

INDUKTIVE STATISTIK

Nach

STATISTIK II,

gelesen von

Prof. Dr. Dr. Helge Toutenburg

an der

**WIRTSCHAFTSWISSENSCHAFTLICHEN FAKULTÄT
der
UNIVERSITÄT BASEL**

im

Sommersemester 2001

zusammengefasst von

Daniel Frank

Inhalt

| | |
|-----------------------------------------------------------------------|-----------|
| Teil I: Kombinatorik | 3 |
| Grundbegriffe und Verfahren | 3 |
| Permutation | 4 |
| Variation mit Wiederholung..... | 4 |
| Variation ohne Wiederholung | 4 |
| Kombination ohne Wiederholung | 4 |
| Kombination mit Wiederholung..... | 5 |
| Schlussbemerkungen | 5 |
| | |
| Teil II: Wahrscheinlichkeitstheorie | 7 |
| Einführung | 7 |
| Ereignisse | 7 |
| Mengeninterpretation von Ereignissen..... | 8 |
| Laplace'sche Wahrscheinlichkeit und die Axiome von Kolmogorov..... | 8 |
| Bedingte Wahrscheinlichkeit, Unabhängigkeit und Unvereinbarkeit | 10 |
| Mehrstufige Wahrscheinlichkeiten..... | 11 |
| Stochastische Funktionen | 12 |
| Mathematische Wahrscheinlichkeit und Verteilungsfunktion..... | 12 |
| Stetige und diskrete Funktionen | 13 |
| Erwartungswert und Varianz..... | 13 |
| Zweidimensionale Zufallsvariablen | 16 |
| Wahrscheinlichkeitsverteilungen | 16 |
| Diskrete Gleichverteilung..... | 16 |
| Einpunktverteilung | 17 |
| Zweipunktverteilung / Null-Eins-Verteilung..... | 17 |
| Binomialverteilung..... | 17 |
| Normalverteilung | 18 |
| | |
| Teil III: Testen von Hypothesen | 20 |
| Prüfverteilungen | 20 |
| Schätzen von Parametern | 21 |
| Punktschätzung | 21 |
| Konfidenzintervall..... | 21 |
| Schätzen einer Binomialwahrscheinlichkeit..... | 23 |
| Parametrische Tests | 23 |
| Nullhypothese, Fehler 1. und 2. Art | 23 |
| Ein-Stichproben-Verfahren | 25 |
| Zwei-Stichproben-Verfahren..... | 26 |
| Gleichheit zweier Binomialverteilungen..... | 27 |
| Nichtparametrische Tests | 28 |
| Anpassungstest..... | 28 |
| Homogenitätstest..... | 30 |
| | |
| Teil IV: Ursache-Wirkungs Beziehungen | 32 |
| Lineare Regression | 32 |
| Vierfeldertafel | 34 |
| χ^2 -Kontingenztest | 34 |
| Odds, Odds Ratio, Log-Odds und Log-Odds Test | 36 |
| Mehrfeldertafeln, G^2 -Statistik | 38 |

TEIL I: KOMBINATORIK

Grundbegriffe und Verfahren

Die Kombinatorik befasst sich mit der Anordnung von Elementen aus einer Grundgesamtheit, also einer Menge. Sie bildet das Grundgerüst der Wahrscheinlichkeitstheorie; insbesondere der Laplace'schen Wahrscheinlichkeit. Die Laplace'sche Wahrscheinlichkeit wird ausgedrückt als günstige pro mögliche Ausgänge eines Experiments. Dieser Quotient gibt dann die Wahrscheinlichkeit für einen Erfolg an:

$$P_{LAPLACE} = \frac{\text{günstige}}{\text{mögliche}}$$

Beispiel 1: In einem Sack befinden sich insgesamt 24 Kugeln, davon 22 schwarze und 2 weisse. Wer bei einmal ziehen eine weisse Kugel erwischt, gewinnt. Die Laplace Wahrscheinlichkeit für einen Erfolg ergibt sich aus 'günstige' (2) pro 'mögliche' (24), also $2/24 = 1/12$.

Im Beispiel sind sowohl die Anzahl der günstigen als auch die Anzahl der möglichen Fälle leicht zu erkennen. Dies ist jedoch nicht selbstverständlich, wie das nachfolgende Beispiel zeigt:

Beispiel 2: Wieder werden Kugeln gezogen, nacheinander zwei Kugeln pro Spiel. Im Sack befinden sich neu goldene, silberne und bronzene Kugeln. Es gewinnt, wer zwei goldene Kugeln zieht. Im Sack befinden sich 8 goldene, 8 silberne und 8 bronzene Kugeln.

Zuerst zu den günstigen Fällen. Da es mehr als zwei goldene Kugeln im Beutel hat, kann man auf mehrere Arten gewinnen. Seien die goldenen Kugeln durchnummeriert von $g_1 \dots g_8$. Daraus folgt das kombinatorische Problem, aus diesen acht Kugeln jeweils Paare zu bilden (2 von 8 Elementen anordnen). Mögliche Paare wären etwa: $g_1g_2, g_1g_4, g_3g_8, g_8g_3$. Offensichtlich spielt die Reihenfolge, in der die goldenen Kugeln gezogen werden, keine Rolle: g_3g_8 ist gleichwertig mit g_8g_3 , da die Gewinnbedingung lediglich 'zwei goldene Kugeln' verlangt.

Bevor der Spieler die erste Kugel zieht, ist die Wahrscheinlichkeit, eine goldene zu erwischen gleich $8/24 = 0.333$. Wenn er die goldene Kugel aus dem Spiel nimmt, bevor er den zweiten Zug wagt, so ist die Wahrscheinlichkeit für eine goldene Kugel im zweiten Zug noch $7/23 = 0.304$ und damit kleiner als im ersten Zug. Legt der Spieler die Kugel jedoch vor dem zweiten Zug in den Beutel zurück, so bleibt die Wahrscheinlichkeit auch im zweiten Zug $8/24 = 0.333$.

Das Zurücklegen hat auch einen Effekt auf die Anzahl der Gewinnkombinationen, denn es kommen beim Zurücklegen alle Fälle hinzu, in denen in beiden Zügen die gleiche Kugel gezogen wird: $g_1g_1, g_2g_2 \dots$

Anhand der dargestellten Sachverhalte wird eine Systematisierung der kombinatorischen Probleme möglich. Es stellen sich immer drei Fragen:

- Werden alle Elemente aus der Gesamtheit angeordnet?
- Ist die Reihenfolge in der Anordnung wesentlich?

→ Sind Wiederholungen gestattet?

Permutation

Eigenschaften: → Aus n Elementen werden n (also alle) angeordnet.
→ Die Reihenfolge ist wesentlich.
→ Wiederholungen sind nicht gestattet.

Typisches Problem: Wie viele mögliche Ranglisten hat ein Schachturnier mit 8 Mitspielern?

Formel: $P(n) = n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-2) \cdot (n-1) \cdot n$

Rechenbeispiel: $P(n=8) = 8! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 = 40'320$

Variation mit Wiederholung

Eigenschaften: → Aus n Elementen werden k angeordnet ($k \leq n$).
→ Die Reihenfolge ist wesentlich.
→ Wiederholungen sind gestattet.

Typisches Problem: Wie viele Abkürzungen lassen sich aus den Buchstaben A, B, C erzeugen?

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

Rechenbeispiel: $\bar{V}(k=3, n=26) = 26^3 = 17'576$

Variation ohne Wiederholung

Eigenschaften: → Aus n Elementen werden k angeordnet ($k < n$).
→ Die Reihenfolge ist wesentlich.
→ Wiederholungen sind nicht gestattet.

Typisches Problem: Wie viele Siegerphotos (mit den ersten 3 Plätzen) gibt es bei einem Schachturnier mit 8 Personen?

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

Rechenbeispiel: $V(k=3, n=8) = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{8 \cdot \cancel{7} \cdot \cancel{6} \cdot \cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1}}{5 \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1}} = 6 \cdot 7 \cdot 8 = 336$

Bemerkung: Die Permutation stellt also einen Spezialfall der Variation ohne Wiederholung dar, bei dem $n = k$. Der Nenner des obigen Bruches wird damit 0, und $0! = 1$. Es bleibt nur der Zähler → $n!$.

Kombination ohne Wiederholung

Eigenschaften: → Aus n Elementen werden k angeordnet ($k \leq n$).
→ Die Reihenfolge ist unwesentlich.
→ Wiederholungen sind nicht gestattet.

Typisches Problem: Wie viele Möglichkeiten gibt es bei einer Lotterie "2 aus 8"? Die Reihenfolge ist unwesentlich, die gezogenen Kugeln bleiben draussen.

Formel:
$$C(k, n) = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}{k \cdot (k-1) \cdot (k-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1}$$

Rechenbeispiel:
$$C(k=2, n=8) = \binom{8}{2} = \frac{8!}{2!6!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8}{(1 \cdot 2) \cdot (1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6)} = \frac{7 \cdot 8}{1 \cdot 2} = \frac{56}{2} = 28$$

kürzen

Bemerkung: $\binom{n}{k}$ (sprich "n tief k") heisst Binomialkoeffizient. Die wichtigsten Eigenschaften des Binomialkoeffizienten sind:

$$\binom{0}{0} = 1 \quad \binom{n}{m} = 0 \text{ für } m > n$$

$\binom{n}{0} = \binom{n}{n} = 1$, der Binomialkoeffizient ist also symmetrisch. Es folgt:

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

Kombination mit Wiederholung

Eigenschaften:
 → Aus n Elementen werden k angeordnet ($k \leq n$).
 → Die Reihenfolge ist unwesentlich.
 → Wiederholungen sind gestattet.

Typisches Problem: Wie viele ungeordnete Stichproben vom Umfang 3 mit Zurücklegen lassen sich aus einer Menge mit 12 Elementen ziehen?

Formel:
$$\bar{C}(k, n) = \binom{n+k-1}{k} = \binom{n+k-1}{n-1} = \frac{n \cdot (n+1) \cdot (n+2) \cdot \dots \cdot (n+k-1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot k}$$

Rechenbeispiel:
$$\bar{C}(k=3, n=12) = \binom{12+3-1}{3} = \binom{14}{3} = \frac{12 \cdot 13 \cdot 14}{1 \cdot 2 \cdot 3} = \frac{2184}{6} = 364$$

Schlussbemerkungen

Anhand der Systematisierung lässt sich nun das leicht abgewandelte *Beispiel 2* lösen:

Beispiel 2: Wieder werden Kugeln gezogen, nacheinander zwei Kugeln pro Spiel, wobei die Kugel nach jedem Zug zurückgelegt wird. Im Sack befinden sich neu goldene, silberne und bronzene Kugeln. Die Kugeln einer Farbe sind untereinander nicht unterscheidbar. Es gewinnt, wer zwei goldene Kugeln zieht. Im Sack befinden sich 8 goldene, 8 silberne und 8 bronzene Kugeln. Wie hoch ist die Gewinnwahrscheinlichkeit?

Um die Laplace Wahrscheinlichkeit zu bestimmen, muss zuerst die Anzahl der günstigen Fälle berechnet werden, wie weiter oben bereits angedeutet. Da 'mit Zurücklegen' gezogen

wird, sind Wiederholungen gestattet. Weil die Kugeln nicht unterscheidbar sind, kann man zwar das Paar g_1g_3 nicht von g_3g_1 unterscheiden, jedes Paar trägt allerdings eine Möglichkeit zu den günstigen Fällen bei, die Reihenfolge ist deshalb wesentlich. Dies führt zu einer Variation mit Wiederholung. Für die Bestimmung der günstigen Fälle werden lediglich die goldenen Kugeln betrachtet, also ist $n = 8$. Dabei werden $k = 2$ Kugeln gezogen. Die Formel dazu lautet:

$$n^k = 8^2 = 64$$

Werden die gleichen Überlegungen für die Anzahl der möglichen Fälle angewendet, so gelangt man wiederum auf eine Variation mit Wiederholung, denn sg und gs tragen jeweils ein Element zu den möglichen Fällen bei (andernfalls wäre es eine Variation). Allerdings bezieht sich das n nun auf alle Kugeln im Beutel, also $n = 24$. Das k ist weiterhin $k = 2$. Die Formel lautet:

$$n^k = 24^2 = 576$$

Für die Laplace'sche Gewinnwahrscheinlichkeit ergibt sich in diesem Spiel also:

$$P_L = \frac{64}{576} = \frac{1}{9} = 0.\bar{1} \approx 11\%$$

Dieses Problem lässt sich übrigens auch auf einfacherem Wege lösen. Die Gewinnbedingung '2 mal Gold' lässt sich aufspalten in 'Gold im 1. Zug' UND 'Gold im 2. Zug'. Die Wahrscheinlichkeit für Gold beträgt in beiden Zügen $8/24 = 1/3$. Für zwei mal Gold hintereinander ergibt sich somit eine Wahrscheinlichkeit von $1/3 \cdot 1/3 = 1/9$ (vgl. Teil II, Mehrstufige Wahrscheinlichkeiten, S. 11)

Obwohl es in diesem Beispiel nicht sehr ausgeprägt ist, kommen in kombinatorischen Problemen oft sehr grosse Zahlen vor, die durchaus auch das Fassungsvermögen eines normalen Taschenrechners zu sprengen vermögen. Auch ist die Berechnung eines Binomialkoeffizienten unter Umständen mit grossem Rechenaufwand verbunden. In so einem Fall sind manchmal die Eigenschaften der Binomialkoeffizienten nützlich, insbesondere ihre Symmetrie.

TEIL II: WAHRSCHEINLICHKEITSTHEORIE

Einführung

Ereignisse

In der Alltagssprache werden Begriffe wie 'wahrscheinlich' oder 'zufällig' anders verwendet als in der Stochastik (der Lehre vom zufälligen Ereignis). Insbesondere wird ein Ereignis in der Alltagssprache als 'zufällig' bezeichnet, wenn ihm eine geringe Wahrscheinlichkeit zugemessen wird ("Ich habe ihn rein zufällig getroffen"). In der Statistik ist dies nicht so.

*Ein Ereignis heisst zufällig, wenn es das Ergebnis eines Zufallsexperimentes ist.
Ein Zufallsexperiment ist ein Experiment mit ungewissem Ausgang, das heisst, es hat mindestens zwei mögliche Ausgänge.*

Ein solches Zufallsexperiment mit zwei möglichen Ausgängen ist zum Beispiel 'eine Münze werfen'. Die Ausgänge lauten dann 'Kopf' und 'Zahl'. Ein weiteres Zufallsexperiment ist 'einmal mit einem Würfel würfeln'. Die möglichen Ausgänge lauten $\{1, 2, 3, 4, 5, 6\}$. Als drittes Beispiel diene 'mit zwei Würfeln gleichzeitig würfeln'. Die möglichen Ausgänge hier lauten $\{(1,1); (1,2); (1,3) \dots (6,5); (6,6)\}$. Hier wird auch die Bedeutung der Kombinatorik für die Stochastik deutlich. Die Kombinatorik gibt nämlich auf einfache Art die Anzahl der möglichen Ausgänge für das dritte Experiment an: $\bar{C}(2,6) = 21$ (Reihenfolge unwesentlich, Wiederholungen gestattet). Mit zwei Würfeln gleichzeitig würfeln ist gleichbedeutend wie mit einem Würfel zweimal hintereinander zu würfeln, wobei es sich bei jedem Wurf um eine 'Ziehung' aus sechs möglichen Ausgängen handelt.

Jeder mögliche Ausgang eines Experiments wird als Elementarereignis ω (omega) bezeichnet. Beim Würfeln mit einem Würfel wäre ein Elementarereignis beispielsweise $\omega_1 = \{1\}$ oder $\omega_3 = \{3\}$. Beim Würfeln mit zwei Würfeln wäre ein Elementarereignis beispielsweise $\omega_1 = \{1,1\}$ oder $\omega_{14} = \{3,2\}$. Im Gegensatz dazu stehen zusammengesetzte Ereignisse. So ist das Ereignis $A = \{2, 4, 6\}$ 'eine gerade Zahl würfeln' zusammengesetzt aus den Elementarereignissen $\omega_1 = \{2\}$, $\omega_2 = \{4\}$ und $\omega_3 = \{6\}$.

*Das Elementarereignis ω ist das kleinste zufällige Ereignis.
Ein Ereignis A tritt ein, wenn eines seiner Elementarereignisse $\omega(A)$ eintritt.*

Die Grenzen des Zufalls bilden das sichere Ereignis Ω (Omega), das alle Elementarereignisse eines Experiments enthält und somit immer eintritt: Ω ('eine Zahl von 1 bis 6 würfeln') = $\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\} = \{1, 2, 3, 4, 5, 6\}$.

Das sichere Ereignis $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ enthält alle Elementarereignisse.

Am anderen Ende der Skala steht das unmögliche Ereignis \emptyset . Das unmögliche Ereignis kann nie eintreten, weil es kein Elementarereignis enthält.

Das unmögliche Ereignis $\emptyset = \{ \}$ enthält keine Elementarereignisse.

Das unmögliche Ereignis bildet das Komplementär- oder Gegenereignis zum sicheren Ereignis.

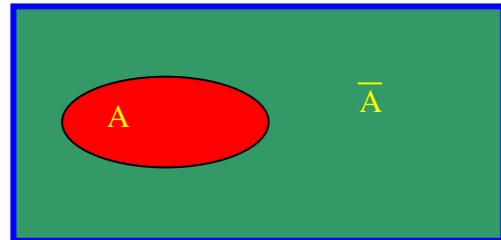
Das Komplementärereignis \bar{E} tritt ein, wenn das Ereignis E nicht eintritt.

Mengeninterpretation von Ereignissen

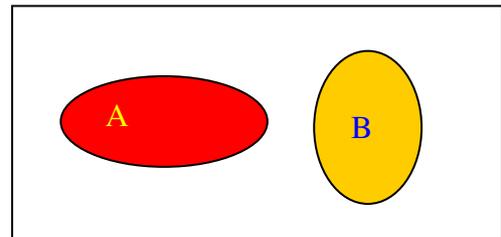
Ereignisse, insbesondere auch Elementarereignisse, lassen sich als Elemente oder Teilmengen einer Menge auffassen, genauer als Elemente des Ereignisraums Ω . Der Ereignisraum Ω umfasst alle Elementarereignisse, somit also alle möglichen Ausgänge eines Experiments. Damit entspricht der Ereignisraum dem sicheren Ereignis.



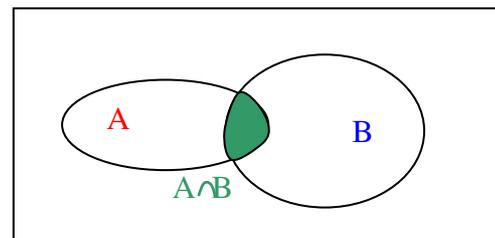
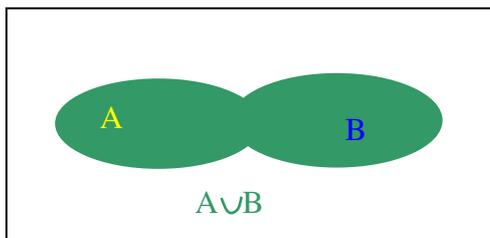
Ein Ereignis A ist Teil des Ereignisraums (A bildet eine Teilmenge von Ω : $A \subseteq \Omega$). Auch \bar{A} ist Teil des Ereignisraumes. Das Gegenereignis \bar{A} von A umfasst alle jene Elemente des Ereignisraumes, die nicht Teil von A sind ($A \cup \bar{A} = \Omega$)



Zwei Ereignisse A und B seien Teil des Ereignisraumes Ω . A und B heissen disjunkt, wenn sie keine gemeinsamen Elemente haben ($A \cap B = \{ \}$).

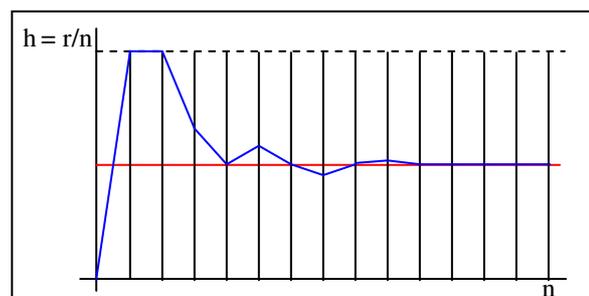


Mit Mengen kann auch gerechnet werden. Die wichtigsten Rechenoperationen sind $A \cup B$ 'A vereinigt mit B' (manchmal auch als $A+B$ geschrieben) und $A \cap B$ 'A geschnitten mit B' (manchmal auch als AB geschrieben).



Laplace'sche Wahrscheinlichkeit und die Axiome von Kolmogorov

Bei der Laplace'schen Wahrscheinlichkeit steht die relative Häufigkeit im Vordergrund. Das Experiment 'Münze werfen' werde immer wieder durchgeführt. Dabei wird die relative Häufigkeit des Ereignisses 'Kopf' betrachtet, also der Quotient 'Anzahl Kopf durch Anzahl Würfe insgesamt'. Nach jedem Wurf wird die aktuelle relative Häufigkeit in Abhängigkeit zur Anzahl Würfe abgetragen. Es ergibt sich das nebenstehende Bild. Die relative Häufigkeit h_i strebt mit wachsendem i einem bestimmten Wert zu. Diese auffallende Stabilität eignet sich, um die Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses zu umschreiben.



Die Laplace'sche Wahrscheinlichkeit ist definiert als $P_L(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl günstige Fälle}}{\text{Anzahl mögliche Fälle}}$.

Dieser empirische Ansatz zur Wahrscheinlichkeitsdefinition weist allerdings theoretische und logische Lücken auf. Erst im 20. Jahrhundert wurde durch Kolmogorov eine formale Grundlage der Wahrscheinlichkeitstheorie geschaffen, das Axiomensystem der Wahrscheinlichkeitsrechnung.

Axiom 1: Jedem zufälligen Ereignis A eines zufälligen Versuchs ist eine Wahrscheinlichkeit $P(A)$ zugeordnet, die Werte zwischen 0 und 1 annehmen kann.

$$0 \leq P(A) \leq 1$$

Axiom 2: Das sichere Ereignis hat die Wahrscheinlichkeit 1:

$$P(\Omega) = 1$$

Axiom 3: Sind A_1 und A_2 disjunkte Ereignisse, so ist

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

Auf den Axiomen gründet die gesamte Wahrscheinlichkeitsrechnung. Die wichtigsten Folgerungen, die sich direkt aus ihnen ableiten lassen sind:

Folgerung 1: Die Wahrscheinlichkeit für das zu A komplementäre Ereignis \bar{A} ist

$$P(\bar{A}) = 1 - P(A)$$

Folgerung 2: Die Wahrscheinlichkeit des unmöglichen Ereignisses \emptyset ist gleich Null.

$$P(\emptyset) = 0$$

Folgerung 3: (Additionssatz für beliebige Ereignisse) Die Wahrscheinlichkeit, dass von zwei Ereignissen A und B , die sich nicht notwendig gegenseitig ausschliessen (nicht disjunkt sind), mindestens eines Eintritt, ist

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

Folgerung 4: (Additionssatz für disjunkte Ereignisse) Wird Axiom 3 auf drei und mehr disjunkte Ereignisse ausgedehnt, so folgt daraus

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

Folgerung 5: Wenn $A_1 \dots A_n$ eine vollständige Zerlegung des Ereignisses B bilden, d.h. wenn $A_1 \cup A_2 \cup \dots \cup A_n = B$, so gilt

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

Folgerung 6: Ebenso gilt, wenn $A_1 \dots A_n$ eine vollständige Zerlegung des Ereignisraumes Ω bilden, für die Wahrscheinlichkeit eines Ereignisses $B \subseteq \Omega$

$$P(B) = \sum_{i=1}^n P(B \cap A_i)$$

Folgerung 7: Wenn das Ereignis A Teilmenge des Ereignisses B ist, so gilt

$$A \subseteq B \Rightarrow P(A) \leq P(B)$$

Bedingte Wahrscheinlichkeit, Unabhängigkeit und Unvereinbarkeit

Betrachtet wird eine Situation, dass von zwei Ereignissen A und B z.B. das Ereignis A eine Vorinformation dahingehend liefert, dass sein Eintreten den möglichen Ereignisraum von B reduziert, d.h. einen Einfluss auf die Wahrscheinlichkeit von B hat.

Beispiel 3: Eine ideale Münze wird 3 mal geworfen. Untersucht werden drei Ereignisse:

A: 1. Wurf liefert Zahl

B: 2. Wurf liefert Zahl

C: genau 2 mal Zahl hintereinander

Damit sind (K = Kopf, Z = Zahl):

$$\Omega = \{ZZZ, ZZK, ZKZ, ZKK, KZZ, KZK, KKZ, KKK\}$$

$$A = \{ZZZ, ZZK, ZKZ, ZKK\} \quad P(A) = 0.5$$

$$B = \{ZZZ, ZZK, KZZ, KZK\} \quad P(B) = 0.5$$

$$C = \{ZZK, KZZ\} \quad P(C) = 0.25$$

Es sei nun bekannt, dass A eingetreten ist (der erste Wurf liefert Zahl). Dies hat einen Einfluss auf die Elementarereignisse von B und C:

$$A \text{ eingetreten} \rightarrow B = \{ZZZ, ZZK, \text{KZZ}, \text{KZK}\} \rightarrow B|A = \{ZZZ, ZZK\}$$

$$\rightarrow \text{aber: } \Omega_{\text{neu}} = \{ZZZ, ZZK, ZKZ, ZKK\}$$

$$\rightarrow P(B|A) = 0.5 \text{ ('Wahrscheinlichkeit von B unter der Bedingung A')}$$

$$\rightarrow P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{0.25}{0.5} = 0.25 \cdot 2 = 0.5 = P(B)$$

$$A \text{ eingetreten} \rightarrow C = \{ZZK, \text{KZZ}\} \rightarrow C|A = \{ZZK\}$$

$$\rightarrow \text{aber: } \Omega_{\text{neu}} = \{ZZZ, ZZK, ZKZ, ZKK\}$$

$$\rightarrow P(C|A) = 0.25$$

$$\rightarrow P(C|A) = \frac{P(C \cap A)}{P(A)} = \frac{0.125}{0.5} = 0.25 = P(C)$$

Das Eintreten von A hat also keinen Einfluss auf die Wahrscheinlichkeit von B oder C.

$$B \text{ eingetreten} \rightarrow A = \{ZZZ, ZZK, \text{ZKZ}, \text{ZKK}\}$$

$$\rightarrow \Omega_{\text{neu}} = \{ZZZ, ZZK, KZZ, KZK\}$$

$$\rightarrow P(A|B) = 0.5$$

$$\rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.25}{0.5} = 0.5 = P(A)$$

$$B \text{ eingetreten} \rightarrow C = \{ZZK, KZZ\}$$

$$\rightarrow \text{aber: } \Omega_{\text{neu}} = \{ZZZ, ZZK, KZZ, KZK\}$$

$$\rightarrow P(C|B) = 0.5 \neq P(C)$$

$$\rightarrow P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{0.25}{0.5} = 0.5 \neq P(C)$$

Offensichtlich hat das Eintreten von B einen Einfluss auf die Wahrscheinlichkeit für C. Noch klarer wird es, wenn das Gegenereignis zu B betrachtet wird: 'im zweiten Wurf keine Zahl'. Damit wird nämlich C unmöglich. Man sagt

Wenn das Eintreten von B das Eintreten von C verhindert, so sind B und C unvereinbar.

Wenn das Eintreten von A keinen Einfluss auf die Wahrscheinlichkeit für B hat, so sind A und B unabhängig.

Wenn A und B unabhängig sind, so gilt $P(A|B) = P(A)$.

Sind A und B nicht unabhängig, so heisst die Wahrscheinlichkeit von B unter der Bedingung, dass A eingetreten ist, 'bedingte Wahrscheinlichkeit' und berechnet sich nach der Formel

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

Für zwei beliebige Ereignisse A und B folgt daraus der Multiplikationssatz

für beliebige Ereignisse: $P(A \cap B) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$

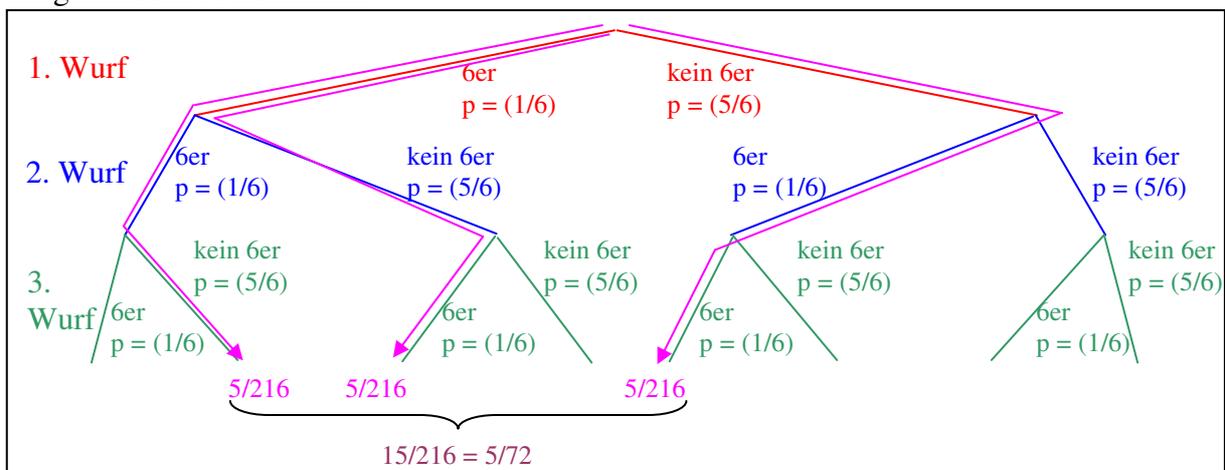
und für unabhängige Ereignisse: $P(A \cap B) = P(A) \cdot P(B)$

Mehrstufige Wahrscheinlichkeiten

Oftmals treten stochastische Probleme als mehrstufig ineinander gefügte Themenkomplexe auf. Solche Probleme lassen sich – wenn sie nicht zu kompliziert sind – einfach mit sogenannten Wahrscheinlichkeitsbäumen lösen.

Beispiel 4: Mit einem idealen Würfel ($p = 1/6$) wird drei mal gewürfelt. Wie gross ist die Wahrscheinlichkeit, davon zwei 6er zu werfen?

Das Problem wird in drei Schritte aufgegliedert. Jeder Schritt entspricht einem Wurf. Das Problem wird ausserdem binär kodiert: statt die Varianten $\{1, 2, 3, 4, 5, 6\}$ zu zeichnen, wird das Problem auf die beiden Varianten $\{6, \text{nicht } 6\}$ reduziert. So bleibt der Baum einigermassen übersichtlich.



Der Baum liefert eine klare Lösung für das Beispiel 4. Jeder Pfad, der die Bedingung 'zwei mal 6er' erfüllt, wird durchlaufen. Wahrscheinlichkeiten, die im Baum untereinander liegen (also auf zwei verschiedenen Ebenen), werden beim Durchlaufen des Baumes multipliziert. Wahrscheinlichkeiten auf gleicher Ebene (insbesondere die jeweiligen Wahrscheinlichkeiten

der drei 'Gewinnpfade') werden addiert. Der Baum in dieser Form, also mit nur zwei Möglichkeiten in jedem Knoten und gleichen Wahrscheinlichkeiten auf jeder Ebene, bildet das Grundgerüst der Binomialverteilung. Diese Einschränkungen sind allerdings nicht notwendig, um das Prinzip des Wahrscheinlichkeitsbaumes anwenden zu können. Auch Probleme mit bedingter Wahrscheinlichkeit können mit diesem Verfahren analysiert werden (vgl. Toutenburg S. 26).

Stochastische Funktionen

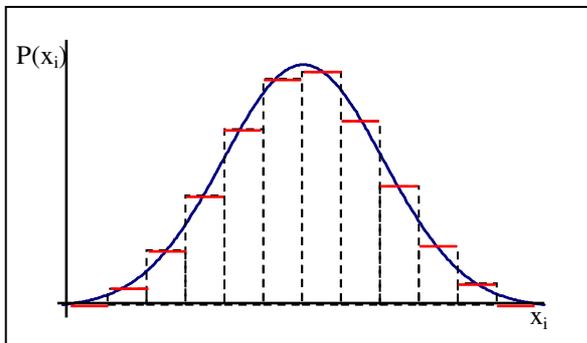
Mathematische Wahrscheinlichkeit und Verteilungsfunktion

Um mit Ereignissen rechnen zu können, ist es zweckmässig, jedem Ereignis ω_i des Ereignisraumes eine reelle Zahl x_i zuzuordnen. Dies erreicht man durch die Abbildung (Funktion) $\omega_i \mapsto X(\omega_i) = x_i$. Während alle möglichen (Elementar-) Ereignisse ω_i zusammen den Ereignisraum Ω ergeben, bilden alle Bilder x_i den Zustandsraum S . Die Variable x_i wird als Zufallsvariable bezeichnet, weil sich die Unsicherheit von ω_i auf x_i überträgt.

Die Abbildung $\omega_i \mapsto X(\omega_i) = x_i$ erzeugt eine Zufallsvariable.

In einem zweiten Schritt wird wiederum jedem x_i eine reelle Zahl $P(x_i)$ zugeordnet, die die Wahrscheinlichkeit von x_i bzw. ω_i beschreibt. Die Funktion $x_i \mapsto P(x_i)$ wird durch die Axiome von Kolmogorov umschrieben.

Die Funktion $x_i \mapsto P(x_i)$ heisst Wahrscheinlichkeitsfunktion.

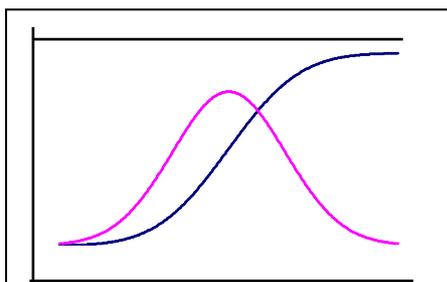


Die Wahrscheinlichkeitsfunktion ist eng mit der Häufigkeitsverteilung aus der deskriptiven Statistik verknüpft, da (gemäss dem Wahrscheinlichkeitsansatz nach Laplace) relative Häufigkeit und Wahrscheinlichkeit verwandt sind. Für stetige Funktionen bildet die Wahrscheinlichkeitsfunktion häufig eine Glockenkurve. Diskrete Funktionen weisen hingegen eine treppenförmige

Wahrscheinlichkeitsfunktion auf.

Die kumulierte Wahrscheinlichkeit führt auf die Verteilungsfunktion. Die Verteilungsfunktion $F(x_i)$ gibt für ein beliebiges $x_i \in S$ (eine beliebige Ausprägung) die Wahrscheinlichkeit für das Eintreten von $x \leq x_i$ an.

Beispiel 5: Die Graphik zeigt die Wahrscheinlichkeits- und Verteilungsfunktion (rosa resp.



blau) für 'Anzahl 6er in 300 Würfeln'. Die Wahrscheinlichkeitsfunktion weist ein Maximum bei 50 auf. Die Wahrscheinlichkeit, genau die Würfelsumme 50 zu werfen, beträgt allerdings verschwindend kleine 0.006; die rosa Kurve ist gegenüber der blauen Kurve um den Faktor 100 gestreckt. Die blaue Kurve erreicht an der Stelle 50 einen Wert von annähernd 0.6. Dies ist die Wahrscheinlichkeit, in 300 Würfeln insgesamt 0 oder 1 oder 2 oder 3 ... oder 49 oder 50 6er zu werfen.

Anmerkung: die obige Darstellung ist nicht ganz korrekt, weil eine stetige Wahrscheinlichkeitsfunktion unterstellt wird, obwohl das Würfelproblem gar keine solche liefert. Ausserdem wurden die Graphen der Darstellung wegen skaliert. Bei 300 Würfeln ist die Abweichung von einer stetigen Funktion jedoch nur noch sehr gering.

Die Verteilungsfunktion ist definiert durch $F(x) = P(X \leq x) = P(-\infty < X \leq x)$.

Die Verteilungsfunktion ist schwach monoton wachsend, und alle Werte von $F(x)$ liegen zwischen 0 und 1. Ausserdem gilt $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$.

Stetige und diskrete Funktionen

Stetige Funktionen unterscheiden sich von diskreten Wahrscheinlichkeitsverteilungen dadurch, dass diskrete Funktionen endlich viele oder abzählbar unendlich viele Ausprägungen für x_i enthalten (zum Beispiel 0 6er, 1 6er, 2 6er...; diese Funktion bleibt diskret, weil auch bei unendlich vielen Würfeln die Anzahl abzählbar bleibt). Bei stetigen Funktionen ist dies nicht der Fall. Zwischen eine x_i und einem $(x_i + \Delta x)$ gibt es immer unendlich viele zulässige Werte für x_i , egal wie klein Δx gewählt wird (Beispiel Körpergrösse: zwischen 180cm und 181cm [$\Delta x = 1\text{cm}$] gibt es unendlich viele Schritte, die alle zulässig sind; es handelt sich nur um ein Messproblem).

Im Grunde führt ein Experiment, das aus einer endlichen Anzahl von Versuchen besteht, immer auf eine Treppenfunktion. Im Beispiel 5 hatte die Treppe 300 Stufen. Die Wahrscheinlichkeit, eine bestimmte Stufe, also genau eine bestimmte Anzahl 6er, zu treffen, war schon recht klein (0.006 für den Erwartungswert).

Für stetige Funktionen ist $P(x) = 0$.

Ab einer gewissen Anzahl von Versuchen ist es deshalb zulässig, eine diskrete Verteilung (in diesem Falle die Binomialverteilung) durch eine stetige Verteilung (meist die Normalverteilung) zu approximieren.

Die Ableitung einer stetigen Verteilungsfunktion liefert die sogenannte Dichtefunktion. Die Integralfunktion $F(x) = \int_{-\infty}^x f(t)dt$ über die Dichtefunktion $f(t)$ liefert an der Stelle x_i also wieder den Wert der kumulierten Wahrscheinlichkeit $P(x < x_i) = f(x_i)$, den Wert der Verteilungsfunktion in x_i .

Erwartungswert und Varianz

Erwartungswert und Varianz bilden die beiden wichtigsten Masszahlen oder Parameter einer Zufallsvariablen. Der Erwartungswert $E(x)$ ist ein Lageparameter. Er charakterisiert den Schwerpunkt einer Verteilung.

Ist x diskret verteilt, so gilt für den Erwartungswert $E(X) = \sum_{i=1}^n x_i \cdot p(x_i)$.

Ist x stetig, so gilt für den Erwartungswert $E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)$.

Damit entspricht der Erwartungswert dem arithmetischen Mittel aus der deskriptiven Statistik.

Beispiel 6: Ein beobachtetes Ereignis lautet: 'mit zwei Würfeln werfen und die Augensumme bestimmen'. Dieses Ereignis wird $m = 36$ mal beobachtet. Die Ausprägungen (Elementarereignisse) des Ereignisses entsprechen der Augensumme: $S = \{2; 3; 4; \dots; 10; 11; 12\}$. Die Verteilung der Resultate entspreche den unten aufgeführten Werten (es tritt jede mögliche Ausprägung gerade einmal auf). Der Erwartungswert dieser diskreten Verteilung berechnet sich anhand der obenstehenden Formel. Der Übersichtlichkeit halber ist eine Tabelle eingefügt, die das Ergebnis jedes Wurfes enthält. Diese Tabelle heisst Urliste.

| Wurf j | Würfel 1 | Würfel 2 | Summe | Wurf j | Würfel 1 | Würfel 2 | Summe |
|----------|----------|----------|-------|----------|----------|----------|-------|
| 1 | 1 | 1 | 2 | 19 | 4 | 1 | 5 |
| 2 | 1 | 2 | 3 | 20 | 4 | 2 | 6 |
| 3 | 1 | 3 | 4 | 21 | 4 | 3 | 7 |
| 4 | 1 | 4 | 5 | 22 | 4 | 4 | 8 |
| 5 | 1 | 5 | 6 | 23 | 4 | 5 | 9 |
| 6 | 1 | 6 | 7 | 24 | 4 | 6 | 10 |
| 7 | 2 | 1 | 3 | 25 | 5 | 1 | 6 |
| 8 | 2 | 2 | 4 | 26 | 5 | 2 | 7 |
| 9 | 2 | 3 | 5 | 27 | 5 | 3 | 8 |
| 10 | 2 | 4 | 6 | 28 | 5 | 4 | 9 |
| 11 | 2 | 5 | 7 | 29 | 5 | 5 | 10 |
| 12 | 2 | 6 | 8 | 30 | 5 | 6 | 11 |
| 13 | 3 | 1 | 4 | 31 | 6 | 1 | 7 |
| 14 | 3 | 2 | 5 | 32 | 6 | 2 | 8 |
| 15 | 3 | 3 | 6 | 33 | 6 | 3 | 9 |
| 16 | 3 | 4 | 7 | 34 | 6 | 4 | 10 |
| 17 | 3 | 5 | 8 | 35 | 6 | 5 | 11 |
| 18 | 3 | 6 | 9 | 36 | 6 | 6 | 12 |

Es muss nun die absolute Häufigkeit aus der Tabelle herausgelesen werden. Diese, dividiert durch die Anzahl Ausprägungen ($m = 36$), ergibt die Wahrscheinlichkeit p .

| Augensumme = x_j | absolute Häufigkeit | Wahrscheinlichkeit = p |
|--------------------|---------------------|--------------------------|
| 2 | 1 | 0.028 |
| 3 | 2 | 0.056 |
| 4 | 3 | 0.083 |
| 5 | 4 | 0.111 |
| 6 | 5 | 0.139 |
| 7 | 6 | 0.167 |
| 8 | 5 | 0.139 |
| 9 | 4 | 0.111 |
| 10 | 3 | 0.083 |
| 11 | 2 | 0.056 |
| 12 | 1 | 0.028 |

Der Erwartungswert des Ereignisses '1 mal mit zwei Würfeln werfen und Augensumme beobachten' berechnet sich also:

$$E(X) = \sum x_j \cdot p(x_j) = 2 \cdot 0.028 + 3 \cdot 0.056 + 4 \cdot 0.083 + \dots + 11 \cdot 0.056 + 12 \cdot 0.028 = 7.007 \approx 7$$

Die induktive Statistik geht nun allerdings einen Schritt weiter. Anstatt dass ein Experiment nur 1 mal durchgeführt wird (mit zwei Würfeln nur 1 mal 36 mal geworfen wird), wird das Experiment aus Beispiel 6 zum Beispiel $n = 30$ mal wiederholt (insgesamt wird nun also $30 \cdot 36 = 1080$ mal gewürfelt). Für jede Runde ergibt sich ein etwas anderes $E(X_i)$.

Der Erwartungswert einer Menge von Erwartungswerten ist: $E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$.

Bedingung ist, dass alle X_i i.i.d. sind (die gleiche Verteilung aufweisen und unabhängig sind). Graphisch bedeutet dies, dass jede Runde eine Kurve von gleicher Form, aber unterschiedlicher Lage auf der X-Achse aufweist.

Die Varianz sagt etwas über die Konzentration der Verteilung um den Erwartungswert aus. Die Varianz ist definiert als: $Var(X) = E[X - E(X)]^2 = \sigma^2$, also als mittlere quadratische Abweichung der Variablen X vom Erwartungswert $E(X)$. In Formeln bedeutet dies:

$$Var(X) = \sum_{i=1}^n (x_i - E(X))^2 \cdot p_i \text{ im diskreten Fall.}$$

$$Var(X) = \int_{-\infty}^{+\infty} (x - E(x))^2 \cdot f(x) dx \text{ im stetigen Fall.}$$

Die (positive) Wurzel aus der Varianz heisst Standardabweichung.

Beispiel 7: Anhand der Urliste aus Beispiel 6 soll nun die Varianz der Verteilung mit $m=36$ Würfeln berechnet werden.

| x_j | $[x_j - E(X)]^2 \cdot p_j$ | abs. Häuf. | x_j | $[x_j - E(X)]^2 \cdot p_j$ | abs. Häuf. |
|-------|----------------------------|------------|-------|----------------------------|------------|
| 2 | 0.7 | 1 | 8 | 0.139 | 5 |
| 3 | 0.896 | 2 | 9 | 0.444 | 4 |
| 4 | 0.747 | 3 | 10 | 0.747 | 3 |
| 5 | 0.444 | 4 | 11 | 0.896 | 2 |
| 6 | 0.139 | 5 | 12 | 0.7 | 1 |
| 7 | 0 | 6 | | | |

Der einfacheren Berechnung wegen wurde die zusammengefasste Urliste verwendet. Für die Varianz ergibt sich:

$$Var(X) = 1 \cdot 0.7 + 2 \cdot 0.896 + 3 \cdot 0.747 + \dots + 2 \cdot 0.896 + 1 \cdot 0.7 = 9.204$$

Für eine Menge von Zufallsvariablen berechnet sich die Gesamtvarianz analog zum Mittelwert nach der Formel:

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

Ein wichtiges Instrument in der Statistik ist die Standardisierung, weil sie es ermöglicht, sehr viele Probleme auf die gleichen bekannten Probleme zurückzuführen (Normalverteilung!).

Eine Zufallsvariable heisst standardisiert, wenn gilt $\mu = 0$, $\sigma^2 = 1$.

Um eine beliebige Verteilung zu standardisieren, wird die Gleichung $Y = \frac{X - \mu}{\sigma}$ angewendet, wobei Y die standardisierte Zufallsvariable darstellt.

Zweidimensionale Zufallsvariablen

Das wichtigste Mass, das den Zusammenhang zwischen zwei Zufallsvariablen X und Y beschreibt, heisst Kovarianz.

Die Kovarianz berechnet sich nach: $\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$

Spezielle Eigenschaften der Kovarianz sind vor allem:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

Die Kovarianz ist symmetrisch in X und Y.

$$\text{Cov}(X, X) = \text{Var}(X)$$

Wenn $X = Y$ bzw. $Y = X$ ist.

$$\text{Cov}(X, Y) = 0$$

Wenn X und Y unabhängig sind.

Die gemeinsame Varianz zweier Zufallsvariablen X und Y berechnet sich nach der Formel:

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2 \cdot \text{Cov}(X, Y)$$

Für unabhängige X und Y ist die Cov = 0. Damit ergibt sich:

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

Die Umkehrung letzterer Beziehung ($\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] \rightarrow X, Y$ unabhängig) gilt jedoch im allgemeinen nicht.

Die Kovarianz ist lageabhängig. Auf Basis der Kovarianz wird der Korrelationskoeffizient als normiertes Mass für die Abhängigkeit zwischen den Zufallsvariablen definiert.

Der Korrelationskoeffizient ist definiert durch $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}$.

Da für unabhängige X, Y die Kovarianz $\text{Cov}(X, Y) = 0$ wird, wird auch der Korrelationskoeffizient 0. Besteht hingegen eine exakte lineare Abhängigkeit, d.h. $Y = aX + b$ mit $a \neq 0$, so ist $\rho = 1$ für $a > 0$ und $\rho = -1$ für $a < 0$.

Wahrscheinlichkeitsverteilungen

Diskrete Gleichverteilung

Eine diskrete Zufallsvariable X mit den Ausprägungen $x_1 \dots x_k$ heisst gleichverteilt, wenn für ihre Wahrscheinlichkeitsfunktion gilt:

$$P(x_i) = \frac{1}{k} \text{ für alle } i = 1 \dots k$$

$$\text{Für den Erwartungswert gilt: } E(X) = \frac{k+1}{2}$$

$$\text{Für die Varianz gilt: } \text{Var}(X) = \frac{1}{12}(k^2 - 1)$$

Einpunktverteilung

Die Einpunktverteilung bildet die Grenze des Zufalls. Eine Zufallsvariable X hat die Einpunktverteilung im Punkt a , wenn sie nur eine Ausprägung a mit $P(a) = 1$ besitzt. In diesem Fall ist der Erwartungswert $E(X) = a$ und die Varianz $\text{Var}(X) = 0$.

Zweipunktverteilung / Null-Eins-Verteilung

Eine zufällige Variable besitzt die Zweipunktverteilung, wenn sie nur zwei Werte x_1 und x_2 mit jeweils positiver Wahrscheinlichkeit annehmen kann. Die Zufallsvariable wird durch ihre beiden möglichen Werte und die zugehörigen Wahrscheinlichkeiten beschrieben:

$$P(X = x_1) = p, P(X = x_2) = 1 - p$$

Wird $x_1 = 1$ und $x_2 = 0$ gewählt, so spricht man von einer Null-Eins-Verteilung.

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

Der Erwartungswert der Null-Eins-Verteilung ist: $E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$

Die Varianz der Null-Eins-Verteilung ist: $\text{Var}(X) = p \cdot (1 - p)$

Binomialverteilung

Beobachtet man in einem Zufallsexperiment, ob ein bestimmtes Ereignis A eingetreten ist oder nicht, und wiederholt man dieses Experiment n mal unabhängig, so ist die Anzahl der eingetretenen Ereignisse A binomial verteilt. Der Binomialverteilung liegt also eine Zweipunktverteilung zugrunde. Die Binomialverteilung kann als eine Summe von identischen, unabhängigen Null-Eins-Verteilungen aufgefasst werden.

Typische Beispiele für binomial verteilte Experimente sind:

Beispiel 8: Anzahl gezogene blaue Kugeln aus einer Urne mit blauen und roten Kugeln in n Zügen mit Zurücklegen.

Anzahl 'Kopf geworfen' einer zweiseitigen Münze in n Würfen.

Anzahl funktionstüchtige Glühbirnen in einer Stichprobe von n Stück.

Die Wahrscheinlichkeit für das Eintreffen des Ereignisses 'k mal blau gezogen' berechnet sich nach der Formel:

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

Der Erwartungswert der Binomialverteilung berechnet sich: $\mu = n \cdot p$

Ihre Varianz berechnet sich: $\sigma^2 = n \cdot p \cdot (1 - p)$

Normalverteilung

Die Normalverteilung ist die wichtigste stetige Wahrscheinlichkeitsverteilung. Die Dichtefunktion der Normalverteilung lautet:

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

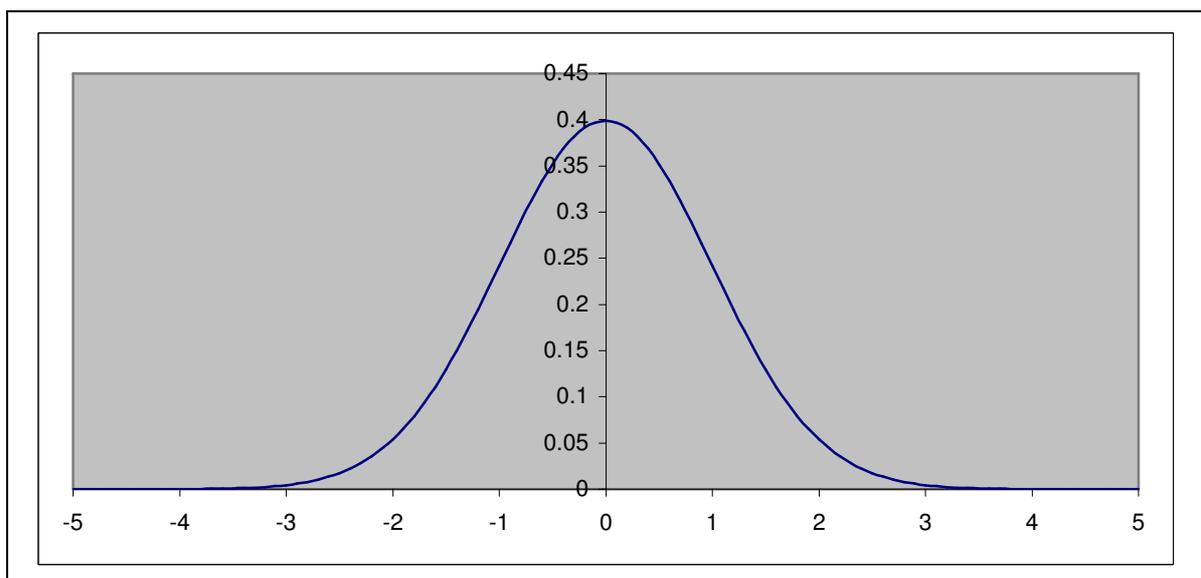
Die Parameter dieser Gleichung sind μ und σ . Für die Momente einer normalverteilten Zufallsgrösse gilt:

$$\begin{aligned} E(X) &= \mu \\ \text{Var}(X) &= \sigma^2 \end{aligned}$$

Sind speziell $\mu = 0$ und $\sigma^2 = 1$, so heisst X standardnormalverteilt: $X \sim N(0,1)$.

Um eine allgemeine Normalverteilung $N(\mu, \sigma)$ in eine standardisierte Normalverteilung $N(0, 1)$ zu überführen, wird die Standardisierungs-Transformation angewandt:

$$\begin{aligned} X \sim N(\mu, \sigma^2) &\rightarrow Z \sim N(0, 1) \\ Z &= \frac{X - \mu}{\sigma} \end{aligned}$$



Wie bereits bekannt ist die Punktwahrscheinlichkeit bei stetigen Verteilungen gleich 0. Deshalb interessiert die Dichtefunktion auch nur eingeschränkt. Interessant hingegen ist die Verteilungsfunktion $\Phi(z)$. Sie gibt die Wahrscheinlichkeit eines Bereiches an, der von $-\infty$ bis z reicht.

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z \exp\left(-\frac{x^2}{2}\right) dx$$

Dieses Integral lässt sich nicht analytisch lösen. Lösungen für Z ($0 \leq Z \leq 3$) der standardisierten Verteilungsfunktion sind deshalb tabelliert (Toutenburg S. 376). Durch die Symmetrie der Dichtefunktion um den Nullpunkt gilt:

$$\Phi(-Z) = 1 - \Phi(Z)$$

Eine Stichprobe von identisch $N(\mu, \sigma^2)$ normalverteilten Zufallsvariablen ist auch wieder normalverteilt mit den Momenten:

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$$

und

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

Damit gilt dann insgesamt für das Stichprobenmittel einer normalverteilten Zufallsvariablen:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Wie oben erwähnt, macht die Punktwahrscheinlichkeit wegen deren Stetigkeit bei der Normalverteilung keinen Sinn. Dafür kann die Wahrscheinlichkeit eines beliebigen Bereichs $[z_1, z_2]$ leicht über die Verteilungsfunktion $\Phi(z)$ bestimmt werden. Bei der Binomialverteilung ist zwar die Bestimmung einer Bereichswahrscheinlichkeit grundsätzlich nicht schwierig, sie ist jedoch unter Umständen mit einem erheblichen Rechenaufwand verbunden, weil für jede Klasse innerhalb des Bereichs die Klassenwahrscheinlichkeit ausgerechnet und diese dann aufaddiert werden muss. Um dieses Problem zu vermeiden, bedient man sich für ausreichend grosse n , p , $(1 - p)$ der Normalverteilung als Approximation der Binomialverteilung.

Für die Binomialverteilung lautet die Approximationsbedingung durch die Normalverteilung:

$$n \cdot p \cdot (1 - p) > 9 \rightarrow B(n, p) \approx N(\mu, \sigma^2).$$

Der zentrale Grenzwertsatz der Wahrscheinlichkeitsrechnung sagt sogar aus, dass jede Summe von i.i.d. Zufallsvariablen X_i normalverteilt ist für eine ausreichend grosse Menge von X_i .

TEIL III: TESTEN VON HYPOTHESEN

Prüfverteilungen

Aus der Normalverteilung lassen sich drei wesentliche Verteilungen, die sogenannten Prüfverteilungen, gewinnen. Diese Verteilungen werden zum Testen von Hypothesen eingesetzt. Je nach dem, über welchen Parameter eine Hypothese gestellt wird, kommt eine unterschiedliche Prüfverteilung zum Zug. Es sind dies im Detail:

Hypothese über

- die Varianz σ^2 einer Normalverteilung: χ^2 -Verteilung.
- den Erwartungswert μ oder zum Vergleich zweier Erwartungswerte: t-Verteilung.
- das Verhältnis von Varianzen zweier normalverteilter Variablen: F-Verteilung.

Die t- und die χ^2 -Verteilung hängen beide ab von einem ganzzahligen Parameter ν ab, der sich aus dem Stichprobenumfang berechnen lässt und dessen Werte Freiheitsgrade genannt werden.

$$H_0: \sigma^2 = \sigma_0^2 \qquad \chi^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2}$$

Getestet wird, ob die Stichprobe mit der Varianz s^2 aus der normalverteilten Grundgesamtheit mit Varianz σ_0^2 stammt. Es ist $s^2 = \sum (x_i - \bar{x})^2$. Getestet wird anhand der χ^2 -Prüfverteilung. Dabei ist $\nu = (n - 1)$. Liegt das χ^2 der Stichprobe über dem Wert der χ^2 -Prüfverteilung, so ist die Abweichung signifikant. Die Hypothese, dass die Stichprobe aus der Grundgesamtheit stammt, wird somit verworfen (nicht akzeptiert).

Ist das σ_0^2 der Grundgesamtheit unbekannt, der Erwartungswert μ_0 hingegen schon, so wird mit dem t-Test geprüft, ob eine Stichprobe mit Mittelwert \bar{x} aus der Grundgesamtheit mit bekanntem Erwartungswert μ_0 stammt.

$$H_0: \mu = \mu_0 \qquad t = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n}$$

Prüfverteilung ist hier die Student t-Verteilung t_ν mit $\nu = (n - 1)$. Liegt der t-Wert der Stichprobe über dem Signifikanzwert der Prüfverteilung, so wird die Nullhypothese verworfen.

Die Fisher F-Verteilung macht eine Aussage über zwei unabhängige Zufallsvariablen X und Y, die χ_n^2 bzw. χ_m^2 verteilt sind. Der Quotient

$$F_{m,n} = \frac{\chi_n^2/m}{\chi_m^2/n}$$

liefert die Fisher F-Verteilung mit (m, n) Freiheitsgraden. Ist X eine χ_1^2 -verteilte Zufallsvariable, so ist der obige Quotient $F_{1, n}$ verteilt. Die Wurzel aus dem Quotienten ist dann t_n -verteilt, da die Wurzel aus einer χ_1^2 -verteilten Zufallsvariablen $N(0,1)$ -verteilt ist.

Ist eine Zufallsvariable W nach $F_{m,n}$ -verteilt, so ist $1/W$ nach $F_{n,m}$ -verteilt. Deshalb sind die Tabellen der $F_{m,n}$ -Verteilung im allgemeinen auf den Fall $m \leq n$ beschränkt.

Schätzen von Parametern

Punktschätzung

Ein wesentliches Charakteristikum der statistischen Schlussweise ist die Tatsache, dass die aus einer Stichprobe gezogenen Schlussfolgerungen im allgemeinen nicht fehlerfrei sind. Die bisher vorgestellten Verteilungen für die Beschreibung von Zufallsvariablen hängen von Parametern ab (Erwartungswert μ , Varianz σ^2 , Wahrscheinlichkeit p), die unbekannt sind, solange nicht eine Vollerhebung durchgeführt wird. Aus einer Stichprobe werden deshalb Masszahlen (Stichprobenmittel \bar{x} , Stichprobenvarianz s^2 , relative Häufigkeit k/n) ermittelt, die als Schätzwerte der Parameter μ , σ^2 , p der Grundgesamtheit bezeichnet werden. Die Ermittlung der Schätzwerte erfolgt aufgrund einer Schätzfunktion.

Schätzfunktion: $f(\text{Stichprobe}) = \text{Parameter}$

Die Schätzfunktion soll dabei so gewählt werden, dass die gemeinsame Dichte von Grundverteilung und Schätzfunktion möglichst gross wird. Man spricht vom Maximum-Likelihood-Prinzip.

Für die Normalverteilung lauten die Maximum-Likelihood-Schätzer der Parameter:

$$\begin{aligned} \text{ML}(\mu) &= \bar{x} \\ \text{ML}(\sigma^2) &= s^2 \\ \text{ML}(p) &= \hat{p} = k/n \end{aligned}$$

Konfidenzintervall

Da bei den angegebenen Schätzverfahren nur ein einziger Zahlenwert als Schätzgrösse resultiert, spricht man dabei von Punktschätzungen. Eine Punktschätzung hat den Nachteil, dass kein Hinweis auf die Genauigkeit dieser Schätzung gegeben wird. Die Abweichung zwischen Punktschätzung und wahren Parameter (z.B. $|\bar{x} - \mu|$) kann erheblich sein, insbesondere bei kleinem Stichprobenumfang. Aussagen über die Genauigkeit einer Schätzung liefert die Konfidenzmethode. Bei ihr wird für den unbekanntem wahren Parameter ein Intervall bestimmt, das den wahren Parameter mit einer Wahrscheinlichkeit $1 - \alpha$ enthält.

Beispiel 9: Eine Stichprobe wird aus der Grundgesamtheit gezogen, um deren wahren Erwartungswert zu bestimmen (μ_0). Die Stichprobe liefert einen ML-Schätzer von $\bar{x} = 99.8$ für den wahren Erwartungswert μ_0 . Durch die Konfidenzmethode wird ein Intervall $[97.8; 101.8]$ bestimmt. Als Signifikanzniveau sei $\alpha = 5\%$ gewählt worden.

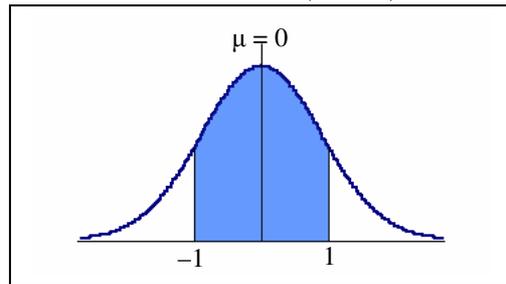
Es ist einsichtig, dass die Stichprobe nicht zwingend den wahren Wert von μ_0 liefert. Soll dieser wahre Wert mit Sicherheit bestimmt werden, so ist eine Vollerhebung notwendig, die aber aus naheliegenden Gründen meist nicht möglich ist.

Das Konfidenzintervall besagt, dass mit 95% Wahrscheinlichkeit ($= 1 - \alpha$) der wahre Wert des Erwartungswertes in der Grundgesamtheit (μ_0) zwischen 97.8 und 101,8 liegt. Die 5%, die zur Sicherheit fehlen, stammen aus der Unsicherheit über die Qualität der Stichprobe; vielleicht wurde eine Stichprobe gezogen, deren arithmetisches Mittel signifikant vom Erwartungswert abweicht. Das wäre dann "Pech", mit dem aufgrund seiner geringen Wahrscheinlichkeit ($< 5\%$) nicht gerechnet wurde.

Um die mathematische Behandlung des Vertrauensintervalls besser verstehen zu können, wird das Problem umgekehrt.

Beispiel 10: Gesucht ist die Wahrscheinlichkeit α , dass eine $N(0, 1)$ -verteilte Zufallsvariable X eine Stichprobe liefert, deren arithmetisches Mittel zwischen $[-1; 1]$ liegt. Die Stichprobe bestehe in diesem Fall aus einem einzigen Element x_1 , wodurch $\bar{x} = x_1$. Die gesuchte Wahrscheinlichkeit entspricht der blauen Fläche im nebenstehenden Diagramm einer Normalverteilung. Zur Berechnung wird auf die Tabelle der Verteilungsfunktion $\Phi(z)$ zurückgegriffen (Toutenburg S. 376f). In diesem Beispiel sei $\Phi(1)$ die Wahrscheinlichkeit des Intervalls $[-1; 1]$. Es ist:

$$\Phi(1) = \Phi(1) - \Phi(-1) = \Phi(1) - [1 - \Phi(1)] = 2 \cdot \Phi(1) - 1 = 2 \cdot 0.8413 - 1$$

$$\Phi(1) = 0.6826$$


Eine beliebige Stichprobe aus X liefert also mit rund 68% Wahrscheinlichkeit ein Stichprobenmittel im Intervall $[-1; 1]$. Das bedeutet aber auch, dass rund 32% aller i.i.d. Stichproben aus dieser Grundgesamtheit ein Stichprobenmittel liefert, das ausserhalb dieses Intervalls liegt (bei einem wahren Erwartungswert $\mu_0 = 0$!)

Bei der Berechnung des Konfidenzintervalls wird normalerweise umgekehrt vorgegangen: es wird für eine fixe Wahrscheinlichkeit (z.B. 95%) das Intervall bestimmt. Charakterisiert wird ein solches Intervall durch das Signifikanz- oder Konfidenzniveau α , das der Wahrscheinlichkeit entspricht, mit der eine Stichprobe aus der Grundgesamtheit nicht innerhalb des Intervalls liegt (man spricht in diesem Fall vom Fehler 1. Art). Das Signifikanzniveau rechnet sich $= 1 - \text{Intervallwahrscheinlichkeit}$.

Das Konfidenzintervall für μ bei bekanntem $\sigma^2 = \sigma_0^2$ berechnet sich direkt aus der Normalverteilung:

$$P\left(\left|\frac{\bar{x} - \mu}{\sigma_0} \cdot \sqrt{n}\right| \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

In Worten: die Wahrscheinlichkeit, dass der Abstand des Stichprobenmittels vom wahren Erwartungswert kleiner ist als die halbe Intervallbreite (beim zweiseitigen Testen) ist gleich $1 - \text{dem Signifikanzniveau}$ ($= 1 - 5\% = 95\%$). $z_{1-(\alpha/2)}$ ist der Wert der Standardisierungsfunktion $X \rightarrow Z$ der Standardnormalverteilung für die durch $(1 - (\alpha/2))$ festgelegte Wahrscheinlichkeit.

Damit ergibt sich für das Intervall selbst:

$$\bar{x} - \left[z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right] \leq \mu_0 \leq \bar{x} + \left[z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

Für $\alpha = 5\%$ ist $z_{1-(\alpha/2)} = 1.960$. Für $\alpha = 2.5\%$ ist $z_{1-(\alpha/2)} = 2.241$. Mit kleinerem α wird somit das Intervall grösser (vgl. Fehler 1. und 2. Art, S. 23f).

Bei unbekanntem σ_0^2 wird das Konfidenzintervall für μ an sich gleich berechnet, statt der Gesamtvarianz muss nun allerdings die Stichprobenvarianz s^2 verwendet werden, was auf die t-Verteilung (mit $v = n - 1$ Freiheitsgraden) führt. Es ergibt sich:

$$\bar{x} - \left[t_v(1-\alpha) \cdot \frac{s}{\sqrt{n}} \right] \leq \mu_0 \leq \bar{x} + \left[t_v(1-\alpha) \cdot \frac{s}{\sqrt{n}} \right]$$

Schätzen einer Binomialwahrscheinlichkeit

Für die Schätzung der wahren Wahrscheinlichkeit p in einem Binomialexperiment wird der ML-Schätzer von $p = x/n = p^\wedge$ verwendet. Dabei ist x/n die relative Häufigkeit der Ausprägung 1 (zum Beispiel 'Kopf' beim Münzwurf). Ist über die Varianz der Stichprobe nichts bekannt, so kann diese aufgrund der Wahrscheinlichkeit und des Stichprobenumfangs geschätzt werden. Da die Grundgesamtheit die Varianz $\sigma^2 = n \cdot p \cdot (1 - p)$ besitzt (wobei p unbekannt ist), gilt für die Varianz der Schätzung von $p^\wedge = X/n$:

$$Var(p^\wedge) = \frac{p(1-p)}{n}$$

Diese wird geschätzt durch:

$$s_{p^\wedge}^2 = \frac{p^\wedge(1-p^\wedge)}{n}$$

Bei der Varianzschätzung handelt es sich – wie bei der ML-Schätzung für p – um eine Punktschätzung.

Oftmals ist man allerdings an einem Vertrauensintervall für p interessiert, um zum Beispiel die wahre Wahrscheinlichkeit einer unbekannt binomischen Verteilung anhand einer Stichprobe zu bestimmen. Man bedient sich dazu der Normalverteilung und wendet den Approximationssatz an, dass für $n \cdot p \cdot (1 - p) > 9$ die Binomial- durch die Normalverteilung approximiert werden kann (vgl. S. 19). In diesem Fall ergibt sich für das Vertrauensintervall von p :

$$p^\wedge - \left[z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p^\wedge(1-p^\wedge)}{n}} \right] \leq p \leq p^\wedge + \left[z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p^\wedge(1-p^\wedge)}{n}} \right]$$

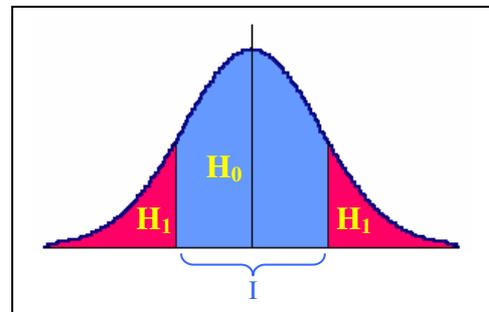
Parametrische Tests

Nullhypothese, Fehler 1. und 2. Art

Bei vielen Tests geht es um die Frage: 'Stammt die Stichprobe mit dem Stichprobenparameter θ (z.B. \bar{x} , s^2 , x/n) aus der Grundgesamtheit mit bekanntem Parameter Θ (μ , σ^2 , p)?' Man spricht dann von parametrischen Tests. Grundlage eines jeden Tests ist die Nullhypothese H_0 . Charakteristikum der Nullhypothese ist, dass sie die Gleichheit von Stichprobenparameter θ und Gesamtheitsparameter Θ unterstellt ($\bar{x} = \mu$, $s^2 = \sigma^2$, $x/n = p$).

A priori gilt die Nullhypothese als richtig. Erst wenn die Wahrscheinlichkeit für die Richtigkeit der Nullhypothese unter einen bestimmten Wert α sinkt (das Signifikanzniveau), wird H_0 verworfen. In diesem Fall wird (nach dem Prinzip des *tertium non datur*) die Arbeitshypothese H_1 akzeptiert, die das Gegenereignis zu H_0 darstellt (z.B. $x/n \neq p$). Dies ist gleichbedeutend mit der Aussage: 'Die Stichprobe stammt nicht aus der Grundgesamtheit' oder 'Die Annahme über die Parameter der Grundgesamtheit ist falsch.'

Aus dem Signifikanzniveau und der Nullhypothese lässt sich analog zum Konfidenzintervall ein Bereich I bestimmen, in dem die Nullhypothese akzeptiert wird.



Beispiel 11: Bei der Qualitätskontrolle der Münzerei in Sikinien geht man davon aus, dass der mittlere Goldgehalt einer Münze $\mu = 50\%$ ist. Der tatsächliche Goldgehalt sei normalverteilt mit $\mu = 50$, $\sigma^2 = 25$. Eine Stichprobe von $n = 100$ Münzen liefert einen mittleren Goldgehalt $\bar{x} = 42\%$. Die Nullhypothese lautet $H_0: \bar{x} = \mu$. α sei 0.05. Wie gross ist die Wahrscheinlichkeit, aus einer normalverteilten Grundgesamtheit $N(50, 25)$ eine Stichprobe mit $\bar{x} = 42$ zu ziehen?

$$z = \frac{\bar{x} - \mu}{\sigma}, z = -1.6. \rightarrow \Phi(z) = \Phi(-1.6) = 1 - \Phi(1.6) = 0.0548 > 0.05.$$

Die Wahrscheinlichkeit, eine Stichprobe mit $\bar{x} = 42$ zu ziehen bei einem wahren $\mu = 50$ ist grösser als 5%. Die Nullhypothese wird deshalb nicht verworfen, die Münzen in Sikinien scheinen also tatsächlich einen Goldgehalt von 50% zu haben.

Bei der Testentscheidung können zwei Arten von Fehler begangen werden. Entweder wird H_0 verworfen (die Münzen seien schlecht), obwohl H_0 richtig ist (das wahre μ ist wirklich 50%), oder H_0 wird akzeptiert (die Münzen seien gut), obwohl H_0 falsch ist (das wahre μ ist nicht 50%).

Fehler 1. Art: H_0 wird verworfen, obwohl H_0 richtig ist.

Die Wahrscheinlichkeit, den Fehler 1. Art zu begehen, entspricht gerade dem Signifikanzniveau α . Es spricht systematisch nichts dagegen, aus einer Grundgesamtheit mit wahren $\mu = 50$ eine Stichprobe mit $\bar{x} = 10$ zu ziehen. Die Wahrscheinlichkeit für diesen Fall ist allerdings so gering, dass unter diesen Bedingungen die Annahme, H_0 sei trotzdem richtig, unhaltbar ist.

Fehler 2. Art: H_0 wird akzeptiert, obwohl H_0 falsch ist.

H_0 ist falsch, das heisst, der wahre Erwartungswert μ_0 der Grundgesamtheit ist nicht 50. Man bezeichnet die Wahrscheinlichkeit, den Fehler 2. Art zu begehen als β . Anders als die Signifikanz α lässt sich das β jedoch nicht pauschal angeben. Das β ist abhängig vom wahren μ_0 der Grundgesamtheit: $\beta = \beta(\mu_0)$. Im Grenzfall, in dem das wahre μ_0 mit dem erwarteten μ übereinstimmt ($\mu = \mu_0 = 50$), ist $\beta = 1 - \alpha$.

Die Güte eines Tests ist definiert durch seine Fähigkeit, den Fehler 2. Art *nicht* zu machen. Auch die Güte eines Tests ist abhängig vom tatsächlichen μ_0 . Es ist

$$g(\mu_0) = 1 - \beta(\mu_0)$$

Ein idealer Test hat eine Gütefunktion der Form $g(\mu_0) = \begin{cases} 1, & \text{wenn } \mu_0 \neq \mu \\ \alpha, & \text{wenn } \mu_0 = \mu \end{cases}$.

Ein-Stichproben-Verfahren

Der einfachste Fall eines Ein-Stichproben Parameter-tests für den Mittelwert μ wurde bereits weiter oben in Beispiel 11 vorgeführt (Prüfen des Mittelwerts bei bekannter Varianz). Es soll geprüft werden, ob der unbekannte Erwartungswert μ einer $N(\mu, \sigma^2)$ -verteilten Zufallsvariable X einen bestimmten Wert $\mu = \mu_0$ besitzt. Dabei ist zunächst die Varianz $\sigma^2 = \sigma_0^2$ bekannt. Unter diesen Bedingungen kann direkt die (Standard-) Normalverteilung als Prüfverteilung angewendet werden. Man spricht vom einfachen Gauss-Test.

$H_0:$ $\mu = \mu_0$

Prüfgrösse: $t = \frac{\bar{x} - \mu_0}{\sigma_0} \cdot \sqrt{n} \sim N(0,1)$

Testentscheidung: H_0 verwerfen, wenn $t > z_{1-(\alpha/2)}$

Getestet werden kann einseitig und zweiseitig. Einseitig wird getestet, wenn eine Vermutung über die Richtung der Abweichung des wahren μ_0 vom geschätzten μ besteht (z.B. $\mu < \mu_0$), oder nur eine Abweichungsrichtung interessiert / relevant ist. Ist nichts darüber bekannt, so wird zweiseitig getestet ($\mu \neq \mu_0$). Wird einseitig getestet, so hat dies einen Einfluss auf die Nullhypothese. Diese lautet dann nicht mehr $\mu = \mu_0$, sondern z.B. $H_0: \mu \geq 2$. Es ist dann die Arbeitshypothese, zu deren Gunsten H_0 abgelehnt wird, $H_1: \mu < 2$.

Schwellenwerte sind (für $\alpha = 5\%$) $z_{1-(\alpha/2)} = 1.960$ (beim zweiseitigen Test) und $z_{(1-\alpha)} = 1.645$ (beim einseitigen Test).

Sehr oft ist das σ_0^2 der Grundgesamtheit nicht bekannt. In diesem Fall nutzt man anstelle der Gesamtvarianz die Stichprobenvarianz s^2 als Schätzung. Dies führt auf den einfachen Student t-Test.

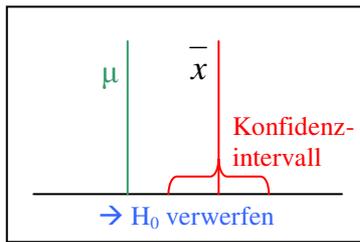
$H_0:$ $\mu = \mu_0$

Prüfgrösse: $t = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n} \sim t_{v=n-1}$

Testentscheidung: H_0 verwerfen, wenn $t > t_{v; (1-\alpha)}$

Dabei ist stammen $\bar{x} = \mu$, s^2 und n aus der Stichprobe (Stichprobenmittel, Stichprobenvarianz, Stichprobenumfang). μ_0 ist der Erwartungswert der Grundgesamtheit. Der Student t-Test beantwortet z.B. die Frage 'Stammt die Stichprobe mit (\bar{x}, s^2, n) aus der Grundgesamtheit mit Parameter μ_0 ?' bzw. 'Ist die Annahme über den wahren Wert von μ_0 in der Grundgesamtheit richtig?'.

Aufgrund der Prüfverteilung lässt sich ein Konfidenzintervall bestimmen. Für den Student t-Test ist dieses:



$$\bar{x} \pm t_{v; 1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

In diesem Fall wird H_0 verworfen, wenn μ_0 nicht im Konfidenzintervall liegt.

Zwei-Stichproben-Verfahren

Grundlage dieser Tests sind zwei Zufallsvariablen X, Y . Diese seien (bis auf weiteres) i.i.d. normalverteilt, also unabhängig. Es sei $X \sim N(\mu_x, \sigma_x^2)$ und $Y \sim N(\mu_y, \sigma_y^2)$. Als erstes soll auf die Gleichheit der Mittelwerte $\mu_x = \mu_y$ getestet werden, wobei $\sigma_x^2 = \sigma_y^2 = \sigma^2$ bekannt ist. Dies führt auf den sogenannten doppelten Gauss-Test.

$H_0: \mu_x = \mu_y$

Prüfgrösse:
$$t(x, y) = \frac{\bar{x} - \bar{y}}{\sqrt{n_y \sigma_x^2 + n_x \sigma_y^2}} \cdot \sqrt{n_x \cdot n_y} \sim N(0,1)$$

Testentscheidung: H_0 verwerfen, wenn $t > z_{1-(\alpha/2)}$

Es sind n_x, n_y die Stichprobenumfänge der unabhängigen Stichproben aus X bzw. Y . Die Testentscheidung erfolgt analog zum Ein-Stichproben-Gauss-Test.

Sind die Varianzen zweier Verteilungen unbekannt, aber gleich, so führt dies auf den doppelten t-Test. Die gemeinsame Varianz σ^2 wird durch die sogenannte gepoolte Stichprobenvarianz geschätzt, die beide Stichproben mit einem Gewicht relativ zu ihrer Grösse verwendet.

Gemeinsame Varianz:
$$S^2 = \frac{(n_1 - 1) \cdot s_x^2 + (n_2 - 1) \cdot s_y^2}{n_1 + n_2 - 2}$$

$H_0: \mu_x = \mu_y$

Prüfgrösse:
$$t(x, y) = \frac{\bar{x} - \bar{y}}{S} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \sim t_{n_1 + n_2 - 2};$$

Testentscheidung: H_0 verwerfen, wenn $t > t_{v; (1-\alpha)}$
 mit $v = n_1 + n_2 - 2$

Sind die Varianzen von X und Y ungleich und unbekannt, so gibt es keine exakte Lösung, um die Nullhypothese $\mu_x = \mu_y$ zu prüfen. Als Näherung verwendet man in diesem Fall den Welch-Test.

Wie oben werden wieder zwei unabhängige Zufallsvariablen mit $E(X) = \mu_x$ und $E(Y) = \mu_y$ betrachtet. Die Annahme der Unabhängigkeit der Variablen wird nun aufgegeben, die beiden

Variablen als abhängig angenommen. Diese Abhängigkeit kann in der Praxis beispielsweise dadurch entstehen, dass an einem Objekt zwei Merkmale gleichzeitig beobachtet werden oder ein Merkmal an einem Objekt zu verschiedenen Zeitpunkten beobachtet wird. Man spricht dann von einer gepaarten oder verbundenen Stichprobe oder von einem matched-pair Design.

Da beide Zufallsvariablen zum selben Objekt gehören, ergibt das Bilden einer Differenz einen Sinn. Mit $D = X - Y$, wobei $D = \sum D_i = \sum x_i - y_i$. Unter $H_0: \mu_x = \mu_y$ ist die erwartete Differenz gleich Null, es gilt $E(D) = \mu_D = 0$. Es wird vorausgesetzt, dass D unter $H_0: \mu_x = \mu_y$ bzw. $H_0: \mu_D = 0$ normalverteilt ist, d.h. es ist $D \sim N(0, \sigma_D^2)$.

$H_0: \mu_x = \mu_y \rightarrow \mu_D = 0$

Prüfgrösse:
$$t = \frac{\bar{D}}{s_D} \cdot \sqrt{n} \sim t_{v=n-1}$$

Differenzenvarianz:
$$s_D^2 = \frac{\sum [(D_i - \bar{D})^2]}{n-1}$$

Testentscheidung: H_0 verwerfen, wenn $t > t_{v; (1-\alpha)}$

Bei diesem Test werden die zwei abhängigen Variablen X und Y zu einer Variablen D zusammengefasst. Damit wird ein Zwei-Stichproben-Problem durch die Differenzenbildung auf ein Ein-Stichproben-Problem zurückgeführt. Die Differenzvariable D ist wie ein Ein-Stichproben-Problem zu behandeln. Da die Varianz der Differenz unbekannt ist, wird die Stichprobenvarianz aus D ermittelt (Differenzenvarianz). Es folgt ein einfacher t-Test.

Gleichheit zweier Binomialverteilungen

Erneut wird ein Zwei-Stichproben-Problem betrachtet, diesmal wieder mit zwei *unabhängigen* Zufallsvariablen X und Y . Dabei ist $X_i \sim B(1; p_1)$ und $Y_i \sim B(1; p_2)$ -verteilt. Somit ist $X = \sum X_i \sim B(n_1; p_1)$ und $Y = \sum Y_i \sim B(n_2; p_2)$. Geprüft wird die Nullhypothese, dass den beiden Verteilungen die gleiche Wahrscheinlichkeit $p_1 = p_2 = p$ zugrunde liegt.

Ähnlich wie beim matched-pair Design wird dazu die Differenz $D = \frac{X}{n_1} - \frac{Y}{n_2} = p^{\wedge}_1 - p^{\wedge}_2$ der

beiden Variablen gebildet. Für hinreichend grosse n_1 und n_2 sind $\frac{X}{n_1}$ und $\frac{Y}{n_2}$ nach dem

Approximationssatz (vgl. S. 19) normalverteilt.

$$\frac{X}{n_1} \sim N\left(p_1; \frac{p_1 \cdot (1-p_1)}{n_1}\right)$$

$$\frac{Y}{n_2} \sim N\left(p_2; \frac{p_2 \cdot (1-p_2)}{n_2}\right)$$

$H_0: p^{\wedge}_1 = p^{\wedge}_2 = p^{\wedge} \Leftrightarrow D = p^{\wedge}_1 - p^{\wedge}_2 = 0$

Gemeinsame Wahrscheinlichkeit:
$$p^{\wedge} = \frac{X + Y}{n_1 + n_2}$$

Prüfgrösse:
$$t = \frac{D}{\sqrt{p^{\wedge} \cdot (1 - p^{\wedge}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0;1)$$

Testentscheidung: H_0 verwerfen, wenn $t > z_{1-(\alpha/2)}$.

Nichtparametrische Tests

In die bisherigen Prüfverfahren ging der Verteilungstyp der Stichprobenvariablen ein, der Typ der Verteilung war also bekannt. Die zu prüfende Hypothesen bezogen sich auf Parameter dieser Verteilung. Die für Parameter bekannter Verteilungen konstruierten Prüfverfahren heissen parametrische Tests, da die Hypothesen Parameterwerte festlegen. Möchte man hingegen Lage- oder Streuungsalternativen bei stetigen Variablen prüfen, deren Verteilung nicht bekannt ist, so sind die im folgenden dargestellten nichtparametrischen Tests zu verwenden.

Anpassungstest

Kennt man den Verteilungstyp der der Stichprobe zugrunde liegenden Zufallsvariablen nicht, so kann man mit Hilfe der Anpassungstests prüfen, ob diese Variable eine bestimmte Verteilung (z.B. die Normalverteilung) besitzt. Es soll also untersucht werden, wie 'gut' sich eine beobachtete Verteilung der hypothetischen Verteilung anpasst.

Bei der Konstruktion des Tests ist es notwendig, die Verteilung der Testgrösse unter der Nullhypothese zu kennen. Daher sind alle Anpassungstests so aufgebaut, dass die eigentlich interessierende Hypothese als Nullhypothese und nicht – wie sonst üblich – als Alternative formuliert wird. Deshalb kann man mit einem Anpassungstest auch kein statistischer Nachweis geführt werden, dass ein bestimmter Verteilungstyp vorliegt, sondern es kann nur nachgewiesen werden, dass ein bestimmter Verteilungstyp nicht vorliegt.

Der wohl bekannteste Anpassungstest ist der χ^2 -Anpassungstest. Die Teststatistik wird so konstruiert, dass sie die Abweichungen der unter H_0 erwarteten von den tatsächlich beobachteten absoluten Häufigkeiten misst. Hierbei ist jedes Skalenniveau zulässig. Um jedoch die erwarteten Häufigkeiten zu berechnen, ist es bei ordinalem oder stetigem Datenniveau notwendig, die Stichprobe X in k Klassen einzuteilen. Dabei sollte die Klasseneinteilung nicht zu fein gewählt werden, um eine genügend grosse Anzahl an Beobachtungen in den einzelnen Klassen zu gewährleisten.

Im folgenden sei die hypothetische Verteilung, auf die getestet wird, die Normalverteilung:
 $F_0(x) = \Phi(\mu, \sigma^2)$

H_0 :
$$F(x) = \Phi(\mu, \sigma^2)$$

$$X \sim N(\bar{x}, s^2)$$

 'Die unbekannte Verteilung X entspricht der Normalverteilung'

Prüfgrösse:
$$t = \sum_{i=1}^k \frac{(n_i - N_i)^2}{N_i} \sim \chi_{k-1-r}^2$$

Testentscheidung: H_0 verwerfen, wenn $t > \chi_{k-1-r; (1-\alpha)}^2$

Dabei ist n_i die absolute Häufigkeit der Stichprobe in der Klasse i und N_i die absolute Häufigkeit der hypothetischen Verteilung. Die Prüfgrösse t ist t -verteilt mit $v = k-1-r$ Freiheitsgraden, wobei k die Anzahl Klassen und r die Anzahl Parameter der Verteilungssypothese ist.

Beispiel 12: Eine Stichprobe mit unbekannter Verteilungsfunktion vom Umfang $n = 1000$ weise die Parameter ($\bar{x} = 51.67$; $s^2 = 337.15$) auf. Als hypothetische Verteilung wird die Normalverteilung mit den Parameter $N(\mu = \bar{x}, \sigma^2 = s^2)$ angenommen. Die Stichprobe wird in 10 Klassen unterteilt. Die anschliessende Tabelle enthält die absolute Häufigkeit der Klasse i aus der Stichprobe [= n_i], die erwartete absolute Häufigkeit aufgrund der hypothetischen Normalverteilung [= N_i], die Differenz der beiden und die quadrierte und durch die absolute Häufigkeit der Hypothese dividierte Differenz der beiden:

| i | k | n_i | $\Phi(k)$ | N_i | $n_i - N_i$ | $\frac{(n_i - N_i)^2}{N_i}$ |
|-----|-------------------|-------|-----------|-------|-------------|-----------------------------|
| 1 | $0 \leq x < 10$ | 5 | 0.00248 | 9 | -4 | 1.778 |
| 2 | $10 \leq x < 20$ | 42 | 0.0116 | 31 | 11 | 3.903 |
| 3 | $20 \leq x < 30$ | 98 | 0.0427 | 76 | 22 | 6.368 |
| 4 | $30 \leq x < 40$ | 122 | 0.119 | 145 | -23 | 3.648 |
| 5 | $40 \leq x < 50$ | 142 | 0.2643 | 200 | -58 | 16.820 |
| 6 | $50 \leq x < 60$ | 285 | 0.4641 | 210 | 75 | 26.786 |
| 7 | $60 \leq x < 70$ | 143 | 0.6736 | 165 | -22 | 2.933 |
| 8 | $70 \leq x < 80$ | 102 | 0.8389 | 99 | 3 | 0.091 |
| 9 | $80 \leq x < 90$ | 51 | 0.9382 | 43 | 8 | 1.488 |
| 10 | $90 \leq x < 100$ | 10 | 0.98124 | 14 | -4 | 1.143 |

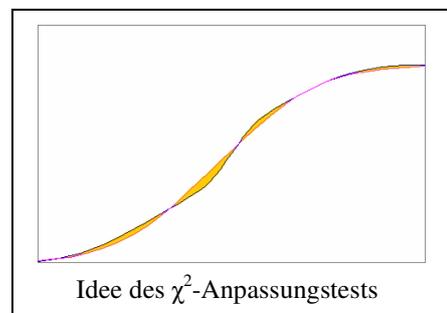
Diese Tabelle liefert $t = 64.867$.

Die Testentscheidung wird anhand der χ^2 -Verteilung gefällt. Das Signifikanzniveau sei $\alpha = 5\%$. Das $(1 - \alpha)$ -Quantil der χ^2 -Verteilung mit $v = 7$ Freiheitsgraden ist $\chi^2_{v; (1-\alpha)} = 14.1 < t$.

Die Nullhypothese, dass die unbekannte Verteilung einer Normalverteilung entspricht, wird also verworfen. Über die wirkliche Verteilung der Stichprobe kann allerdings aufgrund dieses Tests eine Aussage gemacht werden.

Zur genauen Berechnung der Werte siehe auch im Anhang.

Der χ^2 -Anpassungstest prüft also, wie stark zwei Verteilungsfunktionen über alles gesehen voneinander abweichen. Der Kolmogorov-Smirnov-Anpassungstest demgegenüber sucht den grössten Abstand der beiden Verteilungsfunktionen (empirische und hypothetische). Der χ^2 -Anpassungstest hat bei stetigen Variablen den Nachteil, dass eine Gruppierung der Werte notwendig ist. Insbesondere kann die Klassenbildung auch die Teststatistik und damit das Testergebnis beeinflussen.



Idee des χ^2 -Anpassungstests

Dieses Problem wirkt sich besonders stark bei kleinen Stichproben aus. In diesen Fällen ist der Kolmogorov-Smirnov-Anpassungstest dem χ^2 -Anpassungstest vorzuziehen.

Homogenitätstest

Im Gegensatz zu den Anpassungstest vergleichen die Homogenitätstests die Verteilungen zweier Zufallsvariablen miteinander.

Der Kolmogorov-Smirnov-Homogenitätstest (das Zweistichproben-Pendant zum K-S-Anpassungstest) prüft allgemeine Hypothesen der Art 'Die beiden Verteilungen sind gleich'. Seien X und Y zwei unabhängige Stichproben.

H_0 : $F(t) = G(t)$
 X und Y sind gleich verteilt.

Prüfgrösse: $K = \max_{t \in S} |F(t) - G(t)|$

Testentscheidung: H_0 wird verworfen, wenn $K > k_{n,m,(1-\alpha)}$

(Die kritischen Werte für k sind tabelliert, allerdings nicht im Toutenburg.)

Wenn zwei Verteilungen sich nur bezüglich der Lage unterscheiden, kann der U-Test von Mann und Whitney angewendet werden. Der U-Test von Mann und Whitney ist ein Rangtest. Er ist das nichtparametrische Gegenstück zum t-Test und wird angewendet, wenn die Voraussetzungen für den t-Test fehlen oder begründete Zweifel bestehen.

H_0 : $P(x_i > y_i) = 0.5$
Die Lage der Verteilungen ist gleich.

Prüfgrösse: $Z = \frac{U - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}} \sim N(0,1)$

Testentscheidung: H_0 wird verworfen, wenn $|z| > z_{1-(\alpha/2)}$.

Zur Berechnung von U siehe Toutenburg S. 175.

Der Vorzeichen-Test liefert ein Testverfahren für Lagealternativen im matched-pair Design, also bei zwei verbundenen Stichproben. Voraussetzung ist mindestens ein ordinales Skalenniveau in beiden Stichproben.

H_0 : $P(X < Y) = P(Y < X)$
Die Lage der Verteilungen ist in beiden Stichproben gleich.

Prüfgrösse: $t = \sum D_i$
mit $D_i = \begin{cases} 1 & \text{falls } X_i < Y_i \\ 0 & \text{sonst} \end{cases}$

Unter H_0 ist t binomialverteilt mit den Parametern $p = 0.5$ und n .

Testentscheidung: H_0 wird verworfen, wenn $t < b_{n, 1-(\alpha/2)}$

Wobei $b_n :_{1-(\alpha/2)}$ das $(1-(\alpha/2))$ -Quantil einer Binomialverteilung mit den Parametern $B(n, p = 0.5)$ bildet.

Für $n \geq 20$ ist t approximativ $N(n/2, n/4)$ -verteilt. In diesem Fall wird H_0 abgelehnt, falls gilt:

$$|z| = \frac{|2t - n|}{\sqrt{n}} > z_{1-\frac{\alpha}{2}}$$

Als letztes nichtparametrisches Verfahren sei noch die Rangkorrelation nach Spearman erwähnt.

H_0 : $r_s = 0$
Es besteht keine Korrelation.

Prüfgrösse (für $n \geq 30$): $z = r_s \sqrt{n-1}$

mit $r_s = 1 - \frac{6 \cdot \sum_{i=1}^n (r_{x;i} - r_{y;i})^2}{n(n^2 - 1)}$

Testentscheidung: H_0 verwerfen, wenn $z > z_{1-(\alpha/2)}$

Es bedeuten:

$r_{x;i}$ = Rang in der Stichprobe X

$r_{y;i}$ = Rang in der Stichprobe Y

n = Stichprobenumfang.

z ist $N(0,1)$ normalverteilt.

Die Ablehnung der Nullhypothese zeigt lediglich, dass eine Korrelation zwischen X und Y besteht, diese also nicht unabhängig sind. Der Test macht weder eine Aussage über die Stärke der Korrelation noch über ihre Richtung.

TEIL IV: URSACHE-WIRKUNGS BEZIEHUNGEN

Lineare Regression

Mit dem Korrelationskoeffizienten ρ wurde bereits ein dimensionsloses Mass eingeführt, das die Stärke und Richtung des linearen Zusammenhangs zwischen X und Y misst. Ziel der Linearen Regression ist es, diesen Zusammenhang zwischen den beiden stetigen Zufallsvariablen X und Y in einem einfachen Modell zu erfassen.

Die beiden Variablen X und Y werden am gleichen Untersuchungsobjekt erhoben. Dabei kann eine Variable als gegeben oder beeinflussbar angesehen werden (X), während die andere Variable (Y) als Reaktion auf X beobachtet wird. Man spricht deshalb auch von Response.

Dies ist die allgemeine Struktur einer Ursache-Wirkungs Beziehung zwischen X und Y . Mathematisch lässt sich eine solche Beziehung als Funktion auffassen. Es ist:

$$Y = f(X)$$

Das einfachste Modell einer Ursache-Wirkungs Beziehung zwischen X und Y ist die lineare Gleichung:

$$Y = \beta_0 + \beta_1 X$$

Eine lineare Funktion liefert einen einfach zu handhabenden mathematischen Ansatz und ist auch insofern gerechtfertigt, als sich viele Funktionstypen gut durch lineare Funktionen stückweise approximieren lassen.

Das einfachste lineare Modell ist das univariate lineare Modell

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Dabei sind (im Gegensatz zur deskriptiven Regression) β_0 und β_1 unbekannte Modellparameter und ε eine zufällige Fehlervariable für die gelten soll

$$\begin{aligned} E(\varepsilon) &= 0 \\ \text{Var}(\varepsilon) &= \sigma^2 \end{aligned}$$

Für Y folgt für ein gegebenes X aus diesen Bedingungen sofort:

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1 X \\ \text{Var}(Y) &= \sigma^2 \end{aligned}$$

Die univariate Regression stellt nur einen Spezialfall dar. Im allgemeinen führen Probleme im Problemkreis der Wirtschafts- und Sozialwissenschaften auf multivariate lineare Regressionsmodelle.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Es stellt sich nun die Frage, welches Modell (welche Gleichung) sinnvoll Y beschreibt, d.h. welche Regressoren X_i ins Modell aufgenommen werden sollen. Im Grunde wird dazu mit

jedem Regressor ein Test durchgeführt unter der Nullhypothese $H_0: x_i$ hat keinen Einfluss auf Y (d.h. $\beta_i = 0$). Wird diese Nullhypothese verworfen, so wird der Regressor x_i ins Modell aufgenommen.

Aus theoretischen Gründen erfolgt die Modellbildung schrittweise, d.h. in einem ersten Schritt werden die Korrelation aller zur Auswahl stehenden x_i mit Y bestimmt und dasjenige x_i mit der betragsmässig höchsten Korrelation als erstes ins Modell aufgenommen, wenn das zugehörige β_i signifikant von der Nullhypothese $H_0: \beta_i = 0$ abweicht.

Im zweiten Schritt werden die Korrelation der verbliebenen (nicht ins Modell aufgenommenen) Regressoren mit Y ermittelt, dann der Regressor mit der grössten Korrelation ins Modell aufgenommen. Dieser Prozess wiederholt sich so lange, bis die Abweichung für β_j von der Nullhypothese $H_0: \beta_j = 0$ nicht mehr signifikant ist.

Das Prinzip der Modellwahl lässt sich auch anhand der SQ (Sum of Squares) formulieren. Es gilt:

$$SQ_{\text{Total}} = SQ_{\text{Modell}} + SQ_{\text{Residual}}$$

$$SQ_{\text{Total}} = \sum_{i=1}^n SQ(x_i) + SQ_{\text{Residual}}$$

SQ_{Residual} wird nicht durch das Modell erklärt und geht in die Fehlervariable ε ein. Es werden so lange zusätzliche Variablen ins Modell aufgenommen, bis die Differenz von SQ_{Residual} ; vorher und SQ_{Residual} , Nachher nicht mehr signifikant ist. Zur Prüfung der Signifikanz wird der F-Change Test verwendet.

Beispiel 13: Es soll ein Regressionsmodell über die Einflussfaktoren der abhängigen Variable $Y =$ aktuelles Gehalt erstellt werden. In einer Stichprobe wurden vier mögliche Einflussgrössen ermittelt, die in der folgenden Tabelle enthalten sind.

| | X_1 : Einstiegsgehalt | X_2 : Alter | X_3 : Berufserfahrung | X_4 : Geschlecht |
|-------------|----------------------------|------------------|----------------------------|-----------------------|
| Korrelation | -0.88 | 0.73 | 0.70 | -0.92 |
| Signifikanz | 0.000 | 0.001 | 0.017 | 0.000 |

Als erstes wird also X_4 ins Modell aufgenommen. Es ergibt sich somit im ersten Schritt: $Y = \beta_0 + \beta_4 X_4$.

Ausgehend von dieser ersten Schätzung für Y werden nun die verbleibenden Regressoren auf ihre Korrelation und Signifikanz hin überprüft:

| | X_1 : Einstiegsgehalt | X_2 : Alter | X_3 : Berufserfahrung |
|-------------|----------------------------|------------------|----------------------------|
| Korrelation | -0.79 | 0.82 | 0.66 |
| Signifikanz | 0.020 | 0.006 | 0.044 |

Schritt 2: $Y = \beta_0 + \beta_4 X_4 + \beta_2 X_2$

| | | |
|--------------------|-----------------------------------|-----------------------------------|
| | X_1 : <i>Einstiegsgehalt</i> | X_3 : <i>Berufserfahrung</i> |
| <i>Korrelation</i> | -0.72 | 0.66 |
| <i>Signifikanz</i> | 0.058 | 0.120 |

Weder X_1 noch X_3 tragen eine signifikante Verbesserung zum Modell bei und finden deshalb keinen Eingang. Ein gutes Modell für Y ist also:

$$Y = \beta_0 + \beta_4 X_4 + \beta_2 X_2$$

Ein komplexes Problem wird im Buch (Toutenburg, S. 214-218) besprochen.

Vierfeldertafel

χ^2 -Kontingenztest

Im (linearen) Regressionmodell sind X und Y stetige Zufallsvariablen. In der Vierfeldertafel soll eine Ursache-Wirkungs Beziehung zwischen zwei binären Variablen X und Y dargestellt werden. Das heisst, X und Y können jeweils nur die codierten Werte $\{0; 1\}$ annehmen. Als Konvention bezeichnet das Feld oben links in der Vierfeldertafel (1,1) die Risikoausprägung (im Bild rechts gelb dargestellt).

| | | |
|---|---|---|
| | 1 | 0 |
| X | 1 | 0 |
| | 1 | 0 |
| | Y | |

Ein verbreitetes Verfahren zur Analyse einer Vierfeldertafel ist die χ^2 -Statistik. Abstrakt formuliert prüft die χ^2 -Statistik nach der Methode

$$\chi^2 = \frac{(\text{Realität} - \text{Hypothese})^2}{\text{Hypothese}}$$

In der Vierfeldertafel wird χ^2_1 , also mit $v = 1$ Freiheitsgraden getestet. Dies ist so, weil ein bekanntes Feld durch die bekannten Randsummen alle anderen Felder der Tafel determiniert.

Die (Null-)Hypothese des χ^2 -Kontingenztests ist die Unabhängigkeit von X und Y . Wird diese Nullhypothese abgelehnt, so sind X und Y nicht unabhängig, es besteht also ein Zusammenhang zwischen den beiden.

Beispiel 14 1000 Personen wurden befragt nach ihren Ernährungsgewohnheiten X (ordinal skaliert in 10 Stufen von 'nur Junkfood' bis 'vegetarische Diät' und ihrem Body Mass Index (in der Stichprobe traten BMI zwischen 18 und 30 auf). Daraus wurde eine Kontingenztafel mit $X = 10$ Zeilen und $Y = 12$ Spalten angefertigt. Diese Kontingenztafel wurde reduziert auf die nebenstehende Vierfeldertafel mit den Ernährungsausprägungen X : $\{1 = \text{ungesund}; 0 = \text{gesund}\}$ und die Gewichtsausprägungen Y : $\{1 = \text{Übergewicht}; 0 = \text{Normalgewicht}\}$.

| | | | |
|----------|-----|-----|----------|
| | 1 | 0 | Σ |
| X | 312 | 234 | 546 |
| | 198 | 256 | 454 |
| Σ | 510 | 490 | 1000 |
| | 1 | 0 | Σ |
| | Y | | |

Die Risikoausprägung lautet hier {1 = ungesund; 1 = Übergewicht}, es wird also angenommen, dass jemand der sich ungesund ernährt zu Übergewicht neigt. Um den χ^2 -Test durchführen zu können, muss zuerst die Unabhängigkeitshypothese ausformuliert werden.

Die Konstruktion einer Vierfeldertafel unter Unabhängigkeit basiert auf den Randsummen und dem Multiplikationssatz für unabhängige Ereignisse (vgl. S.11). Die Randsummen liefern die relative Häufigkeit (und damit die [Laplace-] Wahrscheinlichkeit) für das Eintreten eines isoliert beobachteten Ereignisses in der Stichprobe (zum Beispiel 'Übergewicht'). Die Wahrscheinlichkeit einer Kombination von Ausprägungen (zum Beispiel 'ungesund UND Übergewicht') berechnet sich nach dem Multiplikationssatz für unabhängige Ereignisse ($P[A \text{ UND } B] = P[A] \cdot P[B]$). Im Beispiel:

Einfache Wahrscheinlichkeit in der Stichprobe:

$$P(X = 0) = 454/1000 = 0.454$$

$$P(X = 1) = 546/1000 = 0.546$$

$$P(Y = 0) = 490/1000 = 0.490$$

$$P(Y = 1) = 510/1000 = 0.510$$

Kombinierte Wahrscheinlichkeit bei Unabhängigkeit:

$$P(\{0;0\}) = 0.454 \cdot 0.490 = 0.22246$$

$$P(\{1;0\}) = 0.546 \cdot 0.490 = 0.26754$$

$$P(\{0;1\}) = 0.454 \cdot 0.510 = 0.23154$$

$$P(\{1;1\}) = 0.546 \cdot 0.510 = 0.27846$$

Aus der Wahrscheinlichkeit unter Unabhängigkeit ergibt sich eine Vierfeldertafel, indem die Wahrscheinlichkeit für ein bestimmtes Feld (eine bestimmte Kombination) mit dem Stichprobenumfang n (im Beispiel: $n = 1000$) multipliziert wird. Da die vier Kombinationen alle möglichen Ausprägungen darstellen, summieren sich die Einzelwahrscheinlichkeiten gerade zu 1. So ergibt sich eine Vierfeldertafel unter Unabhängigkeit, wie sie in der nebenstehenden Graphik dargestellt ist. Die Randsummen ändern sich unter Unabhängigkeit nicht, genauso wenig wie die Gesamtsumme (= Stichprobenumfang!), weil eine Veränderung der Randsumme einer Veränderung der Einzelwahrscheinlichkeiten implizieren würde. Die Hypothese der Unabhängigkeit bezieht sich jedoch auf gleiche Parameter, sprich gleiche Wahrscheinlichkeit.

| | | | | |
|---|----------|-----|-----|----------|
| | | 1 | 0 | Σ |
| X | 1 | 278 | 268 | 546 |
| | 0 | 232 | 222 | 454 |
| | Σ | 510 | 490 | 1000 |
| | | 1 | 0 | Σ |
| Y | | | | |

Vierfeldertafel bei Unabhängigkeit

Aufgrund der Musterformel für den χ^2 -Test (vgl. S. 34) wird nun die Prüfgrösse bestimmt:

H_0 : X, Y unabhängig

Prüfgrösse:

$$\chi_1^2 = \frac{(\text{Realität} - \text{Hypothese})^2}{\text{Hypothese}}$$

$$\chi_1^2 = \frac{(312 - 278)^2}{278} + \frac{(234 - 268)^2}{268} + \frac{(198 - 232)^2}{232} + \frac{(490 - 222)^2}{222}$$

$$\chi_1^2 = 4.158 + 4.313 + 4.983 + 323.532 = 333.986$$

Testentscheidung: $333.986 \gg \chi_{1;(1-\alpha)}^2 = 3.84$

Das Ergebnis ist also (hoch-) signifikant. Unglücklicherweise lässt der χ^2 -Test keine Rückschlüsse auf die Richtung des Zusammenhanges zu, weil das χ^2 unabhängig von der Anordnung der Tafel ist. Dies lässt sich auf zweierlei Arten einsehen: zum einen ist die Nullhypothese H_0 : 'X und Y sind unabhängig' ungerichtet (die Gegenhypothese H_1 lautet lediglich 'X und Y sind abhängig', ohne Richtungsangabe), zum anderen stellt die Prüfgrösse χ_1^2 lediglich die Addition von 4 Summanden dar, wobei bei der Addition die Reihenfolge beliebig ist (womit die Anordnung in der Tafel irrelevant wird).

Odds, Odds Ratio, Log-Odds und Log-Odds Test

Die Odds geben für eine festgelegte X-Ausprägung das Verhältnis an, die Ausprägung Y = 1 statt Y = 0 zu erhalten. Dies äussert sich in einem Verhältnis von (bedingten) Wahrscheinlichkeiten.

$$Odds_{X=1} = \frac{P([Y = 1] | [X = 1])}{P([Y = 0] | [X = 1])} \quad Odds_{X=0} = \frac{P([Y = 1] | [X = 0])}{P([Y = 0] | [X = 0])}$$

Äquivalent lassen sich auch absolute Häufigkeiten verwenden, weil sich die Randsumme (auf der die Wahrscheinlichkeit wie in Beispiel 13 gezeigt beruht) aus dem Bruch herauskürzt zur Illustration sind hier nochmals die Zahlen aus Beispiel 13 angegeben:

$$Odds_{X=1} = \frac{312}{234} = \frac{Y=1}{Y=0} \forall (X=1) \quad Odds_{X=0} = \frac{198}{256} = \frac{Y=1}{Y=0} \forall (X=0)$$

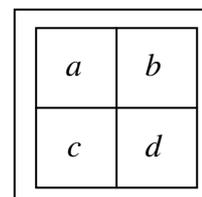
Die Odds-Ratio (kurz OR) ist definiert als

$$OR = \frac{Odds_{X=1}}{Odds_{X=0}}$$

Falls der Ausgang des Experiments (ob Y = 1 oder Y = 2) unabhängig von X ist, so sind die Odds in beiden Fällen gleich. Damit wird der OR = 1. Ergibt eine Kontingenztafel also einen OR = 1, so ist dies gleichbedeutend mit Unabhängigkeit.

Die Odds Ratio lässt sich auch unmittelbar aus der Vierfeldertafel berechnen. Die Formel dazu lautet:

$$OR = \frac{\text{Hauptdiagonal}}{\text{Nebendiagonale}} \\ OR = \frac{a \cdot d}{b \cdot c}$$



Die Odds Ratio ist immer positiv. Sie liefert ausserdem eine Aussage über die Richtung des Zusammenhangs zwischen Y und X:

- OR > 1 X wächst → Y wächst
- OR < 1 X wächst → Y fällt

Um eine Aussage über die Stärke des Zusammenhanges machen zu können, und um die Signifikanz des Zusammenhanges zu überprüfen, müsste die Verteilung der OR bekannt sein. Dies ist jedoch – vor allem aus Symmetriegründen – nicht der Fall. Die Lösung zu diesem Problem bietet der sogenannte Log-Odds bzw. der Log-Odds Test. Der Log-Odds ist nicht anderes als der natürliche Logarithmus aus der Odds Ratio. Die Verteilung des Log-Odds ist annähernd bekannt:

$$\ln[OR] \sim N(0, \sigma_*^2)$$

Die Varianz dieser Verteilung ist unbekannt. Der Varianzschätzer lautet:

$$\sigma_*^2 \wedge = \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)$$

Es ergibt sich das folgende Testverfahren, der sogenannte Log-Odds Test.

H_0 : 'X und Y unabhängig'

Prüfgrösse:
$$z = \frac{\ln(OR)}{\sigma_* \wedge} = \frac{\ln(OR)}{\sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)}} \sim N(0;1)$$

Testentscheidung: H_0 verwerfen, wenn $z > z_{1-(\alpha/2)}$.

Die Buchstaben a, b, c, d beziehen sich auf die Vierfeldertafel S. 36 und bezeichnen ihre Felder.

Äquivalent zu diesem Test lässt sich auch das Konfidenzintervall der Log-Odds bilden:

$$\ln[OR] - (z_{1-(\alpha/2)} \cdot \sigma_* \wedge) < \ln[OR] < \ln[OR] + (z_{1-(\alpha/2)} \cdot \sigma_* \wedge)$$

H_0 wird hier verworfen, wenn die 0 (die Unabhängigkeitsbedingung für den $\ln[OR]$) nicht im Intervall liegt.

Aus dem Konfidenzintervall für den $\ln[OR]$ lässt sich durch Rücktransformation auch ein Konfidenzintervall für die Odds Ratio OR gewinnen:

$$\exp[\ln(OR) - (z_{1-(\alpha/2)} \cdot \sigma_* \wedge)] < OR < \exp[\ln(OR) + (z_{1-(\alpha/2)} \cdot \sigma_* \wedge)]$$

In diesem Fall wird H_0 verworfen, wenn 1 (die Unabhängigkeitsbedingung für die OR) nicht im Intervall liegt.

Wie gesagt sind alle drei Testverfahren gleichwertig.

Unter Zuhilfenahme des Log-Odds wird auch die Terminologie von positivem und negativem Zusammenhang zwischen X und Y klar. Besteht ein negativer Zusammenhang, d.h. fällt Y wenn X wächst, so ist der Log-Odds $\ln[OR] < 0$. Besteht umgekehrt ein positiver Zusammenhang zwischen X und Y, d.h. wächst Y zusammen mit X, so ist der Log-Odds $\ln[OR] > 0$.

Mehrfeldertafeln, G^2 -Statistik

Die Vierfeldertafel werde nun erweitert. Das Y und das X seien nun nicht mehr binär, sondern ordinal in z.B. (X: 5) und (Y: 3) Ausprägungen. Es ergibt sich eine Kontingenztabelle der folgenden Form:

| | | X ordinal | | | | |
|-----------|---|-----------|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| Y ordinal | 0 | | | | | |
| | 1 | | | | | |
| | 2 | | | | | |

Die G^2 -Statistik ist dem χ^2 verwandt. Um den G^2 -Wert der gesamten Kontingenztabelle [G_{total}^2] zu bestimmen, wird die Tabelle in Vierfeldertabellen aufgegliedert und der χ^2 -Wert jeder dieser Untertabellen bestimmt. Die Aufgliederung folgt dabei dem unten dargestellten Schema:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|-----|--|--|---|--|--|----------------------------------------------------------------------------------------------------------------------------------------------------------|--|-----|---|-----|--|--|---|--|--|------------------------------------------------------------------------------------------------------------------------------------------------------------|--|-------|---|-----|--|--|---|--|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------|--|---------|---|-----|--|--|---|--|--|
| G_1 <table border="1"> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td></tr> </table> | | 0 | 1 | 0 | | | 1 | | | G_2 <table border="1"> <tr><td></td><td>0;1</td><td>2</td></tr> <tr><td>0</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td></tr> </table> | | 0;1 | 2 | 0 | | | 1 | | | G_3 <table border="1"> <tr><td></td><td>0;1;2</td><td>3</td></tr> <tr><td>0</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td></tr> </table> | | 0;1;2 | 3 | 0 | | | 1 | | | G_4 <table border="1"> <tr><td></td><td>0;1;2;3</td><td>4</td></tr> <tr><td>0</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td></tr> </table> | | 0;1;2;3 | 4 | 0 | | | 1 | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0;1 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0;1;2 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0;1;2;3 | 4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| G_5 <table border="1"> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0;1</td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td></tr> </table> | | 0 | 1 | 0;1 | | | 2 | | | G_6 <table border="1"> <tr><td></td><td>0;1</td><td>2</td></tr> <tr><td>0;1</td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td></tr> </table> | | 0;1 | 2 | 0;1 | | | 2 | | | G_7 <table border="1"> <tr><td></td><td>0;1;2</td><td>3</td></tr> <tr><td>0;1</td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td></tr> </table> | | 0;1;2 | 3 | 0;1 | | | 2 | | | G_8 <table border="1"> <tr><td></td><td>0;1;2;3</td><td>4</td></tr> <tr><td>0;1</td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td></tr> </table> | | 0;1;2;3 | 4 | 0;1 | | | 2 | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0;1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0;1 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0;1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0;1;2 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0;1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0;1;2;3 | 4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0;1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Für jede Einzeltabelle wird der G_i^2 -Wert bestimmt, der gerade dem χ^2 -Wert der i -ten Untertabelle entspricht. Um das G_{total}^2 der gesamten Kontingenztabelle zu bestimmen, werden die einzelnen G_i^2 -Werte summiert:

$$G_{\text{total}}^2 = \sum G_i^2$$

Für eine Kontingenztabelle von n Ausprägungen in X und m Ausprägungen in Y ergeben sich $v = (n - 1) \cdot (m - 1)$ Untertabellen. Entsprechend ist

$$G_{\text{total}}^2 \sim \chi_v^2\text{-verteilt}$$

mit $v = (n - 1) \cdot (m - 1)$ Freiheitsgraden.

Der Test erfolgt analog zur Vierfeldertabelle. Die Testentscheidung wird anhand des χ_v^2 -Wertes im Vergleich zu $\chi_{v, (1-\alpha)}^2$ gefällt.

Anhang

Zur Berechnung von Beispiel 12 (S. 29) wurde die folgende Excell-Tabelle verwendet. x bezeichnet die Klassengrenzen, z die Standardisierung von x . Die Wahrscheinlichkeit von z ($= \text{phi}(z)$) ist der standardisierten Normalverteilung entnommen. Der Mittelwertschätzer $\hat{\mu}$ für die Verteilungshypothese (Normalverteilung) stammt aus der Stichprobe und stellt ein mit der absoluten Häufigkeit gewichtetes arithmetisches Mittel der Klassenmitten dar ($\hat{\mu} = \frac{\sum \text{Klassenmitte}(i) \cdot n(i)}{n}$). Der Varianzschätzer (hier als $\hat{\sigma}$ bezeichnet) wurde nach

der Formel für die Stichprobenvarianz berechnet: $\hat{\sigma}^2 = \frac{1}{n-1} \cdot \sum (\text{Klassenmitte}(i) - \hat{\mu})^2$.

Die absolute Häufigkeit der hypothetischen Verteilung in der Klasse i [$= N_i$] ist der der Wahrscheinlichkeit entsprechende Anteil am Stichprobenumfang: $N_i = 1000 \cdot p(k)$. Aufgrund von Rundungsfehlern ist die Summe über $N(i)$ nur 992 statt 1000. Dies führt zwar zu einer leichten Verzerrung des Tests, kann aber das Testresultat aufgrund seiner Eindeutigkeit in diesem Fall nicht beeinflussen.

| x | z | phi(z) | k | p(k) | N(i) |
|-----|-------------|---------|----|---------|------|
| 0 | -2.81402537 | 0.00248 | 1 | 0.00912 | 9 |
| 10 | -2.26941043 | 0.0116 | 2 | 0.03110 | 31 |
| 20 | -1.7247955 | 0.0427 | 3 | 0.07630 | 76 |
| 30 | -1.18018056 | 0.119 | 4 | 0.14530 | 145 |
| 40 | -0.63556563 | 0.2643 | 5 | 0.19980 | 200 |
| 50 | -0.09095069 | 0.4641 | 6 | 0.20950 | 210 |
| 60 | 0.45366424 | 0.6736 | 7 | 0.16530 | 165 |
| 70 | 0.99827917 | 0.8389 | 8 | 0.09930 | 99 |
| 80 | 1.54289411 | 0.9382 | 9 | 0.04304 | 43 |
| 90 | 2.08750904 | 0.98124 | 10 | 0.01449 | 14 |
| 100 | 2.63212398 | 0.99573 | | Summe = | 992 |

$\hat{\mu} = 51.67$
 $\hat{\sigma} = 18.3615971$

$p(k[i]) = \text{phi}(z[i+1]) - \text{phi}(z[i])$

| Klassenmitte | n(i) | x - $\hat{\mu}$ | (x - $\hat{\mu}$) ² | gewichtet mit n(i) |
|--------------|------|-----------------|---------------------------------|-------------------------------|
| 5 | 5 | -46.67 | 2178.0889 | 10890.4445 |
| 15 | 42 | -36.67 | 1344.6889 | 56476.9338 |
| 25 | 98 | -26.67 | 711.2889 | 69706.3122 |
| 35 | 122 | -16.67 | 277.8889 | 33902.4458 |
| 45 | 142 | -6.67 | 44.4889 | 6317.4238 |
| 55 | 285 | 3.33 | 11.0889 | 3160.3365 |
| 65 | 143 | 13.33 | 177.6889 | 25409.5127 |
| 75 | 102 | 23.33 | 544.2889 | 55517.4678 |
| 85 | 51 | 33.33 | 1110.8889 | 56655.3339 |
| 95 | 10 | 43.33 | 1877.4889 | 18774.889 |
| | | Summe: | | 336811.1 |
| | | | | 337.148248 = $\hat{\sigma}^2$ |
| | | | | 18.3615971 = $\hat{\sigma}$ |