 70%) wurden Werte > 3.5 angenommen.

Jede empirische Verteilungsfunktion F_h^* hat folgende Eigenschaften:

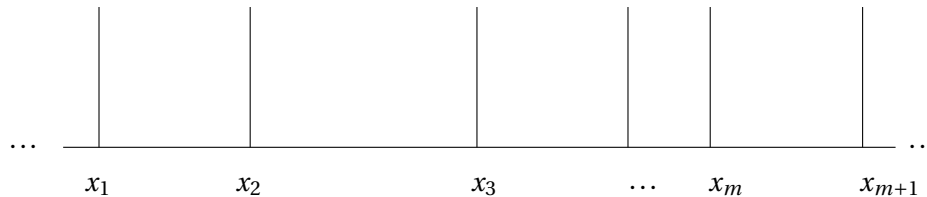
- i) F_h^* ist monoton wachsend (i.A. jedoch nicht streng monoton wachsend)
- ii) $\lim_{x \rightarrow -\infty} F_h^*(x) = 0$
- iii) $\lim_{x \rightarrow \infty} F_h^*(x) = 1$
- iv) F_h^* ist rechtsseitig stetig, d.h. $\lim_{x \downarrow x_0} F_h^*(x) = F_h^*(x_0)$ (Das ist vor allem in Sprungstellen relevant. In "Nichtsprungstellen" ist F_h^* sogar stetig.)

Abschliessend möchte ich nochmal erwähnen, dass sowohl Häufigkeitsfunktion als auch emp.

Verteilungsfunktion die Verteilung der aufgetretenen Ausprägungen eines Merkmals eindeutig beschreiben.

1.4 Häufigkeitsdichte und Histogramm

Oft werden Merkmalsausprägungen eines Merkmals X zu *Grössenklassen* oder *Schichten* zusammengefasst. Dazu wird das von den möglichen Merkmalsausprägungen belegte reelle Intervall durch geeignet gewählte Klassengrenzen x_1, x_2, \dots, x_{m+1} in m Klassen K_1, K_2, \dots, K_m zerlegt.



Für die m Grössenklassen definieren wir die *Klassenbreite* wie folgt:

$$\Delta_i := x_{i+1} - x_i, \quad i = 1, \dots, m.$$

Für jede der m Grössenklassen definieren wir desweiteren ihre relative Häufigkeit

$$n_i^* := \text{relH}(x_i \leq X < x_{i+1}).$$

Diese berechnet sich wie üblich aus der Anzahl der Beobachtungswerte in der entsprechenden Grössenklasse geteilt durch die Gesamtzahl der Beobachtungen.

Diese relativen Häufigkeiten werden wir den Klassenmitten $M_i = \frac{x_{i+1} + x_i}{2}$ zuordnen (Man könnte sie z.B. auch den rechten oder linken Klassengrenzen zuordnen.), sodass sich folgende (relative) Häufigkeitstabelle ergibt:

M_1	M_2	\dots	M_m
n_1^*	n_2^*	\dots	n_m^*

Mittels dieser Tabelle kann die Verteilungsfunktion $F_K^*(x)$ im Kontext klassierter Daten definiert werden.

$$F_K^*(x) = \begin{cases} 0 & \text{für } x < M_1 \\ \sum_{M_i \leq x} n_i^* & \text{sonst} \end{cases}$$

Bemerkung 1.15

Die Klassenmitten M_i übernehmen also die Rolle der Ausprägungen a_i . Die Sprungstellen im

Graphen der Funktion $F_K^*(x)$ sind jeweils in den Klassenmitten.

Beispiel 1.16: Klassierte Daten

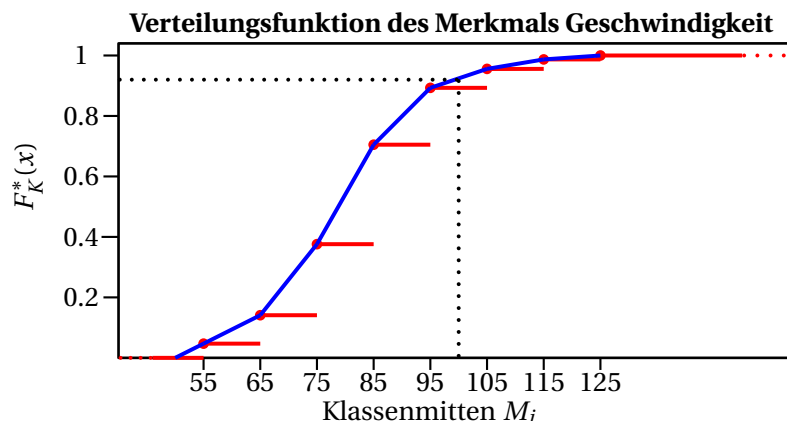
Bei einer Geschwindigkeitsmessung an einem Strassenquerschnitt wurden folgende Geschwindigkeiten (in km/h) ermittelt:

52, 57, 58, 61, 63, 65, 65, 68, 69, 70.5, 71, 71, 72.5, 73, 73, 73.5, 74, 75, 75, 77, 78, 78, 78, 79, 80.5, 80.5, 82, 82, 82, 82.5, 83, 84, 84, 85, 85, 85, 87, 87, 87.5, 88, 88.5, 88.5, 89.5, 89.5, 89.5, 91, 93, 95, 95, 97, 98, 98, 99, 99, 99, 99.5, 103, 103.5, 108, 110, 115, 116, 122

Wir haben also $n = 64$ Messwerte. Die Geschwindigkeiten erstrecken sich in einem Intervall von ungefähr $[50, 130]$. Die Wahl der Grössenklassen ist situations- und problembedingt. In unserem Beispiel wählen wir 8 Klassen mit Breite 10. Die Information der (geordneten) Beobachtungsreihe bereiten wir tabellarisch auf.

Nr. der Klasse	Geschw.-Intervall	absH	relH	$F_K^*(M_i)$, $i = 1, \dots, 8$
1	50-60	3	0.047	0.047
2	60-70	6	0.094	0.141 (=0.047+0.094)
3	70-80	15	0.234	0.375 (=0.047+0.094+0.234)
4	80-90	21	0.328	0.703
5	90-100	12	0.188	0.891
6	100-110	4	0.062	0.953
7	110-120	2	0.031	0.984
8	120-130	1	0.016	1.000

Daraus ergibt sich die folgende Verteilungsfunktion (roter Graph):



Manchmal wird die (unstetige) Treppenfunktion $F_K^*(x)$ stetig approximiert. Hierfür gibt es verschiedene Möglichkeiten. Ich möchte die Approximation durch Polygonzug erwähnen. Wie mittels Polygonzug approximiert wird, ist aus der obigen Graphik ([blauer Polygonzug](#)) ersichtlich. Der Anstieg des Geradenstücks im Intervall $[M_i, M_{i+1})$, $i = 1, \dots, m-1$ (m =Anzahl Klassen) berechnet sich dabei wie folgt:

$$\frac{F_K^*(x_{i+1}) - F_K^*(x_i)}{\Delta_i} = \frac{n_i^*}{\Delta_i}.$$

Er wird auch als *durchschnittliche Häufigkeitsdichte der i -ten Grössenklasse* bezeichnet.

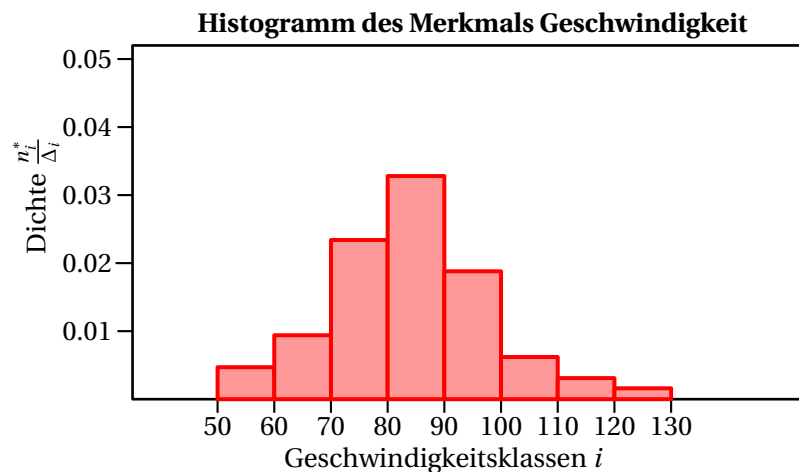
Will man jetzt z.B. wissen, wieviel Prozent der Messungen ≤ 100 km/h waren, kann man das approximativ an der (durch Polygonzug approximierten) Verteilungsfunktion ablesen. Man kommt auf rund 92%.

Bemerkung 1.17: Stetige Approximation der Verteilungsfunktion bei klassierten Daten

Durch das Klassieren verliert man die Information über die Häufigkeitsverteilung innerhalb einer Grössenklasse. Die Approximation durch Polygonzug suggeriert eine gleichmässige Häufigkeitsverteilung innerhalb der Grössenklassen. Das ist oft nicht realistisch.

Die Verteilung der Merkmalsausprägungen im Kontext gruppierter Daten wird häufig durch *Histogramme* graphisch dargestellt. Das sind Säulendiagramme (Balkendiagramme) mit Säulen über den Grössenklassen mit einer Höhe, die so gewählt wird, dass der Flächeninhalt der Säule proportional zu $\frac{n_i}{\Delta_i}$ oder $\frac{n_i^*}{\Delta_i}$ ist. Falls alle Klassenbreiten Δ_i **gleich** sind, kann man also z.B. n_i oder auch n_i^* als Höhe über der Klasse K_i wählen.

Was immer funktioniert, also auch im Fall von unterschiedlichen Klassenbreiten, ist die Wahl der Säulenhöhe als n_i^*/Δ_i . Der Flächeninhalt der Säule über der i -ten Grössenklasse ist dann ganz einfach die relative Häufigkeit n_i^* dieser Grössenklasse. Mit dieser Wahl ergibt sich für unser Geschwindigkeitsbeispiel das folgende Histogramm. (Säulenhöhen sind 0.0047, 0.0094, 0.0234, 0.0328, 0.0188, 0.0062, 0.0031 und 0.0016).



Eine andere Möglichkeit des Umgangs mit unterschiedlichen Klassenbreiten ist das Arbeiten mit einer *Normklassenbreite* (s. Lehrbuch).

2 Kenngrößen einer Stichprobe

2.1 Lagekennwerte

2.1.1 Arithmetisches Mittel

Es gibt verschiedene Wege, um das *arithmetische Mittel* oder den *Mittelwert* \bar{x} einer statistischen Variablen zu bestimmen: (Macht euch klar, worauf diese unterschiedlichen Versionen beruhen!)

Geg.: Beobachtungsreihe x_1, x_2, \dots, x_n vom Umfang n bzw. die unterschiedlichen Ausprägungen a_1, a_2, \dots, a_k mit abs. und rel. Häufigkeiten n_1, n_2, \dots, n_k bzw. $n_1^*, n_2^*, \dots, n_k^*$.

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

$$2) \quad \bar{x} = \frac{1}{n} (n_1 a_1 + n_2 a_2 + \dots + n_k a_k) = \frac{1}{n} \sum_{i=1}^k n_i a_i \quad \text{(Gewichtetes arithm. Mittel)}$$

$$3) \quad \bar{x} = (n_1^* a_1 + n_2^* a_2 + \dots + n_k^* a_k) = \sum_{i=1}^k n_i^* a_i$$

In unserem ersten Beispiel führt das zu:

- 1) Geg.: Beobachtungsreihe 5.0, 1.6, 3.0, 4.1, 1.6, 3.0, 4.1, 4.1, 5.0, 5.0, 4.1, 5.0, 3.0, 4.1, 3.0, 5.0, 4.1, 5.0, 4.1, 4.1 vom Umfang $n = 20$.

$$\bar{x} = \frac{1}{20}(5.0 + 1.6 + \dots + 4.1) = 3.9$$

- 2) Geg.: aufgetretene Ausprägungen 1.6, 3.0, 4.1, 5.0 mit abs. Häufigkeiten 2, 4, 8, 6.

$$\bar{x} = \frac{1}{20}(2 \cdot 1.6 + 4 \cdot 3.0 + 8 \cdot 4.1 + 6 \cdot 5.0) = 3.9$$

- 3) Geg.: aufgetretene Ausprägungen 1.6, 3.0, 4.1, 5.0 mit rel. Häufigkeiten 0.1, 0.2, 0.4, 0.3.

$$\bar{x} = (0.1 \cdot 1.6 + 0.2 \cdot 3.0 + 0.4 \cdot 4.1 + 0.3 \cdot 5.0) = 3.9$$

Für unsere Geschwindigkeitsliste erhalten wir

$$\bar{x} = \frac{1}{64}(52 + 57 + 58 + \dots + 122) = 84.36.$$

Bei der Mittelwertberechnung für klassierte Daten geht man wie folgt vor. Sind die Klassenmittelwerte \bar{x}_i für jede der m Grössenklasse bekannt, dann ergibt sich der exakte Mittelwert durch:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_i \bar{x}_i = \sum_{i=1}^m n_i^* \bar{x}_i. \quad (1)$$

Hier bezeichnen n_i und n_i^* die absoluten bzw. relativen Klassenhäufigkeiten.

Sind die Klassenmittelwerte nicht bekannt, verwendet man die Klassenmitten $M_i = \frac{x_{i+1} + x_i}{2}$ und berechnet geschätzte Mittelwerte mit den Formeln (1), in denen \bar{x}_i durch M_i ersetzt wird. Es werden also sozusagen die (unbekannten) Beobachtungswerte in K_i jeweils durch M_i ersetzt.

Beispiel 2.1: Arithmetisches Mittel für klassierte Daten

Für unser Geschwindigkeitsbeispiel sind die Klassenmitten gegeben durch $M_1 = 55$, $M_2 = 65, \dots$, $M_8 = 125$. Mit den schon berechneten rel. Klassenhäufigkeiten erhalten wir den geschätzten Mittelwert

$$\begin{aligned} \bar{x} &= 0.047 \cdot 55 + 0.094 \cdot 65 + 0.234 \cdot 75 + 0.328 \cdot 85 + 0.188 \cdot 95 + 0.062 \cdot 105 + 0.031 \cdot 115 + 0.016 \cdot 125 \\ &= 84.06. \end{aligned}$$

2.1.2 Geometrisches Mittel

Das geometrische Mittel dient zur Berechnung des Durchschnittswertes bei verhältnisskalierten Merkmalen, deren Merkmalsausprägungen multiplikativ miteinander verbunden sind. Wachstumsraten sind hierfür gute Beispiele.

Sei x_1, x_2, \dots, x_n eine Beobachtungsreihe mit $x_i \geq 0$. Dann wird das geometrische Mittel folgendermassen definiert:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}.$$

Nehmen Sie z.B. an, Sie besässen eine Aktie im Wert von 100CHF, deren Wert im ersten Jahr auf 150CHF steigt und im zweiten Jahr wieder auf ihren Ursprungswert von 100CHF fällt. Was ist die durchschnittliche Rendite? Gesunder Menschenverstand sagt uns 0%.

Es gilt: $100 \cdot 1.5 = 150$, $100 = 150 \cdot \frac{2}{3}$. Ihr Anfangswert (100CHF) und ihr Endwert (ebenfalls 100CHF) sind also multiplikativ über die Renditefaktoren 1.5 und $\frac{2}{3}$ miteinander verbunden. Berechnen wir das arithmetische Mittel der Renditefaktoren, erhalten wir $\bar{x} = (1.5 + \frac{2}{3})/2 \approx 1.083$. Das entspricht einer durchschnittlichen Rendite von rund 8.3%. Das spiegelt also nicht die Realität wider. Versuchen wir es mit dem geometrischen Mittel: $\bar{x}_g = \sqrt{1.5 \cdot \frac{2}{3}} = 1$. Das entspricht einer durchschnittlichen Rendite von 0%.

Ein anderes

Beispiel 2.2: Geometrisches Mittel

Renditen von Anlagen mit wechselnden Zinssätzen. Ein Kapital K_n entstehe in n Zinsperioden aus einem Kapital K_0 mittels wechselnder Aufzinsfaktoren q_1, q_2, \dots, q_n , also $K_n = K_0 q_1 \cdot q_2 \cdot \dots \cdot q_n$. Was ist der Effektivzinssatz p_{eff} ? Dazu muss man

$$K_0 q_{\text{eff}}^n = K_0 \left(1 + \frac{p_{\text{eff}}}{100}\right)^n = K_0 q_1 \cdot q_2 \cdot \dots \cdot q_n$$

lösen. Es ergibt sich $q_{\text{eff}} = \sqrt[n]{q_1 \cdot q_2 \cdot \dots \cdot q_n}$ und damit $p_{\text{eff}} = (q_{\text{eff}} - 1) \cdot 100\%$.

2.1.3 Harmonisches Mittel

Das harmonische Mittel ist das Mittel der Wahl für Verhältniszahlen (Quotient zweier Grössen wie km/h, CHF/Woche etc.), wenn Zusatzinformationen gegeben sind, die sich inhaltlich auf den Zähler beziehen. (Beziehen sie sich auf den Nenner, dann kommt das arithmetische Mittel zum Einsatz.) Das harmonische Mittel ist nur für positive Werte $x_i > 0$ wie folgt definiert:

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

Das harmonische Mittel ist also der Kehrwert des arithmetischen Mittels der Kehrwerte der Merkmalswerte.

Sucht man nach Beispielen für die Anwendung des harmonischen Mittels, wird man unweigerlich auf das Problem der Durchschnittsgeschwindigkeiten stossen. Ich möchte Ihnen das natürlich auch nicht vorenthalten.

Stellen Sie sich vor, Sie fahren mit dem Velo 1km mit Geschwindigkeit $x_1 = 10 \frac{\text{km}}{\text{h}}$ und einen weiteren Kilometer mit $x_2 = 20 \frac{\text{km}}{\text{h}}$. Was war Ihre Durchschnittsgeschwindigkeit?

Versuchen wir es physikalisch. Sie haben $s = 2\text{km}$ zurückgelegt und haben dafür $t = 0.15\text{h}$ benötigt. Es gilt also

$$\bar{v} = \frac{s}{t} = \frac{2\text{km}}{0.15\text{h}} = 13.\bar{3} \frac{\text{km}}{\text{h}}.$$

Wenden wir das arithmetische Mittel an, erhalten wir

$$\bar{x} = \frac{x_1 + x_2}{2} = \frac{10 + 20}{2} \frac{\text{km}}{\text{h}} = 15 \frac{\text{km}}{\text{h}} \neq \bar{v}.$$

Das funktioniert also nicht. Erinnern Sie sich bitte an das eingangs Gesagte. Wenn man Verhältniszahlen mitteln will, für die zählerbezogene Zusatzinformationen gegeben sind, wird das harmonische Mittel verwendet. Das ist hier aber der Fall. Wir haben Zusatzinformationen gegeben, die sich auf den Zähler beziehen, nämlich die mit den jeweiligen Geschwindigkeiten gefahrenen Strecken. Also müssen wir rechnen:

$$\bar{x}_h = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} = \frac{2}{\frac{1}{10 \frac{\text{km}}{\text{h}}} + \frac{1}{20 \frac{\text{km}}{\text{h}}}} = \frac{2}{\frac{3}{20 \frac{\text{km}}{\text{h}}}} = 13.\bar{3} \frac{\text{km}}{\text{h}} = \bar{v}.$$

Um Ihnen den Unterschied zu nennerbezogenen Zusatzinformationen zu verdeutlichen, nehmen Sie jetzt an, Sie seien 1h mit $x_1 = 10 \frac{\text{km}}{\text{h}}$ und eine weitere Stunde mit $x_2 = 20 \frac{\text{km}}{\text{h}}$ gefahren. Was war jetzt Ihre Durchschnittsgeschwindigkeit?

Physikalisch: Sie sind $s = 30\text{km}$ in $t = 2\text{h}$ gefahren. Ihre Durchschnittsgeschwindigkeit betrug also $\bar{v} = 15 \frac{\text{km}}{\text{h}} = \bar{x}$.

Für die Praxis wesentlich wichtiger ist das gewichtete harmonische Mittel. Sind den x_i positive Gewichte g_i zugeordnet, so berechnet man das gewichtete harmonische Mittel wie folgt:

$$\bar{x}_h = \frac{g_1 + g_2 + \dots + g_n}{\frac{g_1}{x_1} + \frac{g_2}{x_2} + \dots + \frac{g_n}{x_n}} = \frac{\sum_{i=1}^n g_i}{\sum_{i=1}^n \frac{g_i}{x_i}}.$$

Sind alle Gewichte gleich, d.h. $g_1 = g_2 = \dots = g_n$, erhält man als Spezialfall das (gewöhnliche) harmonische Mittel.

Beispiel 2.3: Gewichtetes harmonisches Mittel

4 Studierende mit Nebenjobs verdienen 18, 20, 15 bzw. 19 CHF/h. (Situation: Verhältniszahlen!). Ihre Gesamtverdienste pro Woche beliefen sich auf 180, 300, 270 bzw. 380 CHF (d.h. zählerbezogene Zusatzinformation!). Wie hoch ist der DurchschnittsStundenlohn?

$$\bar{x}_h = \frac{180 + 300 + 270 + 380}{\frac{180}{18} + \frac{300}{20} + \frac{270}{15} + \frac{380}{19}} = 17.94.$$

Wir wollen verifizieren, dass das in der Tat die richtige Antwort ist.

Der Gesamtwochenverdienst der Studierenden beträgt 1130 CHF. Gemeinsam haben sie dafür $10 + 15 + 18 + 20 = 63$ h gearbeitet. Der Durchschnittslohn beläuft sich also auf $1130/63$ CHF/h = 17.94 CHF/h.

Bemerkungen 2.4

- i) Ist einer (oder mehrere) der Beobachtungswerte x_i gleich 0, wird \bar{x}_h häufig 0 gesetzt.
- ii) Sind x_{\min} und x_{\max} der kleinste bzw. grösste Beobachtungswert in der Beobachtungsreihe, so gilt immer:

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

Alle unsere Mittelwerte liegen also stets zwischen x_{\min} und x_{\max} .

Das gilt auch für

2.1.4 Modus und Median

Der *Modus* oder *Modalwert* \bar{x}_{Mod} einer statistischen Verteilung ist (sind) der (die) am häufigsten auftretende(n) Merkmalswert(e). Für unser erstes Beispiel haben wir daher $\bar{x}_{\text{Mod}} = 4.1$. Das zweite Beispiel hat 5 Modi, nämlich 78, 82, 85, 89.5 und 99, die alle jeweils dreimal auftreten.

Der Modus bezieht sich also immer auf die Beobachtungsreihe und ist immer eine Ausprägung. Gibt es einen einzigen Modus, spricht man von einer *unimodalen* Verteilung, andernfalls von *bimodaler*, *trimodaler*, ..., *multimodaler* Verteilung.

Der *Median* oder *Zentralwert* \bar{x}_{Med} teilt die **geordnete** Beobachtungsreihe in zwei Hälften. Dazu definieren wir:

$$\bar{x}_{\text{Med}} = \begin{cases} x_{\frac{n+1}{2}} & , \text{ falls } n \text{ ungerade} \\ \frac{1}{2} (x_{\frac{n}{2}} + x_{(\frac{n}{2}+1)}) & , \text{ falls } n \text{ gerade} \end{cases}$$

n ist hier natürlich wieder der Umfang der Beobachtungsreihe.

Vorteil Median gegenüber arithmetischem Mittel: Robust gegenüber “Ausreissern”. Vergleichen Sie die Beobachtungsreihen 1,2,3 und 1,2,1000. Die 1000 passt grössenordnungsmässig nicht zum “übrigen” Datenmaterial. Bestimmen Sie jeweils Median und arithmetisches Mittel. Was beobachten Sie? Darüber hinaus kann der Median schon für ordinalskalierte Merkmale angewandt werden, das arithmetische Mittel (eigentlich) nur für metrische Merkmale.

Die Beobachtungsreihen beider unserer Beispiele haben n gerade. Im ersten Fall ist

$$\bar{x}_{\text{Med}} = \frac{1}{2}(x_{\frac{20}{2}} + x_{(\frac{20}{2}+1)}) = \frac{1}{2}(x_{10} + x_{11}) = \frac{1}{2}(4.1 + 4.1) = 4.1.$$

Für das Geschwindigkeitsbeispiel erhalten wir

$$\bar{x}_{\text{Med}} = \frac{1}{2}(x_{\frac{64}{2}} + x_{(\frac{64}{2}+1)}) = \frac{1}{2}(x_{32} + x_{33}) = \frac{1}{2}(84 + 84) = 84.$$

2.2 Quantile

Der Median “zerlegt” eine geordnete Beobachtungsreihe in 2 Hälften. Wir wollen das verallgemeinern.

Betrachten wir eine **geordnete** Beobachtungsreihe vom Umfang n :

$$x_{[1]} \leq x_{[2]} \leq x_{[3]} \leq \dots \leq x_{[n]}$$

Definition 2.5: Quantil

Sei p eine reelle Zahl zwischen 0 und 1 (d.h. $0 < p < 1$). Eine reelle Zahl $x_{Q[p]}$ heisst p -Quantil (oder $(p \cdot 100\%)$ -Quantil) der Beobachtungsreihe, falls mindestens $p \cdot 100\%$ der Beobachtungswerte $\leq x_{Q[p]}$ und mindestens $(1 - p) \cdot 100\%$ der Beobachtungswerte $\geq x_{Q[p]}$ sind.

Zur Sprechweise: Sei p z.B. 0.75. Dann spricht man vom 0.75-Quantil bzw. $(0.75 \cdot 100\%) = 75\%$ -Quantil. Ist $x_{Q[p]}$ das 75%-Quantil, dann sind mindestens 75% der Beobachtungswerte $\leq x_{Q[p]}$ und mindestens $(1 - p) \cdot 100\% = 25\%$ (also der Komplementanteil) der Beobachtungswerte $\geq x_{Q[p]}$. $x_{Q[p]}$ “zerlegt” also die Beobachtungsreihe im Verhältnis 3:1.

Die *Gauss-Klammer*. Wir definieren

$$[x] := \text{kleinste ganze Zahl } \geq x, \text{ d.h. } x \text{ wird auf die nächste ganze Zahl aufgerundet.}$$

Also gilt z.B. $[2.3] = 3$, $[-2.7] = -2$ und $[5] = 5$.

Man hat dann:

$$x_{Q[p]} = \begin{cases} x_{[np]} & , \text{ falls } np \text{ keine ganze Zahl ist;} \\ \frac{1}{2}(x_{np} + x_{np+1}) & , \text{ falls } np \text{ eine ganze Zahl ist.} \end{cases}$$

Beispiel 2.6: Quantil

Wir verwenden begleitend die folgende Situation. 17 FH-Absolventen wurden nach ihrer Studiendauer befragt. Dabei ergaben sich die unten dargestellten Antworten:

Semester	7	9	10	11	12	13
Anz. Nennungen	1	1	2	2	5	6

Die geordnete Beobachtungsreihe sieht also wie folgt aus:

$x_{[1]}$	$x_{[2]}$	$x_{[3]}$	$x_{[4]}$	$x_{[5]}$	$x_{[6]}$	$x_{[7]}$	$x_{[8]}$	$x_{[9]}$	$x_{[10]}$	$x_{[11]}$	$x_{[12]}$	$x_{[13]}$	$x_{[14]}$	$x_{[15]}$	$x_{[16]}$	$x_{[17]}$
7	9	10	10	11	11	12	12	12	12	12	13	13	13	13	13	13

Versuchen wir, das 0.25-Quantil, 0.5-Quantil und 0.75-Quantil zu bestimmen.

$$np = 0.25 \cdot 17 = 4.25 \notin \mathbb{Z} \implies x_{Q[0.25]} = x_{[4.25]} = x_5 = 11$$

$$np = 0.5 \cdot 17 = 8.5 \notin \mathbb{Z} \implies x_{Q[0.5]} = x_{[8.5]} = x_9 = 12$$

$$np = 0.75 \cdot 17 = 12.75 \notin \mathbb{Z} \implies x_{Q[0.75]} = x_{[12.75]} = x_{13} = 13$$

Versuchen wir auch nochmal, anhand des 0.75-Quantils die Definition zu überprüfen. 17 Beobachtungswerte (nämlich $x_{[1]}, \dots, x_{[17]}$) sind kleiner gleich $x_{Q[0.75]} = x_{[13]}$. Das entspricht aber 100% der Beobachtungswerte. 6 Beobachtungswerte (nämlich $x_{[12]}, \dots, x_{[17]}$) sind grösser gleich $x_{Q[0.75]} = x_{[13]}$. Das entspricht 35.3%. Es sind also sowohl mindestens 75% der Beobachtungswerte $\leq x_{Q[0.75]}$ als auch mindestens 25% der Beobachtungswerte $\geq x_{Q[0.75]}$, wie gefordert. Dass die Prozentwerte jeweils grösser als die geforderten Werte sind, liegt daran, dass die Ausprägung 13 mehrfach auftritt und in beiden Gruppen liegt.

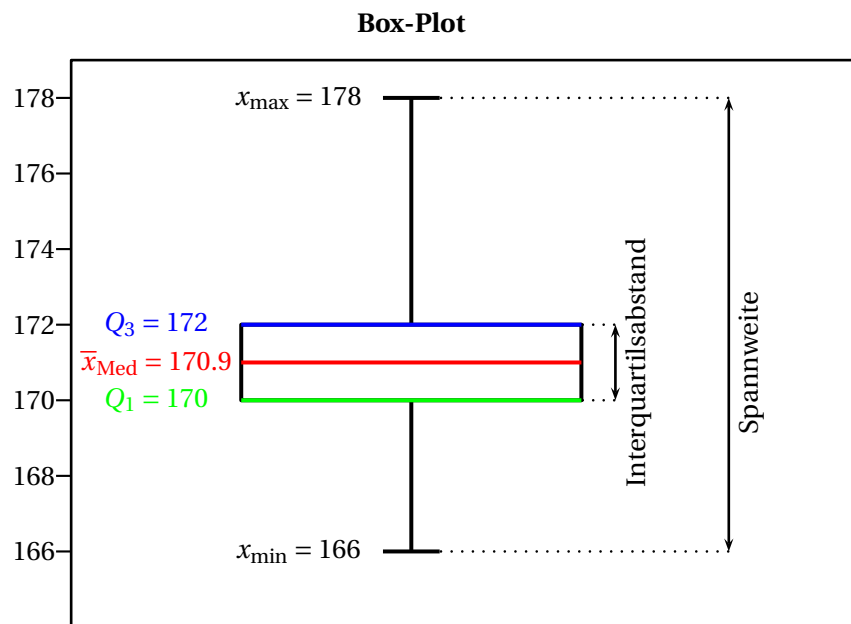
0.25-Quantil, 0.5-Quantil und 0.75-Quantil werden auch mit Q_1 , Q_2 und Q_3 bezeichnet und 1., 2. bzw. 3. *Quantil* genannt. Machen Sie sich auch klar, dass Q_2 nichts anderes als der Median \bar{x}_{Med} ist.

Ausser den 25%-, 50%- und 75%-Quantilen, bekommen häufig auch die 10%-, 20%-, ..., 90%-Quantile und 1%-, 2%-, ..., 99%-Quantile eigene Namen. Sie heissen *Dezile* bzw. *Perzentile*. Anstatt vom 60%-Quantil spricht man also auch vom 6. Dezil.

Q_1 , $Q_2 = \bar{x}_{\text{Med}}$ und Q_3 werden zusammen mit dem kleinsten Beobachtungswert x_{\min} und grösstem Beobachtungswert x_{\max} häufig in einem *Box-Plot* veranschaulicht.

Beispiel 2.7: Box-Plot

Das Datenmaterial bezieht sich auf die Körpergrössen einer Schulklasse von 18-jährigen Mädchen. Obwohl wir die Einzeldaten nicht kennen, verschafft uns der Box-Plot einen relativ guten Überblick über die Verteilung der Körpergrössen.



Die geringste gemessene Grösse war 166cm, die Maximalgrösse 178cm. Ca. 50% der Messungen lagen zwischen $Q_1 = 170\text{cm}$ und $Q_3 = 172\text{cm}$. Ca. 50% der Mädchen hatten Körpergrössen von $\geq 170.9\text{cm}$, die anderen 50% kleinere als dieser Wert. etc.

Bemerkungen 2.8

- i) Für die Begriffe "Spannweite" und "Interquartilsabstand" siehe Abschnitt 2.3.
- ii) Die meisten Statistikprogramme kennzeichnen "Ausreisser" separat. Das geschieht meist, wenn Werte mehr als 1.5 IQA vom Median entfernt liegen.

2.3 Streuungskenngrößen

Betrachten Sie die Beobachtungsreihen $-0.1, 0, 0.1$ und $-1000000, 0, 1000000$. Beide Beobachtungsreihen haben Mittelwert 0. Bei der zweiten streuen die Werte jedoch wesentlich stärker um diesen Mittelwert als bei der ersten. In diesem Abschnitt wollen wir versuchen, das Streuungsverhalten zu quantifizieren.

2.3.1 Spannweite

Die *Spannweite* einer Beobachtungsreihe ist ganz einfach die Differenz aus grösstem und kleinstem Beobachtungswert: $R := \max x_i - \min x_i$.

In unserem Studiendauerbeispiel beträgt die Spannweite also $13 - 7 = 6$.

2.3.2 Empirische Varianz

Wie das arithmetische Mittel als Lagemass spielt die *empirische Varianz* bei metrischen Merkmalen eine herausragende Rolle als Streuungsmass. Sie ist wie folgt definiert:

$$\begin{aligned} s^2 &= \frac{1}{n} \left((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k n_i (a_i - \bar{x})^2 \\ &= \sum_{i=1}^k n_i^* (x_i - \bar{x})^2 \quad (n_i \text{ abs. H., } n_i^* \text{ rel. H.}). \end{aligned}$$

Für unser Studiendauerbeispiel erhalten wir:

$$\begin{aligned} s^2 &= \frac{1}{17} \left((7 - 11.53)^2 + (9 - 11.53)^2 + 2 \cdot (10 - 11.53)^2 + 2 \cdot (11 - 11.53)^2 \right. \\ &\quad \left. + 5 \cdot (12 - 11.53)^2 + 6 \cdot (13 - 11.53)^2 \right) = 2.719. \end{aligned}$$

Effizientere Berechnung der Varianz:

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^k n_i a_i^2 - \bar{x}^2$$

In Worten: Bilde das arithmetische Mittel der quadrierten Beobachtungswerte und subtrahiere davon das Quadrat des arithmetischen Mittels. Das wird häufig kompakt geschrieben als: $s^2 = \overline{x^2} - \bar{x}^2$.

Nachteil Varianz: Bei Ausprägungen mit Einheiten erscheinen diese in der Varianz quadriert.

Abhilfe: Quadratwurzel aus Varianz = *Standardabweichung* $s = \sqrt{s^2}$.

Für unser Beispiel gilt also $s = 1.649$.

Bemerkung 2.9

Varianz und Standardabweichung einer Beobachtungsreihe sind Zahlen, die für sich relativ wenig aussagen oder interpretierbar sind, wenn man sie nicht auf etwas “beziehen” kann. Eine gute Bezugsgrösse ist der Mittelwert der Beobachtungsreihe zu dem die Varianz/Standardabweichung oft ins Verhältnis gesetzt wird.

Um die Streuung in Beobachtungsreihen miteinander vergleichen zu können, wird meist der *Variationskoeffizient* verwendet. Er misst die Streuung relativ zur absoluten Grössenordnung der Beobachtungsreihe:

$$VK = \frac{\text{Standardabweichung}}{\text{Absolutbetrag des arithmetischen Mittels}} = \frac{s}{|\bar{x}|}.$$

Berechnet man den VK für das Beispiel aus diesem Abschnitt, erhält man rund 0.143. Je grösser der VK, desto grösser die relative Streuung um das Mittel.

2.3.3 Interquartilsabstand

Der *Interquartilsabstand* d_Q ist nichts anderes als die Differenz zwischen 3. und 1. Quartil:

$$d_Q = Q_3 - Q_1 = x_{Q[0.75]} - x_{Q[0.25]}$$

Er findet vor allem bei ordinalskalierten Merkmalen Anwendung.

3 Lineare Korrelation

Bisher haben wir nur die Verteilung eines Merkmals in einer Population betrachtet. Häufig ist man jedoch an mehr als einem Merkmal interessiert. Diese können entweder einzeln oder in ihren Beziehungen zueinander untersucht werden. Letzteres führt zu mehrdimensionalen Verteilungen. Im folgenden beschränken wir uns jedoch auf zwei Merkmale.

3.1 Streudiagramm

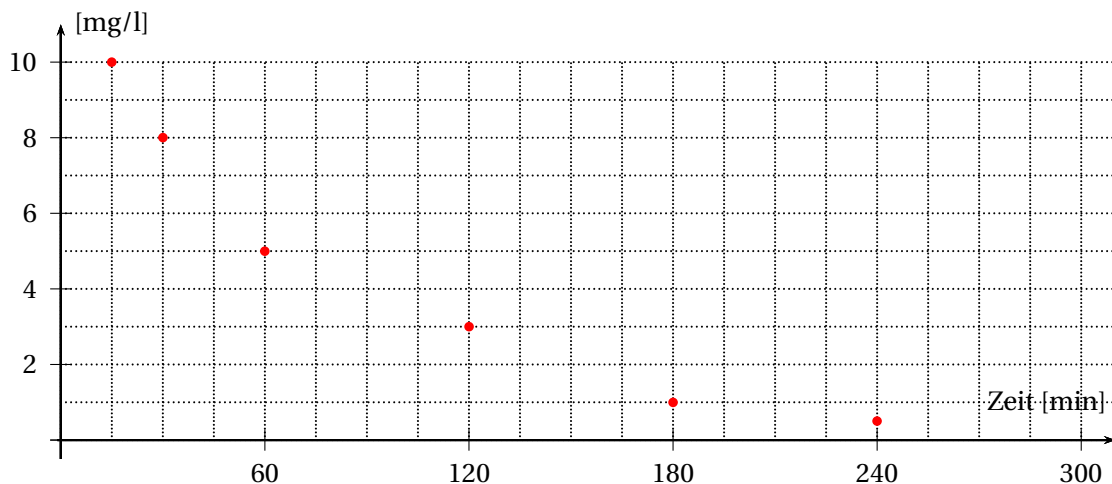
Wir betrachten zwei Merkmale X und Y und ihre Ausprägungen in den statistischen Einheiten e_i einer Grundgesamtheit. Das liefert eine Menge von Wertepaaren $(x_i, y_i) = (X(e_i), Y(e_i))$, d.h. jeder statistischen Einheit werden die Ausprägungen bzgl. der 2 Merkmale zugeordnet. Sind beide Merkmale quantitativ, kann man diese Wertepaare in einem rechtwinkligen Koordinatensystem gegeneinander abtragen und erhält ein sogenanntes *Streudiagramm*.

Beispiel 3.1: Streudiagramm

Ein Medikament besteht aus einem Arzneistoff, welcher ab einer gewissen Konzentration im Blut arzneilich wirksam wird und sogenannten pharmazeutischen Hilfsstoffen, welche den Arzneistoff umhüllen. Wird ein Medikament oral eingenommen, wird es vom Körper metabolisiert (abgebaut) und später ausgeschieden. Für die richtige Dosierung eines Medikaments ist es daher von entscheidender Bedeutung, die Konzentration im Blut zu jedem Zeitpunkt nach der Einnahme zu kennen, um die Dauer der Wirksamkeit vorhersagen zu können. Die Tabelle stellt die Konzentration des Arzneistoffs im Blut in mg/l dar, nachdem zum Zeitpunkt Null 50mg des im Medikament *Nixdrin forte* enthaltenen Arzneistoffs oral eingenommen wurde:

Zeit in min nach Einnahme	15	30	60	120	180	240
Konzentration in mg/l	10	8	5	3	1	0.5

Daraus resultiert das folgende Streudiagramm:



3.2 Kennzahl für das Ausmass des linearen Zusammenhangs

In Streudiagrammen hat man oft den Eindruck, dass die Datenpunkte mehr oder weniger auf einer Geraden liegen, also zumindest näherungsweise ein linearer Zusammenhang zwischen den Merkmalen vorliegt. Dieser Zusammenhang muss jedoch keineswegs kausal sein! Wir wollen versuchen, das Ausmass des linearen Zusammenhangs zu charakterisieren. Dazu definieren wir zu-

nächst einmal die *empirische Kovarianz* s_{XY} zweier metrischer Merkmale X und Y .

Definition 3.2: Empirische Kovarianz

Die empirische Kovarianz wird definiert durch

$$s_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \left(= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y} \right).$$

Hier bezeichnen wie üblich \bar{x} und \bar{y} die arithmetischen Mittel der Merkmale X bzw. Y .

Die Formel in Klammern stellt wieder eine effizientere Möglichkeit der Kovarianzberechnung dar. Bilde das arithmetische Mittel der Produkte der Wertepaarkomponenten und subtrahiere das Produkt der arithmetischen Mittel. Das schreibt man auch suggestiv als: $s_{XY} = \overline{xy} - \bar{x} \cdot \bar{y}$.

Wie bei der Varianz eines Merkmals sagt die Grösse der Kovarianz zweier Merkmale allein relativ wenig aus. Interessanter ist hier eher das Vorzeichen. Bei positiver Kovarianz besteht tendenziell eine gewisse "Gleichläufigkeit" in der Paarung der Ausprägungen der Merkmale, d.h. grosse (kleine) Ausprägungen bzgl. X treten in der Regel mit grossen (kleinen) Ausprägungen bzgl. Y auf, bei negativer Kovarianz herrscht tendenziell "Gegenläufigkeit".

Beispiel 3.3: Kovarianz

In einer Monokultur von 380 Hartholzbäumen wurde zu Kontrollzwecken der Stammdurchmesser D (in m) in Brusthöhe sowie die nutzbare Höhe H (in m) für jeden Baum ermittelt. Da es sich um eine grobe Erhebung handelte, wurden Stammdurchmesser nur auf $0.25m$ und Höhen auf $5m$ genau bestimmt. Dabei fand man z.B. 61 Bäume mit Stammdurchmesser $1m$ und nutzbarer Höhe von $20m$. Die restlichen Werte sind in der sog. *Kontingenztafel* aufgeführt.

	1.0	1.25	1.5	1.75	
20	61	57	11	0	129
25	34	114	65	38	251
	95	171	76	38	380

Man berechnet leicht das Mittel der Stammdurchmesser $\bar{d} = 1.2875$ und der Baumhöhen $\bar{h} = 23.3$. Das ergibt für die Kovarianz:

$$\begin{aligned} s_{HD} &= \frac{1}{380} [61 \cdot (20 - 23.3)(1.0 - 1.2875) + 34 \cdot (25 - 23.3)(1.0 - 1.2875) \\ &\quad + 57 \cdot (20 - 23.3)(1.25 - 1.2875) + 114 \cdot (25 - 23.3)(1.25 - 1.2875) \\ &\quad + 11 \cdot (20 - 23.3)(1.5 - 1.2875) + 65 \cdot (25 - 23.3)(1.5 - 1.2875) \\ &\quad + 0 \cdot (20 - 23.3)(1.75 - 1.2875) + 38 \cdot (25 - 23.3)(1.75 - 1.2875)] \approx 0.225 \end{aligned}$$

Beispiel 3.4: Beispiel 3.1 fortgesetzt

Im Folgenden verwende ich die alternative Methode, die Kovarianz zu berechnen. Es ist ratsam, bei der Bestimmung der Kovarianz (und natürlich auch sonst) systematisch vorzugehen, z.B. mit Hilfe einer Tabelle, in der Sie alle benötigten Grössen eintragen bzw. berechnen.

x_i	y_i	$x_i \cdot y_i$
15	10	150
30	8	240
60	5	300
120	3	360
180	1	180
240	0.5	120
$\bar{x} = 107.5$	$\bar{y} = 4.583$	$\Sigma : 1350$

Hieraus erhält man sofort: $s_{XY} = \frac{1}{6} \cdot 1350 - 107.5 \cdot 4.583 \approx -267.7$.

Wir kommen jetzt zu unserem “linearen Zusammenhangsmass”.

Definition 3.5: Metrischer oder Empirischer oder Pearson'scher Korrelationskoeffizient

Der *Pearson'sche Korrelationskoeffizient* wird folgendermassen definiert:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{XY}}{s_X s_Y}.$$

Bemerkung 3.6

Kovarianzen und Pearson'scher Korrelationskoeffizient können nur für metrische Merkmale berechnet werden. Für ordinalskalierte Merkmale gibt es den sogenannten Rangkorrelationskoeffizienten.

Für den Korrelationskoeffizienten gilt stets

$$-1 \leq r_{XY} \leq 1.$$

Interpretation: Für $r_{XY} > 0$ liegen die Punkte im Streudiagramm tendenziell auf einer Geraden mit positivem Anstieg, für $r_{XY} < 0$ tendenziell auf einer Geraden mit negativem Anstieg. Je näher $|r_{XY}|$ an 1 liegt, um so besser ist die Anpassung der Punkte an eine Gerade. Man kann auch zeigen, dass

r_{XY} dann und nur dann ± 1 ist, wenn die Datenpunkte exakt auf einer Geraden liegen. Für $r_{XY} \approx 0$ ist kein linearer Zusammenhang erkennbar.

Beispiel 3.7: Korrelationskoeffizient

Für unser Hartholzbaum-Beispiel berechnet man $s_D^2 \approx 0.0517$ und $s_H^2 \approx 5.61$. (Überprüfen Sie das!) Das ergibt für den Korrelationskoeffizienten

$$r_{HD} = \frac{s_{HD}}{s_D \cdot s_H} = \frac{0.225}{\sqrt{0.0517} \cdot \sqrt{5.61}} \approx 0.42.$$

4 Lineare Regression

Das Ziel der linearen Regression ist es, den durchschnittlichen linearen Zusammenhang zwischen zwei (metrischen) Merkmalen X und Y durch eine Gerade $y(x) = a + bx$ darzustellen, die das gegebene Datenmaterial in einem gewissen Sinn optimal approximiert. Dabei wird davon ausgegangen, dass ein Merkmal Y von einem Merkmal X abhängt (Körpergrösse \rightarrow Körpergewicht, Trainingszeit \rightarrow Marathonzeit etc.) (Diese Abhängigkeit muss keinesfalls offensichtlich sein!)

Wie bestimmt man die Parameter a und b ? Ganz einfach:

$$b = r_{XY} \frac{s_Y}{s_X} = \frac{s_{XY}}{s_X^2} \quad \text{und} \quad a = \bar{y} - b\bar{x}.$$

Beachten Sie bitte, dass die Berechnung von b $s_X^2 > 0$ voraussetzt. In allen anderen Fällen kann eine Regressionsgerade bestimmt werden, auch wenn das vielleicht keinen Sinn macht.

Beispiel 4.1: Beispiel 3.1 fortgesetzt

Für die Berechnung von b benötigen wir noch s_X^2 und, da ich später auch r_{XY} berechnen will ausserdem s_Y^2 . Für deren Berechnung benutze ich jeweils wieder die effizientere Methode.

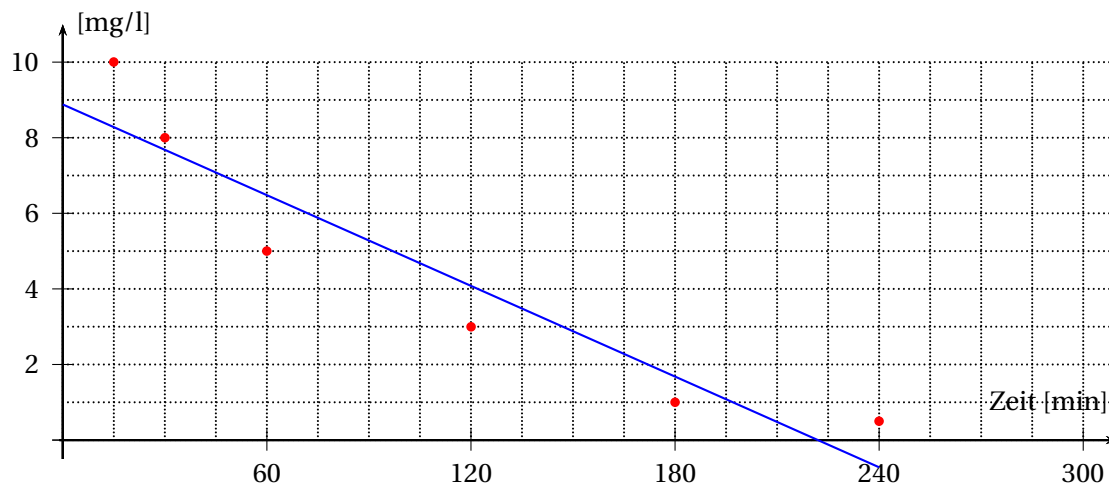
x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
15	10	150	225	100
30	8	240	900	64
60	5	300	3600	25
120	3	360	14400	9
180	1	180	32400	1
240	0.5	120	57600	0.25
$\bar{x} = 107.5$	$\bar{y} = 4.583$	$\Sigma : 1350$	$\Sigma : 109125$	$\Sigma : 199.25$

Somit ergibt sich $s_X^2 = \overline{x^2} - \bar{x}^2 = \frac{1}{6} \cdot 109125 - 107.5^2 = 6631.25$ und $s_Y^2 = \overline{y^2} - \bar{y}^2 = \frac{1}{6} \cdot 199.25 - 4.583^2 \approx 12.20$.

Mit der schon berechneten Kovarianz erhalten wir nun

$$b = \frac{s_{XY}}{s_X^2} = \frac{-267.7}{6631.25} \approx -0.04 \quad \text{und} \quad a = \bar{y} - b\bar{x} = 4.583 - (-0.04) \cdot 107.5 \approx 8.88$$

und somit für die Regressionsgerade von Y bzgl. X : $y = 8.88 - 0.04x$. Die Graphik zeigt nochmal das Streudiagramm mit eingezeichneter Regressionsgeraden.



Regressionsgeraden werden oft zur Interpolation oder Extrapolation genutzt.

Beispiel 4.2: Beispiel 3.1 fortgesetzt

Aus anderen klinischen Studien weiss man, dass die Konzentration des Arzneistoffs im Blut mindestens 1.5 mg/l betragen muss, damit er wirksam ist. Wie lange wirkt das oral eingenommene Medikament, wenn Sie Ihr Modell als Berechnungsgrundlage heranziehen?

Es muss gelten: $1.5 = 8.88 - 0.04x$. Daraus folgt $x = 184.5$. Das Medikament wirkt also rund 3 Stunden und 5 Minuten.

Wie gut ist die Anpassung der Gerade an die Daten? Dazu könnte man den Korrelationskoeffizienten benutzen. Das Mass der Wahl ist aber meist das *Bestimmtheitsmass* $B_{XY} = r_{XY}^2$. (In der Literatur findet man häufig auch die Bezeichnung R^2 !)

Wir hatten bereits festgestellt, dass immer $-1 \leq r_{XY} \leq 1$ gilt. Das liefert sofort $0 \leq B_{XY} \leq 1$. Je grösser B_{xy} ist, d.h. je dichter das Bestimmtheitsmass bei 1 liegt, desto besser ist die Anpassung der

Regressionsgeraden an die Beobachtungspaare.

Beispiel 4.3: Beispiel 3.1 fortgesetzt

Mit unseren schon berechneten Werten ergibt sich

$$B_{XY} = r_{XY}^2 = \left(\frac{s_{XY}}{s_X s_Y} \right)^2 = \frac{s_{XY}^2}{s_X^2 s_Y^2} = \frac{71663.29}{6631.25 \cdot 12.2} \approx 0.89.$$

Wir haben es hier also mit einem hohen Bestimmtheitsmass zu tun. Die Variation der y -Werte wird gut mit dem linearen Modell erklärt. Es besteht ein starker linearer Zusammenhang zwischen abhängiger und unabhängiger Variablen.

Betrachtet man nochmal das Streudiagramm, bleibt trotz des hohen Bestimmtheitsmasses ein gewisses Unbehagen. Die Lage der Datenpunkte deutet eher auf einen fallend exponentiellen Verlauf hin. Aus der Medizin ist bekannt, dass der Abbau von Medikamenten in der Tat exponentiell erfolgt. Eine geeignete Exponentialfunktion mit negativem Exponenten (anstatt einer linearen Funktion) würde das Verhalten offensichtlich besser beschreiben. Solche Betrachtungen führen zur nichtlinearen Regression. Eine visuelle Begutachtung des Datenmaterials sollte also bei jeder Auswertung routinemässig erfolgen.

5 Wahrscheinlichkeitsrechnung

Obwohl es auch schon weit früher Überlegungen zu Wahrscheinlichkeiten gab, begann die eigentliche Entwicklung der (klassischen) Wahrscheinlichkeitsrechnung im 17. Jhd. insbesondere in Verbindung mit Glücksspielen. Als Geburtsstunde wird häufig ein Briefwechsel zwischen Blaise Pascal und Pierre de Fermat aus dem Jahr 1654 angesehen. Einen weiteren Meilenstein lieferte das Lehrbuch *“Grundbegriffe der Wahrscheinlichkeitsrechnung”* von Andrej Kolmogoroff aus dem Jahr 1933, in dem der axiomatische Zugang zu Wahrscheinlichkeiten eingeführt wurde.

In der klassischen Wahrscheinlichkeitsrechnung werden Probleme oft mittels kombinatorischer Methoden behandelt. Wir wollen also auch damit beginnen.

5.1 Kombinatorik

Die Hauptfrage der Kombinatorik kann grob wie folgt umrissen werden: Gegeben eine Menge A mit n Elementen gegeben. (Da die Natur der Elemente keinerlei Rolle spielt, kann man z.B. $A = \{1, 2, \dots, n\}$ nehmen). Wieviele Möglichkeiten gibt es, k Elemente in einer wohldefinierten Weise aus A zu wählen? Bevor wir spezifizieren, was in diesem Kontext “wohldefiniert” heisst, müssen wir gewisse Begriffe bereitstellen.

5.1.1 Fakultäten und Binomialkoeffizienten

Definition 5.1: Fakultät, Binomialkoeffizient

Sei n eine natürliche Zahl. Wir definieren $n!$ (gesprochen: n Fakultät) wie folgt:

$$0! = 1, \quad n! = 1 \cdot 2 \cdot \dots \cdot n \quad \text{für } n > 0.$$

Für $n, k \in \mathbb{N}, k \leq n$, definieren wir den *Binomialkoeffizienten* $\binom{n}{k}$ durch

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Beispiele 5.2: Fakultät, Binomialkoeffizient

i) $3! = 1 \cdot 2 \cdot 3 = 6$, $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$, $10! = 1 \cdot 2 \cdot \dots \cdot 10 = 3628800$.

ii) $\binom{3}{0} = \frac{3!}{0! \cdot 3!} = 1$, $\binom{3}{1} = \frac{3!}{1! \cdot (3-1)!} = 3$, $\binom{3}{2} = 3$ und $\binom{3}{3} = 1$.

Bemerkungen 5.3

- i) Für $n > 0$ ist $n!$ also nichts anderes als das Produkt der ersten n natürlichen Zahlen.
- ii) Beachtet, dass gilt $n! = n \cdot (n-1)!$ oder allgemeiner $n! = n \cdot (n-1) \cdot \dots \cdot (n-k+1) \cdot (n-k)!$ (z.B. $10! = 10 \cdot 9 \cdot 8!$).
- iii) Man erhält eine alternative Berechnung der Binomialkoeffizienten durch geschicktes Kürzen unter Verwendung der obigen Beobachtung.

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

z.B.

$$\binom{20}{3} = \frac{20 \cdot 19 \cdot 18 \cdot 17!}{3! \cdot 17!} = \frac{20 \cdot 19 \cdot 18}{3!} = \frac{6840}{6} = 1140$$

Für praktische Berechnungen ist es fast immer vorteilhaft, zunächst so viel wie möglich zu kürzen.

Binomialkoeffizienten treten beim Potenzieren sogenannter Binome $(a+b)$ auf und haben daher ihren Namen. Man kann zeigen:

$$(a+b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1} b^1 + \dots + \binom{n}{n-1} a^1 b^{n-1} + \binom{n}{n} a^0 b^n$$

für beliebige $a, b \in \mathbb{R}, n \in \mathbb{N}_0$.

Beispiel 5.4: Potenzierung von Binomen, Binomialkoeffizienten

$$(a+b)^3 = \binom{3}{0} a^3 b^0 + \binom{3}{1} a^2 b + \binom{3}{2} a b^2 + \binom{3}{3} a^0 b^3 = a^3 + 3a^2 b + 3ab^2 + b^3.$$

Man kann Binomialkoeffizienten in einem sogenannten *Pascalschen Dreieck* anordnen. Dabei entspricht jede Zeile den Binomialkoeffizienten mit einem festen oberen Wert. Das resultierende Schema ist:

$$\begin{array}{ccccccc} & & & & 1 & & & & \\ & & & & 1 & & 1 & & \\ & & & 1 & & 2 & & 1 & \\ & & 1 & & 3 & & 3 & & 1 \\ 1 & & 4 & & 6 & & 4 & & 1 \\ & & & & \vdots & & & & \end{array}$$

Die vierte Zeile z.B. entspricht $n = 3$. Die Grössen von links nach rechts sind

$\binom{3}{0} = 1, \binom{3}{1} = 3, \binom{3}{2} = 3$ und $\binom{3}{3} = 1$. Pascals Dreieck suggeriert die folgenden allgemeingültigen Eigenschaften von Binomialkoeffizienten.

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}, \quad \binom{n}{k} = \binom{n}{n-k}, \quad \binom{n}{0} = \binom{n}{n} = 1$$

und

$$\binom{n}{1} = \binom{n}{n-1} = n.$$

Die erste Eigenschaft besagt anschaulich, dass sich jeder Wert im Pascalschen Dreieck, der nicht 1 ist, aus der Summe der unmittelbar darüberstehenden Werte ergibt. Die zweite Eigenschaft besagt, dass die Werte in einer Zeile des Pascalschen Dreiecks symmetrisch sind, von links erhält man dieselben Werte wie von rechts (z.B. $\binom{20}{3} = \binom{20}{17}$).

5.1.2 Fundamental- oder Multiplikationsprinzip der Kombinatorik

Erfordert die Erfüllung eines Sachverhaltes oder eines Prozesses S die Erfüllung von k sukzessiven Teilsachverhalten bzw. Teilprozessen S_1, S_2, \dots, S_k und gibt es für die Erfüllung dieser Teilsachverhalte unabhängig voneinander n_1, n_2, \dots, n_k Möglichkeiten, dann gibt es

$$T = n_1 \cdot n_2 \cdot \dots \cdot n_k$$

Möglichkeiten, Sachverhalt S zu erfüllen.

Häufig gilt $n_1 = n_2 = \dots = n_k = n$, d.h. jeder Teilsachverhalt hat die gleiche Anzahl der Möglichkeiten der Erfüllung. Dann gilt offensichtlich

$$T = n_1 \cdot n_2 \cdot \dots \cdot n_k = n \cdot n \cdot \dots \cdot n = n^k.$$

Bemerkung 5.5

Oft ist die Voraussetzung der Unabhängigkeit der Erfüllungsmöglichkeiten der Teilsachverhalte nicht gegeben. Das Fundamentalprinzip kann dann trotzdem angewandt werden, sofern man die aufgrund von aus der Erfüllung vorgelagerter Teilsachverhalte resultierenden Einschränkungen berücksichtigt.

Beispiele 5.6: Fundamentalprinzip der Kombinatorik

- i) (Unabhängigkeit der Erfüllungsmöglichkeiten der Teilsachverhalte) Auf einer Speisekarte gibt es 6 Vorspeisen, 9 Hauptgerichte und 4 verschiedene Nachspeisen. Wieviele Möglichkeiten gibt es, ein Menü mit drei Gängen zusammenzustellen?

Antwort: Sachverhalt S_1 besteht hier aus der Wahl der Vorspeise, d.h. $n_1 = 6$, S_2 entspricht Wahl des Hauptgerichts, d.h. $n_2 = 9$ und S_3 Wahl der Nachspeise, d.h. $n_3 = 4$. Es gibt demnach $T = 6 \cdot 9 \cdot 4 = 216$ Möglichkeiten, ein Menü mit drei Gängen zusammenzustellen.

- ii) Wieviele 3-buchstabile Code-Wörter kann man mit den ersten 8 Buchstaben des Alphabets auf folgende Weisen bilden?

1. Ohne Wiederholung von Buchstaben (Teilsachverhalte nicht unabhängig!): S_1 Wahl des 1. Buchstaben $\Rightarrow n_1 = 8$, S_2 Wahl des 2. Buchstaben $\Rightarrow n_2 = 7$, S_3 Wahl des 3. Buchstaben $\Rightarrow n_3 = 6$.

Antwort: $T = 8 \cdot 7 \cdot 6 = 336$.

2. Buchstaben können wiederholt werden (Teilsachverhalte sind unabhängig!): S_1 Wahl des 1. Buchstaben $\Rightarrow n_1 = 8$, S_2 Wahl des 2. Buchstaben $\Rightarrow n_2 = 8$, S_3 Wahl des 3. Buchstaben $\Rightarrow n_3 = 8$.

Antwort: $T = 8 \cdot 8 \cdot 8 = 8^3 = 512$.

3. Benachbarte Buchstaben sind nicht gleich (Teilsachverhalte nicht unabhängig!): S_1 Wahl des 1. Buchstaben $\Rightarrow n_1 = 8$, S_2 Wahl des 2. Buchstaben $\Rightarrow n_2 = 7$, S_3 Wahl des 3. Buchstaben $\Rightarrow n_3 = 7$.

Antwort: $T = 8 \cdot 7 \cdot 7 = 392$.

In vielen kombinatorischen Fragestellungen wird das Fundamentalprinzip mehr oder weniger explizit genutzt.

5.1.3 Permutationen

Sei eine n -elementige Menge gegeben. Eine *Permutation* dieser Menge ist eine Zusammenstellung dieser Elemente in irgendeiner Reihenfolge (Reihenfolge wichtig! Wiederholungen nicht gestattet, d.h. “kein Zurücklegen”!). (Vom mathematischen Standpunkt aus betrachtet ist eine Permutation also nichts anderes als eine eineindeutige Funktion dieser Menge auf sich selbst.)

Beispiel 5.7: Permutationen

Sei $A = \{1, 2, \dots, 5\}$. Beispiele für Permutationen sind $1, 2, 3, 4, 5$; $2, 1, 3, 4, 5$; $2, 3, 1, 4, 5$ etc.

Beispiel 5.8: Permutationen

Wir betrachten die 3 Lebensbereiche *Familie, Beruf, Sport*. Permutationen dieser Menge sind *Familie, Beruf, Sport*; *Beruf, Sport, Familie*; *Familie, Sport, Beruf* etc.

Wieviele Permutationen ${}_nP$ einer n -elementigen Menge existieren?

$$\text{Anzahl Permutationen einer } n\text{-elementigen Menge: } {}_nP = n!$$

Das folgt u.a. aus dem Fundamentalprinzip mit nicht unabhängigen Teilsachverhalten. Man hat n Möglichkeiten, die erste Position in der Zusammenstellung zu besetzen. Hat man einmal die erste Position besetzt, bleiben $n - 1$ Möglichkeiten, die zweite Position zu besetzen (Wiederholungen sind nicht erlaubt!) usw.

Beispiele 5.9: Permutationen

- i) 4 Bilder sollen an einer Wand in einer Galerie von links nach rechts angeordnet werden. Wieviele Möglichkeiten gibt es? Antwort: ${}_4P = 4! = 24$.
- ii) Ein 100 Meter Sprintfinale besteht aus 8 Teilnehmern. Wieviele Möglichkeiten der Platzverteilung gibt es? Antwort: ${}_8P = 8! = 40320$.
- iii) Wieviele Möglichkeiten der Priorisierung der 3 Lebensbereiche aus dem Beispiel von oben gibt es? Antwort: ${}_3P = 3! = 6$.

Häufig kommt es vor, dass Mengen mit n Elementen nicht unterscheidbare Elemente enthalten, z.B. “ein Buch mehrmals in der gleichen Ausgabe”. Dann gibt es zwar wiederum $n!$ “physische” Permutationen dieser n Elemente, es werden aber gewisse Zusammenstellungen wegen der Gleichartigkeit von Elementen nicht unterscheidbar sein.

Beispiel 5.10

Wieviele Wörter lassen sich aus A_1, A_2, B, C theoretisch (Sie müssen nicht unbedingt sinnvoll sein.) bilden? Um die Problematik zu verdeutlichen: $A_1 A_2 B C$ lässt sich nicht von $A_2 A_1 B C$ unterscheiden, wenn man die Indizes an den A s weglässt. Eine bessere Fragestellung ist die folgende: Wieviele Wörter lassen sich bilden, die das A genau zweimal, B und C genau einmal enthalten?

Überlegungen dieser Art führen zum Begriff der *Permutationen mit Wiederholung* (Reihenfolge wichtig! Wiederholungen erlaubt!)

Sei eine n -elementige Menge gegeben. Von diesen n Elementen seien n_1 vom Typ 1, n_2 vom Typ 2, ..., n_m vom Typ m . (Es ist also $n = n_1 + n_2 + \dots + n_m$). Dann gibt es

$${}_n P_{n_1, n_2, \dots, n_m} = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_m!}$$

unterscheidbare Anordnungsmöglichkeiten.

Beispiele 5.11: Permutationen mit Wiederholungen

- i) Wieviele Möglichkeiten gibt es, sechs Bücher anzuordnen, wenn von einem zwei Exemplare und von einem anderen drei Exemplare vorhanden sind? Antwort: ${}_6 P_{1,2,3} = \frac{6!}{1! \cdot 2! \cdot 3!} = 60$.

In diesem Beispiel handelt es sich in der Tat um eine physische Menge mit 6 Elementen.

- ii) Gegeben sei die 3-elementige Menge $A = \{1, 2, 3\}$. Wie viele 7-ziffrige Zahlen lassen sich bilden, die die 1 und 2 genau zweimal und die 3 genau dreimal enthalten? Antwort: ${}_7 P_{2,2,3} = \frac{7!}{2! \cdot 2! \cdot 3!} = 210$.

Es wäre hier mathematisch nicht exakt zu sagen ...aus den Elementen der Menge $A = \{1, 1, 2, 2, 3, 3, 3\}$..., weil bei Mengen keine Elementwiederholungen erlaubt sind.

5.1.4 Anzahl von Auswahlmöglichkeiten k aus n mit Berücksichtigung der Reihenfolge (Variationen)

Permutationen verwenden stets alle zur Verfügung stehenden Elemente in der Zusammenstellung. Es gibt jedoch Situationen bzw. Fragestellungen, wo das nicht sinnvoll ist. Wenn ihr an das Sprintfinale von oben denkt, wäre z.B. eine sinnvolle Fragestellung, wieviele Möglichkeiten der Medailenverteilung es gibt.

Gegeben sei wieder eine Menge mit n Elementen.

Auswahl mit Zurücklegen und mit Berücksichtigung der Reihenfolge

Anzahl Möglichkeiten: ${}_n\tilde{V}_k = n^k$.

Das folgt unmittelbar aus dem Fundamentalprinzip. k kann hier auch grösser als n sein, da “zurückgelegt” wird.

Auswahl ohne Zurücklegen und mit Berücksichtigung der Reihenfolge

Anzahl Möglichkeiten: ${}_nV_k = \frac{n!}{(n-k)!} (= n \cdot (n-1) \cdot \dots \cdot (n-k+1))$

Man kann sich das wieder leicht plausibel machen. Da hier Reihenfolge wichtig ist, Wiederholungen jedoch nicht erlaubt sind, hat man n Möglichkeiten, die erste Position in der Zusammenstellung zu besetzen. Hat man einmal die erste Position besetzt, bleiben $n-1$ Möglichkeiten, die zweite Position zu besetzen (Wiederholungen sind nicht erlaubt!) usw. Für die k -te Position bleiben dann noch $(n-k+1)$ Möglichkeiten. k darf hier natürlich nicht grösser als n sein!

Bei den zwei obigen Arten der Auswahl spricht man auch von *Variationen*. Ich habe deshalb die “V-Notation” für die entsprechenden Anzahlen verwendet. (Auf euren TR gibt es häufig die Taste nPk für ${}_nV_k$.) Beachtet auch: ${}_nP = {}_nV_n$, da Permutationen als Variation (ohne Zurücklegen) von n Elementen aus n betrachtet werden kann.

Beispiele 5.12: Variationen

- i) Wieviele Möglichkeiten gibt es, die Medaillen Gold, Silber und Bronze zu verteilen? Antwort:
 ${}_8V_3 = \frac{8!}{5!} = 336$.
- ii) Eine Seriennummer besteht aus 2 Buchstaben (aus A, B, \dots, H) gefolgt durch 3 Ziffern (aus $0, 1, 2, \dots, 9$). Wiederholungen sind nicht erlaubt. Wieviele mögliche Seriennummern gibt es?

Beachtet bitte, dass es sich hier um einen “zweistufigen” Prozess handelt. Man muss zunächst einmal die zwei Buchstaben wählen und danach die drei Ziffern. Das Ergebnis folgt dann aus dem Fundamentalprinzip.

$$\text{Antwort: } {}_8V_2 \cdot {}_{10}V_3 = \frac{8!}{(8-2)!} \cdot \frac{10!}{(10-3)!} = 56 \cdot 720 = 40320.$$

5.1.5 Anzahl von Auswahlmöglichkeiten k aus n ohne Berücksichtigung der Reihenfolge (Kombinationen)

Häufig spielt die Reihenfolge bei Auswahlen keine Rolle. Das klassische Beispiel ist die Lotterie. Hier ist es unerheblich, in welcher Reihenfolge die Gewinnzahlen gezogen wurden. Man spricht in diesem Fall meist von *Kombinationen*.

Auswahl ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge

Anzahl Möglichkeiten: ${}_nC_k = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

(Auf euren TR gibt es häufig die Taste nCk dafür!)

Im Gegensatz zu Variationen spielt bei Kombinationen die Reihenfolge keine Rolle. Da man k Elemente auf $k!$ Weisen permutieren kann, wird hier zwischen $k!$ Variationen nicht unterschieden. Das erklärt den Faktor $k!$ im Nenner von ${}_nC_k$.

Beispiele 5.13: Kombinationen ohne Zurücklegen

- i) Wieviele Möglichkeiten gibt es überhaupt, die Medaillen zu verteilen, d.h. ohne Rücksicht auf die Reihenfolge bzw. Art der Medaille? Da hier die Reihenfolge der Medaillen keine Rolle mehr spielt, haben wir: Antwort: ${}_8C_3 = \frac{8!}{3! \cdot 5!} = 56$.
- ii) Beim Lotto "6 aus 49" werden 6 Kugeln aus 49 nummerierten Kugeln gezogen. Wieviele Möglichkeiten gibt es hierfür?

Antwort: ${}_{49}C_6 = \binom{49}{6} = \frac{49!}{6! \cdot 43!} = 13983816$.

- iii) Gegeben sei ein Standard 32-Blatt Kartenspiel (7,8,9,10,B,D,K,A sind jeweils viermal enthalten). Wieviele Möglichkeiten gibt es, 5 Karten auf der Hand zu haben, die 3 Damen (D) und 2 Könige (K) enthalten?

Auch hier haben wir es wieder mit einem "mehrstufigen" Prozess zu tun. Wir brauchen also das Fundamentalprinzip.

Antwort: ${}_4C_3 \cdot {}_4C_2 = \frac{4!}{3! \cdot 1!} \cdot \frac{4!}{2! \cdot 2!} = 4 \cdot 6 = 24$.

(Man muss 3 D aus 4 und 2 K aus 4 wählen!)

Auswahl mit Zurücklegen und ohne Berücksichtigung der Reihenfolge

Anzahl Möglichkeiten: ${}_n\tilde{C}_k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k! \cdot (n-1)!}$

Auch hier kann natürlich wieder $k > n$ sein, da Wiederholungen zugelassen sind.

Beispiele 5.14: Kombinationen mit Zurücklegen

- i) Auf wieviele Weisen können 4 (unterscheidbare) Lehrer auf 3 Schulen verteilt werden?

Antwort: ${}_3\tilde{C}_4 = \binom{6}{4} = 15$

Man muss hier aus $n = 3$ Schulen für jeden der $k = 4$ Lehrer eine (nicht notwendigerweise unterschiedliche) Schule wählen.

- ii) Ein B&B Betreiber hat 10 Gäste und möchte deren Getränkewahl für das Frühstück wissen. Zur Auswahl stehen *Kaffee*, *Tee* und *heisse Schokolade*. Auf wieviele Arten können sich die

Gäste entscheiden? Antwort: ${}_3\tilde{C}_{10} = \binom{12}{10} = 66$

Man möchte meinen, dass das “C” in der Notation von engl. “Combination” herrührt. Wahrscheinlicher ist jedoch, dass ${}_nC_k$ für engl. *n choose k* steht.

Alle für uns wichtigen Kombinatorikprobleme sind Kopien der oben angesprochenen. Nur die konkrete Situation ändert sich. Die Schwierigkeit besteht darin zu erkennen, um welches Auswahlprinzip es sich handelt.

5.2 Grundlagen der Wahrscheinlichkeitsrechnung

Die Wahrscheinlichkeitsrechnung befasst sich mit der Aufdeckung der Gesetzmässigkeiten von zufälligen Ereignissen.

5.2.1 Ereignisraum und Ereignismenge

Dazu werden *Zufallsexperimente* (im weitesten Sinn) untersucht. Resultate von Zufallsexperimenten sind *Elementarereignisse*, d.h. nicht mehr zerlegbare, sich gegenseitig ausschliessende Ausgänge. Die Gesamtheit der Elementarereignisse wird *Ereignisraum* oder *Ergebnisraum* (im Englischen auch *sample space*) genannt und häufig mit S oder Ω bezeichnet.

Beispiele 5.15: Ereignisraum

- i) Einmaliges Würfeln: $S = \{1, 2, 3, 4, 5, 6\}$.
- ii) Einmaliges Werfen einer Münze: $S = \{K, Z\}$.
- iii) Zufällige Wahl eines Punktes auf Kreisscheibe (Schiessen auf Kreisscheibe):

$$S = \{\text{Punkte auf Kreisscheibe}\}$$

Diese Beispiele demonstrieren, was mit “sich gegenseitig ausschliessend” gemeint ist. Man kann z.B. nicht gleichzeitig eine 1 und eine 6 würfeln.

Jede Teilmenge A des Ereignisraums S ist ein *zufälliges Ereignis*. Die Gesamtheit der zufälligen Ereignisse wird häufig als *Ereignismenge* bezeichnet. Notation: $E(S)$.

Ein Ereignis A tritt genau dann ein, wenn eines der Elementarereignisse, aus denen A besteht, eintritt. Im Kontext von zufälligen Ereignissen sind die Elementarereignisse gerade die einelementigen zufälligen Ereignisse. Sie können nicht weiter zerlegt werden. Es ist oft hilfreich, Ereignisse in Verbindung mit Versuchsausgängen zu betrachten. Einige Beispiele hierfür im Kontext "Würfeln".

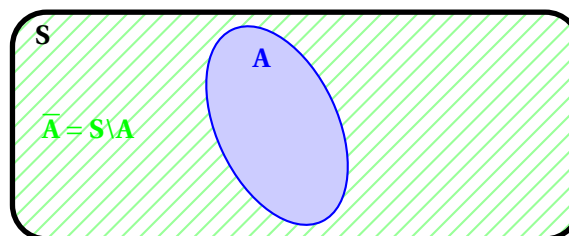
Beispiele 5.16: Ereignisse

- i) Ereignis A_1 : "Es fällt eine gerade Zahl." $\Rightarrow A_1 = \{2, 4, 6\}$. A_1 tritt ein, wenn eine 2, 4 oder 6 gewürfelt wird, also eines der Elementarereignisse, aus denen das (zusammengesetzte) Ereignis A_1 besteht.
- ii) A_2 : "Es fällt eine Zahl ≥ 5 ." $\Rightarrow A_2 = \{5, 6\}$.
- iii) A_3 : "Es fällt die 1 oder 3." $\Rightarrow A_3 = \{1, 3\}$.

Tritt ein Ereignis bei einem Zufallsexperiment immer ein, so liegt ein *sicheres Ereignis* vor. Da S alle möglichen Versuchsausgänge umfasst, ist S ein sicheres Ereignis und als solches gehört S immer zur Ereignismenge $E(S)$. Andererseits gibt es Ereignisse, die nie auftreten können, sogenannte *unmögliche Ereignisse*. Die leere Menge \emptyset repräsentiert ein solches unmögliches Ereignis. Da sie keine Elemente enthält, kann auch keines ihrer Elemente als Versuchsausgang auftreten. Das Ereignis tritt also nie ein. Auch \emptyset gehört immer zur Ereignismenge $E(S)$.

Wie wir gesehen haben, ist jedes Ereignis als Teilmenge von S definiert. Sind zwei Teilmengen A und B , also zwei Ereignisse, gegeben, führt jede Mengenoperation zu einer neuen Teilmenge, also einem neuen Ereignis. (Die folgenden Beispiele beziehen sich auf Beispiel 5.16 oben.) Zur Verdeutlichung dieser Operationen werden sie schematisch durch sogenannte *Venn-Diagramme* dargestellt.

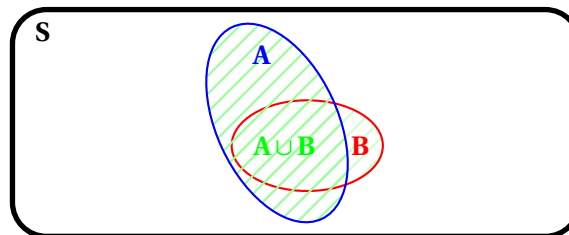
- $\bar{A} := S \setminus A$ ($S \setminus A$ ist die sogenannte *Mengendifferenz* zwischen S und A oder das *Komplement* von A in S und besteht aus allen Elementen von S , die nicht in A enthalten sind.) Tritt also \bar{A} ein, so tritt A nicht ein $\Rightarrow \bar{A}$ entspricht dem Ereignis **nicht** A . Man nennt A und \bar{A} auch *entgegengesetzte* oder *komplementäre* Ereignisse.



Beispiel 5.17: Komplementäres Ereignis

$\bar{A}_1 = S \setminus A_1 = \{1, 3, 5\}$. Das entspricht dem Ereignis “Eine ungerade Zahl fällt.”

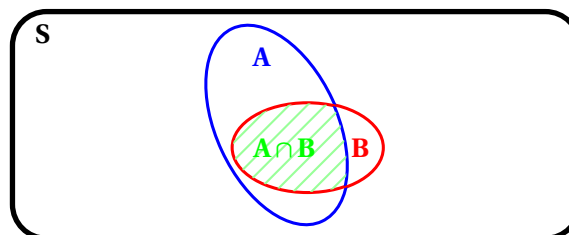
- $A \cup B$ entspricht dem Ereignis A **oder** B . Es tritt ein, wenn A oder B (oder beide) eintreten.



Beispiel 5.18: Vereinigung von Ereignissen

$A_1 \cup A_3 = \{1, 2, 3, 4, 6\}$. Das entspricht dem Ereignis “Die 5 fällt nicht.” Wenn dieses Ereignis eintritt, dann treten A_1 ein oder A_3 ein (oder beide, was hier jedoch ausgeschlossen ist).

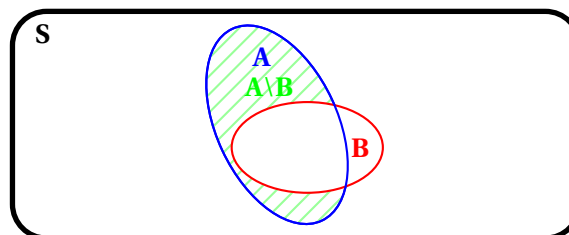
- $A \cap B$ entspricht dem Ereignis A **und** B . Es tritt ein, wenn sowohl A als auch B eintreten.



Beispiel 5.19: Durchschnitt von Ereignissen

$A_1 \cap A_2 = \{6\}$. Das entspricht dem Ereignis “Es wird eine 6 gewürfelt.” Wenn dieses Ereignis eintritt, dann treten sowohl A_1 als auch A_2 ein.

- $A \setminus B$ entspricht dem Ereignis A **ohne** B . Es tritt ein, wenn A eintritt, B jedoch nicht.



Beispiel 5.20

$A_2 \setminus A_1 = \{5\}$. Das entspricht dem Ereignis “Die 5 fällt.” Wenn dieses Ereignis eintritt, tritt A_2 ein, aber nicht A_1 .

Zwei Ereignisse A und B heissen *unvereinbar* oder *konträr* oder *disjunkt*, wenn $A \cap B = \emptyset$. Solche Ereignisse können also nie gleichzeitig eintreten. (A und \bar{A} sind z.B. immer unvereinbar!) A_1 und A_3 von oben sind ebenfalls unvereinbar, da sie nicht gleichzeitig eintreten können.

5.2.2 Zuordnung von Wahrscheinlichkeiten

Im letzten Abschnitt haben wir zufällige Ereignisse definiert. Wir wollen jetzt solchen Ereignissen Wahrscheinlichkeiten zuordnen. Dafür gibt es verschiedene Ansätze.

Klassische Wahrscheinlichkeit

Dieser Wahrscheinlichkeitsbegriff ist der historisch älteste und entwickelte sich in der 2. Hälfte des 17. Jahrhunderts. Er wird auf Zufallsexperimente mit

endlich vielen, gleichwahrscheinlichen Elementarereignissen

angewandt. Solche Zufallsexperimente heissen auch *Laplace-Experimente*. Das Würfeln mit einem fairen Würfel und das Werfen einer fairen Münze sind Beispiele hierfür. Für die Wahrscheinlichkeit (WSK) eines Ereignisses A gilt dann:

$$P(A) = \frac{\text{Anzahl Elemente in } A}{\text{Anzahl Elemente in } S} = \frac{|A|}{|S|}.$$

Bei der Anzahlbestimmung kommt häufig die Kombinatorik zum Zuge. Das ist der Grund, warum wir uns überhaupt kurz mit diesem Thema beschäftigen haben.

Beispiele 5.21: Klassische Wahrscheinlichkeiten

- i) Es werde mit zwei (fairen) Würfeln gleichzeitig gewürfelt. Ereignis A : “Augensumme ≥ 10 ”, Ereignis B : “Augensumme gerade.” Wir wollen die WSK dieser Ereignisse bestimmen.
 $S = \{e_{11}, e_{12}, \dots, e_{66}\}$ wobei e_{ij} das Elementarereignis “1. Würfel Zahl i und 2. Würfel Zahl j ” bezeichnet. Beachtet bitte, dass alle diese Elementarereignisse gleichwahrscheinlich sind. Unser Ereignisraum besteht also aus 36 Elementarereignissen, also $|S| = 36$.
Dem Ereignis A entspricht die Teilmenge $A = \{e_{46}, e_{64}, e_{55}, e_{56}, e_{65}, e_{66}\}$, also $|A| = 6$ und somit $P(A) = \frac{6}{36} = \frac{1}{6}$.
Zur Berechnung von $P(B)$ macht euch bitte klar, dass genau die Hälfte der Elementarereignisse, also 18, zu B gehören. Daher haben wir $P(B) = \frac{18}{36} = \frac{1}{2}$.

- ii) Im Lotto-Beispiel hatten wir gesehen, dass es 13983816 Möglichkeiten gibt, 6 Kugeln aus 49 nummerierten Kugeln zu ziehen. Auch hier ist jede Kombination von 6 Kugeln gleichwahrscheinlich. Die WSK, dass eine bestimmte Kombination gezogen wird, dass man also "einen Sechser im Lotto" hat, beträgt demnach

$$P(\text{"Sechser"}) = \frac{1}{13983816} = 0.000000071,$$

ist also verschwindend gering.

- iii) Im Karten-Beispiel wurde die Anzahl der Möglichkeiten, eine 5-blättrige Hand, bestehend aus 3 Damen und 2 Königen, zu erhalten mit 24 bestimmt. In einem Spiel mit 32 Karten gibt es jedoch $\binom{32}{5}$ Möglichkeiten einer 5-blättrigen Hand, wieder alle gleichwahrscheinlich. Wir haben daher:

$$P(\text{"5-blättrige Hand mit 3 Damen und 2 Königen"}) = \frac{24}{\binom{32}{5}} = 0.000119.$$

Ich hoffe, ihr erkennt, warum die Kombinatorik in diesem Zusammenhang so immens wichtig ist.

Statistische Wahrscheinlichkeit

Um die statistische Wahrscheinlichkeit eines Ereignisses A zu bestimmen, führt man ein Zufallsexperiment viele Male aus und zählt das Eintreten des Ereignisses A . Man kann dann die relative Häufigkeit $h_n(A)$ des Eintretens von A bestimmen bei n Durchführungen des Experiments (d.h. $h_n(A) = (\text{"Anzahl Eintreten von } A") / n$). Man definiert dann

$$P(A) = \lim_{n \rightarrow \infty} h_n(A).$$

Ausser im eng begrenzten Kontext von klassischen WSKen ist das eine der wenigen praktikablen Methoden, WSKen von Ereignissen wirklich zu bestimmen.

Axiomatische Wahrscheinlichkeit (Kolmogoroff-Axiome)

Dieser Zugang postuliert Eigenschaften, die WSKen von Ereignissen zu erfüllen haben. Er lehnt sich dabei an gewisse entsprechende Eigenschaften der relativen Häufigkeiten an, die man als Erfahrungstatsachen betrachten kann und die man von WSKen sinnvollerweise erwartet. Für die praktische Berechnung von WSKen hat dieser Zugang keine Bedeutung. Er lehrt mehr, wie man mit WSKen rechnet. Der Vollständigkeit halber seien sie hier erwähnt:

(K1) Für jedes $A \in E(S)$ gilt $0 \leq P(A) \leq 1$.

(K2) Für das sichere Ereignis S gilt $P(S) = 1$.

(K3) Sind A_1, A_2, A_3, \dots gegenseitig unvereinbar, gilt also $A_i \cap A_j = \emptyset$ für alle i, j mit $i \neq j$, so gilt

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + \dots$$

Bemerkungen 5.22: Axiomatik in der Mathematik

Die folgenden Bemerkungen könnt ihr gerne überspringen. Axiomatik in der Mathematik gehört zur mathematischen Logik/Grundlagen der Mathematik und spielt auch für die meisten Mathematiker im "Alltagsgeschäft" keine Rolle. Aber vielleicht findet es ja der eine oder andere von euch interessant. Das Folgende ist alles sehr informell!

i) Der axiomatische Zugang zur Mathematik versucht, (Teil)Bereiche der Mathematik auf der Basis von nichtbewiesenen, also postulierten, meist selbstevidenten Eigenschaften der Objekte dieses Bereichs und gewisser ihrer Beziehungen zueinander, sogenannten Axiomen oder Postulaten, zu entwickeln. Dieser Zugang geht zurück auf Euklid (wahrscheinlich 3.Jhd. v.u.Z.), der in seinen "Elementen" versucht hat, die Geometrie der Ebene axiomatisch zu entwickeln. Ausgehend von Objekten wie Punkten und Geraden und einigen grundlegenden (unbewiesenen) Tatsachen (Axiomen) hat er versucht (Erfolgreich!), "nachgelagerte" Resultate logisch zu beweisen. Eines seiner Postulate war z.B., dass durch zwei nichtzusammenfallende Punkte genau eine Gerade geht. (Logisch, oder?) Ein anderes Axiom, das sogenannte Parallelenaxiom, postulierte, dass es, gegeben eine Gerade und einen Punkt, der nicht auf dieser Geraden liegt, genau eine Gerade durch diesen Punkt gibt, die parallel zur gegebenen Geraden ist. (Genau so logisch?). Lange hat niemand gross über das Parallelenaxiom nachgedacht. Erst als die Axiomatik wieder in den Blickpunkt der Mathematiker gerückt war (Ende 19., Anfang 20. Jhd.), begann man, wieder über dieses Axiom nachzudenken. Man versuchte, dieses Postulat z.B. dadurch als abhängig von den anderen zu identifizieren, dass man versuchte, es auf Basis der anderen Axiome zu beweisen. Vergeblich. Das führte letztendlich zur Entdeckung sogenannter nichteuklidischer Geometrien, in denen das Parallelenaxiom nicht mehr gültig ist und die heute in der theoretischen Physik und Kosmologie eine bedeutende Rolle spielen.

ii) Ein Kennmerk des axiomatischen Zugangs zur Mathematik ist, dass die Objekte der betreffenden Theorie undefiniert bleiben. Die Kolmogoroff-Axiome sagen z.B. nicht, was Wahrscheinlichkeit ist. Es werden nur gewisse Eigenschaften postuliert.

Euklid hat noch versucht, Begriffe wie "Punkt" (*Ein Punkt ist, was keine Teile hat.*) oder "Gerade" (*Eine Gerade ist eine Linie, die bezüglich der Punkte auf ihr stets gleich liegt.*) zu definieren. Das Problem besteht darin, dass in diesen "Definitionen" wieder Begriffe auftauchen, die aufs Neue definiert werden müssten (Was ist ein "Teil", was eine "Linie" etc.). Wenn man sich nicht auf undefinierte Objekte beziehen würde, würde das zu einem unendlichen "Definitionsprozess" führen. Das ist auch der Grund dafür, warum man sich beim Aufbau der Theorie auf unbewiesene Axiome stützen muss. Der Beweisprozess würde sonst unendlich.

Eine mathematische Theorie basiert also sozusagen auf undefinierten Objekten und unbewiesenen Postulaten. Nur insofern man diese Postulate akzeptiert, hat die betreffende Theorie auch eine “absolute Wahrheit”. Den englischen Logiker Bertrand Russell (1872-1970) hat das dazu bewogen, die Mathematik als Wissenschaft zu definieren, bei der man nicht weiss, wovon man spricht, noch ob das, was man sagt, wahr ist. (Kommt euch bekannt vor??)

iii) Auf der Basis der Kolmogoroff-Axiome kann im Prinzip die ganze Wahrscheinlichkeitstheorie aufgebaut werden. Aus (K2) und (K3) folgt z.B. unmittelbar, dass die WSK des unmöglichen Ereignisses gleich 0 sein muss: $P(\emptyset) = 0$. Wie?

iv) Bei (K3) ist nicht unbedingt klar, warum man sich auf eine unendliche Menge von vereinbaren Ereignissen bezieht. Vergleiche (W4) später. Der Hauptgrund hierfür ist, dass es die Theorie so viel reicher macht.

v) Sowohl die Definition der klassischen Wahrscheinlichkeit als auch die frequentistische Definition haben gewisse logische Mängel. Bei der klassischen Wahrscheinlichkeit wird z.B. “Gleichwahrscheinlichkeit” der Elementarereignisse vorausgesetzt. Was heisst das, wenn “Wahrscheinlichkeit” noch nicht definiert ist? Die Definition ist also gewissermassen “zirkulär”. Trotzdem, sowohl “Klassische Wahrscheinlichkeiten” als auch “Frequentistische Wahrscheinlichkeiten” genügen den Kolmogoroff-Axiomen. Sie bilden sozusagen “Modelle” für dieses Axiomensystem.

vi) Andere bekannte Axiomensysteme sind das Peano-Axiomensystem, auf der sich die Arithmetik und Ordnung der natürlichen Zahlen entwickeln lässt und das Zermelo-Fraenkel-Axiomensystem für die Mengenlehre.

vii) Äusserst interessant in diesem Zusammenhang ist auch der “Gödelsche Unvollständigkeitssatz”. Hier geht es um die Frage, ob man alle Hypothesen, die man im Rahmen eines Axiomensystems sinnvollerweise formulieren kann, auch innerhalb des Axiomensystems beweisen kann. Die Antwort ist Nein. Es gibt in jedem (hinreichend komplexen) Axiomensystem Aussagen, die man “dort” weder beweisen noch widerlegen kann. Stichwort “Kontinuumshypothese”.

viii) Sorry, I got a bit carried away.

5.2.3 Wichtige Regeln für Wahrscheinlichkeiten

(W1): $0 \leq P(A) \leq 1$. $P(A)$ ist demnach immer eine Zahl zw. 0 und 1. (Oft wird die WSK eines Ereignisses auch in Prozenten ausgedrückt. Einer WSK von $p = 0.1$ entspricht z.B. eine WSK von 10%, allgemein einer WSK von p eine WSK von $p \cdot 100\%$ in Prozenten.)

(W2): $P(\bar{A}) = 1 - P(A)$. Die WSK des entgegengesetzten Ereignisses \bar{A} und die WSK von A addieren sich immer zu 1 auf. (Beachtet bitte in diesem Zusammenhang auch, dass $\bar{\bar{A}} = A$, dass also das entgegengesetzte Ereignis von \bar{A} wieder A ist.)

Beispiel 5.23

$$P(\bar{A}_3) = P(\{2, 4, 5, 6\}) = \frac{2}{3} = 1 - \frac{1}{3} = 1 - P(A_3) .$$

Diese vermeintlich unscheinbare Regel ist von grosser Wichtigkeit. Es ist häufig einfacher, die WSK des entgegengesetzten Ereignisses zu berechnen, als die eigentlich gesuchte WSK!!

Beispiel 5.24: Geburtstagsproblem

(Für unsere Verhältnisse schon relativ komplex, aber interessant.)

Wir betrachten eine Gruppe von 6 Personen. Wie gross ist die WSK, dass mindestens 2 Personen am gleichen Tag Geburtstag haben (Ereignis A)? (Dabei schliessen wir den 29. Februar als Geburtstag aus.) Wir verunterstellen auch, dass alle Geburtstage gleichwahrscheinlich sind. (Das ist in der Realität nicht ganz der Fall!)

Unser Ereignisraum ist gegeben durch:

$$S = \{\text{Jede Anordnung von 6 Geburtstagen}\}, \quad \text{also} \quad |S| = 365^6 .$$

(Für jede Position stehen 365 Tage zur Auswahl.) Die direkte Berechnung von $P(A)$ ist unendlich schwierig. Wir versuchen es mit dem entgegengesetzten Ereignis.

\bar{A} : Keine Personen haben am gleichen Tag Geburtstag.

Wir erhalten relativ einfach

$$|\bar{A}| = {}_{365}V_6 = 365 \cdot 364 \cdot 363 \cdot 362 \cdot 361 \cdot 360 .$$

Mit obiger Regel erhalten wir

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{365 \cdot 364 \cdot 363 \cdot 362 \cdot 361 \cdot 360}{365^6} = 0.04 .$$

Man überschreitet zum ersten Mal eine WSK von 0.5, wenn man diese Berechnung mit einer Gruppe von 23 Personen ausführt. Die Antwort auf die Frage: Wieviel Personen sollte eine Gruppe mindestens umfassen, damit die WSK, dass (mindestens) 2 Personen am gleichen Tag Geburtstag haben, grösser als 0.5 ist, ist also 23.

(W3): $P(S) = 1$, $P(\emptyset) = 0$. Die WSK des sicheren Ereignisses ist immer 1, die des unmöglichen Ereignisses immer 0.

(W4): Für Ereignisse A_1, A_2, \dots, A_n , die paarweise unvereinbar sind, d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$ gilt:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) .$$

Wenn also keines der Ereignisse zusammen mit einem anderen Ereignis eintreten kann, dann ist die WSK, dass überhaupt eines der Ereignisse eintritt gleich der Summe der WSKen der Ereignisse.

Beispiel 5.25

$$P(A_1 \cup A_3) = P(\{1, 2, 3, 4, 6\}) = \frac{5}{6} = \frac{1}{2} + \frac{1}{3} = P(A_1) + P(A_3) .$$

Beachtet bitte, dass das für Ereignisse, die nicht disjunkt sind, i.A. nicht richtig ist!

Beispiel 5.26

A_1 und A_2 sind nicht disjunkt, da $A_1 \cap A_2 = \{6\} \neq \emptyset$. Betrachten wir nun

$$P(A_1 \cup A_2) = P(\{2, 4, 5, 6\}) = \frac{2}{3} \neq \frac{5}{6} = \frac{1}{2} + \frac{1}{3} = P(A_1) + P(A_2) .$$

(W5): Allgemein gilt

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

Beispiel 5.27

$$P(A_1 \cup A_2) = P(\{2, 4, 5, 6\}) = \frac{2}{3} = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = P(A_1) + P(A_2) - P(A_1 \cap A_2) .$$

Dieses Resultat lässt sich unter Verwendung des Inklusions-/Exklusions-Prinzips auf mehr als 2 Ereignisse erweitern. Wir wollen es bei zwei Ereignissen belassen.

(W6): $P(A \setminus B) = P(A) - P(A \cap B)$

Beispiel 5.28

$$P(A_1 \setminus A_2) = P(\{1, 3\}) = \frac{1}{3} = \frac{1}{2} - \frac{1}{6} = P(A_1) - P(A_1 \cap A_2) .$$

(W7): Gilt für zwei Ereignisse A und B , dass $A \subseteq B$, so sagt man auch *das Ereignis A zieht das Ereignis B nach sich*. Wenn immer A eintritt, tritt auch B ein. In diesem Fall gilt: $P(A) \leq P(B)$.

Ein abschliessendes Beispiel:

Beispiel 5.29: Bestimmen von Wahrscheinlichkeiten

In einem Kleinbetrieb werden Geschlecht (M-männlich; W-weiblich) und Zivilstand (L-ledig; V-verheiratet) der Mitarbeiter erfragt. Das Resultat ist in der folgenden Tabelle festgehalten:

Person	1	2	3	4	5	6	7	8	9	10
Geschlecht	M	M	M	W	M	W	W	W	M	W
Zivilstand	L	L	V	V	V	L	L	L	L	V

Wir wählen zufällig eine Person. Wie gross ist die WSK, dass

- die Person männlich ist? $P(M) = \frac{5}{10} = \frac{1}{2}$ (Entspricht Personen 1,2,3,5,9)
- die Person verheiratet ist? $P(V) = \frac{4}{10} = \frac{2}{5}$ (Entspricht Personen 3,4,5,10)
- die Person männlich und verheiratet ist? $P(M \cap V) = \frac{2}{10} = \frac{1}{5}$ (Entspricht Personen 3,5)
- die Person männlich oder verheiratet ist? $P(M \cup V) = P(M) + P(V) - P(M \cap V) = \frac{1}{2} + \frac{2}{5} - \frac{1}{5} = \frac{7}{10}$ (Entspricht Personen 1,2,3,4,5,9,10)
- die Person weder männlich noch verheiratet ist? $P(\overline{M} \cap \overline{V}) = P(W \cap L) = \frac{3}{10}$ (Entspricht Personen 6,7,8)

5.3 Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit

Bei einem Pferderennen werden die Siegchancen von Pferd 1 und Pferd 2 auf jeweils 20% und die des Pferdes 3 auf 60% geschätzt. Die anderen mitlaufenden Pferde haben keine realistischen Chancen auf den Sieg. Es sei A_i das Ereignis: "Pferd i gewinnt". Wir haben also $P(A_1) = P(A_2) = 0.2$ und $P(A_3) = 0.6$. Der Jockey von Pferd 2 wurde dahingehend bestochen, dass er (bzw. sein Pferd) das Rennen auf keinen Fall gewinnt. Wir wissen also (oder zumindest der betreffende Jockey und der Bestecher), dass A_2 nicht eintreten wird, oder anders ausgedrückt, dass $\overline{A_2}$ eintritt. Es dürfte klar sein, dass das Eintreten von $\overline{A_2}$ die Siegchancen von Pferd 1 und 2 verändert, nämlich vergrössert (Ein Konkurrent fällt weg!). Sie steigen auf 25% bzw. 75%. (Würdet ihr das auch so sehen?) Was sich nicht ändert ist das Verhältnis der Siegchancen untereinander. Pferd 3 hat immer noch eine dreimal grössere Gewinnchance als Pferd 1.

Die Wahrscheinlichkeit für das Eintreten eines zufälligen Ereignisses A ändert sich also im Allgemeinen, wenn bekannt ist, dass ein anderes zufälliges Ereignis B (mit $P(B) \neq 0$) eintreten wird oder bereits eingetreten ist. Man nennt dies die *bedingte Wahrscheinlichkeit* von A unter der Bedingung B . Notation: $P(A|B)$.

Beispiel 5.30: Bedingte Wahrscheinlichkeit

Betrachten wir wieder die Situation aus Beispiel 5.21 i). Wie verändert sich die WSK des Ereignisses A , wenn wir schon wissen, dass Ereignis B eingetreten ist, die Augensumme also gerade ist? Das Ereignis B ist durch die folgende Menge von Elementarereignissen gegeben:

$$B = \{e_{11}, e_{13}, e_{15}, e_{22}, \dots, e_{64}, e_{66}\}.$$

Wenn B eingetreten ist, dann ist eines von 18 Elementarereignissen eingetreten (und nicht eines von 36). Von A haben aber nur die 4 Elemente $\{e_{46}, e_{64}, e_{55}, e_{66}\}$ eine gerade Augensumme, können also gleichzeitig mit B eintreten. Wir haben also gemäss klassischer Wahrscheinlichkeitsdefinition $P(A|B) = \frac{4}{18} = \frac{2}{9}$. Hier wird also sozusagen die Anzahl der für A und B günstigen Resultate ins Verhältnis gesetzt zur Anzahl der Elemente in B . In WSKen ist das aber genau $P(A \cap B)/P(B)$.

Der letzte Ausdruck wird zur Definition von bedingten WSKen verwendet.

Definition 5.31: Bedingte Wahrscheinlichkeit

Seien A und B Ereignisse aus einer Ereignismenge $E(S)$, sodass $P(B) > 0$. Die *bedingte Wahrscheinlichkeit* von A unter der Bedingung B ist definiert als

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

Wird die WSK des Eintretens eines Ereignisses A nicht vom Eintreten eines anderen Ereignisses B mit $P(B) \neq 0$ beeinflusst, nennt man A und B *stochastisch unabhängig*. Das ist aber genau dann der Fall, wenn $P(A|B) = P(A)$ ist, was wiederum zu $P(A \cap B) = P(A)P(B)$ äquivalent ist.

Teilt man $P(A \cap B) = P(A)P(B)$ durch $P(A)$ (vorausgesetzt $P(A) \neq 0$), erhält man auch $P(B|A) = P(B)$. Wenn also A unabhängig von B ist, ist automatisch auch B unabhängig von A . Das motiviert die folgende Definition.

Definition 5.32: Stochastische Unabhängigkeit

Seien A und B Ereignisse aus einer Ereignismenge $E(S)$. A und B heissen *stochastisch unabhängig*, falls $P(A \cap B) = P(A)P(B)$.

Beachtet bitte, dass diese Definition auch dann sinnvoll ist, wenn $P(A)$ oder/und $P(B)$ gleich 0 sind.

Bemerkung 5.33

Es gilt also nur bei Unabhängigkeit von A und B , dass $P(A \cap B) = P(A)P(B)$. Im allgemeinen hat man $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$!

Beispiele 5.34: Stochastische Unabhängigkeit

- i) In Beispiel 5.29 sind die Ereignisse M : "Person ist männlich." und V : "Person ist verheiratet." unabhängig, da

$$P(M \cap V) = \frac{1}{5} = \frac{1}{2} \cdot \frac{2}{5} = P(M) \cdot P(V).$$

- ii) Aus einer Urne, die 2 weisse Kugeln und 3 schwarze Kugeln enthält, werden nacheinander (Reihenfolge also wichtig!) zwei Kugeln gezogen. Wir betrachten die zwei Ereignisse

E : "Erste gezogene Kugel ist weiss." und F : "Zweite gezogene Kugel ist weiss."

In diesem Beispiel haben wir (die Bedeutung sollte klar sein):

$$S = \{ww, ws, sw, ss\}, \quad E = \{ww, ws\} \quad \text{sowie} \quad F = \{ww, sw\}.$$

Beachtet auch, dass wir uns nicht in einer klassischen Wahrscheinlichkeitssituation befinden, da die Elementarereignisse in S nicht gleichwahrscheinlich sind.

Wir wollen $P(E)$ und $P(F)$ bestimmen. Dazu müssen wir festlegen, wie gezogen wird.

- a) Mit Zurücklegen: (Ich führe die Überlegungen einmal detailliert aus. Der Rest sollte dann (hoffentlich) klar sein.)

$$P(E) = P(\{ww\} \cup \{ws\}) = P(\{ww\}) + P(\{ws\}),$$

da beide Ereignisse Elementarereignisse und somit sich gegenseitig ausschliessend, also unvereinbar, sind (vgl. Eigenschaft (W4)).

$$P(\{ww\}) = \frac{2}{5} \cdot \frac{2}{5} = \frac{4}{25}.$$

Da in der Urne 2 weisse und 3 schwarze, also insgesamt 5 Kugeln sind, ist die WSK dafür, dass beim ersten Ziehen w auftritt nach klassischer WSK gerade $\frac{2}{5}$. Da die gezogene Kugel wieder zurückgelegt wird, die Situation sich also nicht ändert, gilt das auch für w beim zweiten Ziehen. Dass $P(\{ww\})$ gerade das Produkt dieser zwei WSKen ist, folgt aus einem Prinzip analog dem Fundamentalprinzip der Kombinatorik.

Ähnlich erhalten wir

$$P(\{ws\}) = \frac{2}{5} \cdot \frac{3}{5} = \frac{6}{25} \quad \text{und damit}$$

$$P(E) = P(\{ww\}) + P(\{ws\}) = \frac{4}{25} + \frac{6}{25} = \frac{2}{5}.$$

Die gleiche WSK erhalten wir über analoge Betrachtungen für Ereignis F : $P(F) = \frac{2}{5}$.
 Wir haben dann

$$P(E \cap F) = P(\{ww\}) = \frac{4}{25} = \frac{2}{5} \cdot \frac{2}{5} = P(E) \cdot P(F).$$

In dieser Situation sind also E und F unabhängig.

b) Ohne Zurücklegen: Man erhält wieder:

$$P(E) = P(F) = \frac{2}{5} \quad (\text{Versucht bitte, das nachzuvollziehen.})$$

Aber

$$P(E \cap F) = P(\{ww\}) = \frac{2}{5} \cdot \frac{1}{4} \neq \frac{2}{5} \cdot \frac{2}{5} = P(E) \cdot P(F).$$

In dieser Situation sind also E und F nicht unabhängig, also abhängig.

5.3.1 Totale Wahrscheinlichkeit und Bayes' Theorem

Häufig tritt ein Ereignis A immer zusammen mit gewissen anderen Ereignissen ein.

Beispiel 5.35

Sei A das Ereignis "Person ist Raucher". Wir "Zerlegen" die Gesamtbevölkerung in drei Gruppen. (Die Zerlegung und alle Zahlenwerte sind willkürlich gewählt.)

H_1 : Personen jünger als 25 Jahre (Anteil 35% an Gesamtbevölkerung)

H_2 : Personen zwischen 25 Jahren und 50 Jahren (Anteil 45% an Gesamtbevölkerung)

H_3 : Personen älter als 50 Jahre (Anteil 20% and Gesamtbevölkerung)

Das wichtige an der Zerlegung ist: Jede Person gehört zu einer und nur einer Altersgruppe.

Das Ereignis A tritt nun entweder zusammen mit H_1 oder H_2 oder H_3 auf (ausschliessende "oder").

Wir kennen darüber hinaus den Anteil der Raucher in den verschiedenen Altersgruppen: in H_1 sind $\frac{3}{14}$ Raucher, in H_2 sind es $\frac{1}{2}$ und in H_3 sind es $\frac{2}{5}$.

Wir wollen folgende Frage beantworten: Was ist die WSK, dass eine zufällig aus der Gesamtbevölkerung herausgegriffene Person Raucher ist? Gesucht ist also $P(A)$. Dazu zerlegen wir die

“Rauchergruppe” in solche, die zu H_1 , H_2 bzw. H_3 gehören:

$$A = (A \cap H_1) \cup (A \cap H_2) \cup (A \cap H_3).$$

Da sich die Altersgruppen ausschliessen, sind auch die in die Vereinigungsbildung eingehenden Mengen paarweise disjunkt. Nach (W4) haben wir:

$$\begin{aligned} P(A) &= P(A \cap H_1) + P(A \cap H_2) + P(A \cap H_3) \\ &= P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) + P(A|H_3) \cdot P(H_3) \\ &= \frac{3}{14} \cdot 0.35 + \frac{1}{2} \cdot 0.45 + \frac{2}{5} \cdot 0.2 = 0.38. \end{aligned}$$

Das ist die totale WSK, dass eine zufällig aus der Bevölkerung herausgegriffene Person Raucher ist.

Allgemein gilt das folgende Resultat.

Sei H_1, H_2, \dots, H_n eine Zerlegung des Ereignisraumes S , d.h. paarweise disjunkte Mengen ($H_i \cap H_j = \emptyset$ für alle $i \neq j$), so dass $P(H_k) \neq 0$, $k = 1, 2, \dots, n$ und ihre Vereinigung $\bigcup_{k=1}^n H_k = S$ ergibt. (Die Mengen zerlegen also wie oben den Ereignisraum.). Eine solche Zerlegung wird auch *Partition* von S genannt. Dann gilt für jedes Ereignis $A \in E(S)$:

$$P(A) = P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) + \dots + P(A|H_n) \cdot P(H_n) = \sum_{k=1}^n P(H_k) P(A|H_k)$$

(Totale-Wahrscheinlichkeits-Formel)

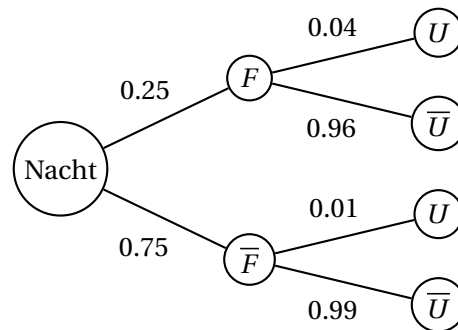
Beispiel 5.36: Totale Wahrscheinlichkeiten

Die WSK, dass eine Januarnacht in Regensdorf frostig ist, beträgt 0.25. In einer frostigen Nacht beträgt die WSK eines Unfalls in der Althardstrasse 0.04 und 0.01 in frostfreien Nächten. Wie gross ist die WSK, dass in der Nacht des 13. Januars ein Unfall in der Althardstrasse passiert?

$$\begin{aligned} P(\text{Unfall}) &= P(\text{Unfall}|\text{Frost}) \cdot P(\text{Frost}) + P(\text{Unfall}|\text{Frostfrei}) \cdot P(\text{Frostfrei}) \\ &= 0.04 \cdot 0.25 + 0.01 \cdot 0.75 = 0.0175. \end{aligned}$$

Dass in der Frage nach der Unfallwahrscheinlichkeit am 13. Januar gefragt wird, ist natürlich vollkommen irrelevant.

Es ist häufig nützlich, Wahrscheinlichkeiten in “mehrstufigen Zufallsprozessen” der besseren Übersicht wegen in Wahrscheinlichkeitsbäumen darzustellen. Für das letzte Beispiel sähe das ungefähr so aus (Bezeichnung müsste klar sein.):



Beachtet nochmal, dass es sich bei $0.04 = P(U|F)$, $0.96 = P(\bar{U}|F)$, $0.01 = P(U|\bar{F})$ und $0.99 = P(\bar{U}|\bar{F})$ um bedingte WSKen handelt.

Die WSK eines Pfades ist gleich dem Produkt der Einzelwahrscheinlichkeiten entlang dieses Pfades. Die WSK eines Ereignisses in einem Wahrscheinlichkeitbaum ist die Summe der WSK aller Pfade, die zum Ereignis führen. (Für $P(U)$ haben wir als WSK des Pfades nach U über F $0.25 \cdot 0.04$ und als WSK des Pfades nach U über \bar{F} $0.75 \cdot 0.01$. Es gilt demnach $P(U) = 0.01 + 0.0075 = 0.0175$.)

Was wäre $P(\bar{U})$, also die WSK, dass eine Januarnacht unfallfrei ist? Wie sieht der Wahrscheinlichkeitsbaum für unser Raucherbeispiel aus?

Kommen wir nochmal zum "Raucherbeispiel". Wir nehmen jetzt zusätzlich an, dass $P(A) \neq 0$. Wir wollen die WSK von H_i finden, wenn wir wissen dass A eingetreten ist, also die Frage klären: Gegeben eine Person ist Raucher, mit welcher WSK kommt sie aus einer der 3 Altersgruppen?

Dazu merken wir an, dass $P(H_i \cap A) = P(A)P(H_i|A) = P(H_i)P(A|H_i)$. Das ergibt

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}.$$

Drücken wir jetzt noch in dieser Formel $P(A)$ durch die Totale-Wahrscheinlichkeits-Formel aus, erhalten wir

$$\begin{aligned} P(H_i|A) &= \frac{P(H_i)P(A|H_i)}{P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) + \dots + P(A|H_n) \cdot P(H_n)} \\ &= \frac{P(H_i)P(A|H_i)}{\sum_{k=1}^n P(H_k)P(A|H_k)} \quad \textbf{(Bayes' Formel)}. \end{aligned}$$

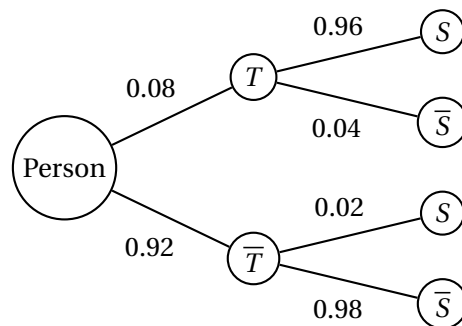
Die $P(H_i|A)$ heissen *a posteriori Wahrscheinlichkeiten*. (Wir schliessen von der Wirkung auf die Ursache!) Wir erhalten also z.B. $P(H_3|A) = 0.211$, d.h. wenn eine Person Raucher ist, dann kommt sie mit ca. 21% WSK aus der Gruppe der über 50-Jährigen.

Ein weiteres

Beispiel 5.37: Totale Wahrscheinlichkeit, Bayes' Formel

Nehmen wir an, dass 8% der Bevölkerung an Tuberkulose leiden. Es existiert ein Hauttest, der Tuberkulose in 96% der Menschen diagnostiziert, die Tuberkulose haben. Er zeigt aber auch in 2% der Personen die Krankheit an, die nicht an Tuberkulose leiden.

Wir wollen zunächst der besseren Übersicht halber einen Wahrscheinlichkeitsbaum zeichnen. Es stehe T bzw. \bar{T} für die Ereignisse "hat Tuberkulose" bzw. "hat keine Tuberkulose" und S bzw. \bar{S} für "positiver Hauttest" bzw. "negativer Hauttest". Dann haben wir:



Wir berechnen zunächst die WSK, dass der Hauttest positiv ausfällt. Mittels Totaler-WSK-Formel erhalten wir

$$P(S) = P(\bar{T})P(S|\bar{T}) + P(T)P(S|T) = 0.92 \cdot 0.02 + 0.08 \cdot 0.96 = 0.0952.$$

Wir interessieren uns jetzt für die WSK, dass eine Person keine Tuberkulose hat unter der Bedingung, dass der Hauttest positiv war, d.h. $P(\bar{T}|S)$ (falsches Positivresultat). Nach Bayes' Formel gilt:

$$P(\bar{T}|S) = \frac{P(\bar{T})P(S|\bar{T})}{P(\bar{T})P(S|\bar{T}) + P(T)P(S|T)} = \frac{0.92 \cdot 0.02}{0.92 \cdot 0.02 + 0.08 \cdot 0.96} = 0.193.$$

Dieses Ereignis tritt also in fast 20% der Fälle ein und macht den Hauttest vollkommen wertlos. (Das wird den Hersteller aber wenig bis gar nicht kümmern.)

Zum gleichen Resultat kommt man unter Zuhilfenahme des WSK-Baums.

$$P(\bar{T}|S) = \frac{\text{Pfadwahrscheinlichkeit des Pfades nach } S \text{ über } \bar{T}}{\text{Summe der WSK aller Pfade, die nach } S \text{ führen}}.$$

Zum Abschluss kommen wir noch mal auf unser Unfallbeispiel zurück.

Beispiel 5.38: Bayes' Formel

Wie gross ist die Wahrscheinlichkeit, dass die Nacht des 13. Januars in Regensdorf frostig war, gegeben dass in der Althardstrasse ein Unfall passierte?

Antwort: $P(F|U) = \frac{P(U|F) \cdot P(F)}{P(U)} = \frac{0.04 \cdot 0.25}{0.0175} = 0.571$.

6 Zufallsvariablen

Achtung: Die Beschreibung von Zufallsvariablen (ZV) ähnelt der von statistischen Variablen. Oft braucht man nur "relative Häufigkeiten" durch "Wahrscheinlichkeiten" zu ersetzen. Das gilt insbesondere im Kontext von diskreten ZV.

6.1 Zufallsvariablen und ihre Verteilungsfunktionen

Wir betrachten einen sogenannten Wahrscheinlichkeitsraum $[S, E(S), P(\cdot)]$ für ein zufälliges Experiment. Hier bezeichnet E wieder die Menge der Elementarereignisse, $E(S)$ den Ereignisraum und P ein Wahrscheinlichkeitsmass auf $E(S)$, d.h. eine Funktion die jedem Ereignis $A \in E(S)$ seine Wahrscheinlichkeit $P(A)$ zuordnet.

Eine *Zufallsvariable* X ist nichts anderes als eine Funktion, die als Definitionsgebiet S und als Funktionswerte reelle Zahlen hat. Also

$$X : S \longrightarrow \mathbb{R}.$$

Eine ZV ordnet also den Elementarereignissen eines Zufallsexperiments reelle Zahlen zu.

Beispiel 6.1: Diskrete Zufallsvariable

Wir betrachten das dreimalige Werfen einer Münze. Die Menge der Elementarereignisse für dieses Experiment ist die folgende Menge:

$$\begin{aligned} S &= \{KKK, KKZ, KZK, ZKK, KZZ, ZKZ, ZZK, ZZZ\} \\ &= \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\} \end{aligned}$$

Eine mögliche ZV X wäre die Anzahl der K im Versuchsausgang, also z.B. $X(e_1) = 3$, $X(e_2) = 2$, etc. Der Wertebereich dieser ZV ist die Menge $W = \{0, 1, 2, 3\}$.

Beispiel 6.2: Stetige Zufallsvariable

Wir wählen zufällig einen Punkt e aus dem Intervall $I = \{e \in \mathbb{R} : -1 \leq e \leq 1\}$. Eine mögliche ZV

wäre der Abstand von e von 1, also z.B. $X(-0.5) = 1.5$, $X(0.3) = 0.7$, etc. Der Wertebereich dieser ZV ist $W = [0, 2]$.

Zufallsvariablen, deren Wertebereiche endlich oder abzählbar unendlich sind, heissen *diskret* (erstes Bsp. oben). Ihre Wertebereiche bestehen also aus “isolierten” Punkten. Zufallsvariablen, deren Wertebereiche ganze Intervalle umfassen, nennt man *stetig* (zweites Bsp. oben).

Bevor wir zur Beschreibung von ZV durch ihre Verteilungsfunktionen kommen, möchte ich einige spezielle “Ereignisse” bzgl. ZV und deren Wahrscheinlichkeiten einführen. Beachtet bitte nochmal, dass Ereignisse wie immer Teilmengen des Ereignisraums S sind. Gewöhnt euch unbedingt an diese Notationen!! Spielt ein bisschen mit den ZV, die wir schon eingeführt haben, herum.

Für reelle Zahlen a, b, x verstehen wir unter:

Ereignis	Bedeutung	Wahrscheinlichkeit
$X = a$	$\{e \in S : X(e) = a\}$	$P(\{e \in S : X(e) = a\})$
$a \leq X \leq b$	$\{e \in S : a \leq X(e) \leq b\}$	$P(\{e \in S : a \leq X(e) \leq b\})$
$X \leq x$	$\{e \in S : X(e) \leq x\}$	$P(\{e \in S : X(e) \leq x\})$

Zur Verdeutlichung vielleicht die dritte Zeile nochmal in Worten: Für eine gegebene reelle Zahl x bezeichnet die Schreibweise $X \leq x$ die Menge aller Elementarereignisse $e \in S$, deren Funktionswerte $X(e)$ unter X kleiner gleich x sind.

Ereignisse der Form $a < X \leq b$, $a \leq X < b$, $a < X < b$, $X < x$, $X \geq x$ bzw. $X > x$ werden analog definiert.

Definition 6.3: Verteilungsfunktion

Definition: Die *Verteilungsfunktion* (VF) F_X einer ZV X über einem Wahrscheinlichkeitsraum $[S, E(S), P(\cdot)]$ ist definiert durch

$$F_X(x) = P(X \leq x) \text{ für alle reellen Zahlen } x.$$

Bemerkungen 6.4

- Der Definitionsbereich für Verteilungsfunktionen ist also grundsätzlich \mathbb{R} . Genauer gilt $F_X : \mathbb{R} \rightarrow [0, 1]$.
- Will man den Funktionswert $F_X(x)$ der Verteilungsfunktion einer ZV X für ein gewisses $x \in \mathbb{R}$ berechnen, muss man sich fragen: Was ist die WSK der Elementarereignisse e , für die $X(e) \leq x$ gilt?
- Sei X eine diskrete ZV mit Wertebereich $W = \{x_1, x_2, \dots, x_n\}$ bzw. $W = \{x_1, x_2, \dots\}$, wobei wir annehmen wollen, dass $x_1 < x_2 < \dots$. Für diskrete ZV ist der Graph der VF immer eine Trep-

penfunktion, die bei 0 beginnt und in jedem $x_i \in W$ jeweils um den Wert $P(X = x_i)$ springt.
 Vergleiche statistische Variablen!!

iv) Die Eigenschaften einer VF sind analog denen einer empirischen VF.

Beispiel 6.5: Verteilungsfunktion einer diskreten Zufallsvariablen

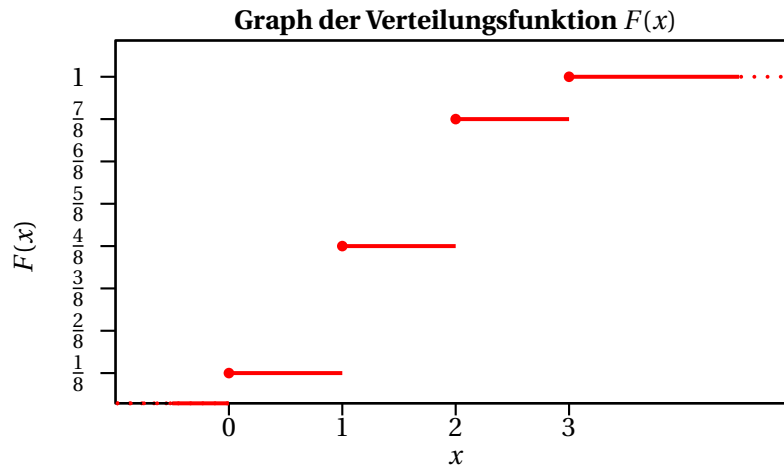
Wir wollen die VF der ZV aus dem Münzwurfbeispiel bestimmen.

$$F(x) = \begin{cases} 0, & x < 0, & \text{da für diese } x \{e \in S : X(e) \leq x\} = \emptyset \\ \frac{1}{8}, & x = 0, & \text{da für diese } x \{e \in S : X(e) \leq x\} = \{e_8\} \\ \frac{1}{8}, & 0 < x < 1, & \text{da für diese } x \{e \in S : X(e) \leq x\} = \{e_8\} \\ \frac{4}{8}, & x = 1, & \text{da für diese } x \{e \in S : X(e) \leq x\} = \{e_5, e_6, e_7, e_8\} \\ \frac{4}{8}, & 1 < x < 2, & \text{da für diese } x \{e \in S : X(e) \leq x\} = \{e_5, e_6, e_7, e_8\} \\ \frac{7}{8}, & x = 2, & \text{da für diese } x \{e \in S : X(e) \leq x\} = \{e_2, e_3, e_4, e_5, e_6, e_7, e_8\} \\ \frac{7}{8}, & 2 < x < 3, & \text{da für diese } x \{e \in S : X(e) \leq x\} = \{e_2, e_3, e_4, e_5, e_6, e_7, e_8\} \\ \frac{8}{8}, & x = 3, & \text{da für diese } x \{e \in S : X(3) \leq x\} = S \\ 1, & x > 3, & \text{da für diese } x \{e \in S : X(e) \leq x\} = S \end{cases}$$

Wir haben also für jedes reelle x den Funktionswert $F(x)$ explizit angegeben.

Verbale Erklärung für den 5. und 6. Wert: Für den 5. Wert gibt man sich ein beliebiges x zwischen 1 und 2 vor und fragt sich: Für welche Elementarereignisse e ist $X(e) \leq x$, ist also die Zahl der K ein Wert kleiner gleich einer Zahl zwischen 1 und 2? Antwort siehe oben. Für den 6. Wert nimmt man $x = 2$ und fragt sich wieder: Für welche Elementarereignisse e ist $X(e) \leq 2$, ist also die Zahl der K kleiner gleich 2? Antwort siehe oben.

Die Graphik der VF hat die für diskrete ZV typische Treppenform mit Sprüngen in 0, 1, 2, 3 um $P(X = 0)$, $P(X = 1)$, $P(X = 2)$ bzw. $P(X = 3)$.



6.2 Diskrete Zufallsvariablen

Wir hatten diskrete ZV dadurch charakterisiert, dass ihr Wertebereich W entweder eine endliche oder abzählbar unendliche Menge ist. Für eine diskrete ZV X definieren wir die *Wahrscheinlichkeitsmassenfunktion* (oder *Massenfunktion* oder *Wahrscheinlichkeitsfunktion*) $f(x)$ wie folgt (Vgl. relative Häufigkeitsfunktion von statistischen Variablen!):

$$f(x) = \begin{cases} P(X = x) & \text{falls } x \in W \\ 0 & \text{sonst} \end{cases}$$

Es gilt also $f : \mathbb{R} \rightarrow [0, 1]$. Darüber hinaus haben wir $\sum_{x_i \in W} f(x_i) = 1$.

Der Zusammenhang zwischen Verteilungsfunktion $F(x)$ und Massenfunktion $f(x)$ einer diskreten Zufallsvariablen X stellt sich wie folgt dar:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x, x_i \in W} f(x_i) = \sum_{x_i \leq x, x_i \in W} P(X = x_i).$$

Um also einen Funktionswert $F(x)$ der Verteilungsfunktion zu bestimmen, addiert man ganz einfach die Wahrscheinlichkeiten aller Funktionswerte x_i der ZV X auf, für die $x_i \leq x$ gilt.

Umgekehrt gilt

$$f(x) = F(x) - \lim_{\Delta x \downarrow 0} F(x - \Delta x).$$

Beispiel 6.6: Massenfunktion

Für unser Beispiel hat die Massenfunktion folgendes Aussehen:

$$f(x) = \begin{cases} P(X=x) = \frac{1}{8} & \text{falls } x = 0 \\ P(X=x) = \frac{3}{8} & \text{falls } x = 1 \text{ oder } x = 2 \\ P(X=x) = \frac{1}{8} & \text{falls } x = 3 \\ 0 & \text{sonst} \end{cases}$$

Die Massenfunktion ist also, wie die Häufigkeitsfunktion im Kontext von Merkmalen, für **alle** $x \in \mathbb{R}$ definiert und nicht nur in den Ausprägungen, die zum Wertebereich von X gehören.

Wir wollen einige WSK bestimmen:

$$P(0 \leq X \leq 2) = f(0) + f(1) + f(2) = \frac{7}{8}.$$

$$P(0 < X \leq 2) = f(1) + f(2) = \frac{6}{8} = \frac{3}{4}.$$

$$P(X \leq 1.5) = F(1.5) = f(0) + f(1) = \frac{4}{8} = \frac{1}{2}.$$

Beachtet, dass es in der Regel einen Unterschied macht, ob man $<$ oder \leq bzw. $>$ oder \geq in der Berechnung der WSKen hat! Man muss im Kontext von diskreten ZV dbzgl. sehr achtsam sein.

Bemerkung 6.7: Wichtig!

$$P(a \leq X \leq b) = F(b) - F(a-1), \quad P(a < X \leq b) = F(b) - F(a)$$

6.3 Stetige Zufallsvariablen

Eine Zufallsvariable X heisst *stetig*, falls es eine sogenannte *Wahrscheinlichkeitsdichtefunktion* (oder auch nur *Dichtefunktion* genannt) mit folgenden Eigenschaften gibt:

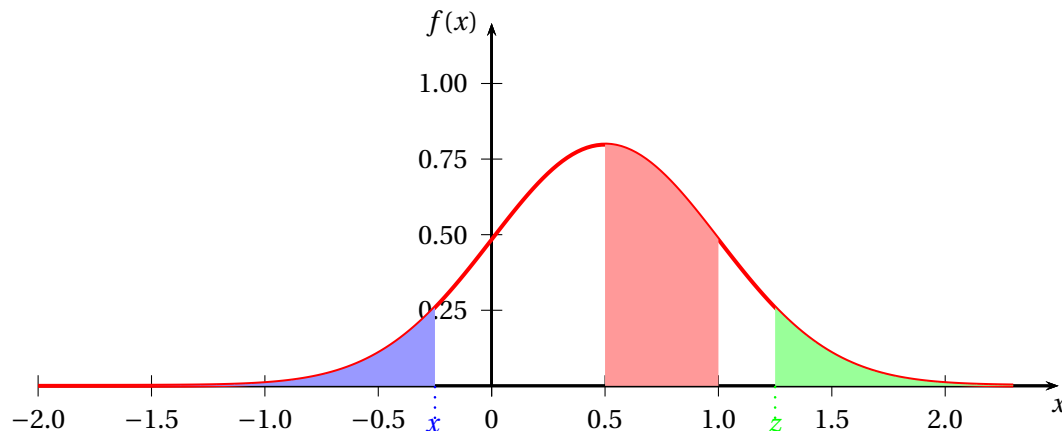
$$(1) \quad f(x) \geq 0 \qquad (2) \quad \int_{-\infty}^{\infty} f(x) dx = 1,$$

sodass

$$P(X \leq x) = F(x) = \int_{-\infty}^x f(t) dt.$$

Die Verteilungsfunktion ist also eine Stammfunktion der Dichtefunktion $f(x)$, d.h. $F'(x) = f(x)$.

Die folgende Graphik zeigt ein Beispiel.



Bemerkungen 6.8

- i) Eine Dichtefunktion hat also nie negative Funktionswerte.
- ii) Wer sich etwas mit Integration auskennt, weiss, dass für nichtnegative Funktionen $f(x)$ Integrale der Form $\int_a^b f(t) dt$ (a kann hier auch $-\infty$ und $b = \infty$ sein.) eine Flächeninterpretation haben. Das Integral ist gleich der Fläche, die vom Graphen der Funktion und der horizontalen Achse über dem Intervall $[a, b]$ eingeschlossen wird. Eigenschaft (2) sagt also nichts anderes, als dass die Fläche unter einer Dichtefunktion immer 1 ist. Das kann übersetzt werden in: $P(X \in \mathbb{R}) = 1$. Die Wahrscheinlichkeit, dass die stetige ZV X Werte in \mathbb{R} annimmt, beträgt 1.
- iii) Für stetige ZV X mit Dichtefunktion $f(x)$ werden die Funktionswerte der Verteilungsfunktion mittels Integration berechnet: $F(x) := P(X \leq x) = \int_{-\infty}^x f(t) dt$. Auch dieses Integral (also $F(x)$) hat eine Flächeninterpretation. Siehe blaue Fläche oben. Sie ist die Fläche links von x unter dem Graphen von $f(x)$.
- iv) Auch andere Wahrscheinlichkeiten können flächenmässig interpretiert werden. Die rote Fläche oben entspricht z.B. $P(0.5 \leq X \leq 1) = F(1) - F(0.5)$, die grüne Fläche $P(X > z) = 1 - P(X \leq z) = 1 - F(z)$.
- v) Nachdem, was wir gesagt haben, entspricht der WSK $P(X = a) = P(a \leq X \leq a)$ die “Fläche” unter $f(x)$ über a . Das ist aber eine Gerade, die keine Fläche besitzt. Es gilt also für stetige ZV stets $P(X = a) = 0$. Diese Tatsache kann man sich auch anders plausibel machen. Wenn $P(X = a) = 0$ nicht gelten würde, dann würden sich die WSKen der Elementarereignisse zu ∞ aufaddieren. Es gibt also im Kontext von stetigen ZV Ereignisse mit WSK 0, die durchaus eintreten können. Für diskrete ZV ist das nicht möglich.
- vi) Keine Angst. Ihr werdet nicht Integrieren müssen. Wir werden WSKen im Kontext stetiger

ZV anders bestimmen.

Achtung: Für **stetige** ZV macht es **keinen** Unterschied, ob man in WSKen \leq oder $<$ bzw. \geq oder $>$ hat! Vergleicht das mit den WSKen nach Beispiel 6.6!

Es ist wichtig zu verstehen, wie man mittels Dichte-/Verteilungsfunktion im Kontext von stetigen ZV Wahrscheinlichkeiten berechnet!! Hier also nochmal kurz zusammengefasst:

$$P(X \leq x) = F(x) = \int_{-\infty}^x f(t) dt \quad P(X > x) = 1 - P(X \leq x) = 1 - \int_{-\infty}^x f(t) dt = 1 - F(x)$$

$$P(a \leq X \leq b) = \int_a^b f(t) dt = F(b) - F(a).$$

Ich hatte zwar gesagt, dass ihr nicht mit Integration konfrontiert werdet, ich möchte euch aber zumindest an einem Beispiel zeigen, wie das funktionieren würde.

Beispiel 6.9

Gegeben sei eine ZV X mit Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{falls } x \leq 0 \\ 4x^3 - 3x^4 & \text{falls } 0 < x \leq 1 \\ 1 & \text{falls } x > 1 \end{cases}$$

Leiten wir diese Funktion in den jeweiligen Bereichen ab, erhalten wir die Dichtefunktion f :

$$f(x) = \begin{cases} 12x^2 - 12x^3 & \text{falls } 0 < x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Somit erhalten wir z.B. für die WSK $P(0.4 \leq X \leq 0.7)$

$$P(0.4 \leq X \leq 0.7) = \int_{0.4}^{0.7} f(x) dx = \int_{0.4}^{0.7} 12x^2 - 12x^3 dx = (4x^3 - 3x^4)|_{0.4}^{0.7} =$$

$$F(0.7) - F(0.4) = 0.4725.$$

Mathematische Folklore: Die Massenfunktion ist für diskrete ZV was die Dichtefunktion für stetige ZV ist.

6.4 Der Erwartungswert einer Zufallsvariablen

Erwartungswerte werden auch *Mittelwerte*, *mathematische Erwartung* oder *1. Moment* genannt.

6.4.1 Der Erwartungswert einer diskreten Zufallsvariablen

Sei X eine diskrete ZV mit Wertebereich W . Wir definieren den *Erwartungswert* EX wie folgt:

$$EX = \sum_{x_i \in W} x_i f(x_i) = \sum_{x_i \in W} x_i P(X = x_i)$$

In Worten: Um den Erwartungswert einer diskreten Zufallsvariablen zu berechnen, multipliziert man jede "Ausprägung" der ZV mit ihrer Wahrscheinlichkeit und summiert diese Produkte.

Ausser dem Symbol EX sieht man auch μ , μ_X oder $\mu(X)$ als Bezeichnung für den Erwartungswert von X .

Beispiel 6.10: Erwartungswert einer diskreten Zufallsvariablen

Für die ZV aus unserem Würfelbeispiel haben wir (in diesem Beispiel war $W = \{0, 1, 2, 3\}$):

$$EX = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$

Das hätten wir wahrscheinlich auch intuitiv erwartet.

6.4.2 Der Erwartungswert einer stetigen Zufallsvariablen

Ist X eine stetige ZV mit Dichtefunktion $f(x)$, berechnet sich der Erwartungswert als

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

6.4.3 Eigenschaften des Erwartungswerts

Sei X eine ZV mit Erwartungswert μ und g eine Funktion. Dann gilt:

- i) $E(X - \mu) = 0$.
- ii) $E(g(X)) = \sum_{x_i \in W} g(x_i) f(x_i)$ (falls X diskret) bzw. $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$ (falls X stetig).
- iii) $E(a) = a$, d.h. der EW einer Konstanten a ist gleich dieser Konstanten.
- iv) $E(b \cdot g(X)) = b \cdot E(g(X))$ für jede Konstante b , insbesondere ist $E(bX) = bE(X)$.
- v) $E(g_1(X) + g_2(X)) = E(g_1(X)) + E(g_2(X))$, insbesondere ist $E(X_1 + X_2) = E(X_1) + E(X_2)$.
- vi) $E(X + a) = E(X) + a$.

Bemerkung 6.11: Eigenschaften des Erwartungswerts

Bei einer linearen Transformation einer ZV X transformiert sich also der EW dementsprechend. $X \longrightarrow Z = aX + b \implies EX \longrightarrow EZ = aEX + b$.

Beispiel 6.12: Eigenschaften des Erwartungswerts

In unserem Würfelbeispiel hatten wir die ZV X mit Wertebereich $W_X = \{0, 1, 2, 3\}$. Ihr EW war $\frac{3}{2}$. Wir bilden die ZV $Z = 2X - 2$. Diese ZV hat als Wertebereich $W_Z = \{-2, 0, 2, 4\}$. Diese Werte treten mit den gleichen WSKen auf, wie die entsprechenden Werte der ZV X . Nach dem oben Gesagten gilt $EZ = 2EX - 2 = 1$. Überprüft das!

6.5 Die Varianz einer Zufallsvariablen

Die Varianz einer ZV wird auch 2. (zentrales) Moment genannt.

6.5.1 Die Varianz einer diskreten Zufallsvariablen

Sei X eine diskrete ZV mit Erwartungswert μ und Wertebereich W . Wir definieren die *Varianz* $\text{var } X$ wie folgt:

$$\text{var } X = \sum_{x_i \in W} (x_i - \mu)^2 f(x_i) = \sum_{x_i \in W} (x_i - \mu)^2 P(X = x_i) = E(X - \mu)^2$$

In Worten: Um die Varianz einer diskreten Zufallsvariablen zu berechnen, multipliziert man die quadratische Abweichung $(x_i - \mu)^2$ jeder "Ausprägung" x_i der ZV vom Erwartungswert mit der Wahrscheinlichkeit der "Ausprägung" und summiert diese Produkte.

Ausser dem Symbol $\text{var } X$ sieht man auch σ^2 , σ_X^2 , $\sigma^2(X)$ oder $\text{Var}(X)$ als Bezeichnung für die Varianz von X .

Beispiel 6.13: Varianz einer diskreten Zufallsvariablen

Für unser Würfelbeispiel haben wir:

$$\text{var } X = \left(0 - \frac{3}{2}\right)^2 \cdot \frac{1}{8} + \left(1 - \frac{3}{2}\right)^2 \cdot \frac{3}{8} + \left(2 - \frac{3}{2}\right)^2 \cdot \frac{3}{8} + \left(3 - \frac{3}{2}\right)^2 \cdot \frac{1}{8} = \frac{3}{4}.$$

6.5.2 Die Varianz einer stetigen Zufallsvariablen

Ist X eine stetige ZV mit Dichtefunktion $f(x)$ und $EX = \mu$, berechnet sich die Varianz als

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

Die (positive) Wurzel aus der Varianz einer ZV X , $\sqrt{\text{var } X} = \sigma_X$, wird als *Standardabweichung* von X bezeichnet.

6.5.3 Eigenschaften der Varianz

Auch für die Varianz gelten gewisse Rechenregeln, die ich hier kurz zusammenfassen will: Sei X eine ZV mit Erwartungswert μ und b eine reelle Zahl. Dann gilt:

- i) $\text{var}(a) = 0$ (Eine konstante ZV streut nicht!) .
- ii) $\text{var}(X + a) = \text{var } X$ (Wird eine ZV um einen konstanten Betrag verschoben, ändert sich die Varianz nicht.) .
- iii) $\text{var}(bX) = b^2 \text{var } X$ bzw. $\sigma_{bX} = |b| \sigma_X$ (Ein konstanter Faktor wird also mit dem Quadrat aus der Varianz “herausgezogen”)! .
- iv) $\text{var } X = E(X^2) - \mu^2$.
- v) $\text{var } X = E(X - b)^2 - (\mu - b)^2$ (Das ist der sogenannte Steinersche Verschiebungssatz. 4. ist davon ein Spezialfall ($b = 0$).) .
- vi) $\text{var}(bX + a) = b^2 \text{var } X$ (Das ist eine Zusammenfassung von Eigenschaften 2. und 3.) .

Achtung: Es gilt also i.A. nicht $\text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2)$!

6.6 Standardisieren von Zufallsvariablen

Sei X eine ZV mit $EX = \mu$ und Standardabweichung σ . Wir wollen X so modifizieren, dass wir eine ZV Z erhalten mit $EZ = 0$ und $\text{var } Z = 1$. Das wird im weiteren Verlauf des Moduls aus verschiedenen Gründen wichtig sein.

Wir definieren die ZV Z durch $Z := \frac{X - \mu}{\sigma}$. Ich behaupte, dass Z die geforderten Eigenschaften hat. Versucht bitte nachzuvollziehen, welche allgemeinen Eigenschaften des EW ich jeweils anwende.

$$EZ = E\left(\frac{X - \mu}{\sigma}\right) = E\left(\frac{X}{\sigma}\right) + E\left(\frac{-\mu}{\sigma}\right) = \frac{1}{\sigma} E(X) + \left(\frac{-\mu}{\sigma}\right) = \frac{1}{\sigma} \mu + \left(\frac{-\mu}{\sigma}\right) = 0 .$$

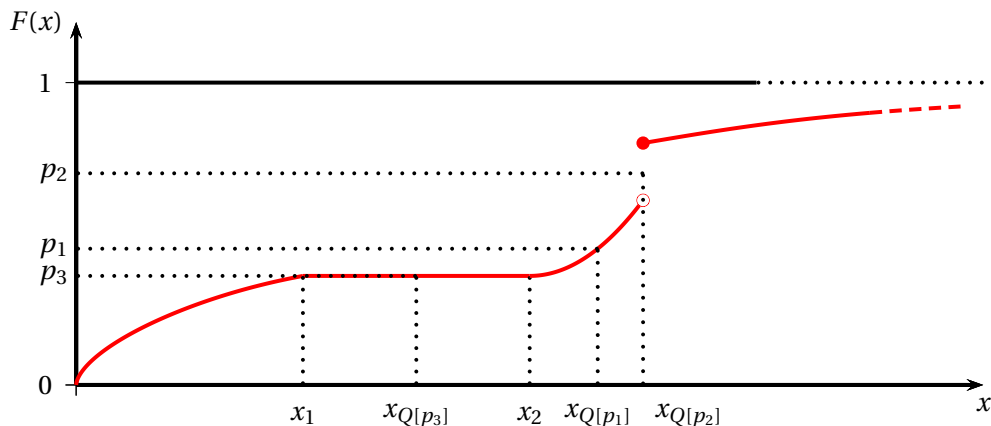
$$\text{var } Z = E\left(\frac{X - \mu}{\sigma} - 0\right)^2 = E\left(\frac{X^2 - 2\mu X + \mu^2}{\sigma^2}\right) = \frac{1}{\sigma^2} (E(X^2) - \mu^2) = 1 .$$

6.7 Quantil, Median und Modus (Kein Modulstoff!)

Sei p eine reelle Zahl mit $0 < p < 1$ und X eine Zufallsvariable. Eine Zahl $x_{Q[p]}$ mit der Eigenschaft, dass sowohl $P(X \leq x_{Q[p]}) = F(x_{Q[p]}) \geq p$ als auch $P(X \geq x_{Q[p]}) \geq 1 - p$ heisst *p-Quantil* der ZV X .

Für $p = \frac{1}{2}$ erhält man als Spezialfall den *Median* oder *Zentralwert* der ZV X . Der Median ist also wieder nichts anderes als das 0.5- oder auch 50%-Quantil.

Die drei typischen Situationen sind in der folgenden Graphik dargestellt.



- i) Es gibt nur **eine** q mit $p = F(q)$. Dann gilt $x_{Q[p]} = q$. (z.B. $p = p_1$ in Graphik)
Dieser Fall tritt z.B. immer dann ein, wenn die VF F (zumindest lokal) eine Umkehrfunktion F^{-1} besitzt (Das ist z.B. dann der Fall, wenn F (zumindest lokal) streng monoton wachsend ist.). Dann gilt selbstverständlich $x_{Q[p]} = F^{-1}(p)$. Das kann aber nur für stetige ZV passieren.
- ii) p fällt genau in einen Sprung von F (z.B. $p = p_2$ in Graphik). Dann wird $x_{Q[p]}$ gewählt wie in der Graphik gezeigt.
- iii) p fällt mit einer Stufenhöhe zusammen (z.B. $p = p_3$ in Graphik). Dann sind nach unserer Definition eigentlich alle q mit $p = F(q)$ p -Quantile. Um Eindeutigkeit zu erzielen, wählt man oft das arithmetische Mittel: $x_{Q[p_3]} = \frac{x_1 + x_2}{2}$.

Bemerkung 6.14: Quantile einer Zufallsvariablen

- i) Die Fälle ii) und iii) treten vor allem bei diskreten Zufallsvariablen auf, da die Graphen ihrer Verteilungsfunktionen Treppenfunktionen sind.
- ii) Wie im Fall von statistischen Variablen teilen auch die Quantile im Kontext von ZV den Wertebereich dieser ZV in Teilbereiche. Ist $x_{Q[p]}$ das p -Quantil einer ZV X , dann beträgt die WSK, dass X Werte $\leq x_{Q[p]}$ annimmt ca. $p \cdot 100\%$ und entsprechend die WSK für Werte $> x_{Q[p]}$ ca. $(1 - p) \cdot 100\%$.

Wie im Kontext von statistischen Variablen, bekommen bestimmte Quantile spezielle Bezeichnungen:

- $p = 0.25, 0.5$ oder $0.75 \rightarrow$ Quartile

- $p = 0.1, 0.2, \dots, 0.9 \rightarrow$ Dezile
- $p = 0.01, 0.02, \dots, 0.99 \rightarrow$ Perzentile

Zum Schluss kommen wir noch zum Begriff des *Modus* einer Zufallsvariablen. Eine reelle Zahl \bar{x}_{Mod} ist Modus einer Zufallsvariablen, falls \bar{x}_{Mod} Maximum der Wahrscheinlichkeitsfunktion (diskreter Fall) bzw. der Dichtefunktion ist (stetiger Fall). Der Modus muss nicht eindeutig bestimmt sein. Für die ZV aus dem Münzwurfbeispiel haben wir z.B. $\bar{x}_{\text{Mod}} = 1$ und $\bar{x}_{\text{Mod}} = 2$.

Beispiel 6.15: Quantile einer diskreten Zufallsvariablen

Nochmal zum Münzwurfbeispiel. Falls ihr euch die Verteilungsfunktion dieser ZV ansieht, dann ergibt sich sofort:

$$x_{Q[p]} = \begin{cases} 0 & , \text{für } p \in (0, \frac{1}{8}) \\ 0.5 & , \text{für } p = \frac{1}{8} \\ 1 & , \text{für } p \in (\frac{1}{8}, \frac{4}{8}) \\ 1.5 & , \text{für } p = \frac{4}{8} \\ 2 & , \text{für } p \in (\frac{4}{8}, \frac{7}{8}) \\ 2.5 & , \text{für } p = \frac{7}{8} \\ 3 & , \text{für } p \in (\frac{7}{8}, 1) \end{cases}$$

Für $x_{Q[\frac{1}{8}]}$, $x_{Q[\frac{4}{8}]}$ und $x_{Q[\frac{7}{8}]}$ beachtet bitte nochmal (iii) oben. Ich habe hier das arithmetische Mittel genommen. Mit dieser Festlegung gilt für den Median $x_{Q[0.5]} = 1.5$.

Beispiel 6.16: Quantile einer stetigen Zufallsvariablen

Wir wollen die p -Quantile der Exponentialverteilung bestimmen. Die Verteilungsfunktion $F_e(x|\lambda)$ dieser Verteilung ist gegeben durch

$$F_e(x|\lambda) = \begin{cases} 0 & , \text{für } x < 0 \\ 1 - e^{-\lambda x} & , \text{für } x \geq 0 \end{cases}$$

für einen gewissen Parameter $\lambda > 0$. Im relevanten Bereich $x \geq 0$ ist diese Funktion streng monoton wachsend, es gibt also zu jedem vorgegebenen $p \in (0, 1)$ ein eindeutig bestimmtes q mit $F_e(q|\lambda) = p$. Dieses q berechnet sich folgendermassen.

$$F_e(q|\lambda) = p \iff 1 - e^{-\lambda q} = p \iff e^{-\lambda q} = 1 - p \iff -\lambda q = \ln(1 - p) \iff q = -\frac{\ln(1 - p)}{\lambda}.$$

Das p -Quantil ist also gegeben durch $x_{Q[p]} = -\frac{\ln(1 - p)}{\lambda}$. Insbesondere gilt für den Median

$$x_{Q[0.5]} = -\frac{\ln \frac{1}{2}}{\lambda} = -\frac{\ln 1 - \ln 2}{\lambda} = \frac{\ln 2}{\lambda}. \text{ Logarithmengesetze!!}$$

7 Spezielle Verteilungen

Diskrete ZV $X \Rightarrow$ Massenfunktion f :

$$f(x) = \begin{cases} P(X = x) & , \text{ falls } x \in \text{WB von } X \\ 0 & , \text{ sonst.} \end{cases}$$

Stetige ZV $X \Rightarrow$ Dichtefunktion f . Dann

$$F(x) := P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

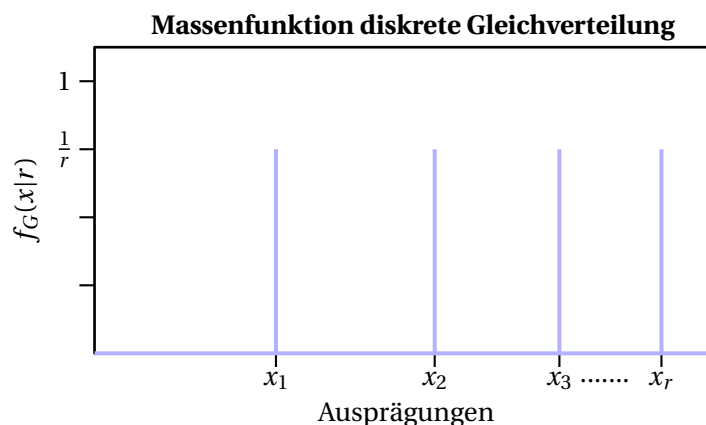
Gewisse Klassen von ZV treten in Anwendungen immer wieder auf und werden durch die Angabe ihrer Massenfunktion (diskreter Fall) oder Dichtefunktion (stetiger Fall) beschrieben. Diese hängen normalerweise von gewissen Parametern ab.

7.1 Diskrete Verteilungen

7.1.1 Gleichförmige Verteilung oder Gleichverteilung (diskret)

Die Gleichverteilung beschreibt Situationen, in denen die ZV X_G nur endlich viele Werte x_1, x_2, \dots, x_r (häufig die Werte $1, 2, \dots, r$) annehmen kann, deren Auftreten alle gleichwahrscheinlich sind. Es gilt also $P(X_G = x_1) = P(X_G = x_2) = \dots = P(X_G = x_r)$. Typische Situationen sind wieder das Würfeln mit einem fairen Würfel oder das Werfen einer fairen Münze. Da die Summe dieser WSKen 1 ergeben muss, folgt für diese konstante WSK: $P(X_G = x_i) = \frac{1}{r}$ für alle $i = 1, \dots, r$. Die Massenfunktion einer gleichverteilten ZV ist demnach gegeben durch

$$f_G(x|r) = \begin{cases} \frac{1}{r} & , \text{ falls } x = x_1, x_2, \dots, x_r \\ 0 & , \text{ sonst.} \end{cases}$$



Wir werden in einfachen Fällen versuchen, jeweils Erwartungswert und Varianz als wichtige Charakteristika von ZV explizit zu berechnen. Für gleichverteilte ZV ist das relativ einfach.

$$EX_G = \sum_{i=1}^r x_i f_G(x_i|r) = \frac{1}{r} \sum_{i=1}^r x_i.$$

Bei der Berechnung des EW einer gleichverteilten ZV werden also alle möglichen Funktionswerte von X aufaddiert und das Ergebnis durch deren Anzahl geteilt.

Für den Spezialfall $x_1 = 1, x_2 = 2, \dots, x_r = r$ liefert das

$$EX_G = \frac{1}{r} \frac{r(r+1)}{2} = \frac{r+1}{2}.$$

Bei der letzten Berechnung wurde benutzt: $\sum_{i=1}^r i = 1 + 2 + \dots + r = \frac{r(r+1)}{2}$.

Im allgemeinen Fall berechnet man die Varianz gemäss Formel. Im Fall $x_1 = 1, x_2 = 2, \dots, x_r = r$ erhält man speziell

$$\text{var } X_G = \frac{r^2 - 1}{12}.$$

7.1.2 Bernoulli-Verteilung

Bernoulli-Experiment: Nur das Eintreten eines Ereignisses A oder dessen Nichteintreten interessiert (d.h. das komplementäre Ereignis \bar{A} tritt ein).

Häufig Umschreibung mit Wendungen wie “Erfolg” \longleftrightarrow “Misserfolg”, “Hat Eigenschaft” \longleftrightarrow “Hat Eigenschaft nicht”, “Tritt auf” \longleftrightarrow “Tritt nicht auf”, u.s.w.

Beispiele 7.1: Bernoulli-Experiment

- i) Münzwurf: Es fällt Kopf (Ereignis A) oder nicht.
- ii) Würfeln: 6 wird gewürfelt (Ereignis A) oder nicht.
- iii) Wahlen: Wählt “Ja” (Ereignis A) oder nicht.

Wir definieren die Zufallsvariable X_B , die den Wert 1 annimmt, falls A eingetreten ist und den Wert 0, falls nicht: $X_B(A) = 1$ bzw. $X_B(\bar{A}) = 0$.

Erfolgswahrscheinlichkeit: $P(A) = \Theta \implies P(\bar{A}) = 1 - \Theta$.

Beispiele 7.2: Erfolgswahrscheinlichkeit in Bernoulli-Experimenten

- i) Ereignis A – “Kopf fällt”: $\Theta = \frac{1}{2}$, falls Münze fair.

ii): Ereignis A – “6 wird gewürfelt”: $\Theta = \frac{1}{6}$, falls Würfel fair.

Was sind Massenfunktion und Verteilungsfunktion von X_B ?

Wir beobachten: $P(X_B = 1) = P(A) = \Theta$ und $P(X_B = 0) = P(\bar{A}) = 1 - \Theta \implies$. Also

$$f_B(x|\Theta) = \begin{cases} 1 - \Theta & , \text{für } x = 0 \\ \Theta & , \text{für } x = 1 \\ 0 & , \text{sonst} \end{cases}$$

und

$$F_B(x|\Theta) = \begin{cases} 0 & , \text{für } x < 0 \\ 1 - \Theta & , \text{für } 0 \leq x < 1 \\ 1 & , \text{für } x \geq 1 \end{cases}$$

Zum Schluss wollen wir noch Erwartungswert und Varianz einer Bernoulli-verteilten ZV berechnen.

$$EX_B = 0 \cdot P(X_B = 0) + 1 \cdot P(X_B = 1) = 0 \cdot f(0) + 1 \cdot f(1) = 0 \cdot (1 - \Theta) + 1 \cdot \Theta = \Theta$$

$$\text{var } X_B = (0 - \Theta)^2 \cdot (1 - \Theta) + (1 - \Theta)^2 \cdot \Theta = \Theta(1 - \Theta).$$

7.1.3 Binomialverteilung

Eine unabhängige n -fache Wiederholung von Bernoulli-Experimenten nennt man auch *Bernoulli-Kette*.

Was ist die Wahrscheinlichkeit, dass das interessierende Ereignis A k -mal eintritt? Das führt auf binomialverteilte ZV.

Es gelte wieder $P(A) = \Theta$.

Wir definieren $X_b :=$ “Anzahl der Erfolge (d.h. das Auftreten von A) in n Wiederholungen”.

Der Wertebereich von X_b ist gegeben durch $W_{X_b} = \{0, 1, 2, \dots, n\}$, da bei n Wiederholungen das Ereignis A mindestens 0-mal und höchstens n -mal auftreten kann.

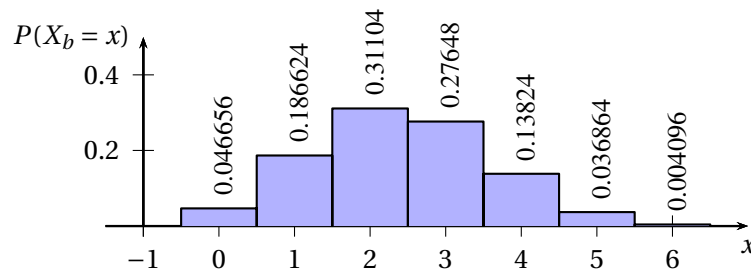
Es gilt

$$f_b(x|\Theta; n) = P(X_b = x) = \binom{n}{x} \Theta^x (1 - \Theta)^{n-x} \text{ für } x = 0, 1, \dots, n \text{ und } 0 \text{ sonst.}$$

Zur Motivation sei Folgendes gesagt. Nehmen wir an, die ersten k Versuche waren Erfolge und die restlichen $n - k$ Misserfolge. Die WSK hierfür ist $\Theta^k (1 - \Theta)^{n-k}$. Jetzt können aber die k Erfolge auf $\binom{n}{k}$ Weisen auf die n Versuche verteilt werden. Dadurch kommt der Faktor $\binom{n}{k}$ zustande.

Für binomialverteilte ZV gilt $EX_b = n\Theta$ und $\text{var } X_b = n\Theta(1 - \Theta)$.

Die folgende Graphik zeigt die Massenfunktion einer binomialverteilten ZV mit $n = 6$ und Erfolgswahrscheinlichkeit $\Theta = 0.4$.



Bemerkungen 7.3

i) Der Werte 0.31104 ist also z.B. die WSK von 2 Erfolgen in 6 Versuchswiederholungen mit $\Theta = 0.4$:

$$0.31104 = \binom{6}{2} 0.4^2 \cdot 0.6^4$$

ii) Die Graphen von Massenfunktionen hatten wir immer als “Stabdiagramme” dargestellt (siehe z.B. Gleichverteilung), mit WSKen, die in den Ausprägungen der Zufallsvariablen konzentriert waren. Die Darstellung oben verwendet eine “histogrammartige” Darstellung, in der die WSKen über Intervalle “verschmiert” werden. Das ist eine alternative Methode, die insbesondere in Statistik-Packages Anwendung findet.

Beispiele 7.4: Binomialverteilung

i) Wir würfeln 5-mal. Wie gross ist die WSK, dass die 6 (Ereignis A , $\Theta = P(A) = \frac{1}{6}$) genau 2-mal auftritt?

Antwort:

$$f_b(2 | \frac{1}{6}; 5) = \binom{5}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 \approx 0.161.$$

ii) Wie gross ist die WSK, dass zufällig mindestens 14 von 15 Messergebnissen eine gerade Zahl als letzte Ziffer haben (Ereignis A : “Letzte Ziffer gerade”), wenn angenommen wird, dass gerade und ungerade Ziffern an letzter Stelle gleichwahrscheinlich (d.h. $\Theta = P(A) = \frac{1}{2}$) sind?

Antwort:

$$f_b(14|\frac{1}{2}; 15) + f_b(15|\frac{1}{2}; 15) = \binom{15}{14} \left(\frac{1}{2}\right)^{14} \left(\frac{1}{2}\right)^1 + \binom{15}{15} \left(\frac{1}{2}\right)^{15} \left(\frac{1}{2}\right)^0 = \frac{1}{2048}$$

iii) Bei der Produktion von Transistoren sind 0.2% defekt (Ereignis A: "Transistor defekt", $\Theta = P(A) = 0.002$). Wie gross ist die Wahrscheinlichkeit, dass unter 5000 Stück genau 10 defekt sind?

Antwort:

$$f_b(10|0.02; 5000) = \binom{5000}{10} \cdot 0.002^{10} \cdot 0.998^{4990} = 0.125235$$

iv) Man bestimme die WSK, dass bei 10 Würfeln einer regulären Münze zweimal oder öfter, höchstens jedoch 8-mal Kopf erscheint.

Antwort: $P(2 \leq X \leq 8) = P(X=2) + P(X=3) + \dots + P(X=8) = F_b(8|\frac{1}{2}; 10) - F_b(1|\frac{1}{2}; 10) = 0.9893 - 0.0107 = 0.9786$. (Vgl. Bemerkung 6.7)

7.1.4 Poisson-Verteilung

Oft tritt ein gewisses Phänomen in einer Zeit-, Raum-, Flächen- oder sonstigen Einheit mit einer mittleren Häufigkeit oder Intensität λ auf. Beispiele hierfür:

- Anzahl Kunden pro Zeiteinheit an Bedienungsschaltern
- Zerfall radioaktiver Substanzen mit grosser Halbwertszeit
- Anzahl von Druckfehlern pro Seite
- Anzahl von Sprüngen in Aktienpreisen in einem gewissen Zeitintervall
- Anzahl von Mutationen in einem DNA-Abschnitt nach Einwirkung einer gewissen Strahlendosis
- etc. etc.

ZV X_p - "Anzahl des Auftretens des Ereignisses in einer Einheit"

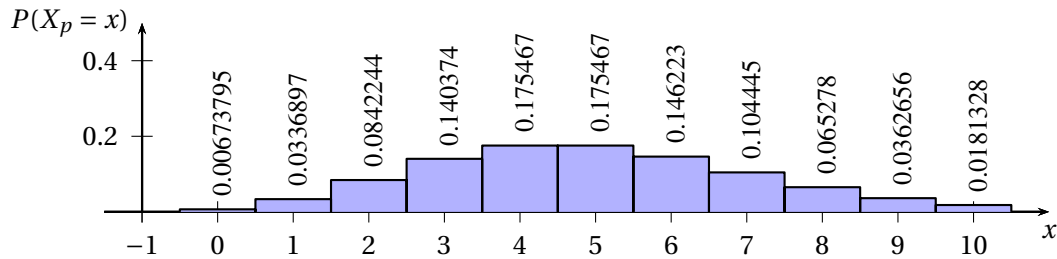
$W_{X_p} = \{0, 1, 2, 3, 4, 5, \dots\}$.

Für die Massenfunktion gilt:

$$f_p(x|\lambda) = P(X_p = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \text{ für } x = 0, 1, 2, 3, \dots \text{ und } 0 \text{ sonst,}$$

wobei λ die oben erwähnte mittlere Häufigkeit oder Intensitätsrate ist.

Die folgende Graphik zeigt (einen Ausschnitt) der Massenfunktion einer Poisson-verteilten ZV mit $\lambda = 5$. Der Wert 0.146223 errechnet sich als $\frac{5^6}{6!}e^{-5}$.



Für Poisson-verteilte ZV gilt $EX_p = \lambda$ und $\text{var } X_p = \lambda$.

Beispiele 7.5: Poisson-Verteilung

i) An einem Bankschalter treffen pro Stunde durchschnittlich 22 Kunden ein. Wie gross ist WSK, dass innerhalb von 15min genau 4 Kunden eintreffen?

Antwort: Die Intensitätsrate bezogen auf eine Stunde ist 22. Die mittlere Anzahl pro 15min ist also 5.5. Das ist unser λ . Für die gesuchte WSK erhalten wir:

$$P(X_p = 4) = f_p(4|5.5) = \frac{5.5^4}{4!} e^{-5.5} = 0.156.$$

ii) In einer Stadt gibt es im Durchschnitt zehn Verkehrsunfälle pro Tag. Man bestimme mit Hilfe der Poissonverteilung die Wahrscheinlichkeit, dass an einem bestimmten Tag in dieser Stadt weniger als vier Verkehrsunfälle passieren!

Antwort: $\lambda = 10$. Daraus folgt

$$\begin{aligned} P(X_p < 4) &= P(X_p = 0) + P(X_p = 1) + P(X_p = 2) + P(X_p = 3) \\ &= \left(\frac{10^0}{0!} + \frac{10^1}{1!} + \frac{10^2}{2!} + \frac{10^3}{3!} \right) e^{-10} = 0.0103 \end{aligned}$$

iii) Zum Gepäckaufbewahrungsschalter eines kleinen Bahnhofs kommen pro Stunde durchschnittlich 5 Kunden. Wie gross ist die Wahrscheinlichkeit, dass in einer Stunde

a) genau 3 Kunden,

Antwort: $P(X_p = 3) = \frac{5^3}{3!} e^{-5} = 0.1404.$

b) höchstens 4 Kunden,

Antwort: $P(X_p \leq 4) = P(X_p = 0) + P(X_p = 1) + P(X_p = 2) + P(X_p = 3) + P(X_p = 4) = F_p(4|5) = 0.4405$.

c) mindestens 5 Kunden

Antwort: $P(X_p \geq 5) = 1 - P(X_p \leq 4) = 1 - 0.4405 = 0.5595$.

kommen?

Beispiel 7.6: Poisson-Verteilung

Interessant. Könnt ihr aber überspringen, falls ihr das anders seht!!

In der Preussischen Armee hat man die Zahl der pro Jahr durch Pferdetritte getöteten Kavalleriesoldaten in 10 Regimentern über eine Periode von 20 Jahren dokumentiert. Das liefert 200 Daten, die unten tabelliert sind:

Zahl der Opfer	0	1	2	3	4
Beobachtete Häufigkeit	109	65	22	3	1
Relative Häufigkeit	0.545	0.325	0.110	0.015	0.005

Wenn die ZV X die Anzahl von Opfern pro Jahr pro Regiment repräsentiert, dann erhält man auf der Basis der beobachteten Daten eine "mittlere Häufigkeit"

$$\lambda = 0 \cdot \frac{109}{200} + 1 \cdot \frac{65}{200} + 2 \cdot \frac{22}{200} + 3 \cdot \frac{3}{200} + 4 \cdot \frac{1}{200} = 0.61.$$

Betrachten wir die Poisson-Verteilung mit $\lambda = 0.61$, erhalten wir die folgenden WSKen:

$$P(X = 0) = 0.543, P(X = 1) = 0.331, P(X = 2) = 0.101,$$

$$P(X = 3) = 0.021, P(X = 4) = 0.003.$$

Der Vergleich dieser Werte mit den beobachteten relativen Häufigkeiten legt nahe, dass die Zahl der Toten pro Regiment pro Jahr durch eine Poisson-Verteilung mit Intensitätsrate $\lambda \approx 0.61$ beschrieben werden kann.

7.1.5 Approximation der Binomialverteilung durch Poisson-Verteilung (Kein Modulstoff!)

Die Poisson-Verteilung spielt auch bei der Approximation der Binomialverteilung eine grosse Rolle. Das ist häufig sehr nützlich, da sich die Berechnung von WSKen für binomialverteilte ZV sehr aufwendig gestalten kann. (Na ja, mit EXCEL oder R zur Hand ist das vielleicht kein stichhaltiges Argument mehr.)

Faustregel: Für $n \geq 100$ und $p \leq \frac{1}{10}$ liefert die Poisson-Verteilung eine gute Näherung für die Binomialverteilung in dem Sinne, dass

$$f_b(x|p; n) \approx f_p(x|np) \quad \text{d.h. mit } \lambda = np.$$

Beispiel 7.7: Approximation Binomialverteilung durch Poisson-Verteilung

Sei $n = 100$ und $p = 0.01$. Das liefert $\lambda = 1$.

k	$f_b(k 0.01; 100)$	$f_p(k 1)$
0	0.3660	0.3679
1	0.3697	0.3679
2	0.1849	0.1839
3	0.0610	0.0613
4	0.0149	0.0153
5	0.0029	0.0031

Die Approximation ist also recht anständig.

Beispiel 7.8: Approximation Binomialverteilung durch Poisson-Verteilung

Siehe Beispiel Transistoren.

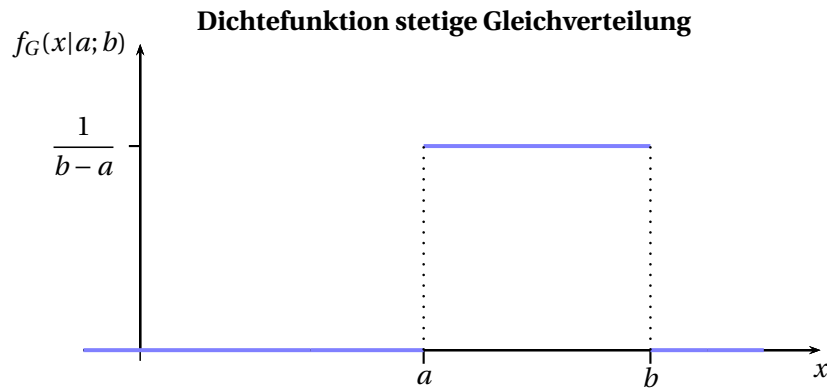
Antwort: $n = 5000 \geq 100$ und $p = 0.002 \leq \frac{1}{10}$, d.h. die Faustregel für Approximierbarkeit der Binomialverteilung durch die Poisson-Verteilung ist erfüllt. Daher gilt mit $\lambda = np = 10$:

$$f_b(10|0.002; 5000) \approx f_p(10|10) = 0.1251.$$

7.2 Stetige Verteilungen

7.2.1 Stetige Gleichverteilung (oder Rechteckverteilung)

Sei X_G eine stetige ZV, die Werte auf einem Intervall $[a, b]$ annimmt. Man unterteile $[a, b]$ in eine beliebige Anzahl gleichlanger Teilintervalle. Ist die WSK, dass X_G Werte in einem dieser Teilintervalle annimmt für jedes dieser Teilintervalle gleich, spricht man von einer (stetig) gleichverteilten ZV. Wenn ihr euch an die Flächeninterpretation von Intervallwahrscheinlichkeiten unter Dichtefunktionen erinnert, kann eine X_G nur eine Dichte der Form



haben. Hier sieht man nochmal schön, dass die Fläche unter der Dichtefunktion 1 ergibt (und woher der Name “Rechteckverteilung” kommt).

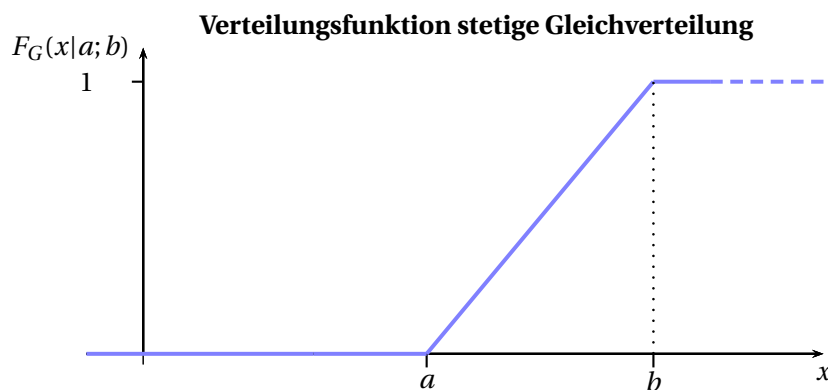
In Formeln:

$$f_G(x|a; b) = \begin{cases} \frac{1}{b-a} & , \text{für } a \leq x \leq b \\ 0 & , \text{sonst} \end{cases}$$

Für die Verteilungsfunktion ergibt sich

$$F_G(x|a; b) = \begin{cases} 0 & , \text{für } x < a \\ \frac{x-a}{b-a} & , \text{für } a \leq x \leq b \\ 1 & , \text{für } x > b \end{cases}$$

Daraus resultiert der folgende Graph:



Ist eine ZV X auf dem Intervall $[a, b]$ gleichverteilt und sind x_1 und x_2 Werte in diesem Intervall mit $a \leq x_1 < x_2 \leq b$, dann gilt $P(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{b - a}$. Die WSK, dass X Werte im Intervall $[x_1, x_2]$

annimmt, ist also gleich dem Verhältnis der Länge dieses Teilintervalls zur Länge des Gesamtintervalls.

Man kann relativ einfach zeigen:

$$EX_G = \frac{a+b}{2} \quad \text{und} \quad \text{var} X_G = \frac{(b-a)^2}{12}$$

Beispiel 7.9: Rechteckverteilung

Wir wählen zufällig einen Punkt e aus dem Intervall $I = [-1, 1]$. Die ZV X_G “Abstand von e von 1” ist auf $[0, 2]$ gleichverteilt. Es gilt z.B. $P(X_G \in [0, 0.5]) = P(X_G \in [1.5, 2]) = \frac{1}{4}$.

7.2.2 Exponentialverteilung (Kein Modulstoff!)

Die Exponentialverteilung wird häufig zur Modellierung von “Zeitlücken” verwendet, z.B.

- Zeit zwischen zwei Anrufen
- Lebensdauer von Atomen beim radioaktiven Zerfall
- Lebensdauer von Bauteilen, Maschinen und Geräten, wenn Alterungserscheinungen nicht betrachtet werden müssen.
- Abstände von Fahrzeugen bei einer Momentaufnahme
- etc. etc.

Wenn eine ZV X_e die Länge dieser “Lücken” beschreibt, dann hat sie eine Dichte der Form:

$$f_e(x|\lambda) = \begin{cases} 0 & , \text{für } x < 0 \\ \lambda e^{-\lambda x} & , \text{für } x \geq 0 \end{cases}$$

Damit ergibt sich eine Verteilungsfunktion der Form

$$F_e(x|\lambda) = \begin{cases} 0 & , \text{für } x < 0 \\ 1 - e^{-\lambda x} & , \text{für } x \geq 0 \end{cases}$$

Für EW und Varianz einer exponentialverteilten ZV mit Parameter λ ergibt sich

$$EX_e = \frac{1}{\lambda} \quad \text{und} \quad \text{var} X_e = \frac{1}{\lambda^2}.$$

Man kann darüber hinaus zeigen, dass die “Zeitlücken” für einen Poisson-Prozess mit Intensität λ gerade nach $F_e(x|\lambda)$ verteilt sind.

Beispiel 7.10: Exponentialverteilung

Eine Maschine hatte in den letzten Jahren pro Jahr im Schnitt 9 Ausfälle. Die Zufallsvariable X beschreibe die Dauer zwischen den Ausfällen in Monaten!

a) Man gebe eine geeignete Verteilung an und bestimme die möglichen Parameter.

Antwort: Wir gehen davon aus, dass die Anzahl der Ausfälle pro Jahr Poisson-verteilt ist mit Intensität 9. Für die Intensität pro Monat gilt dann $\lambda = \frac{9}{12} = \frac{3}{4}$. Nach dem oben gesagten gilt $X \sim F_e(x|\frac{3}{4})$.

b) Wie gross ist die WSK, dass die Dauer zwischen den Ausfällen grösser als 6 Monate ist?

Antwort:

$$P(X > 6) = 1 - P(X \leq 6) = 1 - F_e(6|\frac{3}{4}) = 1 - (1 - e^{-\frac{3}{4} \cdot 6}) \approx 0.0111.$$

Da die Exponentialverteilung auch als Lebensdauerverteilung verwendet wird, ist es möglich, damit zusammenhängende Grössen wie Überlebenswahrscheinlichkeit, die Restlebensdauer und die Ausfallrate mit Hilfe der Verteilungsfunktion anzugeben. So nennt man das Komplement der Verteilungsfunktion die Überlebenswahrscheinlichkeit: $P(X > x) = 1 - F_e(x|\lambda) = e^{-\lambda x}$. Da die Exponentialverteilung eine "gedächtnislose" Verteilung ist (im Zusammenhang mit Überlebenswahrscheinlichkeit bedeutet dies z.B., dass die Überlebenswahrscheinlichkeit unabhängig vom erreichten Alter ist), kann die Exponentialverteilung also eigentlich nur für sogenannte ermüdungsfreie Systeme verwendet werden (Menschen gehören da sicher nicht dazu, es sei denn, sie trinken genug Red Bull® oder andere Energy Drinks).

Beispiel 7.11: Exponentialverteilung

In einer Elektronikfirma werden Funkwecker produziert. Im Rahmen der Qualitätssicherung wird anhand von Reklamationen die Funktionsdauer der Wecker untersucht. Die ZV X beschreibe die Zeitdauer der Funktionsfähigkeit eines Funkweckers in Tagen.

X ist exponentialverteilt mit dem Parameter $\lambda = 0.005$ (Statistisch ermittelt!). In diesem Zusammenhang wird λ als Ausfallrate bezeichnet; es fallen also durchschnittlich pro Tag 5 Promille der Wecker aus. Entsprechend ist $\frac{1}{\lambda}$ die durchschnittliche Zeitdauer, bis ein Wecker ausfällt. Es fällt also im Mittel alle 200 Tage ein Wecker aus.

Die Wahrscheinlichkeit, dass ein Wecker höchstens 20 Tage hält, ist

$$P(X \leq 20) = F_e(20|0.005) = 1 - e^{-0.005 \cdot 20} = 0.0952,$$

d.h. nach 20 Tagen sind ca. 10% der Wecker ausgefallen.

Entsprechend ist der Anteil der Wecker, die mindestens 180 Tage durchhalten

$$P(X \geq 180) = 1 - P(X < 180) = 1 - F_e(180|0.005) = 1 - (1 - e^{-0.005 \cdot 180}) = 0.4066,$$

also halten ca. 40% der Wecker länger als 180 Tage.

Man sieht, dass sehr viele Wecker schon nach kurzer Zeit defekt sind, ein Phänomen, das man bei elektronischen Geräten häufig beobachtet. Wenn sie kaputt gehen, dann meistens sofort. Ab dann halten sie quasi ewig.

7.2.3 Normalverteilung

Die Normalverteilung ist die mit Abstand wichtigste Verteilung, da sie quasi alle natürlichen Phänomene beschreibt, die das Resultat der Überlagerung von vielen unabhängigen Faktoren sind. Normalverteilte ZV haben eine von zwei Parametern abhängige Dichte der Form

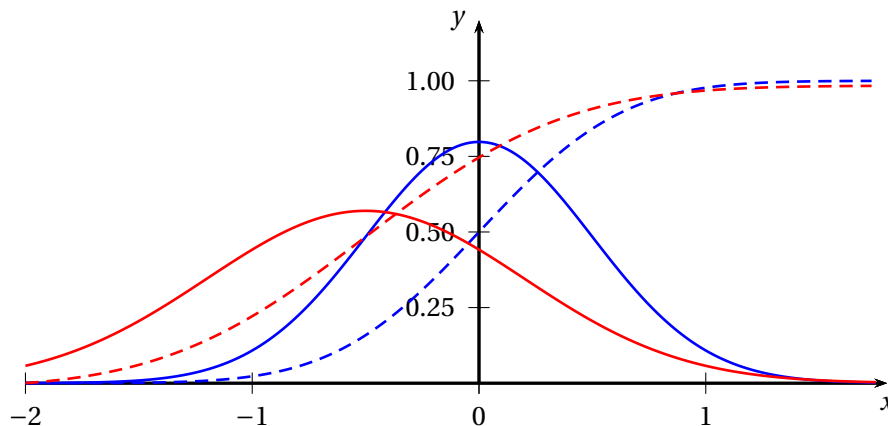
$$f_N(x|\mu;\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

und Verteilungsfunktion

$$F_N(x|\mu;\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2\sigma^2}(t-\mu)^2} dt.$$

Man kann zeigen, dass für eine ZV X_n mit dieser Verteilung gilt $EX_n = \mu$ und $\text{var } X_n = \sigma^2$.

Typische Verläufe von Dichtefunktionen und den dazugehörigen Verteilungsfunktionen (gestrichelt) sind in der folgenden Graphik dargestellt. Die Dichtefunktionen zeigen dabei die bekannte Glockenform. Insbesondere nehmen sie ihr Maximum an der Stelle μ an und die vertikale Gerade durch $x = \mu$ ist Symmetrieachse.



Die blauen Graphen zeigen Dichte- bzw. Verteilungsfunktion der sogenannten *Standardnormalverteilung*, d.h. der Normalverteilung mit Parametern $\mu = 0$ und $\sigma = 1$.

Standardnormalverteilte ZV werden häufig mit Z bezeichnet, Dichte und Verteilungsfunktion der Standardnormalverteilung mit $f_N(z)$ bzw. $F_N(z)$.

$F_N(z)$ ist im Lehrbuch tabelliert!!

Wollen wir also Wahrscheinlichkeiten im Kontext von normalverteilten ZV mittels dieser Tabelle bestimmen, müssen wir diese standardisieren. Wie das gemacht und angewandt wird, zeigt die folgende Zusammenstellung:

Sei die ZV X_n nach $F_N(x|\mu;\sigma)$ verteilt (Man schreibt dafür auch $X_n \sim F_N(x|\mu;\sigma)$.), dann ist die ZV $Z := \frac{X - \mu}{\sigma}$ standardnormalverteilt, d.h. $Z \sim F_N\left(z = \frac{x - \mu}{\sigma}\right)$

Man bestimmt dann

$$\text{i) } P(a \leq X_n \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = F_N\left(\frac{b - \mu}{\sigma}\right) - F_N\left(\frac{a - \mu}{\sigma}\right)$$

(Das gleiche gilt für \leq ersetzt durch $<$, wie ihr euch vielleicht noch erinnert.)

$$\text{ii) } P(X_n > a) = P\left(Z > \frac{a - \mu}{\sigma}\right) = 1 - P\left(Z \leq \frac{a - \mu}{\sigma}\right) = 1 - F_N\left(\frac{a - \mu}{\sigma}\right)$$

iii) $F_N(z) = 1 - F_N(-z)$. Diese Beziehung ist wichtig, da im Lehrbuch nur $F_N(z)$ für $z \geq 0$ tabelliert ist. Will man z.B. $F_N(-1)$ bestimmen, so gilt $F_N(-1) = 1 - F_N(1)$. Aber $F_N(1) = 0.8413$ findet man in der Tabelle.

Die folgenden Beispiele sollen verdeutlichen, wie man obige Fakten anwendet. Beachtet auch hier bitte wieder: Die konkrete Situation ist absolut irrelevant. Das einzig wichtige sind die Zahlenwerte.

Beispiele 7.12: Normalverteilung

i) Der Durchmesser der in einer Fabrik hergestellten Stahlscheiben ist normalverteilt mit dem Mittelwert $\mu = 131\text{mm}$ und der Varianz $\sigma^2 = 5\text{mm}^2$. Mit welcher Wahrscheinlichkeit hat eine zufällig ausgesuchte Scheibe einen Durchmesser von mindestens 128mm?

Antwort:

$$\begin{aligned} P(X_n \geq 128) &= 1 - P(X_n < 128) = 1 - P\left(Z < \frac{128 - 131}{\sqrt{5}}\right) = 1 - F_N(-1.3416) \\ &= 1 - (1 - F_N(1.3416)) \approx 0.91 \end{aligned}$$

ii) Die Rendite einer Aktie sei normalverteilt mit dem Erwartungswert $\mu = 18\%$ und einer Standardabweichung (Risiko) von $\sigma = 5\%$. Wie gross ist die WSK, dass die Rendite der Aktie zwischen 10% und 20% liegen wird?

Antwort:

$$P(10 \leq X_n \leq 20) = P\left(\frac{10 - 18}{5} \leq Z \leq \frac{20 - 18}{5}\right) = F_N(0.4) - F_N(-1.6) = 0.6006$$

iii) Die Länge von Eukalyptusblättern ist normalverteilt mit einem Erwartungswert von 151mm und einer Standardabweichung von 15mm. Mit welcher Wahrscheinlichkeit ist ein zufällig ausgewähltes Blatt

a) zwischen 121 und 154 mm lang

Antwort: $P(X_n \leq 154) - P(X_n \leq 121) = 0.5793 - 0.0228 = 0.5565$.

b) mindestens 169 mm lang

Antwort: $P(X_n \geq 169) = 1 - 0.8849 = 0.1151$.

c) höchstens 130 mm lang

Antwort: $P(X_n \leq 130) = 0.0808$.

d) höchstens 127 mm oder mindestens 166 mm lang?

Antwort: $P(X_n \leq 127) + P(X_n \geq 166) = 0.0548 + (1 - 0.8413) = 0.2135$.

$k\sigma$ -Regeln

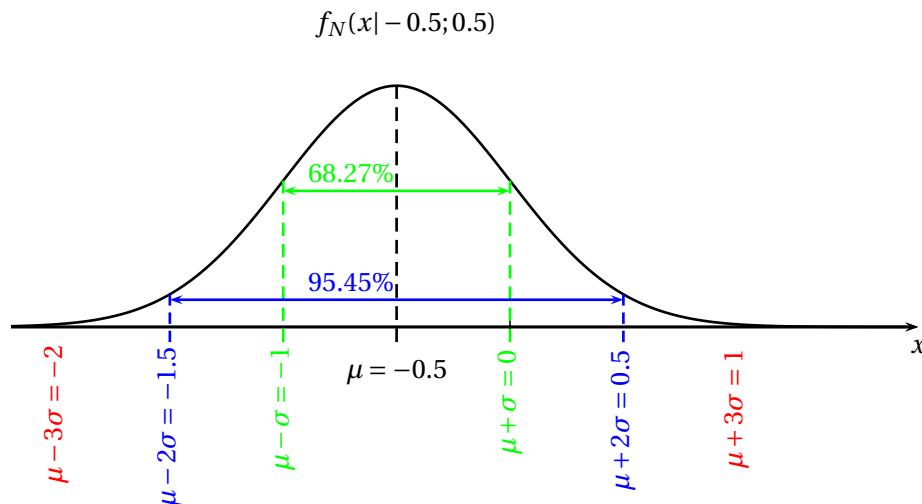
Für eine normalverteilte ZV X_n ist es möglich, allein auf Basis von Erwartungswert μ und Standardabweichung σ , genauere Aussagen über WSKen zu treffen, mit der X_n Werte in gewissen Intervallen annimmt. Konkret wollen wir die Frage klären, wie gross die WSK ist, dass X_n weniger als $k\sigma$ von ihrem Mittelwert μ abweicht?

Es gilt

$$\begin{aligned} P(\mu - k\sigma \leq X_n \leq \mu + k\sigma) &= F_N(\mu + k\sigma | \mu, \sigma) - F_N(\mu - k\sigma | \mu, \sigma) \\ &= F_N\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - F_N\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\ &= F_N(k) - F_N(-k) = 2F_N(k) - 1. \end{aligned}$$

Für $k = 1, 2, 3$ ergeben sich die Werte 0.6827, 0.9545 bzw. 0.9973.

Es gilt also für normalverteilte ZV, dass sie z.B. höchstens mit einer WSK von 0.27% um mehr als 3σ von ihrem Erwartungswert abweichen (Es gibt sowohl sehr wenige riesenwüchsige als auch zwerghüchsig Menschen!). Die folgende Grafik zeigt das nochmal anschaulich am Beispiel einer Normalverteilung mit $\mu = -0.5$ und $\sigma = 0.5$.



68.27% der Beobachtungswerte dieser ZV werden also ins Intervall $[-1, 0]$, 95.45% ins Intervall $[-1.5, 0.5]$ und 99.37% ins Intervall $[-2, 1]$ fallen.

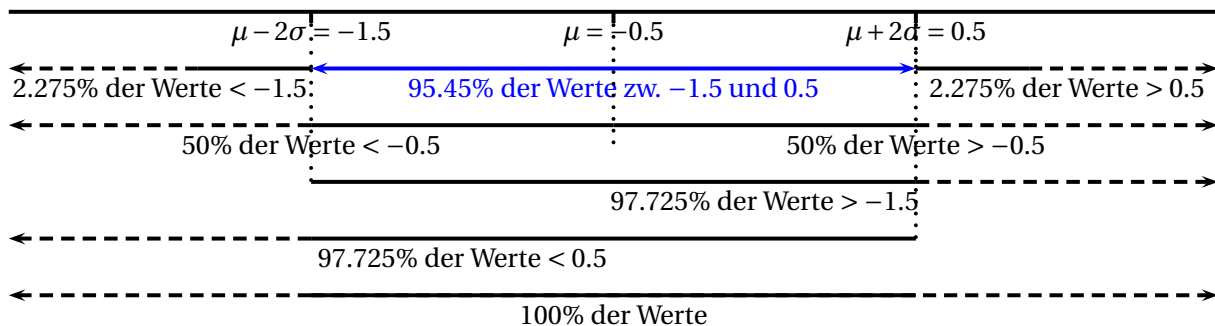
Beispiel 7.13: $k\sigma$ -Regel

Sei X_n normalverteilt mit $\mu = 10$ und $\sigma = 3$. Wie gross ist die WSK, dass X_n Werte grösser als 16 oder kleiner als 4 ist?

$$P(X_n > 16 \text{ oder } X_n < 4) = 1 - P(4 \leq X_n \leq 16)$$

Wir beobachten jetzt, dass $4 = 10 - 6 = \mu - 2\sigma$ und $16 = 10 + 6 = \mu + 2\sigma$. Das heisst, $P(4 \leq X_n \leq 16) = 0.9545$ und somit die gesuchte WSK $1 - 0.9545 = 0.0455$. Die WSK beträgt also ca. 4.5%.

Zum Schluss exemplarisch an der 2σ -Regel noch einige Wahrscheinlichkeiten, die unmittelbar folgen:



8 Grundlegende Problemstellung der schliessenden Statistik

Die schliessende Statistik befasst sich, vereinfacht gesagt, mit der folgenden Problemstellung:

- Es liegen Beobachtungen (Merkmalsausprägungen) x_1, x_2, \dots, x_n in einem zufälligem Experiment vor.
- Die Beobachtungen x_1, x_2, \dots, x_n sind Realisierungen von unabhängigen und identisch verteilten (uiv) ZV X_1, X_2, \dots, X_n .
- Die Verteilung von X_1, X_2, \dots, X_n ist (zumindest teilweise) unbekannt.

Gesucht sind Aussagen über die Verteilung und/oder deren Eigenschaften (Art der Verteilung, Erwartungswert, Varianz,...).

Wie muss man sich das vorstellen?

Beispiel 8.1

Wir betrachten das zufällige Experiment "Fangen eines Fisches aus einem See." Die möglichen Versuchsausgänge (Ereignisraum, Grundgesamtheit) bestehen also aus allen Fische in diesem See. Wir interessieren uns für die Verteilung des Merkmals (ZV) "Gewicht der Fische". Wir könnten den See trockenlegen, alle Fische einsammeln, ihre Gewichte bestimmen und wüssten so alles über die Gewichte der Fische (aller "statistischer Einheiten"). Nicht praktisch. Wir schicken fünf Angler los. Diese fangen je einen Fisch (hoffentlich) und bestimmen dessen Gewicht. Die ZV X_i ($i = 1, 2, \dots, 5$) repräsentieren die Fischgewichte der 5 Angler. Sind die Fische gefangen, ordnen wir jeder ZV X_i das konkrete Fischgewicht (Realisierung) x_i zu. Was können wir auf Basis der Realisierungen x_1, x_2, \dots, x_5 über die "Gesamtsituation" schliessen? Welches mittlere Fischgewicht können wir auf Basis der gefangenen Fische erwarten? -> Punktschätzungen. Können wir einen Bereich angeben, in dem das mittlere Fischgewicht mit einer gewissen Wahrscheinlichkeit liegt? -> Intervallschätzungen. Ist das Gewicht der Fische im See

normalverteilt? -> Anpassungstests etc. etc.

Viele Verfahren der schliessenden Statistik stützen sich mehr oder weniger explizit auf Grenzwertsätze. Wir wollen zwei davon näher betrachten.

9 Grenzwertsätze

Wir betrachten eine Folge von unabhängigen und identisch verteilten (uiv) ZV X_1, X_2, \dots, X_n . Identisch verteilt heisst nichts anderes, als dass sie alle dieselbe Verteilung haben. Sie haben also insbesondere alle denselben Erwartungswert und dieselbe Varianz.

$$EX_i = \mu \quad \text{und} \quad \text{var} X_i = \sigma^2 \quad \text{für alle } i = 1, \dots, n.$$

Bei uns werden uiv ZV vor allem im Zusammenhang mit Stichproben vom Umfang n aus einer Grundgesamtheit auftreten (siehe oben).

Unser Ziel wird sein, Aussagen über die neue Zufallsvariable (die sogenannte Mittelwertvariable)

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

zu treffen, wenn n gegen unendlich strebt.

Aus allgemeinen Eigenschaften von Erwartungswert und Varianz folgt unmittelbar

(E1) $E\bar{X}_n = \mu$

(E2) $\text{var} \bar{X}_n = \frac{\sigma^2}{n}$ (Hierfür braucht man die Unabhängigkeit der ZV. Im Allgemeinen gilt das nicht!).

Was passiert nun für $n \rightarrow \infty$?

Wir betrachten zwei Grenzwertsätze.

9.1 Das Schwache Gesetz der grossen Zahlen

Seien X_1, X_2, \dots, X_n uiv ZV mit $E(X_i) = \mu$. Für alle $\epsilon > 0$ gilt:

$$P(|\bar{X}_n - \mu| \geq \epsilon) \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

In Worten: Für $n \rightarrow \infty$ wird es beliebig unwahrscheinlich, dass sich das arithmetische Mittel von uiv ZV um mehr als eine beliebig kleine Grösse $\epsilon > 0$ vom Erwartungswert μ der ZV abweicht.

Übergang zum Komplementärereignis liefert:

$$P(|\bar{X}_n - \mu| < \epsilon) \rightarrow 1$$

Das Schwache Gesetz der grossen Zahlen ist Motivation der Aussage, dass es um so wahrscheinlicher für ein Stichprobenmittel ist, nahe beim Mittelwert der Grundgesamtheit zu liegen, je grösser die Stichprobe ist.

Spezialfall: *Bernoulli-Gesetz der grossen Zahlen.*

Wir betrachten jetzt ausschliesslich Bernoulli-Variablen.

Zur Erinnerung: Eine Bernoulli-Variable X_B beschreibt das Auftreten (Erfolg) bzw. Nichtauftreten (Misserfolg) eines Ereignisses A in einem Bernoulli-Experiment. Es gilt $X_B = 1$, falls A eingetreten ist und $X_B = 0$ sonst. Wir haben schon abgeleitet, dass $EX_B = \Theta$ und $\text{var } X_B = \Theta(1 - \Theta)$, wobei Θ die Erfolgswahrscheinlichkeit ist, d.h. $\Theta = P(A)$.

Jetzt: n unabhängige Wiederholungen des Bernoulli-Experiments. Die ZV $(X_B)_1, (X_B)_2, \dots, (X_B)_n$ sollen die jeweiligen Versuchsausgänge beschreiben. Sie sind also unabhängig und identisch (Bernoulli-) verteilt. Wir definieren das Gegenstück zu \bar{X}_n von vorhin, nämlich

$$H_n = \frac{1}{n} ((X_B)_1 + (X_B)_2 + \dots + (X_B)_n).$$

H_n repräsentiert die relative Häufigkeit des Auftretens von A bei n Versuchswiederholungen!

Aus (E1) und (E2) folgt:

$$EH_n = \Theta \quad \text{und} \quad \text{var } H_n = \frac{\Theta(1 - \Theta)}{n}.$$

Einsetzen ins Schwache Gesetz der grossen Zahlen liefert das Bernoulli-Gesetz der grossen Zahlen:

$$P(|H_n - \Theta| \geq \epsilon) \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Beachtet bitte nochmal, dass $\Theta = P(A)$. Das Resultat besagt also nichts anderes, als dass die relativen Häufigkeiten H_n des Auftretens von A stochastisch gegen die Wahrscheinlichkeit von A streben. Das ist wichtig für die praktische Bestimmung von Wahrscheinlichkeiten. Um das aber wirklich numerisch machen zu können, braucht es noch ein anderes Resultat, nämlich die

9.1.1 Tschebychev-Ungleichung

Die Tschebyschev-Ungleichung ist eine der wenigen Resultate über WSK im Kontext von Zufallsvariablen, die für alle ZV unabhängig von ihrer Verteilung gelten. Allerdings hat diese Allgemeinheit ihren Preis. Approximationen von WSK mittels Tschebyschev-Ungleichung sind i.A. recht grob, d.h. werden häufig überschätzt.

Das Resultat lautet wie folgt: Sei X eine ZV mit $EX = \mu$ und $\text{var } X = \sigma^2$ und $\epsilon > 0$. Dann gilt

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

In Worten: Die WSK, dass eine ZV um mehr als einen gewissen Betrag ϵ von ihrem EW abweicht, ist kleiner gleich dem Quotienten auf der rechten Seite.

Für das Komplementärereignis $|X - \mu| < \epsilon$ liefert diese Ungleichung unmittelbar:

$$P(|X - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}.$$

Das folgende Beispiel demonstriert, wie T-U im Kontext des Schwachen Gesetzes der grossen Zahlen angewandt wird.

Beispiel 9.1: Tschebychev-Ungleichung

Wie häufig muss man würfeln, um mit WSK grösser gleich 0.95 mit dem gewürfelten Mittel nicht mehr als 0.5 vom wahren Mittel (3.5) abzuweichen?

Die ZV X_i beschreibe die Augenzahl im i -ten Wurf. Dann kann man davon ausgehen, dass X_1, X_2, \dots, X_n unabhängig und identisch verteilt sind. \bar{X}_n sei wie üblich definiert. Dann gilt:

$$EX_i = E\bar{X}_n = 3.5 \quad \text{und} \quad \text{var } X_i \approx 2.92 \quad \text{woraus mit (E2) folgt} \quad \text{var } \bar{X}_n \approx \frac{2.92}{n}.$$

Nach T-U gilt:

$$P(|\bar{X}_n - 3.5| \leq 0.5) \geq 1 - \frac{2.92}{n \cdot 0.5^2}.$$

n muss jetzt so bestimmt werden, dass der letzte Ausdruck grösser gleich 0.95 wird. Das führt auf $n \geq 234$.

Ein weiteres Beispiel:

Beispiel 9.2: Tschebychev-Ungleichung

Die durchschnittliche Länge eines Wikipedia-Artikels beträgt 1000 Zeichen bei einer Standardabweichung $\sigma = 200$. Wie gross ist die WSK, dass ein Artikel zw. 600 und 1400 Zeichen hat? (Hier beschreibt die ZV X also die Anzahl der Zeichen eines zufällig ausgewählten Artikels; $EX = \mu = 1000$ und $\text{var } X = \sigma^2 = 40000$. Die Frage ist nach der WSK, dass X um weniger als $\epsilon = 400$ von ihrem Mittelwert abweicht.) Nach T-U gilt:

$$P(|X - 1000| < 400) \geq 1 - \frac{40000}{160000} = 0.75.$$

Anders ausgedrückt: Mehr als 75% aller Wikipedia-Artikel liegen im Bereich zw. 600 und 1400

Zeichen.

Häufig ist man an Abweichungen interessiert, die ein Vielfaches $k\sigma$ der Standardabweichung sind. Ersetzt man in den obigen Ungleichungen ε durch $k\sigma$ erhält man:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{bzw.} \quad P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Beispiel 9.3

Wie gross ist beim Würfeln die WSK eine 1 oder 6 zu werfen? (Hier beschreibt die ZV X also die Augenzahl; $EX = \mu = 3.5$ und $\text{var } X = \sigma^2 \approx 2.92$. Die Frage ist nach der WSK, dass X um $\varepsilon = 2.5$ (oder mehr) von ihrem Mittelwert abweicht.)

Man kann diese WSK natürlich direkt mit der (allgemeinen) T-U abschätzen. Wir wollen jedoch den Spezialfall benutzen. Dazu müssen wir $\varepsilon = 2.5$ als Produkt $k\sigma$ mit $\sigma = \sqrt{\sigma^2} = \sqrt{2.92} \approx 1.7088$ darstellen. Das liefert $k \approx 1.463$. Nach T-U gilt dann:

$$P(|X - 3.5| \geq 1.463\sigma) \leq \frac{1}{1.463^2} \approx 0.467.$$

Natürlich wissen wir, dass die exakte WSK für dieses Ereignis $\frac{1}{3}$ beträgt. Die Abschätzung mittels T-U ist also recht grob.

Bemerkungen 9.4

- i) Im letzten Beispiel hätten wir ebenso, wie schon angedeutet, die erstere Form der T-U anwenden können. Es ergibt sich natürlich die Frage, warum die alternative Form? Für $k = 1, 2, 3, \dots$ ist die alternative Form rechnerisch einfacher. Zudem braucht man nicht die Standardabweichung zu kennen, um Abschätzungen zu erhalten. Man weiss dadurch z.B. für beliebige ZV, dass sie mit einer WSK von weniger als $\frac{1}{9}$ um mehr als 3σ von ihrem EW abweichen. Dazu ist die Kenntnis von σ nicht nötig.
- ii) Für $k = 1$ liefert die Ungleichung “triviale” Resultate ($P(\dots) \leq 1$ bzw. $P(\dots) \geq 0$). Das gilt auch für $0 < k < 1$.
- iii) Obwohl T-U i.A. zu grosse WSK liefert, kann sie nicht verbessert werden. Man kann für alle zulässigen μ, σ und ε Zufallsvariablen finden, für die die Ungleichungen in Gleichungen übergehen, für die also “verschärfte” Ungleichungen nicht mehr gültig wären.

Wie sieht die Situation im “Bernoulli-Fall” aus? In dieser speziellen Situation haben wir:

$$P(|H_n - \Theta| \geq \varepsilon) \leq \frac{\Theta(1 - \Theta)}{n \cdot \varepsilon^2}$$

bzw.

$$P(|H_n - \Theta| < \varepsilon) \geq 1 - \frac{\Theta(1 - \Theta)}{n \cdot \varepsilon^2}.$$

Informell kann man z.B. die erste Ungleichung wie folgt interpretieren: Die Wahrscheinlichkeit, dass die relative Häufigkeit des Auftretens von A bei n Versuchswiederholungen um mehr als ε von $\Theta = P(A)$ abweicht, ist nach oben durch $\frac{\Theta(1 - \Theta)}{n \cdot \varepsilon^2}$ beschränkt. Eine ähnliche Interpretation kann man für die zweite Ungleichung formulieren.

Im “Bernoulli-Fall” braucht man nicht einmal die Kenntnis von Θ , da immer gilt $\Theta(1 - \Theta) \leq \frac{1}{4}$, wie man sehr einfach mit den Mitteln der Differentialrechnung (angewandt auf die Funktion $f(\Theta) = \Theta - \Theta^2$, die ihr Maximum in $\Theta = \frac{1}{2}$ hat) verifizieren kann. Man hat also auch

$$P(|H_n - \Theta| \geq \varepsilon) \leq \frac{\Theta(1 - \Theta)}{n \cdot \varepsilon^2} \leq \frac{1}{4n \cdot \varepsilon^2}$$

bzw.

$$P(|H_n - \Theta| < \varepsilon) \geq 1 - \frac{\Theta(1 - \Theta)}{n \cdot \varepsilon^2} \geq 1 - \frac{1}{4n \cdot \varepsilon^2}.$$

Dadurch werden die Abschätzungen jedoch i.A. noch mal schlechter.

9.2 Der Zentrale Grenzwertsatz (ZGWS)

Wir standardisieren die ZV \bar{X}_n .

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}.$$

Sei F_n die Verteilungsfunktion der ZV Z_n , dann gilt:

$$F_n(z) \longrightarrow F_N(z) \quad \text{für } n \rightarrow \infty, \text{ d.h.}$$

$$\text{(E3)} \quad P(Z_n \leq z) = P\left(\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \leq z\right) \rightarrow F_N(z).$$

Bemerkungen 9.5

- i) Die ZV Z_n ist also asymptotisch standardnormalverteilt.
- ii) Ohne Standardisieren ist \bar{X}_n ebenfalls asymptotisch normalverteilt, nun allerdings mit den Parametern $(\mu, \sigma / \sqrt{n})$.
- iii) Der ZGWS ist der Grund dafür, warum die Normalverteilung eine so herausragende Rolle spielt. Überall, wo sich viele unabhängige zufällige “Effekte” überlagern (Messfehler, Intelligenzquotienten, Grössen bzw. Gewichte von Lebewesen und und und ...) kann man für die Summe näherungsweise eine Normalverteilung erwarten.

iv) Der ZGWS lässt sich dahingehend verallgemeinern, dass die Voraussetzungen an die ZV, nämlich unabhängig und identisch verteilt zu sein, wesentlich abgeschwächt werden können.

Wie wird der ZGWS angewandt?

Für grosse n (Faustregel $n \geq 30$) wird “ \rightarrow ” durch “ \approx ” ersetzt!!

Konkret: Sei $Z_n = \frac{\bar{X}_n - \mu}{\sigma}$, ($n \geq 30$) und F_n die Verteilungsfunktion von Z_n . Dann:

- $P(Z_n \leq z) = F_n(z) \approx F_N(z)$
- $P(z_1 < Z_n \leq z_2) = F_n(z_2) - F_n(z_1) \approx F_N(z_2) - F_N(z_1)$
(Beachtet hier bitte das “ $<$ ” nach z_1 . Das ist wichtig für diskrete ZV X_1, \dots, X_n .)
- $P(Z_n > z) = 1 - P(Z_n \leq z) = 1 - F_n(z) \approx 1 - F_N(z)$
(Beachtet auch hier bitte wieder das “ $>$ ” vor z . Das ist wichtig für diskrete ZV X_1, \dots, X_n .)

Wichtig: Wir approximieren nur Verteilungsfunktionswerte, nichts anderes! Ihr müsst also, falls ihr mittels Normalverteilung approximieren wollt, alles mittels Verteilungsfunktionswerten ausdrücken. Das ist insbesondere bei der Berechnung von Einzelwahrscheinlichkeiten im diskreten Kontext wichtig!

Beispiel 9.6

$P(X = a) = P(a - 1 < X \leq a) = P(X \leq a) - P(X \leq a - 1) = F_X(a) - F_X(a - 1)$, wobei $F_X(x)$ die Verteilungsfunktion der diskreten ZV X ist.

Bemerkung 9.7

Falls die ZV X_1, X_2, \dots, X_n unabhängig und identisch normalverteilt sind, kann man zeigen, dass auch \bar{X}_n **exakt** normalverteilt ist mit den entsprechenden Parametern. In diesem Fall wird also “ \rightarrow ” bzw. “ \approx ” durch “ $=$ ” ersetzt.

Eine Anwendung des ZGWS:

9.2.1 Stichprobenverteilungen

Wir betrachten ein metrisches Merkmal X und n uiv ZV X_1, X_2, \dots, X_n (denkt wieder an eine Stichprobe vom Umfang n oder n (unabhängige) Wiederholungen eines Experiments) mit $EX = EX_i = \mu$ und $\text{var } X = \text{var } X_i = \sigma^2$.

- $E\bar{X} = \mu$ wegen (E1);
- $\text{var } \bar{X} = \frac{\sigma^2}{n} =: \sigma_{\bar{X}}^2$ wegen (E2) (bzw. Standardabweichung $\sigma_{\bar{X}}$ von \bar{X} ist gleich σ/\sqrt{n});
- Die **standardisierte** ZV $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ ist für grosse n annähernd standardnormalverteilt wegen (E3).

$$\Rightarrow P\left(-z < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) \approx F_N(z) - F_N(-z) =: D(z). \quad (2)$$

Wenn man in (2) in den Ungleichungen alle Ausdrücke zuerst mit $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ multipliziert und anschliessend μ addiert, erhält man letztendlich

$$P(\mu - z\sigma_{\bar{X}} < \bar{X} \leq \mu + z\sigma_{\bar{X}}) \approx F_N(z) - F_N(-z) = D(z). \quad (3)$$

Der links von “ \approx ” stehende Ausdruck ist die Wahrscheinlichkeit, mit der ein Stichprobenmittel in ein gewisses symmetrisches Intervall um das Grundgesamtheitmittel μ fällt. Der rechts von “ \approx ” gibt dafür einen ungefähren Zahlenwert.

Aussagen dieser Art werden als *direkter Schluss* bezeichnet, da hier von der Grundgesamtheit auf die Stichprobe (hier genauer den Stichprobenmittelwert) geschlossen wird.

Es gibt drei typische Fragestellungen, die mit (3) beantwortet werden können, je nachdem was gegeben bzw. gesucht ist:

- Gegeben ist eine Grundgesamtheit und ein “Merkmal” X mit bekanntem $EX = \mu$ und $\text{var } X = \sigma^2$. Ziehe eine Stichprobe vom Umfang n . Wie gross ist die WSK p , dass das Stichprobenmittel \bar{X} in einem symmetrischen Intervall $[\mu - a, \mu + a]$ mit Mittelpunkt μ liegt, d.h. $p = P(\mu - a < \bar{X} \leq \mu + a)$?

Ein Vergleich mit (3) zeigt, dass wir z so bestimmen müssen, dass $a = z\sigma_{\bar{X}}$ ist. Das ergibt $z = \frac{a}{\sigma_{\bar{X}}}$. Mit diesem z liefert dann die rechte Seite von (3) die gesuchte WSK (zumindest eine Näherung), d.h.

$$p = D\left(\frac{a}{\sigma_{\bar{X}}}\right) = F_N\left(\frac{a}{\sigma_{\bar{X}}}\right) - F_N\left(-\frac{a}{\sigma_{\bar{X}}}\right).$$

- Gegeben ist eine Grundgesamtheit und ein “Merkmal” X mit bekanntem $EX = \mu$ und $\text{var } X = \sigma^2$. Der Umfang einer Stichprobe sei n . Vorgegeben ist eine WSK p . Wie gross muss man a in $[\mu - a, \mu + a]$ wählen, so dass das Stichprobenmittel mit einer WSK von p in diesem Intervall liegt?

Hier wollen wir, dass gilt:

$$P(\mu - a < \bar{X} \leq \mu + a) \approx D(z) = F_N(z) - F_N(-z) = 2F_N(z) - 1 = p.$$

Aus der letzten Gleichung kann man z bestimmen. Es gilt dann für das gesuchte a : $a = z\sigma_{\bar{X}}$.

- Gegeben ist eine Grundgesamtheit und ein “Merkmal” X mit bekanntem $EX = \mu$ und $\text{var } X = \sigma^2$. Darüber hinaus sind ein Intervall $[\mu - a, \mu + a]$ und eine WSK p vorgegeben. Wie gross muss der Umfang n der Stichprobe gewählt werden, damit $P(\mu - a < \bar{X} \leq \mu + a) = p$ gilt?

Wie gerade eben haben wir:

$$P(\mu - a < \bar{X} \leq \mu + a) \approx D(z) = F_n(z) - F_N(-z) = 2F_N(z) - 1 = p.$$

Das liefert z . Hier ist allerdings $a = z\sigma_{\bar{X}} = z\frac{\sigma}{\sqrt{n}}$ vorgegeben. n muss also so bestimmt werden, dass die letzte Gleichung erfüllt ist. Daraus folgt $n = \left(\frac{\sigma z}{a}\right)^2$.

Wir wollen die 3 Fälle mit Hilfe der folgenden Situation praktisch umsetzen:

Zur Anfängervorlesung in Statistik sind $N = 800$ Studenten ins Audimax gekommen. Ihre durchschnittliche Körpergrösse μ beträgt 183cm, bei einer Standardabweichung σ von 10cm. Ziehen wir daraus eine Zufallsstichprobe vom Umfang $n = 25$ mit Zurücklegen.

Bemerkung 9.8

Obwohl n mit $n = 25$ noch nicht “gross” ist, kann als Stichprobenverteilung die Normalverteilung angenommen werden, da die Ausgangsverteilung schon annähernd normalverteilt ist.

- Mit welcher Wahrscheinlichkeit wird das Stichprobenmittel \bar{X} ins Intervall [182cm, 184cm] fallen?

Zunächst gilt $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10\text{cm}}{5} = 2\text{cm}$. Beachtet, dass z so gewählt werden muss, dass gilt

$$[182\text{cm}, 184\text{cm}] = [183\text{cm} - z \cdot 2\text{cm}, 183\text{cm} + z \cdot 2\text{cm}].$$

Das liefert $z = \frac{1}{2}$. Für die gesuchte (approximative) WSK ergibt sich $F_N(0.5) - F_N(-0.5) = D(0.5) = 0.3830$.

- Wie gross ist das Intervall, in das der Stichprobenmittelwert mit einer Wahrscheinlichkeit von 0.9 fällt?

Dazu müssen wir den z -Wert so bestimmen, dass $D(z) = F_N(z) - F_N(-z) = 2F_N(z) - 1 = 0.9$, also $F_N(z) = 0.95$ gilt. Das ist für $z = 1.645$ der Fall. Nun gilt:

$$\begin{aligned} [183\text{cm} - z \cdot 2\text{cm}, 183\text{cm} + z \cdot 2\text{cm}] &= [183\text{cm} - 1.645 \cdot 2\text{cm}, 183\text{cm} + 1.645 \cdot 2\text{cm}] \\ &= [179.71\text{cm}, 186.29\text{cm}]. \end{aligned}$$

- Wie gross muss die Stichprobe gewählt werden, damit das Stichprobenmittel mit einer WSK von 95% ins Intervall [181.5cm, 184.5cm] fällt?
 z muss so bestimmt werden, dass $2F_N(z) - 1 = 0.95$ bzw. $F_N(z) = 1.95/2 = 0.975$. Das liefert $z = 1.96$. Da $a = 1.5$ ist, erhalten wir: $n = \left(\frac{\sigma z}{a}\right)^2 = \left(\frac{19.6}{1.5}\right)^2 = 170.74$, also $n = 171$.

Eine weitere Anwendung des ZGWS.

9.2.2 Approximation der Binomialverteilung durch die Normalverteilung

Sei eine binomialverteilte ZV X_b mit Verteilungsfunktion $F_b(x|\Theta; n)$ gegeben.

Faustregel: Falls $n\Theta(1 - \Theta) (= \sigma_X^2) \geq 9$ gilt, können Wahrscheinlichkeiten für X_b hinreichend genau mit der Normalverteilung $F_N(x|n\Theta; \sqrt{n\Theta(1 - \Theta)})$ berechnet werden. Man nimmt also zur Approximation von $F_b(x|\Theta; n)$ die Normalverteilung, die den gleichen EW $n\Theta$ und die gleiche Standardabweichung $\sqrt{n\Theta(1 - \Theta)}$ wie X_b hat.

Beispiel 9.9: Approximation Binomialverteilung durch Normalverteilung

In einer Fabrik werden täglich 3000 Artikel produziert. Die Wahrscheinlichkeit, dass in der Tagesproduktion ein Artikel fehlerhaft ist, beträgt 0.1. Es werde aus der Tagesproduktion eine zufällige Stichprobe von 100 Stück gezogen. Es sei X_b die Zufallsvariable, die die Anzahl der fehlerhaften Artikel in der Stichprobe angibt.

Gesucht ist die Wahrscheinlichkeit, dass in der Stichprobe mindestens 6 Artikel fehlerhaft sind.

Antwort: $P(X_b \geq 6) = 1 - P(X_b \leq 5) = 1 - F_b(5|0.1; 100) = 1 - 0.0576 = 0.9424$

Da $n\Theta(1 - \Theta) = 100 \cdot 0.1 \cdot 0.9 = 9$ die Faustregel erfüllt, wollen wir diese Wahrscheinlichkeit mittels Normalverteilung approximieren. Da $EX_b = n\Theta = 100 \cdot 0.1 = 10$ und $\text{var } X_b = n\Theta(1 - \Theta) = 9$, wählen wir die Normalverteilung mit den Parametern $\mu = 10$ und $\sigma = 3$.

$$P(X_b \geq 6) = 1 - P(X_b - B \leq 5) = 1 - F_b(5|0.1; 100) \approx 1 - F_N(5|10; 3)$$

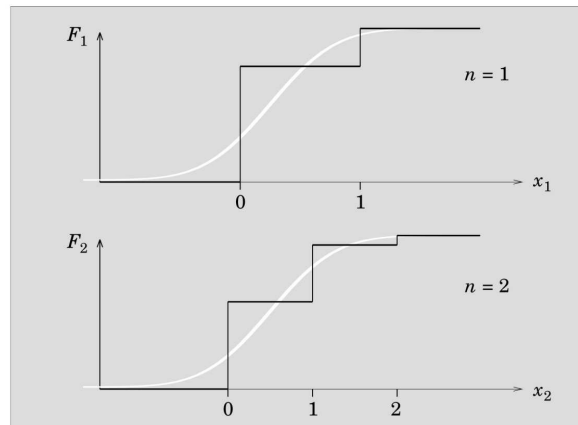
Wir bestimmen jetzt $F_N(5|10; 3) = P(X \leq 5)$. Standardisieren liefert:

$$\begin{aligned} P(X \leq 5) &= P\left(Z \leq \frac{5 - 10}{3}\right) = P\left(Z \leq -\frac{5}{3}\right) = 1 - P\left(Z \leq \frac{5}{3}\right) = 1 - F_N\left(\frac{5}{3}\right) \\ &= 1 - 0.9522 = 0.0478 \end{aligned}$$

Als Approximation für $P(X_b \geq 6)$ ergibt sich damit $1 - 0.0478 = 0.9522$.

Vergleicht man den exakten Wert 0.9424 mit der Approximation durch Normalverteilung (0.9522) fällt auf, dass die Approximation durch Normalverteilung "relativ schlecht" ist.

Möglicher Grund: Die Grafiken zeigen die typische Treppenform einer diskreten Verteilung und



den Verlauf der Normalverteilungsfunktion. Man sieht deutlich, dass die Werte der Normalverteilungsfunktion die Treppen der diskreten Verteilung ungefähr in der Mitte zwischen zwei Ausprägungen der diskreten ZV schneiden. Es ist daher bei der Approximation diskreter Verteilungen im Allgemeinen günstiger (obwohl es hierfür keine Garantie gibt), eine diskrete WSK der Form $P(X \leq k)$ im Normalverteilungskontext durch $P(X \leq k + Sk)$ zu approximieren, wobei Sk die Hälfte der Schrittweite zw. k und der nächsten Ausprägung der diskreten ZV beträgt. Das bezeichnet man als *Stetigkeitskorrektur*. Für die Binomialverteilung ist Sk immer 0.5!

Approximation mit Stetigkeitskorrektur:

Man ersetzt ganz einfach $F_N(5|10;3)$ durch $F_N(5 + 0.5|10;3)$. Man erhält dann nach Standardisierung $F_N(-1.5) = 1 - F_N(1.5) = 1 - 0.9332 = 0.0668$ und damit als Approximation der gesuchten ursprünglichen WSK 0.9332. Das ist zwar nur eine geringfügige Verbesserung, aber immerhin...

Ein weiteres Beispiel:

Beispiel 9.10

Ein fairer Würfel wird $n = 20$ mal geworfen. Wie gross ist die Wahrscheinlichkeit, dass die Zahl der 6er zwischen 3 und 10 liegt (3 und 10 eingeschlossen)? (Die Faustregel $n\Theta(1 - \Theta) \geq 9$ ist hier nicht erfüllt. Es geht nur um die Demonstration der Stetigkeitskorrektur!)

Exakt:

$$P(3 \leq X \leq 10) = F_b(10|1/6; 20) - F_b(2|1/6; 20) = 0.671236$$

Beachtet hier bitte die fettgedruckten Werte!

Mittels Normalverteilung ohne Stetigkeitskorrektur:

$$(\mu = n\Theta = 20/6, \sigma^2 = n\Theta(1 - \Theta) = 100/36)$$

$$\begin{aligned} P(3 \leq X \leq 10) &\approx F_N(10|20/6; 10/6) - F_N(2|20/6; 10/6) \\ &= F_N\left(\frac{10 - 20/6}{10/6}\right) - F_N\left(\frac{2 - 20/6}{10/6}\right) = 0.78811. \end{aligned}$$

Mittels Normalverteilung und Stetigkeitskorrektur:

$$\begin{aligned} P(3 \leq X \leq 10) &\approx F_N(10 + 0.5|20/6; 10/6) - F_N(2 + 0.5|20/6; 10/6) \\ &= F_N\left(\frac{10 + 0.5 - 20/6}{10/6}\right) - F_N\left(\frac{2 + 0.5 - 20/6}{10/6}\right) = 0.69145. \end{aligned}$$

Die Verbesserung der Approximation ist sehr deutlich.

10 Punktschätzungen (Parameterschätzungen)

Punktschätzungen sind Schätzverfahren, mit denen für gewisse Verteilungsparameter ein numerischer Wert produziert wird.

Allgemeines Prinzip:

θ gesucht \implies Schätzfunktion $T = g(X_1, X_2, \dots, X_n) \implies$ numerischer Wert $\hat{\theta} = g(x_1, x_2, \dots, x_n)$ durch Realisierungen x_1, \dots, x_n der Stichprobenvariablen X_1, \dots, X_n als Schätzwert für θ

10.1 Punktschätzung für den Mittelwert

Die Schätzfunktion für den Mittelwert hat die Form

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j. \quad (4)$$

Um eine Schätzung $\hat{\mu}$ für den Mittelwert zu bestimmen, braucht man also nur die Realisierungen x_1, x_2, \dots, x_n der ZV X_1, X_2, \dots, X_n in (4) einzusetzen:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j.$$

Beispiel 10.1: Siehe Beispiel 8.1

Wir wollen eine Schätzung $\hat{\mu}$ für das mittlere Gewicht μ aller Fische im See anhand unserer Stichprobe vom Umfang $n = 5$ bestimmen.

Die fünf gefangenen Fische hatten folgende Gewichte: $x_1 = 415g$, $x_2 = 603g$, $x_3 = 466g$, $x_4 = 530g$ und $x_5 = 591g$. (Nochmal: Jedes x_i ist Realisierung einer ZV X_i , die das Gewicht des i -ten gefangenen Fisches repräsentiert.) Daraus ergibt sich als Schätzung $\hat{\mu}$:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j = \frac{1}{5} (415g + 603g + 466g + 530g + 591g) = 521g.$$

Informell: Als Schätzung für den Mittelwert eines metrischen Merkmals wird das arithmetische Mittel einer Stichprobe verwendet.

10.1.1 Eigenschaften der Punktschätzung für den Mittelwert

Die Schätzfunktion (4) definiert (Das gilt allgemein für Schätzfunktionen.) als Kombination der ZV X_1, X_2, \dots, X_n eine neue ZV \bar{X} . Für diese ZV wollen wir den Erwartungswert und die Varianz bestimmen. Aus der Linearität des Erwartungswertes folgt:

$$E\bar{X} = E\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n} E\left(\sum_{j=1}^n X_j\right) = \frac{1}{n} (EX_1 + \dots + EX_n) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

Eine Schätzfunktion, deren Erwartungswert gleich dem zu schätzenden Parameter ist, heisst *erwartungstreu*. Unsere Schätzfunktion für den Mittelwert ist also erwartungstreu.

Varianz von \bar{X} : (Auch hier wird wieder die Unabhängigkeit der ZV X_1, \dots, X_n verwendet.)

$$\text{var } \bar{X} = \text{var}\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n^2} \text{var}\left(\sum_{j=1}^n X_j\right) = \frac{1}{n^2} (\text{var } X_1 + \dots + \text{var } X_n) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Wenn der Stichprobenumfang n grösser wird, wird die Varianz also immer kleiner, d.h. je grösser der Stichprobenumfang ist, desto grösser ist die Genauigkeit der Schätzung. Diese Eigenschaft wird als *Konsistenz* bezeichnet.

10.1.2 Schätzung des Anteilswerts als Spezialfall

Erinnert euch bitte an den Begriff des Bernoulli-Experiments. Das waren Experimente mit nur zwei interessierenden Ausgängen ("Erfolg" \longleftrightarrow "Misserfolg", "Hat Eigenschaft" \longleftrightarrow "Hat Eigenschaft nicht", "Tritt auf" \longleftrightarrow "Tritt nicht auf",...). Bernoulli-Experimente werden durch Bernoulli-Variablen X_B beschrieben, die nur die zwei Werte 1 (falls Erfolg etc.) und 0 (falls Misserfolg etc.) annehmen können. Die allgemeinen iuv ZV X_1, X_2, \dots, X_n von oben werden jetzt ersetzt durch n Bernoulli-Variablen $X_{B_1}, X_{B_2}, \dots, X_{B_n}$, die die Ergebnisse von n unabhängigen Wiederholungen des betrachteten Bernoulli-Experiments beschreiben. (Denkt z.B. wieder an das n -malige Werfen einer

Münze: X_{B_1} beschreibt das Ergebnis des ersten Wurfs, X_{B_2} das Ergebnis des zweiten Wurfs usw.) In diesem Kontext wird $EX_B = EX_{B_j} = \Theta$ häufig als *Anteilswert* bezeichnet. Als Mittelwert einer ZV wird er genau so geschätzt wie oben beschrieben.

Beispiel 10.2: Schätzung des Anteilswerts

In einer grossen Gruppe von Menschen soll der Anteil von Frauen geschätzt werden. Dazu werden 30 Personen zufällig gezogen und ihr Geschlecht ermittelt. (Bernoulli-Experiment: Ziehen einer Person; 30 Wiederholungen dieses Experiments, $X_{B_i} = 1$ falls i -te gezogene Person weiblich ist, sonst $X_{B_i} = 0$.) Nehmen wir an, 12 der gezogenen Personen waren weiblich (oder anders ausgedrückt: genau 12 der X_{B_i} hatten den Wert/Realisierung $b_i = 1$, die restlichen den Wert 0). Als Schätzung $\hat{\Theta}$ des wahren Anteilswerts Θ erhalten wir wie oben:

$$\hat{\Theta} = \frac{1}{30} \sum_{i=1}^{30} b_i = \frac{1}{30} \cdot 12 = 0.4$$

Der Wert $\hat{\Theta} = 0.4$ ist also eine (erwartungstreue) Schätzung für den Anteil von Frauen in der grossen Gruppe.

10.2 Punktschätzung für die Varianz

Es ist zunächst durchaus naheliegend, als Schätzung $\hat{\sigma}^2$ für die Varianz σ^2 einer ZV die Stichprobenvarianz $s^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$ zu verwenden, wobei \bar{x} wie immer das Stichprobenmittel bezeichnet. Das entspricht einer Schätzfunktion der Form

$$S^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2. \quad (5)$$

Nachteil:

$$ES^2 = \frac{n-1}{n} \sigma^2 = \sigma^2 - \frac{\sigma^2}{n}.$$

Der Erwartungswert der Schätzfunktion ist nicht gleich dem zu schätzenden Parameter σ^2 , sondern um den Wert $\frac{\sigma^2}{n}$ kleiner. Diese Schätzfunktion “unterschätzt” also systematisch den zu schätzenden Parameter! Diese Schätzfunktion ist nicht erwartungstreu!

Dieser Wert (d.h. der Wert, um den der Erwartungswert der Schätzfunktion vom zu schätzenden Parameter abweicht) heisst *Verzerrung* oder *bias*.

Behebung des Problems:

Schätzfunktion aus (5) “korrigieren”:

$$S^{*2} = \frac{n}{n-1} S^2 = \frac{n}{n-1} \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2. \quad (6)$$

Es gilt nun in der Tat: $ES^{*2} = \sigma^2$, d.h. diese Schätzfunktion **ist** erwartungstreu! (Der Faktor $\frac{n}{n-1}$ wird manchmal als *Bessel-Korrektur* bezeichnet.)

Was heisst das nun praktisch? Gegeben seien Realisierungen x_1, x_2, \dots, x_n unserer ZV X_1, X_2, \dots, X_n . Um eine erwartungstreue Schätzung $\hat{\sigma}^2$ der Varianz σ^2 zu bestimmen, geht man wie folgt vor:

- 1) Bestimme das Stichprobenmittel $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$;
- 2) Bestimme alle quadratischen Abweichungen $(x_j - \bar{x})^2$;
- 3) Addiere diese quadratischen Abweichungen: $\sum_{j=1}^n (x_j - \bar{x})^2$;
- 4) Teile das Resultat durch $n-1$: $\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2 = \hat{\sigma}^2$.

(Diese erwartungstreue Varianzschätzung wird im Bärtl mit s^{*2} bezeichnet!)

Beispiel 10.3: Wieder Situation von Beispiel 8.1

Wir führen diese Schritte anhand des Fischbeispiels durch:

- 1) Stichprobenmittel hatten wir schon bestimmt: $\bar{x} = 521 \text{ g}$.
- 2) $(415-521)^2 = 11236$, $(603-521)^2 = 6724$, $(466-521)^2 = 3025$, $(530-521)^2 = 81$, $(591-521)^2 = 4900$ (Alle Zahlen in Klammern haben Einheit g, die rechts von einem “=” die Einheit g^2).
- 3) Die Summe dieser Werte ergibt 25966 g^2 .
- 4) Teilen durch $n-1 = 4$ liefert $\hat{\sigma}^2 = 6491.5 \text{ g}^2$.

Bemerkungen 10.4

i) Als **Schätzung $\hat{\sigma}$ für die Standardabweichung σ** wird ganz einfach die Wurzel aus $\hat{\sigma}^2$ genommen: $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ ($= s^*$ im Bärtl). Beachtet aber bitte, dass diese Schätzung für σ **nicht** erwartungstreu ist! Auch hier wird systematisch “*unterschätzt*”. Für normalverteilte ZV lässt sich allerdings auch hier wieder ein Korrekturfaktor finden. Darauf wollen wir jedoch nicht eingehen.

ii) Wie schon erwähnt, ist auch $\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$ eine Schätzung für die Varianz. Sie ist aber **nicht erwartungstreu**. Passt also auf, ob nach einer Schätzung oder einer erwartungstreuen Schätzung gefragt wird. Im letzteren Fall **müsst** ihr $\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2$ verwenden, im ersteren Fall **könnt** ihr $\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2$ aber auch $\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$ verwenden. Nehmt also am besten immer die erwartungstreue Schätzung.

10.3 Ziehen mit und ohne Zurücklegen: Korrekturfaktor

Die Herleitungen der bisherigen Abschnitte basierten massgeblich auf der Voraussetzung, dass die ZV X_1, X_2, \dots, X_n unabhängig und identisch verteilt sind. Ob diese Voraussetzungen erfüllt sind, hängt u.a. davon ab, wie eine Stichprobe gezogen wurde. Im “Frauenanteilbeispiel” kann man die 30 Personen nacheinander ziehen, ohne sie wieder der Grundgesamtheit zuzuführen oder aber auch ziehen, Geschlecht bestimmen und wieder zurück in den “Topf” werfen. Im ersten Fall sind die ZV $X_{B_1}, X_{B_2}, \dots, X_{B_{30}}$, die die Versuchsausgänge beschreiben, weder unabhängig noch identisch verteilt. Zum Beispiel wird EX_{B_1} im Allgemeinen nicht gleich EX_{B_2} sein. Wie dem auch sei, in der Situation ohne Zurücklegen ändert sich auch z.B. $\text{var} \bar{X}$. Man kann nämlich zeigen, dass nun

$$\text{var} \bar{X} = \frac{N-n}{N-1} \frac{\sigma^2}{n}$$

ist, wobei N der Umfang der Grundgesamtheit ist. Vergleiche das bitte mit (E2)! Die Ergebnisse unterscheiden sich um den Faktor $\frac{N-n}{N-1}$, den sogenannten *Korrekturfaktor*. (Dieser ist für $n > 1$ immer kleiner Eins.)

Bemerkung 10.5

Beim Ziehen ohne Zurücklegen verkleinert sich also die Varianz der Mittelwertsvariable. Das ist plausibel, da man keine schon gewonnene Information “zerstört”.

Wir bestimmen jetzt wieder den Erwartungswert von S^2 :

$$ES^2 = E\left(\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2\right) = \frac{1}{n} \sum_{j=1}^n \text{var} X_j - \text{var} \bar{X} = \sigma^2 - \frac{N-n}{N-1} \frac{\sigma^2}{n} = \frac{N-1}{N} \frac{n-1}{n} \sigma^2$$

Die Verzerrung ist also nun $\frac{N-n}{N-1} \frac{\sigma^2}{n}$. Durch “Nachkorrigieren” kann man aber auch in diesem Fall eine erwartungstreue Schätzfunktion erhalten. Sie lautet nun:

$$S^{**2} = \frac{N-1}{N} \frac{n}{n-1} S^2 = \frac{N-1}{N} \frac{n}{n-1} \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 = \frac{N-1}{N} \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2 \quad (7)$$

Falls ihr also im **Kontext ohne Zurücklegen** operiert, geht ihr in der Praxis folgendermassen vor. Gegeben seien wieder Realisierungen x_1, x_2, \dots, x_n unserer (nicht mehr uiv) ZV X_1, X_2, \dots, X_n und der Umfang N der Grundgesamtheit. Um eine erwartungstreue Schätzung $\hat{\sigma}^2 (= s^{**2})$ der Varianz σ^2 zu bestimmen.

- 1) Bestimme das Stichprobenmittel $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$;
- 2) Bestimme alle quadratischen Abweichungen $(x_j - \bar{x})^2$;

- 3) Addiere diese quadratischen Abweichungen: $\sum_{j=1}^n (x_j - \bar{x})^2$;
- 4) Multipliziere das Resultat mit $\frac{N-1}{N} \cdot \frac{1}{n-1} \cdot \frac{N-1}{(n-1)N} \sum_{j=1}^n (x_j - \bar{x})^2 = \hat{\sigma}^2 (= s^{**2})$.

Bemerkungen 10.6

- i) Um diese Korrektur durchzuführen, muss natürlich wieder der Umfang N der Grundgesamtheit bekannt sein. Das ist oft nicht der Fall.
- ii) Für unendliche oder sehr grosse Grundgesamtheiten, d.h. $\frac{N-1}{N}$ dicht bei 1, ist dieser Korrekturfaktor irrelevant.
- iii) In der Praxis wird der Korrekturfaktor (häufig) vernachlässigt, falls $\frac{n}{N} < 0.05$. Es wird dann die Schätzfunktion aus (6) mit dem dort beschriebenen praktischen Vorgehen verwendet.
- iv) Da der Aufwand der Korrektur sehr gering ist, sollte bei bekanntem N immer (7) verwendet werden.

11 Intervallschätzungen

Nachteil von Punktschätzungen: Es ist im Prinzip nichts bekannt über die Genauigkeit der Schätzung (vgl. aber T-U). Abhilfe schaffen hier sogenannte Intervallschätzungen.

Ziel: Finde ein Schätzverfahren, dass mit Wahrscheinlichkeit $(1 - \alpha)$ ein Intervall liefert, in dem der zu schätzende Parameter liegt.

α wird hier die *Irrtumswahrscheinlichkeit* genannt und $(1 - \alpha)$ das *Konfidenzniveau*.

Häufige Werte für $(1 - \alpha)$: 0.99 (bzw. 99%), 0.95 (bzw. 95%) und 0.9 (bzw. 90%).

11.1 Intervallschätzungen für Mittelwerte

Wie ein Konfidenzintervall für Mittelwerte bestimmt wird, hängt in starkem Masse davon ab, welche Daten man gegeben hat! Ich möchte darauf kurz eingehen. Das wird auch wieder im Kontext von Hypothesentests für Mittelwerte relevant werden.

11.1.1 Fall 1: Stichprobenvariablen X_1, \dots, X_n unabh. normalverteilt mit Parametern μ und bekanntem σ^2

Wir wissen dann: $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim F_N$, d.h. die standardisierte Mittelwertsvariable ist exakt standardnormalverteilt.

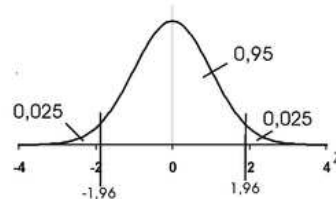
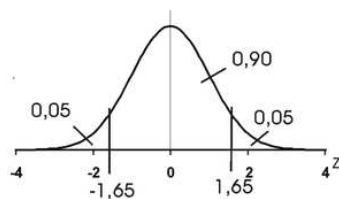
$$\Rightarrow P\left(-z \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) = P\left(\bar{X} - z \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z \frac{\sigma}{\sqrt{n}}\right) = F_N(z) - F_N(-z) = D(z)$$

(Die erste Gleichung folgt durch Umstellen der Ungleichungen nach μ .) Beachtet bitte auch nochmal, dass σ/\sqrt{n} nach (E2) nichts anderes ist als die Standardabweichung $\sigma_{\bar{X}}$ von \bar{X} . $\sigma_{\bar{X}}$ wird auch als der *Standardfehler* des Mittelwerts bezeichnet.

z muss jetzt so bestimmt werden, dass $D(z) = 1 - \alpha$ gilt. Wie wird das gemacht? Erinnert euch:

$$F_N(z) - F_N(-z) = F_N(z) - (1 - F_N(z)) = 2F_N(z) - 1 \stackrel{!}{=} 1 - \alpha \Rightarrow F_N(z) \stackrel{!}{=} 1 - \frac{\alpha}{2}$$

z muss also so bestimmt werden, dass $P(Z \leq z) = F_N(z) = (1 - \frac{\alpha}{2})$. Ich bezeichne diesen z -Wert (Es handelt sich hier um das $(1 - \frac{\alpha}{2})$ -Quantil der Standardnormalverteilung!) im Weiteren mit $z[1 - \frac{\alpha}{2}]$. Siehe Abbildungen.



Für die wichtigsten Konfidenzniveaus erhält man die folgenden Werte:

$$(1 - \alpha) = 0.9 \Rightarrow z = 1.645; (1 - \alpha) = 0.95 \Rightarrow z = 1.96; (1 - \alpha) = 0.99 \Rightarrow z = 2.575$$

Für die eigentliche Berechnung des Konfidenzintervalls braucht man natürlich noch einen Zahlenwert für \bar{X} . Den bekommt man wie immer als Mittelwert einer gegebenen Stichprobe. Nach Ersetzung von \bar{X} durch eine Schätzung \bar{x} schreibt man:

$$KI(\mu, 1 - \alpha) = \left[\bar{x} - z\left[1 - \frac{\alpha}{2}\right] \frac{\sigma}{\sqrt{n}}, \bar{x} + z\left[1 - \frac{\alpha}{2}\right] \frac{\sigma}{\sqrt{n}} \right]$$

Bitte macht euch klar, dass mit steigendem Konfidenzniveau die Intervalle immer grösser werden. Das gilt allgemein für Intervallschätzungen. Darüber hinaus hängt die Breite des Konfidenzintervalls natürlich auch von der Stichprobengrösse ab. Wie?

11.1.2 Fall 2: Stichprobenvariablen X_1, \dots, X_n unabh. normalverteilt mit Parametern μ und unbekanntem σ^2

Hier muss σ^2 geschätzt werden. Man nimmt dafür die erwartungstreue Schätzfunktion

$$S^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ aus dem Abschnitt über Punktschätzungen.}$$

Man kann dann zeigen: $\frac{\bar{X} - \mu}{S^*/\sqrt{n}} \sim T_{n-1}$, wobei T_{n-1} die t -Verteilung mit $n-1$ Freiheitsgraden ist.

Die Bestimmung eines $(1 - \alpha)$ -Konfidenzintervalls folgt jetzt exakt wie in Fall 1 mit dem Unterschied, dass jetzt anstatt der Normalverteilung die t -Verteilung verwendet wird.

$$P\left(-t \leq \frac{\bar{X} - \mu}{S^*/\sqrt{n}} \leq t\right) = P\left(\bar{X} - t \frac{S^*}{\sqrt{n}} \leq \mu \leq \bar{X} + t \frac{S^*}{\sqrt{n}}\right) = T_{n-1}(t) - T_{n-1}(-t)$$

t muss jetzt wieder so bestimmt werden, dass die Differenz $T_{n-1}(t) - T_{n-1}(-t)$ gerade wieder $1 - \alpha$ beträgt. Dazu bestimmt man das $(1 - \frac{\alpha}{2})$ -Quantil $t_{n-1}[1 - \frac{\alpha}{2}]$ der t -Verteilung mit $n-1$ Freiheitsgraden. Um ein konkretes Konfidenzintervall zu bestimmen, ersetzt man nun wieder \bar{X} und S^* durch ihre Realisierungen aus der Stichprobe:

$$\begin{aligned} \text{KI}(\mu, 1 - \alpha) &= \left[\bar{x} - t_{n-1}\left[1 - \frac{\alpha}{2}\right] \frac{s^*}{\sqrt{n}}, \bar{x} + t_{n-1}\left[1 - \frac{\alpha}{2}\right] \frac{s^*}{\sqrt{n}} \right] \\ &= \left[\bar{x} - t_{n-1}\left[1 - \frac{\alpha}{2}\right] \frac{s}{\sqrt{n-1}}, \bar{x} + t_{n-1}\left[1 - \frac{\alpha}{2}\right] \frac{s}{\sqrt{n-1}} \right] \end{aligned}$$

Beachtet bitte nochmal den Unterschied:

$$s^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}; \quad s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Wenn diese Grössen nicht selber berechnet werden müssen, sollte aus der Aufgabenstellung hervorgehen, was gegeben ist.

Bemerkung 11.1

In den Fällen 1 und 2 werden exakte Konfidenzintervalle berechnet.

11.1.3 Fall 3: X_1, \dots, X_n unabh. beliebig identisch verteilte ZV und $n \geq 30$

Hier kommt wieder der ZGWS ins Spiel. Nach diesem gilt (E3): $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \underset{\text{appr.}}{\sim} F_N$, d.h. die standardisierte Mittelwertsvariable ist **approximativ** standardnormalverteilt. Das gilt ebenfalls, wenn die

(exakte) Standardabweichung σ durch eine Schätzung s^* ersetzt wird. Zusammenfassend haben wir mit den schon eingeführten Bezeichnungen:

Falls σ^2 bekannt:

$$KI(\mu, 1 - \alpha) = \left[\bar{x} - z\left[1 - \frac{\alpha}{2}\right] \frac{\sigma}{\sqrt{n}}, \bar{x} + z\left[1 - \frac{\alpha}{2}\right] \frac{\sigma}{\sqrt{n}} \right]$$

Falls σ^2 unbekannt:

$$\begin{aligned} KI(\mu, 1 - \alpha) &= \left[\bar{x} - z\left[1 - \frac{\alpha}{2}\right] \frac{s^*}{\sqrt{n}}, \bar{x} + z\left[1 - \frac{\alpha}{2}\right] \frac{s^*}{\sqrt{n}} \right] \\ &= \left[\bar{x} - z\left[1 - \frac{\alpha}{2}\right] \frac{s}{\sqrt{n-1}}, \bar{x} + z\left[1 - \frac{\alpha}{2}\right] \frac{s}{\sqrt{n-1}} \right] \end{aligned}$$

Da wir hier nur mit einer approximativen Verteilung arbeiten, erhalten wir auch nur ein approximatives Konfidenzintervall,... but who cares.

Wie sieht's nun im Bärli aus. Eigentlich genau so. (Alles andere wäre auch fatal.) Er bewegt sich ausschliesslich im Kontext der Fälle 2 und 3, macht aber noch einen Unterschied im Hinblick auf die Art und Weise, wie die Stichprobe gezogen wird, mit oder ohne Zurücklegen, bzw. ob das praktisch relevant ist oder nicht. Das hat Auswirkungen auf die Schätzung von σ .

11.1.4 Intervallschätzungen für Mittelwerte mit grossen Stichproben (also $n \geq 30$), d.h. Fall 3)

Wir setzen weiter voraus $\frac{N-n}{N-1} \geq 0.9$. (Es macht also praktisch keinen grossen Unterschied, ob die Stichprobe mit oder ohne Zurücklegen gezogen wurde, da sich s^* und s^{**} wenig unterscheiden. (Vgl. Abschnitt 10.3 Ziehen mit oder ohne Zurücklegen))

Praktisch geht man folgendermassen vor:

- Gegeben eine Stichprobe vom Umfang n . Bestimme Stichprobenmittelwert \bar{x} (falls nicht schon gegeben).
- Gegeben ein Konfidenzniveau $(1 - \alpha)$ bzw. eine Irrtumswahrscheinlichkeit α . Bestimme den dazugehörigen z -Wert $z\left[1 - \frac{\alpha}{2}\right]$.
- Ist die Standardabweichung σ (oder die Varianz σ^2) in der Grundgesamtheit gegeben? Wenn ja, dann ist $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ und das $(1 - \alpha)$ -Konfidenzintervall ist nach Fall 3 gegeben durch:

$$\left[\bar{x} - z\left[1 - \frac{\alpha}{2}\right] \frac{\sigma}{\sqrt{n}}, \bar{x} + z\left[1 - \frac{\alpha}{2}\right] \frac{\sigma}{\sqrt{n}} \right]$$

- Falls nicht (Dann ist häufig die Stichprobenvarianz s^2 gegeben oder man muss sie aus der Stichprobe berechnen.), muss eine Schätzung $\hat{\sigma}_{\bar{X}}$ bestimmt werden.

$$\hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}_X}{\sqrt{n}} = \frac{\sqrt{\frac{n}{n-1}}s}{\sqrt{n}} = \frac{s}{\sqrt{n-1}} \quad (\text{für } s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2})$$

bzw.

$$\hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}_X}{\sqrt{n}} = \frac{s^*}{\sqrt{n}} \quad (\text{für } s^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2})$$

Das $(1 - \alpha)$ -Konfidenzintervall ist also (wieder nach **Fall 3**) gegeben durch:

$$\left[\bar{x} - z\left[1 - \frac{\alpha}{2}\right] \frac{s}{\sqrt{n-1}}, \bar{x} + z\left[1 - \frac{\alpha}{2}\right] \frac{s}{\sqrt{n-1}} \right] \quad \text{bzw.} \quad \left[\bar{x} - z\left[1 - \frac{\alpha}{2}\right] \frac{s^*}{\sqrt{n}}, \bar{x} + z\left[1 - \frac{\alpha}{2}\right] \frac{s^*}{\sqrt{n}} \right]$$

Typische Situation: Vorgegeben ist ein Konfidenzniveau $(1 - \alpha)$ und ein “Merkmal” X . Gegeben ist weiterhin der Mittelwert \bar{x} einer Stichprobe oder die Stichprobe selber (sodass man \bar{x} bestimmen kann). Weiterhin ist oft die empirische Stichprobenvarianz s^2 gegeben (oder eben die Stichprobe, sodass man diese berechnen kann). Man bestimme ein $(1 - \alpha)$ -Konfidenzintervall für den Mittelwert μ !

Beispiel 11.2: Intervallschätzung für Mittelwert bei grosser Stichprobe

In einer grossen Stadt (d.h. N gross) hat eine Stichprobe von $n = 180$ Personen ergeben, dass jede Person im Mittel $\bar{x} = 0.82$ Fast-Food-Essen in den vergangenen 7 Tagen hatte bei einer Standardabweichung $s = 0.48$. Man bestimme ein 95% Konfidenzintervall für das Mittel der Fast-Food-Essen in der Gesamtpopulation! (95% Konfidenzintervall entspricht selbstverständlich dem $(1 - 0.05)$ -Konfidenzintervall).

- N nicht gegeben. Wir können also $\frac{N-n}{N-1} \geq 0.9$ nicht überprüfen. Da N aber als gross angenommen wird, können wir davon ausgehen, dass die Bedingung erfüllt ist.
- Stichprobenmittel schon gegeben.
- Zu einem Konfidenzniveau von 95% gehört ein z -Wert von $z[0.975] = 1.96$.
- Die exakte Standardabweichung σ_X ist nicht gegeben, also müssen wir schätzen: $\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n-1}} = \frac{0.48}{\sqrt{179}} \approx 0.0359$.
- Das 95%-Konfidenzintervall ist also gegeben durch

$$[0.82 - 1.96 \cdot 0.0359, 0.82 + 1.96 \cdot 0.0359] = [0.7497, 0.8903].$$

11.1.5 Intervallschätzungen für den Mittelwert eines normalverteilten Merkmals bei kleinen Stichproben (Verletzung von $n \geq 30$), d.h. Fall 2

Beispiel 11.3: Intervallschätzung für Mittelwert bei kleiner Stichprobe

Wir betrachten die Produktion von Brotlaiben und nehmen eine Zufallsstichprobe vom Umfang $n = 12$ (d.h. n ist **nicht** gross!). Dabei wird ein Gesamtgewicht von 5904g bei einer Standardabweichung von 4.85g ermittelt. Vorausgesetzt das Einzelgewicht ist **normalverteilt**, bestimme man ein 95% Konfidenzintervall für den Mittelwert μ des Gewichts der Brotlaibe in der Gesamtproduktion!

- N nicht gegeben. Wir können also $\frac{N-n}{N-1} \geq 0.9$ nicht überprüfen. Da N aber als gross angenommen werden kann, können wir davon ausgehen, dass die Bedingung erfüllt ist.
- Stichprobenmittelwert bestimmen: $\bar{x} = 5904g/12 = 492g$.
- $t_{11}[1 - \frac{\alpha}{2}] = t_{11}[0.975]$. **Diesen Wert findet man in der Bärtl-Tabelle unter 0.95 (zweiseitig), wobei man natürlich in der Zeile mit 11 Freiheitsgraden schauen muss. Für ein gegebenes α muss man also im Bärtl immer unter $1 - \alpha$ (zweiseitig) schauen.** Daraus folgt $t_{11}[0.975] = 2.201$.
- $\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{11}} = 1.462$
- Für das gesuchte Konfidenzintervall erhalten wir also:

$$[492 - 2.201 \cdot 1.462, 492 + 2.201 \cdot 1.462] = [488.78, 495.22].$$

11.1.6 Intervallschätzungen für den Mittelwert bei Verletzung von $\frac{N-n}{N-1} \geq 0.9$

Ist diese Bedingung verletzt, wird wieder “nachkorrigiert”. Konkret:

$$\begin{aligned} \text{var } \bar{X} &= \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \frac{N-n}{N-1}, \quad \text{falls } \sigma \text{ in Grundgesamtheit bekannt ist bzw.} \\ \text{var } \bar{X} &= \sigma_{\bar{X}}^2 \approx \frac{s^{*2}}{n} \frac{N-n}{N-1} \quad (\text{oder } \sigma_{\bar{X}}^2 \approx \frac{s^2}{n-1} \frac{N-n}{N-1}), \quad \text{falls das nicht der Fall ist.} \end{aligned}$$

Alles andere ist genauso wie in Fall 2 und Fall 3 beschrieben, nur wird jetzt folgendermassen ersetzt:

$$\frac{\sigma}{\sqrt{n}} \quad \text{durch} \quad \sqrt{\frac{\sigma^2}{n} \frac{N-n}{N-1}} \quad \text{bzw.} \quad \frac{s^*}{\sqrt{n}} \quad \text{durch} \quad \sqrt{\frac{s^{*2}}{n} \frac{N-n}{N-1}}$$

11.1.7 Intervallschätzungen für den Anteilswert als Spezialfall von Fall 3

Die zugrundeliegenden Stichprobenvariablen X_{B_1}, \dots, X_{B_n} sind hier Bernoulli-verteilt (D.h. natürlich insbesondere, dass wir **nie** in der Situation von Fall 2 sind, da dort Normalverteilung vorausgesetzt wird!) mit (unbekannter) Erfolgswahrscheinlichkeit (Auftrittenswahrscheinlichkeit, Anteilswert) Θ .

Intervallschätzungen für den Anteilswert sind also Spezialfälle des allgemeinen Falls und werden analog durchgeführt. erinnert euch bitte, dass der Anteilswert aus einer Stichprobe wie folgt geschätzt wird:

$$\hat{\Theta} = p = \frac{\text{Anzahl Elemente mit gewünschter Eigenschaft}}{\text{Anzahl Elemente in Stichprobe}} = \frac{m}{n}$$

p fungiert hier als Analogon zu \bar{x} im allgemeinen Fall.

Für die Varianz der entsprechenden Mittelwertsvariable P (Notation Bär!!) (Analogon zu \bar{X}) erhält man dann

$$\text{var} P \approx \frac{p(1-p)}{n}$$

Ihr solltet inzwischen in der Lage sein, das nachzuvollziehen. Ich werde euch aber nochmal auf die Sprünge helfen. Beachtet dazu nochmal, dass $E(X_{B_i}) = \Theta$ und $\text{var} X_{B_i} = \Theta(1 - \Theta)$.

$$\text{var} P = \text{var} \left(\frac{1}{n} (X_{B_1} + \dots + X_{B_n}) \right) \stackrel{(E2)}{\implies} \text{var} P = \frac{\Theta(1 - \Theta)}{n}.$$

Jetzt wird $\text{var} P$ dadurch approximiert, dass man Θ durch seine Schätzung $\hat{\Theta} = p$ ersetzt.

Bemerkung 11.4

Es macht in diesem Kontext keinen Sinn, von bekanntem $\sigma^2 (= \Theta(1 - \Theta))$ auszugehen, da sich daraus der zu bestimmende Parameter Θ berechnen lässt.

Diese Größen ersetzen die entsprechenden Größen aus dem allgemeinen Fall im Konfidenzintervall. Man erhält also als $(1 - \alpha)$ -Konfidenzintervall für den wahren Anteilswert Θ in der Grundgesamtheit:

$$\left[p - z \left[1 - \frac{\alpha}{2} \right] \sqrt{\frac{p(1-p)}{n}}, p + z \left[1 - \frac{\alpha}{2} \right] \sqrt{\frac{p(1-p)}{n}} \right]$$

Bei Verletzung von $\frac{N-n}{N-1} \geq 0.9$ wird auch hier häufig “nachkorrigiert”:

$$\text{var} P \approx \frac{p(1-p)}{n} \frac{N-n}{N-1}.$$

Beispiel 11.5

Betrachten wir nochmal das “Frauenbeispiel” 10.2. Dort wurde eine Stichprobe von $n = 30$ gezogen. $m = 12$ der gezogenen Personen waren weiblich. Nehmen wir an, der Umfang der Grundgesamtheit war $N = 130$. Da $\frac{N-n}{N-1} = \frac{100}{129} \approx 0.78 \leq 0.9$, sollte korrigiert werden. Anteilswert wurde bereits geschätzt: $p = \frac{12}{30} = 0.4$. Für die Schätzung der Varianz von P (mit Korrektur) erhalten wir:

$$\text{var}P \approx \frac{p(1-p)}{n} \frac{N-n}{N-1} = \frac{0.4 \cdot (1-0.4)}{30} \frac{130-30}{129} = 0.0062 \quad \text{und somit} \quad \sqrt{\text{var}P} \approx 0.079$$

Wir wollen das 99% Konfidenzintervall für den wahren Anteilswert Θ der Frauen bestimmen. z -Wert ist rund 2.58. Wir erhalten als Konfidenzintervall:

$$\left[p - z \sqrt{\frac{p(1-p)(N-n)}{n(N-1)}}, p + z \sqrt{\frac{p(1-p)(N-n)}{n(N-1)}} \right] = [0.4 - 2.58 \cdot 0.079, 0.4 + 2.58 \cdot 0.079] = [0.196, 0.604]$$

12 Zwei-dimensionale Verteilungen

12.1 Streudiagramme und Kontingenztafeln

Wir hatten Streudiagramme und Kontingenztafeln schon im Abschnitt 3 im Zusammenhang mit “Zusammenhangsmassen” von zwei Merkmalen (etwas voreilig) eingeführt. Im Beispiel 3.3 hatten wir z.B. mit folgender Kontingenztafel der (absoluten) Häufigkeiten zu tun:

	1.0	1.25	1.5	1.75	
20	61	57	11	0	129
25	34	114	65	38	251
	95	171	76	38	380

Die entsprechende Kontingenztafel bzgl. der relativen Häufigkeiten sieht demnach so aus (Alle Werte durch 380 teilen! Resultate gerundet):

	1.0	1.25	1.5	1.75	
20	0.16	0.15	0.03	0	0.34
25	0.09	0.3	0.17	0.1	0.66
	0.25	0.45	0.2	0.1	1

Im Allgemeinen hat also eine Kontingenztafel die folgende Form:

	y_1	y_2	\cdots	y_j	\cdots	y_l	
x_1	n_{11}	n_{12}	\cdots	n_{1j}	\cdots	n_{1l}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	\cdots	n_{2j}	\cdots	n_{2l}	$n_{2\bullet}$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
x_i	n_{i1}	n_{i2}	\cdots	n_{ij}	\cdots	n_{il}	$n_{i\bullet}$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
x_k	n_{k1}	n_{k2}	\cdots	n_{kj}	\cdots	n_{kl}	$n_{k\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	\cdots	$n_{\bullet j}$	\cdots	$n_{\bullet l}$	n

Es folgt die Beschreibung der Grössen in obiger Tabelle.

- n ist die Anzahl der (nicht notwendigerweise verschiedenen) Wertepaare in der Urliste.
- Die x_i bzw. y_j sind die unterschiedlichen bzgl. der Merkmale X und Y beobachteten Ausprägungen. (Sie werden oft der Grösse nach geordnet (falls die Merkmale zumindest ordinal sind). Das ist aber nicht zwingend nötig.)
- n_{ij} gibt an, wie häufig Ausprägung x_i zusammen mit Ausprägung y_j aufgetreten ist, also die Anzahl der Wertepaare (x_i, y_j) in der Urliste oder noch anders ausgedrückt:

$$n_{ij} = \text{absH}(X = x_i \text{ und } Y = y_j).$$

- $n_{i\bullet}$ ist die Summe der Werte in der i -ten Zeile, d.h.

$$n_{i\bullet} = n_{i1} + n_{i2} + \cdots + n_{il} = \text{absH}(X = x_i).$$

- $n_{\bullet j}$ ist die Summe der Werte in der j -ten Spalte, d.h.

$$n_{\bullet j} = n_{1j} + n_{2j} + \cdots + n_{kj} = \text{absH}(Y = y_j).$$

Beachtet bitte, dass sich, wie immer in doppelindizierten Grössen, der erste Index auf die Zeile und der zweite Index auf die Spalte bezieht.

Die Grössen mit einem Punkt als Index an den "Rändern" der Tabelle sind für die Beschreibung der Einzelmerkmale wichtig. Darauf kommen wir in Kürze zurück.

Anstatt die absoluten Häufigkeiten aufzuzeichnen, kann man gerade so gut, wie wir das im Eingangsbeispiel schon gemacht haben, die relativen Häufigkeiten verwenden. Dazu muss man nur sämtliche Werte in der obigen Tabelle durch n teilen. Der Vollständigkeit halber sei die allgemeine Form angegeben:

	y_1	y_2	\cdots	y_j	\cdots	y_l	
x_1	n_{11}^*	n_{12}^*	\cdots	n_{1j}^*	\cdots	n_{1l}^*	$n_{1\bullet}^*$
x_2	n_{21}^*	n_{22}^*	\cdots	n_{2j}^*	\cdots	n_{2l}^*	$n_{2\bullet}^*$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
x_i	n_{i1}^*	n_{i2}^*	\cdots	n_{ij}^*	\cdots	n_{il}^*	$n_{i\bullet}^*$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
x_k	n_{k1}^*	n_{k2}^*	\cdots	n_{kj}^*	\cdots	n_{kl}^*	$n_{k\bullet}^*$
	$n_{\bullet 1}^*$	$n_{\bullet 2}^*$	\cdots	$n_{\bullet j}^*$	\cdots	$n_{\bullet l}^*$	1

Es ist also z.B. $n_{ij}^* = \frac{n_{ij}}{n}$. Bei der Interpretation der neuen Werte muss man nur “absolut” durch “relativ” ersetzen.

Beachtet bitte, dass in der “absoluten” Kontingenztafel die Summe der Randzeilen- bzw. Randspaltenwerte immer gleich n ist, in der “relativen” Kontingenztafel jedoch immer 1.

Bemerkung 12.1

Die Werte der relativen Kontingenztafel lassen sich auch wahrscheinlichkeitstheoretisch interpretieren, wenn man die rel. Häufigkeiten als Wahrscheinlichkeiten interpretiert. Für Beispiel 3.3: Man zieht zufällig einen Baum. Was ist z.B. die WSK, dass er 20m hoch ist? Antwort: 0.34. Was ist die WSK, dass er 25m hoch ist und einen Stammumfang von 1.5m hat? Antwort: 0.17 etc.

12.2 Randverteilungen

Im letzten Abschnitt haben wir Merkmale gemeinsam betrachtet und ihre Verteilung mittels Kontingenztafel beschrieben. Natürlich kann man auch jedes Merkmal separat betrachten und seine Verteilung so bestimmen, wie wir das in früheren Abschnitten getan haben. Die Frage ist, ob das anhand der Kontingenztafel möglich ist. Die Antwort ist ein klares Ja. Und genau hier kommen die “Ränder” der Kontingenztafel ins Spiel. erinnert euch bitte:

$$n_{i\bullet}^* = \text{relH}(X = x_i) \quad \text{und} \quad n_{\bullet j}^* = \text{relH}(Y = y_j).$$

Durch diese relativen Häufigkeiten, die sogenannten *Randverteilungen* von X und Y (nomen est omen) sind aber sowohl X als auch Y eindeutig bestimmt. Wenn ihr euch jetzt noch eben die Berechnung des arithmetischen Mittels mit Hilfe der relativen Häufigkeiten vergegenwärtigt (und analog der Varianz) sollten die folgenden Formeln nicht überraschen:

$$\bullet \quad \bar{x} = \sum_{i=1}^k n_{i\bullet}^* x_i, \quad \bar{y} = \sum_{j=1}^l n_{\bullet j}^* y_j$$

$$\bullet \quad s_X^2 = \sum_{i=1}^k n_{i\bullet}^* (x_i - \bar{x})^2, \quad s_Y^2 = \sum_{j=1}^l n_{\bullet j}^* (y_j - \bar{y})^2$$

Beispiel 12.2

Wir wollen für unser Hartholz-Beispiel das Mittel und die Varianz für das Merkmal Stamm-durchmesser D berechnen. Hier haben wir die Ausprägungen 1.0, 1.25, 1.5 und 1.75 mit den in der letzten Zeile der relativen Kontingenztafel gegebenen relativen Häufigkeiten.

$$\bar{d} = 0.25 \cdot 1.0 + 0.45 \cdot 1.25 + 0.2 \cdot 1.5 + 0.1 \cdot 1.75 = 1.2875.$$

Für die empirische Varianz erhalten wir:

$$s_D^2 = 0.25 \cdot (1.0 - 1.2875)^2 + 0.45 \cdot (1.25 - 1.2875)^2 + 0.2 \cdot (1.5 - 1.2875)^2 + 0.1 \cdot (1.75 - 1.2875)^2 \approx 0.0517.$$

13 Testen von Hypothesen: Binomialtest, t -Tests, χ^2 -Tests

Bei statistischen Tests wird überprüft, ob bestimmte Annahmen über einen Parameter oder Verteilung in der Grundgesamtheit (wahrscheinlich) zutreffen oder nicht. Dabei ist es unerheblich, wie diese Annahmen gewonnen wurden. Die Überprüfung basiert wieder auf der Grundlage einer Stichprobe. Bei Hypothesentests wird einer *Hypothese* H_0 eine *Alternativ-* oder *Gegenaussage* H_1 gegenübergestellt, die jedoch nicht das logische Komplement von H_0 zu sein braucht. Die Rolle der Stichprobe kann man wie folgt zusammenfassen:

- Sie liefert eine konkrete Realisierung, anhand derer man die Aussage überprüft.
- Wenn das konkrete Resultat der Stichprobe **sehr** unwahrscheinlich ist unter den Modellannahmen von H_0 , dann wird die Nullhypothese H_0 zugunsten der Alternativhypothese H_1 verworfen, andernfalls wird sie beibehalten.
- Was unwahrscheinlich heisst, wird bestimmt durch ein willkürlich vorgegebenes *Signifikanz-* oder *Testniveau* α . Häufig wird $\alpha = 1\%$, 5% oder 10% gewählt.

Hypothesentests laufen immer nach dem gleichen groben Schema ab:

Aufstellen Nullhypothese $H_0 \rightarrow$ Aufstellen Alternativhypothese $H_1 \rightarrow$ Festlegen des Testniveaus $\alpha \rightarrow$ Wahl einer geeigneten Teststatistik \rightarrow Testentscheid.

Die Teststatistik ist situationsbedingt, je nach Testproblem.

Die Teststatistik muss eine Entscheidung anhand der Stichprobe ermöglichen, ob H_0 oder H_1 akzeptiert wird.

Verteilung der Teststatistik unter H_0 muss bekannt sein, zumindest approximativ.

Was das heisst, wollen wir anhand einiger ausgewählter Hypothesentests erläutern.

13.1 Binomialtest

Ein Binomialtest ist ein statistischer Test, bei dem die Teststatistik binomialverteilt ist. Er wird verwendet, um Hypothesen über Merkmale zu prüfen, die genau zwei Ausprägungen annehmen können (*dichotome Merkmale*).

Wir werden diesen Test und insbesondere das konkrete Beispiel benutzen, um die Grundideen von Hypothesentests zu beleuchten. Diese gelten gleichermassen für alle anderen Tests.

Situation: Gesamtpopulation. Babies (einer Stadt, eines Landes etc.)

Nullhypothese H_0 : Babies haben keine Farbpräferenz.

Gegenhypothese H_1 : Babies haben Farbpräferenz.

Stichprobe: Wählen zufällig 30 Babies. 25 wählen rote Rassel, 5 grüne Rassel.

Ist das genug Evidenz, um H_0 zugunsten H_1 zu verwerfen?

Quantifizierung des Problems:

Bernoullivariablen für $i = 1, \dots, 30$:

$$X_i = \begin{cases} 1, & \text{falls Baby } i \text{ 'R' wählt} \\ 0, & \text{falls Baby } i \text{ 'G' wählt} \end{cases}$$

Die X_i sind unabhängige, Bernoulli-verteilte ZV mit Parameter Θ , der exakt bestimmt werden könnte, wenn alle Babies “durchgetestet” würden.

Stichprobe suggeriert: $\Theta = \frac{5}{6}$. Ist das statistisch signifikant?

Aufstellen Hypothesen:

$H_0 : \Theta = \frac{1}{2}$ (entspricht “Keine Farbpräferenz”)

$H_1 : \Theta \neq \frac{1}{2}$ (entspricht “Farbpräferenz”, zweiseitiger Test) oder

$H_1' : \Theta > \frac{1}{2}$ (entspricht “Farbpräferenz für Rot”, rechts- oder oberseitiger Test) oder

$H_1'' : \Theta < \frac{1}{2}$ (entspricht “Farbpräferenz für Grün”, links- oder unterseitiger Test)

Wir wollen das anhand von Gegenhypothese H_1 “durchexerzieren”.

Teststatistik: Die Teststatistik ist allg. eine ZV, die von den mit der Stichprobe verbundenen ZV

abhängt und situationsbezogen ist: $T(X_1, X_2, \dots, X_n)$. (Keine Angst, ihr werdet in WS nie eine Teststatistik “basteln” müssen.) Eine konkrete Stichprobe liefert eine Realisierung $t = T(x_1, x_2, \dots, x_n)$ dieser Teststatistik.

Bei “Keiner Farbpräferenz” sollte man bei 30 Tests 15 mal “Rot” erwarten. Weichen die Werte “zu sehr” nach oben oder unten von 15 ab, sollte man H_0 zugunsten von H_1 verwerfen. Es ist also relativ naheliegend, als Teststatistik die absolute Abweichung der tatsächlich aufgetretenen “Rot-Wahlen” von dieser erwarteten Anzahl zu benutzen. Die aufgetretene Anzahl wird aber beschrieben durch die ZV $X = \sum_{i=1}^{30} X_i$. Die absolute Abweichung von 15 ist also gegeben durch:

$$T(X_1, X_2, \dots, X_n) = |X - 15| = \left| \sum_{i=1}^{30} X_i - 15 \right|$$

In unserer konkreten Situation liefert das $t = |25 - 15| = 10$.

Wir hatten gesagt, dass die Verteilung der Teststatistik unter H_0 zumindest approximativ bekannt sein muss. Wir wissen aber:

i) $X \sim f_b(x | \frac{1}{2}; 30)$ (unter H_0) \Rightarrow exakter Binomialtest

ii) $X \stackrel{\text{appr.}}{\sim} F_N(x | n\Theta; \sqrt{n\Theta(1-\Theta)}) = F_N\left(x | 30 \cdot \frac{1}{2}; \sqrt{30 \cdot \frac{1}{2} \cdot (1 - \frac{1}{2})}\right)$

iii) Man könnte auch mit Poissonverteilung approximieren.

Damit kann man arbeiten!

Wenn P_{H_0} (Stichprobenresultat) $<$ Testniveau α , dann wird H_0 zugunsten H_1 verworfen, ansonsten nicht. Für den exakten Binomialtest erhält man:

$$i) P_{H_0}(T = 10) = \binom{30}{25} \cdot \frac{1}{2}^{25} \cdot \frac{1}{2}^5 + \binom{30}{5} \cdot \frac{1}{2}^5 \cdot \frac{1}{2}^{25} = 2 \cdot 0.000133 = 0.000266 \begin{cases} < \alpha \Rightarrow H_1 \\ \geq \alpha \Rightarrow H_0 \end{cases}$$

Beachtet hier bitte, dass man $t = 10$ sowohl für $X = 25$ als auch für $X = 5$ bekommt.

Bemerkungen 13.1: Allgemeine Betrachtungen über Hypothesentests

- Je kleiner das Testniveau α gewählt wird, desto unwahrscheinlicher ist es, dass H_0 abgelehnt wird. Manchmal kommt es vor, dass H_0 abgelehnt wird, obwohl H_0 wahr ist. Man nennt dies den *Fehler 1. Art* oder α -*Fehler*. Das Testniveau α kann auch dahingehend interpretiert werden, mit welcher Wahrscheinlichkeit man einen solchen Fehler tolerieren will. Andererseits kommt es vor, dass H_0 nicht verworfen wird, obwohl H_1 wahr ist. Diesen Fehler nennt man *Fehler 2. Art* oder β -*Fehler*.
- Im *R Commander* lassen sich viele Hypothesentests menübasiert durchführen. Als Ausgabe

erhält man abhängig vom Typ des Hypothesentests eine Reihe von Informationen. Bei Mittelwerttests wird z.B. automatisch ein $(1 - \alpha)$ -Konfidenzintervall bestimmt. Alle Tests erzeugen jedoch den ominösen p -Wert, der häufig falsch interpretiert wird. Dadurch sah sich z.B. die American Statistical Association im Jahr 2016 genötigt, eine Mitteilung über den Umgang mit p -Werten und statistischer Signifikanz zu veröffentlichen. Der p -Wert (oder die *Überschreitungswahrscheinlichkeit*) ist definiert als die Wahrscheinlichkeit dafür, unter der Voraussetzung, dass H_0 wahr ist, den aus der Stichprobe bestimmten Wert t der Teststatistik oder einen in Richtung der Alternative "extremere" Wert zu erhalten. Bei der Berechnung des p -Wertes ist also die Überlegung entscheidend, welche Teststatistikwerte noch stärker gegen H_0 sprechen. Es gilt:

- rechtsseitiger Test: $p_r = P_{H_0}(T \geq t)$
- linksseitiger Test: $p_l = P_{H_0}(T \leq t)$
- zweiseitiger Test: $p_z = 2 \cdot \min\{p_r, p_l\}$

Ich will ein einziges Mal den p -Wert per Hand und zwar für unser konkretes Beispiel berechnen.

Im Beispiel hatten wir $t = 10$. Was ist die WSK, dass $t = 10$ oder grösser ist (Obergrenze für t ist natürlich 15!), die Teststatistikwerte also noch stärker gegen H_0 sprechen?

Antwort:

$$\begin{aligned} p &= P(X \leq 5) + P(X \geq 25) = P(X \leq 5) + (1 - P(X \leq 24)) = F_b\left(5 \mid \frac{1}{2}; 30\right) - (1 - F_b(24 \mid \frac{1}{2}; 30)) \\ &= 0.000162457 + (1 - 0.999837543) = 0.000324914 \end{aligned}$$

Diesen Wert produziert auch R!

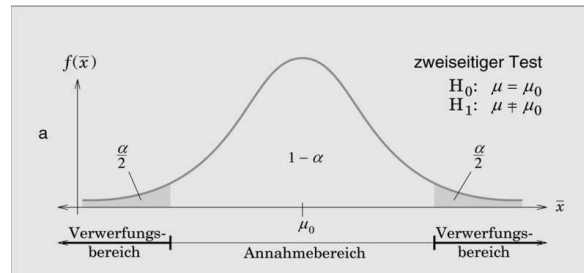
Der p -Wert hängt also nur vom Stichprobenresultat und der zugrundeliegenden Verteilung ab, nicht vom Testniveau!

Ist der p -Wert kleiner oder gleich dem vorgegebenen Testniveau α , so wird H_0 verworfen, ansonsten behält man H_0 bei. p ist also sozusagen das kleinste Testniveau, bei dem das Stichprobenresultat nicht zur Ablehnung von H_0 führen würde.

Wird X durch die Normalverteilung approximiert sieht alles folgendermassen aus:

ii) Standardisieren von X : $Z = \frac{X - n\Theta}{\sqrt{n\Theta(1-\Theta)}} = \frac{X - 15}{\sqrt{30 \cdot 0.5}}$. Z.B. $\alpha = 0.01$: $z[1 - \frac{0.01}{2}] = z[0.995] = 2.575$.

Für $x = 25$ erhält man aber $z = 3.66 \Rightarrow H_1$.



13.2 Einstichproben t -Test

Testsituation: Test von Annahmen und Behauptungen zu durchschnittlichen Merkmalsausprägungen.

Betrachten wieder n Stichprobenvariablen X_1, X_2, \dots, X_n mit $EX_i = \mu$ und $\text{var } X_i = \sigma^2$. Wir wissen:

- $\frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$ ist für grosse n approximativ standardnormalverteilt (folgt aus ZGWS).
- $\frac{\bar{X} - \mu}{\hat{\sigma}_{\bar{X}}}$ ist für beliebige n t -verteilt mit $(n - 1)$ Freiheitsgraden, falls das zugrunde liegende Merkmal (annähernd) normalverteilt ist und $\sigma_{\bar{X}}$ mittels einer Stichprobe geschätzt werden muss.

Wir kennen also approximativ die Verteilung von \bar{X} , wenn wir annehmen, dass der Mittelwert μ einen konkreten Wert μ_0 hat!

Realisierung x_1, x_2, \dots, x_n aus Stichprobe gegeben \Rightarrow Für Teststatistik kann numerischer Wert berechnet werden.

Anhand dieses Wertes wird entschieden, ob das konkrete Stichprobenresultat sehr unwahrscheinlich unter den Modellannahmen von H_0 ist oder eben nicht. Je nachdem, ob dieser Wert grösser oder kleiner einem kritischen Wert x_t ist (der vom Signifikanzniveau α und der (approximativen) Verteilung der Teststatistik abhängt), wird H_0 zugunsten H_1 verworfen oder beibehalten.

13.2.1 Zweiseitiges Testniveau

Testniveau: α

Nullhypothese: $H_0: \mu = \mu_0$, **Alternativhypothese:** $H_1: \mu \neq \mu_0$

Bemerkung 13.2: Zweiseitiges Testniveau

Der Test heisst zweiseitig, weil sowohl Abweichungen von μ_0 nach oben als auch nach unten gegen H_0 sprechen.

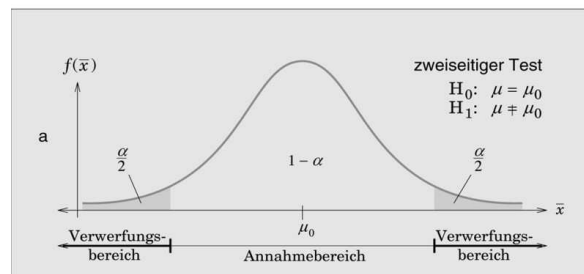
Testverteilung: Unter der Annahme, dass das “Merkmal” annähernd normalverteilt ist, ist $\frac{\bar{X} - \mu_0}{\hat{\sigma}_{\bar{X}}}$ annähernd t -verteilt, falls H_0 wahr ist und $\hat{\sigma}_{\bar{X}} = \frac{s^*}{\sqrt{n}}$ bzw. $\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n-1}}$ (jedenfalls, was gegeben ist).

Teststatistik: $T = \frac{\bar{X} - \mu_0}{\hat{\sigma}_{\bar{X}}}$, **realisierte Teststatistik (Prüfgrösse):** $t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{X}}}$

Testurteil: Kritischen x_t -Wert liest man aus Anhang C (zweiseitig) in der Zeile mit dem entsprechenden Freiheitsgrad unter dem entsprechenden α ab. Wenn

$$-x_t \leq t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{X}}} \leq x_t \quad (\text{Annahmebereich } \bar{V})$$

wird H_0 beibehalten, ansonsten zugunsten H_1 verworfen. Siehe nachstehende Grafik.



Beispiel 13.3: Zweiseitiger t -Test

Eine Maschine produziert Verschlüsse für Flaschen. Der Durchmesser der Verschlüsse soll 1.85cm sein. Aus früheren Studien weiss man, dass der Durchmesser annähernd normalverteilt ist. Die Maschine muss justiert werden, wenn der Durchschnittsdurchmesser einer Stichprobe bei einem Testniveau von $\alpha = 0.01$ signifikant von 1.85cm abweicht. Eine Stichprobe von 37 Verschlüssen ergab einen Durchschnittsdurchmesser von 1.87cm bei einer Standardabweichung von 0.05cm. Deutet dieses Resultat darauf hin, dass die Maschine justiert werden muss, d.h. dass der Durchmesser der produzierten Verschlüsse signifikant von 1.85cm verschieden ist?

Antwort: Hier ist $\alpha = 0.01$, $\bar{x} = 1.87$ und $s = 0.05$. Man hat demnach $\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n-1}} = \frac{0.05}{6} \approx 0.0083$.

Nullhypothese: $H_0 : \mu = 1.85$ ($= \mu_0$) Gegenhypothese: $H_1 : \mu \neq 1.85$

Kritischer Wert: $x_t = 2.7195 (= t_{n-1}[1 - \frac{\alpha}{2}] = t_{36}[0.995])$

$$\text{Teststatistik: } t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{1.87 - 1.85}{0.0083} = 2.41$$

Testentscheid: Da $-x_t = -2.7195 \leq t = 2.41 \leq 2.7195 = x_t$, kann die Nullhypothese nicht verworfen werden. Die Abweichungen der Durchmesser sind zufällig (nicht statistisch relevant).

13.2.2 Einseitiges Testniveau

Häufig ist man nur an Abweichungen des Mittelwerts von einem angenommenen Wert nach einer Seite interessiert.

Nehmt z.B. die folgende Situation. Das Lieblingsgetränk eurer Statistikgruppe ist Bier, das ihr in der Regel in Gläsern mit Normfüllung von 0.4l zu euch nehmt. Beim letzten Kneipenbesuch macht ihr euch den Spass und misst die Füllmenge eurer 100 (Frei erfundene Anzahl oder grosse Gruppe!) konsumierten Bierchen. Dabei kommt ihr auf eine mittlere Füllmenge von 0.38l. Als in Statistik bewanderte Personen fragt ihr euch, ob das rein zufällig ist oder der Wirt euch tatsächlich systematisch betrügt. Abweichungen nach oben sind für euch dabei irrelevant. Das führt auf t -Tests mit einseitigem Testniveau.

Einseitig rechtes Testniveau (oder einseitig oberseitig)

Hier führen nur grosse Abweichungen von μ_0 nach oben zur Ablehnung von H_0 .

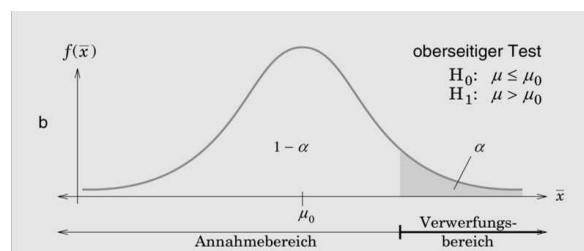
Dieser Test wird analog zum 2-seitigem Test durchgeführt. Die einzigen Unterschiede sind:

Nullhypothese: $H_0: \mu \leq \mu_0$, **Alternativhypothese:** $H_1: \mu > \mu_0$

Testurteil: Den kritischen x_t -Wert liest man aus Anhang C (einseitig) in der Zeile mit dem entsprechenden Freiheitsgrad unter dem entsprechenden α ab. Wenn

$$t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}} \leq x_t \quad (\text{Annahmebereich } \bar{V})$$

wird H_0 beibehalten, ansonsten zugunsten H_1 verworfen. Siehe nachstehende Grafik.



Beispiel 13.4: Einseitig oberseitiger t-Test

Eine Versicherung kalkuliert eine Spartenversicherung auf der Basis einer durchschnittlichen Schadenserwartung von 750CHF pro Schadensfall. Eine Stichprobe von 101 Schadensfällen ergibt eine durchschnittliche Schadenssumme von 778CHF bei einer Standardabweichung von 160CHF. Testet, ob die Versicherung bei einem Signifikanzniveau von $\alpha = 10\%$ weiterhin davon ausgehen kann, dass die durchschnittlichen Kosten eines Schadensfalls höchstens 750CHF betragen!

Antwort: Es gilt $\mu_0 = 750$ und $\alpha = 0.1$. Die (empirische) Standardabweichung $s = 160$ wurde über eine Stichprobe bestimmt. Wir schätzen also $\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n-1}} = \frac{160}{\sqrt{100}} = 16$.

Nullhypothese: $H_0 : \mu \leq 750$ ($= \mu_0$) Gegenhypothese: $H_1 : \mu > 750$

Kritischer Wert: $x_t = 1.2901$ ($= t_{n-1}[1 - \alpha] = t_{100}[0.9]$)

Teststatistik: $t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{X}}} = \frac{778 - 750}{16} = 1.75$

Testentscheid: Da $x_t = 1.2901 < t = 1.75$, muss die Nullhypothese verworfen werden. Die durchschnittlichen Kosten eines Schadenfalls sind signifikant höher als 750CHF.

Einseitig linkes Testniveau (oder einseitig unterseitig)

Hier führen nur grosse Abweichungen von μ_0 nach unten zur Ablehnung von H_0 .

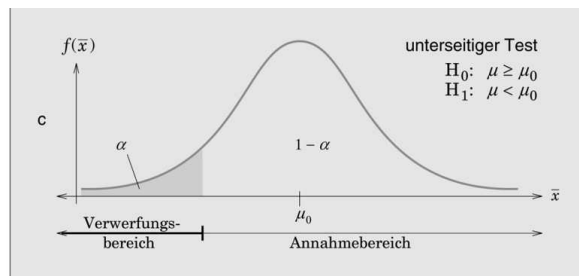
Auch dieser Test wird analog zum 2-seitigen Test durchgeführt. Die einzigen Unterschiede sind:

Nullhypothese: $H_0 : \mu \geq \mu_0$, **Alternativhypothese:** $H_1 : \mu < \mu_0$

Testurteil: Den kritischen x_t -Wert liest man aus Anhang C (einseitig) in der Zeile mit dem entsprechenden Freiheitsgrad unter dem entsprechenden α ab. Wenn

$$t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{X}}} \geq -x_t \quad (\text{Annahmebereich } \bar{V})$$

wird H_0 beibehalten, ansonsten zugunsten H_1 verworfen. Siehe nachstehende Grafik.



Beispiel 13.5: Einseitig unterseitiger t-Test

Es wird behauptet, dass das monatliche Einkommen von Lehrlingen einer Branche mindestens 725CHF betrage. Diese Behauptung soll bei einem Signifikanzniveau von 0.05 getestet werden. Bei einer Stichprobe unter $n = 40$ Lehrlingen wurde ein Durchschnittseinkommen von 700CHF bei einer Varianz von $s^2 = 2000\text{CHF}^2$ ermittelt.

Antwort: Hier ist $s = \sqrt{2000} \approx 44.72$ die empirische Standardabweichung, die über eine Stichprobe bestimmt wurde. Wir schätzen also $\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n-1}} \approx \frac{44.72}{\sqrt{39}} \approx 7.16$. $\mu_0 = 725$ und $\alpha = 0.05$.

Nullhypothese: $H_0 : \mu \geq 725$ ($= \mu_0$) Gegenhypothese: $H_1 : \mu < 725$

Kritischer Wert: $x_t = 1.6849 (= t_{n-1}[1 - \alpha] = t_{39}[0.95])$

Teststatistik: $t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{X}}} = \frac{700 - 725}{7.16} = -3.49$

Testentscheid: Da $-x_t = -1.6849 > t = -3.49$, muss die Nullhypothese verworfen werden. Das durchschnittliche monatliche Einkommen ist signifikant kleiner als 725CHF.

Bemerkung 13.6

Für $n \geq 30$ könnte man anstatt der t -Verteilung auch die Normalverteilung verwenden. Ich gehe darauf nicht näher ein, da Bärthl immer nur die t -Verteilung benutzt. Vielleicht denkt ihr aber mal kurz darüber nach.

13.3 t-Tests über Anteilswerte

Wir haben in der Vergangenheit häufig Parallelen wahrgenommen zwischen statistischen Theorien bzgl. Mittelwerten und Anteilswerten. Diese setzen sich bei Hypothesentests fort. Für t -Tests über Anteilswerte braucht man eigentlich nur den Stichprobenmittelwert \bar{x} durch den Stichprobenanteilswert p zu ersetzen (bzw. μ_0 durch den hypothetischen Anteilswert p_0) und $\sigma_{\bar{X}}$ durch

$\hat{\sigma}_{\bar{X}} = \sqrt{\frac{p_0(1-p_0)}{n}}$ schätzen. Ein Beispiel sollte ausreichen, um dies zu verdeutlichen.

Beispiel 13.7: t-Test für Anteilswerte

Ein Dozierender behauptet, 50% aller Studierenden würden morgens unausgeschlafen zum Präsenzunterricht an der FFHS erscheinen. Ihr glaubt aber, der Anteil sei noch höher als 50%. Eine Befragung von 35 Studierenden ergab einen Anteilswert von $p = 0.6$. Welche Hypothese wird bei einem Testniveau von 0.01 akzeptiert?

Antwort:

Nullhypothese: $H_0 : p \leq 0.5$ ($= p_0$) Gegenhypothese: $H_1 : p > 0.5$

(Es handelt sich also um einen einseitig oberseitigen Test.)

Kritischer Wert: $x_t = 2.4411$

$$\text{Teststatistik: } t = \frac{p - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.6 - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{35}}} = 0.0845$$

Testurteil: Da $t = 0.0845 < x_t = 2.4411$, kann die Nullhypothese also nicht verworfen werden.
Der Stichprobenanteilstwert von 0.6 ist rein zufällig.

13.4 χ^2 -Anpassungstest

Anpassungstests dienen der Beantwortung der Frage, ob gegebene Daten hinreichend an eine vorgegebene Verteilung angepasst sind. Der Wichtigste unter ihnen ist der χ^2 -Anpassungstest. Das Prinzip stellt sich folgendermassen dar.

- Die für ein Merkmal X angenommene hypothetische Verteilung wird als Nullhypothese H_0 formuliert. Die Alternativhypothese H_1 ist die logische Verneinung von H_0 .
- n Beobachtungen von X liegen in m verschiedenen Kategorien.
- Falls man $p_j = P_{H_0}(X \text{ liegt in } j)$ (die WSK, dass X unter der hypothetischen Verteilung in der j -ten Kategorie liegt) definiert, dann müssen unter der hypothetischen Verteilung im Mittel $n_{ej} := np_j$ Beobachtungen in der j -ten Kategorie liegen.
- Ist n_{oj} die tatsächliche Anzahl von Beobachtungen in der j -ten Kategorie, dann ist die Teststatistik

$$\chi_{emp}^2 = \sum_{j=1}^m \frac{(n_{oj} - n_{ej})^2}{n_{ej}}$$

annähernd χ^2 -verteilt mit $(m - 1 - \text{Anz. aus Stichprobe geschätzter Parameter})$ Freiheitsgraden, falls in jeder Kategorie mindestens 5 Beobachtungen liegen.

- H_0 wird verworfen, falls bei einem Testniveau von α die Teststatistik χ_{emp}^2 grösser als ein kritischer Wert x_k ist. Andernfalls wird H_0 beibehalten. Der kritische Wert kann aus Anhang B ähnlich wie beim t -Test ermittelt werden.

Beispiel 13.8: χ^2 -Anpassungstest auf Gleichverteilung

In einem Werk liegen Angaben über die Anzahl der Maschinenstillstände in den einzelnen Stunden einer 8-Stunden-Schicht vor (siehe Tabelle unten). Vermutet wird, dass die Wahrscheinlichkeit eines Maschinenstillstands für gewisse Stunden der Schicht besonders gross,

also nicht gleichverteilt ist.

Schichtstunden	1	2	3	4	5	6	7	8
Maschinenstillstände	15	10	11	18	19	14	12	11

Man prüfe bei einem Testniveau von $\alpha = 0.05$, ob diese Vermutung berechtigt ist.

Antwort:

Die $m = 8$ einzelnen Schichtstunden sind hier die Kategorien.

Nullhypothese: H_0 : Die Maschinenstillstände sind gleichverteilt.

Gegenhypothese: H_1 : Die Maschinenstillstände sind nicht gleichverteilt.

Die totale Anzahl von Maschinenstillständen beträgt $n = 110$. Unter der Annahme der Gleichverteilung (d.h. $p_j = p = \frac{1}{8}$ für jede Kategorie) würden pro Schichtstunde $\frac{110}{8} = 13.75$ Maschinenstillstände erwartet.

Kritischer Wert: $x_k = 14.067 (= \chi^2_{m-1}[1 - \alpha] = \chi^2_7[0.95])$

Teststatistik:

$$\chi^2_{emp} = \sum_{j=1}^8 \frac{(n_{oj} - n_{ej})^2}{n_{ej}} = \frac{(15 - 13.75)^2}{13.75} + \frac{(10 - 13.75)^2}{13.75} + \dots + \frac{(11 - 13.75)^2}{13.75} = 5.78$$

Testurteil: Da die Teststatistik $\chi^2_{emp} = 5.78 \leq x_k = 14.067$ gilt, kann H_0 mit diesem Test nicht verworfen werden. Die Schwankungen sind "zufällig".

Beispiel 13.10: χ^2 -Anpassungstest auf Normalverteilung

Es liegen von 81 börsennotierten Unternehmen die Umsätze vor (in Mio \$). Mit den bekannten Methoden aus der Beschreibenden Statistik wurde daraus ein Mittel von $\bar{x} = 6982$ bei einer Standardabweichung von $s = 14984$ berechnet. Für den Anpassungstest wird das Datenmaterial wie folgt klassiert:

Anmerkung: Klassengrenzen in Einheiten von Tsd., es entspricht also z.B. dem Intervall (10, 15] das reale Intervall (10000, 15000] (und das natürlich immer noch in Mio \$).

Klasse	$(-\infty, 5]$	$(5, 10]$	$(10, 15]$	$(15, 20]$	$(20, 25]$	$(25, 30]$	$(30, 35]$	$(35, \infty)$
n_{oj}	26	12	6	8	7	10	7	5

Bemerkung 13.10

Das Datenmaterial spricht natürlich nicht unbedingt für eine Normalverteilung (Warum?). Es geht im Weiteren um die Demonstration des allgemeinen Vorgehens.

Sei X die Zufallsvariable, die den Firmenumsatz beschreibt. Ist X normalverteilt?

Da keine Parameter vorgegeben sind, schätzen wir diese:

$$\hat{\mu} = 6982 \quad \text{und} \quad \hat{\sigma} = \sqrt{\frac{n}{n-1}} s = \sqrt{\frac{81}{80}} \cdot 14984 \approx 15077$$

Man prüfe bei einem Testniveau von $\alpha = 0.05$, ob die Vermutung der Normalverteilung mit diesen Parametern gerechtfertigt ist.

Antwort:

Nullhypothese: H_0 : Die Umsätze sind nach $F_N(x|6982; 15077)$ normalverteilt.

Gegenhypothese: H_1 : Die Umsätze sind nicht nach $F_N(x|6982; 15077)$ normalverteilt.

Ich berechne exemplarisch die zu erwartende Häufigkeit n_{e3} für die 3. Klasse $(10, 15]$. Was ist die WSK $P(10 < X \leq 15)$ unter der Voraussetzung, dass X die Verteilung aus H_0 hat?

Standardisieren liefert:

$$\begin{aligned} P(10 < X \leq 15) &= P\left(\frac{10000 - 6982}{15077} < Z \leq \frac{15000 - 6982}{15077}\right) = F_N(0.531) - F_N(0.2) \\ &= 0.7019 - 0.5793 = \mathbf{0.1226} \quad (= p_3) \\ \Rightarrow n_{e3} &= np_3 = 81 \cdot 0.1226 = \mathbf{9.93} \end{aligned}$$

Die anderen Wahrscheinlichkeiten und die daraus resultierenden zu erwartenden Häufigkeiten sind in der folgenden Tabelle zusammengefasst:

Klasse	$(-\infty, 5]$	$(5, 10]$	$(10, 15]$	$(15, 20]$	$(20, 25]$	$(25, 30]$	$(30, 35]$	$(35, \infty)$
WSK p_j	0.4483	0.1310	0.1226	0.1032	0.0779	0.0540	0.0316	0.0314
$n_{ej} = np_j$	36.31	10.61	9.93	8.40	6.31	4.37	2.56	2.54

Mit den beobachteten Häufigkeiten n_{oj} und den erwarteten Besetzungszahlen n_{ej} berechnet man jetzt wie oben die Teststatistik $\chi_{emp}^2 = 22.1$.

Zur Bestimmung des kritischen Wertes beachtet bitte, dass zwei Parameter (μ und σ) aus der Stichprobe geschätzt wurden. Wir benötigen also die χ^2 -Verteilung mit $8-1-2=5$ Freiheitsgraden.

Kritischer Wert: $x_k = 11.07 (= \chi_{m-1-2}^2[1-\alpha] = \chi_5^2[0.95])$

Da die Teststatistik $\chi_{emp}^2 = 22.1$ grösser als der kritische Wert $x_k = 11.07$ ist, muss die Nullhypothese zugunsten der Alternativhypothese verworfen werden.

13.5 χ^2 -Unabhängigkeitstest

Es werden zwei Merkmale X und Y betrachtet und man interessiert sich dafür, ob die Merkmale statistisch unabhängig sind. Dabei geht man davon aus, dass die Beobachtungen von X in q verschiedene Kategorien fallen und jene von Y in r Kategorien.

Beispiel 13.11: χ^2 -Unabhängigkeitstest

In einer Statistikprüfung konnten für Aufgabe A 16 Punkte und für Aufgabe B 22 Punkte erreicht werden. Als gut wurde ein Punktbereich von 14-16 Punkten für Aufgabe A bzw. 17-22 Punkten für Aufgabe B definiert. Das Abschneiden der Studierenden ist in der folgenden Tabelle wiedergegeben:

Aufgabe B	Aufgabe A		Gesamt
	0-13 (schlecht)	14-16 (gut)	
17-22 (gut)	$n_{o11} = 6$	$n_{o12} = 13$	$n_{Z1} = 19$
0-16 (schlecht)	$n_{o21} = 17$	$n_{o22} = 8$	$n_{Z2} = 25$
Gesamt	$n_{S1} = 23$	$n_{S2} = 21$	$n = 44$

(Ich habe in der Tabelle zur Deutlichkeit die allgemeinen Bezeichnungen dieser Werte gemäss Bärthl zugefügt.)

Die Merkmale X und Y sind hier also die erreichten Punktkategorien (gut-schlecht) in Aufgabe A bzw. B. Beachtet bitte, dass hier im Gegensatz zur Kontingenztafel in der Kopfspalte bzw. Kopfzeile nicht einzelne Merkmalsausprägungen angegeben sind, sondern gewisse Kategorien, in welche diese fallen!

Unabhängigkeit zweier Merkmale X und Y in der Kontingenztafel der absoluten Häufigkeiten ist dadurch gekennzeichnet, dass gilt:

$$n_{oij} = \frac{n_{Zi} \cdot n_{Sj}}{n} =: n_{eij}$$

für **alle** $i = 1, \dots, q$ und $j = 1, \dots, r$.

Wir bezeichnen also die **bei Unabhängigkeit zu erwartenden** absoluten Häufigkeiten mit n_{eij} (Diese müssen nicht ganzzahlig sein!).

Für unser Beispiel ergibt sich damit folgende Tabelle der bei Unabhängigkeit zu erwartenden Häufigkeiten. (Ich habe zur besseren Verdeutlichung die Rechnungen angegeben. Alle Werte sind auf eine Dezimalstelle gerundet.):

Aufgabe B	Aufgabe A		Gesamt
	0-13	14-16	
17-22	$n_{e11} = 9.9 (= \frac{19 \cdot 23}{44})$	$n_{e12} = 9.1 (= \frac{19 \cdot 21}{44})$	19
0-16	$n_{e21} = 13.1 (= \frac{25 \cdot 23}{44})$	$n_{e22} = 11.9 (= \frac{25 \cdot 21}{44})$	25
Gesamt	23	21	44

Ihr erkennt hoffentlich die Systematik. Sind die Merkmale in mehr Kategorien unterteilt, verhält sich alles vollkommen analog.

Der χ^2 -Unabhängigkeitstest verläuft nun wie folgt:

Nullhypothese: H_0 : Die Merkmale X und Y sind unabhängig.

Gegenhypothese: H_1 : Die Merkmale X und Y sind nicht unabhängig.

Teststatistik: Mit den beobachteten Häufigkeiten n_{oij} und hypothetischen Häufigkeiten n_{eij} wird nun analog zum Anpassungstest die Teststatistik

$$\chi_{emp}^2 = \sum_{i=1}^q \sum_{j=1}^r \frac{(n_{oij} - n_{eij})^2}{n_{eij}}$$

gebildet. Diese ist für hinreichend grosse n_{eij} (Die n_{eij} sollten wieder wenigstens 5 sein.) annähernd χ^2 -verteilt mit $(q-1)(r-1)$ Freiheitsgraden.

Kritischer Wert: x_k wird genauso bestimmt wie im Anpassungstest.

Testurteil: Wie beim Anpassungstest wird H_0 zugunsten der Gegenhypothese abgelehnt, falls $\chi_{emp}^2 > x_k$ gilt. Andernfalls wird H_0 beibehalten.

Für unser Beispiel ergibt das:

$$\chi_{emp}^2 = \frac{(6-9.9)^2}{9.9} + \frac{(13-9.1)^2}{9.1} + \frac{(17-13.1)^2}{13.1} + \frac{(8-11.9)^2}{11.9} = 5.65.$$

In unserem Fall gilt $q = r = 2$. Wir benutzen daher die χ^2 -Verteilung mit $(q-1)(r-1) = 1$ Freiheitsgraden. Bei einem Testniveau von z.B. $\alpha = 0.05$ ergibt sich $x_k = 3.841$ und H_0 würde abgelehnt. Es besteht keine Unabhängigkeit der Merkmale. Für ein Testniveau von $\alpha = 0.01$ hingegen wäre $x_k = 6.635$. Bei diesem Testniveau müsste also die Behauptung der Unabhängigkeit beibehalten werden.

Index

- p -Wert, 113
- Überschreitungswahrscheinlichkeit, 113

- Alternativhypothese, 110
- Anteilswert, 97

- Beobachtungsreihe, 9
- Bernoulli-Gesetz der grossen Zahlen, 86
- Bernoulli-Kette, 71
- Bessel-Korrektur, 98
- Bestimmtheitsmass, 32
- bias, 97
- Binomialkoeffizient, 34
- Box-Plot, 24

- Dezil, 24
- Dichtefunktion, 61

- Einheit, statistische, 5
- Elementarereignisse, 41
- Ereignis, entgegengesetztes, 42
- Ereignis, sicheres, 42
- Ereignis, unmögliches, 42
- Ereignis, zufälliges, 42
- Ereignismenge, 42
- Ereignisraum, 41
- Ereignisse, disjunkte, 44
- Ereignisse, komplementäre, 42
- Ereignisse, konträre, 44
- Ereignisse, unvereinbare, 44
- Ergebnisraum, 41
- Erhebungseinheit, 5

- Fakultät, 34
- Fehler 1. Art, 112
- Fehler 2. Art, 112

- Gauss-Klammer, 23
- Gegenhypothese, 110
- Grössenklasse, 15

- Grundgesamtheit, 6

- Häufigkeit, relative, 11
- Häufigkeitsfunktion, absolute, 12
- Häufigkeitsfunktion, relative, 13
- Häufigkeitsverteilung, absolute, 10
- Häufigkeitsverteilung, relative, 11
- Häufigkeitsdichte, durchschnittliche, 17
- Histogramm, 17

- Identifikationskriterium, 5
- Interquartilsabstand, 27
- Irrtumswahrscheinlichkeit, 100

- Klassenbreite, 15
- Kollektiv, 6
- Kombination, 39
- Komplement, 42
- Konfidenzniveau, 100
- Kontingenztafel, 29
- Korrekturfaktor, 99
- Korrelationskoeffizient, Pearson'scher, 30
- Kovarianz, empirische, 29

- Laplace-Experiment, 44

- Masse, statistische, 6
- Massenfunktion, 60
- Median, 22
- Mengendifferenz, 42
- Merkmal, 6
- Merkmal, dichotomes, 111
- Merkmal, diskretes, 6
- Merkmal, intervallskaliertes, 7
- Merkmal, nominal skaliertes, 7
- Merkmal, ordinal skaliertes, 7
- Merkmal, qualitatives, 6
- Merkmal, quantitatives, 6
- Merkmal, stetiges, 6
- Merkmal, verhältnisskaliertes, 8

- Merkmaleswert, 6
Merkmalsausprägung, 6
Messniveau, 7
Mittel, arithmetisches, 18
Mittelwert, 18
Mittelwerts, Standardfehler des, 101
Modalität, 6
Modalwert, 22
Modus, 22

Nullhypothese, 110

Partition, 54
Pascalsches Dreieck, 35
Permutation, 37
Permutation mit Wiederholung, 38
Perzentil, 24
Population, 6

Quantil, 23
Quartil, 24

Randverteilung, 109

Schätzung, erwartungstreue, 96
Schätzung, konsistente, 96
Schicht, 15
Schluss, direkter, 91
Schwaches Gesetz der grossen Zahlen, 85
Signifikanzniveau, 110
Skala, kardinale, 8
Skala, metrische, 8
Skalenniveau, 7
Spannweite, 26
Standardabweichung, 27
Standardnormalverteilung, 80
Stetigkeitskorrektur, 94
Stichprobe, 8
Stichprobe, reine, 8
Stichprobe, repräsentative, 9
Streudiagramm, 27

Teilerhebung, 8

Teilgesamtheit, 8
Testniveau, 110
Totalerhebung, 8
Tschebychev-Ungleichung, 86

Unabhängigkeit, stochastische, 51
Unit, 5
Untersuchungseinheit, 5
Urliste, 9

Varianz, empirische, 26
Variation, 39
Variationskoeffizient, 27
Venn-Diagramm, 42
Verteilung, bimodale, 22
Verteilung, multimodale, 22
Verteilung, trimodale, 22
Verteilung, unimodale, 22
Verteilungsfunktion, empirische, 13
Verzerrung, 97
Vollerhebung, 8

Wahrscheinlichkeit, a posteriori, 55
Wahrscheinlichkeit, bedingte, 50, 51
Wahrscheinlichkeitsdichtefunktion, 61
Wahrscheinlichkeitsfunktion, 60
Wahrscheinlichkeitsmassenfunktion, 60

Zentralwert, 22
Zufallsvariable, Median, 66
Zufallsexperiment, 41
Zufallsvariable, 57
Zufallsvariable, diskrete, 58
Zufallsvariable, Erwartungswert einer, 64
Zufallsvariable, Modus einer, 68
Zufallsvariable, Standardabweichung einer, 66
Zufallsvariable, stetige, 58
Zufallsvariable, Varianz einer, 65
Zufallsvariable, Verteilungsfunktion einer, 58
Zufallsvariable, Zentralwert einer, 66
Zufallsvariablen, Quantil einer, 66