

Prof. Dr. Th. Royen

**Einführung in die
Wahrscheinlichkeitsrechnung
und Statistik**

Inhaltsverzeichnis

1	Diskrete Wahrscheinlichkeitsräume	4
1.1	Vorgänge mit zufälligen Ergebnissen	4
1.1.1	Beispiele	4
1.1.2	Ereignisse	5
1.1.3	Wahrscheinlichkeit	5
1.2	Elemente der Kombinatorik	8
1.2.1	Variationen	8
1.2.2	Kombinationen	8
1.2.3	Permutationen	9
1.3	Kombination unabhängiger Vorgänge, Produkträume	10
1.3.1	Das Gesetz der großen Zahlen	10
1.3.2	Stochastische Konvergenz	11
1.3.3	Unabhängige Zufallsvariable	11
1.4	Verteilung von Zufallsvariablen	12
1.4.1	Wahrscheinlichkeits- und Verteilungsfunktion	12
1.4.2	Mittelwert und Erwartungswert	13
1.4.3	Varianz und Standardabweichung	15
1.4.4	Streuung	17
1.5	Einige diskrete Verteilungen und erzeugende Funktionen	17
1.5.1	Die Binominalverteilung ($B(n; p)$ -Verteilung)	17
1.5.2	Die Poisson-Verteilung ($P(\mu)$ -Verteilung)	18
1.5.3	Geometrische Verteilung und negative Binominalverteilung	21
1.5.4	Die hypergeometrische Verteilung $H(n; M, N)$ -Verteilung	22
1.6	Bivariate und multivariate Verteilungen, Kovarianz und Korrelation	23
1.6.1	Kovarianz und Korrelation	24
1.6.2	Multivariate Verteilungen	26
1.6.3	Die Multinomialverteilung	27
1.6.4	Die multivariate hypergeometrische Verteilung	28
1.7	Bedingte Wahrscheinlichkeit, Bayes-Formel	28
1.7.1	Bedingte Wahrscheinlichkeit	28
1.7.2	Die Bayes-Formel	32
2	Stetige Verteilungen	33
2.1	Wahrscheinlichkeitsdichte, Parameter von Verteilungen	33
2.1.1	Wahrscheinlichkeitsdichte	33
2.1.2	Einige Parameter (Kenngrößen) von Verteilungen	33
2.1.3	Tabelle einiger stetige Verteilungen	35
2.1.4	Beispiele stetiger Verteilungen	35

2.1.5	Charakteristische Funktionen	37
2.1.6	Einige Eigenschaften charakteristischer Funktionen	37
2.1.7	Laplacetransformierte	38
2.1.8	Beispiele für char. Funktionen und Laplacetransformierte	39
2.2	Der Zentrale Grenzwertsatz	39
2.3	Bivariate und multivariate stetige Verteilungen	44
2.3.1	Marginalverteilungen („Randverteilungen“)	45
2.3.2	Bedingte Verteilungen und bedingte Dichten	46
2.3.3	Die multivariate Normalverteilung $N(\vec{\mu}; \Sigma)$ -Verteilung	47
2.3.4	Die wichtigsten Eigenschaften der $N_q(\vec{\mu}; \Sigma)$ -Verteilung	47
3	Statistische Methoden	52
3.1	Konfidenzintervalle für unbekannte Parameterwerte	52
3.1.1	$(1 - \alpha)$ -Konfidenzintervall	52
3.1.2	Einige Beispiele	52
3.2	Testen von Hypothesen	59
3.2.1	Nullhypothese	59
3.2.2	Alternativhypothese	59
3.2.3	Einstichproben t -Test	60
3.2.4	2-Stichproben t -Tests	61
3.2.5	Test für den Wert einer unbekanntem Wahrscheinlichkeit p	63
3.2.6	Vierfelder-Tests zum Vergleich zweier unbekannter Wahrscheinlichkeiten	63
3.3	Regression, Maximum-Likelihood-Schätzung von Parametern, Methode der kleinsten Quadrate	64
3.3.1	Regression	64
3.3.2	Das Maximum-Likelihood-Schätzprinzip	65
3.3.3	Die Methode der kleinsten Quadrate	66
3.3.4	Die Regressionsgerade	67
3.3.5	Tests und Konfidenzbereiche für die Parameter von Regressiongeraden	68
3.4	Chi-Quadrat-Tests zur Analyse von Häufigkeiten	71
3.4.1	Goodness of fit - Test für unbekannte Wahrscheinlichkeitswerte	71
3.4.2	Allgemeiner Goodness of fit - Test für Wahrscheinlichkeiten mit unbekanntem Parameterwerten	72
3.4.3	Der χ^2 -Test auf Unabhängigkeit zweier diskreter Merkmale	73
3.5	Nichtparametrische Verfahren - Rangtests	75
3.5.1	Rangtests	75
3.5.2	Wilcoxon-Zwei-Stichproben-Rangtest zum Vergleich zweier unabhängiger Stichproben	75
3.5.3	Grenzwertsätze von Kolmogoroff und Smirnow	77
3.5.4	Anwendungsbeispiele	78
3.5.5	Test auf Normalverteilung nach Lilliefors	79
3.5.6	Zur Simulation von $N(0; 1)$ -Zufallsgrößen	79
A	Anhang	81
A.1	Tabelle der Standardnormalverteilung	81
A.2	Tabelle der χ^2 - Verteilung	83
A.3	Tabelle der t -Verteilung	84

1 Diskrete Wahrscheinlichkeitsräume

1.1 Vorgänge mit zufälligen Ergebnissen

1.1.1 Beispiele

Vorgang	Ergebnis	beobachtete Merkmale (= Zufallsvariable)
1. Wurf eines Würfels	geworfene Augenzahl ω	Augenzahl $X(\omega) = \omega$
2. zweimaliger Wurf eines Würfels	Paar $\omega = (\omega_1, \omega_2)$ geworfener Augenzahlen	z.B. $S(\omega) :=$ Summe der Augenzahlen oder $H(\omega) :=$ Häufigkeit der Sechs bei zwei Würfeln
3. n -maliger Wurf eines Würfels	Folge ^I $\omega = (\omega_1, \dots, \omega_n)$ der geworfenen Augenzahlen	$S(\omega) := \omega_1 + \dots + \omega_n$ $H(\omega) :=$ Häufigkeit der Sechs in der Folge ω
4. zufälliges Ziehen einer Karte aus einer Kartei mit lauter verschiedenen Karten	gezogene Karte ω	auf der Karte ω eingetragene Merkmale, z.B. $X(\omega) =$ Körperlänge $Y(\omega) =$ Geschlecht $Z(\omega) =$ Augenfarbe, usw. in einer Personenkartei
5. n -maliges zufälliges Ziehen einer Karte aus einer Kartei		
(a) mit Zurücklegen, d.h. n -malige Ausführung von Vorgang Nr. 4	Folge $\omega = (\omega_1, \dots, \omega_n)$ der gezogenen Karten (gleiche Karte mehrfach möglich)	zugehörige zufällige Werte $X_1 = X(\omega_1), X_2 = X(\omega_2), \dots$ $Y_1 = Y(\omega_1), Y_2 = Y(\omega_2), \dots$ usw.
(b) ohne Zurücklegen	Folge $\omega = (\omega_1, \dots, \omega_n)$ (alle Karten verschieden) oder <u>Teilmenge</u> $\omega = \{\omega_1, \dots, \omega_n\}$	zugehörige zufällige Werte $X_1 = X(\omega_1), X_2 = X(\omega_2), \dots$ $Y_1 = Y(\omega_1), Y_2 = Y(\omega_2), \dots$ wie vorher

^IFolge: n -Tupel

Allgemeiner:

Vorgang	Ergebnis	zugehörige Merkmalswerte
zufälliges Ziehen einer <u>Stichprobe</u> mit n Elementen $\omega_1, \dots, \omega_n$ aus einer Grundgesamtheit	gezogene Stichprobe	X_1, \dots, X_n ^{II} Y_1, \dots, Y_n ... wie vorher
(a) mit Zurücklegen	Folge $\omega = (\omega_1, \dots, \omega_n)$ (gleiche Elemente mehrfach möglich)	
(b) ohne Zurücklegen	Folge ω oder Teilmenge $\omega = \{\omega_1, \dots, \omega_n\}$ wenn Reihenfolge unerheblich ist ^{III}	

Vorgänge mit zufälligen Ergebnissen können sich über eine Zeitdauer erstrecken. Es sei z.B. $X(t)$ die zufällige Anzahl der Elemente in einer Warteschlange zum Zeitpunkt t , in der zu jedem Zeitpunkt zufällige Zu- und Abgänge erfolgen können. Allgemeiner kann $Z(t)$ (reell- oder vektorwertig) den zufallsabhängigen Zustand eines Systems zum Zeitpunkt t bezeichnen. Vorgänge die durch zufallsabhängige Funktionen $Z(t)$ beschrieben werden, heißen stochastische Prozesse. Die Wahrscheinlichkeitsrechnung befaßt sich mit der mathematischen Beschreibung (Modellierung) von Vorgängen mit zufälligen Ergebnissen.

1.1.2 Ereignisse

Beispiel zu 1. (einmaliger Wurf eines Würfels):

Das Ereignis $A :=$ „Es wird eine gerade Augenzahl ω geworfen“ tritt genau dann ein, wenn $\omega = 2$ oder $\omega = 6$. Wir fassen das Ereignis A einfach als die Teilmenge $\{2, 4, 6\}$ derjenigen Ereignisse ω auf, bei denen A eintritt. Allgemein: Ist V ein Vorgang mit den möglichen Ergebnissen $\omega \in \Omega$, so ist ein Ereignis A irgendeine Teilmenge von Ω . Die Sprechweise „ A tritt ein“ ist dann gleichbedeutend mit „Das Ergebnis ω gehört zur Teilmenge A “.

Extremfälle:

Das „unmögliche Ereignis“ ist die leere Menge \emptyset , tritt aber nie ein, da kein $\omega \in \emptyset$.

Das „sichere Ereignis“ ist die ganze Menge Ω aller möglichen Ergebnisse, tritt also immer ein, da stets $\omega \in \Omega$.

Der Durchschnitt zweier Ereignisse $A, B \subseteq \Omega$ ist das Ereignis $A \cap B = \{\omega \in \Omega | \omega \in A \text{ und } \omega \in B\}$, d.h. A und B treten gleichzeitig zusammen ein.

Die Vereinigung zweier Ereignisse $A, B \subseteq \Omega$ ist das Ereignis $A \cup B = \{\omega \in \Omega | \omega \in A \text{ oder } \omega \in B\}$, d.h. A oder B tritt ein.

Komplementärereignis zu A :

$\bar{A} = \{\omega \in \Omega | \omega \notin A\}$, d.h. „ A tritt nicht ein“.

Beispiel zu 1. (einmaliger Wurf eines Würfels):

$A := \{2, 4, 6\}$, d.h. gerade Augenzahl und $B := \{1, 2, 3\}$, d.h. Augenzahl ≤ 3 .

$A \cap B = \{2\}$, d.h. Augenzahl ist gerade und ≤ 3 .

$A \cup B = \{1, 2, 3, 4, 6\}$, d.h. Augenzahl ist gerade oder ≤ 3 .

$\bar{A} = \{1, 3, 5\}$

1.1.3 Wahrscheinlichkeit

Ein Vorgang V mit zufälligen Ergebnissen wird vollständig beschrieben durch die Menge Ω der möglichen Ergebnisse und die Eintrittswahrscheinlichkeiten p_i für die einzelnen Ergebnisse ω_i . Die Wahrscheinlichkeiten sind

^{II} (X_1, \dots, X_n) wird auch als Zufallsstichprobe (mit oder ohne Zurücklegen) vom Umfang n des Merkmals X bezeichnet.

^{III}z.B. interessiert bei einer Befragung von Einwohnern einer Stadt (= Grundgesamtheit) nur die Menge der ausgewählten befragten Personen, nicht deren Reihenfolge.

Werte zwischen 0 und 1^{IV} , und die Summe aller Einzelwahrscheinlichkeiten hat stets den Wert 1. Es gilt also immer

$$\boxed{\sum_{\omega_i \in \Omega} p_i = 1}$$

Die Wahrscheinlichkeit für ein beliebiges Ereignis $A \subseteq \Omega$ ist die Summe der Wahrscheinlichkeiten für die einzelnen Ergebnisse $\omega \in A$, kurz:

$$\boxed{P(A) = \sum_{\omega_i \in \Omega} p_i} \quad (P : \text{Probability})$$

Ω zusammen mit einem Wahrscheinlichkeitsmaß P für die Teilmengen (= Ereignisse) von Ω heißt dann ein diskreter^V Wahrscheinlichkeitsraum (Ω, P) .

Für obiges Ereignis $A := \{2, 4, 6\} \subseteq \Omega = \{1, 2, 3, 4, 5, 6\}$ gilt allgemein $P(A) = p_2 + p_4 + p_6$. Insbesondere haben bei einem Wurf mit einem idealen Würfel alle Ergebnisse (= Augenzahlen) ω_i die gleiche Wahrscheinlichkeit $p_i = \frac{1}{6}$. Dann ist die Wahrscheinlichkeit für eine gerade Augenzahl gegeben durch

$$P(A) = p_2 + p_4 + p_6 = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Allgemeiner:

Laplacevorgänge sind Vorgänge mit nur gleichwahrscheinlichen Ergebnissen. Gleichwahrscheinliche Ergebnisse wollen wir auch Fälle nennen. Bezeichnet $|\Omega|$ die Anzahl der Elemente (Fälle) einer endlichen Ergebnismenge Ω und $|A|$ die Anzahl der Fälle aus denen das Ereignis A zusammengesetzt ist, so gilt für jedes Ereignis eines Laplacevorgangs:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der Fälle von } A}{\text{Anzahl aller möglichen Fälle}}$$

Diese sogenannte Laplacewahrscheinlichkeit ist somit auf das Abzählen von Fällen zurückgeführt, d.h. auf Kombinatorik.

Im allgemeinen Fall sind die Wahrscheinlichkeiten p_i der Ergebnisse ω_i oft unbekannt und müssen aus Beobachtungen geschätzt werden. Dies ist eine Grundaufgabe der Statistik.

Für die Vereinigung zweier Ereignisse A und B gilt:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Beispiel: zweimaliger Wurf eines idealen Würfels

$X_i :=$ Augenzahl beim i -ten Wurf, $i = 1, 2$. Jeder mögliche Fall entspricht in der Veranschaulichung einem Kästchen ω mit der Wahrscheinlichkeit $p = \frac{1}{36}$.

	1	2	3	4	5	6	
1	/	/	/	/	/	/	
2	/	/	/	/	/	/	
3	/	/	/	/	/	/	
4					/	/	$A := \{X_1 \leq 3\}$
5					/	/	
6					/	/	

$B := \{X_2 \geq 5\}$

^{IV}Gelegentlich gibt man Wahrscheinlichkeiten in Prozent an, also $p \cdot 100\%$

^Vdiskret bedeutet, dass man die Ergebnisse zählen oder nummerieren kann. Ω ist höchstens abzählbar unendlich

$$\begin{aligned}
 P(A) &= \frac{18}{36} = \frac{1}{2} \\
 P(B) &= \frac{12}{36} = \frac{1}{3} \\
 P(A \cap B) &= \frac{6}{36} = \frac{1}{6} \\
 P(A \cup B) &= \frac{18 + 12 - 6}{36} = \frac{24}{36} = \frac{2}{3}
 \end{aligned}$$

Die Fälle in $A \cap B$ sind in A und in B enthalten und müssen daher wieder abgezogen werden, damit sie nicht doppelt gezählt werden!

Zwei Ereignisse $A, B \subseteq \Omega$ heißen disjunkt oder unvereinbar, wenn $A \cap B = \emptyset$, d.h. wenn A und B nicht gleichzeitig eintreten können. Offenbar gilt für disjunkte Ereignisse $P(A \cup B) = P(A) + P(B)$. Allgemein gilt für mehrere paarweise disjunkte Ereignisse A_1, \dots, A_n die

Additivität der Wahrscheinlichkeit

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

kürzer:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (\text{falls } A_i \cap A_j = \emptyset \text{ für } i \neq j).$$

Wahrscheinlichkeiten von Ereignissen A_i aus Ω verhalten sich daher wie Massen von Teilstücken A_i eines Körpers Ω mit Gesamtmasse $P(\Omega) = 1$.

Zwei Ereignisse $A, B \subseteq \Omega$ heißen (stochastisch) unabhängig, wenn

$$P(A \cap B) = P(A) \cdot P(B)$$

Für obiges Beispiel:

$$P(A \cap B) = P(\{X_1 \leq 3\} \cap \{X_2 \geq 5\}) = \frac{6}{36} = \frac{3}{6} \cdot \frac{2}{6} = P(A) \cdot P(B)$$

Deutung: A ist ein Ereignis beim 1. Wurf und B ein Ereignis beim 2. Wurf. Ein Ereignis beim 1. Wurf beeinflusst nicht ein Ereignis beim 2. Wurf und umgekehrt. Solche Ereignisse, die sich nicht gegenseitig beeinflussen können, also in einem intuitivem Sinn „unabhängig“ sind, die sind auch unabhängig gemäß obiger Definition. Entsprechend heißen mehrere Ereignisse $A_1, \dots, A_n \subseteq \Omega$ vollständig unabhängig, wenn für jede Auswahl $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ aus diesen n Ereignissen gilt:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k})$$

Es gilt also die Multiplikativität der Wahrscheinlichkeiten vollständig unabhängiger Ereignisse.

Eine etwas schwächere Eigenschaft ist die paarweise Unabhängigkeit der Ereignisse A_1, \dots, A_n . Hier wird nur $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ für alle Paare $1 \leq i < j \leq n$ gefordert. Aus vollständiger Unabhängigkeit folgt paarweise Unabhängigkeit. Die Umkehr gilt nicht!

Beispiel:

Beim Wurf zweier Würfel sei $A := \{X_1 \leq 3\}$, $B := \{X_2 \geq 3\}$ und $C := (A \cap B) \cup (\bar{A} \cap \bar{B})$. Es gilt: $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4} \Rightarrow$ paarweise Unabhängig von A, B und C . Aber $P(A \cap B \cap C) = P(A \cap B) = \frac{1}{4} \neq P(A) \cdot P(B) \cdot P(C) = \frac{1}{8}$, also keine vollständige Unabhängigkeit von A, B, C .

Da Ereignisse als Teilmengen von Ω aufgefaßt werden, gelten für das Rechnen mit Ereignissen alle Rechenregeln der BOOL'schen Algebra für die Mengenoperationen z.B. die Distributivgesetze:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Nach dem Dualitätsprinzip der Mengenlehre gibt der vollständige Tausch der Zeichen \cap und \cup in einer richtigen Gleichung wieder eine richtige Gleichung.

$$\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \overline{A_i} \quad \text{d.h. „keines der } A_i \text{ tritt ein“}$$

$$\overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \overline{A_i} \quad \text{d.h. „nicht alle } A_i \text{ treten ein“}$$

Bei einer unendlichen Folge A_1, A_2, \dots von Ereignissen interessieren oft die Ereignisse

$$\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \quad \text{d.h. „unendlich viele der } A_i \text{ treten ein“}$$

$$\bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i \quad \text{d.h. „schließlich alle } A_i \text{ treten ein“ (d.h. alle mit } i \geq n \text{ ab irgendeinem } n)$$

1.2 Elemente der Kombinatorik

Es gibt sechs häufig auftretende Grundaufgaben für das Abzählen von „Fällen“. Hierzu liege eine Menge M mit n verschiedenen Elementen vor (z.B. ein Kartenspiel).

1.2.1 Variationen

(a) Variationen ohne Wiederholung von k aus n Elementen

Es wird k -mal zufällig ein Element ohne Zurücklegen aus M gezogen. Das Ergebnis ist ein k -Tupel gezogener Elemente. (k gezogene Spielkarten liegen nebeneinander.)

Solche durch zufälliges Ziehen entstehenden k -Tupel heißen auch Variationen von k aus n Elementen. Die Anzahl der verschiedenen k -Tupel (Variationen), die hierbei entstehen können, ist gleich

$$\boxed{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}$$

(b) Variationen mit Wiederholung von k aus n Elementen

Es wird k -mal zufällig ein Element mit Zurücklegen aus M gezogen. Das Ergebnis ist ein k -Tupel von Einzelziehungsergebnissen. (Die k Ziehungsergebnisse je einer Karte können nebeneinander notiert werden.)

Im Unterschied zu 1.2.1 (a) kann das gleiche Element (die gleiche Spielkarte) dabei wiederholt vorkommen! Die Anzahl der hierbei möglichen verschiedenen k -Tupel (Variationen mit Wiederholung) ist gleich

$$\boxed{n^k}$$

1.2.2 Kombinationen

(a) Kombinationen ohne Wiederholung von k aus n Elementen

Es werden wieder k Elemente ohne Zurücklegen aus M gezogen. Als Ziehungsergebnis wird nun die Teilmenge der k gezogenen Elementen aufgefasst. Die Reihenfolge der Einzelziehungsergebnisse wird ignoriert. Die Anzahl der möglichen verschiedenen Kombinationen ohne Wiederholung von k aus n Elementen ist also einfach die Anzahl

aller verschiedenen Teilmengen vom Umfang k aus einer Menge vom Umfang n (d.h. mit n Elementen). Diese Anzahl ist gegeben durch den Binominalkoeffizienten

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

Begründung: Die k Elemente einer Teilmenge vom Umfang k können in $k!$ verschiedene Reihenfolgen angeordnet werden. Einer Teilmenge (Kombination) entsprechen also $k!$ verschiedene Variationen (ohne Wiederholung) ihrer k Elemente, somit folgt das Ergebnis aus 1.2.1 (a).

(b) Kombinationen mit Wiederholung von k aus n Elementen

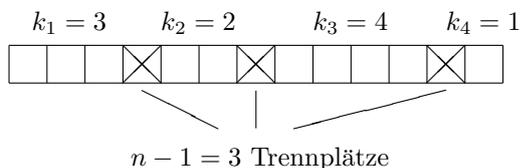
Es wird k -mal ein Element mit Zurücklegen aus M gezogen. Wieder interessiert nicht die Reihenfolge der Einzelziehungsergebnisse, sondern nur wie oft das j -te Element von M gezogen wurde. Die Ziehungsergebnisse sind also die Häufigkeiten k_1, \dots, k_n für die einzelnen Elemente von M . Offenbar gilt: $k = k_1 + \dots + k_n$. Die Anzahl der dabei verschiedenen möglichen Ziehungsergebnisse (Kombinationen mit Wiederholung) ist also gleich der Anzahl der unterscheidbaren additiven Zerlegungen $k = k_1 + \dots + k_n$ von k mit Summanden $k_j \in \mathbb{N}_0, j = 1, \dots, n$. Diese Anzahl ist gegeben durch

$$\binom{k+n-1}{n-1}$$

Begründung: Jeder Zerlegung einer Reihe mit k Plätzen in Abschnitte mit den Längen k_1, \dots, k_n entspricht (umkehrbar eindeutig) einer Auswahl von $n-1$ „Trennplätzen“ in einer Reihe mit $k+n-1$ Plätzen, und es gibt genau $\binom{k+n-1}{n-1}$ verschiedene Möglichkeiten für die Auswahl von $n-1$ Trennplätzen unter $k+n-1$ Plätzen.

Beispiel:

Darstellung einer Zerlegung von $k = 10$ in $n = 4$ Summanden $k_1 = 3, k_2 = 2, k_3 = 4, k_4 = 1$ mit Hilfe von $n-1 = 3$ Trennplätzen:



Unmittelbar benachbarte Trennplätze würden Summanden $k_j = 0$ entsprechen.

1.2.3 Permutationen

(a) Permutationen ohne Wiederholung

Die Anzahl der verschiedenen mögliche Reihenfolgen (Umordnungen) von n verschiedenen Elementen ist die Anzahl der Permutationen ohne Wiederholung von n Elementen. Diese Anzahl ist gegeben durch

$$n! = 1 \cdot 2 \cdot \dots \cdot n \quad (\text{ergibt sich aus 1.2.1 (a) mit } k = n)$$

$(0! := 1)$

Für große n hat man als Näherungsformel die Stirling-Formel:

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \cdot e^{\frac{1}{12n}} \quad (\text{rechte Seite ist genauer!})$$

z.B. mit $n = 20$:

$$2,42278 \dots 10^{18} < 20! = 2,432902008 \dots 10^{18} < 2,432902852 \dots 10^{18}$$

mit $n = 100$: Um „Overflow“ im Taschenrechner zu umgehen wird logarithmiert:

$$\log(100!) \approx \log(\sqrt{200\pi}) + 100 \cdot \log\left(\frac{100}{e}\right) = 157,96964$$

$$100! \approx 10^{157,96964} = 10^{0,96964} \cdot 10^{157} = 9,32485 \cdot 10^{157}$$

Mit der genaueren Approximation: $100! \approx 9,33262 \cdot 10^{157}$

(b) Permutationen mit Wiederholung

Es wird k -mal mit Zurücklegen aus M mit n verschiedenen Elementen gezogen. Dabei sei k_j wieder die Häufigkeit mit der das j -te Element gezogen wird. (Man denke sich $n = 4$ Farben, z.B. $M := \{\text{rot, grün, blau, gelb}\}$ und k Ziehungsergebnisse als Farbfolge). Die Anzahl der verschiedenen möglichen Reihenfolgen (unterscheidbaren Farbfolgen) mit den gleichen Häufigkeiten k_1, \dots, k_n ist gegeben durch

$$\frac{k!}{k_1! \cdot k_2! \cdot \dots \cdot k_n!}$$

Begründung: Man denke sich die k Einzelziehungsergebnisse als eine Farbsequenz mit k Plätzen, auf denen die Farbe Nr. j genau k_j -mal auftritt, $j = 1, \dots, n$. Permutationen (Vertauschungen) von Plätzen mit gleicher Farbe untereinander ändern die Farbsequenz nicht. (z.B. bleibt der Tausch zweier roter Plätze unbemerkt).

Es gibt genau $k_1! \cdot k_2! \cdot \dots \cdot k_n!$ solcher Permutationen, welche nur Plätze gleicher Farbe untereinander vertauschen. Von den insgesamt $k!$ Permutationen aller k Plätze ergeben also jeweils $\prod_{j=1}^n k_j!$ Permutationen die gleiche Farbsequenz. Es gibt also insgesamt $\frac{k!}{k_1! \cdot \dots \cdot k_n!}$ unterscheidbare Farbsequenzen, d.h. verschiedene Reihenfolgen mit den gegebenen Häufigkeiten k_1, \dots, k_n .

Zahlenbeispiel: Bei $k = 100$ Ziehungen mit Zurücklegen aus einem Kartenspiel wurde 27-mal „Herz“, 21-mal „Karo“, 29-mal „Pik“ und 23-mal „Kreuz“ erhalten. Es gibt

$$\frac{k!}{k_1! \cdot k_2! \cdot k_3! \cdot k_4!} = \frac{100!}{27! \cdot 21! \cdot 29! \cdot 23!} \approx 7,3391 \cdot 10^{56}$$

unterscheidbare Reihenfolgen (= Permutationen mit Wiederholung) für diese Ziehungsergebnisse.

1.3 Kombination unabhängiger Vorgänge, Produkträume

Der Vorgang V bestehe aus aufeinander folgenden unabhängigen Ausführungen (d.h. ohne irgendwelche gegenseitige Beeinflussungen) der Vorgänge V_i mit zufälligen Ergebnissen aus den diskreten Ergebnismengen Ω_i und den zugehörigen Wahrscheinlichkeitsmaßen P_i auf Ω_i , ($i = 1, \dots, n$). Dann wird V vollständig beschrieben durch (Ω, P) mit dem kartesischen Produkt $\Omega = \Omega_1 \times \dots \times \Omega_n$, der Menge aller „Ergebnistupel“ $\omega = (\omega_1, \dots, \omega_n)$ und dem „Produktmaß“ $P = P_1 \times \dots \times P_n$. P ist bereits vollständig bestimmt durch die Werte $P(\{\omega\}) = P_1(\{\omega_1\}) \cdot \dots \cdot P_n(\{\omega_n\})$.

Für jedes Ergebnis $A \subseteq \Omega$ ist wieder $P(A) = \sum_{\omega \in A} P(\{\omega\})$, da Ω höchstens abzählbar ist.

Speziell für Ereignisse $A \subseteq \Omega$ der Form $A_1 \times \dots \times A_n$ mit $A_i \subseteq \Omega_i$ gilt:

$$P(A) = P(A_1 \times \dots \times A_n) = \prod_{i=1}^n P_i(A_i)$$

(daher der Name „Produktmaß“).

Die Anzahl der möglichen Ergebnisse $\omega \in \Omega$ beträgt offenbar $|\Omega| = \prod_{i=1}^n |\Omega_i|$, falls alle Ω_i endlich sind. Im Laplacefall ist $P(A_1 \times \dots \times A_n) = \prod_{i=1}^n \frac{|A_i|}{|\Omega_i|}$.

Speziell können die Vorgänge V_i - und damit die (Ω_i, P_i) - alle identisch sein. Dann wird V beschrieben durch das n -dimensionale kartesische Produkt $\Omega = \Omega^n = \Omega_1 \times \Omega_1 \times \dots \times \Omega_1$ und das entsprechende Produktmaß $P = P^{(n)} = P_1 \times P_1 \times \dots \times P_1$.

1.3.1 Das Gesetz der großen Zahlen

Es sei V_1 ein Vorgang mit Ergebnismenge Ω_1 und $A \subseteq \Omega_1$ ein Ereignis mit der Wahrscheinlichkeit $p = P(A)$.

$V^{(n)}$ bezeichne die n -fache Ausführung von V_1 mit beliebigem $n \in \mathbb{N}$ und $H_n(A)$ sei die absolute Häufigkeit des Auftretens von A bei $V^{(n)}$. Dann ist $h_n(A) := \frac{1}{n} H_n(A)$ die entsprechende relative Häufigkeit von A .

Man wird vermuten, dass für große n die Zufallsgröße $h_n(A)$ mit hoher Wahrscheinlichkeit dicht am Wert $P(A)$ liegt. Genauer gilt das schwache Gesetz der großen Zahlen:

$$\lim_{n \rightarrow \infty} P\left\{ \underbrace{h_n(A)}_{\text{zufällig!}} - \underbrace{P(A)}_{\text{feste Zahl } p!} > \varepsilon \right\} = 0 \quad \text{für alle } \varepsilon > 0 \quad (\text{Beweis später})$$

Man sagt hierfür auch: Die Folge der Zufallsgrößen $h_n(A)$ konvergiert stochastisch gegen $P(A)$, kurz $h_n(A) \xrightarrow{p} P(A)$.

Praktisch bedeutet das: Die Wahrscheinlichkeit $P(A)$ für jedes Ergebnis lässt sich (zumindest prinzipiell) durch die relativen Häufigkeiten $h_n(A)$ für dieses Ergebnis in hinreichend langen Wiederholungsfolgen $V^{(n)}$ mit immer größerer Wahrscheinlichkeit („Sicherheit“) immer genauer approximieren!

1.3.2 Stochastische Konvergenz

Allgemeine Definition der stochastischen Konvergenz von Zufallsgrößen

Ist X eine Zufallsgröße (die auch zu einer festen Zahl ausarten kann) und (X_n) eine unendliche Folge von Zufallsgrößen, so konvergiert die Folge X_n stochastisch gegen X , kurz: $X_n \xrightarrow{p} X$ für $n \rightarrow \infty$, genau dann, wenn

$$\lim_{n \rightarrow \infty} P\{|X_n - X| > \varepsilon\} = 0 \quad \text{für alle } \varepsilon > 0.$$

Zahlenbeispiel zur stochastischen Konvergenz:

Ein idealer Würfel wird n -mal geworfen, dabei sei h_n die relative Häufigkeit der Sechs. Typische Ergebnisse von fünf Serien mit je n Würfeln (durch Zufallszahlen mit Werten $1, \dots, 6$ im Computer simuliert):

Serien-Nr	1	2	3	4	5
$h_n (n = 100)$	0,17	0,21	0,12	0,19	0,14
$h_n (n = 10^6)$	0,166982	0,166421	0,166230	0,167056	0,167215

$$h_n \xrightarrow{p} \frac{1}{6} = 1, \bar{6} \quad \text{für } n \rightarrow \infty$$

1.3.3 Unabhängige Zufallsvariable

Zwei (auf dem gleichen Wahrscheinlichkeitsraum (Ω, P) definierte) diskrete Zufallsvariable X mit den möglichen Werten x_i und Y mit den möglichen Werten y_j heißen (stochastisch) unabhängig, wenn für alle x_i, y_j die entsprechende Ereignisse $A_i := \{X = x_i\}, B_j := \{Y = y_j\}$ stochastisch unabhängig sind, also stets

$$P\{X = x_i \text{ und } Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j\}$$

gilt. (Also $P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$.)

Entsprechend heißen mehrere Zufallsvariable X_i mit den möglichen Werten $x_{ik}, i = 1, \dots, n$, vollständig bzw. paarweise unabhängig, wenn für alle möglichen Wertekombinationen (n -Tupel) $x_{1,k_1}, \dots, x_{n,k_n}$ die entsprechenden Ereignisse $\{X_i = x_{i,k_i}, i = 1, \dots, n$, vollständig bzw. paarweise unabhängig sind. Für vollständig unabhängige Zufallsvariable X_i folgt dann auch, dass die für beliebige Teilmengen B_i der Wertebereiche der X_i definierten Ereignisse $A_i := \{X_i \in B_i\}$ vollständig unabhängig sind und daher gilt:

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

Wichtiger Sonderfall:

Vollständig unabhängig, identisch verteilte Zufallsvariable:

(englisch: i.i.d.r.v., d.h. independently, identically distributed random variables)

X sei eine Zufallsvariable bei einem Vorgang V . Der Vorgang $V^{(n)}$ bestehe aus n vollständig unabhängigen Ausführungen von V . Die Zufallsvariable X_i (auf dem Produktraum $\Omega^n = \Omega \times \dots \times \Omega$) sei definiert durch den zufälligen Wert von X bei der i -ten Ausführung von V , $\forall i = 1, \dots, n$. Dann sind X_1, \dots, X_n i.i.d.r.v.

Ist X eine Zufallsgröße (d.h. reellwertig), dann gilt mit der Verteilungsfunktion $F(x) := P\{X \leq x\}, x \in \mathbb{R}$, (vgl. 1.4) von X , dass auch alle X_i die gleiche Verteilungsfunktion besitzen und für beliebige Zahlen $x_1, \dots, x_n \in \mathbb{R}$:

$$P\{X_1 \leq x_1, \dots, X_n \leq x_n\} = \prod_{i=1}^n P\{X_i \leq x_i\} = \prod_{i=1}^n F(x_i)$$

Umgekehrt folgt aus dieser Gleichung, dass X_1, \dots, X_n vollständig unabhängig, identisch verteilte Zufallsgrößen sind.

\forall also $X_i(\omega) = X_i((\omega_1, \dots, \omega)) := X(\omega_i), \quad i = 1, \dots, n, \quad \text{alle } \omega_i \in \Omega$

1.4 Verteilung von Zufallsvariablen

1.4.1 Wahrscheinlichkeits- und Verteilungsfunktion

Es sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum und $X = X(\omega)$ eine Zufallsvariable auf Ω mit den möglichen Werten x_i (nicht notwendige Zahlen). Sehr oft handelt es sich um eine zufallsabhängige Häufigkeit X . Dann sind die Werte $x_i \in \mathbb{N}_0$.

Definition: Die auf der Wertemenge von X definierte Funktion

$$f : x_i \rightarrow p_i = f(x_i) := P\{X = x_i\}$$

(Ausführlicher: $P(\{\omega \in \Omega | X(\omega) = x_i\})$.)

heißt die Wahrscheinlichkeitsfunktion von X .

Man sagt: Die (Wahrscheinlichkeits)verteilung der Zufallsvariablen X ist bekannt, wenn die Wahrscheinlichkeiten $p_i = f(x_i)$ für alle möglichen Werte x_i von X bekannt sind.

Die Verteilung (und damit das Verhalten) einer Zufallsgröße (d.h. mit reellen Zahlen x_i als Werten) kann auch durch ihre (kumulative) Verteilungsfunktion F beschrieben werden:

Definition:

$$F(x) := P\{X \leq x\}, \quad x \in \mathbb{R}$$

(auch $F_X(x)$) heißt die kumulative Verteilungsfunktion der Zufallsgröße X .

Offenbar gilt: $F(x)$ ist monoton wachsend auf \mathbb{R} , und es ist $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$. F ist eine rechtstetige „Treppenfunktion“ mit Sprüngen $p_i = f(x_i)$ an den Sprungstellen x_i , den Werten von X .

Beispiel: Eine ideale Münze mit den Seiten $(Z, W) = (\text{Zahl, Wappen})$ wird sechsmal geworfen. Man denke sich die Ergebnisfolge $\omega = (\omega_1, \dots, \omega_6)$ notiert, z.B. $\omega = (Z, W, W, W, W, W)$ oder auch $\omega = (Z, Z, W, W, W, W)$. Es gibt $2^6 = 64$ verschiedene mögliche Ergebnisfolgen ω , die bei einer idealen Münze alle die gleiche Wahrscheinlichkeit $p = \frac{1}{2^6} = \frac{1}{64} = 0,015625$ haben. Die interessierende Zufallsgröße sei die Häufigkeit X für „Zahl“ bei sechs Würfeln. Die möglichen Werte von X sind $x_k = k$ mit $k = 0, \dots, 6$.

Das Ereignis $\{X = k\}$ tritt genau dann ein, wenn sich in der Ergebnisfolge ω an genau k Positionen Z befindet und an den übrigen $6 - k$ Positionen W . Da es $\binom{n}{k} = \binom{6}{k}$ Möglichkeiten gibt, k aus 6 Positionen auszuwählen, erhält man die Wahrscheinlichkeitsfunktion von X durch

$$p_k := P\{X = k\} = \binom{6}{k} \frac{1}{2^6} = \frac{1}{64} \binom{6}{k}, \quad k = 0, 1, \dots, 6.$$

(sogenannte Binominalverteilung mit $n = 6$ und $p = \frac{1}{2}$, kurz $B(n; p)$ -Verteilung, vgl. Abschnitt 1.5)

k	$\binom{6}{k}$	$p_k = P\{X = k\} = \frac{1}{64} \binom{6}{k}$
0	1	$\frac{1}{64} = 0,015625$
1	6	$\frac{6}{64} = 0,093750$
2	15	$\frac{15}{64} = 0,234375$
3	20	$\frac{20}{64} = 0,312500$
4	15	$\frac{15}{64} = 0,234375$
5	6	$\frac{6}{64} = 0,093750$
6	1	$\frac{1}{64} = 0,015625$

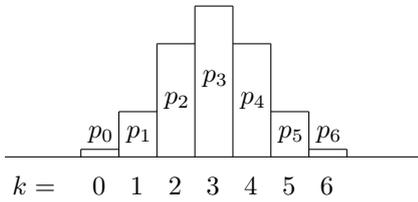
Die mathematische Beschreibung durch die $B(6; \frac{1}{2})$ -Verteilung bleibt gleich für jedes dichotome^{VII} Merkmal Y , d.h. mit zwei möglichen Werten $y_0 = N(\text{ein})$, $y_1 = J(\text{a})$ und $P\{Y = „N“\} = P\{Y = „J“\} = \frac{1}{2}$. Graphische

^{VII}gabelartig, zweiteilig; [grch. dica „zweifach“]

Veranschaulichung der Wahrscheinlichkeitsfunktion einer $B(n; p)$ -Verteilung mit $n = 6$ und $p = \frac{1}{2}$,

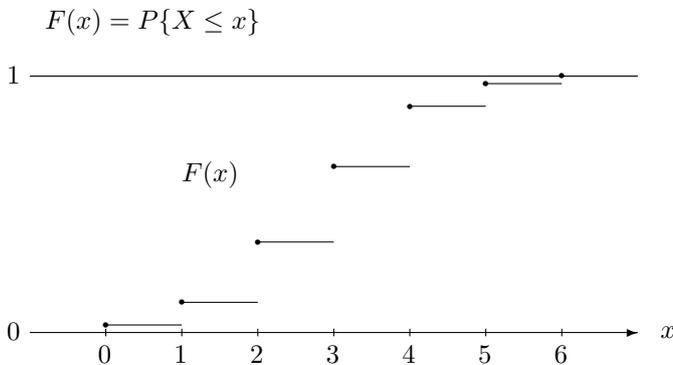
also

$$p_k = \frac{1}{2^6} \binom{6}{k}, \quad k = 0, 1, \dots, 6 :$$



Die Säulenflächen entsprechen den Wahrscheinlichkeiten p_k . Die Gesamtfläche aller Säulen entspricht daher 1!

Graphische Veranschaulichung der entsprechenden Verteilungsfunktion mit Sprunghöhen p_k :



1.4.2 Mittelwert und Erwartungswert

Ein idealer Würfel wurde 100-mal geworfen mit folgenden Ergebnissen:

Augenzahl	1	2	3	4	5	6
beob. Häufigkeit	14	12	20	15	21	18

Mittelwert der geworfenen Augenzahlen:

$$\bar{x} = \frac{14 \cdot 1 + 12 \cdot 2 + 20 \cdot 3 + 15 \cdot 4 + 21 \cdot 5 + 18 \cdot 6}{100} = 3,71$$

Für eine viel größere Anzahl n von Würfeln würde man

$$\bar{x} \approx \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \dots + \frac{1}{6} \cdot 6 = \frac{1}{6} \cdot (1 + 2 + \dots + 6) = 3,5$$

erwarten, da die relativen Häufigkeiten $h_{n,j} = h_n(\{\text{Augenzahl} = j\})$ nach dem Gesetz der großen Zahlen stochastisch gegen $p_j = P(\{\text{Augenzahl} = j\}) = \frac{1}{6}, j = 1, \dots, 6$, konvergieren.

Allgemein: Es sei V ein Vorgang und X eine dabei anfallende Zufallsgröße mit den möglichen Werten x_j und $p_j = P\{X = x_j\}$. $V^{(n)}$ bestehe aus n unabhängigen Wiederholungen von V mit n Beobachtungen (Realisierungen) von X . Die relative Häufigkeit für das Auftreten des Wertes x_j sei $h_{n,j}$. Dann ist der zufallsabhängige Mittelwert der beobachteten X -Werte gegeben durch

$$\bar{X} = \sum_j x_j h_{n,j}$$

d.h. das mit den zufälligen relativen Häufigkeiten h_{nj} gewichtete Mittel aller möglichen Werte x_j von X . Für $n \rightarrow \infty$ konvergieren die h_{nj} stochastisch gegen die Wahrscheinlichkeiten $p_j = P\{X = x_j\}$ für alle j . Somit konvergieren für wachsende n die Mittelwerte $\bar{X} = \bar{X}^{(n)}$ stochastisch gegen den festen Zahlenwert $\mu := \sum_j x_j p_j$. Dies ist für endlich viele verschiedene Werte x_j klar. Bei abzählbar unendlich vielen Werten x_j soll die absolute Konvergenz der Reihe vorausgesetzt werden. Dann gilt für $n \rightarrow \infty$:

$$\bar{X}^{(n)} = \sum_j x_j h_{nj} \xrightarrow{p} \mu = \sum_j x_j p_j.$$

Definition: Ist X eine diskrete Zufallsgröße mit den möglichen Werten x_j und der durch die Wahrscheinlichkeiten $p_j = P\{X = x_j\}$ gegebenen Verteilung, dann heißt die Zahl

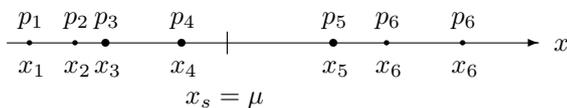
$$\mu = E(X) := \sum_j x_j p_j \quad \text{der Erwartungswert von } X.$$

(oder μ_X)

Bei unendlich vielen verschiedenen unbeschränkt großen Werten $|x_j|$ muss der Erwartungswert nicht existieren. Er existiert, wenn die Reihe $\sum_j x_j p_j$ absolut konvergent ist (d.h. $\sum_j |x_j| p_j < \infty$), was in diesem Kurs meist stillschweigend vorausgesetzt wird. Bei nur positiven Werten x_j und divergenter Reihe $\sum_j x_j p_j$ wird ergänzend $\mu := \infty$ gesetzt.

Der Erwartungswert ist ein wichtiger Parameter (= Kenngröße) der Verteilung einer Zufallsgröße. Er kann als eine Maßzahl für den „Mittelpunkt“ oder das „Zentrum“ einer Verteilung aufgefaßt werden. Er entspricht dem Schwerpunkt einer analogen Massenverteilung. Die folgende Skizze hat daher zwei Deutungen:

1. diskrete eindimensionale Massenverteilung mit „Massenprodukten“ der Masse p_j bei Position x_j auf der x -Achse und der Gesamtmasse $\sum_j p_j = 1$ mit der Schwerpunktskoordinate $x_s = \sum_j x_j p_j$.
2. diskrete eindimensionale Wahrscheinlichkeitsverteilung einer Zufallsgröße X mit den Wahrscheinlichkeiten $p_j = P\{X = x_j\}$ und dem Erwartungswert $\mu = \sum_j x_j p_j$.



Der aus n i.i.d.r.v. X_1, \dots, X_n gebildete Mittelwert $\bar{X}^{(n)} = \sum_{i=1}^n X_i$ ist ein zufallsabhängiger Schätzwert für den (oft unbekannt) Erwartungswert $\mu = E(X) = E(X_i)$ mit folgenden Eigenschaften:

1. $\bar{X}^{(n)}$ ist ein erwartungstreuer Schätzer^{VIII} für den Erwartungswert, d.h. $E(\bar{X}^{(n)}) = \mu$
2. Die Folge der Mittelwerte liefert konsistente Schätzer für μ , d.h. $\bar{X}^{(n)} \xrightarrow{p} \mu$ für $n \rightarrow \infty$ ^{IX}

Rechenregeln für Erwartungswerte

1. $E(cX) = cE(X)$ mit beliebiger Konstante c
2. $E(X + Y) = E(X) + E(Y)$ (Additivität) (d.h. der „Erwartungsoperator“ E ist ein linearer Operator) \Rightarrow
Für beliebige Linearkombinationen $\sum_j c_j X_j$ von Zufallsgrößen mit endlichen Erwartungswerten gilt:

$$E\left(\sum_j c_j X_j\right) = \sum_j c_j E(X_j)$$

3. Für unabhängige Zufallsgrößen X, Y mit endlichen Erwartungswerten gilt (Multiplikativität):

$$E(XY) = E(X) \cdot E(Y)$$

^{VIII} $\hat{=}$ unverzerrter Schätzer, englisch: unbiased estimate

^{IX} ausführlicher: $\lim_{n \rightarrow \infty} P\left\{|\bar{X}^{(n)} - \mu| > \varepsilon\right\} = 0$ für alle $\varepsilon > 0$ (stochastische Konvergenz)

Beweis zu 3.: Es sei $p_i = P\{X = x_i\}$, $q_j = P\{Y = y_j\}$. Wegen der Unabhängigkeit ist

$$E(XY) = \sum_i \sum_j x_i y_j P\{X = x_i \text{ und } Y = y_j\} = \sum_i \sum_j x_i y_j p_i q_j = \sum_i x_i p_i \cdot \sum_j y_j q_j = E(X) \cdot E(Y).$$

Beispiel: Die Zufallsgröße X habe nur die Werte $x_0 = 0$ und $x_1 = 1$ und es sei $p := P\{X = 1\}$:

$$E(X) = \sum_j x_j p_j = 0 \cdot (1 - p) + 1 \cdot p = p$$

Nun sei S $B(n; p)$ -verteilt mit $n = 6$ und $p = \frac{1}{2}$. S hat also die möglichen Werte $k = 0, \dots, 6$ mit $p_k = P\{S = k\} = \binom{6}{k} \frac{1}{2^6}$

$$E(S) = \sum_{k=0}^6 k \cdot \binom{6}{k} \frac{1}{2^6} = \frac{1}{64} \sum_{k=0}^6 k \cdot \binom{6}{k} = 3$$

Einfacher erhält man $E(S)$ durch Ausnutzen der Additivität des Erwartungswertes:

Es sei hierzu I_k die Indikatorvariable für das Auftreten von „Zahl“ beim k -ten Wurf (allgemeiner: für das Auftreten des interessierenden Ereignisses A mit Wahrscheinlichkeit $p = P(A)$ bei der k -ten Ausführung von Vorgang V). Es ist

$$I_k := \begin{cases} 1, & \text{wenn } A \text{ bei der } k\text{-Ausführung von } V \text{ eintritt} \\ 0, & \text{sonst} \end{cases}$$

und $E(I_k) = 1p + 0(1 - p) = p$.

Offenbar ist $S = \sum_{k=1}^6 I_k$, (z.B. $S = 1 + 0 + 0 + 1 + 0 + 0 = 2$)
und somit

$$E(S) = \sum_{k=1}^6 E(I_k) = np = 6 \cdot \frac{1}{2} = 3$$

Offenbar hätte man in dieser sehr einfachen Situation das Ergebnis auch ohne Rechnung „intuitiv“ richtig erraten.

1.4.3 Varianz und Standardabweichung

Als ein Maß für die Schwankung einer Zufallsgröße X um ihren Erwartungswert $\mu = E(X)$ kann man die zu erwartende quadratische Abweichung zwischen X und μ ansehen oder besser noch die Wurzel hieraus. Als weitere Parameter der Verteilung von X definieren wir:

$$\sigma^2 = \text{Var}(X) := E(X - \mu)^2 \text{ heißt die Varianz von } X \text{ und} \\ \sigma = \sqrt{\text{Var}(X)} \text{ die Standardabweichung von } X^X$$

Ferner heißt $\mu_n := E(X^n)$ das n -te Moment von X , $n \in \mathbb{N}$. (sofern μ_n noch einen endlichen Wert hat)

Es ist

$$\sigma^2 = E(X^2 - 2\mu X + \mu^2) = \underbrace{E(X^2)}_{\mu_2} - 2\mu \underbrace{E(X)}_{\mu=\mu_1} + \mu^2 = \mu_2 - \mu_1^2$$

Zur Berechnung einer Varianz ist dies oft einfacher.

Sind x_j die möglichen Werte von X mit $p_j = P\{X = x_j\}$, so ist

$$\mu_2 = \sum_j x_j^2 p_j \quad \text{und} \quad \sigma^2 = \sum_j (x_j - \mu)^2 p_j = \mu_2 - \mu^2$$

Rechenregeln für die Varianz:

^Xenglisch: standard deviation, abgekürzt SD oder sd

1. $\text{Var}(a + bX) = b^2 \text{Var}(X)$ für beliebige Zahlen a, b (speziell also $\text{Var}(-X) = \text{Var}(X)$)
2. Für unabhängige Zufallsgrößen X, Y gilt: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ (Additivität)

Allgemeiner gilt für nur paarweise unabhängige Zufallsgrößen X_i und Zahlen $a, b_i, i = 1, \dots, n$:

$$\text{Var} \left(a + \sum_{i=1}^n b_i X_i \right) = \sum_{i=1}^n b_i^2 \text{Var}(X_i)$$

Regel Nr.1 folgt unmittelbar aus der Definition von $\text{Var}(X)$. Man beachte, dass $E(a + x) = a + E(X)$, also $X - \mu = X - E(X)$ sich bei Addition einer Konstanten a zu X nicht ändert.

Beweis zu Regel Nr.2:

Mit

$$\begin{aligned} p_i &= P\{X = x_i\}, \\ q_j &= P\{Y = y_j\} \text{ und} \\ P\{X = x_i, Y = y_j\} &= p_i q_j \end{aligned}$$

und

$$\begin{aligned} \mu_x &:= E(X), \\ \mu_y &:= E(Y) \end{aligned}$$

ist

$$\begin{aligned} E(X - \mu_x)(Y - \mu_y) &= \sum_i \sum_j (x_i - \mu_x)(y_j - \mu_y) p_i q_j = \sum_i (x_i - \mu_x) p_i \cdot \sum_j (y_j - \mu_y) q_j = \\ &= (E(X) - \mu_x)(E(Y) - \mu_y) = 0, \end{aligned}$$

folglich

$$\begin{aligned} \text{Var}(X + Y) &= E(X - \mu_x + Y - \mu_y)^2 = E(X - \mu_x)^2 + E(Y - \mu_y)^2 + 2E(X - \mu_x)(Y - \mu_y) = \\ &= \text{Var}(X) + \text{Var}(Y) + 0, \end{aligned}$$

was zu zeigen war.

Speziell für i.i.d.r.v. X_1, \dots, X_n mit $\sigma^2 = \text{Var}(X_i)$ folgt:

$$\text{Var}(\bar{X}) = \text{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}$$

Die Standardabweichung des Mittelwerts \bar{X} ist daher:

$$\sigma_{\bar{X}} = \sqrt{\text{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

(englisch: standard error of mean, abgekürzt: SEM)

Beispiel für Varianzberechnung:

Für eine Indikatorzufallsgröße X mit $p = P\{X = 1\}$ und $1 - p = P\{X = 0\}$ gilt $\mu = p$ und $\mu_2 = E(X^2) = 1^2 p + 0^2 (1 - p) = p$, also $\sigma_x^2 = \mu_2 - \mu^2 = p - p^2 = p(1 - p)$. Die oben schon behandelte $B(6; \frac{1}{2})$ -verteilte Zufallsgröße $S = \sum_{k=1}^6 I_k$ hat dann die Varianz $\text{Var}(S) = \sum_{k=1}^6 \text{Var}(I_k) = 6p(1 - p) = 6 \cdot \frac{1}{2}(1 - \frac{1}{2}) = \frac{3}{2}$ und somit die Standardabweichung $\sqrt{\frac{3}{2}}$. Kennt man die Verteilung der Zufallsgröße X nicht, also auch nicht die Parameter $\mu = E(X)$ und $\sigma^2 = \text{Var}(X)$, so läßt sich aus wiederholten unabhängigen Beobachtungs- oder Meßwerten X_1, \dots, X_n (einer Zufallsstichprobe für X) μ durch den Mittelwert \bar{X} schätzen und σ^2 durch das Streuungsquadrat s^2 .

1.4.4 Streuung

Definition:

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

heißt das Streuungsquadrat (oder die Stichprobenvarianz) der Stichprobe X_1, \dots, X_n .

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

heißt die Streuung oder die Stichprobenstandardabweichung der Stichprobe.

s^2 ist ein konsistenter und erwartungstreuer Schätzer für σ^2 . Letzteres wird gerade durch den Faktor $\frac{1}{n-1}$ vor der Summe erzwungen. Schreibt man nämlich

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2) = \sum_{i=1}^n X_i^2 - 2\bar{X} \underbrace{\sum_{i=1}^n X_i}_{n\bar{X}} + n\bar{X}^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2,$$

so folgt mit $E(X_i) = E(\bar{X}) = \mu$:

$$\begin{aligned} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) &= E\left(\sum_{i=1}^n (X_i - \mu - (\bar{X} - \mu))^2\right) = E\left(\sum_{i=1}^n (X_i - \mu)^2\right) - n \underbrace{E(\bar{X} - \mu)^2}_{\text{Var}(\bar{X})} \\ &= n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2. \end{aligned}$$

Standardisierung einer Zufallsgröße X mit $\mu = E(X)$ und $\sigma = \sqrt{\text{Var}(X)}$:

$X^* := \frac{X-\mu}{\sigma}$ ist die zu X standardisierte Zufallsgröße. Es ist $E(X^*) = 0$ und $\text{Var}(X^*) = \frac{1}{\sigma^2} \text{Var}(X) = 1$. Haben die X -Werte eine Maßeinheit, so hat σ die gleiche Maßeinheit. Somit sind die X^* -Werte ohne Einheiten (X^* ist eine dimensionslose Größe). Zusammenhang zwischen den Verteilungsfunktionen F von X und F^* von X^* :

$$F(x) = P\{X \leq x\} = P\left\{X^* = \frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right\} = F^*\left(\frac{x-\mu}{\sigma}\right).$$

1.5 Einige diskrete Verteilungen und erzeugende Funktionen

1.5.1 Die Binominalverteilung ($B(n; p)$ -Verteilung)

Das Bernoulli-Schema:

V sei ein Vorgang mit zufälligen Ergebnissen und A ein Ereignis hierbei mit der Wahrscheinlichkeit $p = P(A)$. Der Vorgang $V^{(n)}$ bestehe in n unabhängigen Ausführungen von V und $S = S_n$ sei die absolute Häufigkeit des Eintretens von A hierbei. S hat die möglichen Werte $k = 0, 1, \dots, n$ und es gilt für alle k :

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

Definition:

Eine Zufallsgröße S mit den möglichen Werten $k = 0, \dots, n$ und der durch obige Formel gegebenen Verteilung heißt binominalverteilt mit den Parametern n und p , kurz: $B(n; p)$ -verteilt.

Statt „ S ist $B(n; p)$ -verteilt“ schreiben wir kürzer: $S \sim B(n; p)$.

Begründung für obige Formel: Zunächst sei $n = 5$ und $k = 2$.

Es sei A_{11000} das Ereignis „ A tritt ein bei den Ausführungen Nr. 1 und 2 und A tritt nicht ein bei den folgenden Ausführungen Nr. 3, 4 und 5.“. Wegen der vorausgesetzten Unabhängigkeit der Ausführungen von V gilt:

$P(A_{11000}) = p \cdot p(1-p)(1-p)(1-p) = p^2(1-p)^3$. Das Ereignis $\{S = 2\}$ ist zusammengesetzt aus den $\binom{n}{2} = \binom{5}{2}$ gleichwahrscheinlichen Ereignissen $A_{11000}, A_{10100}, \dots, A_{00011}$, je nach den $\binom{5}{2}$ Möglichkeiten, die 2 Positionen für 1 ($\hat{=}$ Eintritt von A) unter 5 Positionen auszuwählen. Somit ist $P\{S = 2\} = \binom{n}{k} p^k (1-p)^{n-k} = \binom{5}{2} p^2 (1-p)^3$. Der allgemeine Fall ist analog erhältlich.

Wenn in einer Anwendung eine zufällige Häufigkeit auftritt, so hat sie eine $B(n; p)$ -Verteilung, wenn sie durch ein Bernoulli-Schema obiger Art zustande gekommen ist!

Beispiel 1:

In einem System mit n Leitungen sei jede Leitung unabhängig von den übrigen mit der Wahrscheinlichkeit p frei und mit Wahrscheinlichkeit $1-p$ belegt. Die zufällige Anzahl X freier Leitungen ist dann $B(n; p)$ -verteilt. Mit $n = 10$ und $p = 0,1$ erhält man z.B.

$$P\{X = 0\} = \binom{10}{0} p^0 (1-p)^{10} = (1-p)^{10} = 0,9^{10} = 0,348678 \dots$$

d.h. mit einer 34,87%-igen Wahrscheinlichkeit sind alle Leitungen belegt.

$$\begin{aligned} P\{X \geq 2\} &= 1 - P\{X = 0\} - P\{X = 1\} = 1 - \sum_{k=0}^1 \binom{10}{k} p^k (1-p)^{10-k} \\ &= 1 - 0,9^{10} - 10 \cdot 0,1 \cdot 0,9^9 = 0,263901 \dots (= \text{Wahrsch. für mindestens zwei freie Leitungen}) \end{aligned}$$

Beispiel 2:

In einem Raum seien n Personen versammelt, davon haben X Personen gerade Geburtstag. X ist zufällig und mit sehr guter Näherung (Schaltjahre!) $B(n; \frac{1}{365})$ -verteilt. Berechne für $n = 100$:

$$P\{X = 0\}, P\{X = 1\}, P\{X \geq 2\}.$$

Die obige $B(n; p)$ -verteilte Größe S_n lässt sich als Summe von Indikatorvariablen darstellen. Dies vereinfacht die Berechnung der Verteilungsparameter $\mu = E(S_n)$ und $\sigma^2 = \text{Var}(S_n)$.

Die Indikatorvariable I_j zeigt das Eintreten von A bei der j -ten Ausführung von V an, d.h.

$$I_j := \begin{cases} 1, & \text{wenn } A \text{ bei der } j\text{-ten Ausführung von } A \text{ eintritt} \\ 0, & \text{sonst} \end{cases}$$

$$\mu = E(S_n) = E\left(\sum_{j=1}^n I_j\right) = \sum_{j=1}^n E(I_j) = \sum_{j=1}^n (1 \cdot p + 0(1-p)) = np$$

Mit $E(I_j^2) = E(I_j) = p$ folgt $\text{Var}(I_j) = E(I_j^2) - (E(I_j))^2 = p - p^2 = p(1-p)$, also

$$\sigma^2 = \text{Var}(S_n) = \sum_{j=1}^n \text{Var}(I_j) = np(1-p)$$

$$\sqrt{\text{Var}(S_n)} = \sqrt{np(1-p)}$$

1.5.2 Die Poisson-Verteilung ($P(\mu)$ -Verteilung)

Partikel oder „Punkte“ seien rein zufällig und homogen^{XI} über einen sehr großen Bereich des \mathbb{R}^n (anschaulich $n \leq 3$) verteilt. Es sei X die zufällige Anzahl Punkte in einem Teilbereich von einer Längeneinheit (Flächen- oder Volumeneinheit). Für gegen ∞ wachsende Bereichsgröße gilt immer genauer, dass X poissonverteilt ist mit dem Parameter $\mu = E(X)$, kurz: $X \sim P(\mu)$, d.h.

$$P\{X = k\} = e^{-\mu} \cdot \frac{\mu^k}{k!}, \quad k \in \mathbb{N}_0 \quad \text{XII}$$

^{XI}homogen bedeutet, dass die Punkte keine Teilregion systematisch „bevorzugen“, also überall mit gleicher Dichte bis auf rein zufällige Schwankungen verteilt sind

Der Bereich \mathbb{R} oder $[0, \infty)$ wird oft als Zeitachse betrachtet. Die „Punkte“ sind dann punktuelle Ereignisse oder „Signale“ in der Zeit. Bezeichnet X_t die Anzahl Signale in einem Zeitintervall der Länge t [Zeiteinheiten], so ist $E(X_t) = \mu t = \mu t$ und $X \sim P(\mu t)$, d.h. $P\{X_t = k\} = e^{-\mu t} \cdot \frac{(\mu t)^k}{k!}$.

Beispiele:

- X := Anzahl der Rosinen in einem Stück Rosinenbrot vom Gewicht 50 g, das aus einer großen Menge Teig mit Rosinen stammt.
- Y := Anzahl von Leukozythen (weiße Blutkörperchen) unter dem Gesichtsfeld eines Mikroskops (oder andere Partikel in einem Ausstrich).
- Z := Anzahl Signale pro Zeiteinheit in einem Geigerzähler nahe einer radioaktiv strahlenden Substanz.
- usw.

Eine $B(n; p)$ -Verteilung mit großem n und kleinem p und $\mu = np$ kann durch eine $P(\mu)$ -Verteilung approximiert werden.

Es gilt nämlich mit festem $\mu = np$ für $n \rightarrow \infty$ und $p = \frac{\mu}{n} \rightarrow 0$ für jedes $k \in \mathbb{N}_0$:

$$\begin{aligned} \binom{n}{k} p^k (1-p)^{n-k} &= \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-k} \\ &= \underbrace{\left(1 - \frac{\mu}{n}\right)^n}_{\rightarrow e^{-\mu}} \cdot \frac{\mu^k}{k!} \cdot \underbrace{\frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{n^k}}_{\rightarrow 1} \cdot \underbrace{\left(1 - \frac{\mu}{n}\right)^{-k}}_{\rightarrow 1} \rightarrow e^{-\mu} \cdot \frac{\mu^k}{k!} \end{aligned}$$

Zahlenbeispiel mit $n = 1000$, $p = 0,002$, $\mu = np = 2$:

k	$p_k = \binom{1000}{k} 0,002^k \cdot 0,998^{1000-k}$	$e^{-2} \frac{2^k}{k!}$
0	0,135065	0,135335
1	0,270670	0,270671
2	0,270942	0,270671
3	0,180628	0,180447
4	0,090223	0,090224
5	0,036017	0,036089
6	0,011970	0,012030
7	0,003406	0,003437
8	0,000847	0,000859
Σ	0,999768	0,999763
	(d.h. $P\{9 \leq X \leq 1000\} = 0,000232$)	

Wegen der Exponentialreihe $e^\mu = \sum_{k=0}^\infty \frac{\mu^k}{k!}$ gilt tatsächlich

$$\sum_{k=0}^\infty e^{-\mu} \frac{\mu^k}{k!} = e^{-\mu} \cdot e^\mu = 1$$

und auch

$$E(X) = e^{-\mu} \sum_{k=0}^\infty k \cdot \frac{\mu^k}{k!} = e^{-\mu} \sum_{k=1}^\infty k \cdot \frac{\mu^k}{k!} = \mu e^{-\mu} \sum_{k=1}^\infty \frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu} \sum_{k=0}^\infty \frac{\mu^k}{k!} = \mu e^{-\mu} \cdot e^\mu = \mu.$$

Ferner ist auch $\text{Var}(X) = \mu$. Dies zeigt man leichter durch erzeugende Funktionen:

Definition:

Ist X eine Zufallsgröße mit Werten in \mathbb{N}_0 und $p_k := P\{X = k\}$, so heißt die Funktion

$$\boxed{\Psi(z) := E(z^X) = \sum_{k=0}^\infty p_k z^k}$$

^{XII}Der Parameter μ der Poisson-Verteilung wird oft mit λ bezeichnet.

die erzeugende Funktion von X , bzw. die erzeugende Funktion der durch die Folge $(p_k)_{k \in \mathbb{N}_0}$ gegebenen Verteilung. (Falls $p_k = 0$ für $k > n$ ist ψ ein Polynom vom Grad $\leq n$.) Da die Taylorkoeffizienten der Potenzreihendarstellung durch $\psi(z)$ eindeutig bestimmt sind, ist die Abbildung $(p_k)_{k \in \mathbb{N}_0} \rightarrow \psi(z)$ umkehrbar, d.h. $\psi(z)$ bestimmt die Verteilung von X eindeutig. Wegen $\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} |p_k| = 1$ ist die Potenzreihe für $\psi(z)$ mindestens für alle $|z| \leq 1$ absolut konvergent!

Oft ist die Bestimmung von Parametern wie μ , $\mu_n = E(X^n)$, σ^2 einer Verteilung einfacher mit Hilfe der entsprechenden erzeugenden Funktion. Es ist

$$\begin{aligned} \psi'(z) &= \sum_{k=1}^{\infty} k \cdot p_k \cdot z^{k-1} & , \text{ also } \psi'(1) &= \sum_{k=1}^{\infty} k \cdot p_k = \mu \\ \psi''(z) &= \sum_{k=2}^{\infty} k \cdot (k-1) \cdot p_k \cdot z^{k-2} & , \text{ also } \psi''(1) &= E(X(X-1)) = E(X^2) - E(X) = \mu_2 - \mu \end{aligned}$$

folglich

$$\sigma^2 = \text{Var}(X) = \mu_2 - \mu^2 = \psi''(1) + \psi'(1) - (\psi'(1))^2 = \psi''(1) + \mu(1 - \mu)$$

Beispiel:

Die erzeugende Funktion einer $P(\mu)$ -Verteilung ist

$$\psi(z) = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} z^k = e^{-\mu} e^{\mu z} = e^{\mu(z-1)}$$

$$\psi'(z) = \mu \psi(z)$$

$$\psi''(z) = \mu^2 \psi(z)$$

also $\sigma^2 = \mu^2 + \mu - \mu^2 = \mu$

Die erzeugende Funktion einer $B(n; p)$ -Verteilung ist mit $q := 1 - p$:

$$\psi(z) = \sum_{k=0}^n \binom{n}{k} p^k \cdot q^{n-k} \cdot z^k = (q + pz)^n$$

Berechne durch Ableitung von ψ den Erwartungswert und die Varianz einer $B(n; p)$ -Verteilung.

Erzeugende Funktionen helfen auch bei der Berechnung der Verteilung einer Summe $S = X + Y$ unabhängiger Zufallsgrößen mit Werten in \mathbb{N}_0 . Mit den erzeugenden Funktionen ψ_X , ψ_Y und ψ_S von X , Y und S gilt:

$$\psi_S(z) = E(z^{X+Y}) = E(z^X \cdot z^Y) = E(z^X) \cdot E(z^Y) = \psi_X(z) \cdot \psi_Y(z)$$

da mit X, Y auch z^X, z^Y unabhängige Zufallsgrößen sind. Es gilt also für unabhängige Zufallsgrößen mit Werten in \mathbb{N}_0 :

$$\boxed{\psi_{X+Y}(z) = \psi_X(z) \cdot \psi_Y(z)}$$

Mit $p_i = P\{X = i\}$, $q_j = P\{Y = j\}$, $r_n = P\{S = X + Y = n\}$ ist

$$\psi_X(z) \cdot \psi_Y(z) = \sum_{i=0}^{\infty} p_i \cdot z^i = \sum_{j=0}^{\infty} q_j \cdot z^j = \sum_{n=0}^{\infty} \left(\sum_{i+j=n} p_i \cdot q_j \right) z^n = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n p_k \cdot q_{n-k} \right) z^n = \sum_{n=0}^{\infty} r_n \cdot z^n = \psi_S(z)$$

Es ist also $r_n = \sum_{k=0}^n p_k \cdot q_{n-k}$. Die Folge (Verteilung) $(r_n)_{n \in \mathbb{N}_0}$ heißt die Faltung der Folgen (Verteilungen) $(p_i)_{i \in \mathbb{N}_0}$ und $(q_j)_{j \in \mathbb{N}_0}$. Der Faltung der Verteilungen entspricht also die Multiplikation der entsprechenden erzeugenden Funktionen!

Allgemein gilt für die erzeugenden Funktionen ψ_{X_i} mehrerer vollständig unabhängiger Zufallsgrößen mit Werten in \mathbb{N}_0 :

$$\boxed{\psi_{\sum_i X_i}(z) = \prod_i \psi_{X_i}(z)}$$

Sind z.B. die $X_i \sim P(\mu_i)$ und vollständig unabhängig, so ist

$$\psi_{\sum_i X_i}(z) = \prod_i \psi_{X_i}(z) = \prod_i e^{(z-1)\mu_i} = e^{(z-1)\sum_i \mu_i}$$

also ist $\sum_i X_i \sim P(\sum_i \mu_i)$.

Oft werden Verteilungen nach dem Typ ihrer erzeugenden Funktionen benannt, z.B.:

1.5.3 Geometrische Verteilung und negative Binominalverteilung

A sei ein als „Erfolg“ bezeichnetes Ereignis mit $p = P(A)$ und $q = P(\bar{A}) = 1 - p$ bei einem Vorgang V . In einer Folge unabhängiger Wiederholungen sei T die Anzahl Misserfolge (\bar{A}) bis zum ersten Erfolg (A). Bei gleichen Zeitabständen zwischen den Wiederholungen kann T auch als die Wartezeit bis zum ersten Erfolg angesehen werden. Offenbar gilt:

$$P\{T = k\} = q^k p, \quad k \in \mathbb{N}_0$$

Die erzeugende Funktion $\psi_T(z)$ dieser sogenannten geometrischen Verteilung ist durch eine geometrische Reihe gegeben:

$$\psi_T(z) = p \sum_{k=0}^{\infty} q^k z^k = \frac{p}{1 - qz}$$

Somit ist $\psi'_T(z) = \frac{pq}{(1-qz)^2}$ und $\mu = E(T) = \psi'_T(1) = \frac{pq}{(1-q)^2} = \frac{q}{p} = \frac{1}{p} - 1$. Mit $\psi''(1) = 2\mu^2$ ergibt sich $\sigma^2 = \text{Var}(T) = \frac{q}{p^2}$.

Variante der geometrischen Verteilung:

Es wird die Anzahl Y der Versuche bis zum ersten Erfolg einschließlich gezählt, also $Y \geq 1$!

$$P\{Y = k\} = q^{k-1} p, \quad k \in \mathbb{N}$$

$$\psi_Y(z) = p \sum_{k=1}^{\infty} q^{k-1} z^k = \frac{pz}{1 - qz}$$

$$\Rightarrow \mu = E(Y) = E(T + 1) = \frac{1}{p}$$

$$\sigma^2 = \text{Var}(Y) = \text{Var}(T) = \frac{q}{p^2}$$

Die negative Binominalverteilung mit den Parametern $\alpha > 0$, $0 < p < 1$, ($NB(\alpha; p)$)-Verteilung ist definiert durch die erzeugende Funktion

$$\psi(z) = p^\alpha (1 - qz)^{-\alpha} = p^\alpha \sum_{k=0}^{\infty} \binom{-\alpha}{k} (-q)^k z^k \stackrel{\text{XIII}}{=} p^\alpha \sum_{k=0}^{\infty} \binom{\alpha + k - 1}{k} q^k z^k$$

Beachte, dass $\psi(1) = p^\alpha (1 - q)^{-\alpha} = p^\alpha p^{-\alpha} = 1$, wie es für jede Wahrscheinlichkeitsfunktion auf \mathbb{N}_0 gelten muss. Es ist

$$\psi'(z) = \alpha q \frac{p^\alpha}{(1 - qz)^{\alpha+1}}$$

$$\mu = \psi'(1) = \alpha \frac{q}{p}$$

$$\psi''(z) = \alpha(\alpha + 1)q^2 \frac{p^\alpha}{(1 - qz)^{\alpha+2}}$$

$$\psi''(1) = \alpha(\alpha + 1) \frac{q^2}{p^2} \rightarrow \sigma^2 = \alpha \frac{q}{p^2}$$

^{XIII}Dies ist die binomische Reihe mit den verallgemeinerten Binominalkoeffizienten

$$\binom{-\alpha}{k} := \frac{-\alpha(-\alpha-1)\dots(-\alpha-k+1)}{k!} = (-1)^k \frac{(\alpha+k-1)\dots(\alpha+1)\alpha}{k!} = (-1)^k \binom{\alpha+k-1}{k}$$

Für $X \sim NB(\alpha; p)$ gilt also:

$$P\{X = k\} = \binom{\alpha + k - 1}{k} p^\alpha q^k.$$

Ist z.B. $S_r = \sum_{i=1}^r T_i$ eine Summe unabhängiger Wartezeiten mit der obigen geometrischen Verteilung, so gilt $\psi_{S_r}(z) = \left(\frac{p}{1-qz}\right)^r$ und $S_r \sim NB(r; p)$. Die Anzahl Y_r der Wiederholungen von V bis zum r -ten Erfolg einschließlich ist $S_r + r$, also

$$\psi_{Y_r}(z) = \left(\frac{pz}{1-qz}\right)^r \quad \text{und} \quad P\{Y_r = k + r\} = \binom{r + k - 1}{k} q^k p^r.$$

1.5.4 Die hypergeometrische Verteilung $H(n; M, N)$ -Verteilung

Aus einer Grundgesamtheit mit N Elementen, die aus M „Merkmalsträgern“ besteht und $N - M$ „Nichtmerkmalsträgern“ werden zufällig n Elemente ohne Zurücklegen gezogen. Das Ergebnis ist eine Teilmenge $\omega = \{\omega_1, \dots, \omega_n\}$ gezogener Elemente (= Zufallsstichprobe vom Umfang n ohne Zurücklegen).

$X :=$ zufällige Anzahl der „Merkmalsträger“ in der Stichprobe ω .

Gesucht: Verteilung von X , also $P\{X = k\}$ für alle möglichen Werte k .

Es gibt $\binom{N}{n}$ mögliche Ergebnisse (Teilmengen ω), die alle mit gleicher Wahrscheinlichkeit $P\{\omega\} = \frac{1}{\binom{N}{n}}$ erhalten

werden können. Bei zufälliger Auswahl von n Elementen gibt es genau $\binom{M}{k} \cdot \binom{N-M}{n-k}$ verschiedene Möglichkeiten genau k Elemente aus den M Merkmalsträgern zu erhalten und $n - k$ Elemente aus den $N - M$ Nichtmerkmalsträgern. Es ergibt sich also durch Abzählen der Ergebnisse (Fälle) mit genau k Merkmalsträgern die Laplacewahrscheinlichkeit:

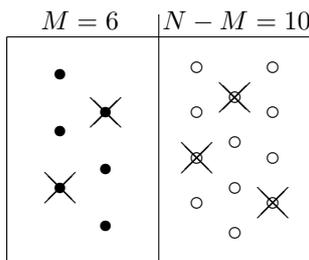
$$P\{X = k\} = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad \max(0; n - (N - M)) \leq k \leq \min(n; M)$$

Falls $n \leq \min(M; N - M)$: $0 \leq k \leq n$.

(Beachte: $n - k \leq N - M \Rightarrow k \geq n - (N - M)$)

Zahlenbeispiel:

$M = 6$ Merkmalsträger, $N - M = 10$ Nichtmerkmalsträger und $n = 5$ gezogene Elemente



$$P\{X = 2\} = \frac{\binom{6}{2} \binom{10}{3}}{\binom{16}{5}} = \frac{15 \cdot 120}{4368} = 0,4120879 \dots$$

Es ist $\mu = E(X) = n \frac{M}{N}$, $\sigma^2 = \text{Var}(X) = n \frac{M}{N} \frac{N-M}{N} \frac{N-n}{N-1}$ oder mit der Abkürzung $p = \frac{M}{N} =$ Anteil der Merkmalsträger in der Grundgesamtheit = Wahrscheinlichkeit, bei der ersten Ziehung einen Merkmalsträger zu ziehen:

$$\mu = np$$

$$\sigma^2 = np(1-p) \frac{N-n}{N-1}.$$

Falls $\frac{n}{M}$ und $\frac{n}{N-M}$ beide klein sind - d.h. praktisch etwa $\max\left(\frac{n}{M}, \frac{n}{N-M}\right) \leq 0,1$ -, kann die Veränderung der

Grundgesamtheit durch das Ziehen ohne Zurücklegen vernachlässigt werden, und man erhält dann die Approximation der hypergeometrischen Verteilung durch die Binominalverteilung, d.h.

$$P\{X = k\} = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \approx \binom{n}{k} p^k (1-p)^{n-k}, \text{ XIV } \quad k = 0, \dots, n$$

$$p = \frac{M}{N} \quad n \leq 0,1M \text{ und } n \leq 0,1(N-M).$$

Zahlenbeispiel:

Unter $N = 200$ Geräten eines Typs weisen $M = 80$ Mängel auf (Merkmalsträger). Wie groß ist die Wahrscheinlichkeit, in einer Zufallsstichprobe ohne Zurücklegen mit $n = 10$ Geräten genau $X = 3$ mangelhafte Geräte zu finden?

$$P\{X = 3\} = \frac{\binom{80}{3} \binom{120}{7}}{\binom{200}{10}} = 0,217696 \dots \approx \binom{10}{3} 0,4^3 \cdot 0,6^7 \sim 0,215$$

Bemerkung:

Die $H(n; M, N)$ -Verteilung stimmt mit der $H(M; n, N)$ -Verteilung überein! Drückt man alle Binominalkoeffizienten durch die Fakultäten aus (also z.B. $\binom{M}{k} = \frac{M!}{k!(M-k)!}$), so verifiziert man leicht, dass stets

$$\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \frac{\binom{n}{k} \binom{N-n}{M-k}}{\binom{N}{M}}$$

gilt!

1.6 Bivariate und multivariate Verteilungen, Kovarianz und Korrelation

Die gemeinsame Verteilung zweier diskreter Zufallsvariablen X und Y mit den möglichen Werten x_i bzw. y_j wird durch die gemeinsame Wahrscheinlichkeitsfunktion (WF)

$$f : (x_i, y_j) \rightarrow p_{ij} = f(x_i, y_j) := P\{X = x_i, Y = y_j\}$$

vollständig beschrieben.

Für Zufallsgrößen (reelle Werte!) ist die gemeinsame Verteilungsfunktion (VF) definiert durch

$$F(x, y) := P\{X \leq x, Y \leq y\} = \sum_{x_i \leq x, y_j \leq y} f(x_i, y_j)$$

für beliebige $x, y \in \mathbb{R}$.

Sind f_X, f_Y die Wahrscheinlichkeitsfunktionen von X bzw. Y allein, so ist die Unabhängigkeit von X und Y äquivalent zu

$$f(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j) \quad \text{für alle möglichen Werte } x_i, y_j.$$

Sind die Werte $x_i = i, y_j = j$ aus \mathbb{N}_0 , so ist die gemeinsame erzeugende Funktion ψ von (X, Y) definiert durch

$$\psi(s, t) := E(s^X \cdot t^Y) = \sum_{i=0}^{\infty} \cdot \sum_{j=0}^{\infty} p_{ij} s^i t^j. \quad (|s|, |t| \leq 1)$$

XIV rechnerischer Nachweis:

$$\begin{aligned} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} &= \frac{M(M-1)\dots(M-k+1)}{k!} \cdot \frac{(N-M)\dots(N-M-(n-k)+1)}{(n-k)!} \\ &= \frac{n!}{k!(n-k)!} \cdot \underbrace{\frac{M(M-1)\dots(M-k+1)}{N(N-1)\dots(N-k+1)}}_{k \text{ Faktoren}} \cdot \underbrace{\frac{(N-M)\dots(N-M-(n-k)+1)}{(N-k)\dots(N-n+1)}}_{n-k \text{ Faktoren}} \\ &\rightarrow \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für feste Zahlen } n, k \text{ und } M, N \rightarrow \infty \text{ mit } \frac{M}{N} \rightarrow p \end{aligned}$$

Im Fall der Unabhängigkeit von X und Y sind auch s^X , t^Y unabhängig, und es ist mit den erzeugenden Funktionen ψ_X , ψ_Y von X bzw. Y :

$$\psi(s, t) = E(s^X \cdot t^Y) = E(s^X) E(t^Y) = \psi_X(s) \cdot \psi_Y(t)$$

Aus der gemeinsamen Verteilung von X und Y erhält man auch die Verteilungen von X bzw. Y allein: Wegen

$$\{X = x_i\} = \{X = x_i, Y \text{ beliebig}\} = \bigcup_j \{X = x_i, Y = y_j\}$$

ist

$$P\{X = x_i\} = \sum_j P\{X = x_i, Y = y_j\} = \sum_j p_{ij}$$

Kurzschreibweise hierfür:

$$p_{i.} := \sum_j p_{ij}$$

Der Punkt in $p_{i.}$ steht also für die Summation über den weggelassenen Index! Analog ist

$$P\{Y = y_j\} = p_{.j} = \sum_i p_{ij}.$$

Weil in einer Tabelle mit den Wahrscheinlichkeiten p_{ij} die Werte $p_{i.}$, $p_{.j}$ oft als Randsummen eingetragen werden, nennt man die Wahrscheinlichkeiten $p_{i.}$, $p_{.j}$ auch die Rand- oder Marginalwahrscheinlichkeiten (engl. margin = Rand).

Beispiel:

In einer bestimmten Region finden sich in der Gesamtheit der Arbeitnehmer folgende Anteile^{XV} $p_{ij} = P\{X = x_i, Y = y_j\}$ für die Merkmale $X =$ Einkommensklasse und $Y =$ Geschlecht:

X : Einkommensklasse	Y : Geschlecht		Verteilung von X : $p_{i.} = \sum_j p_{ij}$
	$y_1 = m$	$y_2 = w$	
$x_1 =$ niedrig	0,188	0,226	0,414
$x_2 =$ mittel	0,415	0,118	0,533
$x_3 =$ hoch	0,042	0,011	0,053
Verteilung von Y : $p_{.j} = \sum_i p_{ij}$	0,645	0,355	1,000

1.6.1 Kovarianz und Korrelation

Für bivariate Verteilungen definieren wir zwei weitere Parameter:

Definition:

Sind X, Y zwei Zufallsgrößen mit den Erwartungswerten $\mu_X = E(X)$, $\mu_Y = E(Y)$ und den Varianzen $\sigma_X^2 = \text{Var}(X)$, $\sigma_Y^2 = \text{Var}(Y)$, so heißt

$$\sigma_{XY} = \text{Kov}(X, Y) := E((X - \mu_X)(Y - \mu_Y))$$

die Kovarianz von X und Y und

$$\rho = \text{Korr}(X, Y) := \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{Kov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}$$

der Korrelationskoeffizient oder kürzer die Korrelation von X und Y . Es ist

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

^{XV} Anteil einer Merkmalskombination (x_i, y_j) in der Grundgesamtheit Ω dieser Arbeitnehmer = Wahrscheinlichkeit bei zufälliger Ziehung eines Arbeitnehmers $\omega \in \Omega$ gerade diese Merkmalskombination (x_i, y_j) zu erhalten.

also

$$\text{Kov}(X, Y) = E(XY) - E(X)E(Y) = \mu_{XY} - \mu_X \mu_Y$$

Die Kovarianz hat folgende Eigenschaften:

1. X, Y unabhängig $\Rightarrow \text{Kov}(X, Y) = 0$
Die Umkehr gilt nicht allgemein!
2. $\text{Kov}(X, X) = \text{Var}(X) \quad (\sigma_{X,X} = \sigma_X^2)$
3. $\text{Kov}(a_0 + a_1 X, b_0 + b_1 Y) = a_1 b_1 \text{Kov}(X, Y)$
allgemeiner: $\text{Kov}(a_0 + \sum_i a_i X_i, b_0 + \sum_j b_j Y_j) = \sum_i \sum_j \text{Kov}(X_i, Y_j)$
4. $|\text{Kov}(X, Y)| \leq \sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)} \quad (|\sigma_{X,X}| \leq \sigma_X \sigma_Y)$

Das Gleichheitszeichen steht hier nur, wenn eine lineare Beziehung $Y = a + bX$ zwischen X und Y besteht mit Konstanten a und b .

Entsprechend gilt für die Korrelation von X mit Y :

1. X, Y unabhängig $\Rightarrow \rho = 0$
2. $|\rho| = \frac{|\sigma_{XY}|}{\sigma_X \sigma_Y} \leq 1$ und $|\rho| = 1$ nur, falls $Y = a + bX$ mit $b \neq 0$

Definition:

X, Y heißen unkorreliert, wenn $\rho = 0$

Aus der Unabhängigkeit folgt die Unkorreliertheit von X und Y , die Umkehr muß nicht gelten!

Haben die abhängigen Zufallsgrößen X, Y die Abweichungen $X - \mu_X$ und $Y - \mu_Y$ die Tendenz, häufiger gleiche und seltener verschiedene Vorzeichen anzunehmen, dann ist die Kovarianz σ_{XY} und damit auch $\rho > 0$. Eine positive Korrelation zwischen X und Y lässt sich als eine „positive“ Abhängigkeit zwischen X und Y deuten. Entsprechendes gilt für σ_{XY} und $\rho < 0$.

Kovarianzen treten als Zusatzglieder in der Varianz einer Summe $S = \sum_{i=1}^n X_i$ abhängiger Zufallsgrößen X_i auf: Mit $\mu_i = E(X_i)$, $\sigma_i^2 = \text{Var}(X_i)$ und $\sigma_{ij} = \text{Kov}(X_i, X_j)$ ist

$$\begin{aligned} \text{Var}(S) &= E(S - E(S))^2 = E\left(\sum_{i=1}^n (X_i - \mu_i)\right)^2 = E\left(\sum_{i=1}^n (X_i - \mu_i)^2 + 2 \sum_{1 \leq i < j \leq n} (X_i - \mu_i)(X_j - \mu_j)\right)^{\text{XVI}} \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Kov}(X_i, X_j) = \sum_{i=1}^n \sigma_i^2 + 2 \sum_{i < j} \sigma_{ij} \end{aligned}$$

Beispiel:

$S :=$ Anzahl der „Merkmalsträger“, in n Ziehungen ohne Zurücklegen aus einer Grundgesamtheit mit N Elementen, darunter M Merkmalsträgern.

$$S = \sum_{i=1}^n X_i \quad \text{mit } X_i = \begin{cases} 1, & \text{Merkmalsträger bei } i\text{-ter Ziehung gezogen} \\ 0, & \text{sonst} \end{cases}$$

$$E(X_i) = p := \frac{M}{N}$$

$$\text{Var}(X_i) = p(1 - p)$$

$$S \sim H(n; M, N)$$

$$\text{Var}(S) = ?$$

$$E(X_i X_j) = P\{X_i = 1, X_j = 1\} = \frac{M(M-1)}{N(N-1)} \quad (i \neq j) \quad \text{vgl. Übung1, Aufgabe 4c}$$

^{XVI} $2 \cdot \sum_{1 \leq i < j \leq n} \hat{=} \sum_{i \neq j}$

$$\begin{aligned} \Rightarrow \operatorname{Kov}(X_i, X_j) &= E(X_i X_j) - E(X_i) E(X_j) = \frac{M(M-1)}{N(N-1)} - \frac{M}{N} \frac{M}{N} = \\ &= \frac{1}{N-1} \frac{M}{N} \cdot \underbrace{\left(M-1 - \frac{M}{N}(N-1) \right)}_{\frac{M}{N}-1} = -\frac{1}{N-1} \frac{M}{N} \left(1 - \frac{M}{N} \right) = -\frac{p(1-p)}{N-1}. \end{aligned}$$

$$\begin{aligned} \operatorname{Var}(S) &= \sum_{i=1}^n \operatorname{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \operatorname{Kov}(X_i, X_j) = \\ &= np(1-p) - n(n-1) \frac{p(1-p)}{N-1} = np(1-p) \left(1 - \frac{n-1}{N-1} \right) \end{aligned}$$

Die schwach negative Korrelation der Indikatorgrößen X_i bewirkt eine um den Faktor $1 - \frac{n-1}{N-1} = \frac{N-n}{N-1}$ ($\approx 1 - \frac{n}{N}$)^{XVII} kleinere Varianz als beim Ziehen mit Zurücklegen! ($np(1-p)$ = Varianz einer $B(n; p)$ -Verteilung)

Schätzung von Kovarianz und Korrelation aus einer Zufallsstichprobe mit i.i.d.r.v. Wertepaaren (X_i, Y_i) , $i = 1, \dots, n$:

$$\begin{aligned} s_{XY} = \hat{\sigma}_{XY} &:= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \\ r = \hat{\rho} &= \frac{s_{XY}}{s_X \cdot s_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, \quad (|r| \leq 1) \end{aligned}$$

Siehe dazu auch in Kapitel 1.4.4 auf Seite 17.

Bemerkungen:

Für das Skalarprodukt der Vektoren

$$\vec{x} := \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} \quad \text{und} \quad \vec{y} := \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \quad \text{mit} \quad \vec{x}, \vec{y} \in \mathbb{R}^n$$

gilt gemäß der Cauchy-Schwarz-Ungleichung:

$$\frac{|\vec{x} \bullet \vec{y}|}{|\vec{x}| \cdot |\vec{y}|} \leq 1 \quad (= 1 \text{ nur für } \vec{y} = b\vec{x}), \text{ also } |r| \leq 1^{\text{XIX}}.$$

Wegen $r = r_n \xrightarrow{p} \rho$ für $n \rightarrow \infty$ folgt auch $|\rho| \leq 1$ und somit $|\sigma_{XY}| \leq \sigma_X \sigma_Y$ für die Kovarianz. (vgl. obige Eigenschaft Nr. 4 der Kovarianz.)

Die Korrelation ρ zwischen X und Y ist dasselbe wie die Kovarianz der standardisierten Zufallsgrößen $X^* := \frac{X - \mu_X}{\sigma_X}$, $Y^* := \frac{Y - \mu_Y}{\sigma_Y}$:

$$\operatorname{Kov}(X^*, Y^*) = E \left(\frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y} \right) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \rho$$

1.6.2 Multivariate Verteilungen

Sind X_1, \dots, X_r diskrete Zufallsvariable mit möglichen Werten x_{ik} , $i = 1, \dots, r$, so ist die gemeinsame Wahrscheinlichkeitsfunktion

$$f(x_{1k_1}, \dots, x_{rk_r}) = P\{X_1 = x_{1k_1}, \dots, X_r = x_{rk_r}\}$$

^{XVII} $\frac{n}{N}$: Auswahlatz

^{XVIII} Korrelationstaste am Taschenrechner!

^{XIX} wie bekannt:

$$\vec{x} \cdot \vec{y} = |\vec{x}| |\vec{y}| \cdot \cos \varphi \quad \Rightarrow \quad \cos \varphi = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} \quad \Rightarrow \quad |\cos \varphi| \leq 1$$

Haben alle X_i den gleichen Wertebereich mit Werten x_k , so kann man einfacher schreiben:

$$f(x_{k_1}, \dots, x_{k_r}) = P\{X_1 = x_{k_1}, \dots, X_r = x_{k_r}\}$$

Für r Zufallsgrößen X_i ist die gemeinsame Verteilungsfunktion $F(x_1, \dots, x_r) := P\{X_1 \leq x_1, \dots, X_r \leq x_r\}$ für alle $(x_1, \dots, x_r) \in \mathbb{R}^r$.

1.6.3 Die Multinomialverteilung

Es seien A_1, \dots, A_r paarweise disjunkte Ereignisse bei einem Vorgang V mit $p_i = P(A_i)$ und $\sum_{i=1}^r p_i = 1$. Meist sind die A_i durch die r möglichen Werte („Ausprägungen“) z_i eines Merkmals Z definiert, also $A_i = \{Z = z_i\}$, $i = 1, \dots, r$. Es sei Y_i (oder Y_{ni}) die absolute Häufigkeit von A_i (bzw. des Wertes z_i) bei n unabhängigen Ausführungen von V , $i = 1, \dots, r$. Es ist also $0 \leq Y_i \leq n$ und $\sum_{i=1}^r Y_i = n$. Die gemeinsame WF von (Y_1, \dots, Y_r) ist dann durch

$$f(n_1, \dots, n_r) = P\{Y_1 = n_1, \dots, Y_r = n_r\} = \frac{n!}{n_1! \cdot \dots \cdot n_r!} p_1^{n_1} \cdot \dots \cdot p_r^{n_r}$$

gegeben, wobei alle $n_i \geq 0$, $\sum_{i=1}^r n_i = n$ und $\sum_{i=1}^r p_i = 1$.

Für $r = 2$ erhält man mit $p_1 = p$, $p_2 = 1 - p$, $n_1 = k$, $n_2 = n - k$ die frühere Formel $\binom{n}{k} p^k (1 - p)^{n-k}$ für die $B(n; p)$ -Verteilung. $\vec{Y} = (Y_1, \dots, Y_r)$ hat eine Multinomialverteilung mit den Parametern n, p_1, \dots, p_r , kurz $\vec{Y} \sim M(n; p_1, \dots, p_r)$.

Jede einzelne Häufigkeit Y_i ist $B(n; p_i)$ -verteilt, also ist

$$\begin{aligned} E(Y_i) &= n p_i \\ \text{Var}(Y_i) &= n p_i (1 - p_i). \end{aligned}$$

$\text{Kov}(Y_i, Y_j)$ für $i \neq j$ kann mit Hilfe von Indikatorvariablen berechnet werden:

$$X_{ik} := \begin{cases} 1, & \text{wenn } A_i \text{ bei der } k\text{-ten Ausführung von } V \text{ eintritt} \\ 0, & \text{sonst} \end{cases}$$

Entsprechend seien X_{jl} die Indikatorvariablen für A_j , also

$$\begin{aligned} Y_i &= \sum_{k=1}^n X_{ik} \\ Y_j &= \sum_{l=1}^n X_{jl} \end{aligned}$$

Da A_i, A_j einander ausschließen, ist stets $X_{ik} X_{jk} = 0$ und

$$E(Y_i Y_j) = E\left(\sum_{k=1}^n \sum_{l=1}^n X_{ik} X_{jl}\right) = \sum_{k \neq l} E(X_{ik} X_{jl}) = \sum_{k \neq l} E(X_{ik}) E(X_{jl}) \stackrel{n(n-1) \text{ Summanden!}}{=} n(n-1) p_i p_j$$

Somit folgt:

$$\text{Kov}(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i) E(Y_j) = n(n-1) p_i p_j - n p_i n p_j = -n p_i p_j \quad (i \neq j)$$

Alternativ hierzu kann man mit der erzeugenden Funktion

$$\psi(z_1, \dots, z_r) = E\left(z_1^{Y_1} \cdot \dots \cdot z_r^{Y_r}\right) = \left(\sum_{i=1}^r p_i z_i\right)^n = \sum_{n_1 + \dots + n_r = n} \frac{n!}{n_1! \cdot \dots \cdot n_r!} \prod_{i=1}^r (p_i z_i)^{n_i}$$

der $M(n; p_1, \dots, p_r)$ -Verteilung

$$E(Y_i Y_j) = \frac{\partial^2 \psi}{\partial z_i \partial z_j}(1, \dots, 1) = n(n-1) p_i p_j$$

erhalten.

Zahlenbeispiel:

Ein Kartenspiel enthält 16 rote, 12 grüne, 12 blaue und 8 gelbe Karten. Es wird 30-mal zufällig eine Karte mit Zurücklegen gezogen und jedes mal die Farbe notiert. Wie groß ist die Wahrscheinlichkeit, dass 9-mal eine rote, 8-mal eine grüne, 7-mal eine blaue und 6-mal eine gelbe Karte gezogen wird?

$$P\{Y_1 = 9, Y_2 = 8, Y_3 = 7, Y_4 = 6\} = \frac{30!}{9!8!7!6!} \left(\frac{1}{3}\right)^9 \left(\frac{1}{4}\right)^8 \left(\frac{1}{4}\right)^7 \left(\frac{1}{6}\right)^6 = 5,066579 \cdot 10^{-3}.$$

1.6.4 Die multivariate hypergeometrische Verteilung

Eine Grundgesamtheit mit N Elementen enthält M_j Elemente mit der „Merkmalsausprägung“ Nr. j , $j = 1, \dots, r$, $M_1 + \dots + M_r = N$.

Es wird n -mal ein Element zufällig ohne Zurücklegen aus dieser Gesamtheit gezogen. Dabei sei Y_j die Häufigkeit der Merkmalsausprägung Nr. j in dieser Stichprobe, $j = 1, \dots, r$. Die gemeinsame Verteilung von $\vec{Y} = (Y_1, \dots, Y_r)$ ist eine multivariate hypergeometrische Verteilung, abhängig von den Parametern n, M_1, \dots, M_r , kurz $\vec{Y} \sim H(n; M_1, \dots, M_r)^{XX}$. Es ist

$$P\{Y_1 = n_1, \dots, Y_r = n_r\} = \frac{\binom{M_1}{n_1} \cdot \dots \cdot \binom{M_r}{n_r}}{\binom{N}{n}} \quad \text{mit } n_1 + \dots + n_r = n$$

Jede einzelne Häufigkeit Y_j ist $H(n; M_j, N)$ -verteilt, und es ist für $i \neq j$:

$$\text{Kov}(Y_i, Y_j) = -n p_i p_j \left(1 - \frac{n-1}{N-1}\right) \quad \text{mit } p_j := \frac{M_j}{N}, \quad j = 1, \dots, r.$$

Für $n \ll \min(M_1, \dots, M_r)$ kann die $H(n; M_1, \dots, M_r)$ -Verteilung durch die $M(n; p_1, \dots, p_r)$ -Verteilung mit $p_j = \frac{M_j}{N}$ approximiert werden, da sich dann das Ziehen ohne Zurücklegen nicht mehr merklich vom Ziehen mit Zurücklegen unterscheidet, die Veränderung der Grundgesamtheit durch die sukzessiven Entnahmen somit vernachlässigbar ist

Zahlenbeispiel:

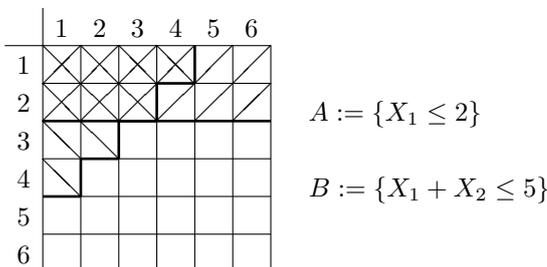
Die gleiche Grundgesamtheit aus Karten wie im Beispiel zur Multinomialverteilung, jetzt aber mit $n = 30$ Ziehungen ohne Zurücklegen:

$$P\{Y_1 = 9, Y_2 = 8, Y_3 = 7, Y_4 = 6\} = \frac{\binom{16}{9} \binom{12}{8} \binom{12}{7} \binom{8}{6}}{\binom{48}{30}} = 0,01717935 \dots$$

1.7 Bedingte Wahrscheinlichkeit, Bayes-Formel

1.7.1 Bedingte Wahrscheinlichkeit

Beispiel: Wurf zweier idealer Würfel



$$P(A) = \frac{12}{36} = \frac{1}{3}$$

$$P(B) = \frac{10}{36} = \frac{5}{18}$$

$$P(A \cap B) = \frac{7}{36}$$

^{XX}Wegen $M_1 + \dots + M_r = N$ kann man auch die Schreibweise $H(n; M_1, \dots, M_{r-1}, N)$ vereinbaren. Für $r = 2$ stimmt diese Notation mit der Bezeichnung $H(n; M, N)$ für die univariate (= eindimensionale) hypergeometrische Verteilung überein

Wie groß ist die Wahrscheinlichkeit für A , wenn schon bekannt ist, dass B bereits eingetreten ist?

Schreibweise: $P(A|B)$

Sprechweise: bedingte Wahrscheinlichkeit von A unter B (nämlich unter der Bedingung, dass B eingetreten ist)

Zähle hierzu nur die Fälle von A innerhalb von B ! B tritt hier an die Stelle der Grundgesamtheit Ω mit allen 36 Fällen!

$$P(A|B) = \frac{7}{10} = \frac{\frac{7}{36}}{\frac{10}{36}} = \frac{P(A \cap B)}{P(B)} \quad \left(\text{zum Vergleich: } P(A) = \frac{1}{3} < P(A|B) \right)$$

$$P(B|A) = \frac{7}{12} = \frac{\frac{7}{36}}{\frac{12}{36}} = \frac{P(B \cap A)}{P(A)} \quad \left(\text{zum Vergleich: } P(B) = \frac{5}{18} < P(B|A) \right)$$

Allgemeine Definition:

A und B seien zwei Ereignisse mit $P(B) > 0$ bei einem Vorgang mit zufälligen Ergebnissen.

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

heißt die bedingte Wahrscheinlichkeit von A unter B .

Stets gilt:

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

Speziell für unabhängige Ereignisse A, B gilt:

$$\begin{aligned} P(A \cap B) &= P(A) P(B) = P(A|B) P(B) = \\ &= P(B \cap A) = P(B) P(A) = P(B|A) P(A) \end{aligned}$$

Nach Division durch $P(B)$ bzw. $P(A)$ folgt:

$$\boxed{A, B \text{ unabhängig} \Leftrightarrow P(A) = P(A|B) \Leftrightarrow P(B) = P(B|A)}$$

Zur Veranschaulichung der Gleichung $P(A) = P(A|B)$, welche die Unabhängigkeit von A und B zum Ausdruck bringt:

Stellen sie sich vor, sie müssten raten, ob A eingetreten ist. Wenn für sie dabei die zusätzliche Information, dass B bereits eingetreten ist, völlig wertlos ist, weil die Wahrscheinlichkeit für das Eintreten von A durch diese Information nicht verändert wird, dann sind A und B unabhängige Ereignisse, und es gilt dann auch umgekehrt: $P(B) = P(B|A)$.

Beispiel:

Gemeinsame Verteilung von Einkommensklasse X und Geschlecht Y (siehe Beispiel auf Seite 24):

Die bedingte Wahrscheinlichkeit, dass eine zufällig aus dieser Grundgesamtheit gezogene Person ein hohes Einkommen hat, wenn sie ein Mann ist, hat den Wert

$$p_{3|1} := P\{X = \text{hoch} | Y = \text{m}\} = \frac{p_{31}}{p_{\cdot 1}} = \frac{0,042}{0,645} \approx 0,0651 > p_3 \stackrel{\text{XXI}}{=} P\{X = \text{hoch}\} = 0,053.$$

(also $p_{31} > p_3 \cdot p_{\cdot 1}$) Die entsprechende Wahrscheinlichkeit, falls die gezogene Person eine Frau ist:

$$p_{3|2} = \frac{p_{32}}{p_{\cdot 2}} = \frac{0,011}{0,355} \approx 0,0310 < p_3 = 0,053.$$

(also $p_{32} < p_3 \cdot p_{\cdot 2}$)

Bei Unabhängigkeit von X, Y hätte für alle Werte x_i, y_j gelten müssen:

$$p_{ij} = p_i \cdot p_{\cdot j}.$$

^{XXI} $p_3 = p_{31} + p_{32}$

Die bedingte Verteilung eines diskreten Merkmals Y mit Werten y_j unter B ist definiert durch die bedingten Wahrscheinlichkeiten

$$f(y_j|B) := P\{Y = y_j|B\} = \frac{P(\{Y = y_j\} \cap B)}{P(B)} \quad \text{für alle möglichen } y_j$$

Beispiel:

Es seien X_1, X_2 unabhängige $P(\mu_i)$ -verteilte Häufigkeiten. Gesucht ist die Verteilung von X_1 unter der Bedingung $\{X_1 + X_2 = n\}$ mit bekanntem $n \in \mathbb{N}$, also die Wahrscheinlichkeiten

$$p_k := P\{X_1 = k|X_1 + X_2 = n\}, \quad k = 0, \dots, n.$$

$$p_k = \frac{P(\{X_1 = k\} \cap \{X_1 + X_2 = n\})}{P\{X_1 + X_2 = n\}} = \frac{P\{X_1 = k, X_2 = n - k\}}{P\{X_1 + X_2 = n\}}$$

Wegen $X_1 + X_2 \sim P(\mu_1 + \mu_2)$ folgt:

$$p_k = \frac{e^{-\mu_1} \cdot \frac{\mu_1^k}{k!} \cdot e^{-\mu_2} \cdot \frac{\mu_2^{n-k}}{(n-k)!}}{e^{-(\mu_1 + \mu_2)} \cdot \frac{(\mu_1 + \mu_2)^n}{n!}} = \frac{n!}{k!(n-k)!} \cdot \left(\frac{\mu_1}{\mu_1 + \mu_2}\right)^k \cdot \left(\frac{\mu_2}{\mu_1 + \mu_2}\right)^{n-k} =$$

$$= \binom{n}{k} p^k (1-p)^{n-k} \quad \text{mit } p := \frac{\mu_1}{\mu_1 + \mu_2}.$$

Kurz: $X_1|X_1 + X_2 = n \sim B(n; p)$.

Eine statistische Anwendung:

In zwei Städten S_1, S_2 mit $N_1 = 57600$ und $N_2 = 86400$ Verkehrsteilnehmern kommen in einem Jahr insgesamt $X_1 + X_2 = 11$ tödliche Verkehrsunfälle vor, davon $X_1 = 7$ in S_1 .

Wie glaubwürdig ist die sogenannte Nullhypothese

$H_0 :=$ Das Risiko (die Wahrscheinlichkeit) eines tödlichen Verkehrsunfalls ist in beiden Städten identisch, d.h.

$$\frac{\mu_1}{\mu_2} = \frac{E(X_1)}{E(X_2)} = \frac{N_1}{N_2}$$

Als Maß für die „Glaubwürdigkeit“ von H_0 kann man die unter Annahme der Gültigkeit von H_0 berechnete Wahrscheinlichkeit ansehen, dass X_1 mindestens so hoch ausfällt wie der beobachtete Wert 7 bei gegebener Gesamtzahl $X_1 + X_2 = 11$. Mit

$$p = \frac{\mu_1}{\mu_1 + \mu_2} = \frac{N_1}{N_1 + N_2} = 0,4$$

erhält man:

$$P_{H_0}\{X_1 \geq 7|X_1 + X_2 = 11\} = \sum_{k=7}^{11} \binom{11}{k} 0,4^k \cdot 0,6^{11-k} = 0,09935 \dots \approx 0,1$$

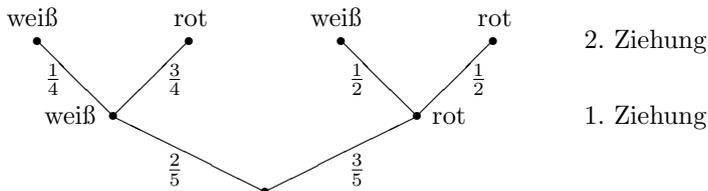
Das, was tatsächlich eingetreten ist - nämlich $\{X_1 \geq 7|X_1 + X_2 = 11\}$ - ist also ein verhältnismäßig unwahrscheinliches Ergebnis, wenn H_0 wahr wäre. Unter der Alternativhypothese H_1 : „das Risiko ist in S_1 größer als in S_2 “ wäre auch die Wahrscheinlichkeit für $\{X_1 \geq 7|X_1 + X_2 = 11\}$ größer.

Solche Alternativentscheidungen zwischen zwei Hypothesen (einander ausschließende Vorstellungen über die Realität, bzw. über Werte unbekannter Parameter) nennt man in der Statistik Tests. Wenn bei einem für das vorliegende Entscheidungsproblem geeigneten Testverfahren das „Glaubwürdigkeitsmaß“ P_{H_0} unter einen vorgegebenen Wert α (meist $\alpha = 0,05$ oder $\alpha = 0,01$) fällt, dann wird H_0 als unglaubwürdig abgelehnt. α heißt das Niveau des Tests oder genauer die Irrtumswahrscheinlichkeit des Tests für den sogenannten Fehler 1. Art, d.h. H_0 abzulehnen, obwohl H_0 wahr ist.

Bedingte Wahrscheinlichkeiten treten auch beim Ziehen ohne Zurücklegen auf. Die möglichen Ziehungen und ihre bedingte Wahrscheinlichkeiten lassen sich durch Ereignisbäume darstellen. Es soll z.B. zweimal aus insgesamt drei

„roten“ und zwei „weißen“ Elementen zufällig ein Element ohne Zurücklegen entnommen werden. Der folgende Ereignisbaum zeigt alle möglichen Ziehungsfolgen als Pfade:

$N = 5$, $M = 3$ „rote“ und $N - M = 2$ „weiße“ Elemente:



Mit $Z_i :=$ Ergebnis der i -ten Ziehung, $i = 1, 2$, ist z.B.:

$$P\{Z_1 = \text{weiß}\} = \frac{2}{5}$$

$$P\{Z_2 = \text{rot} | Z_1 = \text{weiß}\} = \frac{3}{4}$$

$$P\{Z_2 = \text{rot}, Z_1 = \text{weiß}\} = \frac{3}{4} \cdot \frac{2}{5} = 0,3$$

$$P\{Z_2 = \text{rot}\} = P\{Z_2 = r | Z_1 = w\} \cdot P\{Z_1 = w\} + P\{Z_2 = r | Z_1 = r\} \cdot P\{Z_1 = r\} = \frac{3}{4} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{3}{5} = \frac{3}{5}$$

Das letzte Ergebnis folgt einfacher direkt daraus, dass jedes der 5 Elemente mit gleicher Wahrscheinlichkeit bei der zweiten Ziehung gezogen werden kann.

Allgemeiner:

Ein System mit endlich vielen möglichen Zuständen z_1, \dots, z_n nimmt an aufeinander folgenden Zeitpunkten t_0, t_1, \dots, t_k die zufallsabhängigen Zustände Z_0, Z_1, \dots, Z_k an. Es sei $A_j^{(i)} := \{Z_i = z_j\}$. Alle möglichen zeitlichen Entwicklungen des mit dem Zustand z_{j_0} startenden Systems können durch die bei der Wurzel $A_{j_0}^{(0)}$ beginnenden Pfade mit den Knoten $A_{j_i}^{(i)}$ auf der i -ten Stufe (Ebene) dargestellt werden. Z.B. ist mit $k = 2$:

$$P\left(A_{j_0}^{(0)} \cap A_{j_1}^{(1)} \cap A_{j_2}^{(2)}\right) = P\left(A_{j_0}^{(0)}\right) \cdot P\left(A_{j_1}^{(1)} | A_{j_0}^{(0)}\right) \cdot P\left(A_{j_2}^{(2)} | A_{j_0}^{(0)} \cap A_{j_1}^{(1)}\right)$$

Fasst man die bedingten Wahrscheinlichkeiten

$$p_{j_i | j_0 \dots j_{i-1}} := P\left(A_{j_i}^{(i)} | A_{j_0}^{(0)} \cap \dots \cap A_{j_{i-1}}^{(i-1)}\right)$$

als Bewertungen der Pfadabschnitte zwischen $A_{j_{i-1}}^{(i-1)}$ und $A_{j_i}^{(i)}$ auf, so ist die Wahrscheinlichkeit einer bestimmten Systementwicklung = der Wahrscheinlichkeit eines bestimmten bei $A_{j_0}^{(0)}$ startenden Pfades = Produkt der Bewertungen der einzelnen Pfadabschnitte.

Besonders wichtig sind Systeme, bei denen die Übergangswahrscheinlichkeit von einem Zustand $z_{j_{i-1}}$ in einen Zustand z_{j_i} nur von diesen beiden Zuständen und nicht von den früheren Zuständen oder der Zeit abhängen (sogenannte stationäre Markovketten). Mit der vereinfachten Notation $z_i :=$ Wert von Z_i ist dann stets

$$P\{Z_i = z_i | Z_{i-1} = z_{i-1}, \dots, Z_0 = z_0\} = P(\{Z_i = z_i | Z_{i-1} = z_{i-1}\})$$

Einfaches Beispiel:

Zwei Spieler spielen mehrmals hintereinander ein Spiel. Die Gewinnwahrscheinlichkeit für Spieler Nr. 1 sei p . Im Fall eines Gewinns erhält Spieler Nr. 1 einen Euro, umgekehrt verliert Spieler Nr. 1 mit Wahrscheinlichkeit $q = 1 - p$ und muß einen Euro an Spieler Nr. 2 zahlen. Es sei $S_n = X_1 + \dots + X_n$ der „Gewinn“ von Spieler Nr. 1 nach n Spielen, wobei $p = P\{X_i = 1\}$, $q = P\{X_i = -1\}$, alle X_i i.i.d. Offenbar gilt (bei unbeschränkten Kredit):

$$P\{S_n = s_n | S_{n-1} = s_{n-1}, \dots, S_1 = s_1\} = P\{S_n = s_n | S_{n-1} = s_{n-1}\}$$

und zwar

$$P\{S_n = k | S_{n-1} = k - 1\} = p$$

$$P\{S_n = k | S_{n-1} = k + 1\} = q$$

Übung:

Berechne $P\{S_n = k\}$ für $-n \leq k \leq n$.

1.7.2 Die Bayes-Formel

Vorbemerkung:

Ist B_1, \dots, B_n eine Zerlegung des sicheren Ereignisses Ω , d.h.

$$\Omega = \bigcup_{i=1}^n B_i, \quad B_i \cap B_j = \emptyset \quad \text{für } i \neq j,$$

so gilt für jedes Ereignis $A \subseteq \Omega$:

$$P(A) = P(A \cap \Omega) = P\left(\bigcup_{i=1}^n A \cap B_i\right) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A|B_i) P(B_i).$$

Nun seien A_1, \dots, A_m und B_1, \dots, B_n zwei verschiedene Zerlegungen von Ω , $\Omega = \bigcup_{i=1}^m A_i = \bigcup_{j=1}^n B_j$. Ferner seien alle Wahrscheinlichkeiten $P(B_j)$ und $P(A_i|B_j)$ bekannt. Gesucht sind die „umgekehrten“ bedingten Wahrscheinlichkeiten $P(B_k|A_i)$.

Bayes-Formel:

$$P(B_k|A_i) = \frac{P(A_i|B_k) \cdot P(B_k)}{\sum_{j=1}^n P(A_i|B_j) \cdot P(B_j)}$$

Begründung:

Der Zähler ist $P(A_i \cap B_k) = P(B_k \cap A_i)$ und der Nenner ist nach obiger Zerlegungsformel gleich $P(A_i)$.

Beispiel:

Ω := durch Röntgen auf Tuberkulose (TB) geprüfte Grundgesamtheit von Personen.

A := „positiver Befund auf TB bei zufällig ausgewählter Person“

B := „zufällig ausgewählte Person hat tatsächlich TB“

$$\begin{aligned} P(A|B) &= 0,95 && (= \text{Wahrscheinlichkeit für richtig positiven Befund}) \\ P(A|\bar{B}) &= 0,01 && (= \text{Wahrscheinlichkeit für falsch positiven Befund}) \\ P(B) &= 0,001 && (= \text{Anteil der TB-Kranken in } \Omega) \end{aligned}$$

Wie groß ist bei positivem Befund die Wahrscheinlichkeit dafür, dass tatsächlich TB vorliegt? (hier $A_1 = A$, $A_2 = \bar{A}$, $B_1 = B$, $B_2 = \bar{B}$)

$$P(B|A) = \frac{P(A|B) P(B)}{P(A|B) P(B) + P(A|\bar{B}) P(\bar{B})} = \frac{0,95 \cdot 0,001}{0,95 \cdot 0,001 + 0,01 \cdot 0,999} \approx 0,0868.$$

2 Stetige Verteilungen

2.1 Wahrscheinlichkeitsdichte, Parameter von Verteilungen

2.1.1 Wahrscheinlichkeitsdichte

Es sei X eine Zufallsgröße mit einem Intervall (a, b) als Wertebereich ($a = -\infty$ oder $b = \infty$ zulässig):

Definition:

X heißt stetig verteilt, wenn die Verteilungsfunktion (VF) $F(x) = P\{X \leq x\}$, $x \in \mathbb{R}$, stetig ist. Ist $F(x)$ (wenigstens stückweise) differenzierbar, so heißt

$$f(x) := \frac{d}{dx} F(x)$$

die Wahrscheinlichkeitsdichte (WD) von X

Stets gilt dann: $f(x) \geq 0$ und

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx$$

Insbesondere:

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

und

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

$$\left(\int_{-\infty}^{\infty} f(x) dx = \int_a^b f(x) dx, \text{ falls } f(x) = 0 \text{ für } x \notin (a, b) \right)$$

ZEICHNUNG FOLGT!!

Die Wahrscheinlichkeit für das Ereignis $\{x_1 < X \leq x_2\}$ ist also anschaulich die Fläche unter der Dichtekurve zwischen den Stellen x_1 und x_2 . Es ist hier belanglos, ob $x_1 < X$ oder $x_1 \leq X$ geschrieben wird.

2.1.2 Einige Parameter (Kenngrößen) von Verteilungen

1. Die Momente von X :

Für eine Zufallsgröße X mit der W-Dichte $f(x)$ ist das erste Moment μ_1 der Erwartungswert $\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$.¹

Man sagt: μ „existiert“, wenn $\int_{-\infty}^{\infty} x f(x) dx$ absolut konvergent ist, d.h. $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$. Dann ist auch $\int_{-\infty}^{\infty} x f(x) dx$ konvergent und μ ist eine reelle Zahl.

Entsprechend ist das n -te Moment $\mu_n = E(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx$, $n \in \mathbb{N}_0$, falls noch $\int_{-\infty}^{\infty} |x|^n f(x) dx < \infty$.

¹Zur Veranschaulichung: X kann nur diskret als X^* gemessen werden mit möglichen Werten x_j und Abständen $\Delta x = x_{j+1} - x_j$.

($\mu_0 = 1$, $\mu_1 = \mu$). Aus der Existenz von μ_n folgt immer die Existenz von μ_k für alle $k < n$.

In der Regel wird in diesem Kurs für eine Zufallsgröße mindestens die Existenz von μ und μ_2 vorausgesetzt.

Allgemein ist für eine absolut integrierbare Funktion $Y = h(X)$

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

absolute Momente von X :

$$E(|X|^n) = \int_{-\infty}^{\infty} |x|^n f(x) dx \quad n \in \mathbb{N}_0$$

zentrale Momente von X :

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

Insbesondere ist das zweite zentrale Moment

$$\sigma^2 = \text{Var}(X) = E(X - \mu)^2 = E(X^2) - \mu^2 = \mu_2 - \mu^2$$

die Varianz von X !

$$\gamma := E\left(\frac{X - \mu}{\sigma}\right)^3$$

heißt die Schiefe von X . (engl.: skewness)

Ist X um μ herum symmetrisch verteilt, d.h. $f(\mu + z) = f(\mu - z)$, dann ist

$$E(X - \mu)^3 = \int_{-\infty}^{\infty} (x - \mu)^3 f(x) dx = \int_{-\infty}^{\infty} z^3 f(\mu + z) dz = 0, \quad \text{also } \gamma = 0$$

ZEICHNUNGEN!!

2. Die Quantile einer stetigen Verteilung:

Das p -Quantil von X - bzw. der Verteilung von X - ist der Wert q_p , der von X mit der Wahrscheinlichkeit p unterschritten wird, d.h.

$$P\{X \leq q_p\} = F(q_p) = p \quad \Leftrightarrow \quad q_p = F^{-1}(p) \quad (\text{mit der Umkehrfunktion } F^{-1} \text{ zu } F)$$

Speziell heißt $m = q_{0,5}$ der Median von X („50%-Punkt“) und

$q_{0,25}$ das untere Quartil von X

$q_{0,75}$ das obere Quartil von X

ZEICHNUNGEN!!

Z.B. ist $\{X^* = 73 \text{ mm}\} = \{72,5 \text{ mm} \leq X < 73,5 \text{ mm}\}$, wenn mit einer Messgenauigkeit von $\pm \frac{1}{2} \Delta x = 0,5 \text{ mm}$ gemessen wird.
Mit

$$p_j := P\{X^* = x_j\} = P\left\{x_j - \frac{1}{2} \Delta x \leq X < x_j + \frac{1}{2} \Delta x\right\} = \int_{x_j - \frac{1}{2} \Delta x}^{x_j + \frac{1}{2} \Delta x} f(x) dx \simeq f(x_j) \Delta x \quad \text{für } \Delta x \rightarrow 0 \text{ folgt:}$$

$$E(X^*) = \sum_j x_j p_j \simeq \sum_j x_j f(x_j) \Delta x \rightarrow \int_{-\infty}^{\infty} x f(x) dx$$

2.1.3 Tabelle einiger stetige Verteilungen

Name	Symbol	Parameter	Dichte $f(x)$	$\{x f(x) > 0\}$	$E(X)$	$\text{Var}(X)$
Gleich- oder Rechteckverteilung	$R(a; b)$	$a < b$	$\frac{1}{b-a}$	$a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normalverteilung	$N(\mu; \sigma^2)$	$\mu \in \mathbb{R}, \sigma > 0$	$\frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{\sqrt{2\pi}\sigma}$	$x \in \mathbb{R}$	μ	σ^2
Standardnormalverteilung	$N(0; 1)$	$\mu = 0, \sigma = 1$	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$	$x \in \mathbb{R}$	0	1
Exponentialverteilung	$E(\lambda)$	$\lambda > 0$	$\lambda e^{-\lambda x}$	$x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gammaverteilung ^{II}	$G(\alpha; \lambda)$	$\alpha, \lambda > 0$	$\lambda \frac{(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$	$x > 0$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Weibullverteilung	$W(\alpha; \lambda)$	$\alpha, \lambda > 0$	$\alpha \lambda (\lambda x)^{\alpha-1} \cdot e^{-(\lambda x)^\alpha}$	$x > 0$	$\frac{\Gamma(1+\frac{1}{\alpha})}{\lambda}$	$\frac{\Gamma(1+\frac{2}{\alpha})-\Gamma(1+\frac{1}{\alpha})^2}{\lambda^2}$
Betaverteilung ^{III}	$B(\alpha; \beta)$	$\alpha, \beta > 0$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$	$0 < x < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
t -Verteilung mit ν Freiheitsgraden	$t(\nu)$	$\nu > 2$	$\frac{\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)}$	$x \in \mathbb{R}$	0	$\frac{\nu}{\nu-2}$

2.1.4 Beispiele stetiger Verteilungen

1. Gleich- oder Rechteckverteilung ($R(a; b)$ -Verteilung)

Eine Zufallsgröße X heißt gleichverteilt in $(a; b)$, wenn sie eine auf $(a; b)$ konstante Dichte f hat, also wegen $\int_{\mathbb{R}} f(x) dx = 1$:

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

$$\Rightarrow P\{X \leq x\} = F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x \geq b \end{cases}$$

Viele Programme (und auch Taschenrechner) ermöglichen zur Simulation von Vorgängen, bei denen Zufallsgrößen auftreten, die Erzeugung sogenannter **Pseudozufallszahlen**. (Solche Programme heißen Pseudozufallsgeneratoren.) Obwohl eine Folge X_1, X_2, \dots solcher Zahlen deterministisch durch eine Rekursionsformel erzeugt wird, verhalten sich die Werte X_1, X_2, \dots mit sehr guter Näherung wie unabhängige Zufallsgrößen mit einer gewünschten Verteilungsfunktion $F(x)$. Bietet das Programm in $(0; 1)$ gleichverteilte Pseudozufallsgrößen U_i an, so gewinnt man für eine stetige streng wachsende VF $F(x)$ durch die Transformation $X_i := F^{-1}(U_i)$ Pseudozufallsgrößen X_i mit der VF $F(x)$, denn es gilt:

$$P\{X = F^{-1}(U) \leq x\} = P\{U \leq F(x)\} = F(x)$$

da für jede Zahl $u \in (0; 1)$ gilt:

$$P\{U \leq u\} = u$$

Beispiel:

Erzeugung (quasi) exponential verteilter Pseudozufallsgrößen X_i aus $R(0; 1)$ -Pseudozufallsgrößen U_i . Die X_i sollen die VF $F(x) = 1 - e^{-\lambda x}$, ($x > 0$), haben. Dies wird erreicht durch $X_i = F^{-1}(U_i) = -\frac{1}{\lambda} \ln(1 - U_i)$. Da mit

^{II} $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} dx$, $\alpha > 0$ ist die Gammafunktion. Es ist $\Gamma(1) = 1$ und $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ (zeigt man durch partielle Integration). Für $n \in \mathbb{N}$ folgt: $\Gamma(n + 1) = n(n - 1) \cdot \dots \cdot 2 \cdot 1 = n!$, d.h. die Gammafunktion interpoliert die Fakultät! Ferner ist $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\lim_{\alpha \downarrow 0} \Gamma(\alpha) = \infty$.

^{III} $B(\alpha, \beta)$ $\alpha > 0, \beta > 0$, ist die Betafunktion. Es ist $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$

U_i auch $1 - U_i$ $R(0; 1)$ -verteilt ist, kann man auch $X_i = -\frac{1}{\lambda} \ln U$ setzen. Eine Zufallsgröße X mit einer diskreten Gleichverteilung auf $\{0, 1, \dots, n-1\}$, d.h. $P\{X = k\} = \frac{1}{n}$ für $k = 0, \dots, n-1$ läßt sich durch $X := [nU]$ erzeugen. ($[x]$ = größte ganze Zahl $\leq x$)

Übung:

Wie lassen sich Zufallsgrößen Z erzeugen mit Werten $1, \dots, r$ und vorgeschriebenen Wahrscheinlichkeiten $p_j := P\{Z = j\}$, $j = 1, \dots, r$. Wie läßt sich ein $M(n; p_1, \dots, p_r)$ -verteilter Zufallsvektor $\vec{Y} = (Y_1, \dots, Y_r)$ erzeugen? ($Y_j :=$ Häufigkeit von $\{Z = j\}$ bei n unabhängigen Wiederholungen von Z)

2. Die Exponentialverteilung ($E(\lambda)$ -Verteilung)

T sei eine zufallsabhängige Wartezeit mit beliebigen positiven Werten, z.B. die Dauer eines Telefonats. $T \sim E(\lambda)$ bedeutet:

$$\begin{aligned} F(t) &= P\{T \leq t\} = 1 - e^{-\lambda t} \quad (t > 0, \text{ Parameter } \lambda > 0) \\ \Rightarrow f(t) &= \lambda e^{-\lambda t} \quad \text{und} \\ \mu = E(T) &= \int_0^{\infty} \lambda t e^{-\lambda t} dt \stackrel{x:=\lambda t}{=} \frac{1}{\lambda} \int_0^{\infty} x e^{-x} dx = \frac{1}{\lambda} \end{aligned}$$

(Allgemein ist $\int_0^{\infty} x^n e^{-x} dx = n!$ für $n \in \mathbb{N}_0$)

Wenn $T \sim E(\lambda)$ ist, wie ist dann die „Restwartezeit“ T_τ über den Zeitpunkt τ hinaus verteilt, wenn $T > \tau$ wird?

$$P\{T > t + \tau | T > \tau\} = \frac{P\{T > t + \tau \text{ und } T > \tau\}}{P\{T > \tau\}} \stackrel{!}{=} \frac{P\{T > t + \tau\}}{P\{T > \tau\}} = \frac{e^{-\lambda(t+\tau)}}{e^{-\lambda\tau}} = e^{-\lambda t}$$

d.h. die Restwartezeit $T - \tau$ unter der Bedingung $\{T > \tau\}$ hat die gleiche Verteilung wie T ! Dies ist die charakteristische Eigenschaft einer Exponentialverteilung. Die Exponentialverteilung ist daher nicht geeignet (oder bestenfalls eine grobe Näherung) zur Beschreibung der zufallsabhängigen Lebensdauer T von Objekten (oder Lebewesen), welche „altern“. Die Lebensdauer eines radioaktiven Atoms bis zum Zerfall wird dagegen sehr genau durch eine Exponentialverteilung beschrieben. Die Halbwertszeit $t_{0,5}$ ist der Median dieser Verteilung. Aus

$$\begin{aligned} e^{-\lambda t_{0,5}} &= \frac{1}{2} \quad \text{folgt} \\ t_{0,5} &= \frac{\ln 2}{\lambda}. \end{aligned}$$

Zur Beschreibung von „Lebensdauern“ ist die Weibull-Verteilung ($W(\alpha; \lambda)$ -Verteilung in der Liste in diesem Abschnitt) geeigneter. Sie geht für $\alpha = 1$ in eine Exponentialverteilung über.

3. Die Normalverteilung ($N(\mu; \sigma^2)$ -Verteilung)

Die Zufallsgröße X mit $\mu = E(X)$ und $\sigma^2 = \text{Var}(X)$ heißt normalverteilt ($X \sim N(\mu; \sigma^2)$), wenn sie die Dichte

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

hat. f hat sein Maximum bei μ , ist symmetrisch (glockenförmig) um μ und hat Wendepunkte bei $\mu \mp \sigma$. Gelegentlich nennt man die durch $f(x)$ gegebene Kurve auch die Gauß'sche Glockenkurve. Die Standardnormalverteilung hat die Parameter $\mu = 0$ und $\sigma = 1$, also die Dichte $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Die VF der $N(0; 1)$ -Verteilung wird oft mit $\Phi(x)$ bezeichnet. Aus $Z \sim N(0; 1)$ folgt $X := \mu + \sigma Z \sim N(\mu; \sigma^2)$ mit der VF

$$F(x) = P\{X = \mu + \sigma Z \leq x\} = P\left\{Z \leq \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

und der obigen Dichte

$$f(x) = \frac{d}{dx} \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Die Bedeutung der Normalverteilung beruht vor allem darauf, dass die Verteilung von standardisierten Summen unabhängiger Zufallsgrößen unter sehr allgemeinen Voraussetzungen mit wachsender Summandenzahl gegen eine $N(0; 1)$ -Verteilung strebt. (sogenannter zentraler Grenzwertsatz, vgl. folgenden Abschnitt)

Häufig gebrauchte Werte der Funktion Φ :

$$\begin{aligned}\Phi(z_{0,05}) &= 0,95 && \Leftrightarrow && z_{0,05} &= 1,645 \\ \Phi(z_{0,025}) &= 0,975 && \Leftrightarrow && z_{0,025} &= 1,960 \\ \Phi(z_{0,01}) &= 0,99 && \Leftrightarrow && z_{0,01} &= 2,326 \\ \Phi(z_{0,005}) &= 0,995 && \Leftrightarrow && z_{0,005} &= 2,576\end{aligned}$$

2.1.5 Charakteristische Funktionen

Definition:

Die charakteristische Funktion einer beliebigen Zufallsgröße X ist durch $\varphi(t) := E(e^{itX})$ definiert für alle $t \in \mathbb{R}$. Also

$$\varphi(t) = \left\{ \begin{array}{ll} \int_{-\infty}^{\infty} e^{itx} f(x) dx, & \text{falls } X \text{ die Dichte } f \text{ hat} \\ \sum_j e^{itx_j} p_j, & \text{falls } X \text{ diskret ist mit } p_j = P\{X = x_j\} \end{array} \right\} = \int_{-\infty}^{\infty} e^{itx} dF(x).^{IV}$$

Für stetig verteiltes X mit der Dichte f ist somit $\varphi(t)$ gleich der Fouriertransformation $\hat{f}(t)$ der Dichte $f(x)$.

Für Zufallsgrößen mit Werten in \mathbb{N}_0 erhält man $\varphi(t)$ aus der erzeugenden Funktion $\psi(z) = E(z^X)$:

$$\varphi(t) = \psi(e^{it}).$$

Zwei verschiedene Verteilungen haben stets verschiedene charakteristische Funktionen, d.h. eine char. Funktion bestimmt die zugehörige Verteilung eindeutig!

Hat die Dichte $f(x)$ eine absolut integrierbare Fouriertransformierte - d.h. $\int_{-\infty}^{\infty} |\hat{f}(t)| dt < \infty$ - so ist $f(x)$ aus $\hat{f}(t)$ durch die Fourier'sche Umkehrformel bestimmt:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(t) e^{-itx} dt$$

$\varphi(t)$ ist immer gleichmäßig stetig auf \mathbb{R} , und es ist stets $\varphi(0) = 1$. Für Dichten gilt:

$$\lim_{|t| \rightarrow \infty} \hat{f}(t) = 0.$$

2.1.6 Einige Eigenschaften charakteristischer Funktionen

1. Hat X die char. Funktion $\varphi(t)$, so hat $Y = a + bX$ die char. Funktion

$$\varphi_Y(t) = E\left(e^{it(a+bX)}\right) = e^{iat} \varphi(bt) \quad (a, b \text{ konstant})$$

2. $\varphi(-t) = \overline{\varphi(t)}$ (konjugiert komplexer Wert $\bar{\varphi}$)

^{IV}Für eine Dichte f mit zugehöriger VF F ist $dF(x) = f(x) dx$. Für diskretes X mit Werten $x_j \in \mathbb{R}$ ist die VF F eine „Treppenfunktion“ mit Sprunghöhen p_j bei x_j . Definiert man:

$$\lim_{\varepsilon \downarrow 0} \int_{x-\varepsilon}^{x+\varepsilon} dF(t) := \left\{ \begin{array}{ll} p_j, & \text{für } x = x_j \\ 0, & \text{sonst} \end{array} \right\}$$

so ist

$$\int_{\mathbb{R}} e^{itx} dF(x) = \sum_j e^{itx_j} p_j$$

(sogenanntes Stieltjes-Integral). Damit hat man eine einheitliche Schreibweise $\varphi(t) = \int_{\mathbb{R}} e^{itx} dF(x)$ für die char. Fkt. φ der VF F .

3. Für um 0 symmetrische Dichten f ($f(-x) = f(x)$) ist

$$\varphi(t) = 2 \int_0^{\infty} \cos(tx) f(x) dx$$

reell. Dies folgt aus der Euler'schen Formel

$$e^{itx} = \cos(tx) + i \sin(tx)$$

und

$$\int_{-\infty}^{\infty} \sin(tx) f(x) dx = 0$$

wegen $\sin(-tx) = -\sin(tx)$.

4. Multiplikatitivität für unabhängige Zufallsgrößen:

$$\varphi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX} \cdot e^{itY}) = E(e^{itX}) \cdot E(e^{itY}) = \varphi_X(t) \cdot \varphi_Y(t)$$

Allgemeiner gilt für eine Summe $S_n = \sum_{j=1}^n X_j$ unabhängiger Zufallsgrößen X_j :

$$\varphi_{S_n}(t) = \prod_{j=1}^n \varphi_{X_j}(t)$$

Sind alle X_j identisch verteilt mit der char. Funktion $\varphi(t)$:

$$\boxed{\varphi_{S_n}(t) = (\varphi(t))^n}$$

5. Existieren die ersten n Momente $\mu_k = E(X^k)$, $k = 1, \dots, n$, so ist $\varphi(t)$ n -mal stetig differenzierbar, und es gilt für die k -te Ableitung:

$$\varphi^{(k)}(t) = E((iX)^k e^{itX})$$

insbesondere also

$$\varphi^{(k)}(0) = i^k \mu_k.$$

Falls alle Momente μ_k , $k \in \mathbb{N}$, existieren, so erhält man die formale Taylorreihe

$$\varphi(t) = E(e^{itX}) = E\left(\sum_{k=0}^{\infty} X^k \frac{(it)^k}{k!}\right) = \sum_{k=0}^{\infty} \mu_k \frac{(it)^k}{k!}$$

deren Konvergenzradius wie üblich zu bestimmen ist.

2.1.7 Laplacetransformierte

Für positive Zufallsgrößen X wird statt der char. Funktion $\varphi(t)$ oft die stets reelle Laplacetransformierte

$$L(s) := E(e^{-sX}) = \varphi(is) = \int_0^{\infty} e^{-sx} dF(x)$$

mit $s \geq 0$ benutzt. Für $s > 0$ ist $L(s)$ beliebig oft differenzierbar mit

$$L^{(k)}(s) = \int_0^{\infty} (-x)^k e^{-sx} dF(x)$$

Existieren die Momente μ_1, \dots, μ_n , so ist

$$L^{(k)}(0) = (-1)^k \mu_k \quad \text{für } k \leq n.$$

Hat X die Laplacetransformierte (L.T.) $L(s)$, so hat $Y = a + bX$ die L.T. $e^{-as} L(bs)$ ($b > 0$). Natürlich gilt auch die Multiplikatitivität der Laplacetransformierten für unabhängige Zufallsgrößen so wie für die entsprechenden charakteristischen Funktionen.

2.1.8 Beispiele für char. Funktionen und Laplacetransformierte

1. $f(x) = \lambda e^{-\lambda x}$, $x > 0$:

$$L(s) = \lambda \int_0^{\infty} e^{-sx} e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(\lambda+s)x} dx = \lambda \left(-\frac{e^{-(\lambda+s)x}}{\lambda+s} \Big|_{x=0}^{\infty} \right) = \frac{\lambda}{\lambda+s} = \frac{1}{1+\frac{s}{\lambda}}$$

$$\Rightarrow \text{char. Funktion } \varphi(t) = L(-it) = \frac{1}{1-\frac{it}{\lambda}}.$$

2. Gamma-Dichte $f(x) = \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}$, $x > 0$, Parameter $\alpha > 0$:

$$L(s) = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-x} e^{-sx} dx = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-(1+s)x} dx$$

$$\stackrel{y:=(1+s)x}{=} \frac{1}{\Gamma(\alpha)} \int_0^{\infty} \frac{y^{\alpha-1}}{(1+s)^{\alpha-1}} e^{-y} \frac{dy}{1+s} = \frac{1}{(1+s)^{\alpha}} \frac{1}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1} e^{-y} dy = \frac{1}{(1+s)^{\alpha}}$$

Nützliche Beziehungen für einen „Skalenfaktor“ λ : Hat X die VF $F(x)$, die Dichte $f(x)$ und char. Funktion $\varphi(t)$ oder die L.T. $L(s)$, so hat $Y = \frac{1}{\lambda}X$ die VF $F(\lambda y)$, die Dichte $\lambda f(\lambda y)$ und $\varphi_Y(t) = \varphi(\frac{t}{\lambda})$ bzw. $L_Y(s) = L(\frac{s}{\lambda})$. ($\lambda > 0$)

3. $X \sim R(-1; 1)$ mit Dichte $f(x) = \frac{1}{2}$ für $-1 < x < 1$:

$$\varphi(t) = 2 \int_0^1 \cos(tx) \cdot \frac{1}{2} dx = \frac{\sin(tx)}{t} \Big|_{x=0}^1 = \frac{\sin(t)}{t}.$$

4. $X \sim N(0; 1)$:

$$\varphi(t) = \exp\left(-\frac{t^2}{2}\right).$$

(Beweis in 2.2)

2.2 Der Zentrale Grenzwertsatz

Oft benötigt man die Verteilung von Mittelwerten oder Summen unabhängiger Zufallsgrößen. Leider ist die exakte Verteilung einer Summe aus den Verteilungen der Summanden - von einigen Ausnahmen abgesehen - nur mit großem Aufwand berechenbar. Für zwei unabhängige Zufallsgrößen X und Y mit den Verteilungsfunktionen F und G und den Dichten f bzw. g erhält man für die VF H der Summe $Z = X + Y$:

$$H(z) = P\{Z = X + Y \leq z\} = \int_{-\infty}^{\infty} P\{x + Y \leq z | X = x\} f(x) dx =$$

$$= \int_{-\infty}^{\infty} P\{Y \leq z - x | X = x\} f(x) dx = \int_{-\infty}^{\infty} G(z - x) f(x) dx$$

($P\{Y \leq z - x | X = x\} = P\{Y \leq z - x\}$ wegen der Unabhängigkeit von X , Y !)

Die Dichte h von Z ist dann:

$$h(z) = \frac{d}{dz} H(z) = \int_{-\infty}^{\infty} \frac{d}{dz} G(z - x) f(x) dx = \int_{-\infty}^{\infty} g(z - x) f(x) dx.$$

Definition:

Die W -Dichte

$$h(z) = \int_{-\infty}^{\infty} g(z-x) f(x) dx = \int_{-\infty}^{\infty} f(z-y) g(y) dy$$

heißt die Faltung der W -Dichten f und g , kurz:

$$h = f * g = g * f.^V$$

Die Dichte von $S_n = \sum_{j=1}^n X_j$ ist (bei vollständig unabhängigen X_j) die n -fache Faltung $f_1 * f_2 * \dots * f_n$ der Dichten f_j der Summanden X_j . Rechnerisch bedeutet dies eine $(n-1)$ -fache Integration! Für die entsprechenden char. Funktionen $\varphi_j(t) = \hat{f}_j(t)$ der X_j gilt die wesentlich einfachere Beziehung:

$$\varphi_{S_n}(t) = \prod_{j=1}^n \varphi_j(t)$$

Damit lassen sich die folgenden Spezialfälle leicht nachweisen: (alle X_j werden stets als vollständig unabhängig vorausgesetzt)

- (1) $X_j \sim P(\mu_j) \Rightarrow \sum_j X_j \sim P\left(\sum_j \mu_j\right)$
- (2) $X_j \sim B(n_j; p) \Rightarrow \sum_j X_j \sim B\left(n = \sum_j n_j; p\right)$
- (3) $X_j \sim G(\alpha_j; \lambda) \Rightarrow \sum_j X_j \sim G\left(\sum_j \alpha_j; \lambda\right)$
- (4) $X_j \sim E(\lambda) \Rightarrow \sum_{j=1}^k X_j \sim G(k; \lambda)^{VI}$
- (5) $X_j \sim N(\mu_j; \sigma_j^2) \Rightarrow \sum_j X_j \sim N\left(\sum_j \mu_j; \sum_j \sigma_j^2\right)$

z.B. Nachweis von (5):

Ist $Z_j \sim N(0; 1)$ mit der char. Funktion $\varphi(t) = E(e^{itZ}) = e^{-\frac{1}{2}t^2}$ (Nachweis: unten), so ist $X_j := \mu_j + \sigma_j Z_j \sim N(\mu_j; \sigma_j^2)$ mit der char. Funktion

$$\begin{aligned} \varphi_j(t) &= E\left(e^{it(\mu_j + \sigma_j Z_j)}\right) = e^{it\mu_j} \varphi(\sigma_j t) = e^{it\mu_j} \cdot e^{-\frac{1}{2}\sigma_j^2 t^2} \\ \varphi_{S_n}(t) &= \prod_j \varphi_j(t) = \prod_j e^{it\mu_j - \frac{1}{2}\sigma_j^2 t^2} = e^{it\sum_j \mu_j - \frac{1}{2}(\sum_j \sigma_j^2)t^2}. \end{aligned}$$

Der letzte Ausdruck ist aber gerade die char. Funktion einer $N\left(\sum_j \mu_j; \sum_j \sigma_j^2\right)$ -Verteilung!

Ebenso findet man bei (3) mit der Laplacetransformierten $\left(1 + \frac{s}{\lambda}\right)^{-\alpha_j}$ einer $G(\alpha_j; \lambda)$ -Verteilung die Laplacetransformierte von $\sum_j X_j$ als

$$\prod_j \left(1 + \frac{s}{\lambda}\right)^{-\alpha_j} = \left(1 + \frac{s}{\lambda}\right)^{-\sum_j \alpha_j}$$

also

$$\sum_j X_j \sim G\left(\sum_j \alpha_j; \lambda\right).$$

(Gleiche Rechnung mit char. Funktion, wenn s durch $-it$ ersetzt wird.)

$\varphi(t) = e^{-\frac{1}{2}t^2}$ ist die char. Funktion der $N(0; 1)$ -Verteilung, d.h. die Fouriertransformierte der Dichte

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

^VDie kommutative Faltungsoperation ist allgemeiner auf ganz $\mathcal{L}^1(\mathbb{R})$ ($:=$ Menge aller auf \mathbb{R} absolut integrierbarer Funktionen) definiert. Für $f, g \in \mathcal{L}^1(\mathbb{R})$ (d.h. $\int_{\mathbb{R}} |f| dx < \infty$ und $\int_{\mathbb{R}} |g| dx < \infty$) ist auch $h = f * g \in \mathcal{L}^1(\mathbb{R})$.

^{VI}Spezialfall von (3), da $E(\lambda)$ -Verteilung (Exponentialverteilung) mit der $G(1; \lambda)$ -Verteilung übereinstimmt.

Beweis:

$$X \sim N(0; 1) \Rightarrow \mu_{2k-1} = E(X^{2k-1}) = \int_{-\infty}^{\infty} x^{2k-1} f(x) dx = 0 \quad \text{für alle } k \in \mathbb{N}$$

Mit der Substitution: $y = \frac{1}{2}x^2 \Rightarrow dy = x dx \Rightarrow x = \sqrt{2y} \Rightarrow dx = \frac{1}{\sqrt{2y}} dy$ ergibt sich:

$$\begin{aligned} \mu_{2k} &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^{2k} e^{-\frac{1}{2}x^2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} (2y)^k e^{-y} (2y)^{-\frac{1}{2}} dy \stackrel{\text{Sub.}}{=} \frac{2^k}{\sqrt{\pi}} \int_0^{\infty} y^{k-\frac{1}{2}} e^{-y} dy = \\ &= \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) = \frac{2^k}{\Gamma\left(\frac{1}{2}\right)} \Gamma\left(k + \frac{1}{2}\right) = \\ &= \frac{2^k}{\Gamma\left(\frac{1}{2}\right)} \left(k - \frac{1}{2}\right) \left(k - \frac{3}{2}\right) \dots \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = (2k-1)(2k-3) \dots \cdot 3 \cdot 1 \end{aligned}$$

Nach erweitern mit $2 \cdot 4 \cdot \dots \cdot 2k = 2^k k!$:

$$\mu_{2k} = \frac{(2k)!}{2^k k!} \quad (k \in \mathbb{N}_0).$$

$$\varphi(t) = E(e^{itX}) = \sum_{k=0}^{\infty} E(X^{2k}) \frac{(it)^{2k}}{(2k)!} = \sum_{k=0}^{\infty} \frac{1}{2^k k!} (-t^2)^k = \sum_{k=0}^{\infty} \frac{1}{k!} \left(-\frac{1}{2}t^2\right)^k \stackrel{(\text{Exp.reihe!})}{=} e^{-\frac{1}{2}t^2}.$$

Zur Bestimmung der evtl. „Grenzverteilung“ einer Folge von Verteilungen (hier insbesondere der Verteilungen von standardisierten Summen von unabhängigen Zufallsgrößen) benötigen wir folgenden Konvergenzbegriff:

Definition:

Eine Folge $(F_n)_{n \in \mathbb{N}}$ von Verteilungsfunktionen auf \mathbb{R} heißt schwach konvergent gegen eine Verteilungsfunktion F , wenn $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ für alle Stetigkeitsstellen x von F . (Also für alle $x \in \mathbb{R}$, wenn F stetig auf ganz \mathbb{R} ist.)

Eine Folge $(X_n)_{n \in \mathbb{N}}$ heißt konvergent nach Verteilung („in distribution“) gegen eine Zufallsgröße X , wenn die Folge $(F_n)_{n \in \mathbb{N}}$ ihrer Verteilungsfunktionen schwach gegen die Verteilungsfunktion F von X konvergiert, kurz: $X_n \xrightarrow{d} X$.

(Die Werte von X_n müssen also gar nicht konvergieren, sondern nur die entsprechenden Verteilungsfunktionen!)

Bemerkungen:

$$\begin{aligned} X_n \xrightarrow{p} X &\Rightarrow X_n \xrightarrow{d} X \\ X_n \xrightarrow{d} c &\Rightarrow X_n \xrightarrow{p} c \end{aligned}$$

d.h. stochastische Konvergenz impliziert Konvergenz nach Verteilung. Für eine Konstante $X = c$ gilt auch die Umkehr!

Nützlich ist oft: $X_n \xrightarrow{d} a, Y_n \xrightarrow{d} b, Z_n \xrightarrow{d} Z$ mit Konstanten $a, b \Rightarrow X_n + Y_n, Z_n \xrightarrow{d} a + b, Z$.

Man wird erwarten, dass die eindeutige Beziehung zwischen Verteilungsfunktionen F und ihren charakteristischen Funktionen $\varphi(t) = \int_{\mathbb{R}} e^{itx} dF(x)$ „stetig“ im folgenden Sinn ist: „ $\varphi_n \rightarrow \varphi \Leftrightarrow F_n \rightarrow F$ “, was gleich durch den unten folgenden Satz 1 präzisiert wird.

Konvergiert z.B. eine Folge (f_n) von Dichten „im Mittel“ gegen eine Dichte f , d.h.

$$\|f_n - f\| := \int_{\mathbb{R}} |f_n(x) - f(x)| dx \rightarrow 0$$

so gilt offenbar für die zugehörigen char. Funktionen φ_n und φ :

$$|\varphi_n(t) - \varphi(t)| = \left| \int_{\mathbb{R}} (f_n(x) - f(x)) e^{itx} dx \right| \leq \int_{-\infty}^{\infty} |f_n(x) - f(x)| dx = \|f_n - f\| \quad \text{für alle } t \in \mathbb{R}$$

da $|e^{itx}| = 1$, also gleichmäßige Konvergenz der Folge (φ_n) gegen φ auf \mathbb{R} . Außerdem ist

$$|F_n(x) - F(x)| \leq \int_{-\infty}^x |f_n(u) - f(u)| du \leq \|f_n - f\|$$

also

$$F_n(x) \rightarrow F(x) \quad \text{für alle } x \in \mathbb{R}.$$

Allgemeiner gilt folgender wichtiger Satz über die schwache Konvergenz von Verteilungen:

Satz 1:

Ist $(\varphi_n)_{n \in \mathbb{N}}$ eine Folge char. Funktionen von Zufallsgrößen X_n und gilt:

$$\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t) \quad \text{für alle } t \in \mathbb{R}$$

mit einer bei 0 stetigen Funktion φ , so ist auch φ die char. Funktion einer Zufallsgröße X , und es gilt $X_n \xrightarrow{d} X$, also auch die schwache Konvergenz $F_n \rightarrow F$ der entsprechenden Verteilungsfunktionen.

Zum Nachweis von Verteilungskonvergenz wird man daher oft char. Funktionen heranziehen oder bei ausschließlich positiven Zufallsgrößen deren Laplacetransformierte, für die ein analoger Satz gilt.

Der zentrale Grenzwertsatz für Folgen unabhängiger identisch verteilter Zufallsgrößen X_j mit $E(X_j) = \mu$, $\text{Var}(X_j) = \sigma^2$:

Satz 2:

Für die standardisierten Summen

$$S_n^* := \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{\sum_{j=1}^n (X_j - \mu)}{\sigma \sqrt{n}}$$

gilt:

$$\lim_{n \rightarrow \infty} P\{S_n^* \leq z\} = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}x^2} dx \quad \text{für alle } z \in \mathbb{R}.$$

Praktisch also: Für hinreichend große n : $P\{S_n^* \leq z\} \approx \Phi(z)$.

Beweis:

Ist $\varphi(t)$ die char. Funktion von $X_j - \mu$, also $\varphi'(0) = 0$ und $\frac{1}{2}\varphi''(0) = -\frac{1}{2}\sigma^2 t^2$, so gilt für $t \rightarrow 0$ wegen der Taylorformel:

$$\varphi(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2)$$

und folglich

$$\varphi_{S_n^*}(t) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right)^n = \left(1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right) \right)^n \rightarrow e^{-\frac{1}{2}t^2} \quad \text{für jedes feste } t \in \mathbb{R} \text{ und } n \rightarrow \infty.$$

(Wegen $\lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n = e^x$ für jedes x .)

Da $e^{-\frac{1}{2}t^2}$ die char. Funktion der $N(0; 1)$ -Verteilung ist, folgt Satz 2 aus Satz 1.

Anwendung auf Mittelwerte $\bar{X} = \frac{1}{n}S_n$ von i.i.d.r.v. X_j :

Wegen

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

ist $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0; 1)$ -verteilt, bzw. \bar{X} für $n \rightarrow \infty$ asymptotisch $N(\mu; \frac{\sigma^2}{n})$ -verteilt.

Zur Genauigkeit der Approximation der VF F_n^* von S_n^* durch Φ :

Existieren das 3. und 4. Moment μ_3 und μ_4 von X_j , so gilt mit der Schiefe $\gamma = E\left(\frac{X_j - \mu}{\sigma}\right)^3$:

$$F_n^*(x) - \Phi(x) \stackrel{\text{VII}}{\simeq} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \left(\frac{\gamma}{6}(1-x^2)\frac{1}{\sqrt{n}} + R(x)\frac{1}{n} \right) \quad \text{für } n \rightarrow \infty,$$

wobei $R(x)$ ein noch von σ^2, μ_3, μ_4 abhängiges Polynom vom Grad 3 ist. Hat also X_j bereits eine symmetrische Verteilung, so ist $\gamma = 0$ und $F_n^*(x) - \Phi(x) = O\left(\frac{1}{n}\right)$. Die Approximationen sind umso besser, je mehr die Dichte von X_j bereits einer Normaldichte ähnelt.

Beispiel:

Approximation der $B(n; p)$ -Verteilung durch die Normalverteilung:

$$S_n \sim B(n; p) \quad \Rightarrow \quad S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} \approx N(0; 1) - \text{verteilt}$$

für hinreichend große n , da S_n als eine Summe von n i.i.d. Indikatorgrößen X_j mit $\mu = E(X_j) = p$ und $\sigma^2 = \text{Var}(X_j) = p(1-p)$ aufgefaßt werden kann.

„Faustregel“ zur Brauchbarkeit dieser Approximation:

$$\text{mindestens } np(1-p) \geq 10 \quad \left(\text{z.B. } n \geq 40 \text{ für } p = \frac{1}{2} \right)$$

z.B. ergibt sich mit $n = 100, p = \frac{1}{2}, E(S_{100}) = 50$ und $\text{Var}(S_{100}) = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 5$:

(a) exakte Berechnung von

$$P\{S_{100} \leq 55\} = \sum_{j=0}^{55} \binom{100}{j} \left(\frac{1}{2}\right)^j \left(1 - \frac{1}{2}\right)^{100-j} = \frac{1}{2^{100}} \sum_{j=0}^{55} \binom{100}{j} = 0,864373\dots$$

(b) Näherung durch die Normalapproximation:

$$P\{S_{100} \leq 55\} = P\left\{S_{100}^* = \frac{S_{100} - 50}{5} \leq \frac{55 - 50}{5} = 1\right\} \approx \Phi(1) = 0,841345\dots$$

Warum ist diese Approximation so schlecht?

Da die $B(n; p)$ -Verteilung keine stetige Verteilung ist, sollte noch eine sogenannte „Stetigkeitskorrektur“ vorgenommen werden. Betrachte hierzu die beigefügte grafische Darstellung der $B(100; \frac{1}{2})$ -Verteilung. In der Originalskala stehen die möglichen Werte j unter den Mitten der jeweiligen Säulenflächen für $p_j = P\{S_{100} = j\}$.

^{VII}Die Symbole „ \simeq “ („asymptotisch gleich“), „ o “ („klein o “) und „ O “ („groß O “):

$$\begin{aligned} f(x) \simeq g(x) \text{ für } x \rightarrow x_0 &\Leftrightarrow \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 1 \\ f(x) = o(g(x)) \text{ für } x \rightarrow x_0 &\Leftrightarrow \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0 \\ f(x) = O(g(x)) \text{ für } x \rightarrow x_0 &\Leftrightarrow \frac{f(x)}{g(x)} \text{ bleibt beschränkt für } x \rightarrow x_0 \end{aligned}$$

$P\{S_{100} \leq 55\}$ bedeutet Addition bis zur 55. Säule einschließlich, also die „Säulenfläche“ bis zum rechten Rand der 55. Säule. Der j entsprechende standardisierte Wert ist

$$x_j = \frac{j - E(S_{100})}{\sqrt{\text{Var}(S_{100})}} = \frac{j - 50}{5}.$$

Die Säulenfläche wird approximiert durch die Fläche unter der Gauß'schen „Glockenkurve“ bis zu der Stelle x , die in der Mitte zwischen $\frac{55-50}{5}$ und $\frac{56-50}{5}$ liegt, also bis

$$x_{55,5} = \frac{55,5 - 50}{5} = 1,1$$

Mit dieser Stetigkeitskorrektur erhält man die wesentlich bessere Näherung:

$$P\{S_{100} \leq 55\} \approx \Phi(x_{55,5}) = \Phi(1,1) = 0,864334\dots$$

Entsprechend würde z.B. $P\{S_{100} \geq 60\}$ durch

$$1 - \Phi(x_{59,5}) = \Phi(-x_{59,5}) = \Phi\left(-\frac{59,5 - 50}{5}\right) = \Phi(-1,9) \approx 0,028717\dots$$

approximiert.

Der zentrale Grenzwertsatz gilt auch unter gewissen Voraussetzungen für Folgen $(X_j)_{j \in \mathbb{N}}$ unabhängiger nicht identisch verteilter Zufallsgrößen. Haben die X_j noch endliche dritte Momente, so gilt mit

$$\mu_j = E(X_j)$$

$$\sigma_j^2 = \text{Var}(X_j)$$

$$\tau_n^2 = \text{Var}(S_n) = \text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \sigma_j^2$$

unter der Ljapunav-Bedingung: $\frac{1}{\tau_n^3} \sum_{j=1}^n E(|X_j - \mu_j|^3) \rightarrow 0$ für $n \rightarrow \infty$:^{VIII}

$$\lim_{n \rightarrow \infty} P\left\{S_n^* = \frac{\sum_{j=1}^n (X_j - \mu_j)}{\tau_n} \leq z\right\} = \Phi(z) \quad \text{für alle } z \in \mathbb{R}$$

Sogar notwendige und hinreichende Bedingungen sind im Falle unabhängiger X_j bekannt. Sogar für nicht unabhängige X_j kann der zentrale Grenzwertsatz unter gewissen Bedingungen für geeignet standardisierte Summen S_n^* noch gelten, woran noch immer geforscht wird.

2.3 Bivariate und multivariate stetige Verteilungen

Definition:

Ein Zufallsvektor

$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_q \end{pmatrix} \quad \left(\text{hier auch geschrieben als: } \vec{X} = (X_1, \dots, X_q)\right)$$

mit q zufallsabhängigen Koordinaten X_j heißt stetig verteilt, wenn die gemeinsame Verteilungsfunktion

$$F(x_1, \dots, x_q) := P\{X_1 \leq x_1, \dots, X_q \leq x_q\}$$

stetig auf \mathbb{R}^q ist. Falls F nach allen x_j differenzierbar ist, heißt

$$f(x_1, \dots, x_q) := \frac{\partial^q}{\partial x_1 \cdots \partial x_q} F(x_1, \dots, x_q)$$

^{VIII}Dies ist z.B. erfüllt, wenn $E\left|\frac{X_j - \mu_j}{\sigma_j}\right|^3 \leq C$ mit einer von j unabhängigen Konstanten C und wenn $\frac{1}{\tau_n} \max_{1 \leq j \leq n} \sigma_j \rightarrow 0$ für $n \rightarrow \infty$.

die gemeinsame W -Dichte der Zufallsgrößen X_j .

Es ist dann

$$P\left(\bigcap_{j=1}^q \{a_j < X_j < b_j\}\right) = \int_{a_1}^{b_1} \dots \int_{a_q}^{b_q} f(x_1, \dots, x_q) dx_1 \dots dx_q$$

und

$$F(x_1, \dots, x_q) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_q} f(t_1, \dots, t_q) dt_1 \dots dt_q$$

Anschaulich für $q = 2$:

Man denke sich f als Höhe über dem Punkt $(x_1, x_2) \in \mathbb{R}^2$. $P\{a_1 < X_1 < b_1, a_2 < X_2 < b_2\}$ ist dann die „Wahrscheinlichkeitsmasse“

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_1 dx_2$$

über dem Rechteck $(a_1; b_1) \times (a_2; b_2)$.

ZEICHNUNG!!

Die früheren Definitionen für Kovarianz und Korrelation können für stetige Zufallsgrößen beibehalten werden:

Mit $\mu_j = E(X_j)$, $\sigma_j^2 = \text{Var}(X_j)$, $i = 1, \dots, q$ ist

$$\sigma_{ij} = \text{Kov}(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j) = E(X_i X_j) - \mu_i \mu_j = \rho_{ij} \sigma_i \sigma_j$$

mit dem Korrelationskoeffizienten

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} = E\left(\frac{X_i - \mu_i}{\sigma_i} \cdot \frac{X_j - \mu_j}{\sigma_j}\right).$$

Stets ist $-1 \leq \rho_{ij} \leq 1$.

Speziell ist $\sigma_{jj} = \sigma_j^2$ und $\rho_{jj} = 1$.

Vollständige Unabhängigkeit der X_j , $j = 1, \dots, q$ ist äquivalent zur vollständigen Unabhängigkeit der Ereignisse $\{X_j \leq x_j\}$ für alle Werte $x_j \in \mathbb{R}$, somit äquivalent zu

$$F(x_1, \dots, x_q) = F_1(x_1) \cdot \dots \cdot F_q(x_q) \quad \text{für alle } x_j,$$

wobei die F_j die Verteilungsfunktionen der einzelnen Zufallsgrößen X_j sind. Entsprechend ist dann die gemeinsame Dichte

$$f(x_1, \dots, x_q) = f_1(x_1) \cdot \dots \cdot f_q(x_q).$$

2.3.1 Marginalverteilungen („Randverteilungen“)

$\vec{Z} = (X, Y)$ habe die gemeinsame W -Dichte $f(x, y)$. Die Marginalverteilungen von \vec{Z} sind die Verteilungen von X bzw. von Y .

Verteilungsfunktion G von X :

$$G(x) = P\{X \leq x\} = P\{X \leq x, Y < \infty\} = \int_0^x \left(\int_{-\infty}^{\infty} f(u, y) dy \right) du \quad \Rightarrow \quad \text{Marginaldichte von } X :$$

$$g(x) = \frac{d}{dx} G(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Entsprechend für Y :

$$H(y) = P\{Y \leq y\} = P\{X < \infty, Y \leq y\} = \int_0^y \left(\int_{-\infty}^{\infty} f(x, v) dx \right) dv \Rightarrow \text{Marginaldichte von } Y :$$

$$h(y) = \frac{d}{dy} H(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Eine andere Bezeichnung für die Marginaldichten von X bzw. Y :

$$f(x, \cdot) := \int_{-\infty}^{\infty} f(x, y) dy \quad \text{mit VF } F(x, \cdot) = P\{X \leq x\}$$

$$f(\cdot, y) := \int_{-\infty}^{\infty} f(x, y) dx \quad \text{mit VF } F(\cdot, y) = P\{Y \leq y\}$$

(vgl. frühere Schreibweise $p_{i, \cdot}$, $p_{\cdot, j}$)

Der Punkt steht also für Integration über die weggelassene Variable. Allgemein sind die Marginalverteilungen eines Zufallsvektors \vec{X} die gemeinsamen Verteilungen echter Teilmengen seiner Koordinaten. Hat \vec{X} die Dichte $f(\vec{x}) = f(x_1, \dots, x_q)$, so hat z.B. (X_1, \dots, X_k) mit $k < q$ die Marginaldichte

$$g(x_1, \dots, x_k) = f(x_1, \dots, x_k, \dots) = \int_{\mathbb{R}^{q-k}} f(x_1, \dots, x_q) dx_{k+1} \dots dx_q$$

2.3.2 Bedingte Verteilungen und bedingte Dichten

Die gemeinsame Dichte $f(x, y)$ zweier Merkmale (X, Y) beschreibt vollständig das Verhalten beider Merkmale zusammen. Oft will man das Verhalten von Y bei einem gegebenen festen Wert x von X beschreiben.^{IX} Diese sogenannte bedingte Verteilung von Y unter der Bedingung $X = x$, kurz von $Y|x$ hängt bei abhängigen Merkmalen vom x -Wert ab. Gesucht ist die bedingte Dichte $f(y|x)$ von $Y|x$. Es ist

$$\begin{aligned} P\{y \leq Y \leq y + \Delta y | x \leq X \leq x + \Delta x\} &= \frac{P\{y \leq Y \leq y + \Delta y, x \leq X \leq x + \Delta x\}}{P\{x \leq X \leq x + \Delta x\}} \stackrel{\approx}{=} \text{ („asymptotisch =“)} \\ &\approx \frac{f(x, y) \Delta x \Delta y}{f(x, \cdot) \Delta x} = \frac{f(x, y)}{f(x, \cdot)} \Delta y \quad \text{für } \Delta x, \Delta y \rightarrow 0. \end{aligned}$$

Durch den Grenzübergang $\Delta x, \Delta y \rightarrow 0$ erhält man die bedingte Dichte $f(y|x) = \frac{f(x, y)}{f(x, \cdot)}$. Analog: $f(x|y) = \frac{f(x, y)}{f(\cdot, y)}$.

In Analogie zum eindimensionalen („univariaten“) Fall, wird die charakteristische Funktion eines Zufallsvektors $\vec{X} = (X_1, \dots, X_q)$ durch

$$\varphi_{\vec{X}}(\vec{t}) := E\left(e^{i \vec{t}' \vec{X}}\right)$$

definiert. (\vec{t} als Zeilenvektor (t_1, \dots, t_q) , \vec{X} als Spaltenvektor, also $\vec{t}' \vec{X} = \text{Skalarprodukt } \vec{t} \bullet \vec{X} = \sum_{j=1}^q t_j X_j$.)

Ausführlicher geschrieben:

$$\begin{aligned} \varphi_{\vec{X}}(\vec{t}) &= \varphi_{\vec{X}}(t_1, \dots, t_q) = E\left(\exp\left(i \sum_{j=1}^q t_j X_j\right)\right) = \int_{\mathbb{R}^q} \exp\left(i \sum_{j=1}^q t_j X_j\right) dF(x_1, \dots, x_q) = \\ &= \int_{\mathbb{R}^q} \exp\left(i \sum_{j=1}^q t_j X_j\right) f(x_1, \dots, x_q) dx_1 \dots dx_q \end{aligned}$$

^{IX}Bsp.: In einer großen Personengesamtheit (z.B. deutsche Erwachsene) sei $Y := \text{Gewicht}$ und $X := \text{Körperlänge}$. Die bedingte Verteilung von Y unter $X = x = 180 \text{ cm}$ beschreibt dann die Gewichtsverteilung der Personen mit 180 cm Körperlänge.

Für vollständig unabhängige X_j hat man wegen

$$e^{i \sum_j t_j X_j} = \prod_j e^{i t_j X_j} : \quad \varphi_{\vec{X}}(\vec{t}) = \prod_{j=1}^q \varphi_j(t_j)$$

mit den charakteristischen Funktionen

$$\varphi_j(t_j) = E(e^{i t_j X_j}) = \int_{\mathbb{R}} e^{i t_j x} dF_j(x)$$

der einzelnen Zufallsgrößen X_j .

Wieder ist die Beziehung zwischen der VF F von \vec{X} und der charakteristischen Funktion φ von \vec{X} bijektiv (umkehrbar eindeutig). Für positive X_j wird oft die entsprechende Laplacetransformierte $L_{\vec{X}}(\vec{s}) = E(e^{-\vec{s}' \vec{X}}) = \varphi_{\vec{X}}(\vec{s})$ verwendet. Schwierige Verteilungen werden oft zunächst durch ihre charakteristische Funktion $\varphi(\vec{t})$ bestimmt. Aus dieser versucht man dann durch Umkehr die VF $F(\vec{x})$ bzw. die Dichte $f(\vec{x})$ zu bestimmen.

2.3.3 Die multivariate Normalverteilung $N(\vec{\mu}; \Sigma)$ -Verteilung

Der Zufallsvektor $\vec{X} = (X_1, \dots, X_q)$ habe nur $N(\mu_j; \sigma_j^2)$ -verteilte Koordinaten X_j . Ferner sei $\sigma_{ij} = \text{Kov}(X_i, X_j)$. (also $\sigma_{jj} = \sigma_j^2$) Die Varianzen σ_{jj} und Kovarianzen werden in der Kovarianzmatrix $\Sigma = (\sigma_{ij})$ zusammengefaßt und die μ_j im Vektor $\vec{\mu}$.

Definition:

Der obige Zufallsvektor \vec{X} hat eine q -variate (oder q -dimensionale) Normalverteilung, wenn er die gemeinsame Dichte

$$f(\vec{x}) = \frac{1}{(2\pi)^{q/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\vec{X} - \vec{\mu})' \Sigma^{-1} (\vec{X} - \vec{\mu})\right) \quad (\vec{x} \in \mathbb{R}^q)^X$$

hat.

Kurz: $\vec{X} \sim N_q(\vec{\mu}; \Sigma)$.

Diese Verteilung ist also durch die Parameter $\vec{\mu}$ und Σ vollständig bestimmt. Die Bedeutung der $N_q(\vec{\mu}; \Sigma)$ -Verteilung beruht einmal auf ihrer häufigen Eignung als Näherung zur Beschreibung des gemeinsamen Verhaltens von q Zufallsgrößen X_j vor allem aber auch auf dem multivariaten zentralen Grenzwertsatz:

Satz:

Ist $(\vec{X}_k)_{k \in \mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvektoren \vec{X}_k mit Erwartungsvektor $\vec{\mu}$ und Kovarianzmatrix Σ , so sind die Summen $\frac{1}{\sqrt{n}} \sum_{k=1}^n (\vec{X}_k - \vec{\mu})$ asymptotisch $N(\vec{0}; \Sigma)$ -verteilt für $n \rightarrow \infty$.

Anwendung auf ein Stichprobenmittel $\vec{M}_n := \frac{1}{n} \sum_{k=1}^n \vec{X}_k$:

Für nicht zu kleine n ist $\vec{M}_n \approx N(\vec{\mu}; \frac{1}{n} \Sigma)$ -verteilt!

2.3.4 Die wichtigsten Eigenschaften der $N_q(\vec{\mu}; \Sigma)$ -Verteilung

1. Alle k -dimensionalen Marginalverteilungen ($k < q$) sind wieder normalverteilt mit der entsprechenden Submatrix von Σ als Kovarianzmatrix.
2. Die Projektion von \vec{X} auf irgendeine Gerade in \mathbb{R}^q hat eine eindimensionale Normalverteilung. Für $X_a := \vec{a}' \vec{X} = \sum_{j=1}^q a_j X_j$ gilt:

$$\mu_a := E(X_a) = \sum_{j=1}^q a_j \mu_j = \vec{a}' \vec{\mu}$$

^xvgl. mit dem eindimensionalen Fall ($q = 1$):

$$f(x) = \frac{1}{(2\pi)^{1/2} \sqrt{\sigma^2}} \exp\left(-\frac{1}{2}(x - \mu) \frac{1}{\sigma^2} (x - \mu)\right)$$

$$\sigma_a^2 := \text{Var}(X_a) = \sum_{i=1}^q \sum_{j=1}^q a_i a_j \sigma_{ij} = \vec{a}' \Sigma \vec{a}$$

(quadratische Form von \vec{a}) und es ist $X_a \sim N(\mu_a; \sigma_a^2)$.

3. Jede q -dimensionale Normalverteilung kann durch lineare Transformation aus q unabhängigen $N(0; 1)$ -verteilten Zufallsgrößen $\vec{Z} = (Z_1, \dots, Z_q)$ erzeugt werden. $\vec{X} = A \vec{Z} + \vec{\mu}$ mit einer festen regulären $q \times q$ -Matrix A ist $N_q(\vec{\mu}; \Sigma)$ -verteilt mit $\Sigma = A A'$. (A' = Transponierte von A)
Auf diese Art kann man $N(\vec{\mu}; \Sigma)$ -verteilte Zufallsvektoren erzeugen (simulieren).
4. $\vec{X} \sim N_q(\vec{\mu}; \Sigma) \Rightarrow$ Die quadratische Form $(\vec{X} - \vec{\mu})' \Sigma^{-1} (\vec{X} - \vec{\mu})$ ist $\chi^2(q)$ -verteilt! (Anwendung in 3.4)
5. $\vec{X} \sim N_q(\vec{\mu}; \Sigma) \Rightarrow \vec{Y} = B \vec{X} \sim N_q(B \vec{\mu}; B \Sigma B')$ mit beliebiger fester $p \times q$ -Matrix B vom Rang $p \leq q$.
6. $\vec{X} = (X_1, \dots, X_q) \sim N_q(\vec{\mu}; \Sigma)$ mit unabhängigen Koordinaten $X_j \Leftrightarrow$ alle Korrelationen $\rho_{ij} = 0$ und damit alle Kovarianzen $\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j = 0$. Somit

$$\Sigma = \text{Diag}(\sigma_{11}, \dots, \sigma_{qq}) = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_q^2 \end{pmatrix}, \quad \det \Sigma = \sigma_1^2 \cdot \dots \cdot \sigma_q^2,$$

und die gemeinsame Dichte ist

$$f(\vec{x}) = \frac{1}{(2\pi)^{q/2} \sigma_1 \cdot \dots \cdot \sigma_q} \exp\left(-\frac{1}{2} \sum_{j=1}^q \left(\frac{x_j - \mu_j}{\sigma_j}\right)^2\right) = \prod_{j=1}^q \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x_j - \mu_j}{\sigma_j}\right)^2\right)$$

also das Produkt der $N(\mu_j; \sigma_j^2)$ -Dichten der einzelnen X_j .

7. Hat (X, Y) eine bivariate Normalverteilung^{XI} mit

$$\vec{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}, \quad \sigma = \rho \sigma_x \sigma_y,$$

so hat die bedingte Verteilung von $Y|X = x$ eine $N(\mu_y + \frac{\sigma_{xy}}{\sigma_{xx}}(x - \mu_x); \sigma_y^2(1 - \rho^2))$ -Verteilung. Insbesondere hängt die bedingte Erwartung $E(Y|x)$ linear von x ab!

$$E(Y|x) = \alpha + \beta x \text{ mit } \beta = \frac{\sigma_{xy}}{\sigma_{xx}} = \rho \frac{\sigma_y}{\sigma_x} \text{ und } \alpha = \mu_y - \beta \mu_x$$

Diese Beziehung ist die Grundlage der linearen Regression (vgl. 3.3)

8. Charakteristische Funktion φ einer $N_q(\vec{\mu}; \Sigma)$ -Verteilung:

$$\varphi(\vec{t}) = \exp\left(i \vec{\mu}' \vec{t} - \frac{1}{2} \vec{t}' \Sigma \vec{t}\right)$$

ausführlicher:

$$\varphi(t_1, \dots, t_q) = \exp\left(i \sum_{j=1}^q \mu_j t_j - \frac{1}{2} \sum_{j=1}^q \sum_{k=1}^q \sigma_{jk} t_j t_k\right)$$

Speziell für unabhängige X_j , d.h. $\sigma_{jk} = 0$ für $j \neq k$: ($\sigma_{jj} = \sigma_j^2$)

$$\varphi(t_1, \dots, t_q) = \prod_{j=1}^q \exp\left(i \mu_j t_j - \frac{1}{2} \sigma_j^2 t_j^2\right)$$

also das Produkt der charakteristischen Funktionen der einzelnen X_j .

9. \vec{X}, \vec{Y} unabhängig

$$\begin{aligned}\vec{X} &\sim N_q(\vec{\mu}_x; \Sigma_x) \\ \vec{Y} &\sim N_q(\vec{\mu}_y; \Sigma_y) \\ \Rightarrow \vec{X} + \vec{Y} &\sim N_q(\vec{\mu}_x + \vec{\mu}_y; \Sigma_x + \Sigma_y)\end{aligned}$$

Dies kann leicht mit den charakteristischen Funktionen von \vec{X}, \vec{Y} und $\vec{X} + \vec{Y}$ nachgewiesen werden. Speziell gilt für eine Zufallsstichprobe mit n unabhängigen $N_q(\vec{\mu}; \Sigma)$ -verteilten \vec{X}_k , dass der Mittelwert $\vec{M} = \frac{1}{n} \sum_{k=1}^n \vec{X}_k$ die Kovarianzmatrix $\frac{1}{n^2}(n \Sigma) = \frac{1}{n} \Sigma$ hat und $N_q(\vec{\mu}; \frac{1}{n} \Sigma)$ -verteilt ist.

Ein wichtiges Beispiel für den multivariaten zentralen Grenzwertsatz:

$\vec{N} = (N_1, \dots, N_q)$ seien $M(n; p_1, \dots, p_q)$ -verteilte Häufigkeiten N_j für das Eintreten einander ausschließender Ereignisse A_j mit $p_j = P(A_j)$ bei n unabhängigen Ausführungen eines Vorgangs V . Es ist also $\sum_{j=1}^q N_j = n$ und $\sum_{j=1}^q p_j = 1$. Ferner gilt:

$$\begin{aligned}E(N_j) &= n p_j \\ \text{Var}(N_j) &= n p_j (1 - p_j) \\ \text{Kov}(N_i, N_j) &= -n p_i p_j \quad (i \neq j)\end{aligned}$$

Mit der Kovarianzmatrix $\Sigma = (\sigma_{ij})^{\text{XII}}$ (also $\Sigma = \text{Diag}(p_1, \dots, p_q) - \vec{p}\vec{p}'$) wäre der Vektor der Abweichungen $\frac{1}{\sqrt{n}}(\vec{N} - n\vec{p})$ asymptotisch $N_q(\vec{0}; \Sigma)$ -verteilt. Wegen $p_1 + \dots + p_q = 1$ ist aber Σ singulär. Es ist nämlich jede Zeilensumme in Σ gleich

$$\sum_{j=1}^q \sigma_{ij} = p_i - \sum_{j=1}^q p_i p_j = p_i - p_i = 0$$

also $\text{Rang}(\Sigma) < q$. Eine asymptotische Normalverteilung mit regulärer Kovarianzmatrix erhält man durch Weglassen der q -ten überflüssigen Komponente, d.h. mit $r := q - 1$:

$$\vec{X} := \frac{1}{\sqrt{n}} \begin{pmatrix} N_1 - n p_1 \\ \vdots \\ N_r - n p_r \end{pmatrix}$$

ist für $n \rightarrow \infty$ asymptotisch $N_r(\vec{0}; \Sigma_r)$ -verteilt, wobei Σ_r aus obigem Σ durch Weglassen von Zeile und Spalte Nr. q entsteht.

Nach Eigenschaft Nr. 4 ist dann $\vec{\chi}^2 := \vec{X}' \Sigma_r^{-1} \vec{X}$ asymptotisch $\chi^2(r)$ -verteilt. (Chi-Quadrat-Verteilung mit $r = q - 1$ FG'n.) Man rechnet leicht nach, dass $\Sigma_r^{-1} = (\sigma^{ij})$ die Elemente

$$\sigma^{ij} = \begin{cases} \frac{1}{p_i} + \frac{1}{p_q}, & j = i \\ \frac{1}{p_q}, & j \neq i \end{cases}$$

hat. ($\Sigma_r^{-1} \Sigma_r = \text{Einheitsmatrix}$ verifizieren)

^{XI}Mit $\det(\Sigma) = \sigma_x^2 \sigma_y^2 (1 - \rho^2)$,

$$\Sigma^{-1} = \frac{1}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)} \begin{pmatrix} \sigma_y^2 & -\rho \sigma_x \sigma_y \\ -\rho \sigma_x \sigma_y & \sigma_x^2 \end{pmatrix} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x \sigma_y} \\ -\frac{\rho}{\sigma_x \sigma_y} & \frac{1}{\sigma_y^2} \end{pmatrix}$$

ergibt sich aus der allgemeinen Dichteformel:

$$f(x, y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \cdot \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left(\left(\frac{x - \mu_x}{\sigma_x} \right)^2 - 2\rho \left(\frac{x - \mu_x}{\sigma_x} \right) \cdot \left(\frac{y - \mu_y}{\sigma_y} \right) + \left(\frac{y - \mu_y}{\sigma_y} \right)^2 \right) \right\}$$

XII

$$\sigma_{ij} = \begin{cases} p_i(1 - p_i), & j = i \\ -p_i p_j, & j \neq i \end{cases}$$

Damit erhält man für $\tilde{\chi}^2$ den einfachen Ausdruck

$$\tilde{\chi}^2 = \sum_{j=1}^q \frac{(N_j - n p_j)^2}{n p_j} \quad \left(= \sum_{j=1}^q \frac{(N_j - E(N_j))^2}{E(N_j)} \right)$$

Diese Formel ist die Grundlage für die sogenannten χ^2 -Tests für unbekannte Wahrscheinlichkeiten (vgl. 3.4)

Beweis von Eigenschaft Nr. 3:

1. $\vec{Z} = (Z_1, \dots, Z_q)$ mit q unabhängigen $N(0; 1)$ -verteilten Z_j hat die gemeinsame Dichte

$$f(\vec{z}) = f(z_1, \dots, z_q) = (2\pi)^{-q/2} \exp\left(-\frac{1}{2} \sum_{j=1}^q z_j^2\right) = (2\pi)^{-q/2} \exp\left(-\frac{1}{2} \vec{Z}' \vec{Z}\right)$$

(\vec{Z}' : Zeilenvektor, \vec{Z} : Spaltenvektor, $\vec{Z}' \vec{Z} = \text{Skalarprodukt } \vec{Z} \bullet \vec{Z} = \sum_j Z_j^2$.)

Nun sei $\vec{Y} = A \vec{Z}$ mit einer regulären $q \times q$ -Matrix A : Setze $\Sigma := A A'$ mit $A' = A$ -transponiert.

2.

$$\det \Sigma = \det A \cdot \det A' = (\det A)^2$$

und

$$\Sigma^{-1} = A'^{-1} A^{-1} = A^{-1'} A^{-1}$$

3. Bei der linearen Abbildung $\vec{z} = A^{-1} \vec{y}$ gilt für die „Volumenelemente“:

$$dz_1 \dots dz_q = \frac{1}{|\det A|} dy_1 \dots dy_q = \frac{1}{\sqrt{\det \Sigma}} dy_1 \dots dy_q$$

Mit $f(\vec{z})$ und der gesuchten Dichte $g(\vec{y})$ von \vec{Y} muss für zwei entsprechende Bereiche $B \subseteq \mathbb{R}^q$ und $B^* = \{\vec{y} \in \mathbb{R}^q | A^{-1} \vec{y} \in B\}$ gelten:

$$\begin{aligned} P\{\vec{Y} = A \vec{Z} \in B^*\} &= P\{\vec{Z} = A^{-1} \vec{Y} \in B\} \\ &\Leftrightarrow \\ &\int_{B^*} g(\vec{y}) dy_1 \dots dy_q = \int_B f(\vec{z}) dz_1 \dots dz_q \quad \text{(a), (b), (c)} \\ &= \frac{1}{(2\pi)^{q/2} \sqrt{\det \Sigma}} \int_{B^*} \exp\left(-\frac{1}{2} (A^{-1} \vec{y})' (A^{-1} \vec{y})\right) dy_1 \dots dy_q = \\ &= \frac{1}{(2\pi)^{q/2} \sqrt{\det \Sigma}} \int_{B^*} \exp\left(-\frac{1}{2} \vec{y}' \overbrace{A^{-1'} A^{-1}}^{\Sigma^{-1}} \vec{y}\right) dy_1 \dots dy_q \end{aligned}$$

Also hat $\vec{Y} = A \vec{Z}$ die $N_q(\vec{0}, \Sigma)$ -Dichte $g(\vec{y})$ und $\vec{X} = A \vec{Z} + \vec{\mu}$ ist $N_q(\vec{\mu}; \Sigma)$ -verteilt mit $\Sigma = A A'$, was zu zeigen war.

Beweis von Eigenschaft Nr. 4:

Es ist $\vec{Y} := \vec{X} - \vec{\mu} = A \vec{Z}$ mit q unabhängigen $N(0; 1)$ -verteilten Z_j $N_q(\vec{0}; \Sigma)$ -verteilt mit $\Sigma = A A'$, also $\Sigma^{-1} = A'^{-1} A^{-1}$ (Eigenschaft Nr. 3)

$$(\vec{X} - \vec{\mu})' \Sigma^{-1} (\vec{X} - \vec{\mu}) = \vec{Y}' \Sigma^{-1} \vec{Y} = \vec{Z}' A' A'^{-1} A^{-1} A \vec{Z} = \vec{Z}' \vec{Z} = \sum_{j=1}^q Z_j^2 \sim \chi^2(q)$$

Beweis von Eigenschaft Nr. 8:

$X_a := \sum_{j=1}^q a_j X_j \sim N(\mu_a; \sigma_a^2)$ mit $\mu_a = \sum_j a_j \mu_j$, $\sigma_a^2 = \vec{a}' \Sigma \vec{a}$

$$\begin{aligned} \Rightarrow \varphi_{X_a}(t) &= E(e^{it X_a}) = E\left(\exp\left(i \sum_{j=1}^q t a_j X_j\right)\right) \\ &= e^{i \mu_a t} \cdot e^{-\frac{1}{2} \sigma_a^2 t^2} = \exp\left(i \sum_{j=1}^q t a_j \mu_j\right) \cdot \exp\left(-\frac{1}{2} t \vec{a}' \Sigma \vec{a} t\right) \end{aligned}$$

für jeden festen Vektor $\vec{a} \in \mathbb{R}^q$.

Schreibe \vec{t} statt $t \vec{a}$, also t_j statt $t a_j$:

$$\begin{aligned} \Rightarrow \varphi_{\vec{X}}(\vec{t}) &= E(e^{i \vec{t}' \vec{X}}) = E\left(\exp\left(i \sum_{j=1}^q t_j X_j\right)\right) \\ &= \exp\left(i \sum_{j=1}^q t_j \mu_j\right) \cdot \exp\left(-\frac{1}{2} \vec{t}' \Sigma \vec{t}\right) = \exp\left(i \vec{\mu}' \vec{t} - \frac{1}{2} \vec{t}' \Sigma \vec{t}\right) \end{aligned}$$

was zu zeigen war.

3 Statistische Methoden

3.1 Konfidenzintervalle für unbekannte Parameterwerte

3.1.1 $(1 - \alpha)$ -Konfidenzintervall

Wichtige Aufgaben der Statistik sind unter anderem das Schätzen unbekannter Parameterwerte und das Testen von Hypothesen mit Hilfe von Zufallsstichproben. Um eine Vorstellung von der Genauigkeit einer aus der Stichprobe X_1, \dots, X_n berechneten Schätzwertes $\hat{\vartheta} = \hat{\vartheta}(X_1, \dots, X_n)$ für einen unbekanntem Wert des Parameters ϑ der Verteilung von X zu bekommen, will man möglichst ein sogenanntes $(1 - \alpha)$ -Konfidenzintervall (auch: Vertrauensintervall) $(\vartheta_1; \vartheta_2)$ aus der Stichprobe berechnen, welches ϑ mit der Wahrscheinlichkeit $1 - \alpha$ einschließt. Die Grenzen $\vartheta_i = \vartheta_i(X_1, \dots, X_n)$ sind dabei - wie alle aus Zufallsstichproben X_1, \dots, X_n berechneten Größen - zufallsabhängige Werte! Übliche Werte für das sogenannte Konfidenzniveau $1 - \alpha$ (= „statistische Sicherheit“ der Einschließung von ϑ) sind $1 - \alpha = 0,95$ oder $0,99$.

3.1.2 Einige Beispiele

1. Konfidenzintervall für einen unbekanntem Erwartungswert $\mu = E(X)$ bei bekannter Varianz $\sigma^2 = \text{Var}(X)$: (bei unbekannter Varianz: siehe weiter hinten in diesem Abschnitt.)

Nach dem zentralen Grenzwertsatz gilt für das Stichprobenmittel \bar{X} von n i.i.d.r.v. X_j :

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j : \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0; 1) - \text{verteilt}$$

für nicht zu kleine n . Mit der oberen $\frac{\alpha}{2}$ -Schranke $z_{\alpha/2}$ der $N(0; 1)$ -Verteilung folgt:

$$P \left\{ \frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}} \right\} = P \left\{ \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\} \approx 1 - \alpha$$

Damit erhält man das asymptotische $(1 - \alpha)$ -Konfidenzintervall

$$\boxed{(\mu_1; \mu_2) = \left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)} \quad \text{für } \mu = E(X)$$

Mit wachsendem n wird das „nominelle Konfidenzniveau“ $1 - \alpha$ immer besser angenähert. Für normalverteilt X_j wäre das Niveau exakt $= 1 - \alpha$.

Zahlenbeispiel:

$\bar{X} = 12,4$; $n = 16$; $\sigma = 0,8$; $\alpha = 0,05$; $z_{\alpha/2} = 1,96$:

$$(\mu_1; \mu_2) = \left(12,4 - 1,96 \frac{0,8}{\sqrt{16}}; 12,4 + 1,96 \frac{0,8}{\sqrt{16}} \right) = (12,008; 12,792)$$

Zur Deutung eines solchen Konfidenzintervalls:

Bitte nicht $P\{12,008 \leq \mu \leq 12,792\} \approx 0,95$ schreiben! Dies ist unsinnig, da jetzt zwischen den Klammern keine Zufallsgröße mehr steht. Die Aussage „ $12,008 \leq \mu \leq 12,792$ “ ist entweder wahr oder falsch, besitzt aber für sich allein gar keine Wahrscheinlichkeit. Die Situation ist völlig analog zum Wurf einer idealen Münze. Nach dem Wurf - d.h. nach der Realisierung des zufallsabhängigen Merkmals $S =$ „Seite“ - sagt niemand „Mit Wahrscheinlichkeit $\frac{1}{2}$ liegt Wappen oben“ sondern nur „Wappen liegt oben“. Diese Aussage A kann wahr oder falsch sein, und man weiß nicht, was zutrifft, wenn man das Ergebnis nicht sieht. Man weiß aber, dass A bei sehr vielen wiederholten Münzwürfen immer genauer in der Hälfte der Fälle zutrifft. Genauso weiß man, dass die Aussage

A: „Der unbekannte Erwartungswert μ liegt im errechneten Konfidenzintervall“ bei sehr vielen Konfidenzintervallberechnungen mit stets neuen Stichproben des Merkmals X annähernd für $(1 - \alpha) \cdot 100\%$ dieser Intervalle zutrifft. In diesem Sinne quantifizieren wir unser „Vertrauen“ in die Gültigkeit der Aussage A mit dem Wert $1 - \alpha$.

2. Konfidenzintervall für eine unbekannte Wahrscheinlichkeit p

Es sei X die $B(n; p)$ -verteilte Häufigkeit für das Eintreten eines Ereignisses A mit unbekannter Wahrscheinlichkeit $p = P(A)$. Die relative Häufigkeit $\hat{p} := \frac{X}{n}$ ist dann ein zufallsabhängiger Schätzer für p , und nach dem zentralen Grenzwertsatz gilt:

$$\frac{\hat{p} - p}{\sqrt{\text{Var}(\hat{p})}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0; 1) - \text{verteilt für ausreichend große } n.$$

Es gilt daher:

$$P \left\{ \frac{|\hat{p} - p|}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\frac{\alpha}{2}} \right\} \approx 1 - \alpha$$

\Leftrightarrow

$$P \left\{ \hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right\} \approx 1 - \alpha.$$

Hierbei stört, dass die Varianz $\frac{p(1-p)}{n}$ des Schätzers \hat{p} selbst vom unbekanntem p -Wert abhängt. Es werden vier unterschiedlich genaue Konfidenzintervalle für p angegeben:

1. Für sehr große n : Ersetze $p(1-p)$ durch $\hat{p}(1-\hat{p})$:

$\Rightarrow \approx (1 - \alpha)$ -Konfidenzintervall für p mit den Grenzen p_1, p_2 :

$$p_{1/2} = \hat{p} \mp z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

2. Etwas genaueres asymptotisches $(1 - \alpha)$ -Konfidenzintervall:

$$P \left\{ \frac{|\hat{p} - p|}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\frac{\alpha}{2}} \right\} \approx 1 - \alpha \quad \Leftrightarrow \quad P \left\{ |\hat{p} - p|^2 \leq z_{\frac{\alpha}{2}}^2 \frac{p(1-p)}{n} \right\} \approx 1 - \alpha.$$

Um die Menge der p -Werte zu erhalten, welche diese Ungleichung erfüllen (d.h. den Konfidenzbereich), wird die bzgl. p quadratische Gleichung $(\hat{p} - p)^2 = z_{\alpha/2}^2 \frac{p(1-p)}{n}$ gelöst. Ihre Lösungen

$$p_{1/2} = \left(\hat{p} + \frac{z_{\alpha/2}^2}{2n} \mp z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}} \right) / \left(1 + \frac{z_{\alpha/2}^2}{n} \right)$$

sind die Grenzen eines etwas genaueren asymptotischen $(1 - \alpha)$ -Konfidenzintervalles für p . Sie gehen für $n \rightarrow \infty$ in die Grenzen des Intervalls in 1. über.

3. Konfidenzintervall für p mit der arcsin-Transformation:

$$Z := \arcsin \sqrt{\hat{p}} \approx \arcsin p + \frac{1}{2\sqrt{p(1-p)}}(\hat{p} - p)$$

(lineare Taylorapproximation mit $\frac{d}{dp} \arcsin \sqrt{p} = \frac{1}{\sqrt{1-p}} \cdot \frac{1}{2\sqrt{p}}$)
 Z hat annähernd die von p unabhängige Varianz

$$\sigma^2 = \frac{1}{4p(1-p)} \text{Var}(\hat{p} - p) = \frac{1}{4p(1-p)} \cdot \frac{p(1-p)}{n} = \frac{1}{4n},$$

und es gilt:

$$Z \approx N(\mu; \sigma^2) - \text{verteilt} \quad \text{mit } \mu = \arcsin \sqrt{p} \text{ und } \sigma^2 = \frac{1}{4n}$$

$$\Rightarrow \approx (1 - \alpha) - \text{Konfidenzintervall f\u00fcr } \mu : \quad \mu_{1/2} = \arcsin \sqrt{\hat{p}} \mp \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}}$$

Durch R\u00fccktransformation mit der Umkehrfunktion $p = \sin^2 \mu$ erh\u00e4lt man das Konfidenzintervall

$$(p_1; p_2) = (\sin^2 \mu_1; \sin^2 \mu_2)$$

4. Fast exaktes $(1 - \alpha)$ -Konfidenzintervall f\u00fcr p (nach Clopper-Pearson):

Hier wird ein Zusammenhang zwischen der $B(n; p)$ -Verteilung und der $F(m; n)$ -Verteilung ausgen\u00fctzt. Sind $X_{m/2}$ und $Y_{n/2}$ $G(\frac{m}{2}; \frac{1}{2})$ - bzw. $G(\frac{n}{2}; \frac{1}{2})$ -verteilte unabh\u00e4ngige Zufallsgr\u00f6\u00dfen, so hat

$$F := \frac{\frac{1}{m} X_{m/2}}{\frac{1}{n} Y_{n/2}}$$

eine sogenannte F -Verteilung mit den Freiheitsgraden m und n , kurz $F \sim F(m; n)$. Deren $(1 - \alpha)$ -Quantile $F(1 - \alpha; m, n)$ sind in vielen Programmen (z.B. Excel) verf\u00fcgbar. Mit $\hat{p} = \frac{x}{n}$ ist

$$(p_1; p_2) = \left(\frac{1}{1 + \frac{n+1-x}{x} \cdot F(1 - \frac{\alpha}{2}; 2(n+1-x), 2x)}; \frac{1}{1 + \frac{n-x}{x+1} \cdot F(\frac{\alpha}{2}; 2(n-x), 2(x+1))} \right)$$

Zahlenbeispiel: $\hat{p} = 0,3$, $n = 1000$:

Methode 1 : $(p_1; p_2) = (0,2716; 0,3284)$

Methode 2 : $(p_1; p_2) = (0,2724; 0,3291)$

Methode 3 : $(p_1; p_2) = (0,2720; 0,3288)$

Ein mathematisch exaktes Konfidenzniveau $1 - \alpha$ ist f\u00fcr eine diskrete Verteilung im Allgemeinen nicht m\u00f6glich. Bei einer $B(n; p)$ -verteilten Zufallsgr\u00f6\u00dfe X gibt es nur endlich viele Wahrscheinlichkeitswerte $P\{k_1 \leq X \leq k_2\}$ mit $0 \leq k_1 \leq k_2 \leq n$, unter denen $1 - \alpha$ im allgemeinen nicht exakt vorkommt! Die L\u00f6sung Nr. 4. ist die im Rahmen dieser Einschr\u00e4nkung bestm\u00f6gliche Ann\u00e4herung an das nominelle Konfidenzintervall $1 - \alpha$.

Die Normalapproximation f\u00fcr die arcsin-Transformation kann genauer gemacht werden:

$$\arcsin \sqrt{\frac{X + \frac{3}{8}}{n + \frac{3}{4}}} \approx N \left(\arcsin \sqrt{\frac{np + \frac{3}{8}}{n + \frac{3}{4}}}; \frac{1}{4n + 2} \right)$$

(Lohnt sich nur f\u00fcr kleinere n .)

3. Konfidenzintervall f\u00fcr den unbekanntem Wert ρ des Korrelationskoeffizienten zweier Zufallsgr\u00f6\u00dfen X und Y mit einer gemeinsamen Normalverteilung

Bemerkung: F\u00fcr Zufallsgr\u00f6\u00dfen (X, Y) mit einer gemeinsamen (2-dimensionalen) Normalverteilung gilt die \u00c4quivalenz:

$$X, Y \text{ unabh\u00e4ngig} \Leftrightarrow \rho = 0.$$

Zur Sch\u00e4tzung $\hat{\rho} = r$ siehe 1.6.1 auf Seite 26! Die Dichte des zuf\u00e4lligen Sch\u00e4tzers r ist von relativ komplizierter Form und \u00fcberdies vom unbekanntem wahren Wert ρ der Korrelation abh\u00e4ngig. Auch hier hilft eine Transformation mit dem Ziel, eine n\u00e4herungsweise normalverteilte Gr\u00f6\u00dfe zu erhalten:

Die Fisher'sche Z -Transformation:

$$Z := \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = \text{Artanh}(r)^I$$

ist

$$Z \approx N\left(\mu = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right) + \frac{\rho}{2(n-1)}; \sigma^2 = \frac{1}{n-3}\right) - \text{verteilt!}$$

Für hinreichend große n kann $\mu = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right)$ gesetzt werden. Insbesondere gilt für $\rho = 0$ auch $\mu = 0$.

Aus einer Stichprobe mit n Wertepaaren (x_i, y_i) erhält man mit dem Stichprobenkorrelationskoeffizienten r (Taschenrechnertaste!) zunächst das $\approx (1 - \alpha)$ -Konfidenzintervall

$$(\mu_1; \mu_2) = \left(Z - \frac{z_{\alpha/2}}{\sqrt{n-3}}; Z + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) \quad \text{für } \mu = E(Z)$$

und hieraus durch Rücktransformation das $\approx (1 - \alpha)$ -Konfidenzintervall für ρ :

$$(\rho_1; \rho_2) = (\tanh(\mu_1); \tanh(\mu_2)) = \left(\frac{e^{2\mu_1} - 1}{e^{2\mu_1} + 1}; \frac{e^{2\mu_2} - 1}{e^{2\mu_2} + 1}\right)$$

Zahlenbeispiel:

$r = 0,6$, $n = 80$, $\alpha = 0,05$, $z_{\alpha/2} = 1,96$:

$$(\mu_1; \mu_2) = (0,4698; 0,9165)$$

$$(\rho_1; \rho_2) = (0,4380; 0,7242).$$

Für die weiteren Beispiele werden die - nach der Normalverteilung - wichtigsten Verteilungen der elementaren Statistik benötigt:

Def. 1:

Sind Z_1, \dots, Z_n unabhängig $N(0; 1)$ -verteilt, so hat $\chi_n^2 := Z_1^2 + \dots + Z_n^2$ eine Chiquadrat-Verteilung^{II} mit n Freiheitsgraden, kurz:

$$\sum_{j=1}^n Z_j^2 \sim \chi^2(n)$$

(Unterscheide die Zufallsgröße χ_n^2 vom Namen $\chi^2(n)$ der Verteilung!)

Dichte:

$$\frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2}, \quad (x > 0)$$

$$E(\chi_n^2) = n$$

$$\text{Var}(\chi_n^2) = 2n.$$

Diese Verteilung ist offenbar mit der Gammaverteilung $G(\frac{n}{2}; \frac{1}{2})$ identisch (vgl. Tabelle in 2.1.3 auf Seite 35). Ihre Quantile $\chi^2(\alpha, n)$ und $\chi^2(1 - \alpha, n)$ sind tabelliert und in vielen Programmen verfügbar.

Satz 1:

X_1, \dots, X_n unabhängig und identisch $N(\mu; \sigma^2)$ -verteilt \Rightarrow das Streuungsquadrat

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

ist verteilt wie

$$\frac{\sigma^2}{n-1} \chi_{n-1}^2 \quad \left(\Leftrightarrow (n-1) \frac{s^2}{\sigma^2} \sim \chi^2(n-1)\right)$$

^IArea tangens hyperbolicus = Umkehrfunktion zu $\tanh(x) := \frac{e^{2x}-1}{e^{2x}+1}$, $x \in \mathbb{R}$

^{II}griechischen Buchstaben χ („chi“) nicht mit x verwechseln.

(Beweis am Ende des Paragraphen)

Die χ^2 -Verteilung spielt eine wichtige Rolle bei der Analyse beobachteter zufallsabhängiger Häufigkeiten. Eine einfache Anwendung ist:

4. Konfidenzintervall für eine unbekannte Varianz σ^2 einer $N(\mu; \sigma^2)$ -Verteilung

Aus der Stichprobe X_1, \dots, X_n erhält man mit s :

$$P \left\{ \chi^2 \left(\frac{\alpha}{2}; n-1 \right) \leq (n-1) \frac{s^2}{\sigma^2} \leq \chi^2 \left(1 - \frac{\alpha}{2}; n-1 \right) \right\} = 1 - \alpha$$

und durch Umstellen nach σ^2 die Grenzen

$$(\sigma_1^2; \sigma_2^2) = \left(\frac{(n-1)s^2}{\chi^2(1 - \frac{\alpha}{2}; n-1)}; \frac{(n-1)s^2}{\chi^2(\frac{\alpha}{2}; n-1)} \right)$$

für das $(1 - \alpha)$ -Konfidenzintervall für σ^2 . Für nur $\approx N(\mu; \sigma^2)$ -verteilte X_j ist natürlich das Niveau $1 - \alpha$ auch nur als Näherung anzusehen.

Zahlenbeispiel:

Als Maß für die Genauigkeit eines unverzerrten Messverfahrens kann die Standardabweichung σ der zufallsabhängigen Messfehler X_j aufgefaßt werden. Es ergab sich aus einer Stichprobe mit 61 Werten eine Streuung $s = 0,42$ (Maßeinheiten). Mit $\alpha = 0,1$ ergibt sich für σ das 90%-Konfidenzintervall

$$(\sigma_1, \sigma_2) = \left(\sqrt{\frac{60}{79,08}} \cdot 0,42; \sqrt{\frac{60}{43,19}} \cdot 0,42 \right) = (0,3658; 0,4950)$$

Wie man sieht, sind für eine genauere Schätzung von σ oder σ^2 größere Stichproben erforderlich!

Def. 2:

Sind Z_1, \dots, Z_n unabhängig und identisch $N(0; 1)$ -verteilt, so hat

$$t_{n-1} := \frac{Z_1}{\sqrt{\frac{1}{n-1} \sum_{i=2}^n Z_i^2}}$$

eine t -Verteilung mit $\nu = n - 1$ Freiheitsgraden, kurz $t_{n-1} \sim t(n-1)$.
(vgl. Tabelle in 2.1.3 auf Seite 35)

Unterscheide auch hier die Zufallsgröße t_{n-1} vom Namen $t(n-1)$ ^{III} der Verteilung. Die Dichte einer t -Verteilung ist „glockenförmig“ symmetrisch um 0 und in Abhängigkeit vom Freiheitsgrad etwas breiter als die $N(0; 1)$ -Dichte. Für $n \rightarrow \infty$ strebt die $t(n)$ -Dichte gegen die $N(0; 1)$ -Dichte. Praktisch kann man die tabellierten oberen α -Schranken $t(\alpha; n)$ ($= (1-\alpha)$ -Quantile) für $n > 200$ durch die entsprechenden Schranken z_α der $N(0; 1)$ -Verteilung ersetzen.

5. Konfidenzintervall für einen unbekanntem Erwartungswert μ bei unbekannter Varianz σ^2

Es sei X_1, \dots, X_n eine Zufallsstichprobe mit n unabhängigen $N(\mu; \sigma^2)$ -verteilten Zufallswerten X_j . Dann gilt:

$$t_{n-1} = \frac{\bar{X} - \mu}{\sqrt{\hat{\text{Var}}(\bar{X})}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t(n-1).$$

Aus

$$P \left\{ -t \left(\frac{\alpha}{2}; n-1 \right) \leq \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \leq t \left(\frac{\alpha}{2}; n-1 \right) \right\} = 1 - \alpha$$

^{III}Auch: „Student-Verteilung“ nach ihrem Schöpfer W. S. Gosset, der unter dem Pseudonym „Student“ publizierte. Ihre erste Anwendung fand die t -Verteilung in der Qualitätskontrolle.

folgt durch Umstellen nach μ das $(1 - \alpha)$ -Konfidenzintervall

$$\boxed{(\mu_1; \mu_2) = \left(\bar{X} - t\left(\frac{\alpha}{2}; n-1\right) \frac{s}{\sqrt{n}}; \bar{X} + t\left(\frac{\alpha}{2}; n-1\right) \frac{s}{\sqrt{n}} \right)} \quad \text{für } \mu = E(X).$$

Wegen der notwendigen Schätzung von σ durch die Stichprobenstreuung s ist nun die Verteilung des standardisierten Mittelwertes $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ eine t -Verteilung statt einer $N(0; 1)$ -Verteilung. Es sind daher im Konfidenzintervall von Bsp. 1 nur die Schranken $z_{\alpha/2}$ der $N(0; 1)$ -Verteilung durch die Schranken $t(\frac{\alpha}{2}; n-1)$ der $t(n-1)$ -Verteilung zu ersetzen. Glücklicherweise ist die t -Verteilung verhältnismäßig **robust** gegen Abweichungen von der Normalverteilung bei den Stichprobenwerten, so dass auch für nicht normalverteilte Stichprobendaten $1 - \alpha$ noch als Näherung für das Konfidenzniveau angesehen werden darf.

Zahlenbsp.:

wie in Bsp. 1 mit $\bar{X} = 12,4$, $n = 16$, $\alpha = 0,05$ aber mit einer Schätzung $s = 0,86$ aus den 16 Stichprobenwerten, also mit $n - 1 = 15$ Freiheitsgraden (kurz: FG = 15) $t(0,025; 15) = 2,131 \Rightarrow 95\%$ -Konfidenzintervall für μ :

$$\left(12,4 - 2,131 \frac{0,86}{\sqrt{16}}; 12,4 + 2,131 \frac{0,86}{\sqrt{16}} \right) = (11,94; 12,86)$$

Def. 3:

Sind χ_m^2 , χ_n^2 zwei unabhängige $\chi^2(m)$ bzw. $\chi^2(n)$ -verteilte Zufallsgrößen, so hat

$$F_{m,n} := \frac{\chi_m^2/m}{\chi_n^2/n}$$

eine F -Verteilung mit den Freiheitsgraden m und n , kurz: $F_{m,n} \sim F(m, n)$. Die $(1 - \alpha)$ -Quantile sind verfügbar. Sind X_1, \dots, X_m und Y_1, \dots, Y_n zwei unabhängige Stichproben aus einer $N(\mu_x; \sigma^2)$ bzw. $N(\mu_y; \sigma^2)$ -Verteilung (also $\sigma_x^2 = \sigma_y^2 = \sigma^2$), so hat das Verhältnis der Streuungsquadrate

$$F := \frac{s_x^2}{s_y^2} = \frac{\frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2}{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}$$

eine $F(m-1, n-1)$ -Verteilung.

Hat man unbekannte evtl. verschiedene Varianzen σ_x^2 , σ_y^2 , so ist

$$\frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2} \sim F(m-1; n-1)$$

und aus

$$P \left\{ F\left(\frac{\alpha}{2}; m-1, n-1\right) \leq \frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2} \leq F\left(1 - \frac{\alpha}{2}; m-1, n-1\right) \right\} = 1 - \alpha$$

folgt das $(1 - \alpha)$ -Konfidenzintervall für das unbekannte Varianzverhältnis $\frac{\sigma_x^2}{\sigma_y^2}$ mit den Grenzen

$$\boxed{\left(\frac{s_x^2}{s_y^2} \frac{1}{F\left(1 - \frac{\alpha}{2}; m-1, n-1\right)}; \frac{s_x^2}{s_y^2} \frac{1}{F\left(\frac{\alpha}{2}; m-1, n-1\right)} \right)}$$

Tabellen enthalten oft keine unteren Schranken der F -Verteilung. Es ist aber stets

$$F(\alpha; m, n) = \frac{1}{F(1 - \alpha; n, m)}$$

Folgere dies aus der Beziehung $P\{F_{m,n} \leq F(\alpha; m, n)\} = \alpha$ durch Übergang zum offensichtlich $F(n; m)$ -verteilten Kehrwert $\frac{1}{F_{m,n}}$.

Zahlenbsp:

$s_x^2 = 15,5$, $m - 1 = 60$, $s_y^2 = 18,2$, $n - 1 = 120$, $\alpha = 0,1$, $F(0,95; 60; 120) = 1,429$, $F(0,95; 120, 60) = 1,467$ gibt für das unbekannte Varianzverhältnis σ_x^2/σ_y^2 das 90%–Konfidenzintervall

$$\left(\frac{15,5}{18,2} \cdot \frac{1}{1,429}; \frac{15,5}{18,2} \cdot 1,467 \right) = (0,5960; 1,249).$$

Besonders wichtig ist die F –Verteilung für die Varianzanalyse. Diese ist eine allgemeine statistische Methode zur Analyse der Wirkungen einer oder mehrerer Ursachen auf die Verteilung und insbesondere auf die Erwartungswerte eines oder mehrerer Merkmale. Sie wird vor allem für die statistische Analyse von Experimenten im Versuchswesen benötigt. (Z.B wie wirken verschiedene Dosierungen mehrerer Substanzen auf das zu erwartende Wachstum eines Tumors innerhalb eines gewissen Zeitraums?)

Bei allen Beispielen lassen sich auch einseitige $(1 - \alpha)$ –Konfidenzbereiche angeben, wie z.B. nur eine untere $(1 - \alpha)$ –Konfidenzgrenze

$$\bar{X} - t(\alpha; n - 1) \frac{s}{\sqrt{n}} \quad \text{für } \mu = E(X)$$

Zur Verteilung von $\chi_n^2 = \sum_{i=1}^n Z_i^2$ und des Streuungsquadrat $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ von n unabhängigen $N(\mu; \sigma^2)$ –verteilten X_i :

$$Z_1 \sim N(0; 1) \quad \Rightarrow \quad P\{Z_1^2 \leq x\} = P\{|Z_1| \leq \sqrt{x}\} = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1.$$

(Wegen der Symmetrie um 0 ist stets $\Phi(-y) = 1 - \Phi(y)$.)

Also hat Z_1^2 die Dichte

$$f(x) = 2 \frac{d}{dx} \Phi(\sqrt{x}) = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{2\sqrt{x}} e^{-\frac{1}{2}x} = \frac{1}{2^{1/2} \Gamma(\frac{1}{2})} x^{-\frac{1}{2}} e^{-\frac{x}{2}}.$$

Dies ist die Dichte einer $G(\frac{1}{2}; \frac{1}{2})$ –Verteilung (= $\chi^2(1)$ –Verteilung).

Wegen (3) in 2.2 auf Seite 40 hat dann $\sum_{i=1}^n Z_i^2$ eine $G(\frac{n}{2}; \frac{1}{2})$ –Verteilung (= $\chi^2(n)$ –Verteilung).

Zur Verteilung von s^2 :

Der Zufallspunkt $P \in \mathbb{R}^n$ habe den Koordinatenvektor $\vec{Z} = (Z_1, \dots, Z_n)'$ und in einem neuen um den Nullpunkt gedrehten Koordinatensystem den neuen Koordinatenvektor $\vec{Z}^* = (Z_1^*, \dots, Z_n^*)'$,^{IV} wobei das Abstandsquadrat $\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n Z_i^{*2}$ vom Nullpunkt gleich bleibt. Eine Drehung wird rechnerisch durch eine orthogonale Matrix A beschrieben mit zueinander orthogonalen Zeilenvektoren der Länge 1. $\vec{Z}^* = A \vec{Z}$. Wählt man in A als letzte Zeile $(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$, so ist $Z_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i = \sqrt{n} \bar{Z}$ und folglich

$$\chi_{n-1}^2 := \sum_{i=1}^{n-1} Z_i^{*2} = \sum_{i=1}^n Z_i^2 - n \bar{Z}^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 \sim \chi^2(n-1).$$

$X_i \sim N(\mu, \sigma^2) \quad \Rightarrow \quad X_i$ ist verteilt wie $\mu + \sigma Z_i$ und $(X_i - \bar{X})^2$ wie $\sigma^2(Z_i - \bar{Z})^2$. Also ist

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

verteilt wie $\sigma^2 \chi_{n-1}^2$, was zu zeigen war.

^{IV} \vec{Z} und \vec{Z}^* haben die gleiche gemeinsame rotationssymmetrische Dichte

$$\left(\frac{1}{\sqrt{2\pi}} \right)^n \prod_{i=1}^n \exp\left(-\frac{1}{2} Z_i^2\right) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n Z_i^2\right) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n Z_i^{*2}\right)$$

3.2 Testen von Hypothesen

3.2.1 Nullhypothese

Viele Vorstellungen über einen realen Sachverhalt sind bei genauerer Formulierung Vorstellungen über die Werte gewisser Parameter von Wahrscheinlichkeitsverteilungen. Statt Vorstellungen (Meinungen, Vermutungen) sagt man in der Statistik Hypothesen. Ein statistischer Test fällt aufgrund einer oder mehrerer Zufallsstichproben eine Entscheidung über die Glaubwürdigkeit zweier einander ausschließender Hypothesen. Bei Hypothesen über den Wert irgendeines Verteilungsparameters γ - z.B. des Erwartungswertes $\gamma = \mu$ - hat die sogenannte Nullhypothese H_0 die Form

$$H_0 : \gamma = \gamma_0$$

wobei γ_0 ein hypothetischer (d.h. behaupteter oder vermuteter) Wert von γ ist und γ den wahren unbekanntem Parameterwert bezeichnet. Mit der Abweichung $\delta := \gamma - \gamma_0$ wird das Wort „Nullhypothese“ verständlicher. Offenbar bedeutet

$$H_0 : \delta = 0$$

3.2.2 Alternativhypothese

Demgegenüber besagt eine einseitige Alternativhypothese

$$\begin{aligned} H_1^+ : \gamma > \gamma_0 &\Leftrightarrow \delta > 0 \quad \text{oder gegebenenfalls auch} \\ H_1^- : \gamma < \gamma_0 &\Leftrightarrow \delta < 0 \end{aligned}$$

Die zweiseitige Alternativhypothese ist die Negation von H_0 :

$$H_1 : \gamma \neq \gamma_0 \Leftrightarrow \delta \neq 0$$

Die einseitige Alternative H_1^+ ist sinnvoll, wenn man weiß, dass im Fall $\gamma \neq \gamma_0$ nur $\gamma > \gamma_0$ in Frage kommt.

Einfaches Beispiel:

Die durchschnittliche „Lebensdauer“ eines technischen Massenartikels - z.B. die Brenndauer einer Glühlampe eines bestimmten Fabrikats - hat einen bestimmten Wert, z.B. 1150 Stunden. Die Lebensdauer X eines zufällig herausgegriffenen Exemplars dieses Artikels (Brenndauer X einer einzelnen Glühlampe) hängt von vielen (zukünftigen) unbekanntem Einflüssen ab und verhält sich daher wie eine Zufallsgröße. Die obige Hypothese ist also eine Nullhypothese der Form:

$$H_0 : \mu = E(X) = \mu_0 = 1150 \text{ [h]}$$

Eine einseitige Alternativhypothese (z.B. für Werbezwecke) könnte

$$H_1 : \mu > \mu_0$$

lauten.

Vor einem Test sollte sowohl H_0 als auch die Alternativhypothese klar formuliert werden, weil das Testverfahren von beiden Hypothesen abhängt. Bei jedem statistischen Test sind zwei Fehlentscheidungen möglich:

Fehler 1. Art:

H_0 wird abgelehnt (d.h. als „unglaublich“ eingestuft), obwohl H_0 wahr ist.

Fehler 2. Art:

H_0 wird nicht abgelehnt, obwohl die Alternativhypothese wahr ist.

Jeder statistische Test kontrolliert die Wahrscheinlichkeit p für einen Fehler 1. Art, d.h. es wird durch die Entscheidungsregel garantiert, dass $p \leq \alpha$ mit einem vorgegebenen kleinen Wert α , meist $\alpha = 0,01$ oder $\alpha = 0,05$. Diese maximale Irrtumswahrscheinlichkeit α für einen Fehler 1. Art heißt auch das Niveau α des Tests.

Die Wahrscheinlichkeit β für einen Fehler 2. Art ist eine Funktion, die im wesentlichen davon abhängt, wie stark sich der wahre Parameter γ vom hypothetischen Wert γ_0 unterscheidet.

$1 - \beta$ ist dann die Wahrscheinlichkeit einer Entscheidung für die Alternativhypothese und heißt die Power oder Gütefunktion des Tests. Sie kann als „Empfindlichkeit“ des Tests aufgefaßt werden, Abweichungen von der Nullhypothese zu „entdecken“. Bei festem Niveau α sollte die Power („Empfindlichkeit“) möglichst hoch sein, bzw. β möglichst klein. Zu einem Test gehört eine sogenannte aus den vorliegenden Stichprobendaten berechnete Teststatistik (oder Prüfgröße) T , deren Verhalten möglichst empfindlich auf Abweichungen von der Nullhypothese reagiert. Bei den meisten Entscheidungsregeln führen „auffallend große“ Werte von T zur Ablehnung der Nullhypothese.

3.2.3 Einstichproben t -Test

Einstichproben t -Test für den unbekanntem Erwartungswert μ eines $N(\mu; \sigma^2)$ -verteilten Merkmals X bei (zunächst) bekannter Varianz σ^2 ^V

$$H_0 : \mu = \mu_0 \quad \text{mit einem behaupteten hypothetischen Wert } \mu_0$$

$$H_1^+ : \mu > \mu_0 \quad \Leftrightarrow \quad \delta := \mu - \mu_0 > 0$$

Gegeben: Stichprobe aus n „Realisierungen“ („Beobachtungen“) X_1, \dots, X_n von X

Teststatistik:

Mit μ_0 standardisierter Mittelwert

$$t = \frac{\bar{X} - \mu_0}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Unter H_0 (d.h. wenn H_0 wahr ist) ist t ein Zufallswert aus einer $N(0; 1)$ -Verteilung. Unter H_1^+ , wenn also $\mu > \mu_0$ gilt, ist

$$t = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} = \underbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}_{\sim N(0; 1)!} + \frac{\delta}{\sigma}\sqrt{n}$$

ein Wert aus einer $N(\frac{\delta}{\sigma}\sqrt{n}; 1)$ -Verteilung.

ZEICHNUNG!! Entscheidungsregel für den Einstichproben t -Test zum vorgegebenen Niveau α (= Irrtumswahrscheinlichkeit für Fehler 1. Art):

Lehne H_0 ab zugunsten von H_1^+ , wenn $t \geq z_\alpha$, wobei z_α die obere α -Schranke der $N(0; 1)$ -Verteilung ist. ^{VI}

Falls H_0 wahr ist, gilt $P\{t \geq z_\alpha\} = \alpha$, also wird H_0 irrtümlich mit der Wahrscheinlichkeit α abgelehnt.

Falls $\delta > 0$ ist, wird H_0 fälschlich nicht abgelehnt mit der Wahrscheinlichkeit β . Es ist

$$\beta = P\left\{t = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} < z_\alpha\right\} = P\left\{t = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_\alpha - \frac{\delta}{\sigma}\sqrt{n}\right\} = \Phi\left(z_\alpha - \frac{\delta}{\sigma}\sqrt{n}\right)$$

mit der Verteilungsfunktion Φ der $N(0; 1)$ -Verteilung. Somit ist die Power:

$$1 - \beta = 1 - \Phi\left(z_\alpha - \frac{\delta}{\sigma}\sqrt{n}\right) = \Phi\left(\frac{\delta}{\sigma}\sqrt{n} - z_\alpha\right)$$

(Stets gilt $\Phi(-x) = 1 - \Phi(x)$ aus Symmetriegründen.)

Die Kenntnis der Power lässt sich zu Planungszwecken nutzen, insbesondere zur vernünftigen Wahl eines Stichprobenumfangs n .

^VZ.B. in der industriellen Qualitätskontrolle ist meist $\sigma^2 = \text{Var}(X)$ ziemlich genau bekannt aus sehr vielen früheren X -Werten.

^{VI}Computerprogramme benutzen meist keine Schrankenwerte, sondern berechnen unter Annahme der Gültigkeit von H_0 die Überschreitungswahrscheinlichkeit von t über den aus der Stichprobe berechneten t -Wert t_{ber} , d.h. $p := P\{t \geq t_{\text{ber}} | H_0\}$. Offenbar ist $t_{\text{ber}} \geq z_\alpha$ äquivalent zu $p \leq \alpha$. p ist ein Maß für die Glaubwürdigkeit von H_0 .

Beispiel:

Wie groß sollte n bei vorgegebener Irrtumswahrscheinlichkeit $\alpha = 0,01$ gewählt werden, damit eine vorgegebene Abweichung $\delta = \mu - \mu_0$ - z.B. $\delta = \frac{1}{2}\sigma$ - vom Test mit mindestens einer Sicherheit von $1 - \beta = 0,9$ „erkannt“ wird, d.h. zur Ablehnung von H_0 führt?

$$\Phi\left(\frac{\delta}{\sigma}\sqrt{n} - z_\alpha\right) = 1 - \beta \Leftrightarrow \frac{\delta}{\sigma}\sqrt{n} - z_\alpha = z_\beta \quad (\text{aus der } \Phi\text{-Tabelle})$$

$$\Rightarrow n \geq \left(\frac{\sigma}{\delta}(z_\alpha + z_\beta)\right)^2$$

Mit obigen Zahlen: $n \geq (2 \cdot (2,326 + 1,282))^2 = 52,07$, also $n = 53$.

Für den entsprechenden Test von $H_0 : \mu = \mu_0$ gegen die zweiseitige Alternative lautet die Entscheidungsregel zum Niveau α :

$$H_0 \text{ abgelehnt, wenn } |t| \geq z_{\alpha/2}$$

Es führen also alle „zu weit von 0 entfernten“ t -Werte zur Ablehnung von H_0 .

Wenn σ unbekannt ist, wird σ durch die Stichprobenstreuung s ersetzt und die Schranken z_α durch die Schranken $t(\alpha; n-1)$ der t -Verteilung. Mit $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ hat man für die drei Varianten des Einstichproben t -Tests:

$$H_0 \text{ abgelehnt zugunsten von } \begin{cases} H_1^+ (\mu > \mu_0), & \text{wenn } t \geq t(\alpha; n-1) \\ H_1^- (\mu < \mu_0), & \text{wenn } t \leq -t(\alpha; n-1) \\ H_1 (\mu \neq \mu_0), & \text{wenn } |t| \geq t\left(\frac{\alpha}{2}; n-1\right) \end{cases}$$

3.2.4 2-Stichproben t -Tests

Häufig soll anhand von Stichproben geprüft werden, ob die Änderung eines „Einflussfaktors“ die Verteilung eines Merkmals X nach rechts oder links verschiebt. Es soll z.B. geprüft werden, ob die Einnahme eines bestimmten Medikaments die - geeignet gemessene - Konzentrationsfähigkeit X von Personen mindert, oder es soll getestet werden, ob eine technische Modifikation bei der Herstellung von Solarzellen zu einer durchschnittlichen Erhöhung des Wirkungsgrades X im praktischen Dauerbetrieb führt. Für derartige Vergleiche sind grundsätzlich zwei Situationen zu unterscheiden.

(a) Verbundene Stichproben

Es liegt eine Zufallstichprobe von n Einheiten aus der Grundgesamtheit vor, an denen das Merkmal X einmal unter Bedingung A und einmal unter Bedingung B beobachtet oder gemessen wird. Es liegen also stets zwei „verbundene“ Werte X_A, X_B von der gleichen Einheit vor! Falls möglich, sollte man verbundene Stichproben vorziehen. Wird nämlich an der gleichen Einheit ein deutlicher Unterschied zwischen X_A und X_B festgestellt, so wird man diesen Unterschied eher den unterschiedlichen Bedingungen A bzw. B zuschreiben, als wenn solche Unterschiede an verschiedenen Einheiten beobachtet würden, da sich diese auch bei identischen Bedingungen A und B hinsichtlich ihres X -Werts deutlich unterscheiden könnten. Oft sind aber verbundene Stichproben von der Sache her unmöglich. Man kann z.B. die gleiche Solarzelle im obigen Beispiel nicht auf zweierlei Weisen herstellen!

$$\text{Stichprobendaten: } (X_{1A}, X_{1B}), \dots, (X_{nA}, X_{nB}),$$

$$\mu_A = E(X_A), \mu_B = E(X_B) \text{ sind unbekannt.}$$

Zu testen ist $H_0 : \delta = \mu_B - \mu_A = 0$ gegen $H_1^+ : \delta > 0$ oder gegen $H_1^- : \delta < 0$ oder zweiseitig gegen $H_1 : \delta \neq 0$. Im obigen Beispiel mit der Konzentrationsfähigkeit wäre also einseitig gegen H_1^- zu testen, wenn B für Einnahme des Medikaments steht.

Bilde die Stichprobe der Differenzen $D_i := X_{iB} - X_{iA}$, $i = 1, \dots, n$, und prüfe mit dem Einstichproben t -Test die Hypothese

$$H_0 : \delta = E(D) = E(X_B - X_A) = \mu_B - \mu_A = 0$$

gegen die entsprechende Alternativhypothese. Prüfgröße ist also

$$t = \frac{\bar{D}}{s_D} \sqrt{n} \quad \text{mit } \bar{D} = \frac{1}{n} \sum_{i=1}^n D_i \text{ und } s_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$$

Mit diesem t lautet die Entscheidungsregel wörtlich wie vorher beim Einstichproben t -Test. Ein $(1-\alpha)$ -Konfidenzintervall für $\delta = \mu_B - \mu_A$ ist dann:

$$(\delta_1; \delta_2) = \left(\bar{D} - t\left(\frac{\alpha}{2}; n-1\right) \frac{s_D}{\sqrt{n}}; \bar{D} + t\left(\frac{\alpha}{2}; n-1\right) \frac{s_D}{\sqrt{n}} \right)$$

Die Irrtumswahrscheinlichkeit α bzw. die „statistische Sicherheit“ $1 - \alpha$ ist für normalverteilte Ausgangsdaten exakt, andernfalls nur eine Näherung.

(b) t -Test für zwei unabhängige Stichproben

Es ist

$$X_A \sim N(\mu_A; \sigma_A^2), X_B \sim N(\mu_B; \sigma_B^2), \mu_A, \mu_B, \sigma_A^2, \sigma_B^2$$

unbekannt. Zunächst wird angenommen, dass $\sigma_A^2 = \sigma_B^2 = \sigma^2$ (dann liegt für $\mu_B \neq \mu_A$ nur eine Translation (Verschiebung) der Verteilung vor.)

$$\text{Stichprobe 1 unter „Bedingung“ } A: X_{1A}, \dots, X_{mA}$$

$$\text{Stichprobe 2 unter „Bedingung“ } B: X_{1B}, \dots, X_{nB}$$

Zu testen ist wieder $H_0: \delta = \mu_B - \mu_A = 0$ gegen $\delta > 0$ oder gegen $\delta < 0$ oder zweiseitig gegen $\delta \neq 0$. Es gibt aus beiden Stichproben einen Schätzer für σ^2 :

$$s_A^2 = \frac{1}{m-1} \sum_{i=1}^m (X_{iA} - \bar{X}_A)^2 \quad \text{mit } m-1 \text{ Freiheitsgraden,}$$

$$s_B^2 = \frac{1}{n-1} \sum_{j=1}^n (X_{jB} - \bar{X}_B)^2 \quad \text{mit } n-1 \text{ Freiheitsgraden}$$

Bilde hieraus den bestmöglichen Schätzer s^2 für σ^2 :

$$s^2 = \frac{(m-1)s_A^2 + (n-1)s_B^2}{m-1+n-1} \quad \text{mit } m+n-2 \text{ FG'n}$$

s^2 ist also das mit den FG'n gewichtete Mittel von s_A^2 und s_B^2 . ($s^2 = \frac{s_A^2 + s_B^2}{2}$ für $m = n$)

Die Freiheitsgrade addieren sich hierbei, d.h. s^2 ist verteilt wie $\sigma^2 \frac{\chi_{m+n-2}^2}{m+n-2}$. Mit diesem s ist die Prüfgröße

$$t = \frac{\bar{X}_B - \bar{X}_A}{\sqrt{\hat{\text{Var}}(\bar{X}_B - \bar{X}_A)}} = \frac{\bar{X}_B - \bar{X}_A}{\sqrt{\frac{s^2}{m} + \frac{s^2}{n}}} = \frac{\bar{X}_B - \bar{X}_A}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \quad \text{mit } m+n-2 \text{ FG'n}$$

Unter H_0 gilt $t \sim t(m+n-2)$. Daher ist die Entscheidungsregel zum Niveau α wieder

$$H_0 \text{ abgelehnt zugunsten von } \begin{cases} H_1^+, & \text{wenn } t \geq t(\alpha; m+n-2) \\ H_1^-, & \text{wenn } t \leq -t(\alpha; m+n-2) \\ H_1, & \text{wenn } |t| \geq t\left(\frac{\alpha}{2}; m+n-2\right) \end{cases}$$

Das $(1-\alpha)$ -Konfidenzintervall für $\delta = \mu_B - \mu_A$ hat die Grenzen

$$\delta_{1/2} = \bar{X}_B - \bar{X}_A \mp t\left(\frac{\alpha}{2}; m+n-2\right) s \sqrt{\frac{1}{m} + \frac{1}{n}}$$

Zahlenbeispiel:

Vergleich der Wirkungsgrade X_A, X_B zweier Solarzellentypen A und B : Zu testen ist $H_0: \delta = \mu_B - \mu_A = 0$ gegen $\delta > 0$.

Stichprobe A:	13,5	11,7	12,9	12,4	11,8	13,0	13,2	12,1	12,3	12,6
Stichprobe B:	12,8	12,0	11,9	14,1	13,9	12,9	13,5	12,7	13,4	13,8

$$m = n = 10, \alpha := 0,05$$

$$\left. \begin{array}{l} \bar{X}_A = 12,55 \\ \bar{X}_B = 13,10 \end{array} \right\} \begin{array}{l} s_A^2 = 0,358\bar{3} \\ s_B^2 = 0,5911 \end{array} \Rightarrow s = 0,6890 \text{ mit } 18 \text{ FG'n.}$$

$$t = \frac{\bar{X}_B - \bar{X}_A}{s\sqrt{\frac{2}{10}}} = 1,785 > t(0,05; 18) = 1,73$$

d.h. der Test ist signifikant auf dem 5%–Niveau, also Entscheidung für $\mu_B > \mu_A$.

Falls s_A^2, s_B^2 sehr stark voneinander abweichen, ist die Annahme $\sigma_A^2 = \sigma_B^2$ unrealistisch. In diesem Fall bilde

$$\tilde{t} := \frac{\bar{X}_B - \bar{X}_A}{\sqrt{\frac{s_A^2}{m} + \frac{s_B^2}{n}}}$$

\tilde{t} ist unter $H_0 \approx t(\nu)$ –verteilt mit dem FG

$$\nu = \frac{(v_A + v_B)^2}{\frac{v_A^2}{m-1} + \frac{v_B^2}{n-1}}, \quad v_A := \frac{s_A^2}{m}, \quad v_B := \frac{s_B^2}{n}$$

(Näherung nach Welch.)

3.2.5 Test für den Wert einer unbekanntem Wahrscheinlichkeit p

Gegeben: eine relative Häufigkeit \hat{p} aus einer Stichprobe vom Umfang n .

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \text{ ist } \approx N(0; 1) \text{ – verteilt.}$$

Zu testen ist $H_0 : p = p_0$ ($p_0 =$ vorgegebener „hypothetischer“ Wert für p) gegen $H_1^+ : p > p_0$ oder $H_1^- : p < p_0$ oder zweiseitig gegen $H_1 : p \neq p_0$. Testgröße ist die unter H_0 für nicht zu kleine $n \approx N(0; 1)$ –verteilte Prüfgröße

$$Z := \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Die Entscheidungsregeln für die Tests zum Niveau α vergleichen wieder Z mit den Schranken z_α bzw. $-z_\alpha$ und im zweiseitigen Fall $|Z|$ mit $z_{\alpha/2}$ aus der Φ –Tabelle.

3.2.6 Vierfelder-Tests zum Vergleich zweier unbekannter Wahrscheinlichkeiten

Es liegen unter evtl. verschiedenen Bedingungen entstandene unabhängige $B(n_i; p_i)$ –verteilte Häufigkeiten für ein interessierendes Ereignis A vor ($i = 1, 2$). Zu testen ist, ob sich die beiden unbekanntem Wahrscheinlichkeiten p_1, p_2 für A unterscheiden oder nicht, also $H_0 : \delta := p_1 - p_2 = 0$ gegen $H_1^+ : \delta > 0$ oder gegen $H_1^- : \delta < 0$ oder zweiseitig gegen $H_1 : \delta \neq 0$. Die für den Test zur Verfügung stehenden Stichprobendaten können in einer Vierfeldertafel zusammengefaßt werden:

	A eingetreten		Σ
	ja	nein	
Stichprobe 1	N_{11}	N_{12}	n_1
Stichprobe 2	N_{21}	N_{22}	n_2
Σ	$N_{.1}$	$N_{.2}$	n

n_1, n_2 : gegebene Stichprobenumfänge

$N_{.1}, N_{.2}$: zufällig

Schätzer für die p_i : $\hat{p}_i = \frac{N_{i1}}{n_i}$

Unter H_0 würde der gemeinsame Wert $p = p_1 = p_2$ durch $\hat{p} = \frac{N_{.1}}{n}$ geschätzt. Testgröße ist die unter $H_0 \approx t(n-2)$ –verteilte standardisierte Differenz

$$\tilde{t} = \frac{\hat{p}_1 - \hat{p}_2}{\left(\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)^{1/2}}$$

die bei Tests zum asymptotischen Niveau α mit den Schranken $t(\alpha; n-2)$ bzw. zweiseitig mit $t(\frac{\alpha}{2}; n-2)$ zu vergleichen ist. (Für $n \geq 200$ mit den entsprechenden Schranken der Normalverteilung.)

Oft wird auch die Chi-Quadrat-Prüfgröße

$$\tilde{\chi}^2 = n \frac{(N_{11} N_{22} - N_{12} N_{21})^2}{n_1 n_2 N_{.1} N_{.2}} \quad \text{oder} \quad \tilde{\chi} = \sqrt{n} \frac{N_{11} N_{22} - N_{12} N_{21}}{\sqrt{n_1 n_2 N_{.1} N_{.2}}}$$

verwendet. Eine einfache Umrechnung zeigt $\tilde{t} = \tilde{\chi}$!

Unter H_0 ist also $\tilde{\chi}^2 \approx$ verteilt wie Z^2 mit $Z \sim N(0; 1)$, also $\chi^2(1)$ -verteilt (1 Freiheitsgrad). Diese Approximation ist etwas genauer für $\chi^{*2} := \frac{n-1}{n} \tilde{\chi}^2$.

Entscheidungsregeln für diesen χ^2 -Test zum Niveau α :

$$H_0 (\delta = p_1 - p_2 = 0) \text{ abgelehnt zugunsten von } \begin{cases} H_1^+ (\delta > 0), & \text{wenn } \tilde{\chi} \geq z_\alpha \\ H_1^- (\delta < 0), & \text{wenn } \tilde{\chi} \leq -z_\alpha \\ H_1 (\delta \neq 0), & \text{wenn } \tilde{\chi}^2 \geq \tilde{\chi}^2(1 - \alpha; 1) \end{cases}$$

(Obere χ^2 -Schranke $\chi^2(1 - \alpha; 1) = z_{\alpha/2}^2$)

Zahlenbeispiel:

Test für $H_0 : p_2 = p_1$ gegen $H_1 : p_2 \neq p_1$ auf dem 5%-Niveau:

	ja	nein	Σ
1.	148	52	200
2.	160	88	248
Σ	308	140	448

$$\tilde{\chi}^2 = 448 \frac{(148 \cdot 88 - 52 \cdot 160)^2}{200 \cdot 248 \cdot 308 \cdot 140} = 4,635 > 3,8416 = \chi^2(0,95; 1) \quad (= z_{0,025}^2)$$

Der Test ist signifikant auf dem 5%-Niveau, d.h. $H_0 (p_2 = p_1)$ wird als unglaublich abgelehnt.

3.3 Regression, Maximum-Likelihood-Schätzung von Parametern, Methode der kleinsten Quadrate

3.3.1 Regression

Häufig will man wissen, wie eine Größe Y durch ein oder mehrere Einflussgrößen X_1, \dots, X_q beeinflusst wird. Außer den identifizierten bekannten Einflussgrößen X_1, \dots, X_q kann es viele weitere unbekannte unkontrollierbare Einflüsse auf Y geben, so dass der Y -Wert durch die X_j -Werte noch nicht vollständig bestimmt wird, also nicht einfach eine mathematische Funktion von X_1, \dots, X_q ist. Der Y -Wert setzt sich additiv zusammen aus einer Funktion von X_1, \dots, X_q , die eine gewisse Anzahl von Parametern enthalten kann und einen Anteil, der von den vielen unbekanntem Einflüssen herrührt und den man als „Zufallsrest“ auffasst.

Die genaue Lieferzeit Y eines Produkts durch eine Spedition wird linear durch die Entfernung X zum Kunden beeinflusst, hängt aber auch von den zufälligen Schwankungen der Verkehrsdichte ab.

Die mittägliche Ozonkonzentration Y wird durch meteorologische Daten vom Morgen des gleichen Tages stark beeinflusst (z.B. $X_1 =$ Ozonkonzentration am Morgen, $X_2 =$ Bewölkungsgrad, $X_3 =$ Windstärke, ...) Y kann aber nicht exakt als Funktion dieser X_j vorherbestimmt werden, weil es zu viele weitere unbekannte Schwankungsgründe für die einzelnen Y -Messwerte gibt. Man kann aber versuchen, eine möglichst gute Prognose \hat{Y} als Funktion der verfügbaren Einflussgrößen X_j zu erhalten.

Sehr viele derartige Situationen können bereits hinreichend genau durch das Modell der linearen Regression beschrieben werden:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_q X_q + Z$$

mit einem von X_1, \dots, X_q unabhängigen „Zufallsrest“ Z mit $E(Z) = 0$ und $\text{Var}(Z) = \sigma^2$. Die unbekanntem Parameter dieses Modells sind die Regressionskoeffizienten β_0, \dots, β_q und die Restvarianz σ^2 . Das Modell ist nicht so speziell wie es auf den ersten Blick wirkt, da die X_j selbst auch nicht lineare Funktionen anderer

Zufallsgrößen sein können.

Es ist z.B. mit $X_1 := X, X_j := X^j, j = 1, \dots, q,$

$$Y = p(X) + Z = \beta_0 + \beta_1 X + \dots + \beta_q X^q + Z.$$

Eine polynomiale Beeinflussung von Y durch eine (oder auch mehrere) Größen ist als Spezialfall in obigem Modell enthalten! Das Wort linear bezieht sich im linearen Regressionmodell auf die Linearität bzgl. der Parameter β_0, \dots, β_q . Dementsprechend liegt ein nicht lineares Regressionmodell vor, wenn mit einer bzgl. der Parameter $\vec{\beta} = (\beta_0, \beta_1, \dots)$ nicht linearen Funktion $f(X_1, \dots, X_q; \vec{\beta})$ gilt:

$$Y = f(X_1, \dots, X_q; \vec{\beta}) + Z$$

Insbesondere bei Versuchen zur Untersuchung des möglichen Einflusses von X_1, \dots, X_q auf das „Zielmerkmal“ Y wird man Werte x_{ij} der Größen X_j vorgeben. Man hat dann einen Datensatz (Stichprobe) folgender Form:

$$\begin{array}{cccc} Y_1 & x_{11} & \dots & x_{1q} \\ \vdots & \vdots & & \vdots \\ Y_i & x_{i1} & \dots & x_{iq} \\ \vdots & \vdots & & \vdots \\ Y_n & x_{n1} & \dots & x_{nq} \end{array}$$

Mit zufälligen Resten Z_i gilt dann für die zufallsabhängigen Werte Y_i das sogenannte Modell der linearen Regression 1. Art:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_q x_{iq} + Z_i, \quad i = 1, \dots, n, \quad n > q + 1.$$

$$E(Z_i) = 0, \quad \text{Var}(Z_i) = \text{Var}(Y_i) = \sigma^2,$$

alle Z_i (und damit auch alle Y_i) vollständig unabhängig. Es ist also $E(Y_i)$ gleich der bedingten Erwartung von Y unter der Bedingung $X_j = x_{ij}, j = 1, \dots, q,$ also

$$E(Y_i) = E(Y|X_j = x_{ij}, j = 1, \dots, q) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_q x_{iq}.$$

Mit der sogenannten Designmatrix

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1q} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nq} \end{pmatrix} \quad \text{und dem Parametervektor } \vec{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_q \end{pmatrix}$$

läßt sich obiges Modell kürzer in der Form

$$\boxed{\vec{y} = X \vec{\beta} + \vec{z}}$$

schreiben. (Auch zufallsabhängige Vektoren wie \vec{y} und \vec{z} werden oft klein geschrieben.)

Die Grundaufgabe der linearen Regression besteht darin, aus den gegebenen Stichprobendaten die unbekanntem Modellparameter β_0, \dots, β_q und σ^2 möglichst optimal zu schätzen.

Die Parameter β_0, \dots, β_q werden üblicherweise mit der Methode der kleinsten Quadrate geschätzt, die ein Spezialfall der allgemeinen Maximum Likelihood Schätzmethode für unbekanntem Verteilungsparameter aus Stichproben ist. Diese allgemeine Parameterschätzmethode soll hier kurz erklärt werden.

3.3.2 Das Maximum-Likelihood-Schätzprinzip

ZEICHNUNG!!

Angenommen, man weiß, dass ein beobachteter Wert x einer Zufallsgröße X nur aus einer der beiden möglichen Verteilungen mit Dichte f_i und Erwartungswert $\mu_i, i = 1, 2,$ stammen kann. Wenn man sich für eine Dichte bzw. einen Parameterwert μ_i entscheiden muss, wird man sich für jene Dichte f entscheiden, die den größeren der beiden Werte $f_1(x; \mu_1), f_2(x; \mu_2)$ gibt. Unter der anderen Verteilung wäre es ja unwahrscheinlicher, den

tatsächlich eingetretenen Wert x zu bekommen! Dieses Prinzip läßt sich wesentlich verallgemeinern:

Hat man eine Zufallsstichprobe mit Werten y_i aus einer Verteilung mit einer von einem unbekanntem Parameter ϑ (oder einem Parametervektor $\vec{\vartheta} = (\vartheta_1, \dots, \vartheta_q)$, falls es mehrere Parameter sind) abhängigen Dichte $f(y; \vec{\vartheta})$, so ist die gemeinsame Dichte der Stichprobenwerte y_1, \dots, y_n :

$$f_n(\vec{y}; \vec{\vartheta}) = f_n(y_1, \dots, y_n; \vartheta_1, \dots, \vartheta_q) = \prod_{i=1}^n f(y_i; \vec{\vartheta})$$

Der Maximum-Likelihood-Schätzer $\hat{\vec{\vartheta}} = \hat{\vec{\vartheta}}(y_1, \dots, y_n)$ ist derjenige Wert von $\vec{\vartheta}$, der $f_n(\vec{y}; \vec{\vartheta})$ als Funktion von $\vec{\vartheta}$ maximiert (sofern dieser eindeutig bestimmt ist).

Äquivalent hierzu wird (aus rechentechnischen Gründen) meist $L_n(\vec{y}; \vec{\vartheta}) := \ln(f_n(\vec{y}; \vec{\vartheta}))$ bzgl. $\vec{\vartheta}$ maximiert. $f_n(\vec{y}; \vec{\vartheta})$ heißt als Funktion von $\vec{\vartheta}$ auch die Likelihood-Funktion und $L_n(\vec{y}; \vec{\vartheta})$ die Log-Likelihood-Funktion.

Man könnte Maximum-Likelihood-Schätzer verdeutschen als „plausibelster Schätzwert“ auf Basis der verfügbaren Stichprobendaten. Die Maximum-Likelihood-Schätzmethode hat sich wegen ihrer (zumindest asymptotischen) Optimalitätseigenschaften durchgesetzt. Unter sehr allgemeinen Voraussetzungen sind Likelihood-Schätzer asymptotisch effizient, d.h. sie haben asymptotisch minimale Varianz, sind also (zumindest für große n) nicht durch genauere Schätzmethode ersetzbar. Außerdem sind diese Schätzer (fast immer) asymptotisch normalverteilt.

Ein einfaches Beispiel: $Y \sim N(\mu; \sigma^2)$, μ, σ^2 unbekannt: \Rightarrow Dichte der Stichprobenwerte y_1, \dots, y_n ist

$$\begin{aligned} f_n(y_1, \dots, y_n; \mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y_i - \mu}{\sigma}\right)^2\right) = \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right), \quad (\vec{\vartheta} = (\mu, \sigma^2)) \end{aligned}$$

Die Maximum-Likelihood-Schätzer $\hat{\mu}, \hat{\sigma}^2$ ergeben sich aus den notwendigen Bedingungen $\frac{\partial f_n}{\partial \mu} = 0, \frac{\partial f_n}{\partial \sigma^2} = 0$. Man erkennt aber ohne Rechnung, dass zunächst für bel. σ^2 das Maximum von f_n bzgl. μ durch Minimierung der Quadratsumme

$$Q(\mu) := \sum_{i=1}^n (y_i - \mu)^2$$

bzgl. μ erreicht wird, was auf die Lösung

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

führt. Der Stichprobenmittelwert \bar{y} ist also der Maximum-Likelihood-Schätzer für μ . Die Maximierung bzgl. σ^2 führt auf

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Dieser Schätzer ist im Gegensatz zu

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

nicht erwartungstreu, hat aber eine etwas kleinere Varianz als s^2 .

3.3.3 Die Methode der kleinsten Quadrate

Im obigen Regressionsmodell hat Y_i den Erwartungswert $\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_q x_{iq}$. Nimmt man zunächst $N(0; \sigma^2)$ -verteilte Reste Z_i an, so findet man ähnlich wie vorher die Maximum-Likelihood-Schätzer für $\vec{\beta} = (\beta_0, \dots, \beta_q)$ durch Maximieren der gemeinsamen Dichte

$$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_q x_{iq})^2\right)$$

von (Y_1, \dots, Y_n) bzgl. $\vec{\beta}$. Dies ist offenbar äquivalent zur Bestimmung des Minimums der Summe der „Abweichungsquadrate“

$$Q(\vec{\beta}) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_q x_{iq})^2 = |\vec{y} - X \vec{\beta}|^2.$$

Dies ist die „Methode der kleinsten Quadrate“ zur Schätzung von $\vec{\beta}$. Die notwendigen Bedingungen $\frac{\partial Q}{\partial \beta_0} = 0, \dots, \frac{\partial Q}{\partial \beta_q} = 0$ bilden ein lineares Gleichungssystem zur Bestimmung von $\vec{\beta}$. Dieses System erhält man auch geometrisch sehr einfach: Die Vektoren

$$X \vec{\beta} = \beta_0 \vec{x}_0 + \beta_1 \vec{x}_1 + \dots + \beta_q \vec{x}_q$$

mit den Spalten \vec{x}_j von X^{VII} und beliebigen β_j bilden den von den \vec{x}_j aufgespannten Unterraum $[\vec{x}_0, \dots, \vec{x}_q]$ des \mathbb{R}^n (beachte Voraussetzung $q + 1 < n$).

$|\vec{y} - X \vec{\beta}|$ wird minimal, wenn $X \vec{\beta}$ die Orthogonalprojektion von \vec{y} auf diesen Unterraum ist, also $\vec{y} - X \vec{\beta}$ senkrecht zu allen \vec{x}_j ist. Somit sind alle Skalarprodukte $\vec{x}_j'(\vec{y} - X \vec{\beta}) = 0$, zusammengefaßt also:

$$X'(\vec{y} - X \vec{\beta}) = \vec{0} \quad \Leftrightarrow \quad \boxed{X'X \vec{\beta} = X' \vec{y}.}$$

Dieses lineare Gleichungssystem hat für $\text{Rang}(X) = \text{Spaltenzahl von } X = q + 1$ die eindeutige Lösung

$$\boxed{\vec{b} = \hat{\vec{\beta}} = (X'X)^{-1} X' \vec{y}}$$

\vec{b} ist der „Kleinst-Quadrate-Schätzer“ für $\vec{\beta}$, der im Fall normalverteilter Reste mit dem Maximum-Likelihood-Schätzer übereinstimmt. Auch für Zufallsreste Z_i mit unbekannter Verteilung (das ist die in der Praxis meist vorliegende Situation) wird der Kleinst-Quadrate-Schätzer \vec{b} verwendet, weil dieser unter allen linear von den Werten y_i abhängigen unverzerrten Schätzern dann noch immer die kleinste Varianz hat.

3.3.4 Die Regressionsgerade

Hier handelt es sich um lineare Regression mit nur einer Einflußgröße X ($q = 1$).

Modell: $Y_i = E(Y_i) + Z_i = \beta_0 + \beta_1 x_i + Z_i$ mit unabhängigen Resten Z_i , $E(Z_i) = 0$, $\text{Var}(Z_i) = \text{Var}(Y_i) = \sigma^2$.

Man unterscheide die „wahre“ unbekannte Regressionsgerade

$$E(Y|X = x) = \beta_0 + \beta_1 x$$

von der aus der Stichprobe (x_i, y_i) , $i = 1, \dots, n$, geschätzten Regressionsgeraden

$$\hat{y} = b_0 + b_1 x.$$

Mit

$$X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix}$$

erhält man:

$$X'X = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}, \quad X' \vec{y} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

und nach leichter Umformung:

$$\boxed{b_1 = \frac{Q_{xy}}{Q_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad b_0 = \bar{y} - b_1 \bar{x}}$$

^{VII}

$$\vec{x}_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

(Tasten auf Taschenrechner für b_0 und b_1 oft mit a , b bezeichnet.)

Die geschätzte Regressiongerade $\hat{y} = b_0 + b_1 x = \bar{y} + b_1(x - \bar{x})$ geht stets durch den Punkt (\bar{x}, \bar{y}) .

Wegen $Q_{xy} = b_1 Q_{xx}$ erhält man für die „Restquadratsumme“ (= Summe der quadrierten Abweichungen zwischen den y_i und den Werten $\hat{y}_i = \bar{y} + b_1(x_i - \bar{x})$):

$$Q_{R(\text{est})} := \sum_{i=1}^n \underbrace{(y_i - \bar{y}) - b_1(x_i - \bar{x})}^2 = Q_{yy} - 2b_1 Q_{xy} + b_1^2 Q_{xx} = Q_{yy} - b_1^2 Q_{xx}.$$

(oder $Q_{yy} - b_1 Q_{xy}$ mit $Q_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$)

$$s^2 = \frac{1}{n-2} Q_R$$

ist dann ein erwartungstreuer Schätzer für σ^2 mit $n-2$ Freiheitsgraden.

Deutung der Quadratsummenzerlegung:

$$Q_{yy} = b_1^2 Q_{xx} + Q_{\text{Rest}} \quad \text{mit den zugehörigen Freiheitsgraden}$$

$$\text{FG: } n-1 = 1 + n-2.$$

$$b_1^2 Q_{xx} = \sum_i b_1^2 (x_i - \bar{x})^2 = \sum_i (\hat{y}_i - \bar{y})^2$$

erfaßt die Schwankung der Werte \hat{y}_i auf der Regressiongeraden, also den Anteil der Schwankung von Y , der durch die Abhängigkeit von x verursacht wird. Q_R erfaßt die von x unabhängige Restschwankung von Y . Daher bezeichnet man

$$r^2 := \frac{b_1^2 Q_{xx}}{Q_{yy}} = 1 - \frac{Q_R}{Q_{yy}}$$

auch als das (aus der Stichprobe geschätzte) Bestimmtheitsmaß von Y durch x . Rechnerisch ist r^2 identisch mit dem Quadrat $\frac{Q_{xy}^2}{Q_{xx} Q_{yy}}$ des Stichprobenkorrelationskoeffizienten r der Paare (x_i, y_i) .

3.3.5 Tests und Konfidenzbereiche für die Parameter von Regressiongeraden

Für $N(0; \sigma^2)$ -verteilte unabhängige Reste sind alle folgenden t -Prüfgrößen exakt t -verteilt und die nominalen α -Niveaus exakt, andernfalls handelt es sich um Näherungen. Die wahre Regressiongerade wird jetzt durch $\alpha + \beta x$ bezeichnet und die geschätzte durch $\hat{y} = a + b x$.

(1) Konfidenzintervall und Test für die „wahre“ unbekannte Steigung β einer Regressiongeraden:

Es liege eine aus einer Stichprobe geschätzte Regressiongerade $\hat{y} = a + b x$ vor mit einem Reststreuungsquadrat $s^2 = \frac{1}{n-2} Q_R$ mit $n-2$ FG'n.

$$\text{Var}(b) = \text{Var} \left(\frac{1}{Q_{xx}} \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) \right) = \text{Var} \left(\frac{1}{Q_{xx}} \sum_{i=1}^n (x_i - \bar{x}) Y_i \right), \quad \text{da } \bar{Y} \sum_{i=1}^n (x_i - \bar{x}) = 0.$$

(Y_i hier groß geschrieben, um die Zufallsabhängigkeit hervorzuheben; die x_i sind fest vorgegebene nicht zufällige Werte.)

$$\Rightarrow \text{Var}(b) = \frac{1}{Q_{xx}^2} \sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2 = \frac{\sigma^2}{Q_{xx}},$$

geschätzt durch $\frac{s^2}{Q_{xx}}$ mit $n-2$ FG'n. Somit ist mit dem wahren β -Wert $\beta = E(b)$:

$$t := \frac{b - \beta}{\sqrt{\hat{\text{Var}}(b)}} = \frac{b - \beta}{s/\sqrt{Q_{xx}}} \sim t(n-2)$$

⇒ Grenzen $\beta_{1/2}$ für ein $(1 - \alpha)$ -Konfidenzintervall für β :

$$\beta_{1/2} = b \mp t\left(\frac{\alpha}{2}; n - 2\right) \frac{s}{\sqrt{Q_{xx}}}.$$

t -Tests für $H_0 : \beta = \beta_0$ (β_0 : vorgegebener behaupteter Wert) gegen $H^+ : \beta > \beta_0$ oder $H^- : \beta < \beta_0$ oder $H_1 : \beta \neq \beta_0$:

Mit der unter H_0 t -verteilten Prüfgröße $t = \frac{b - \beta_0}{s/\sqrt{Q_{xx}}}$ hat man die Entscheidungsregeln für diese t -Tests zum Niveau α :

$$H_0 \text{ abgelehnt zugunsten von } \begin{cases} H_1^+, & \text{wenn } t \geq t(\alpha; n - 2) \\ H_1^-, & \text{wenn } t \leq -t(\alpha; n - 2) \\ H_1, & \text{wenn } |t| \geq t\left(\frac{\alpha}{2}; n - 2\right). \end{cases}$$

(2) Konfidenzintervall und Tests für den Achsenabschnitt α :

$$\text{Var}(a) = \text{Var}(\bar{Y} - b\bar{x}) = \text{Var}(\bar{Y}) + \bar{x}^2 \text{Var}(b) = \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{Q_{xx}},$$

geschätzt durch

$$s^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{Q_{xx}} \right) \Rightarrow t := \frac{a - \alpha}{s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Q_{xx}}}} \sim t(n - 2).$$

Grenzen $\alpha_{1/2}$ für 95%-Konfidenzintervall für α :

$$a \mp t(0,025; n - 2) s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Q_{xx}}}.$$

Entsprechend kann man Hypothesen $H_0 : \alpha = \alpha_0$ gegen ein- oder zweiseitige Alternativhypothesen $\alpha > \alpha_0$, $\alpha \neq \alpha_0$ mit dem t -Test und der Prüfgröße

$$t = \frac{a - \alpha_0}{s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Q_{xx}}}}$$

testen.

(3) Vergleich der Steigung zweier unabhängiger Regressionsgeraden:

Aus zwei unabhängigen Stichproben mit n_1 bzw. n_2 Wertepaaren (x_i, Y_i) für die gleichen Merkmale X, Y unter evtl. modifizierten (Versuchs)bedingungen liegen zwei geschätzte Regressionsgeraden $\hat{y}_1 = a_1 + b_1 x$, $\hat{y}_2 = a_2 + b_2 x$ vor mit den Reststreuungsquadraten s_1^2 mit $n_1 - 2$ FG'n bzw. s_2^2 mit $n_2 - 2$ FG'n. Die gepoolte Reststreuung

$$s = \sqrt{\frac{(n_1 - 2)s_1^2 + (n_2 - 2)s_2^2}{n_1 - 2 + n_2 - 2}}$$

hat dann $\nu = n_1 + n_2 - 4$ FG'e. Zu testen ist:

$H_0 : \beta_2 = \beta_1$ gegen $H_1^+ : \beta_2 > \beta_1$ oder gegen $H_1^- : \beta_2 < \beta_1$ oder gegen $H_1 : \beta_2 \neq \beta_1$

Mit der unter H_0 $t(\nu)$ -verteilten Prüfgröße

$$t := \frac{b_2 - b_1}{\sqrt{\hat{\text{Var}}(b_2 - b_1)}} = \frac{b_2 - b_1}{s\sqrt{\frac{1}{Q_{xx1}} + \frac{1}{Q_{xx2}}}}$$

hat man die üblichen Entscheidungsregeln für t -Tests zum Niveau α mit $\nu = n_1 + n_2 - 4$ FG'n.^{VIII}

^{VIII}Falls wegen stark unterschiedlicher Reststreuungen nicht gepoolt werden soll, kann für

$$t = \frac{b_2 - b_1}{s\sqrt{\frac{1}{Q_{xx1}} + \frac{1}{Q_{xx2}}}}$$

der FG nach Welch (vgl. 3.2.4) verwendet werden.

(4) Prognoseintervall

Prognoseintervall für einen künftigen Wert Y^* an der Stelle x^* , der unter den gleichen kontrollierten Bedingungen entsteht wie die Stichprobendaten (x_i, Y_i) , $i = 1, \dots, n$ aus denen eine geschätzte Regressionsgerade $\hat{y} = a + bx = \bar{Y} + b(x - \bar{x})$ vorliegt:

$$\text{Var}(Y^* - \hat{y}^*) = \text{Var}(Y^* - \bar{Y} - b(x^* - \bar{x})) = \sigma^2 + \frac{\sigma^2}{n} + (x^* - \bar{x})^2 \frac{\sigma^2}{Q_{xx}},$$

geschätzt durch

$$s^2 \left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{Q_{xx}} \right)$$

mit der Reststreuung s der vorliegenden Regressionsgeraden mit $n - 2$ FG'n.

⇒ $(1 - \alpha)$ -Prognoseintervall für Y^* an der Stelle x^* hat die Grenzen:

$$\bar{Y} + b(x^* - \bar{x}) \mp t\left(\frac{\alpha}{2}; n - 2\right) s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{Q_{xx}}}.$$

Ein Prognoseintervall ist für den künftigen Wert einer Zufallsgröße konstruiert, ein Konfidenzintervall hingegen für den unbekanntem Wert eines Parameters, also einer nicht zufallsabhängigen Zahl!

ZEICHNUNG!!

Zahlenbeispiel:

x = Entfernung [km], Y = Lieferzeit [h], $n = 10$

x_i	215	325	480	550	670	825	920	1070	1215	1350
y_i	3,0	4,5	6,25	7,25	8,75	10,5	12,0	13,75	15,5	18,0

Geschätzte Regressionsgerade: $\hat{y} = a + bx$ mit $a = 0,153310$, $b = 0,0128565$

$$\begin{array}{llll} \bar{x} = 762 & s_x^2 = 144206, \bar{6} & Q_{xx} = 9s_x^2 = 1297860 & Q_R = 0,450632 \\ \bar{Y} = 9,95 & s_y^2 = 23,886\bar{1} & Q_{yy} = 9s_y^2 = 214,975 & s = 0,237337 \quad \text{mit } n - 2 = 8 \text{ FG'n} \end{array}$$

$$t(0,025; 8) = 2,306$$

95%-Prognoseintervall für Y^* an der Stelle $x^* = 600$ km:

$$\begin{aligned} a + bx^* \mp t(0,025; 8) s \sqrt{1 + \frac{1}{10} + \frac{(x^* - \bar{x})^2}{Q_{xx}}} &= \\ = 7,867239 \mp 2,306 \cdot 0,237337 \cdot 1,058405 &= (7,288; 8,4465) \end{aligned}$$

also ungefähr (7 h 17 min; 8 h 27 min).

Bei einem bzgl. der Parameter $\vec{\beta} = (\beta_1, \beta_2, \dots, \beta_q)$ nicht linearen Regressionsmodell $Y_i = f(x_i; \vec{\beta}) + Z_i$ führt die Methode der kleinsten Quadrate, also die Minimierung der „Abweichungsquadratsumme“

$$Q(\vec{\beta}) = \sum_{i=1}^n (Y_i - f(x_i; \vec{\beta}))^2$$

bzgl. $\vec{\beta}$ auf das nicht lineare Gleichungssystem $\frac{\partial Q}{\partial \beta_j} = 0$, $j = 1, \dots, q$ mit den „Kleinst-Quadrate-Schätzern“ $\hat{\beta}_j$ als Lösungen (falls eindeutig). In manchen Fällen ist es möglich, eine Schätzung einfacher durch Linearisierung - d.h. Rückführung auf lineare Regression - zu bestimmen.

Beispiel:

$$f(x; \alpha, \beta) = \alpha e^{\beta x}.$$

Es soll eine Exponentialkurve $\hat{y} = a e^{bx}$ an die Stichprobendaten (x_i, Y_i) , $i = 1, \dots, n$ angepaßt werden. Stattdessen wird eine Regressiongerade $\hat{y}^* = a^* + bx = \ln(a) + bx$ an die transformierten Daten $(x_i, Y_i^* := \ln(Y_i))$ angepaßt. Mit der Rücktransformation $a = e^{a^*}$ erhält man Schätzer (a, b) für (α, β) . Man beachte, dass diese Schätzer nicht exakt mit den Schätzern $(\hat{\alpha}; \hat{\beta})$ übereinstimmen, die mit der direkten Methode der kleinsten Quadrate gewonnen würden, also durch Lösen der Gleichungen $\frac{\partial Q}{\partial \alpha} = 0$, $\frac{\partial Q}{\partial \beta} = 0$ mit

$$Q(\alpha, \beta) = \sum_{i=1}^n (Y_i - \alpha e^{\beta x_i})^2.$$

Die Linearisierungsmethode sollte besser nur bei kleinen Abweichungen von der Regressionslinie angewandt werden!

Falls die Varianzen σ_i^2 der Reste Z_i nicht identisch sind, sondern von der Form $\sigma_i^2 = \sigma^2 v_i$ mit bekannten v_i (z.B. $v_i = y_i$ oder $v_i = y_i^2$), wird die gewichtete Abweichungssumme

$$\sum_i \frac{1}{v_i} (Y_i - f(x_i; \vec{\beta}))^2$$

minimiert.

Im Fall des allgemeinen linearen Regressionsmodells erhält man mit der Schätzung $\vec{b} = (b_0, \dots, b_q)$ die „Restquadratsumme“ $Q_R := |\vec{y} - X \vec{b}|^2$ und hieraus den Schätzer

$$s^2 = \frac{Q_R}{n - 1 - q}$$

mit $n - 1 - q$ FG'n für die unbekannte Restvarianz σ^2 . Die Schätzer \vec{b} sind asymptotisch $N_{q+1}(\vec{\beta}; \sigma^2(X'X)^{-1})$ – verteilt mit der Kovarianzmatrix $(\text{Kov}(b_i, b_j)) = \sigma^2(X'X)^{-1}$. Deren Schätzung $s^2(X'X)^{-1}$ ist in Programmen für Kleinst-Quadrate Schätzung (engl.: least square estimate) meist verfügbar. Insbesondere ist dann $\hat{\text{Var}}(b_j) = s^2 v_{jj}$ mit dem j -ten Diagonalelement v_{jj} von $V = (v_{ij}) := (X'X)^{-1}$. Damit erhält man in der üblichen Weise ein $(1 - \alpha)$ -Konfidenzintervall für β_j mit den Grenzen

$$b_j \mp t\left(\frac{\alpha}{2}; n - 1 - q\right) s \sqrt{v_{jj}}$$

Z.B. ist mit

$$X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix}$$

bei einer Regressionsgeraden:

$$(X'X)^{-1} = \begin{pmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix}^{-1} = \frac{1}{n \Sigma x_i^2 - (\Sigma x_i)^2} \begin{pmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{pmatrix}$$

und damit tatsächlich - wie bereits vorher schon verwendet -

$$\text{Var}(b_1) = \text{Var}(b) = \sigma^2 \frac{n}{n \Sigma x_i^2 - (\Sigma x_i)^2} = \frac{\sigma^2}{\Sigma x_i^2 - n \bar{x}^2} = \frac{\sigma^2}{Q_{xx}}.$$

3.4 Chi-Quadrat-Tests zur Analyse von Häufigkeiten

3.4.1 Goodness of fit - Test^{IX} für unbekannte Wahrscheinlichkeitswerte

Es liege eine Stichprobe vom Umfang n vor mit den Häufigkeiten N_j für die Ausprägungen x_j eines diskreten Merkmals X mit q möglichen Werten x_j . Die Werte der Wahrscheinlichkeiten $p_j = P\{X = x_j\}$ sind unbekannt.

^{IX} „goodness of fit“: „Güte der Anpassung“, nämlich des Modells an die Daten.

Es ist $\sum_{j=1}^q p_j = 1$ und $\sum_{j=1}^q N_j = n$, $E(N_j) = n p_j$.
Zu testen ist eine Nullhypothese über die Werte p_j :

$$H_0 : p_j = p_{0j}, \quad j = 1, \dots, q \text{ mit vorgegebenen Werten } p_{0j}.$$

gegen

$$H_1 : \text{nicht alle } p_j = p_{0j}.$$

z.B. H_0 : X hat eine diskrete Gleichverteilung, d.h. alle p_j sind identisch, also $p_{0j} = \frac{1}{q}$ für alle j .

Gemäß Abschnitt 2.3 ist

$$\tilde{\chi}^2 := \sum_{j=1}^q \frac{(N_j - n p_{0j})^2}{n p_{0j}} \quad \text{unter } H_0 \text{ asymptotisch } \chi^2(q-1) \text{ - verteilt!}$$

Unter H_1 hat $\tilde{\chi}^2$ eine Tendenz zu größeren Werten. Genauer:

$\tilde{\chi}^2$ hat dann eine sogenannte nicht zentrale χ^2 -Verteilung, die außer von r vom Nichtzentralitätsparameter $\delta_n^2 := n \sum_{j=1}^q \frac{(p_j - p_{0j})^2}{p_{0j}}$ abhängt.

Mit den oberen Schranken $\chi^2(1 - \alpha; q - 1)$ der $\chi^2(q - 1)$ -Verteilung hat man die Entscheidungsregel zum asymptotischen Niveau α :

$$H_0 \text{ wird abgelehnt, wenn } \tilde{\chi}^2 \geq \chi^2(1 - \alpha; q - 1).$$

$\tilde{\chi}^2$ lässt sich durch Ausquadrieren umformen zu

$$\tilde{\chi}^2 = \left(\frac{1}{n} \sum_{j=1}^q \frac{N_j^2}{p_{0j}} \right) - n, \quad \left(= \left(\frac{q}{n} \sum_{j=1}^q N_j^2 \right) - n, \text{ falls alle } p_{0j} = \frac{1}{q} \right)$$

Beispiel:

X := Bahnnummer des Siegerpferdes bei Pferderennen mit $q = 8$ Bahnen. Es sollte eigentlich gelten: $p_j = P\{X = j\} = \frac{1}{8}$, $j = 1, \dots, 8$. Diese Nullhypothese H_0 (also gleiche Siegeschancen für jede Bahnnummer) wurde anhand einer Zufallsstichprobe mit den Ergebnissen von $n = 144$ Pferderennen getestet:

Bahn Nr. j	1	2	3	4	5	6	7	8
Anzahl Siege	29	19	18	25	17	10	15	11

Niveau $\alpha = 0,05$, $\chi^2(0,095; 7) = 14,07$

$$\tilde{\chi}^2 = \frac{8}{144}(29^2 + 19^2 + \dots + 11^2) - 144 = 16,33 > 14,07$$

$\Rightarrow H_0$ wird auf dem 5%-Niveau abgelehnt!

3.4.2 Allgemeiner Goodness of fit - Test für Wahrscheinlichkeiten mit unbekanntem Parameterwerten

Wie vorher liegen Stichprobenhäufigkeiten N_j , $j = 1, \dots, q$ vor mit $\sum_{j=1}^q N_j = n$ = Stichprobenumfang. Oft will man eine Modellvorstellung über die zu erwartenden Häufigkeiten $E(N_j) = n p_j(\vec{\vartheta})$ prüfen, wobei die Wahrscheinlichkeiten p_j gegebene Funktionen gewisser Parameter $(\vartheta_1, \dots, \vartheta_k) = \vec{\vartheta}$ sind, deren Werte aber unbekannt sind.

Ein einfaches Beispiel:

Modellvorstellung: Das Merkmal X hat eine Poissonverteilung. (= Nullhypothese)

In einer Stichprobe mit n Realisierungen von X werden die Häufigkeiten N_j des Ereignisses $\{X = x_j = j\}$ gezählt, $j = 0, \dots, q - 1$. Da große Werte für j selten oder gar nicht auftreten, wird für geeignetes q statt $\{X = q\}$ das Ereignis $\{X \geq q\}$ mit der Häufigkeit N_q betrachtet. Bei Gültigkeit einer Poissonverteilung mit dem unbekanntem Erwartungswert $\mu = E(X)$ würde für die zu erwartenden Häufigkeiten gelten:

$$E(N_j) = n p_j = n \cdot e^{-\mu} \frac{\mu^j}{j!} \quad \text{für } j < q$$

und

$$p_q = 1 - \sum_{j < q} p_j.$$

Hier ist also $\vec{\vartheta} = \vartheta = \mu$ ein unbekannter Parameter, von dem die Wahrscheinlichkeiten $p_j = p_j(\mu)$ abhängen.

Im allgemeinen Fall (jetzt wieder $j = 1, \dots, q$) gewinnt man aus den multinominal verteilten Stichprobenhäufigkeiten N_1, \dots, N_q durch Maximierung ihrer Likelihood-Funktion

$$\prod_{j=1}^q (p_j(\vec{\vartheta}))^{N_j} \quad \text{oder besser der Log-Likelihood-Funktion} \quad \sum_j N_j \ln(p_j(\vec{\vartheta}))$$

bzgl. $\vec{\vartheta}$ den Likelihood-Schätzer $\hat{\vec{\vartheta}}$ und hieraus die Likelihoodschätzer $\hat{p}_j := p_j(\hat{\vec{\vartheta}})$ für die p_j .
 Wenn die Nullhypothese H_0 - d.h. die Modellannahme über die p_j - zutrifft, dann ist

$$\tilde{\chi}^2 = \sum_{j=1}^q \frac{(N_j - n \hat{p}_j)^2}{n \hat{p}_j} = \left(\frac{1}{n} \sum_{j=1}^q \frac{N_j^2}{\hat{p}_j} \right) - n \quad \text{asymptotisch } \chi^2(q - 1 - k)\text{-verteilt!}$$

Man beachte die Anzahl der Freiheitsgrade = $q - 1$ - Anzahl geschätzter Parameter!
 Vorsicht: Dieser wichtige Satz gilt im Allgemeinen nicht, wenn die p_j auf eine andere Weise geschätzt werden.

Im folgenden wichtigen Beispiel können die Likelihood-Schätzer praktisch ohne Rechnung erhalten werden:

3.4.3 Der χ^2 -Test auf Unabhängigkeit zweier diskreter Merkmale

Gegeben sind zwei diskrete Merkmale X mit den möglichen Ausprägungen x_1, \dots, x_r und Y mit den Ausprägungen y_1, \dots, y_s . In einer Zufallsstichprobe mit n Realisierungen (Beobachtungen) (x, y) von (X, Y) sei N_{ij} die Häufigkeit für das Ereignis $A_i \cap B_j = \{X = x_i, Y = y_j\}$. Diese Häufigkeiten lassen sich in einer $r \times s$ -Mehrfeldertafel zusammenfassen (auch Kontingenztafel genannt).

	y_1	\dots	y_s	Σ
x_1	N_{11}	\dots	N_{1s}	$N_{1.}$
\vdots	\vdots		\vdots	\vdots
x_r	N_{r1}	\dots	N_{rs}	$N_{r.}$
Σ	$N_{.1}$	\dots	$N_{.s}$	n

$N_{.1}, \dots, N_{.s}$ und $N_{1.}, \dots, N_{r.}$ zufällig.
 Es soll die Nullhypothese der Unabhängigkeit von X und Y anhand der Stichprobenhäufigkeiten N_{ij} getestet werden.

Mit den unbekanntem Wahrscheinlichkeiten $p_{ij} := P\{X = x_i, Y = y_j\}$ und den Marginalwahrscheinlichkeiten $p_{i.} = P\{X = x_i\}$ und $p_{.j} = P\{Y = y_j\}$ ist die Nullhypothese äquivalent zu H_0 : für alle p_{ij} gilt: $p_{ij} = p_{i.} p_{.j}$.
 Die Alternativhypothese H_1 ist die Negation von H_0 . Unter H_0 sind die Wahrscheinlichkeiten p_{ij} Funktionen der $r - 1 + s - 1$ voneinander unabhängigen Parameter $p_{i.}, i < r$ und $p_{.j}, j < s$. Man beachte, dass $p_{r.} = 1 - \sum_{i=1}^{r-1} p_{i.}$ und $p_{.s} = 1 - \sum_{j=1}^{s-1} p_{.j}$.

Unter H_0 stimmen die Likelihoodschätzer $\hat{p}_{i.}, \hat{p}_{.j}$ mit den plausiblen Schätzwerten $\hat{p}_{i.} = \frac{N_{i.}}{n}, \hat{p}_{.j} = \frac{N_{.j}}{n}$ überein. Damit hat man unter H_0 die Schätzer $n \hat{p}_{ij} = n \hat{p}_{i.} \hat{p}_{.j} = \frac{1}{n} N_{i.} N_{.j} = \frac{N_{i.} N_{.j}}{n}$ und die Prüfgröße (= Teststatistik)

$$\begin{aligned} \tilde{\chi}^2 &= \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - n \hat{p}_{ij})^2}{n \hat{p}_{ij}} = n \sum_{i=1}^r \sum_{j=1}^s \frac{\left(N_{ij} - \frac{N_{i.} N_{.j}}{n} \right)^2}{N_{i.} N_{.j}} = \\ &= \left(n \sum_{i=1}^r \sum_{j=1}^s \frac{N_{ij}^2}{N_{i.} N_{.j}} \right) - n \quad \text{mit } r s - 1 - (r - 1) - (s - 1) = (r - 1)(s - 1) \text{ FG'n.} \end{aligned}$$

⇒ Entscheidungsregel für den χ^2 -Test auf Unabhängigkeit zum asymptotischen Niveau α :

$$H_0 \text{ wird abgelehnt, wenn } \tilde{\chi}^2 \geq \chi^2(1 - \alpha; (r - 1)(s - 1)).$$

Bemerkung: Die Approximation durch die χ^2 -Verteilung ist brauchbar, wenn alle $N_{ij} \geq 5$

Beispiel:

X := Tätigkeitskategorie mit den Ausprägungen x_1 = „Beamter“, x_2 = „Angestellter“, x_3 = „Selbstständiger“,
 Y := Zufriedenheit von Informatikern mit ihrem Arbeitsplatz mit den Ausprägungen y_1 = „weniger zufrieden“,
 y_2 = „zufrieden“, y_3 = „sehr zufrieden“

Zufallsstichprobe:

	y_1	y_2	y_3	Σ
x_1	42	50	34	126
x_2	74	95	35	204
x_3	16	20	24	60
Σ	132	165	93	390

$$\alpha = 0,01, \chi^2(0,99; 4) = 13,28$$

$$\tilde{\chi}^2 = 14,4159 > 13,28$$

⇒ Der Test ist signifikant auf dem 1%-Niveau. Mit mindestens 99%-iger Sicherheit sind die Merkmale X und Y nicht unabhängig!

Für Zeilenanzahl $r = 2$ oder Spaltenanzahl $s = 2$ rechnerisch bequeme Umformungen von $\tilde{\chi}^2$:

$$\text{Falls } s = 2: \quad \tilde{\chi}^2 = \frac{n^2}{N_{.1} N_{.2}} \left(\left(\sum_{i=1}^r \frac{N_{i1}^2}{N_{i.}} \right) - \frac{N_{.1}^2}{n} \right) \quad \text{mit } r - 1 \text{ FG'n.}$$

$$\text{Falls } r = 2: \quad \tilde{\chi}^2 = \frac{n^2}{N_{1.} N_{2.}} \left(\left(\sum_{j=1}^s \frac{N_{1j}^2}{N_{.j}} \right) - \frac{N_{1.}^2}{n} \right) \quad \text{mit } s - 1 \text{ FG'n.}$$

Für eine Vierfeldertafel, also $r = s = 2$ gibt weitere Vereinfachung die in 3.2.6 auf Seite 64 angegebene Formel für $\tilde{\chi}^2$ zum Vergleich zweier Wahrscheinlichkeiten.

Rechnerisch in gleicher Weise wie der χ^2 -Test für Unabhängigkeit kann der χ^2 -Homogenitätstest zum Vergleich r unabhängiger Stichproben mit $M(n_i; p_{i1}, \dots, p_{is})$ -verteilten Häufigkeiten N_{i1}, \dots, N_{is} durchgeführt werden. Es ist $n_i = \sum_{j=1}^s N_{ij}$ der Umfang der i -ten Stichprobe und $E(N_{ij}) = n_i p_{ij}$, $i = 1, \dots, r$.

Stichprobe Nr.:	y_1	...	y_j	...	y_s	Σ
1						
⋮			⋮			⋮
i		...	N_{ij}	...		n_i
⋮			⋮			⋮
r						
Σ	...		$N_{.j}$...		n

n_i : feste vorgegebene Stichprobenumfänge, $N_{.j}$ zufällig

Zu testen ist die Nullhypothese (Homogenitätshypothese)

$$H_0: \text{ Für alle Ausprägungen } y_j \text{ von } Y \text{ gilt: } p_{1j} = p_{2j} = \dots = p_{rj} \text{ gegen die Negation } H_1 \text{ von } H_0$$

Prüfgröße ist wieder die unter H_0 asymptotisch $\chi^2((r - 1)(s - 1))$ -verteilte Teststatistik

$$\tilde{\chi}^2 = n \left(\sum_{i=1}^r \sum_{j=1}^s \frac{N_{ij}^2}{n_i N_{.j}} \right) - n.$$

3.5 Nichtparametrische Verfahren - Rangtests

3.5.1 Rangtests

Die bisherigen Testverfahren und insbesondere ihre Irrtumswahrscheinlichkeiten waren nur mathematisch exakt unter Annahme bestimmter - noch von unbekanntem Parameterwerten abhängigen - Verteilungen für die Merkmale (Zufallsgrößen) in den Stichproben, z. B.: X ist $N(\mu; \sigma^2)$ -verteilt mit unbekanntem μ und σ . Realistisch ist aber, dass die Verteilung eines Stichprobenmerkmals X meist völlig unbekannt ist und ein bestimmter Verteilungstyp wie z.B. Normal- oder Exponentialverteilung nur eine Näherung darstellt. Demgegenüber gelten die sogenannten nichtparametrischen (oder „verteilungsfreien“) Testverfahren der Statistik für beliebige - meist nur als stetig vorausgesetzte - Verteilungsfunktionen der Stichprobenwerte, so dass man über deren Verteilung nichts wissen muss. Eine sehr große Klasse nichtparametrischer Tests sind die sogenannten Rangtests. Bei diesen werden nicht die Originalwerte der Stichproben analysiert, sondern nur deren Rangwerte, was einen leichten Informationsverlust zur Folge hat.

Definition:

Ist Y_1, \dots, Y_n eine Zufallsstichprobe mit n verschiedenen Werten einer Zufallsgröße Y , so ist der Rang von Y_i definiert durch

$$R_i = \text{Rg}(Y_i) := \text{Anzahl der Stichprobenwerte } Y_j \leq Y_i.$$

Der kleinste Stichprobenwert erhält also den Rang 1 und der größte den Rang n . Falls in der Stichprobe mehrere identische Werte vorkommen (sogenannte Bindungen, engl: ties), werden mittlere Ränge (midranks) vergeben:

Beispiel:

Stichprobe	4,7	5,4	5,4	6,1
Ränge	1	2,5	2,5	4

also 2,5 als Mittelwert der Ränge 2 und 3.

Man beachte, dass bei stetigen Verteilungen Bindungen nur durch die beschränkte Meßgenauigkeit reeller Merkmalswerte auftreten können. Da in einer Zufallsstichprobe alle Werte Y_i aus der gleichen Verteilung stammen, haben alle R_i eine diskrete Gleichverteilung auf ihrem Wertebereich $\{1, 2, \dots, n\}$, d.h. $P\{R_i = r\} = \frac{1}{n}$ für $r = 1, \dots, n$. Folglich ist der Erwartungswert

$$E(R_i) = \frac{1}{n} \sum_{r=1}^n r = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

$$\sigma^2 := \text{Var}(R_i) = E(R_i^2) - (E(R_i))^2 = \frac{1}{n} \sum_{r=1}^n r^2 - \left(\frac{n+1}{2}\right)^2 = \frac{1}{6}(2n+1)(n+1) - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}.$$

Die Kovarianz $\sigma_{ij} = \text{Kov}(R_i, R_j) = \rho \sigma^2$ ergibt sich wegen $\sum_{i=1}^n R_i = \frac{n(n+1)}{2}$ aus $\text{Var}(\sum_{i=1}^n R_i) = 0 = n\sigma^2 + n(n-1)\rho\sigma^2$ leicht als

$$\sigma_{ij} = -\frac{\sigma^2}{n-1} = -\frac{n+1}{12},$$

$$\text{Korr}(R_i, R_j) = \rho = -\frac{1}{n-1}.$$

Beispiel für einen Rangtest:

3.5.2 Wilcoxon-Zwei-Stichproben-Rangtest zum Vergleich zweier unabhängiger Stichproben

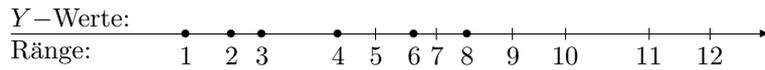
(auch: Mann-Whitney-Test)

Es handelt sich also um ein nichtparametrisches „Gegenstück“ zum t -Test für zwei unabhängige Stichproben.

Gegeben sind zwei unabhängige Stichproben $\mathcal{S}_1 = \{X_{1,1}, \dots, X_{1,n_1}\}$ und $\mathcal{S}_2 = \{X_{2,1}, \dots, X_{2,n_2}\}$ des stetig verteilten Merkmals X , die unter verschiedenen Bedingungen B_1 und B_2 gewonnen werden. Es soll getestet werden, ob sich die Verteilung von X unter B_2 gegenüber der Verteilung unter B_1 um einen Wert δ verschoben hat. Ist $F_1(x)$ die Verteilungsfunktion (VF) von X unter B_1 , so ist $F_2(x) = F_1(x - \delta)$.

Zu testen ist also $H_0 : „\delta = 0“$. gegen $H_1^+ : „\delta > 0“$ oder zweiseitig gegen $H_1 : „\delta \neq 0“$. Bilde hierzu die Gesamtstichprobe $\{Y_1, \dots, Y_n\}$ aus den insgesamt $n = n_1 + n_2$ Einzelwerten und bestimme deren Ränge $R_k = \text{Rg}(Y_k)$ $k = 1, \dots, n$. Als Prüfgröße kann man die Summe $S_2 := \sum_{Y \in \mathcal{S}_2} R_k$ verwenden, also die Summe der Ränge derjenigen Y -Werte, die aus der zweiten Stichprobe stammen oder auch die entsprechende Summe S_1 .

Beispiel:



$$S_2 = 5 + 7 + 9 + 10 + 11 + 12 = 54.$$

Offenbar sprechen auffallend große Werte von S_2 (oder äquivalent hierzu: auffallend kleine Werte von S_1) für die Alternative H_1^+ .

Unter H_0 hat S_i den Erwartungswert $n_i \frac{n+1}{2}$, $i = 1, 2$, und man kann statt S_i auch die zentrierten Prüfgrößen

$$T_i := S_i - n_i \frac{n+1}{2}$$

verwenden, die unter H_0 symmetrisch um 0 verteilt sind. Wegen $S_1 + S_2 = n \frac{n+1}{2}$ gilt stets: $T_1 + T_2 = 0$. Die exakte (diskrete) Verteilung von $T := T_2$ hängt nur von n_1 und n_2 ab. Sie ist in Programmen verfügbar und α -Schranken $w(\alpha; n_1, n_2)$ sind tabelliert. Damit hat man folgende Entscheidungsregeln für obige Tests zum Niveau α :

$$H_0 \text{ abgelehnt zugunsten von } \begin{cases} H_1^+ (\delta > 0), & \text{wenn } T_2 \geq w(\alpha; n_1, n_2) \\ H_1 (\delta \neq 0), & \text{wenn } |T_2| \geq w(\frac{\alpha}{2}; n_1, n_2). \end{cases}$$

Für $n \geq 30$ und $\min(n_1, n_2) \geq 10$ ist die standardisierte Prüfgröße

$$T^* := \frac{T_2}{\sqrt{\text{Var}(T_2)}} = \frac{S_2 - n_2 \frac{n+1}{2}}{\sqrt{n_1 n_2 \frac{n+1}{12}}} \times \text{ unter } H_0 \approx N(0; 1)\text{-verteilt.}$$

T^* ist asymptotisch normalverteilt. Dann folgt mit den Schranken z_α der $N(0; 1)$ -Verteilung:

$$H_0 \text{ abgelehnt zugunsten von } \begin{cases} H_1^+, & \text{wenn } T^* \geq z_\alpha \\ H_1, & \text{wenn } |T^*| \geq z_{\alpha/2}. \end{cases}$$

Zahlenbeispiel: $n_1 = n_2 = 20$, $n = n_1 + n_2 = 40$, $S_2 = 478$, $\alpha = 0,05$:

$$T^* = \frac{478 - 20 \frac{41}{2}}{\sqrt{20 \cdot 20 \cdot \frac{41}{12}}} = 1,8394 > z_{0,05} = 1,645$$

\Rightarrow Der Wilcoxon-Test ist signifikant auf dem 5%-Niveau, also Entscheidung für eine Verschiebung der Verteilung von X unter „Bedingung B_2 “ um einen Wert $\delta > 0$.

Der Mann-Whitney-Test sieht nur formal anders aus, ist aber äquivalent zum Wilcoxon Zwei-Stichproben-Rangtest.

Hinweis: Zum Vergleich verbundener Stichproben dient der Wilcoxon-Vorzeichen-Rangtest. Hierbei geht man wie beim entsprechenden t -Test wieder zur Stichprobe der Differenzen D_1, \dots, D_n über, wobei etwaige Nullen bereits weggelassen seien. Man bildet die Ränge $R_i = \text{Rg}(|D_i|)$, $i = 1, \dots, n$ und die Prüfgröße

$$S_+ := \sum_{i=1}^n V_i R_i \text{ mit } V_i = \begin{cases} 1, & D_i > 0 \\ 0, & D_i < 0 \end{cases}$$

^xBei Bindungen ist die Formel für $\text{Var}(T_2)$ zu korrigieren, was in Programmen berücksichtigt wird.

oder die unter $H_0 : „\delta = E(D) = 0“$ zentrierte Prüfgröße $T_+ = S_+ - \frac{n(n+1)}{4}$, für deren Verteilung unter H_0 wieder entsprechend nur von α und n abhängige Schrankenwerte $c(\alpha; n)$ verfügbar sind. ^{XI}

3.5.3 Grenzwertsätze von Kolmogoroff und Smirnow

Weitere nichtparametrische Verfahren ergeben sich aus den Grenzwertsätzen von Kolmogoroff und Smirnow über die empirischen Verteilungsfunktionen von Stichproben:

Gegeben sei eine Zufallsstichprobe X_1, \dots, X_n eines Merkmals X mit einer stetigen (unbekannten) Verteilungsfunktion F . Die empirische VF. der Stichprobe: $F_n(x) :=$ relative Häufigkeit der Stichprobenwerte $\leq x$, ist eine zufallsabhängige Funktion von x , die alle Informationen der Stichprobe enthält und die für wachsende n gegen die VF $F(n)$ strebt. Genauer gilt der Fundamentalsatz für empirische Verteilungsfunktionen von Gliwienko:

$$P \left\{ \lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0 \right\} = 1$$

Es liegt also mit Sicherheit gleichmäßige Konvergenz von F_n gegen F vor. Die Frage nach der Geschwindigkeit dieser Konvergenz wird durch die folgenden Sätze über die Grenzverteilungen der Zufallsgrößen

$$D_n := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \quad \text{und} \quad D_n^+ := \sup_{x \in \mathbb{R}} (F_n(x) - F(x))$$

beantwortet:

Satz von Kolmogoroff:

Für jede stetige VF. $F(x)$ gilt:

$$\lim_{n \rightarrow \infty} P \{ \sqrt{n} D_n \leq y \} = \begin{cases} K(y) := \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 y^2}, & y > 0 \\ 0 & , y \leq 0 \end{cases}$$

Die durch die VF $K(y)$ gegebene Grenzverteilung ist die Kolmogoroff-Verteilung.

Satz von Smirnow:

Für jede stetige VF. $F(x)$ gilt:

$$\lim_{n \rightarrow \infty} P \{ \sqrt{n} D_n^+ \leq y \} = \begin{cases} 1 - e^{-2y^2}, & y > 0 \\ 0 & , y \leq 0 \end{cases}$$

Von Interesse ist ferner der Vergleich zweier unabhängiger Stichproben anhand ihrer empirischen VF'n $F_m(x)$ und $G_n(x)$, die für wachsende Stichprobenumfänge m, n gegen stetige VF'n F bzw. G streben.

Es sei

$$D_{mn} := \max_{x \in \mathbb{R}} |F_m(x) - G_n(x)|,$$

$$D_{mn}^+ := \max_{x \in \mathbb{R}} (F_m(x) - G_n(x))$$

(Das Maximum wird an einem der Stichprobenwerte angenommen.)

Weitere Grenzwertsätze von Smirnow:

Für stetige VF'n $G = F$ gilt:

$$\lim_{m, n \rightarrow \infty} P \left\{ \sqrt{\frac{m \cdot n}{m+n}} D_{mn} \leq y \right\} = \begin{cases} K(y), & y > 0 \\ 0 & , y \leq 0 \end{cases}$$

$$\lim_{m, n \rightarrow \infty} P \left\{ \sqrt{\frac{m \cdot n}{m+n}} D_{mn}^+ \leq y \right\} = \begin{cases} 1 - e^{-2y^2}, & y > 0 \\ 0 & , y \leq 0 \end{cases}$$

XI

$$T_+^* := \frac{T_+}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad \text{unter } H_0 \text{ asymptotisch } \sim N(0; 1).$$

Man beachte, dass in diesen vier Grenzwertsätzen die Grenzverteilung nicht von der speziellen VF. $F(x)$ abhängt. Dies gilt auch für die exakten Verteilungen von D_n , D_n^+ , D_{mn} und D_{mn}^+ , die nur von n bzw. m und n abhängen und die in Programmen oder Tabellen verfügbar sind. Für Strichprobenumfänge $n \geq 40$ ($m \geq 40$) kann bereits die Grenzverteilung mit guter Näherung verwendet werden.

3.5.4 Anwendungsbeispiele

1. Goodness of Fit - Test für eine vorgegebene spezielle VF $F_0(x)$:

Zu testen ist anhand einer Zufallsstichprobe vom Umfang n aus einer Verteilung mit unbekannter stetiger VF $F(x)$:

$$H_0 : F(x) = F_0(x) \text{ für alle } x \in \mathbb{R} \text{ gegen die Negation } H_1 \text{ von } H_0.$$

Berechne hierzu $D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|$, und vergleiche D_n mit der oberen α -Schranke der Verteilung von D_n unter H_0 . Zu große Werte von D_n führen zur Ablehnung von H_0 . Für $n \geq 40$ hat man mit den $(1 - \alpha)$ -Quantilen $k_{1-\alpha}$ der Kolmogoroff-Verteilung die Entscheidungsregel zum Niveau α :

$$H_0 \text{ wird abgelehnt, wenn } \sqrt{n} \cdot D_n \geq k_{1-\alpha}.$$

2. Anwendung auf die Genauigkeit der Näherung an eine mathematisch schwer oder gar nicht berechenbare stetige Verteilungsfunktion mittels Simulation:

Oft ist die Zufallsgröße X , deren VF. F gebraucht wird, selbst eine Funktion mehrerer evtl. sogar abhängiger Zufallsgrößen Y_1, Y_2, \dots , so dass eine mathematische Berechnung von F (zumindest praktisch) nicht möglich ist. Ein Beispiel ist z.B. die VF. F der multivariaten Spannweite $X = X_{k,d} := \max_{1 \leq i < j \leq k} |\vec{Z}_j - \vec{Z}_i|$, wobei $\vec{Z}_1, \dots, \vec{Z}_k$ unabhängige $N_d(\vec{\mu}; \Sigma)$ -verteilte Zufallspunkte in \mathbb{R}^d sind. Auch für unabhängige $N(0; 1)$ -verteilte Koordinaten - also $\vec{\mu} = 0$ und $\Sigma = d \times d$ -Einheitsmatrix - ist $F(x)$ nicht exakt berechenbar. Man kann aber durch Simulation im Rechner normalverteilte Pseudozufallsgrößen erzeugen und damit auch in einer Schleife n Realisierungen von X erzeugen. Für ein „Wertegitter“ mit Werten x_i wird dabei die Häufigkeit der X -Werte $\leq x_i$, $i = 1, 2, \dots$ gezählt, also die empirische VF. $F_n(x)$ an den (hinreichend dicht liegenden) Stellen x_i ermittelt. Wie groß muss n gewählt werden, damit $D_n = \sup_x |F_n(x) - F(x)|$ mit mindestens 99%-iger Sicherheit $\leq 10^{-4}$ ausfällt? Dies wäre eine Mindestanforderung an eine Wertetabelle mit den Werten $F_n(x_i)$ statt $F(x_i)$. (Zur Simulation von $N(0; 1)$ -Größen siehe Kapitel 3.5.6 auf Seite 79.)

Es soll also gelten:

$$P\{D_n \leq 10^{-4}\} \geq 0,99.$$

Andererseits ist:

$$P\{\sqrt{n}D_n \leq k_{0,99}\} = 0,99.$$

Mit dem tabellierten Wert $k_{0,99} = 1,63$ findet man daher $\frac{k_{0,99}}{\sqrt{n}} \leq 10^{-4}$, also $n \geq (10^4 \cdot 1,63)^2 = 2,6569 \cdot 10^8$.

3. Vergleich zweier unbekannter stetiger Verteilungsfunktionen F und G anhand zweier unabhängiger Stichproben aus diesen Verteilungen vom Umfang m bzw. n mit den empirischen VF'n. F_m und G_n :

Zu testen: $H_0 : G = F$ gegen die Negation $H_1 : G \neq F$.

Hierzu ist die Prüfgröße D_{mn} mit der oberen α -Schranke (= $(1 - \alpha)$ -Quantil) ihrer Verteilung unter H_0 zu vergleichen. Überschreitungen dieser tabellierten Schranken führen zur Ablehnung von H_0 . Für größere Stichproben erhält man wieder mit den Schranken $k_{1-\alpha}$ die Entscheidungsregel zum asymptotischen Niveau α :

$$H_0 \text{ wird abgelehnt, falls } \sqrt{\frac{m \cdot n}{m+n}} D_{mn} \geq k_{1-\alpha}.$$

Für den Test gegen eine einseitige Alternative wird eine allgemeinere Alternative als die „Verschiebungsalternative“ beim Wilcoxon-Test betrachtet.

Definition:

Sind F und G die VF'n. der Zufallsgrößen X und Y , so heißt Y stochastisch größer als X , wenn $G(x) \leq F(x)$ für alle $x \in \mathbb{R}$ und $G(x) < F(x)$ dabei vorkommt. (kurz: $Y \overset{P}{>} X$) Es gilt also für alle $x \in \mathbb{R}$:

$$P\{Y > x\} \geq P\{X > x\} \quad \text{mit „>“ für gewisse } x\text{-Werte.}$$

Zu testen ist $H_0 : G = F$ gegen $H_1^+ : Y \overset{P}{>} X$. Entscheidungsregel mit den $(1 - \alpha)$ -Quantilen $y_{1-\alpha} = \sqrt{\frac{1}{2} \ln(\frac{1}{\alpha})}$ der VF. $1 - e^{-2y^2}$ zum asymptotischen Niveau α :

$$H_0 \text{ abgelehnt zugunsten } H_1^+, \text{ falls } \sqrt{\frac{m \cdot n}{m + n}} D_{mn}^+ \geq y_{1-\alpha}.$$

3.5.5 Test auf Normalverteilung nach Lilliefors

Oft geht es beim Goodness of fit - Test nicht um eine einzelne vollständig bestimmte VF. $F_0(x)$ sondern um den Typ der Verteilung, z.B. die Nullhypothese

H_0 : Das Merkmal X hat eine $N(\mu; \sigma^2)$ -Verteilung mit unbekanntem μ und σ ,
d.h. irgendeine VF. der Form $F(x) = \Phi(\frac{x-\mu}{\sigma})$.

Setzt man hier aus einer vorliegenden Stichprobe x_1, \dots, x_n für X die üblichen Schätzwerte \bar{x} und s für μ und σ ein, so gilt für

$$\tilde{D}_n := \sup_{x \in \mathbb{R}} \left| F_n(x) - \Phi\left(\frac{x - \bar{x}}{s}\right) \right| \quad \text{und } n \rightarrow \infty$$

nicht mehr die kolmogoroff'sche Grenzverteilung! Lilliefors hat für die Verteilung von \tilde{D}_n in Abhängigkeit von α und n $(1 - \alpha)$ -Quantile approximiert und tabelliert und auch $(1 - \alpha)$ -Quantile $l_{1-\alpha}$ für die Grenzverteilung von $\sqrt{n}\tilde{D}_n$ für $n \rightarrow \infty$, die bereits für $n > 35$ verwendet werden können. Mit diesen Schranken $l_{1-\alpha}$ erhält man die Entscheidungsregel zum asymptotischen Niveau α :

H_0 wird abgelehnt, wenn $\sqrt{n}\tilde{D}_n \geq l_{1-\alpha}$.

<u>$(1 - \alpha)$-Quantile nach Lilliefors:</u>	α	0,05	0,01	
	$l_{1-\alpha}$	0,886	1,031	
<u>$(1 - \alpha)$-Quantile der Kolmogoroff-Verteilung:</u>	α	0,05	0,01	0,001
	$k_{1-\alpha}$	1,36	1,63	1,95

Hinweis:

Hält man als Maß für den Abstand zwischen F_n und F nicht das Supremum von $|F_n(x) - F(x)|$ sondern das „mittlere Abstandsquadrat“, so erhält man noch empfindlichere Tests (d.h. mit höherer Power) vom „Cramer-v. Mises-Typ“. Eine typische Testgröße dieser Art ist

$$\omega_n^2 := \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x).$$

Die Grenzverteilung von $n\omega_n^2$ für $n \rightarrow \infty$ unter H_0 (d.h. falls $F_n \rightarrow F$) ist bekannt und damit auch die entsprechenden $(1 - \alpha)$ -Quantile dieser Grenzverteilung, die für asymptotische Tests von H_0 verwendet werden können.

3.5.6 Zur Simulation von $N(0; 1)$ -Zufallsgrößen

Viele Computerprogramme können $R(0; 1)$ -verteilte Pseudozufallszahlen U generieren.

$$U \sim R(0; 1) \quad \Rightarrow \quad Z := \Phi^{-1}(U) \sim N(0; 1).$$

Leider ist die Rechenzeit für die Umkehrfunktion Φ^{-1} der $N(0; 1)$ -VF. Φ bei umfangreichen Simulationen viel zu lang. Auf folgende Art lassen sich aus Paaren (U_1, U_2) unabhängiger $R(0; 1)$ -Größen rasch Paare (Z_1, Z_2)

unabhängiger $N(0; 1)$ -Größen gewinnen:

$R^2 := Z_1^2 + Z_2^2$ hat eine $\chi^2(2)$ -Verteilung, deren Dichte mit der Exponentialdichte $f(x) = \frac{1}{2}e^{-x/2}$ übereinstimmt.

Nach dem der $R(0; 1)$ -Verteilung folgenden Beispiel in Kapitel 2.1.4 auf Seite 35 ist dann $\sqrt{-2\ln(U_1)}$ verteilt wie R . Die Polarkoordinaten (R, Φ) des zufälligen Punkts mit den kartesischen Koordinaten (Z_1, Z_2) hat man dann mit dem in $(0; 2\pi)$ gleichverteilten Winkel $\Phi = 2\pi U_2$. Somit erhält man mit $R := \sqrt{-2\ln(U_1)}$ und Φ :

$$Z_1 = R \cos \Phi, \quad Z_2 = R \sin \Phi$$

A Anhang

A.1 Tabelle der Standardnormalverteilung

z	$\phi(z)$	$\phi(-z)$	$D(z)$	z	$\phi(z)$	$\phi(-z)$	$D(z)$	z	$\phi(z)$	$\phi(-z)$	$D(z)$
0,01	0,50399	0,49601	0,00798	0,51	0,69497	0,30503	0,38994	1,01	0,84375	0,15625	0,68750
0,02	0,50798	0,49202	0,01596	0,52	0,69847	0,30153	0,39694	1,02	0,84614	0,15386	0,69228
0,03	0,51197	0,48803	0,02394	0,53	0,70194	0,29806	0,40388	1,03	0,84849	0,15151	0,69698
0,04	0,51595	0,48405	0,03190	0,54	0,70540	0,29460	0,41080	1,04	0,85083	0,14917	0,70166
0,05	0,51994	0,48006	0,03988	0,55	0,70884	0,29116	0,41768	1,05	0,85314	0,14686	0,70628
0,06	0,52392	0,47608	0,04784	0,56	0,71226	0,28774	0,42452	1,06	0,85543	0,14457	0,71086
0,07	0,52790	0,47210	0,05580	0,57	0,71566	0,28434	0,43132	1,07	0,85769	0,14231	0,71538
0,08	0,53188	0,46812	0,06376	0,58	0,71904	0,28096	0,43808	1,08	0,85993	0,14007	0,71986
0,09	0,53586	0,46414	0,07172	0,59	0,72240	0,27760	0,44480	1,09	0,86214	0,13786	0,72428
0,10	0,53983	0,46017	0,07966	0,60	0,72575	0,27425	0,45150	1,10	0,86433	0,13567	0,72866
0,11	0,54380	0,45620	0,08760	0,61	0,72907	0,27093	0,45814	1,11	0,86650	0,13350	0,73300
0,12	0,54776	0,45224	0,09552	0,62	0,73237	0,26763	0,46474	1,12	0,86864	0,13136	0,73728
0,13	0,55172	0,44828	0,10344	0,63	0,73565	0,26435	0,47130	1,13	0,87076	0,12924	0,74152
0,14	0,55567	0,44433	0,11134	0,64	0,73891	0,26109	0,47782	1,14	0,87286	0,12714	0,74572
0,15	0,55962	0,44038	0,11924	0,65	0,74215	0,25785	0,48430	1,15	0,87493	0,12507	0,74986
0,16	0,56356	0,43644	0,12712	0,66	0,74537	0,25463	0,49074	1,16	0,87698	0,12302	0,75396
0,17	0,56749	0,43251	0,13498	0,67	0,74857	0,25143	0,49714	1,17	0,87900	0,12100	0,75800
0,18	0,57142	0,42858	0,14284	0,68	0,75175	0,24825	0,50350	1,18	0,88100	0,11900	0,76200
0,19	0,57535	0,42465	0,15070	0,69	0,75490	0,24510	0,50980	1,19	0,88298	0,11702	0,76596
0,20	0,57926	0,42074	0,15852	0,70	0,75804	0,24196	0,51608	1,20	0,88493	0,11507	0,76986
0,21	0,58317	0,41683	0,16634	0,71	0,76115	0,23885	0,52230	1,21	0,88686	0,11314	0,77372
0,22	0,58706	0,41294	0,17412	0,72	0,76424	0,23576	0,52848	1,22	0,88877	0,11123	0,77754
0,23	0,59095	0,40905	0,18190	0,73	0,76730	0,23270	0,53460	1,23	0,89065	0,10935	0,78130
0,24	0,59483	0,40517	0,18966	0,74	0,77035	0,22965	0,54070	1,24	0,89251	0,10749	0,78502
0,25	0,59871	0,40129	0,19742	0,75	0,77337	0,22663	0,54674	1,25	0,89435	0,10565	0,78870
0,26	0,60257	0,39743	0,20514	0,76	0,77637	0,22363	0,55274	1,26	0,89617	0,10383	0,79234
0,27	0,60642	0,39358	0,21284	0,77	0,77935	0,22065	0,55870	1,27	0,89796	0,10204	0,79592
0,28	0,61026	0,38974	0,22052	0,78	0,78230	0,21770	0,56460	1,28	0,89973	0,10027	0,79946
0,29	0,61409	0,38591	0,22818	0,79	0,78524	0,21476	0,57048	1,29	0,90147	0,09853	0,80294
0,30	0,61791	0,38209	0,23582	0,80	0,78814	0,21186	0,57628	1,30	0,90320	0,09680	0,80640
0,31	0,62172	0,37828	0,24344	0,81	0,79103	0,20897	0,58206	1,31	0,90490	0,09510	0,80980
0,32	0,62552	0,37448	0,25104	0,82	0,79389	0,20611	0,58778	1,32	0,90658	0,09342	0,81316
0,33	0,62930	0,37070	0,25860	0,83	0,79673	0,20327	0,59346	1,33	0,90824	0,09176	0,81648
0,34	0,63307	0,36693	0,26614	0,84	0,79955	0,20045	0,59910	1,34	0,90988	0,09012	0,81976
0,35	0,63683	0,36317	0,27366	0,85	0,80234	0,19766	0,60468	1,35	0,91149	0,08851	0,82298
0,36	0,64058	0,35942	0,28116	0,86	0,80511	0,19489	0,61022	1,36	0,91309	0,08691	0,82618
0,37	0,64431	0,35569	0,28862	0,87	0,80785	0,19215	0,61570	1,37	0,91466	0,08534	0,82932
0,38	0,64803	0,35197	0,29606	0,88	0,81057	0,18943	0,62114	1,38	0,91621	0,08379	0,83242
0,39	0,65173	0,34827	0,30346	0,89	0,81327	0,18673	0,62654	1,39	0,91774	0,08226	0,83548
0,40	0,65542	0,34458	0,31084	0,90	0,81594	0,18406	0,63188	1,40	0,91924	0,08076	0,83848
0,41	0,65910	0,34090	0,31820	0,91	0,81859	0,18141	0,63718	1,41	0,92073	0,07927	0,84146
0,42	0,66276	0,33724	0,32552	0,92	0,82121	0,17879	0,64242	1,42	0,92220	0,07780	0,84440
0,43	0,66640	0,33360	0,33280	0,93	0,82381	0,17619	0,64762	1,43	0,92364	0,07636	0,84728
0,44	0,67003	0,32997	0,34006	0,94	0,82639	0,17361	0,65278	1,44	0,92507	0,07493	0,85014
0,45	0,67364	0,32636	0,34728	0,95	0,82894	0,17106	0,65788	1,45	0,92647	0,07353	0,85294
0,46	0,67724	0,32276	0,35448	0,96	0,83147	0,16853	0,66294	1,46	0,92785	0,07215	0,85570
0,47	0,68082	0,31918	0,36164	0,97	0,83398	0,16602	0,66796	1,47	0,92922	0,07078	0,85844
0,48	0,68439	0,31561	0,36878	0,98	0,83646	0,16354	0,67292	1,48	0,93056	0,06944	0,86112
0,49	0,68793	0,31207	0,37586	0,99	0,83891	0,16109	0,67782	1,49	0,93189	0,06811	0,86378
0,50	0,69146	0,30854	0,38292	1,00	0,84134	0,15866	0,68268	1,50	0,93319	0,06681	0,86638

z	$\phi(z)$	$\phi(-z)$	$D(z)$	z	$\phi(z)$	$\phi(-z)$	$D(z)$	z	$\phi(z)$	$\phi(-z)$	$D(z)$
1, 51	0, 93448	0, 06552	0, 86896	2, 01	0, 97778	0, 02222	0, 95556	2, 51	0, 99396	0, 00604	0, 98792
1, 52	0, 93574	0, 06426	0, 87148	2, 02	0, 97831	0, 02169	0, 95662	2, 52	0, 99413	0, 00587	0, 98826
1, 53	0, 93699	0, 06301	0, 87398	2, 03	0, 97882	0, 02118	0, 95764	2, 53	0, 99430	0, 00570	0, 98860
1, 54	0, 93822	0, 06178	0, 87644	2, 04	0, 97932	0, 02068	0, 95864	2, 54	0, 99446	0, 00554	0, 98892
1, 55	0, 93943	0, 06057	0, 87886	2, 05	0, 97982	0, 02018	0, 95964	2, 55	0, 99461	0, 00539	0, 98922
1, 56	0, 94062	0, 05938	0, 88124	2, 06	0, 98030	0, 01970	0, 96060	2, 56	0, 99477	0, 00523	0, 98954
1, 57	0, 94179	0, 05821	0, 88358	2, 07	0, 98077	0, 01923	0, 96154	2, 57	0, 99492	0, 00508	0, 98984
1, 58	0, 94295	0, 05705	0, 88590	2, 08	0, 98124	0, 01876	0, 96248	2, 58	0, 99506	0, 00494	0, 99012
1, 59	0, 94408	0, 05592	0, 88816	2, 09	0, 98169	0, 01831	0, 96338	2, 59	0, 99520	0, 00480	0, 99040
1, 60	0, 94520	0, 05480	0, 89040	2, 10	0, 98214	0, 01786	0, 96428	2, 60	0, 99534	0, 00466	0, 99068
1, 61	0, 94630	0, 05370	0, 89260	2, 11	0, 98257	0, 01743	0, 96514	2, 61	0, 99547	0, 00453	0, 99094
1, 62	0, 94738	0, 05262	0, 89476	2, 12	0, 98300	0, 01700	0, 96600	2, 62	0, 99560	0, 00440	0, 99120
1, 63	0, 94845	0, 05155	0, 89690	2, 13	0, 98341	0, 01659	0, 96682	2, 63	0, 99573	0, 00427	0, 99146
1, 64	0, 94950	0, 05050	0, 89900	2, 14	0, 98382	0, 01618	0, 96764	2, 64	0, 99585	0, 00415	0, 99170
1, 65	0, 95053	0, 04947	0, 90106	2, 15	0, 98422	0, 01578	0, 96844	2, 65	0, 99598	0, 00402	0, 99196
1, 66	0, 95154	0, 04846	0, 90308	2, 16	0, 98461	0, 01539	0, 96922	2, 66	0, 99609	0, 00391	0, 99218
1, 67	0, 95254	0, 04746	0, 90508	2, 17	0, 98500	0, 01500	0, 97000	2, 67	0, 99621	0, 00379	0, 99242
1, 68	0, 95352	0, 04648	0, 90704	2, 18	0, 98537	0, 01463	0, 97074	2, 68	0, 99632	0, 00368	0, 99264
1, 69	0, 95449	0, 04551	0, 90898	2, 19	0, 98574	0, 01426	0, 97148	2, 69	0, 99643	0, 00357	0, 99286
1, 70	0, 95543	0, 04457	0, 91086	2, 20	0, 98610	0, 01390	0, 97220	2, 70	0, 99653	0, 00347	0, 99306
1, 71	0, 95637	0, 04363	0, 91274	2, 21	0, 98645	0, 01355	0, 97290	2, 71	0, 99664	0, 00336	0, 99328
1, 72	0, 95728	0, 04272	0, 91456	2, 22	0, 98679	0, 01321	0, 97358	2, 72	0, 99674	0, 00326	0, 99348
1, 73	0, 95818	0, 04182	0, 91636	2, 23	0, 98713	0, 01287	0, 97426	2, 73	0, 99683	0, 00317	0, 99366
1, 74	0, 95907	0, 04093	0, 91814	2, 24	0, 98745	0, 01255	0, 97490	2, 74	0, 99693	0, 00307	0, 99386
1, 75	0, 95994	0, 04006	0, 91988	2, 25	0, 98778	0, 01222	0, 97556	2, 75	0, 99702	0, 00298	0, 99404
1, 76	0, 96080	0, 03920	0, 92160	2, 26	0, 98809	0, 01191	0, 97618	2, 76	0, 99711	0, 00289	0, 99422
1, 77	0, 96164	0, 03836	0, 92328	2, 27	0, 98840	0, 01160	0, 97680	2, 77	0, 99720	0, 00280	0, 99440
1, 78	0, 96246	0, 03754	0, 92492	2, 28	0, 98870	0, 01130	0, 97740	2, 78	0, 99728	0, 00272	0, 99456
1, 79	0, 96327	0, 03673	0, 92654	2, 29	0, 98899	0, 01101	0, 97798	2, 79	0, 99736	0, 00264	0, 99472
1, 80	0, 96407	0, 03593	0, 92814	2, 30	0, 98928	0, 01072	0, 97856	2, 80	0, 99744	0, 00256	0, 99488
1, 81	0, 96485	0, 03515	0, 92970	2, 31	0, 98956	0, 01044	0, 97912	2, 81	0, 99752	0, 00248	0, 99504
1, 82	0, 96562	0, 03438	0, 93124	2, 32	0, 98983	0, 01017	0, 97966	2, 82	0, 99760	0, 00240	0, 99520
1, 83	0, 96638	0, 03362	0, 93276	2, 33	0, 99010	0, 00990	0, 98020	2, 83	0, 99767	0, 00233	0, 99534
1, 84	0, 96712	0, 03288	0, 93424	2, 34	0, 99036	0, 00964	0, 98072	2, 84	0, 99774	0, 00226	0, 99548
1, 85	0, 96784	0, 03216	0, 93568	2, 35	0, 99061	0, 00939	0, 98122	2, 85	0, 99781	0, 00219	0, 99562
1, 86	0, 96856	0, 03144	0, 93712	2, 36	0, 99086	0, 00914	0, 98172	2, 86	0, 99788	0, 00212	0, 99576
1, 87	0, 96926	0, 03074	0, 93852	2, 37	0, 99111	0, 00889	0, 98222	2, 87	0, 99795	0, 00205	0, 99590
1, 88	0, 96995	0, 03005	0, 93990	2, 38	0, 99134	0, 00866	0, 98268	2, 88	0, 99801	0, 00199	0, 99602
1, 89	0, 97062	0, 02938	0, 94124	2, 39	0, 99158	0, 00842	0, 98316	2, 89	0, 99807	0, 00193	0, 99614
1, 90	0, 97128	0, 02872	0, 94256	2, 40	0, 99180	0, 00820	0, 98360	2, 90	0, 99813	0, 00187	0, 99626
1, 91	0, 97193	0, 02807	0, 94386	2, 41	0, 99202	0, 00798	0, 98404	2, 91	0, 99819	0, 00181	0, 99638
1, 92	0, 97257	0, 02743	0, 94514	2, 42	0, 99224	0, 00776	0, 98448	2, 92	0, 99825	0, 00175	0, 99650
1, 93	0, 97320	0, 02680	0, 94640	2, 43	0, 99245	0, 00755	0, 98490	2, 93	0, 99831	0, 00169	0, 99662
1, 94	0, 97381	0, 02619	0, 94762	2, 44	0, 99266	0, 00734	0, 98532	2, 94	0, 99836	0, 00164	0, 99672
1, 95	0, 97441	0, 02559	0, 94882	2, 45	0, 99286	0, 00714	0, 98572	2, 95	0, 99841	0, 00159	0, 99682
1, 96	0, 97500	0, 02500	0, 95000	2, 46	0, 99305	0, 00695	0, 98610	2, 96	0, 99846	0, 00154	0, 99692
1, 97	0, 97558	0, 02442	0, 95116	2, 47	0, 99324	0, 00676	0, 98648	2, 97	0, 99851	0, 00149	0, 99702
1, 98	0, 97615	0, 02385	0, 95230	2, 48	0, 99343	0, 00657	0, 98686	2, 98	0, 99856	0, 00144	0, 99712
1, 99	0, 97670	0, 02330	0, 95340	2, 49	0, 99361	0, 00639	0, 98722	2, 99	0, 99861	0, 00139	0, 99722
2, 00	0, 97725	0, 02275	0, 95450	2, 50	0, 99379	0, 00621	0, 98758	3, 00	0, 99865	0, 00135	0, 99730

A.2 Tabelle der χ^2 - Verteilung

$F(x)$	Anzahl der Freiheitsgrad									
	1	2	3	4	5	6	7	8	9	10
0,001	0,00	0,00	0,02	0,09	0,21	0,38	0,60	0,86	1,15	1,48
0,005	0,00	0,01	0,07	0,21	0,41	0,68	0,99	1,34	1,73	2,16
0,01	0,00	0,02	0,11	0,30	0,55	0,87	1,24	1,65	2,09	2,56
0,025	0,00	0,05	0,22	0,48	0,83	1,24	1,69	2,18	2,70	3,25
0,05	0,00	0,10	0,35	0,71	1,15	1,64	2,17	2,73	3,33	3,94
0,1	0,02	0,21	0,58	1,06	1,61	2,20	2,83	3,49	4,17	4,87
0,25	0,10	0,58	1,21	1,92	2,67	3,45	4,25	5,07	5,90	6,74
0,5	0,45	1,39	2,37	3,36	4,35	5,35	6,35	7,34	8,34	9,34
0,75	1,32	2,77	4,11	5,39	6,63	7,84	9,04	10,22	11,39	12,55
0,9	2,71	4,61	6,25	7,78	9,24	10,64	12,02	13,36	14,68	15,99
0,95	3,84	5,99	7,81	9,49	11,07	12,59	14,07	15,51	16,92	18,31
0,975	5,02	7,38	9,35	11,14	12,83	14,45	16,01	17,53	19,02	20,48
0,99	6,63	9,21	11,34	13,28	15,09	16,81	18,48	20,09	21,67	23,21
0,995	7,88	10,60	12,84	14,86	16,75	18,55	20,28	21,96	23,59	25,19
0,999	10,83	13,82	16,27	18,47	20,52	22,46	24,32	26,13	27,88	29,59

$F(x)$	Anzahl der Freiheitsgrad									
	11	12	13	14	15	16	17	18	19	20
0,001	1,83	2,21	2,62	3,04	3,48	3,94	4,42	4,90	5,41	5,92
0,005	2,60	3,07	3,57	4,07	4,60	5,14	5,70	6,26	6,84	7,43
0,01	3,05	3,57	4,11	4,66	5,23	5,81	6,41	7,01	7,63	8,26
0,025	3,82	4,40	5,01	5,63	6,26	6,91	7,56	8,23	8,91	9,59
0,05	4,57	5,23	5,89	6,57	7,26	7,96	8,67	9,39	10,12	10,85
0,1	5,58	6,30	7,04	7,79	8,55	9,31	10,09	10,86	11,65	12,44
0,25	7,58	8,44	9,30	10,17	11,04	11,91	12,79	13,68	14,56	15,45
0,5	10,34	11,34	12,34	13,34	14,34	15,34	16,34	17,34	18,34	19,34
0,75	13,70	14,85	15,98	17,12	18,25	19,37	20,49	21,60	22,72	23,83
0,9	17,28	18,55	19,81	21,06	22,31	23,54	24,77	25,99	27,20	28,41
0,95	19,68	21,03	22,36	23,68	25,00	26,30	27,59	28,87	30,14	31,41
0,975	21,92	23,34	24,74	26,12	27,49	28,85	30,19	31,53	32,85	34,17
0,99	24,73	26,22	27,69	29,14	30,58	32,00	33,41	34,81	36,19	37,57
0,995	26,76	28,30	29,82	31,32	32,80	34,27	35,72	37,16	38,58	40,00
0,999	31,26	32,91	34,53	36,12	37,70	39,25	40,79	42,31	43,82	45,32

$F(x)$	Anzahl der Freiheitsgrad									
	21	22	23	24	25	26	27	28	29	30
0,001	6,4	7,0	7,5	8,1	8,7	9,2	9,8	10,4	11,0	11,6
0,005	8,0	8,6	9,3	9,9	10,5	11,2	11,8	12,5	13,1	13,8
0,01	8,9	9,5	10,2	10,9	11,5	12,2	12,9	13,6	14,3	15,0
0,025	10,3	11,0	11,7	12,4	13,1	13,8	14,6	15,3	16,0	16,8
0,05	11,6	12,3	13,1	13,8	14,6	15,4	16,2	16,9	17,7	18,5
0,1	13,2	14,0	14,8	15,7	16,5	17,3	18,1	18,9	19,8	20,6
0,25	16,3	17,2	18,1	19,0	19,9	20,8	21,7	22,7	23,6	24,5
0,5	20,3	21,3	22,3	23,3	24,3	25,3	26,3	27,3	28,3	29,3
0,75	24,9	26,0	27,1	28,2	29,3	30,4	31,5	32,6	33,7	34,8
0,9	29,6	30,8	32,0	33,2	34,4	35,6	36,7	37,9	39,1	40,3
0,95	32,7	33,9	35,2	36,4	37,7	38,9	40,1	41,3	42,6	43,8
0,975	35,5	36,8	38,1	39,4	40,6	41,9	43,2	44,5	45,7	47,0
0,99	38,9	40,3	41,6	43,0	44,3	45,6	47,0	48,3	49,6	50,9
0,995	41,4	42,8	44,2	45,6	46,9	48,3	49,6	51,0	52,3	53,7
0,999	46,8	48,3	49,7	51,2	52,6	54,1	55,5	56,9	58,3	59,7

$F(x)$	Anzahl der Freiheitsgrad								> 100 (Naherung)
	40	50	60	70	80	90	100		
0,001	17,9	24,7	31,7	39,0	46,5	54,2	61,9	$\frac{1}{2}(h - 3,09)^2$	
0,005	20,7	28,0	35,5	43,3	51,2	59,2	67,3	$\frac{1}{2}(h - 2,58)^2$	
0,01	22,2	29,7	37,5	45,4	53,5	61,8	70,1	$\frac{1}{2}(h - 2,33)^2$	
0,025	24,4	32,4	40,5	48,8	57,2	65,6	74,2	$\frac{1}{2}(h - 1,96)^2$	
0,05	26,5	34,8	43,2	51,7	60,4	69,1	77,9	$\frac{1}{2}(h - 1,64)^2$	
0,1	29,1	37,7	46,5	55,3	64,3	73,3	82,4	$\frac{1}{2}(h - 1,28)^2$	
0,25	33,7	42,9	52,3	61,7	71,1	80,6	90,1	$\frac{1}{2}(h - 0,67)^2$	
0,5	39,3	49,3	59,3	69,3	79,3	89,3	99,3	$\frac{1}{2}h^2$	
0,75	45,6	56,3	67,0	77,6	88,1	98,6	109,1	$\frac{1}{2}(h + 0,67)^2$	
0,9	51,8	63,2	74,4	85,5	96,6	107,6	118,5	$\frac{1}{2}(h + 1,28)^2$	
0,95	55,8	67,5	79,1	90,5	101,9	113,1	124,3	$\frac{1}{2}(h + 1,64)^2$	
0,975	59,3	71,4	83,3	95,0	106,6	118,1	129,6	$\frac{1}{2}(h + 1,96)^2$	
0,99	63,7	76,2	88,4	100,4	112,3	124,1	135,8	$\frac{1}{2}(h + 2,33)^2$	
0,995	66,8	79,5	92,0	104,2	116,3	128,3	140,2	$\frac{1}{2}(h + 2,58)^2$	
0,999	73,4	86,7	99,6	112,3	124,8	137,2	149,4	$\frac{1}{2}(h + 3,09)^2$	

Mit $h = \sqrt{2m - 1}$ (m = Freiheitsgrade)

A.3 Tabelle der t -Verteilung

$F(x)$ $1 - \alpha$	$\nu = \text{Anzahl der Freiheitsgrad}$									
	1	2	3	4	5	6	7	8	9	10
0,5	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,6	0,33	0,29	0,28	0,27	0,27	0,27	0,26	0,26	0,26	0,26
0,7	0,73	0,62	0,58	0,57	0,56	0,55	0,55	0,55	0,54	0,54
0,8	1,38	1,06	0,98	0,94	0,92	0,91	0,90	0,89	0,88	0,88
0,9	3,08	1,89	1,64	1,53	1,48	1,44	1,42	1,40	1,38	1,37
0,95	6,31	2,92	2,35	2,13	2,02	1,94	1,90	1,86	1,83	1,81
0,975	12,70	4,30	3,18	2,78	2,57	2,45	2,37	2,31	2,26	2,23
0,99	31,80	6,97	4,54	3,75	3,37	3,14	3,00	2,90	2,82	2,76
0,995	63,70	9,93	5,84	4,60	4,03	3,71	3,50	3,36	3,25	3,17
0,999	318,30	22,30	10,20	7,17	5,89	5,21	4,79	4,50	4,30	4,14

$F(x)$ $1 - \alpha$	$\nu = \text{Anzahl der Freiheitsgrad}$									
	11	12	13	14	15	16	17	18	19	20
0,5	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,6	0,26	0,26	0,26	0,26	0,26	0,26	0,26	0,26	0,26	0,26
0,7	0,54	0,54	0,54	0,54	0,54	0,54	0,53	0,53	0,53	0,53
0,8	0,88	0,87	0,87	0,87	0,87	0,87	0,86	0,86	0,86	0,86
0,9	1,36	1,36	1,35	1,35	1,34	1,34	1,33	1,33	1,33	1,33
0,95	1,80	1,78	1,77	1,76	1,75	1,75	1,74	1,73	1,73	1,73
0,975	2,20	2,18	2,16	2,15	2,13	2,12	2,11	2,10	2,09	2,09
0,99	2,72	2,68	2,65	2,62	2,60	2,58	2,57	2,55	2,54	2,53
0,995	3,11	3,06	3,01	2,98	2,95	2,92	2,90	2,88	2,86	2,85
0,999	4,03	3,93	3,85	3,79	3,73	3,69	3,65	3,61	3,58	3,55

$F(x)$ $1 - \alpha$	$\nu = \text{Anzahl der Freiheitsgrad}$									
	22	24	26	28	30	40	50	100	200	∞
0,5	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,6	0,26	0,26	0,26	0,26	0,26	0,26	0,26	0,25	0,25	0,25
0,7	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,52
0,8	0,86	0,86	0,86	0,86	0,85	0,85	0,85	0,85	0,84	0,84
0,9	1,32	1,32	1,32	1,31	1,31	1,30	1,30	1,29	1,29	1,28
0,95	1,72	1,71	1,71	1,70	1,70	1,68	1,68	1,66	1,65	1,65
0,975	2,07	2,06	2,06	2,05	2,04	2,02	2,01	1,98	1,97	1,96
0,99	2,51	2,49	2,48	2,47	2,46	2,42	2,40	2,37	2,35	2,33
0,995	2,82	2,80	2,78	2,76	2,75	2,70	2,68	2,63	2,60	2,58
0,999	3,51	3,47	3,44	3,41	3,39	3,31	3,26	3,17	3,13	3,09