

Peter von der Lippe

Induktive Statistik

Formeln, Aufgaben, Klausurtraining

Ursprünglich verlegt bei Oldenbourg, hier
in überarbeiteter Form als download zur
Verfügung gestellt



Oldenbourg

Teil I

Formelsammlung mit Tabellenanhang

von der Lippe: Induktive Statistik Inhalt von Teil I (Formelteil)

Kap.1:	Einführung, Stichprobenraum	3
Kap.2:	Kombinatorik	4
Kap.3:	Ereignisalgebra, Wahrscheinlichkeit	14
Kap.4:	Zufallsvariablen, Wahrscheinlichkeitsverteilung	25
Kap.5:	Spezielle diskrete Wahrscheinlichkeitsverteilungen	36
Kap.6:	Spezielle stetige Verteilungen	51
Kap.7:	Grenzwertsätze, Stichprobenverteilung	63
Kap.8:	Schätztheorie	73
Kap.9:	Testtheorie	87
Kap.10:	Stichprobentheorie	100
	Tabellenanhang	109

Detailgliederung von Teil II (Aufgabenteil)

Kap.1:	Einführung, Stichprobenraum
Kap.2:	Kombinatorik
Kap.3:	Ereignisalgebra, Wahrscheinlichkeit
	3.1. Mengenoperationen mit Ereignissen
	3.2. Wahrscheinlichkeitsbegriff
	3.3. Additionssätze
	3.4. Multiplikationssätze, bedingte Wahrscheinlichkeiten, Unabhängigkeit
	3.5. Totale Wahrscheinlichkeit, Theorem von Bayes
Kap.4:	Zufallsvariablen, Wahrscheinlichkeitsverteilung
	4.1. Eindimensionale Zufallsvariable
	4.2. Zweidimensionale Zufallsvariable
	4.3. Linearkombination und -transformation
	4.4. Erzeugende Funktionen
Kap.5:	Spezielle diskrete Wahrscheinlichkeitsverteilungen
	5.1. Zweipunktverteilung
	5.2. Geometrische Verteilung, Binomialverteilung
	5.3. Hypergeometrische Verteilung
	5.4. Poissonverteilung
Kap.6:	Spezielle stetige Verteilungen
	6.1. lineare Verteilungen, Gleichverteilung (stetig)
	6.2. Normalverteilung
Kap.7:	Grenzwertsätze, Stichprobenverteilung
	7.1. Tschebyscheffsche Ungleichung, Grenzwertsätze, stochastische Konvergenz
	7.2. Stichprobenverteilungen
Kap.8:	Schätztheorie
	8.1. Maximum-Likelihood-Methode
	8.2. Punktschätzung
	8.3. Intervallschätzung (Mittel- und Anteilswert)
	8.4. Konfidenzintervallschätzung für die Differenz von zwei Mittel-bzw. Anteilswerten
Kap.9:	Testtheorie
	9.1. Test für Mittel- und Anteilswerte (Ein-Stichproben-Fall)
	9.2. Signifikanztests für Mittel- und Anteilswertdifferenzen (zwei unabhängige Stichproben)
Kap.10:	Stichprobentheorie
	10.1. Notwendiger Stichprobenumfang 10.2.Hochrechnung 10.3.Geschichtete Stichproben

Kapitel 1: Einführung

Wahrscheinlichkeitsaussagen beziehen sich auf Zufallsexperimente (ZE), und zwar (gerade wegen der Zufälligkeit) nicht auf den Ausgang eines einzelnen ZE, sondern auf die (zumindest gedanklich) unendliche Folge von Wiederholungen (Realisationen) des ZE unter einem

- unveränderlichen
- exakt beschriebenen

Bedingungskomplex.

Def. 1.1: Zufallsexperiment (ZE)

Ein Zufallsexperiment liegt vor, wenn

1. es wohldefinierte Ereignisse als Ergebnis des ZE gibt
2. das ZE unter denselben Bedingungen unabhängig beliebig oft wiederholbar ist
3. das Ereignis (der Versuchsausgang) im Einzelfall
 - a) nicht voraussagbar ist
 - b) nicht willkürlich (systematisch) zu beeinflussen ist
4. es wohl aber bei einer Vielzahl von Wiederholungen des ZE gewisse Regelmäßigkeiten gibt.

Def. 1.2: Stichprobenraum

- a) Ein Stichprobenraum Ω ist die Menge aller möglichen, sich gegenseitig ausschließender Elementarereignisse $\omega_1, \omega_2, \dots, \omega_n$

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\} \quad \text{bzw.} \quad \{\omega, \omega \in \Omega\}$$

Im folgenden wird von endlichen Stichprobenräumen mit gleichwahrscheinlichen Elementarereignissen (Laplace-Annahme) ausgegangen.

- b) Es sind Elementarereignisse und zusammengesetzte Ereignisse (Def. 3.1) zu unterscheiden.

Kapitel 2: Kombinatorik

1. Grundaufgaben der Kombinatorik	5
2. Binomialkoeffizienten und Multinomialkoeffizienten	7
3. Ergänzungen und Vertiefungen zum Auswahlproblem: Inklusion und Exklusion	11
4. Die Gamma- und die Beta-Funktion	13

Gegenstand der Kombinatorik sind enumerative Probleme, Fragen im Zusammenhang mit endlichen Mengen. Es geht um die Anzahl der Arten, wie eine wohldefinierte Operation (z.B. Auswahl oder Anordnung von Objekten) ausgeführt werden kann.

1. Grundaufgaben der Kombinatorik

Fragestellungen

1. Anordnung von n Elementen (**Permutation**) oder Auswahl von $i \leq n$ Elementen (die Elemente seien a, b, c, \dots)
2. Es ist zu unterscheiden:
 - a) mit Berücksichtigung der Anordnung (**Variation**): $\{a,b\}$ und $\{b,a\}$ sind verschieden
 - b) ohne Berücksichtigung der Anordnung (**Kombination**): $\{a,b\}$ und $\{b,a\}$ sind gleich
3. Wiederholungen. Dabei bedeutet:
 - a) **ohne Wiederholung (oW)**: Elemente a, b und c treten nur einmal auf
 - b) **mit Wiederholung (mW)**: Es kann auch $\{a,a\}$, $\{a,b,b\}$, $\{b,b,b\}$, ... auftreten

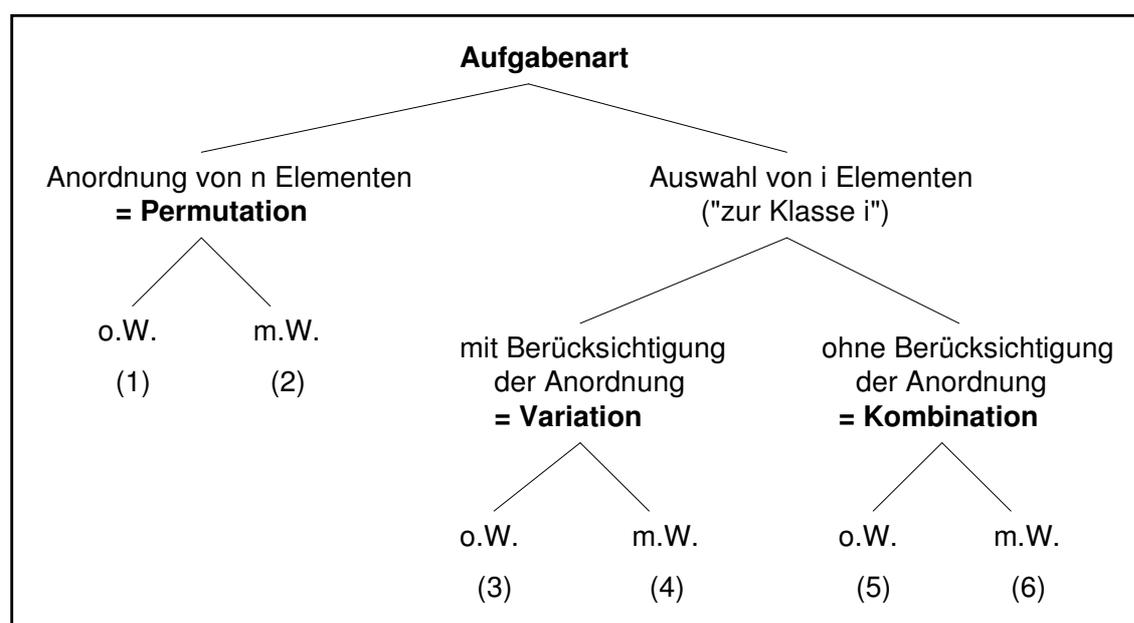
Diese Kriterien werden kombiniert zu sechs Grundaufgaben (vgl. Übers. 2.1).

Anwendungen in der Stichprobentheorie:

- a) Die Anzahl der Stichproben beim Ziehen **ohne Zurücklegen** ist K . Dabei sind die K Stichproben gleich wahrscheinlich.
- b) Die Anzahl verschiedener Stichproben **mit Zurücklegen** ist K_w , davon sind aber nicht alle gleich wahrscheinlich. Durch das Zurücklegen ist die Urne praktisch unendlich, so dass auch $i > n$ (i [sonst n] Umfang der Stichprobe; n [sonst N] Umfang der Grundgesamtheit).

Übersicht 2.1: Die sechs Grundaufgaben der Kombinatorik

a) Fallunterscheidungen



b) Formeln für die Anzahl

(2.1) Anzahl der Permutationen ohne Wiederholung von n Elementen	$P = n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$
(2.2) Anzahl der Permutationen mit Wiederholung von n Elementen, wobei das k -te Element n_k mal auftritt	$P_W = \frac{n!}{\prod_{k=1}^m n_k!}$ mit $\sum_{k=1}^m n_k = n$
(2.3) Anzahl der Variationen ohne Wiederholung (V) von n Elementen zur i -ten Klasse	$V = \frac{n!}{(n-i)!} = i! \binom{n}{i}$
(2.4) Anzahl der Variationen mit Wiederholung (V_W) zur i -ten Klasse	$V_W = n^i$
(2.5) Anzahl der Kombinationen ohne Wiederholung (K)	$K = \binom{n}{i} = \frac{n!}{i! (n-i)!}$
(2.6) Anzahl der Kombinationen mit Wiederholung (K_W)	$K_W = \binom{n+i-1}{i}$

Bemerkungen:

zu (1): Stirlingsche Formel:

Für großes n gilt: $n! \approx n^n \frac{1}{e^n} \sqrt{2 \pi n} = \left(\frac{n}{e}\right)^n \sqrt{2 \pi n} = \sqrt{2 \pi} n^{n+1/2} e^{-n}$.

zu (3), (5) u. (6): $\binom{n}{i}$ ist der Binomialkoeffizient ("n über i") als Spezialfall des Multinomialkoeffizienten (Formel für P_W). Ist $m = 2$, $n_1 = i$ und $n_2 = n - i$, so ist

$$\frac{n!}{\prod n_k!} = \frac{n!}{n_1! n_2!} = \frac{n!}{i! (n-i)!} = \binom{n}{i}.$$

Die Zweiteilung der Elemente kann z.B. bedeuten: $n_1 = i$ Elemente gelangen in die Auswahl $n_2 = n - i$ gelangen nicht in die Auswahl. Weitere Bemerkungen zum Binomialkoeffizienten vgl. Abschnitt 2.

zu (4): Hierbei kann jedes Element bis zu i -mal wiederholt werden (i kann auch größer als n sein).

Zusammenhänge der Formeln untereinander:

$P_W \rightarrow P$ wenn für alle $k = 1, \dots, m$ gilt $n_k = 1$, folgt $P_W = P$

$V \rightarrow P$ wenn $i = n$ gilt (also keine Auswahl), folgt $V = P$ (Permutation ohne Wiederholung als Spezialfall der Variation ohne Wiederholung)

$P_W \rightarrow K$ wenn $n_1 = i$ und $n_2 = n - i$, folgt $P_W = K$ (siehe oben; vgl. auch Gl. 2.15)

$V \rightarrow K$ da jedes Element auf $i!$ Arten permutiert werden kann, gilt (2.3a): $V = i! K$

$K \rightarrow K_W$ Herleitung der Kombinationen ohne Wiederholung aus dem Additionstheorem für Binomialkoeffizienten (Satz 2.1). Im trivialen Fall $i=1$ (keine Wiederholungen möglich) ist

$$K_W = K = \binom{n}{1} = n$$

$P_W \rightarrow V_W$ Variationen mit Wiederholung als Summe aller möglichen Permutationen mit Wiederholung (Satz 2.2)

2. Binomialkoeffizienten und Multinomialkoeffizienten

Def. 2.1: Binomialkoeffizient

Der Ausdruck

$$(2.5) \quad \binom{n}{i} = \frac{n!}{(n-i)!i!}$$

mit $n, i \in \mathbb{N}$, $0 \leq i \leq n$ heißt **Binomialkoeffizient**.

Def. 2.2: Multinomialkoeffizient

Der Ausdruck

$$(2.6) \quad \binom{n}{n_1 \ n_2 \ \dots \ n_m} = \frac{n!}{n_1! n_2! \dots n_m!} = \frac{n!}{\prod_{k=1}^m n_k!}, \quad \sum_{k=1}^m n_k = n$$

heißt **Multinomialkoeffizient** (oder auch Polynomkoeffizient genannt).

Eigenschaften des Binomialkoeffizienten und des Multinomialkoeffizienten

1. Name und Folgerung aus der Definition

a) Die Entwicklung des Binoms $(a+b)^n$ führt zu folgender Summe:

$$\begin{aligned} & b^n + nab^{n-1} + \dots + \binom{n}{i} a^i b^{n-i} + \dots + na^{n-1}b + a^n \\ &= \binom{n}{0} a^0 b^n + \binom{n}{1} a^1 b^{n-1} + \dots + \binom{n}{i} a^i b^{n-i} + \dots + \binom{n}{n-1} a^{n-1} b + \binom{n}{n} a^n b^0 \end{aligned}$$

Setzt man $a=b=1$, so folgt leicht der bekannte Zusammenhang über die Summe von Binomialkoeffizienten von Gleichung 2.9.

Entsprechend erscheint der Multinomialkoeffizient in der Expansion eines Multinoms, etwa ($m = 3$)

$$(a + b + c)^n = \sum_{(n_1, n_2, n_3)} \binom{n}{n_1 \quad n_2 \quad n_3} a^{n_1} b^{n_2} c^{n_3}, \text{ mit } n = n_1 + n_2 + n_3.$$

b) Nach Definition gilt $\binom{n}{0} = \binom{n}{n} = 1$ und $\binom{n}{1} = n$.

c) Symmetrie $\binom{n}{i} = \binom{n}{n-i}$.

2. Binomialkoeffizient als Summe und Produktsumme von Binomialkoeffizienten

$$(2.7a) \quad \binom{n}{i} + \binom{n}{i+1} = \binom{n+1}{i+1} \text{ (Pascalsches Dreieck)}$$

$$(2.7b) \quad \sum_{i=0}^{n-m} \binom{m+i}{m} = \binom{n+1}{m+1}$$

Folgerung:

$$(2.7c) \quad \binom{n}{i} = \sum_{k=0}^{n-i} \binom{i-1+k}{i-1}$$

$$(2.7d) \quad \sum_{i=0}^n \binom{m}{k-i} \binom{n}{i} = \binom{m+n}{k}$$

Satz 2.1 Additionstheorem für Binomialkoeffizienten

Aus Gleichung 2.7d folgt*:

$$(2.8) \quad \binom{n+i-1}{i} = \sum_{k=0}^{i-1} \binom{n}{i-k} \binom{i-1}{k}$$

Aus dieser Formel folgt auch der Zusammenhang zwischen Kombinationen mit und ohne Wiederholung:

$$\binom{n+i-1}{i} = \binom{n}{i} + \sum_{k=1}^{i-1} \binom{n}{i-k} \binom{i-1}{k}.$$

* Man erhält Gleichung 2.8 aus 2.7d, wenn man den Symbolen k , n , m und i in Gleichung 2.7c die Symbole i , $i-1$, n und k zuordnet. Aus Gleichung 2.8 folgt übrigens auch Gleichung 2.9a.

Der zweite Summand gibt an, um wie viel sich die Anzahl der Kombinationen erhöht, durch die Wiederholung von $k = 1, 2, \dots$ der $i - k$ ausgewählten Elemente, um zur Zahl der Kombinationen mit Wiederholung zu gelangen.

Folgerung aus Gleichung 2.8:

$$(2.8a) \quad \binom{n}{i} = \sum_{k=0}^{i-1} \binom{n-(i-1)}{i-k} \binom{i-1}{k}.$$

Dieser Zusammenhang erklärt die Reproduktivität der Binomialverteilung. Aus einer Urne von n Kugeln i auszuwählen läuft auf das gleiche hinaus, wie aus zwei Urnen mit $n - (i - 1)$ und $i - 1$ Kugeln so viele Kugeln herauszunehmen, dass es zusammen i Kugeln sind.

3. Summe von Binomialkoeffizienten

a) Variables i , also $i = 0, 1, \dots, n$,

$$(2.9) \quad \sum_i \binom{n}{i} = 2^n \quad \text{und} \quad \sum_i (-1)^i \binom{n}{i} = 0 \quad (2.9a) \quad \sum_i \binom{n}{i}^2 = \binom{2n}{n}$$

$$(2.10) \quad \sum_i \binom{n}{i} i = 2^{n-1} n \quad (2.10a) \quad \sum_i \binom{n}{i} i^2 = 2^{n-2} n(n+1)$$

$$(2.11) \quad \sum_{i=0}^m \binom{n}{i} \binom{n-i}{m-i} = 2^m \binom{n}{m}, \quad m \leq n$$

b) Variables n (k läuft bis n), konstantes i

Summe der natürlichen Zahlen

$$(2.11a) \quad \sum_{k=1}^n \binom{k}{1} = \binom{n+1}{2} = \frac{n(n+1)}{2}. \quad \text{Das folgt auch aus Gleichung 2.7b mit } m = 1.$$

$$(2.12a) \quad \sum_{k=2}^n \binom{k}{2} = \binom{n+1}{3} \quad (2.13) \quad \sum_{k=3}^n \binom{k}{3} = \binom{n+1}{4} \text{ usw..}$$

Die allgemeinen Zusammenhänge beschreiben die folgenden Formeln:

$$(2.13a) \quad \sum_{i=0}^r \binom{n+i}{n} = \binom{n+r+1}{n+1} \quad \text{und} \quad (2.13b) \quad \sum_{i=0}^r \binom{n+i}{i} = \binom{n+r+1}{r}$$

4. Multinomialkoeffizient als Produkt von Binomialkoeffizienten

$$(2.14) \quad \frac{n!}{n_1! n_2! \dots n_k!} = \binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots \binom{n-n_1-\dots-n_{k-1}}{n_k}$$

Spezialfall der Permutationen ohne Wiederholungen $n_1 = n_2 = \dots = n_k = 1$:

$$(2.14a) \quad \binom{n}{1} \binom{n-1}{1} \binom{n-2}{1} \dots \binom{1}{1} = n!$$

Spezialfall Kombinationen:
$$\binom{n}{i} = \binom{n}{n-i} \binom{n-i}{n-i}$$

Permutationen mit Wiederholungen kann man als wiederholte Kombinationen auffassen: Aus n Elementen werden n_1 ausgewählt, aus den verbleibenden Elementen wieder n_2 usw.

5. Rekursionsformeln für Binomialkoeffizienten

$$(2.15) \quad \binom{n}{i} = \frac{n-i+1}{i} \binom{n}{i-1} \quad (2.15a) \quad \binom{n}{i} = \frac{n}{n-i} \binom{n-1}{i}$$

$$(2.15b) \quad \binom{n}{i} = \frac{n}{i} \binom{n-1}{i-1} = \frac{n}{i} \frac{n-1}{i-1} \binom{n-2}{i-2} \text{ usw..}$$

Folgerung: Aus Gleichung 2.15 folgt, dass bei gegebenem ungeradzahligem n die Binomialkoeffizienten von $i = 0$ bis $i = (n-1)/2$ ansteigen und von $i = (n+1)/2$ an bis $i = n$ abfallen (bei geradzahligem n ist das Maximum $i = n/2$).

Aus Gleichung 2.15a und 2.15b folgt leicht der als Pascalsches Dreieck bekannte Zusammenhang der Gleichung 2.7a bzw. (gleichbedeutend):

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

6. Binomialkoeffizienten mit negativen Elementen

Nach Definition gilt:

$$(2.16) \quad \binom{-n}{i} = (-1)^i \binom{n+i-1}{i} \quad \text{und} \quad (2.16a) \quad \binom{-n}{-i} = (-1)^{n+i} \binom{i-1}{n-1}.$$

7. Summe von Multinomialkoeffizienten

Man kann Variationen mit Wiederholungen als Summe von Permutationen mit Wiederholungen auffassen wegen:

Satz 2.2: Additionstheorem für Multinomialkoeffizienten

$$(2.17) \quad n^i = \sum_{a_1, a_2, \dots} \binom{i}{a_1 \ a_2 \ \dots \ a_n},$$

wobei summiert wird über alle n -Tupel a_1, a_2, \dots, a_n mit $\sum_{k=1}^n a_k = i$. Mit 2-Tupeln, also $a_1 = j$, $a_2 = n - j$, ergibt sich $\sum_{j=0}^n \binom{n}{j} = 2^n$, also Gl. 2.9 als Spezialfall.

8. Rekursive Beziehungen zwischen Kombinationen mit Wiederholung

Satz 2.3: Additionstheorem für Kombinationen mit Wiederholung

Verabredet man $K_w(n, i)$ für die Anzahl der Kombinationen mit Wiederholung zur Klasse i , so gilt:

$$(2.18) \quad K_w(n, i) = K_w(n, i-1) + K_w(n-1, i)$$

Diese Rekursionsformel ist auch Ausgangspunkt für den Beweis von Gl. 2.6.

Ersetzt man in Gleichung 2.18 den Ausdruck $K_w(n-1, i)$ durch $K_w(n-1, i-1) + K_w(n-2, i)$, hierbei wieder $K_w(n-2, i)$ usw., so erhält man:

$$(2.19) \quad K_w(n, i) = \sum_{m=1}^n K_w(m, i-1),$$

was sich übrigens auch aus Gleichung 2.13a ergibt.

3. Ergänzungen und Vertiefungen zum Auswahlproblem: Inklusion und Exklusion

a) Zum Permutationsbegriff

Def. 2.3: Zirkuläre Permutationen

Die Anzahl P_z der zirkulären Permutationen von n Elementen ist die Anzahl der Möglichkeiten, n Elemente im Kreis anzuordnen. Sie beträgt:

$$(2.20) \quad P_z(n) = (n-1)!$$

Da wegen der kreisförmigen Anordnung der erste und der n -te Platz identisch sind, werden faktisch nur $n - 1$ Elemente permutiert.

Def. 2.4: Fixpunktfreie Permutationen

Geht man von einer Sitzordnung von $n \geq 2$ Stühlen und n Personen aus, so kann man nach der Anzahl $P_F(n)$ der "neuen" Sitzordnungen fragen, bei denen keine Person auf ihrem alten Platz bleibt. Sie beträgt:

$$(2.21) \quad P_F(n) = n! \left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - + \dots + \frac{(-1)^n}{n!} \right) = n! \sum_{j=0}^n \frac{(-1)^j}{j!}$$

wobei die Summe den Anfang der Potenzreihenentwicklung von e^{-1} darstellt. Man spricht auch vom Rencontre Problem und nennt die Zahlen $P_F(n)$ auch Rencontre Zahlen, für die die folgende Rekursionsformel gilt:

$$(2.21a) \quad P_F(n) = nP_F(n-1) + (-1)^n$$

b) Zum Auswahlproblem: Inklusion und Exklusion

Für ein Auswahlproblem aus n Elementen kann sich auch die Aufgabe stellen, dass die Auswahl so getroffen werden sollte, dass bei $p < n$ der n Elemente mit

- keinem der p Elemente (**Exklusion**)

- genau p Elementen (**Inklusion**)

kombiniert bzw. variiert wird.

Die hierzu relevanten Formeln sind in Übersicht 2.2 zusammengestellt. Die **Exklusion** von p vorgeschriebenen Elementen läuft darauf hinaus, einfach p Elemente weniger für eine Auswahl zur Verfügung zu stellen. Sie stellt also eine **Reduktion der Auswahlgesamtheit** dar. Inklusion bedeutet ebenfalls, diese Elemente von einer Auswahl auszuschließen und statt i nur noch $i - p$ Elemente frei zu kombinieren bzw. zu variieren. Man reserviert einfach p Plätze in der Auswahlgesamtheit und in der Auswahl selbst und fragt nach den Variations- und Kombinationsmöglichkeiten der übrigen Elemente. **Inklusion** stellt also eine **Reduktion der Auswahlgesamtheit und der Auswahl** dar. Die Differenz zwischen der allgemeinen Formel (etwa Gl. 2.5 bei Kombination ohne Wiederholungen) und der entsprechenden Formel für die Exklusion von p Elementen

$$\binom{n}{i} - \binom{n-p}{i} \quad \text{bzw. im Spezialfall } p=1: \quad \binom{n}{i} - \binom{n-1}{i} = \binom{n-1}{i-1} \quad \text{Gleichung 2.7a}$$

ist die Anzahl der Kombinationen **ohne** Wiederholungen mit mindestens einem bzw. genau einem (wenn $p=1$) von p vorgeschriebenen Elementen (Inklusion von mindestens einem Element = keine Exklusion von allen Elementen). Bei $p=1$ ist die allgemeine Formel für Kombinationen die Summe der Exklusions- und Inklusionsformel (Übersicht 2.2).

Übersicht 2.2

	Exklusion	Inklusion
Kombinationen ohne Wiederholung	(2.22) $\binom{n-p}{i}$	(2.23) $\binom{n-p}{i-p}$
Kombinationen mit Wiederholung	(2.24) $\binom{n+i-p-1}{i}$	(2.25) $\binom{n+i-p-1}{i-p}$
Variationen ohne Wiederholung	(2.22) $i! \binom{n-p}{i}$	(2.23) $i! \binom{n-p}{i-p}$
Variationen mit Wiederholung	(2.24) $(n-p)^i$	(2.25) $(n-p)^{(i-p)^*}$

*) Nur wenn $n > p$ und $i > p$, sonst keine allgemeine Formel möglich, da i nicht beschränkt ist.

4. Die Gamma- und die Beta-Funktion

Def. 2.5: Gammafunktion

Die Funktion

$$(2.30) \quad \Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad 0 < \alpha, \quad 0 \leq x \leq \infty$$

heißt Gamma-Funktion.

Folgerungen:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\sqrt{\pi}, \quad \Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right)$$

$$\Gamma(1) = 1, \quad \Gamma(2) = 1, \quad \Gamma(3) = 2\Gamma(2) = 2!$$

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)!$$

$$\Gamma(n+z) = \frac{(n+z-1)!}{(n-1)!} \Gamma(n) \quad (z \text{ und } n \text{ ganzzahlig})$$

Def. 2. 6: Betafunktion

Die Funktion

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx, \quad \alpha, \beta > 0, \quad 0 \leq x \leq 1$$

heißt Beta-Funktion.

Bemerkungen zu Def. 2.5 und 2.6:

1. Die Gamma- und die Beta-**Funktionen** sollten nicht verwechselt werden mit der Gamma- und Beta-**Verteilung**.
2. Man nennt die oben definierten Funktionen auch vollständige Gamma bzw. Beta-Funktion. Die entsprechenden unvollständigen Funktionen haben eine feste Integrationsgrenze z ($z < \infty$ oder $z \neq 1$).
3. Die beiden Funktionen treten in einigen Dichtefunktionen auf (χ^2 , t , F-Verteilung).

Kapitel 3: Ereignisse und ihre Wahrscheinlichkeit

1. Einführende Konzepte der Mengenlehre	14
2. Wahrscheinlichkeitsbegriff	20
3. Additionssätze	21
4. Multiplikationssätze, stochastische Unabhängigkeit	22
5. Totale Wahrscheinlichkeit, Bayessches Theorem	23

Das "Rechnen" mit Ereignissen, das Inhalt dieses Kapitels ist, wird oft auch als Ereignisalgebra bezeichnet. Der Begriff wird jedoch auch spezieller benutzt (Def. 3.7). Es ist formal äquivalent dem Rechnen mit Mengen. Mit Mengen können Mengensysteme gebildet und hierfür Mengenfunktionen definiert werden. Die Wahrscheinlichkeit ist eine solche Mengenfunktion.

1. Einführende Konzepte der Mengenlehre

1.1. Relationen zwischen und Operationen mit Ereignissen

Ein Ereignis kann als Menge dargestellt werden, so dass auf diese auch Verknüpfungsoperationen für Mengen angewandt werden können. Durch Operationen mit Elementarereignissen entstehen zusammengesetzte Ereignisse (vgl. Def. 3.1 und 3.2).

a) Operationen

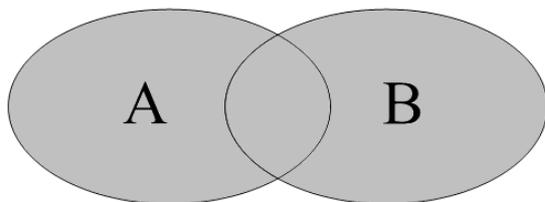
Der Stichprobenraum Ω bestehe aus den Mengen (Ereignissen) A , B und C , dargestellt im Euler-Venn-Diagramm. Das Ergebnis einer Operation wird durch Schattierung angegeben. Ω wird als Kasten dargestellt, die Mengen A , B , C sind Flächen in dem Kasten.

Vereinigung: $A \cup B$ oder $A + B$

(auch Summe genannt)

$$A \cup B := \{x \mid x \in A \vee x \in B\}^1$$

spricht: „A oder B“ (inklusive oder)

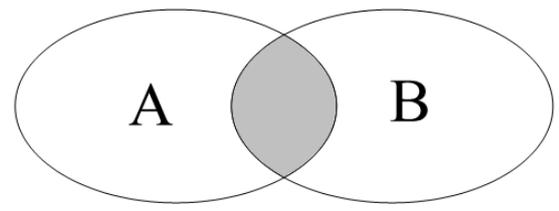


Durchschnitt: $A \cap B$ oder AB

(auch Produkt oder Konjunktion)

$$A \cap B := \{x \mid x \in A \wedge x \in B\}$$

spricht: „sowohl A als auch B“



¹ Dies ist eine Definition. Fast alles, was im folgenden dargestellt wird, sind Definitionen. Nur einige besonders hervorzuhebende Definitionen sind numeriert worden.

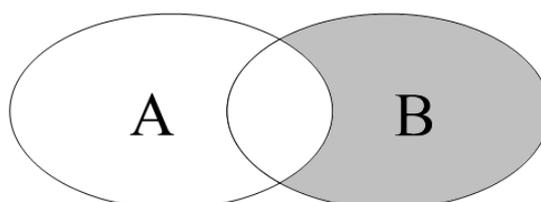
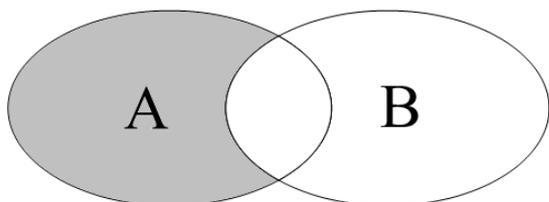
Differenz: $A \setminus B$ oder $A - B$

(auch relatives Komplement)

$$A \setminus B := \{x \mid x \in A \wedge x \notin B\}$$

dagegen $B \setminus A$

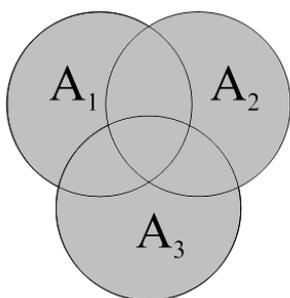
auch zu definieren mit $A \setminus B = A \cap \bar{B}$



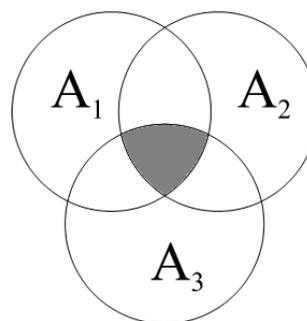
Mehr als zwei Ereignisse:

$$\bigcup_{i=1}^n A_i \quad \text{etwa } n=3: A_1 \cup A_2 \cup A_3$$

$$\bigcap_{i=1}^n A_i \quad \text{etwa } n=3: A_1 \cap A_2 \cap A_3$$



$$A_1 \cup A_2 \cup A_3$$



$$A_1 \cap A_2 \cap A_3$$

Auch die Übertragung auf überabzählbar unendlich viele Ereignisse ist möglich.

Eigenschaften der Operationen \cap (und \cup)

1. Kommutativität: $A \cap B = B \cap A$ ²⁾
2. Assoziativität: $(A \cap B) \cap C = A \cap (B \cap C)$
3. Distributivität: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
4. Adjunktivität: $A \cap (A \cup B) = A$
5. Idempotenz: $A \cap A = A$

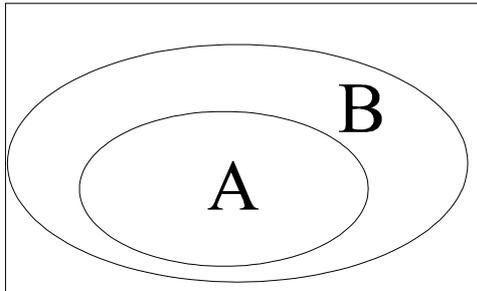
Die Differenz \setminus ist nicht kommutativ und nicht assoziativ.

²⁾ Das Zeichen = bedeutet Gleichwertigkeit (Gleichheit), d.h. gleiche Elemente enthaltend, von Mengen bzw. Ereignissen.

b) Relationen

Teilergebnis $A \subset B$

$x \in A \Rightarrow x \in B$
 $A \cap B = A, A \cup B = B$
 (auch Inklusion genannt)³



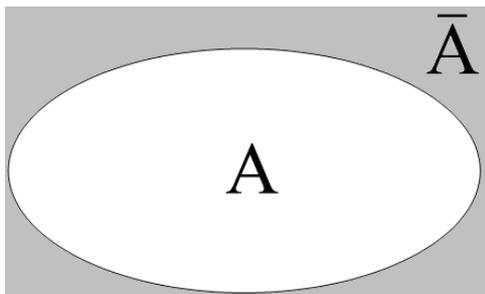
Eigenschaften der Relationen \subset

asymmetrisch: $A \subset B \neq B \subset A$
 antisymmetrisch: $A \subset B \wedge B \subset A \Rightarrow A = B$
 transitiv: $A \subset B \wedge B \subset D \Rightarrow A \subset D$
 reflexiv: wenn $A=B$, dann dann gilt $A \subset A$

Komplementärereignis \bar{A}

$\bar{A} := \{x | x \notin A\}$
 $\bar{A} = \Omega \setminus A$
 $\bar{A} \cup A = \Omega$ und $\bar{A} \cap A = \emptyset$

(auch absolutes Komplement oder Gegenereignis genannt)



$A \cup \bar{A} = \Omega, A \cap \bar{A} = \emptyset, \bar{\bar{A}} = A$
 $A \cup \emptyset = A, A \cap \emptyset = \emptyset$
 $A \cup \Omega = \Omega, A \cap \Omega = A$

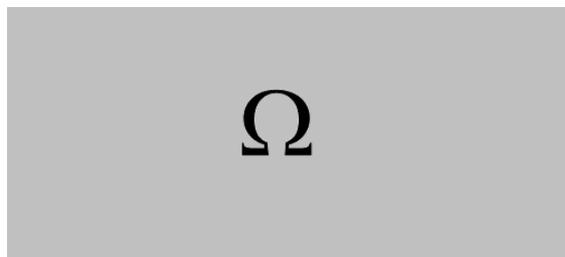
(de Morgansche Gesetze)

$\overline{A \cap B} = \bar{A} \cup \bar{B}$ und $\overline{A \cup B} = \bar{A} \cap \bar{B}$

³ Gemeint sind echte Teilmengen im Sinne der Mathematik.

c) Besondere Ereignisse und Wahrscheinlichkeiten

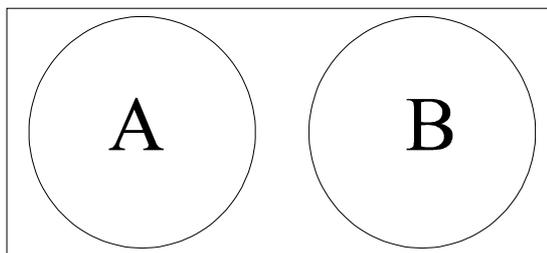
- sicheres Ereignis Ω



- unmögliches Ereignis \emptyset



- disjunkte Ereignisse
(oder elementfremd, unvereinbar,
nicht: unabhängig)
 $A \cap B = \emptyset$, $A \setminus B = A$



Mit dem Alltagsverständnis der Wahrscheinlichkeit als einer Zahl P , die zwischen 0 und 1 liegen muß, lassen sich hieraus bereits folgende Aussagen gewinnen:

1. $P(\emptyset) = 0$
2. $P(\Omega) = 1$, daraus folgt $P(\bar{A}) = 1 - P(A)$, denn wegen $A \cup \bar{A} = \Omega$ soll gelten
 $P(A) + P(\bar{A}) = P(\Omega) = 1$
3. Wenn A und B unvereinbar, dann gilt
 $P(A \cup B) = P(A) + P(B)$ und $P(A \cap B) = P(AB) = 0$
4. Wenn $A \subset B$, dann
 $P(A) \leq P(B)$ und $P(A \cup B) = P(B)$, so dass auch gilt $P(B - A) = P(B) - P(A)$,
statt allgemein $P(B - A) = P(A \cup B) - P(A)$.

d) Zusammengesetzte Ereignisse und Elementarereignisse

Definition 3.1:

Ein Ereignis $A \neq \emptyset$ heißt zusammengesetzt, wenn A dargestellt werden kann als $A = B \cup C$ (mit $B \neq A$ und $C \neq A$), andernfalls liegt ein Elementarereignis vor.

Äquivalente Definition:

Definition 3.2:

Das Ereignis $A \neq \emptyset$ ist genau dann Elementarereignis, wenn es kein Ereignis $B \neq \emptyset$, $B \neq A$ gibt, das Teilereignis von A ist. \emptyset ist kein Elementarereignis. Folgerung: Je zwei Elementarereignisse A_1 und A_2 sind disjunkt.

1.2. Produktmenge, Mengenfunktion

Definition 3.3: Produktmenge, Kartesisches Produkt

Bei zwei Mengen Ω_1, Ω_2 sind beispielsweise (a_1, b_1) oder (a_2, b_3) mit $a_i \in \Omega_1$ und $b_j \in \Omega_2$ jeweils ein geordnetes Paar (Tupel). Die Menge $\Omega_1 \times \Omega_2$ aller geordneten Paare ist die Produktmenge von Ω_1 und Ω_2 .

$$\Omega_1 \times \Omega_2 = \{(a, b) \mid a \in \Omega_1 \wedge b \in \Omega_2\}$$

$$\text{Allgemein: } \Omega_1 \times \Omega_2 \times \dots \times \Omega_n = \prod_{i=1}^n \Omega_i = \{(x_1, \dots, x_n) \mid x_i \in \Omega_i\}.$$

Im Falle von $\Omega_1 = \dots = \Omega_n = \Omega$ schreibt man auch Ω^n . Eine Produktmenge (ihre Elemente) ist mit einem Baumdiagramm darstellbar. Eine "Relation" ist eine Teilmenge der Produktmenge.

Definition 3.4: Mengenfunktion

Wird einer Menge A nach einer Zuordnungsvorschrift eine Zahl $Q(A)$ zugeordnet, so spricht man von einer Mengenfunktion.

1.3. Mengensysteme

Mengen, deren Elemente selbst wieder Mengen darstellen, nennt man Mengensysteme. Sie werden häufig abgeleitet aus einer Menge Ω , die eine Klasseneinteilung ist.

Definition 3.5: Vollständige Zerlegung, Klasseneinteilung, Partition

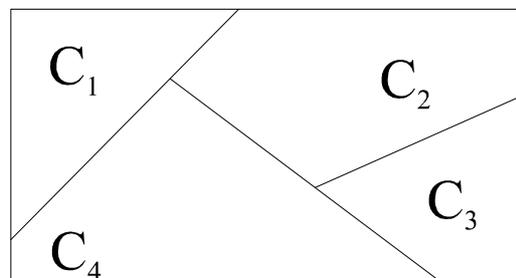
Ein Stichprobenraum Ω wird in n nichtleere, paarweise disjunkte Mengen (Ereignisse C_i) zerlegt, wenn gilt:

1. $\bigcup_{i=1}^n C_i = \Omega$ (Ausschöpfung)
2. $C_i \cap C_j = \emptyset \quad i, j = 1, 2, \dots, n; \quad i \neq j$
3. $C_i \neq \emptyset \quad \text{für alle } i = 1, 2, \dots, n \quad (n \text{ kann auch unendlich sein}).$

Veranschaulichung einer Zerlegung

$$\Omega = \{C_i \mid i = 1, 2, \dots, n\} \quad \text{für } n = 4 \quad (\text{vgl. Abb.})$$

Ein solches vollständiges System von Ereignissen ist die Menge der Elementarereignisse oder auch $\Omega = \{A, \bar{A}\}$.



Folgerung

Ist A ein beliebiges Ereignis und $\{C_1, C_2, \dots, C_n\}$ eine Zerlegung (vollständiges System), so gilt:

$$(3.1) \quad A = \bigcup_{i=1}^n (A \cap C_i)$$

(Darauf beruht der Satz der totalen Wahrscheinlichkeit, Gl. 3.14). Die Mengen (Ereignisse) $A \cap C_1, A \cap C_2, \dots$ sind disjunkt (unvereinbar).

Definition 3.6: Potenzmenge

Das Mengensystem $\mathbf{P}(\Omega)$, dessen Mengen alle möglichen Teilmengen von Ω , einschließlich \emptyset und Ω umfassen, heißt Potenzmenge (Ereignisfeld) von Ω .

Bei einem endlichen Stichprobenraum, etwa $\Omega = \{a, b, c\}$ ist das Ereignisfeld

$$\mathbf{P}(\Omega) = \{\Omega, \{a, b\}, \{a, c\}, \{b, c\}, \{a\}, \{b\}, \{c\}, \emptyset\}.$$

Bei n Elementen von Ω besteht das Ereignisfeld aus $2^n = \sum_{i=0}^n \binom{n}{i}$ Elementen.

Motivation

Es genügt nicht, die Wahrscheinlichkeit eines Ereignisses (Elementarereignisses) von Ω , etwa von A, B usw. zu definieren, sondern es muss für alle durch \cap und \cup zu bildenden, zusammengesetzten Ereignisse, eine Wahrscheinlichkeit definiert sein. Ereignisfelder sind bezüglich Vereinigung, Durchschnitt und Komplementbildung abgeschlossen.

Definition 3.7: Sigma-(σ -)Algebra

Eine Teilmenge M von $\mathbf{P}(\Omega)$, zu der Ω gehört und mit einer Menge A auch deren Komplement \bar{A} , die wegen der folgenden Festlegung Nr. 2 auch \emptyset enthält und die abgeschlossen ist gegenüber der Vereinigung (Summe) \cup von n oder auch abzählbar unendlich vielen Ereignissen A_i [Summe, daher σ ; M ist wegen Nr. 2 auch hinsichtlich \cap abgeschlossen]

- (1) $\Omega \in M$
- (2) wenn $A \in M$, dann auch $\bar{A} \in M$
- (3) $\bigcup_{i=1}^n A_i \in M$

heißt σ -Algebra. Anders als $\mathbf{P}(\Omega)$ muss M nicht einelementige Mengen enthalten.

Motivation

Die Potenzmenge kann sehr groß oder bei nicht endlichem Ω auch unendlich sein. Die Wahrscheinlichkeit ist eine auf ein Mengensystem definierte reellwertige Funktion, eine Mengenfunktion, deren Definitionsbereich eine σ -Algebra ist. Sie gibt an, welche Mengen in ihm mindestens enthalten sein müssen, um die Wahrscheinlichkeit definieren zu können?

Definition 3.8: Wahrscheinlichkeit

Sei M eine σ -Algebra. Eine auf M definierte Funktion $P: M \rightarrow \mathbb{R}$ heißt Wahrscheinlichkeitsmaß, wenn die folgenden (Kolmogoroff'schen) Axiome erfüllt sind.

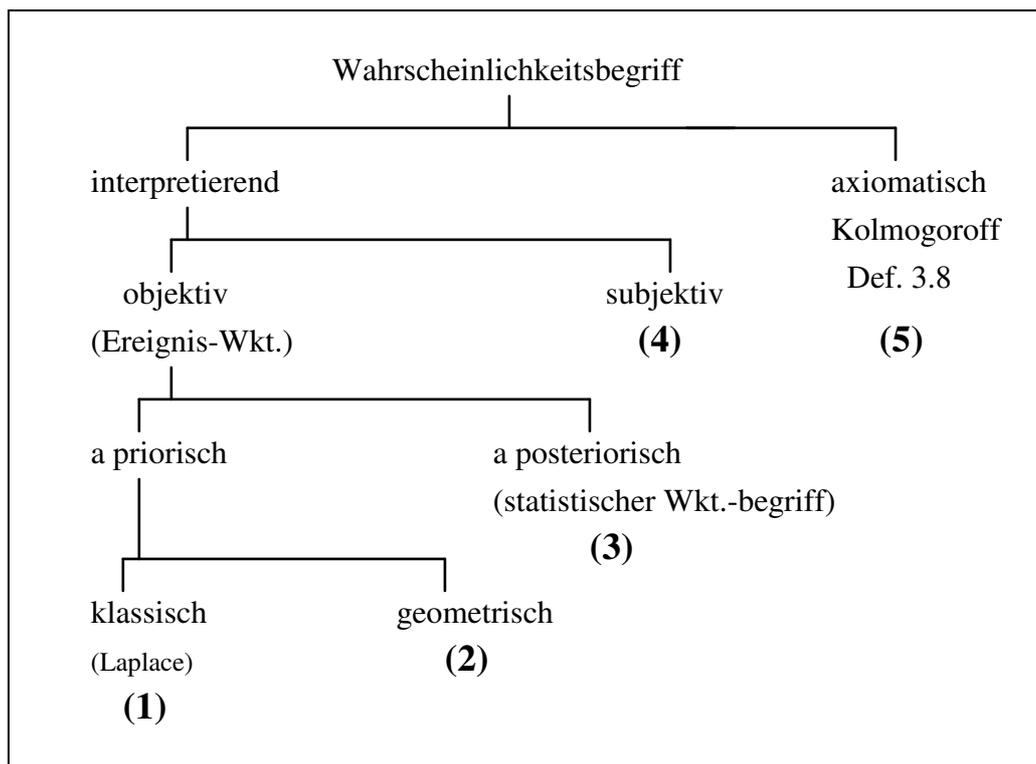
1) $P(A) \geq 0$	Nichtnegativität für alle $A \in \mathcal{M}$
2) $P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$	Volladditivität (σ -Additivität), wobei alle Folgen A_i Zerlegungen von Ω seien also $A_i \cap A_j = \emptyset$
3) $P(\Omega) = 1$	Normierung ($0 \leq P(\cdot) \leq 1$), sicheres Ereignis

Die Wahrscheinlichkeit ist ein normiertes, additives Maß auf eine σ -Algebra. Ist Ω endlich, so genügt es, Wahrscheinlichkeiten für Elementarereignisse zu definieren, alle anderen Wahrscheinlichkeiten folgen daraus. Wegen Axiom 2 und 3 folgt aus Axiom 1 auch $P(A) \leq 1$.

2. Wahrscheinlichkeitsbegriff

Die Axiome von Definition 3.8 legen die mathematischen Eigenschaften von Wahrscheinlichkeiten fest. Sie geben keine Auskunft darüber, wie man den numerischen Wert einer bestimmten Wahrscheinlichkeit erhält und interpretiert (Berechnungsanweisung, Interpretationsproblem). Zu Versuchen, diese Probleme im Wahrscheinlichkeitsbegriff zu lösen, vgl. Übersicht 3.1.

Übersicht 3.1



(1) Klassischer Wahrscheinlichkeitsbegriff

Die Wahrscheinlichkeit des Ereignisses A ist die Häufigkeit n_A des Eintretens von A (oder $|A|$ = Mächtigkeit der Menge A) dividiert durch die Anzahl n aller möglichen Fälle:

$$P(A) = \frac{n_A}{n} = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der gleichmöglichen Fälle}}$$

(2) Geometrischer Wahrscheinlichkeitsbegriff

Auch anwendbar bei überabzählbar unendlichem Stichprobenraum Ω .

(3) A posteriorisch (v. Mises)

Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit bei sehr vielen Beobachtungen ($n \rightarrow \infty$).

(4) Subjektiver Wahrscheinlichkeitsbegriff

Maß für den Grad der Überzeugtheit von der Richtigkeit einer Aussage (logische Wahrscheinlichkeit, Hypothesenwahrscheinlichkeit).

(5) Axiomatischer Begriff

Er ist z.B. insofern allgemeiner als 1, weil nicht eine endliche Menge Ω mit gleichwahrscheinlichen Elementarereignissen vorausgesetzt wird. Begriff 1 ist als Spezialfall enthalten.

3. Additionssätze

Bestimmung der Wahrscheinlichkeit einer Vereinigung. Im Falle unverträglicher Ereignisse Vereinfachung (spezielle Additionssätze, Übersicht 3.2). Allgemein gilt:

$$P(A_1 \cup \dots \cup A_n) \leq \sum P(A_i) \quad i = 1, 2, \dots, n \quad (\text{Boolesche Ungleichung}).$$

Übersicht 3.2: Additionssätze allgemein, darunter (*) speziell ($A_i \cap A_j = \emptyset, \forall i, j$)

	Additionssätze	
Zwei Ereignisse A, B	(3.3)	$P(A \cup B) = P(A) + P(B) - P(AB)$
	(3.3*)	$P(A \cup B) = P(A) + P(B)$
Drei Ereignisse A, B, C	(3.4)	$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$
	(3.4*)	$P(A \cup B \cup C) = P(A) + P(B) + P(C)$
Allgemeine Ereignisse A_1, A_2, \dots, A_n	(3.5)	Formel von Sylvester (siehe unten)
	(3.5*)	$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$

Formel von Sylvester

$$(3.5) \quad P\left(\bigcup_{i=1}^n A_i\right) = (-1)^2 \sum_{i=1}^n P(A_i) + (-1)^3 \sum_{i < j} P(A_i A_j) + \dots + (-1)^{n+1} P(A_1 A_2 \dots A_n)$$

Interpretation: Summanden

erster Summand: $\sum P(A_i) \rightarrow$ Vorzeichen positiv, denn $(-1)^2=+1$ bestehend aus $\binom{n}{1} = n$ Summanden

zweiter Summand: $\sum P(A_i A_j) \rightarrow$ Vorzeichen negativ, denn $(-1)^3=-1$ bestehend aus $\binom{n}{2}$ Summanden $A_1 A_2, A_1 A_3, \dots, A_1 A_n, A_2 A_3, \dots, A_2 A_n$ usw.

letzter Summand: bestehend aus $\binom{n}{n} = 1$ Summanden

4. Multiplikationssätze, stochastische Unabhängigkeit

Definition 3.9: Bedingte Wahrscheinlichkeiten

$P(A|B)$ ist die Wahrscheinlichkeit des Eintreffens des Ereignisses A unter der Voraussetzung, dass Ereignis B eingetreten ist (muß nicht eine zeitliche Folge, "zuerst B, dann A", sein).

$$(3.6) \quad P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{entsprechend} \quad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Für bedingte Wahrscheinlichkeiten gelten die gleichen Axiome und Sätze wie für unbedingte Wahrscheinlichkeiten, z.B. $P(A \cup C|B) = P(A|B) + P(C|B) - P(AC|B)$ (Additionssatz).

Generell gilt:

$$(3.13) \quad P(A|B) + P(\bar{A}|B) = 1 \quad \text{oder} \quad P(AB|C) + P(\bar{A}\bar{B}|C) = 1 \quad \text{usw.}, \text{ aber nicht}$$

$$P(A|B) + P(A|\bar{B}) = 1 \quad \text{oder} \quad P(AB|C) + P(AB|\bar{C}) = 1.$$

Definition 3.10: Stochastische Unabhängigkeit

a) **paarweise** (pairwise) Unabhängigkeit von zwei Ereignissen A, B bedeutet:

$$(3.7) \quad P(AB) = P(A) P(B)$$

oder gleichbedeutend

$$(3.8) \quad P(A) = P(A|B)$$

und wegen der Symmetrie der Unabhängigkeit gilt dann auch $P(B) = P(B|A)$ oder

$$(3.9) \quad P(A|B) = P(A|\bar{B}) = P(A) \quad \text{und} \quad P(B|A) = P(B|\bar{A}) = P(B)$$

b) **wechselseitige** (mutual) Unabhängigkeit bei mehr als zwei, z.B. bei drei Ereignissen bedeutet:

$$(3.10) \quad P(ABC) = P(A) P(B) P(C)$$

Mit Gleichung 3.10 gilt auch paarweise Unabhängigkeit von A und B, A und C sowie B und C, nicht aber umgekehrt. Wechselseitige Unabhängigkeit ist also eine strengere Forderung als paarweise Unabhängigkeit.

Bemerkung zu Definition 3.10

Mehrdeutigkeit des Begriffs Unabhängigkeit in der Statistik:

- unabhängige Züge (bei wiederholter Ziehung aus einer Urne mit Zurücklegen, → Kap. 4, 5)
- bei mehrdimensionaler Zufallsvariable (→ Kap. 4)
- unabhängige und abhängige (verbundene) Stichproben (→ Kap. 9)
- in der Regressionsanalyse: "unabhängige Variablen"

Multiplikationssätze

Gegenstand: Bestimmung der Wahrscheinlichkeit $P(AB)$, $P(ABC)$, usw. aus bedingten und unbedingten Wahrscheinlichkeiten. Bei Unabhängigkeit jeweils Spezialfall des Multiplikationssatzes. Vgl. Übersicht 3.3.

Folgerungen, Verallgemeinerungen

Bedingte Wahrscheinlichkeit allgemeinerer Art: $P(A_1 \dots A_m | B_1 \dots B_k) = \frac{P(A_1 \cap \dots \cap B_k)}{P(B_1 \cap \dots \cap B_k)}$.

Entsprechend läßt sich auch der Multiplikationssatz allgemeiner formulieren. Beispiel:

$$P(A_1 A_2 | B_1 B_2 B_3) = P(A_1 A_2 | B_1 B_2 B_3) P(B_1 | B_2 B_3) P(B_2 | B_3) P(B_3).$$

Übersicht 3.3: Multiplikationssätze allgemein, darunter ()
speziell (A_i, A_j paarweise unabhängig)*

	Multiplikationssätze
Zwei Ereignisse A, B	(3.11) $P(AB) = P(A)P(B A) = P(B)P(A B)$ (3.11*) $P(AB) = P(A) P(B)$
n Ereignisse A_1, \dots, A_n	(3.12) $P(A_1 \dots A_n) = P(A_1 A_2 \dots A_n) \cdot P(A_2 A_3 \dots A_n) \cdot P(A_3 A_4 \dots A_n) \dots P(A_{n-1} A_n) P(A_n)$ (3.12*) $P(A_1 \dots A_n) = P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$

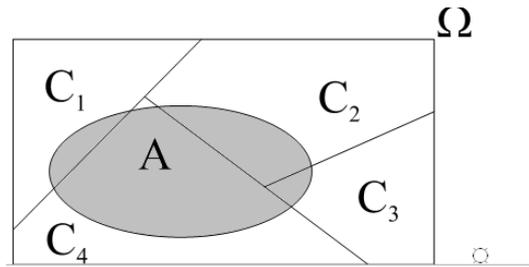
5. Totale Wahrscheinlichkeit, Bayessesches Theorem

1) Satz von der totalen Wahrscheinlichkeit

Es sei $\{C_i | i = 1, \dots, n\}$ eine vollständige Zerlegung von Ω mit sich gegenseitig ausschließenden Ereignissen C_i mit $P(C_i) > 0$. Weiterhin sei $A \subset \Omega$. Dann gilt nach Gl. 3.1:

$$(3.14) \quad P(A) = \sum_i P(C_i) \cdot P(A|C_i) = \sum_i P(C_i \cap A)$$

$$(3.14a) \quad P(AB) = \sum_i P(C_i) \cdot P(AB|C_i)$$



2) Bayessches Theorem

$$(3.15) \quad P(C_i|A) = \frac{P(C_i) \cdot P(A|C_i)}{\sum_i P(C_i) \cdot P(A|C_i)} = \frac{P(A \cap C_i)}{P(A)}$$

Interpretation: Bedeutung für Schätz- und Testtheorie:

C_1, C_2, \dots : Alternativen, Hypothesen, Ursachen (für Beobachtungen)

$P(C_1), P(C_2), \dots$: a priori Wahrscheinlichkeiten $\left(\sum_i P(C_i) = 1 \right)$

$P(C_1|A), P(C_2|A), \dots$: a posteriori Wahrscheinlichkeiten $\left(\sum_i P(C_i|A) = 1 \right)$

$P(A|C_1), P(A|C_2), \dots$: Likelihoods

Spezielle Fälle

- $P(C_i|A) = P(C_i)$ für alle i , wenn Likelihoods gleich sind
- $P(C_i|A) = 0$ oder $P(C_i|A) = 1$, d.h. bei extremen Werten für a priori-Wahrscheinlichkeiten von 0 oder 1 nehmen auch die a posteriori Wahrscheinlichkeiten extreme Werte an; entsprechend
- extreme Werte für die Likelihoods von 0 oder 1: a posteriori-Wahrscheinlichkeiten sind dann auch 0 oder 1, unabhängig von den a priori-Wahrscheinlichkeiten
- alle a priori-Wahrscheinlichkeiten gleich $P(C_1) = P(C_2) = \dots = P(C_n) = 1/n$ (Prinzip des mangelnden Grundes): dann gilt $P(C_i|A) = P(A|C_i) / \sum P(A|C_i)$.

Die totale Wahrscheinlichkeit $P(A)$ ist nach Gl. 3.14 ein gewogenes Mittel der Likelihoods.

Ist $P(A|C_i) > P(A)$, dann ist $P(C_i|A) > P(C_i)$. Ist $P(A|C_i) < P(A)$, dann ist $P(C_i|A) < P(C_i)$.

Beachte: $\sum_i P(C_i|A) = 1$, aber $\sum_i P(A|C_i)$ muss nicht 1 sein.

Kapitel 4: Zufallsvariablen, Wahrscheinlichkeitsverteilungen

1. Eindimensionale Zufallsvariablen	25
2. Mehrdimensionale Zufallsvariablen	28
3. Momente von Funktionen von Zufallsvariablen	31
4. Erzeugende Funktionen	31
5. Intervalltransformationen	35

1. Eindimensionale Zufallsvariable

Def. 4.1: Zufallsvariable

a) diskrete Zufallsvariable

1. Sei $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ der Stichprobenraum eines Zufallsexperiments mit den Ereignissen $\omega_1, \omega_2, \dots, \omega_n$. Die Funktion X , welche jedem Element $\omega_i \in \Omega$ ($i = 1, 2, \dots, n$) eine reelle Zahl $X(\omega_i) = x_i$ zuordnet, heißt **Zufallsvariable** (ZV). Der Wertebereich von X ist die Menge der reellen Zahlen \mathbb{R} . Die dem Ereignis ω_i zugeordnete Wahrscheinlichkeit $P(\omega_i)$ wird auf X übertragen, in dem Sinne, dass für $\omega_i = x_i, i = 1, \dots, n$, $P(\omega_i) = P(X = x_i)$ gilt, und es ist definiert:

$$(4.1) \quad f(x_i) = P(X=x_i) = P(\omega_i) = p_i.$$

2. Die Tupel $x_i, f(x_i)$ bei einer endlichen oder abzählbar unendlichen Folge von Werten $x_1 < x_2 < x_3 \dots < x_n$ mit

$$(4.2) \quad f(x) = \begin{cases} f(x_i) & \text{für } x_1, \dots, x_n \\ 0 & \text{sonst} \end{cases}$$

ist die **Wahrscheinlichkeitsfunktion** (oder: Zähldichte) der diskreten Zufallsvariable X .

3. Die Funktion

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u), \text{ bzw.}$$

$$F(x_j) = P(X \leq x_j) = \sum_{i=1}^j f(x_i) = \sum_{i=1}^j p_i$$

heißt **Verteilungsfunktion** der diskreten Zufallsvariable X .

b) stetige Zufallsvariable

1. Der Stichprobenraum ist überabzählbar unendlich. Definiert ist die Wahrscheinlichkeit dafür, dass X Werte in einem Intervall $x < X \leq x + \Delta x$ annimmt mit

$$(4.3) \quad P(x < X \leq x + \Delta x) = \int_x^{x+\Delta x} f(u) du.$$

2. Die Funktion $f(x)$ mit $-\infty < x < +\infty$ oder $a \leq x \leq b$, heißt **Dichtefunktion** (stetige Wahrscheinlichkeitsfunktion; man beachte aber: $f(x)$ ist keine Wahrscheinlichkeit!).
3. Die Verteilungsfunktion $F(x)$ der stetigen ZV X ist gegeben durch:

$$(4.4) \quad F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du \quad (-\infty < x < \infty).$$

Der Zusammenhang zwischen der Dichtefunktion $f(x)$ und der Verteilungsfunktion $F(x)$ der stetigen Zufallsvariablen X ist gegeben durch:

$$(4.5) \quad f(x) = \frac{dF(x)}{dx} \quad (-\infty < x < \infty).$$

Übersicht 4.1: Eigenschaften der Wahrscheinlichkeitsfunktion und Intervallwahrscheinlichkeiten

Art der Zufallsvariable	Eigenschaften der Wahrscheinlichkeitsfunktion (gem. den Kolmogoroffschen Axiomen)	Intervallwahrscheinlichkeiten
diskrete ZV	(4.6a) $0 \leq f(x) \leq 1$, (4.6b) $\sum_x f(x) = 1$, bzw. $\sum_i f(x_i) = \sum_i p_i = 1$	(4.6c) $P(a < X \leq b) = F(b) - F(a) = \sum_{a < x_i \leq b} f(x_i)^*$
stetige ZV	(4.7a) $0 \leq f(x) < \infty$, nicht aber $f(x) \leq 1$ (4.7b) $\int_{-\infty}^{\infty} f(x) dx = 1$.	(4.7c) $P(c < x \leq d) = \int_c^d f(x) dx = F(d) - F(c) \geq 0$

*) also beispielsweise: $P(x_1 < x \leq x_4) = p_2 + p_3 + p_4 = F(x_4) - F(x_1)$.

Def. 4.2: Momente

1. Die folgende Funktion der Zufallsvariable X

$$E(X-a)^r$$

ist das r -te (theoretische) **Moment** um a . Ist $a = 0$, so spricht man von **Anfangsmomenten**, ist $a = E(X) = \mu$ von **zentralen Momenten**. Zwischen Anfangsmomenten $E_k = E(X^k)$ und zentralen Momenten $M_k = E[(X-\mu)^k]$ besteht die folgende Beziehung:

$$M_k = \sum_{i=0}^k \binom{k}{i} E_{k-i} (-\mu)^i, \text{ mit } E_1 = \mu.$$

2. Das **erste Anfangsmoment** $E(X) = E_1 = \mu$ heißt **Erwartungswert** und ist gegeben mit

$$(4.8) \quad \mu = E(X) = \sum x f(x) = \sum_{i=1}^n x_i f(x_i), \text{ wenn } X \text{ diskret ist, bzw. mit}$$

$$(4.9) \quad \mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx \text{ wenn } X \text{ stetig ist.}$$

3. Das zweite zentrale Moment ist die theoretische **Varianz**.

$$(4.10) \quad V(X) = \sigma^2 = E[X - E(X)]^2 = E(X - \mu)^2 = E(X^2) - \mu^2$$

Wie im einzelnen bei diskretem und stetigem X zu rechnen ist, vgl. Übers. 4.2.

4. Der Erwartungswert

$$(4.11) \quad E^*(X, k) = E[X(X-1)(X-2) \dots (X-k+1)] \\ E^*(X, 1) = E(X), E^*(X, 2) = E[X(X-1)] \text{ usw. , } (k = 1, 2, 3, \dots),$$

ist das k -te **faktorielle Moment** von X (um Null).

Übersicht 4.2: Gegenüberstellung der Terminologie
der Deskriptiven und Induktiven Statistik

1. **diskrete** Variable X

	Induktive Statistik	Deskriptive Statistik
1a) Verteilungs- begriffe	Wahrscheinlichkeitsverteilung $f(x_i) = p_i = P(X=x_i)$	Häufigkeitsverteilung¹⁾ $h_i = \frac{n_i}{n}$
	Verteilungsfunktion²⁾ $F(x_j) = \sum_{i=1}^j f(x_i) = P(X \leq x_i)$	Summenhäufigkeit (kumulierte rel. Häufigkeit) $H_j = \sum_{i=1}^j h_i$
1b) Momente	Erwartungswert $E(X) = \mu = \sum_i x_i f(x_i)$	Mittelwert $\bar{x} = \sum_i x_i h_i$
	Varianz (theoretische) $V(X) = \sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 f(x_i)$	Varianz (empirische) $s^2 = \sum_i (x_i - \bar{x})^2 h_i$

1) relative Häufigkeit h_i .

2) Die Verteilungsfunktion $F(x)$ ist eine monoton nichtfallende, rechtsseitig stetige Treppenfunktion.

2. **stetige** Variable X (nur induktive Statistik)

2a) Verteilungs- begriffe	Dichtefunktion	$f(x) = \frac{dF(x)}{dx}$	Verteilungsfunktion	$F(x) = \int_{-\infty}^x f(u) du$
2b) Momente	Erwartungswert	$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x) dx$		
	Varianz (theoretische)	$V(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$		

Eigenschaften des Erwartungswerts

Der Erwartungswert E ist ein **linearer Operator**:

- Erwartungswert einer Konstanten a: $E[a] = a$.
- Erwartungswert einer Lineartransformation: $E(bX) = b E(X)$.
Bei $Y = a + bX$ (Lineartransformation) gilt: $E(Y) = a + b E(X)$.
- Funktionen ϕ der diskreten Zufallsvariable X:

$$E \left[\sum_i c_i \phi_i (X) \right] = \sum_i c_i E [\phi_i (X)],$$

wobei $\phi (X)$ eine Funktion der diskreten Zufallsgröße X ist (entsprechende Formel bei stetiger Zufallsvariable).

Wichtiger Hinweis zu Übers. 4.2

Auf dem ersten Blick mag es die klare Analogie zur Deskriptiven Statistik geben, zumindest im diskreten Fall.

Wahrscheinlichkeits-verteilung:

X_1	X_2	...	X_m
p_1	p_2	...	p_m

Häufigkeits-verteilung

X_1	X_2	...	X_m
h_1	h_2	...	h_m

Erwartungswert $E(X) = \sum x_i \cdot p_i$

Mittelwert $\bar{x} = \sum x_i \cdot h_i$

so dass manche geneigt sind diese Begriffe praktisch gleich zu setzen und die Unterschiede nicht zu erkennen. Es ist deshalb **unbedingt zu beachten**:

1. relative Häufigkeiten beziehen sich auf **endlich** viele Beobachtungen, Wahrscheinlichkeiten dagegen auf ein prinzipiell **unendlich** oft wiederholbares Zufallsexperiment;
2. das prägt auch den Unterschied zwischen dem **Erwartungswert** und dem Mittelwert. Der Erwartungswert **kann** z.B. auch **nicht endlich sein** (nicht "existieren"), während der Mittelwert einer (empirischen) Häufigkeitsverteilung immer endlich ist.

2. Mehrdimensionale Zufallsvariablen

Def. 4.3: Zweidimensionale Wahrscheinlichkeitsverteilung

1. Die Funktion $f(x,y)$ ist die **gemeinsame Wahrscheinlichkeitsverteilung** (bei diskreten Zufallsvariablen X,Y) bzw. gemeinsame Dichtefunktion (bei stetigen Zufallsvariablen X,Y). Im diskreten Fall ist $f(x_i,y_j)$ eine Wahrscheinlichkeit:

$$f(x_i,y_j) = P(X = x_i \text{ und } Y = y_j), \quad i = 1, 2, \dots, m, j = 1, 2, \dots, k,$$

und die gemeinsame Wahrscheinlichkeitsfunktion lautet:

$$(4.12) \quad f(x,y) = \begin{cases} f(x_i, y_j) & \text{für } X = x_i, Y = y_j \\ 0 & \text{sonst} \end{cases}$$

2. Hieraus abgeleitet werden:

- a) eine zweidimensionale Verteilungsfunktion $F(x,y)$,
- b) zwei eindimensionale **Randverteilungen** $f_1(x)$, $f_2(y)$,
- c) eindimensionale **bedingte** Verteilungen (im diskreten Fall m bedingte Verteilungen f_{by} und k bedingte Verteilungen f_{bx}).

3. Die **Verteilungsfunktion** ist definiert als

$$(4.13a) \quad F(x,y) = P(X \leq x, Y \leq y) = \sum_{v \leq y} \sum_{u \leq x} f(u,v) \quad \text{im diskreten Fall bzw.}$$

$$(4.13b) \quad F(x,y) = \int_{-\infty}^y \int_{-\infty}^x f(u,v) du dv \quad \text{im stetigen Fall.}$$

4. **Randverteilungen** f_1, f_2

$$(4.14a) \quad f_1(x) = \sum_y f(x,y) = P(X = x) \quad \text{bzw.} \quad f_1(x) = \int_{-\infty}^{\infty} f(x,y) dy,$$

$$(4.14b) \quad f_2(y) = \sum_x f(x,y) = P(Y=y) \quad \text{bzw.} \quad f_2(y) = \int_{-\infty}^{\infty} f(x,y) dx .$$

5. Bedingte Verteilungen f_{bx}, f_{by}

$$(4.15) \quad f_{bx}(x|y) = \frac{f(x,y)}{f_2(y)} \quad (\text{bedingte Verteilung der Variable X}),$$

$$(4.15b) \quad f_{by}(y|x) = \frac{f(x,y)}{f_1(x)} \quad (\text{bedingte Verteilung der Variable Y}),$$

im diskreten Fall sind dies die bedingten Wahrscheinlichkeiten $P(X=x|Y=y)$ und $P(Y=y|X=x)$.

Eigenschaften der gemeinsamen Wahrscheinlichkeitsverteilung

im diskreten Fall	1. $0 \leq f(x,y) \leq 1$	im stetigen Fall	1. $0 \leq f(x,y)$
	2. $\sum_y \sum_x f(x,y) = 1$		2. $\iint f(x,y) dx dy = 1$

Def. 4.4: Momente und Produktmomente

1. Die Momente der **Randverteilungen** sind $E(X)$, $E(Y)$, $V(X)$, $V(Y)$ usw.. Es gilt z.B. bei diskretem X

$$E(X) = \sum_x x f_1(x) \quad \text{und} \quad V(X) = \sigma_X^2 = \sum_x x^2 f_1(x) - E[(X)]^2 = E(X^2) - [E(X)]^2$$

oder bei stetigem X

$$E(X) = \int_{-\infty}^{\infty} x f_1(x) dx \quad \text{und} \quad V(X) = \int_{-\infty}^{\infty} x^2 f_1(x) dx - [E(X)]^2$$

und die Momente von Y entsprechend.

2. Bedingte Erwartungswerte

$$(4.16a) \quad E(Y|X=x) = \sum_y y f_b(y|x) = \frac{1}{f_1(x)} \sum_y y f(x,y) \quad \text{im diskreten Fall bzw.}$$

$$(4.16b) \quad E(Y|X=x) = \int_{-\infty}^{+\infty} y f_b(y|x) dy = \frac{1}{f_1(x)} \int_{-\infty}^{+\infty} y f(x,y) dy \quad \text{im stetigen Fall}$$

und $E(X|Y=y)$ entsprechend.

3. Produktmomente (Kovarianz)

Die (theoretische) Kovarianz $C(X,Y)$ als zentrales Produktmoment ist definiert als

$$(4.17) \quad C(X,Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$$

d.h. im diskreten Fall: $\sigma_{XY} = \sum_y \sum_x x \cdot y f(x,y) - \mu_X \mu_Y$

bzw. im stetigen Fall: $\sigma_{XY} = \int_a^b \int_c^d x \cdot y f(x,y) dx dy - \mu_X \mu_Y,$

wenn für den Definitionsbereich gilt: $a \leq y \leq b$ und $c \leq x \leq d$.

4. Der (theoretische) **Korrelationskoeffizient** ρ_{xy} ist die auf den Wertebereich $[-1,+1]$ normierte (theoretische) Kovarianz σ_{xy} : $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

Er ist das Produktmoment der **standardisierten** Zufallsvariablen X^*, Y^* mit:

$$X^* = \frac{X - \mu_x}{\sigma_x} \quad \text{und} \quad Y^* = \frac{Y - \mu_y}{\sigma_y}, \quad \text{also} \quad E(X^* Y^*) = \rho_{XY}$$

5. **Stochastische Unabhängigkeit:** X und Y sind unabhängig, wenn für $f(x,y)$ gilt:

(4.18) $f(x,y) = f_1(x)f_2(y)$	und damit: $f_{bx}(x y) = \frac{f(x,y)}{f_2(y)} = f_1(x), f_{by}(y x) = f_2(y)$
--------------------------------	---

Stochastisch unabhängige Zufallsvariablen X,Y sind stets auch unkorreliert (aber die Umkehrung dieses Satzes ist nicht zulässig).

6. Verallgemeinerung für mehr als zwei Dimensionen

Es sei $\mathbf{x}' = [X_1 \ X_2 \ \dots \ X_m]$ ein m-dimensionaler Zufallsvektor mit den Realisationen $[x_1 \ x_2 \ \dots \ x_m]$. Die Parameter der gemeinsamen Wahrscheinlichkeitsfunktion (sofern sie existieren, d.h. endlich sind) werden in der symmetrischen und positiv definiten Momentenmatrix \mathbf{M} (oder Σ)

$$(4.19) \quad \mathbf{M} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1m} \\ \vdots & & & \vdots \\ \sigma_{m1} & \sigma_{m2} & \dots & \sigma_m^2 \end{bmatrix}$$

zusammengefasst (Varianz-Kovarianz-Matrix der m Zufallsvariablen). Die Determinante $|\mathbf{M}|$ dieser Matrix heißt "**verallgemeinerte Varianz**". Die Bedingung $|\mathbf{M}| = 0$ ist notwendig und hinreichend dafür, dass mit Wahrscheinlichkeit Eins unter den Zufallsvariablen mindestens eine lineare Beziehung besteht (exakt erfüllt ist; also etwa $X_i = a + b X_j$). Ist $|\mathbf{M}| \neq 0$, so gilt $|\mathbf{M}| \leq \sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_m^2$, so dass die Determinante ihren größten Wert dann hat, wenn alle Kovarianzen σ_{ij} (und damit auch Korrelationen ρ_{ij} verschwinden.). Es gilt also $0 \leq |\mathbf{M}| \leq \sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_m^2$.

Entsprechend ist die Korrelationsmatrix:

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \cdot & \cdot & \rho_{1m} \\ \rho_{21} & 1 & \cdot & \cdot & \rho_{2m} \\ \cdot & & & & \\ \cdot & & & & \\ \rho_{m1} & \rho_{m2} & \cdot & \cdot & 1 \end{bmatrix}$$

ebenfalls symmetrisch und positiv definit und es gilt $0 \leq |\mathbf{R}| \leq 1$.

Die Varianzen (Hauptdiagonale von \mathbf{M}) und der Erwartungswertvektor $\boldsymbol{\mu}' = [\mu_1 \ \mu_2 \ \dots \ \mu_m]$ sind Parameter der Randverteilungen.

3. Momente von Funktionen von Zufallsvariablen

a) Lineare Funktionen von Zufallsvariablen

Def. 4.5: Linearkombinationen und -transformationen

Die Zufallsvariable

(4.20)	$Y = a + bX$, a, b : Konstante	ist eine Lineartransformation der Zufallsvariablen X und
(4.21)	$Z = b_1X_1 + b_2X_2 + \dots + b_nX_n$ (b_1, b_2, \dots, b_n konstante "Gewichte")	ist eine (gewogene) Linearkombination der Zufallsvariablen X_1, X_2, \dots, X_n

Bemerkungen zu Def. 4.5:

1. ie Lineartransformation ist ein Spezialfall der Linearkombination, wenn in $Y = aX_0 + bX$ die Zufallsvariable X_0 degeneriert ist zu einer Einpunkt-Verteilung mit $X_0 = x_0 = 1$ und $p_0 = 1$.
2. Spezialfälle von Gl. 4.22 sind die **ungewogene** Linearkombination
 $Z = X_1 + X_2 + \dots + X_n$, mit $b_1 = b_2 = \dots = b_n = 1$ oder
 $\bar{X} = \frac{1}{n} Z$ das arithmetische Mittel, mit $b_1 = b_2 = \dots = b_n = \frac{1}{n}$.
3. Für eine weitere Betrachtung ist entscheidend, ob die Zufallsvariablen X_1, X_2, \dots, X_n paarweise stochastisch unabhängig sind oder nicht (Übers. 4.3, nächste Seite).

b) Produkte von Zufallsvariablen

Bei **unabhängigen** Zufallsvariablen (bei abhängigen sehr komplizierte Formeln) gilt:

$$(4.22) \quad E(X_1 X_2 \dots X_n) = E(X_1) E(X_2) \dots E(X_n).$$

4. Erzeugende Funktionen

Ist X eine ZV, dann ist eine Funktion von X etwa t^X oder e^{tX} , ($t \in \mathbb{R}$) eine ZV mit einer Wahrscheinlichkeitsverteilung und einem Erwartungswert, der dann eine Funktion von t ist. In diesem Abschnitt werden solche Funktionen betrachtet.

Def. 4.6: Erzeugende Funktion, Faltung

1. Eine erzeugende Funktion der Zufallsvariable X ist eine Funktion der reellen Zahl t , deren Ableitungen bestimmte nützliche Eigenschaften haben. Es gibt verschiedene Arten von erzeugenden Funktionen (vgl. Übers. 4.3). In Übersicht 4.3 werden einige erzeugende Funktionen definiert. Darin ist $f^{(n)}(j)$ die n -te Ableitung der erzeugenden Funktion f nach t an der Stelle $t = j$.
2. Unter der **Faltung (convolution)** von zwei unabhängigen Zufallsvariablen X und Y mit den Wahrscheinlichkeitsverteilungen $f_1(x)$ und $f_2(y)$ versteht man eine Zufallsvariable Z , für deren Wahrscheinlichkeitsverteilung f gilt:

Übersicht 4.3: Formeln für Linearkombinationen

a) Momente von Linear-kombinationen

$$Z = b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

1. Erwartungswert	$E(Z) = b_1 E(X_1) + b_2 E(X_2) + \dots + b_n E(X_n)$
2. Varianz a) unabhängige Zufallsvariablen b) keine Unabhängigkeit	$\sigma_Z^2 = b_1^2 \sigma_1^2 + \dots + b_n^2 \sigma_n^2$ $\sigma_Z^2 = b_1^2 \sigma_1^2 + b_2^2 \sigma_2^2 + \dots + b_n^2 \sigma_n^2$ $+ 2(b_1 b_2 \sigma_{12} + b_1 b_3 \sigma_{13} + \dots + b_{n-1} b_n \sigma_{n-1,n})$

Spezialfall: Arithmetisches Mittel (ungewogen): $\bar{X} = \frac{1}{n} X_1 + \frac{1}{n} X_2 + \dots + \frac{1}{n} X_n$

1. Erwartungswert	$E(\bar{X}) = \mu_{\bar{X}} = \frac{1}{n} \sum_{i=1}^n E(X_i)$
2. Varianz	$V(\bar{X}) = \sigma_{\bar{X}}^2 = \left(\frac{1}{n}\right)^2 V(X_1) + \left(\frac{1}{n}\right)^2 V(X_2) + \dots + \left(\frac{1}{n}\right)^2 V(X_n)$ $+ 2\left(\frac{1}{n}\right)^2 \sigma_{12} + 2\left(\frac{1}{n}\right)^2 \sigma_{23} + \dots + 2\left(\frac{1}{n}\right)^2 \sigma_{n-1,n}$

Wenn $E(X_i) = \mu$ und $V(X_i) = \sigma_i^2 = \sigma^2$ für alle $i=1, \dots, n$, und die ZV'en unkorreliert sind ,

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \quad \text{und} \quad V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

b) Produktmoment eines Produkts von Linearkombinationen

Beispiel: $Z_1 = a_1 X_1 + a_2 X_2$ und $Z_2 = b_1 X_1 + b_2 X_2 + b_3 X_3$

$C(Z_1, Z_2) = a_1 b_2 V(X_1) + a_1 b_2 C(X_1, X_2) + a_1 b_3 C(X_1, X_3) + a_2 b_1 C(X_1, X_2)$ $+ a_2 b_2 V(X_2) + a_2 b_3 C(X_2, X_3)$

c) Linear-transformation $Y = a + bX$

$E(Y) = a + b \cdot E(X) = a + b \mu_X \quad \text{und} \quad V(Y) = \sigma_Y^2 = b^2 V(X) = b^2 \cdot \sigma_X^2$
--

(4.23a) $f(z) = \sum_x P(X = x)P(Y = z - x) = \sum_y P(Y = y)P(X = z - y)$ im diskreten Fall

bzw.

(4.23b) $f(z) = \int f_1(x)f_2(z - x) = \int f_2(y)f_1(z - y)dy$ im stetigen Fall.

Man schreibt auch $Z = X * Y$, wenn Z eine Faltung darstellt.

Zu dem Konzept der erzeugenden Funktion führen wir drei für Kap. 5 bis 7 sehr bedeutsame Sätze ohne Beweis an, aus denen Zusammenhänge zwischen Verteilungen und Grenzübergängen deutlich werden. Nicht für nur Beweise sondern auch für die Berechnung von Momenten können erzeugende Funktionen von großem Nutzen sein.

Übersicht 4.4: Erzeugende Funktionen

Name	Definition	Bedeutung der Ableitungen*)
wahrscheinlichkeitserzeugende Funktion $W_x(t)$	$W_x(t) = E(t^X)$, wenn $ t \leq 1$, $t \in \mathbb{R}$, $x \in \mathbb{N}$ (X ist eine nichtnegative ganzzahlige Zufallsvariable)	1. $\frac{W_x^{(k)}(0)}{k!} = p_k$ 2. $W_x^{(k)}(1) = E^*(X, k)$ $= E[X(X-1) \dots (X-k+1)]$ (k-tes faktorielles Moment)
faktorielle momentenerzeugende Funktion $\phi_x(t)$	$\phi_x(t) = W_x(1+t) = E[(1+t)^X]$ $= \sum_{x=0}^{\infty} (1+t)^x p_x$ (X ist eine nichtnegative ganzzahlige Zufallsvariable)	$\phi_x^{(k)}(0) = E^*(X, k) = E[X(X-1) \dots (X-k+1)]$ das k-te faktorielle Moment**)
momentenerzeugende Funktion $M_x(t)$	$M_x(t) = E(e^{tx})$ im diskreten Fall $M_x(t) = \sum e^{tx_i} p_i$ im stetigen Fall $M_x(t) = \int e^{tx} f(x) dx$ (X ist eine beliebige reellwertige Zufallsvariable, $x \in \mathbb{R}$)	$M_x^{(k)}(0) = E(X^k)$ das k-te Anfangsmoment
charakteristische Funktion $\Psi_x(t)$	$\Psi_x(t) = E(e^{itx})$ $i^2 = -1, x \in \mathbb{R}$	$\Psi_x^{(k)}(0) = i^k E(X^k)$ Ψ ist die Fouriertransformierte der Dichtefunktion $f(x)$

*) die nullte Ableitung von W_x an der Stelle 0 beträgt $W_x(0) = p_0$, aber z.B. bei der momentenerzeugenden Funktion $M_x(0) = 1$ (ebenso bei $\phi_x(t)$ und $\Psi_x(t)$).

***) Beispiel: $\phi_x''(0) = E[X(X-1)] = E^*(X, 2)$.

Bemerkungen zu Def.4.6:

1. Ist X eine diskrete Zufallsvariable, die nur ganze, positive Zahlen annehmen kann, also die Wahrscheinlichkeitsverteilung mit

x	0	1	2	...	n
f(x)	p ₀	p ₁	p ₂	...	p _n

gegeben, so ist die wahrscheinlichkeitserzeugende Funktion $W_x(t)$ wie folgt definiert:

$$(4.24) \quad W_x(t) = E(t^X) = p_0 + p_1 t + p_2 t^2 + \dots + p_n t^n = \sum_{x=0}^n p_x t^x.$$

Die Ableitungen nach t sind:

$$W_x'(t) = p_1 + 2 t p_2 + 3 t^2 p_3 + 4 t^3 p_4 + \dots + n \cdot t^{n-1} p_n = \sum_{x=1}^n x t^{x-1} p_x$$

$$W_x''(t) = 2 \cdot 1 t^0 p_2 + 3 \cdot 2 t^1 p_3 + 4 \cdot 3 t^2 p_4 + \dots + n(n-1)t^{n-2} p_n = \sum_{x=2}^n x(x-1)t^{x-2} p_x \quad \text{usw.,}$$

so dass $W_x'(0) = 1! p_1 = p_1$, $W_x''(0) = 2! p_2$ bzw. allgemein:

$$(4.25) \quad W_x^{(k)}(0) = k! p_k.$$

Ferner gilt

$$W_x'(1) = p_1 + 2p_2 + 3p_3 + \dots + n p_n = E(X),$$

$$W_x''(1) = E(X^2) - E(X) = E[X(X-1)] \text{ das zweite faktorielle Moment}$$

(nach Def. 4.2 also $E^*(X,2) = W_x''(1)$, oder allgemein $E^*(X,k) = W_x^{(k)}(1)$).

2. Gegeben sei eine Zufallsvariable mit den vier Ausprägungen 0, 1, 2 und 3, deren Wahrscheinlichkeiten p_0, p_1, p_2 und p_3 sind. Dann ist die faktorielle momenterzeugende Funktion

$$\begin{aligned} \phi_X(t) &= p_0 + (1+t)p_1 + (1+t)^2 p_2 + (1+t)^3 p_3 \\ &= 1 + t p_1 + 2t p_2 + 3t p_3 + t^2 p_2 + 3t^2 p_3 + t^3 p_3 \end{aligned}$$

und die ersten beiden Ableitungen nach t sind an der Stelle $t = 0$

$$\phi_X'(0) = p_1 + 2p_2 + 3p_3 = E(X) = E^*(X,1), \text{ und } \phi_X''(0) = 2p_2 + 6p_3 = E[X(X-1)] = E^*(X,2).$$

Entsprechend erhält man für die k -te Ableitung an der Stelle $t = 0$ das k -te faktorielle Moment $E^*(X,k) = E[X(X-1)\dots(X-k+1)]$ mit der k -ten Ableitung von $\phi_X(t)$

(vgl. Übers. 4.3).

3. Ist X_1 zweipunktverteilt (vgl. Kap. 5) mit der momenterzeugenden Funktion

$$M_X(t) = E(e^{tX}) = e^{0 \cdot t} (1-\pi) + e^{1t} \pi = \pi e^t + (1-\pi)$$

und X_2 identisch verteilt, dann ist die momenterzeugende Funktion der **Summe** $X = X_1 + X_2$ der beiden unabhängig identisch verteilten Zufallsvariablen $[\pi e^t + (1-\pi)]^2$, d.h. nach Satz 4.1 das **Produkt** der momenterzeugenden Funktionen. Das ist aber die momenterz. Funktion der Binomialverteilung mit $n = 2$.

Satz 4.1: Die erzeugende Funktion einer Faltung Z ist das Produkt der erzeugenden Funktionen von X und Y .

So wie mit Satz 4.1 der Nutzen erzeugender Funktionen bei der Betrachtung von Summen unabhängig identisch verteilter Zufallsvariablen (also für das Problem, ob eine Wahrscheinlichkeitsverteilung "reproduktiv" ist) offenbar wird, zeigt sich mit den folgenden Sätzen der Vorzug mit solchen Funktionen Grenzverteilungen (Kap.7) zu untersuchen:

Satz 4.2: (Levy-Cramer; vgl. Bem.6 zu Def. 7.2)

Eine endliche Folge von Verteilungsfunktionen $F_1(x), F_2(x), \dots$ konvergiert genau dann gegen die asymptotische Verteilung $F(x)$ (Grenzverteilung), wenn die Folge der charakteristischen Funktionen auf jedem endlichen Intervall $t_0 \leq t \leq t_1$ gegen die charakteristische Funktion der Grenzverteilung konvergiert.

Satz 4.3: Eine Folge von Wahrscheinlichkeitsverteilungen $f_1(x), f_2(x), \dots$ konvergiert genau dann gegen eine Wahrscheinlichkeitsverteilung $f(x)$ (Grenzverteilung), wenn die Folge der charakteristischen Funktionen auf jedem endlichen Intervall $t_0 \leq t \leq t_1$ gegen die charakteristische Funktion der Grenzverteilung konvergiert.

5. Verteilungen transformierter Variablen, Intervalltransformationen

In diesem Abschnitt wird gezeigt, wie man die Wahrscheinlichkeitsverteilung $f^*(z)$ einer gemäß einer Funktion $Z = g(X)$ transformierten Variable X findet und wie diese Verteilung zusammenhängt mit der Verteilung von X . Das Problem tritt auch auf, wenn es gilt eine für das Intervall (den Definitionsbereich) $a \leq x \leq b$ definierte Zufallsvariable X in eine für das Intervall $[a^*, b^*]$ definierte Zufallsvariable Z zu transformieren. Die Zusammenhänge werden schließlich auf den mehrdimensionalen Fall verallgemeinert.

a) ein- und mehrdimensionale diskrete Zufallsvariable

Gegeben sei die diskrete Wahrscheinlichkeitsverteilung

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

etwa $X_1 = 1$, $X_2 = 2$ usw., und gesucht ist die Wahrscheinlichkeitsverteilung der transformierten Variable Z , etwa $Z = 3X^2 + 4$. Die Lösung ist einfach, weil dann nur jedem konkreten Wert x_i ein z_i gemäß dieser Transformation zuzuordnen ist und die Wahrscheinlichkeiten hiervon nicht berührt werden. Man erhält dann:

$$f(z) = \begin{cases} p_1 & \text{für } z_1 = 7 \\ p_2 & \text{für } z_2 = 16 \text{ usw.} \\ \vdots & \\ 0 & \text{sonst} \end{cases}$$

Bei einer mehrdimensionalen **diskreten** Verteilung ist analog zu verfahren. Schwieriger ist das Problem im Falle einer **stetigen** Zufallsvariable.

b) eindimensionale stetige Zufallsvariable

Satz 4.4: Sei $f(x)$ die Dichtefunktion der stetigen Zufallsvariable X und $Z = g(X)$ eine eindeutige Transformation mit der Umkehrung $X = h(Z)$ und der Ableitung $\frac{dx}{dz} = h'(z)$, dann

lautet die Dichtefunktion von Z :

$$(4.25) \quad f^*(z) = f[h(z)] |h'(z)|$$

In $f(x)$ wird jetzt x gem. $x = h(z)$ eingesetzt und diese Funktion mit der absolut genommenen Ableitung $\frac{dx}{dz} = \frac{dh(z)}{dz} = h'(z)$ multipliziert. Ist x im Intervall $[a, b]$ definiert, so muss für z gelten $a^* \leq z \leq b^*$. Wenn es gilt, das Intervall $[a, b]$ in das Intervall $[a^*, b^*]$ zu transformieren, so lautet die Funktion $g(x)$

$$(4.26) \quad Z = \frac{a^*b - b^*a}{b - a} + \frac{b^* - a^*}{b - a} X \quad \text{mit der Umkehrfunktion } X = h(Z)$$

$$(4.27) \quad X = -\frac{a^*b - b^*a}{b^* - a^*} + \frac{b - a}{b^* - a^*} Z, \text{ so dass } h'(z) = \frac{b - a}{b^* - a^*} \text{ ist.}$$

c) mehrdimensionale stetige Zufallsvariable

Die Verallgemeinerung von Satz 4.4 für zweidimensionale Verteilungen lautet:

Die Variablen (X, Y) haben die gemeinsame Dichte $f(x, y)$ und es seien $u = g_1(x, y)$ und $v = g_2(x, y)$ stetige monotone Transformationen mit den Umkehrungen $x = h_1(u, v)$, $y = h_2(u, v)$, dann gilt:

$$(4.28) \quad f^*(u, v) = f[h_1(u, v), h_2(u, v)] \cdot |J|,$$

wobei J die Jacobische Determinante (Funktionaldeterminante) $J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$ und $|J|$ deren

Betrag ist. Man erkennt, dass die Betrachtung unter b ein Spezialfall ist und wie der Zusammenhang leicht zu verallgemeinern ist bei mehr als zwei Dimensionen.

Kapitel 5: Spezielle diskrete Wahrscheinlichkeitsverteilungen

1.	Übersicht und Einführung	36
2.	Zweipunkt-Z(π), Binomial- $B(n, \pi)$ und hypergeometr. Verteilung $H(n, M, N)$	38
3.	Geometrische Verteilung $GV(\pi)$ und negative Binomialverteilung $NB(\pi, r)$	44
4.	Poissonverteilung $P(\lambda)$	46
5.	Weitere Verteilungen eindimensionaler diskreter Zufallsvariablen	50
6.	Polytome Versuche (mehrdimensionale diskrete Verteilungen)	50

1. Übersicht und Einführung

Def. 5.1: Bernoulli-Experiment, Urnenmodell

- Ein Zufallsexperiment, bei dem nur zwei sich gegenseitig ausschließende Ereignisse eintreten können, heißt **Bernoulli-Experiment**. Der Versuchsausgang ist somit dichotom. Man spricht (ohne Wertung) von "Erfolg" und "Misserfolg". Bei wiederholten Experimenten kann man unabhängige und abhängige Versuche unterscheiden.
- Ein **Urnenmodell** besteht in der Spezifizierung der Zusammensetzung einer Urne und der Art der Ziehung aus der Urne. Die Urne kann aus zwei oder $m > 2$ Arten von Kugeln bestehen und es kann mehrmals mit oder ohne Zurückziehen gezogen werden.

Bemerkungen zu Def. 5.1:

- Beim Bernoulli-Experiment besteht die Urne aus
 - M schwarzen Kugeln ("Erfolg") und
 - $N-M$ weißen Kugeln ("Misserfolg"),

insgesamt also aus N Kugeln. Beim einmaligen Ziehen aus der Urne ist die "Erfolgswahrscheinlichkeit" gegeben durch $\pi = M/N$ und die "Misserfolgswahrscheinlichkeit" mit $1 - \pi = (N - M)/N$. Üblich ist auch die Notation $p = \pi$ für die Erfolgs- und $q = 1 - \pi$ für die Misserfolgswahrscheinlichkeit. Bei $n \geq 2$ -maligem Ziehen aus der Urne ist zu unterscheiden zwischen

- Ziehen mit Zurücklegen (ZmZ, unabhängige Versuche): durch das Zurücklegen wird die Urne praktisch unendlich, so dass N nicht zu beachten ist,
 - Ziehen ohne Zurücklegen (ZoZ, abhängige Versuche)
2. Eine andere Veranschaulichung wiederholter Zufallsversuche mit polytomen und speziell dichotomen (Bernoulli-Experiment) Ausgang ist ein **Baumdiagramm** (Wahrscheinlichkeitsbaum).
 3. Die Zufallsvariable (ZV) kann im Folgenden unterschiedlich definiert sein:
 - der **Anzahl** X der **Erfolge** bei n Versuchen (X ist die ZV, n ist keine ZV)
 - der **Anteil** $p = X/n$ der **Erfolge** bei n Versuchen (Relativierte Verteilungen, p ist eine Lineartransformation von X)
 - die Anzahl X der **nicht erfolgreichen Versuche** bis zum r -ten Mal (oder speziell $r = 1$ ten Mal) ein Erfolg auftritt
 - oder die Anzahl X^* der **Versuche**.

Gerade hinsichtlich der letzten Betrachtung gibt es Unterschiede bei der Darstellung einer Verteilung in den Lehrbüchern, weshalb (wenn nötig) "alternative Formulierung" der Zufallsvariable aufgeführt werden.

4. Eine Variante des Urnenmodells besteht im Ziehen aus k Urnen mit einem Anteil π_i schwarzer Kugeln. Das führt zur verallgemeinerten Binomialverteilung (von Poisson). Ein Modell, das ein Hinzufügen von Kugeln nach Ziehung vorsieht, führt zur Polya-Verteilung, einer Verallgemeinerung von Binomial- und Hypergeometrischer Verteilung.

Eigenschaften der Verteilungen

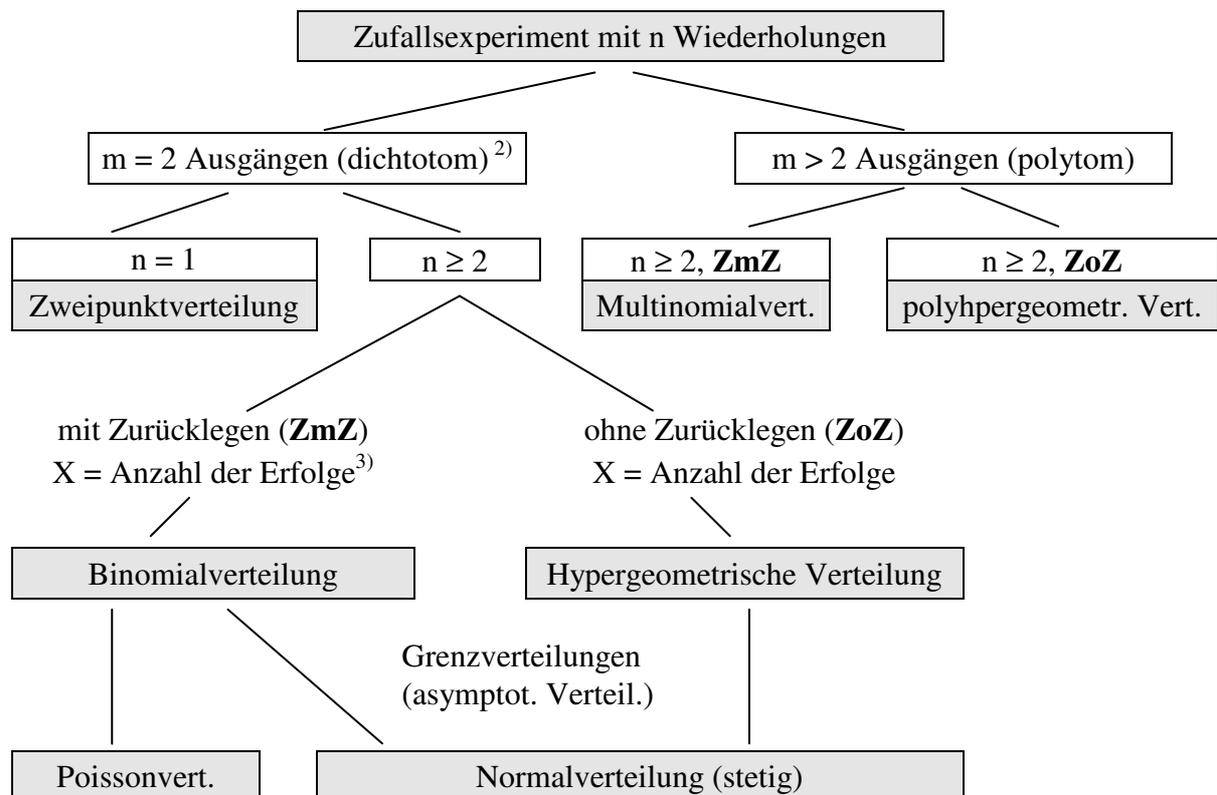
(Die folgenden Bemerkungen gelten auch für stetige Verteilungen), Im Zusammenhang mit Wahrscheinlichkeitsverteilungen interessieren i. d. R. die folgenden Eigenschaften einer Verteilung:

1. Parameter (die die Gestalt der Verteilung bestimmen) und die hierbei zulässigen Wertebereiche
2. Interpretation der Zufallsvariable X und deren zulässiger Wertebereich
3. Momente der Verteilung sowie Median, Modus etc.
4. Erzeugende Funktionen, die u. a. auch Aufschluß geben über die Ziffern 5 und 6
5. Reproduktivität (vgl. Def. 5.2)
6. Zusammenhänge mit anderen Verteilungen
 - eine Verteilung V_1 kann z.B. ein Spezialfall einer anderen allgemeineren Verteilung V_2 sein (etwa bei einer bestimmten Parameterkonstellation von V_2)
 - Approximationsmöglichkeiten (vgl. Def. 5.3).

Def. 5.2: Reproduktivität

Sind die Zufallsvariablen X_1, X_2, \dots, X_n verteilt nach einer bestimmten Verteilung V und ist die Summe unabhängiger Zufallsvariablen $X_1 + X_2 + \dots + X_n$ ebenfalls nach V verteilt, so ist die Verteilung V reproduktiv.

Übersicht 5.0: Urnenmodell von einigen¹⁾ Verteilungen



1) wenn nichts anderes vermerkt ist: **diskrete** Verteilungen

2) Bernoulli-Experiment

3) Andere **Fragestellungen**: GV, NB

GV (geometr. Vert.): Wie groß ist bei unabhängigen Bernoulli-Experimenten die Wahrscheinlichkeit, dass nach $X = 0, 1, 2, \dots$ Mißerfolgen erstmals ein Erfolg auftritt? Die Zufallsvariable X ist die Anzahl der Mißerfolge in einer Folge von Mißerfolgen bis zum ersten und i. d. R. einzigen Erfolg. Die Anzahl der Versuche ist dann $X + 1$.

NB (negative Binomialvert.): $f(x)$ ist die Wahrscheinlichkeit dafür, dass der r -te Erfolg gerade im $(x + r)$ ten Versuch eintritt. Offenbar ist GV der Spezialfall $r = 1$.

Def. 5.3: Approximation

Eine Folge von Verteilungen des gleichen Typs V_1, V_2, \dots, V_n , die sich durch die für die Parameter angenommenen Zahlenwerte unterscheiden, kann gegen eine Grenzverteilung G konvergieren, so dass es möglich ist, Wahrscheinlichkeiten nach V in guter Näherung durch meist leichter zu bestimmende Wahrscheinlichkeiten nach G zu approximieren.

2. Zweipunkt-, Binomial- und hypergeometrische Verteilung

a) Zweipunktverteilung [$Z(\pi) = B(1, \pi)$]

Bei einmaliger Durchführung eines Bernoulli-Experiments kann $x = 0$ (Mißerfolg) mit Wahrscheinlichkeit $1 - \pi$ oder $x = 1$ (Erfolg) mit Wahrscheinlichkeit π auftreten. Man spricht

von einer Zweipunktverteilung, weil die Zufallsvariable zwei Werte, x_1 und x_2 annehmen kann, in diesem speziellen Fall $x_1 = 0$ und $x_2 = 1$.

Es gilt $E(X) = \pi$ und $V(X) = \pi(1-\pi)$

Bemerkenswert ist dass (wie Abb. 5.1 zeigt) die Varianz $\sigma^2 = V(X)$ im Betrag beschränkt ist. Sie beträgt

$$V(X) = \pi(1-\pi) \leq \frac{1}{4} \quad (0 \leq \pi \leq 1)$$

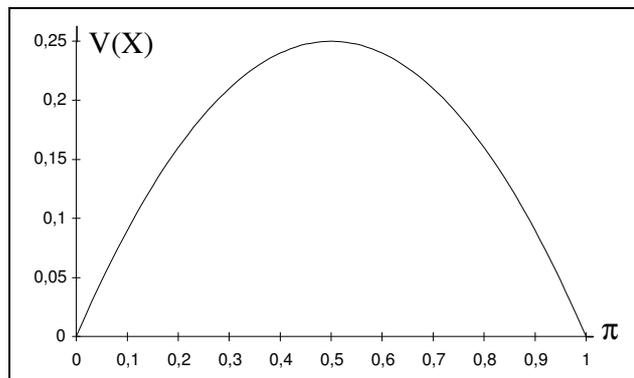
und nimmt ihren maximalen Wert an der Stelle $\pi = 1/2$ an.

Für die Momente der $Z(\pi)$ -Verteilung erhält man ganz allgemein:

$$E(X) = \mu = 0 \cdot (1-\pi) + 1 \cdot \pi = \pi \quad \text{und} \quad E(X^2) = 0^2 \cdot (1-\pi) + 1^2 \cdot \pi = \pi, \text{ allgemein } E(X^k) = \pi.$$

Dass alle Anfangsmomente den Wert π annehmen ergibt sich auch aus den Ableitungen der momentenerzeugenden Funktion $M_X(t)$. Die Verteilung ist linkssteil ($\gamma > 0$), wenn $x < 1/2$ und rechtssteil ($\gamma < 0$), wenn $\pi > 1/2$ (also die Wahrscheinlichkeit des Erfolges größer ist als die des Misserfolges).

Abb 5.1:



Übersicht 5.1: Zweipunktverteilung $Z(\pi) = B(1, \pi)$

Wahrscheinlichkeitsfunktion	$f_Z(x) = \begin{cases} 1-\pi & \text{für } x_1 = 0 \\ \pi & \text{für } x_2 = 1 \\ 0 & \text{sonst} \end{cases}$
Verteilungsfunktion	$F_Z(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1-\pi & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1 \end{cases}$
Parameter	π (Erfolgswahrscheinlichkeit)
Zufallsvariable X	Anzahl der Erfolge bei $n=1$ Versuchen. Realisationen $x_1=0$ und $x_2=1$
Urnenmodell	Bernoulli-Experiment
Summe identisch verteilter Zufallsvariablen	$X_1 + X_2 + \dots + X_n \sim B(n, \pi)$
Momente	Erwartungswert $\mu = E(X) = E(X^2) = \dots = E(X^n) = \pi$, Varianz $\sigma^2 = V(X) = \pi(1-\pi)$, Schiefe $\gamma = (1-2\pi)/\sqrt{\pi(1-\pi)}$
andere Verteilungen	$Z(\pi) = B(1, \pi)$
erzeugende Funktionen	$W_x(t) = \pi^t + (1-\pi)$, $M_x(t) = \pi e^t + (1-\pi)$, $\Psi_x(t) = \pi e^{it} + (1-\pi)$
Bedeutung	Modell für die Grundgesamtheit bei homograde Fragestellung

b) Binomialverteilung [$B(n, \pi)$]

Eine n -malige unabhängige Wiederholung bei konstantem π des Bernoulli-Experiments (n mal ZmZ aus einer Urne) führt zur Binomialverteilung. Die Anzahl X der Erfolge kann Werte zwischen 0 und n annehmen. Sind die unabhängigen Zufallsvariablen X_1, \dots, X_n identisch zweipunktverteilt mit π , so ist deren Summe $X = X_1 + \dots + X_n$ binomialverteilt mit den Parametern n und π . Daraus folgt unter anderem auch $E(X) = \sum E(X_i) = \sum \pi = n\pi$. Die B-Verteilung ist die Stichprobenverteilung für die Anzahl der Erfolge bei (ZmZ-) Stichproben vom Umfang n aus einer $Z(\pi)$ -verteilten Grundgesamtheit.

Der **Begriff** Binomialverteilung (oder: binomische Verteilung) nimmt Bezug darauf, dass die Entwicklung (Expansion) des Binoms $(q + p)^n$ zu folgender Gleichung (binomischer Satz) führt:

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

worin das allgemeine Glied $\binom{n}{x} p^x q^{n-x}$ die Wahrscheinlichkeitsfunktion der B-Verteilung mit $p = \pi$ und $q = 1 - \pi$ darstellt. Daraus folgt auch (da $p + q = 1$), dass die Summe der binomialen Wahrscheinlichkeiten $\sum f_B(x) = 1$ ist. Es gibt $\binom{n}{x}$ Möglichkeiten in einer Versuchsserie (n Versuche) x Erfolge und damit $n - x$ Misserfolge zu erhalten. Jede dieser Möglichkeiten hat die Wahrscheinlichkeit $\pi^x (1 - \pi)^{n-x}$. Das erklärt die Formel für $f(x)$.

Wie bei der Z-Verteilung ist die **Schiefe** $\gamma > 0$ (linkssteil), wenn $\pi < \frac{1}{2}$, also $\pi < 1 - \pi$. Auch

aus $f_B(x) = \binom{n}{x} (1 - \pi)^n$ folgt, dass $\frac{\pi}{1 - \pi} \neq 1$ Asymmetrie impliziert.

Aus $f_B(x|n, \pi) = f_B(n - x|n, 1 - \pi)$ folgt, dass x und $n - x$ sowie π und $1 - \pi$ vertauscht werden können. So genügt es, die Wahrscheinlichkeiten der B-Verteilung bei gegebenen n für $0 \leq \pi \leq 1/2$ zu tabellieren (**vgl. Tabelle im Anhang**, Seite T-1).

Aus der **Rekursionsformel** (vgl. Übers. 5.8) $v = \frac{f_B(x+1)}{f_B(x)} = \frac{n-x}{x+1} \cdot \frac{\pi}{1-\pi}$ folgt, dass die

Wahrscheinlichkeiten der B-Verteilung so lange ansteigen (also $v > 1$), wie x kleiner ist als $\mu - (1 - \pi) = (n + 1) \pi - 1$ ist und dass sie fallen ($v < 1$), wenn $x > (n + 1) \pi - 1$. Die B-Verteilung hat, falls $(n + 1) \pi$ ganzzahlig ist, zwei Modalwerte, nämlich $(n + 1) \pi - 1$ und $(n + 1) \pi$, anderenfalls ist der Modus $[(n + 1) \pi]$ (Gaußklammer []).

Für die faktoriellen Momente gilt:

$$E^*(X, k) = n(n-1) \dots (n-k+1) \pi^k, \text{ also } E^*(X, 1) = E(X) = n\pi \text{ und}$$

$$E^*(X, 2) = E[X(X-1)] = n(n-1)\pi^2.$$

Folglich erhält man für die Varianz: $\sigma^2 = V(X) = n \cdot \pi(1 - \pi)$.

c) Weitere Bemerkungen zur Binomialverteilung

1. **Laplace-Verteilung**, Spezialfall der Binomialverteilung für $\pi = 1 - \pi = 1/2$

(Der Begriff wird auch für eine stetige Verteilung benutzt.)

$$f(x|n) = \binom{n}{x} \frac{1}{2^n}, \quad \mu = \frac{1}{2} n, \quad \sigma = \frac{1}{2} \sqrt{n}.$$

Veranschaulichung: Galtonisches Brett, römischer Brunnen.

2. **Verallgemeinerte Binomialverteilung** $vB(n, p_1, p_2, \dots, p_n)$

Erfolgswahrscheinlichkeiten sind nicht konstant, sondern $\pi_i = p_i$ und $1 - \pi_i = q_i$ (mit $i = 1, \dots, n$). Man erhält vB durch Expansion des Produkts

$$(p_1 + q_1)(p_2 + q_2) \dots (p_n + q_n).$$

Erwartungswert: $\mu = \sum p_i = \sum \pi_i$ statt $n \cdot \pi$, Varianz: $\sigma^2 = \sum p_i q_i$.

Binomialverteilung als Spezialfall, $p_1 = \dots = p_n = \pi$.

3. **Relativierte Binomialverteilung** $rB(n, \pi)$

Ist X binomialverteilt mit $f_B = (x|n, \pi)$, dann ist der Anteilswert $p = \frac{X}{n}$ (eine Lineartransformation von X) relativiert binomialverteilt mit dem Erwartungswert $E(p) = \pi$ und der Varianz $V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot V(X) = \frac{\pi(1-\pi)}{n}$ sowie der (in Kapitel 8 eine große Rolle

spielenden) Standardabweichung $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ (vgl. Abb. 5.2).

Die rB -Verteilung ist die Stichprobenverteilung für den **Anteil** der Erfolge (die B -Verteilung für die **Anzahl** der Erfolge) bei Stichproben **mit** Zurücklegen vom Umfang n .

Übers. 5.2 nächste Seite

d) Hypergeometrische Verteilung [$H(n, M, N)$]

1. Herleitung

Wird bei dem in Definition 5.1 beschriebenen Urnenmodell n mal **ohne** Zurücklegen gezogen, so ist die Wahrscheinlichkeit für das Auftreten von x Erfolgen nach dem klassischen Wahrscheinlichkeitsbegriff.

$$f_H(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, \quad 0 \leq x \leq \min(n, M), \quad 0 \leq M \leq N.$$

Wegen Gl. 2.7d ist

Übersicht 5.2: Binomialverteilung $B(n, \pi)$ (Bernoulli-Verteilung)*)

Wahrscheinlichkeitsfunktion	$f_B(x) = \begin{cases} \binom{n}{x} \pi^x (1-\pi)^{n-x} & x = 0, 1, \dots, n \\ 0 & \text{sonst} \end{cases}$
Verteilungsfunktion	$F_B(x) = \sum_{v=0}^x \binom{n}{v} \pi^v (1-\pi)^{n-v}$
Parameter	n, π
Zufallsvar. X	Anzahl der Erfolge bei n Versuchen ($0 \leq x \leq n$)
Urnenmodell	Bernoulli-Experiment, n-mal Ziehen mit Zurücklegen (ZmZ)
Reproduktivität	Bei konstantem π ist die Summe unabhängig binomialverteilter Zufallsvariablen ebenfalls binomialverteilt
Momente	$\mu = E(X) = n \cdot \pi, \sigma^2 = V(X) = n \cdot \pi(1-\pi), \gamma = \frac{1-2\pi}{\sqrt{n\pi(1-\pi)}}$
andere Verteilungen	relativierte Binomialverteilung Normalverteilung und Poissonverteilung als Grenzverteilung Betaverteilung 1. Art
erzeugende Funktionen	$W_x(t) = [\pi t + (1-\pi)]^n, M_x(t) = [\pi e^t + (1-\pi)]^n, \Psi_x(t) = [\pi e^{it} + (1-\pi)]^n$
Bedeutung	Stichprobenverteilung (homograde Theorie) z.B. Anzahl der Ausschussstücke in der statistischen Qualitätskontrolle

$$\sum_{x=0}^n \binom{M}{x} \binom{N-M}{n-x} = \binom{N}{n}, \text{ so daß } \sum f_H(x) = 1.$$

Bei einer Grundgesamtheit vom Umfang N sind

- $\binom{N}{n}$ Stichproben vom Umfang n durch ZoZ und
- N^n Stichproben durch ZmZ

möglich und gleichwahrscheinlich. Von den Stichproben o.Z. sind $\binom{M}{x} \binom{N-M}{n-x}$ von der Art,

dass sowohl x Erfolge als auch n - x Mißerfolge "gezogen worden sind" ("günstige Fälle"). Bei dieser Betrachtungsweise wird z.B. jede der M schwarzen Kugeln als eine andere Kugel als die übrigen schwarzen Kugeln angesehen.

2. Finite multiplier (Endlichkeitskorrektur)

Die Varianz der H-Verteilung ist mit $V(X) = n\pi(1-\pi) \frac{(N-n)}{(N-1)} \leq n\pi(1-\pi)$ kleiner (bei endlichem N) als die Varianz der B-Verteilung. Der Faktor $\frac{(N-n)}{(N-1)} \approx 1 - n/N$ heißt Endlichkeitskorrektur und strebt mit $N \rightarrow \infty$ gegen 1. Er ist eine Funktion des Auswahlsatzes n/N und gilt als vernachlässigbar, wenn $n/N < 0,05$.

*) Der Begriff wird auch gebraucht für die Zweipunktverteilung (bzw. dem speziellen Fall $x_1=0, x_2=1$ der Z-Verteilung).

Bei Totalerhebung ist $V(X) = 0$ und wegen $N = n$ gilt:

$$f_H(x) = \begin{cases} 1 & \text{wenn } x = M \\ 0 & \text{sonst} \end{cases}$$

was intuitiv verständlich ist.

3. Relativierte hypergeometrische Verteilung $rH(n, M, N)$

Ist die Anzahl X der Erfolge hypergeometrisch verteilt, dann ist der Anteil der Erfolge $p = X/n$ relativiert hypergeometrisch verteilt mit dem Erwartungswert

$$E(p) = \pi \text{ und der Varianz } V(p) = \frac{\pi(1-\pi)}{n} \cdot \frac{N-n}{N-1} = \frac{1}{n^2} V(X).$$

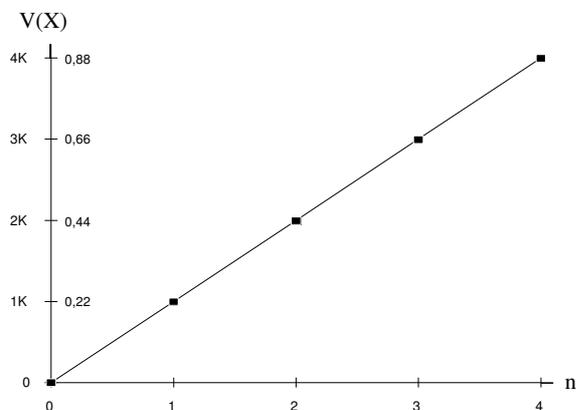
Die rH -Verteilung ist die Stichprobenverteilung für den **Anteil** der Erfolge (die H -Verteilung für die **Anzahl** der Erfolge) bei Stichproben **ohne** Zurücklegen vom Umfang n aus einer Grundgesamtheit vom Umfang N .

Die Varianz von p nimmt mit wachsendem N ab und ist bei $N = 1$ gleich der Varianz von X (der hypergeometrischen Verteilung) und bei $n = N$ ist sie Null (vgl. Abb. 5.2).

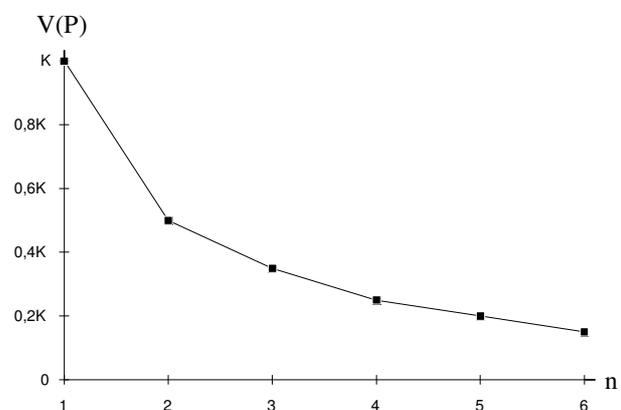
Abb. 5.2: Varianzen der Binomialverteilung und der Hypergeometrischen Verteilung in Abhängigkeit von n

Rechenbeispiel: $\pi = \frac{1}{3}$ und $N = 6$, $\pi(1-\pi) = \frac{2}{9} = 0,222 = K$.

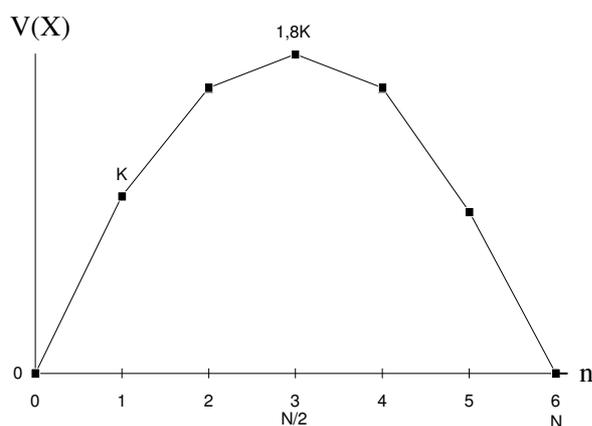
a) Binomialverteilung



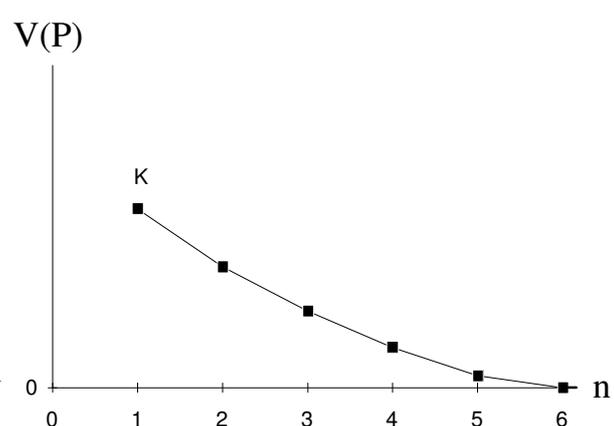
b) Relativierte Binomialv. (anderer Maßstab)



c) Hypergeometrische Verteilung



d) Relativierte hypergeometrische Verteilung



Zahlenangaben zu Abb. c und d

n	0	1	2	3	4	5	6
V(X)	0	K	1,6K < 2K	1,8K < 3K	1,6K < 4K	K < 5K	0 < 6K
V(p)	∞		0,4 K < K/2	0,2 K < K/3	0,1 K < K/4	0,04 K < K/5	0 < K/6

Bei der hypergeometrischen Verteilung ergeben sich also jeweils kleinere Varianzen als bei der Binomialverteilung.

Übersicht 5.3: Hypergeometrische Verteilung $H(n, M, N)$

Wahrscheinlichkeitsfunktion	$f_H(x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \quad \text{für } x = 0, 1, \dots$
Symmetrie	$f(x n, M, N) = f(x M, n, N)$, die Parameter M und n sind vertauschbar
Parameter	n, M, N mit $0 \leq n \leq N$ und $0 \leq M \leq N$ $\left(\pi = \frac{M}{N} \right)$
Zufallsvariable X	X ist die Anzahl der Erfolge bei n mal Ziehen ohne Zurücklegen (ZoZ) mit $\max(0, n - (N - M)) \leq x \leq \min(n, M)$
Urnenmodell	wie B(n, π) aber ZoZ statt ZmZ
Momente ^{*)}	$\mu = E(X) = n \cdot \pi$, $\sigma^2 = V(X) = n \cdot \pi(1 - \pi) \frac{N - n}{N - 1}$, $\gamma = \frac{[(N - 2n)(1 - 2\pi)]}{[(N - 2)\sigma]}$, $E^*(X, k) = \frac{M!n!}{N!} \cdot \frac{(N - k)!}{(M - k)! \cdot (n - k)!}$
andere Verteil.	relativierte Hypergeometrische Verteilung
Grenzübergang	Mit $N \rightarrow \infty$ bzw. $n \ll N$ strebt HV gegen die. B(n, π)-Verteilung
Bedeutung	Stichprobenverteilung homograde Fall bei Ziehen ohne Zurücklegen

3. Geometrische Verteilung und negative Binomialverteilung

a) Wahrscheinlichkeiten für Versuchsfolgen

1. Folge von x Misserfolgen

Urnenmodell (analog: Roulette): die Wahrscheinlichkeit für eine erste schwarze Kugel (Erfolg) nach x roten Kugeln (Misserfolg) bei konstanter Erfolgswahrscheinlichkeit π (also ZmZ) ist $(1 - \pi)^x \pi = f_{GV}(x)$ (Wahrscheinlichkeitsverteilung der geometr. Verteilung GV).

Bedeutung (Interpretation) von X: Wartezeit bis zum Eintritt des Erfolges.

^{*)}Bei übereinstimmenden n und p haben Binomial- und hypergeometrische Verteilung den gleichen Erwartungswert, die Varianzen unterscheiden sich um den Korrekturfaktor (finite multiplier, Endlichkeitskorrektur) $(N-n)/(N-1)$, der für $N \rightarrow \infty$ gegen 1 strebt.

2. Folge von x Misserfolgen und r Erfolgen

Die Wahrscheinlichkeit dafür, dass in einer Serie von $m-1$ Versuchen $r-1$ Erfolge auftreten und x Mißerfolge (so dass $m=x+r$) ist: $\binom{m-1}{r-1} p^{r-1} q^x$, wenn $p = \pi$ die konstante

Erfolgswahrscheinlichkeit ist und $q=1-\pi$. Sollte die Serie von m Versuchen abschließen mit einem Erfolg (dem r -ten Erfolg) so ist die Wahrscheinlichkeit hierfür:

$$f_{\text{NB}}(x) = \binom{m-1}{r-1} p^r q^x \quad \text{gemäß der NB-Verteilung (= negative Binomialverteilung).}$$

Für $r=1$ erhält man wegen $m=x+1$ hierfür $f_{\text{NB}}(x) = \binom{x}{0} p q^x = f_{\text{GV}}(x)$, die Wahrscheinlichkeitsfunktion der GV als Spezialfall.

b) alternative Formulierung für GV und NBV

Die Wahrscheinlichkeit des ersten Erfolges nach $x^* - 1$ Mißerfolgen, also im x^* -ten Versuch ist $f_{\text{GV}}(x^*) = \pi(1-\pi)^{x^*-1}$ (vgl. Übersicht 5.4a). X^* ist die Anzahl der **Versuche**, X dagegen die Anzahl der **Mißerfolge** und es gilt $X^* = X + 1$ (Lineartransformation). Zählt man wie hier als Zufallsvariable die Anzahl der Versuche, so besteht ein Zusammenhang zwischen der (so definierten) GV und der Exponentialverteilung- und Poissonverteilung. Entsprechend ist X^* in der alternativen Formulierung der negativen Binomialverteilung (Übers. 5.5 a) definiert als Anzahl der Versuche bis zum r -ten Erfolg ausschließlich (d.h. mit dem r -ten Erfolg im X^* -ten Versuch).

Übersicht 5.4: Geometrische Verteilung $GV(\pi) = NB(1, \pi)$

Wahrscheinlichkeitsfunktion $p = \pi, q = 1 - \pi$	$f_{\text{GV}}(x) = \pi(1 - \pi)^x = p q^x$ Wahrscheinlichkeiten bilden eine geometrische Reihe
Verteilungsfunktion	$F_{\text{GV}}(x) = 1 - (1 - \pi)^{x+1} = 1 - q^{x+1}$
Parameter	$\pi = p$ (Erfolgswahrscheinlichkeit)
Zufallsvariable X (Mißerfolge)	Anzahl der Misserfolge bis zum ersten Erfolg ($X^* = X + 1 =$ Anzahl der Versuche), $x \in \mathbb{N}$
Urnenmodell	wie $B(n, \pi)$; Bernoulli-Experiment beliebig oft wiederholt (ZmZ)
Summe ident. vert. ZVn	$X_1 + \dots + X_r \sim NB(r, \pi)$
Momente	$\mu = E(X) = \frac{(1 - \pi)}{\pi} = \frac{q}{p}$, $\sigma^2 = V(X) = \frac{(1 - \pi)}{\pi^2} = \frac{q}{p^2}$ $\gamma = \frac{(1 + q)}{\sqrt{q}}$, $E^*(X, k) = k \left(\frac{q}{p} \right)^k$
andere Verteilungen	Spezialfall von NB wenn $r = 1$. Für $\pi \leq 0,1$ ist $X^* = X + 1$ näherungsweise exponentialverteilt mit $\lambda = \pi$
erzeugende Funktionen:	$W_x(t) = \frac{\pi}{1 - (1 - \pi)t} = \frac{p}{1 - qt}$, $M_x(t) = \frac{\pi}{1 - (1 - \pi)e^t} = \frac{p}{1 - qe^t}$ $\Psi_x(t) = \frac{\pi}{1 - (1 - \pi)e^{it}} = \frac{p}{1 - qe^{it}}$
Bedeutung	Wartezeiten

Übersicht 5.4a: Geometrische Verteilung (alternative Formulierung) $GV^*(\pi)$

Wahrscheinlichkeitsfunktion	$f_{GV^*}(x^*) = \pi(1-\pi)^{x^*-1} = \frac{p}{q} q^{x^*}$
Verteilungsfunktion	$F_{GV^*}(x^*) = 1 - (1-\pi)^{x^*} = 1 - q^{x^*}$
Zufallsvariable $X^*=X+1$	Anzahl der Versuche bis zum ersten Erfolg
Momente	$\mu^* = E(X^*) = \frac{1}{\pi}$, $V(X^*) = \frac{(1-\pi)}{\pi^2} = V(X)$

Übersicht 5.5: Negative Binomialverteilung $NB(\pi, r)$ (Pascal-Verteilung)

Wahrscheinlichkeitsfkt. ¹⁾ ($p = \pi$, $q = 1-\pi$)	$f_{NB}(x) = \binom{-r}{x} p^r (-q)^x = \binom{x+r-1}{r-1} p^r q^x$ oder mit $m = x+r$ $f_{NB}(x) = \binom{m-1}{r-1} p^r q^{m-r} = \binom{m-1}{r-1} p^r q^x$
Parameter	$\pi = p$ und r ($r = 1, 2, \dots$)
Zufallsvariable $X^2)$	Anzahl der Mißerfolge bei $m=x+r$ Versuchen und r Erfolgen $x \in \mathbb{N}$
Urnenmodell	wie $B(n, \pi) = B(n, p)$
Reproduktivität	$X_1 \sim NB(r_1, \pi)$ und $X_2 \sim NB(r_2, \pi)$, dann $(X_1 + X_2) \sim NB(r_1 + r_2, \pi)$
Momente	$\mu = E(X) = \frac{rq}{p}$, $\sigma^2 = V(X) = \frac{rq}{p^2}$, $\gamma = \frac{(1+q)}{\sqrt{rq}}$, $E(X^2) = \frac{rq(1+rq)}{p^2}$
andere Verteilungen	Spezialfall von GV wenn $r = 1$. Wenn $r \rightarrow \infty$, $q \rightarrow 0$ und $rq/p = \lambda$ geht NB in Poissonverteilung $P(\lambda)$ über
erzeugende Funktionen:	$W_x(t) = \left(\frac{p}{1-qt}\right)^r$, $M_x(t) = \left(\frac{p}{1-qe^{it}}\right)^r$, $\Psi_x(t) = \left(\frac{p}{1-qe^{it}}\right)^r$

1) Die erste Art, die Wahrscheinlichkeitsfunktion darzustellen erklärt den Namen "negative Binomialverteilung".

2) Wegen X ist auch $M=X+r$ [mit der Realisation m] eine ZV.

Übersicht 5.5a: Negative Binomialverteilung (alternative Formulierung) $NB^*(r, \pi)$

Wahrscheinlichkeitsfunktion	$f_{NB^*}(x^*) = \binom{x^*-1}{r-1} p^r q^{x^*-r}$
Momente	$E(X^*) = \frac{r}{p} = E(X) + r$ $V(X^*) = V(X) = rq/p$

4. Poissonverteilung ($P(\lambda)$)

a) Eigenschaften und Bedeutung

Diskrete Verteilung ($X=0, 1, 2, \dots$) für "seltene Ereignisse" (Erfolgswahrscheinlichkeit π gering). Gleiches Urnenmodell und gleiche Fragestellung wie $B(n, \pi)$. Nur ein Parameter λ , der zugleich μ und σ^2 darstellt. Man gewinnt die $P(\lambda)$ -Verteilung als asymptotische

Verteilung einer Folge von Binomialverteilungen $B_1(n_1, \pi_1), B_2(n_2, \pi_2), \dots$, wenn $n_j > n_{j-1}$ und $\pi_j < \pi_{j-1}$ und $n_j \pi_j = n_{j-1} \pi_{j-1} = \lambda = \text{const.}$. Dass sich so viele empirische Häufigkeitsverteilungen in guter Näherung durch die $P(\lambda)$ -Verteilung darstellen lassen, ist mit dem Poisson-Prozess zu erklären. Die Anzahl X der Erfolge ist naturgemäß diskret, die Wartezeit zwischen den "Erfolgen" (z.B. den Ausfällen einer Maschine) kann diskret (Anzahl der Einheitsintervalle) oder stetig gemessen werden. Ist X poissonverteilt, so ist die Wartezeit geometrisch verteilt (GV^* , Zeit diskret) bzw. exponentialverteilt ($E(\lambda)$, Zeit stetig). Die Poissonverteilung ist reproduktiv. Ferner gilt: Ist $X \sim P(\lambda)$ dann ist mit der Konstanten c $cX \sim P(c\lambda)$.

b) Herleitung

1) Als Grenzverteilung der Binomialverteilung:

Bei konstantem $\lambda = n\pi$ geht die Binomialverteilung für $n \rightarrow \infty$ (und somit $\pi \rightarrow 0$) in die Poissonverteilung über.

2) Poisson-Prozess (zeitlich):

Die Häufigkeit des Eintretens von X Ereignissen in einem vorgegebenen Zeitintervall T ist (angenähert) poissonverteilt mit λ als mittlere Ereignishäufigkeit, wenn

- T in Intervalle der Länge Δt zerlegt wird und in jedem Intervall ein Versuch stattfindet, wobei $X = 0$ oder $X = 1$ Erfolg möglich ist mit den Wahrscheinlichkeiten $P_0(t) = 1 - \lambda t$ und $P_1(t) = \lambda t$. Die Wahrscheinlichkeit des Auftretens eines Erfolges ist proportional zur Länge t des Intervalls und nur von der Länge, nicht von der Lage des Intervalls abhängig (Stationarität), und

- die Versuchsausgänge in den Intervallen unabhängig sind, d.h. die Wahrscheinlichkeit für 0 Erfolge im Intervall $t + \Delta t$ ist

$$P_0(t + \Delta t) = P_0(t) \cdot P_0(\Delta t) \text{ oder bei } X = 1 \text{ Erfolg erhält man:}$$

$$P_1(t + \Delta t) = P_0(t) \cdot P_1(\Delta t) + P_1(t) \cdot P_0(\Delta t) \text{ usw.}$$

Mit $\Delta t \rightarrow 0$ erhält man Differentialgleichungen und für den allgemeinen Fall die Gleichung

$$P'_x(t) + \lambda P_x(t) = \lambda P_{x-1}(t) \text{ deren Lösung } P_x(t) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \text{ ist.}$$

Das ist aber beim Einheitsintervall $t = 1$ die Wahrscheinlichkeitsfunktion der Poissonverteilung. Im Poissonprozeß ist die Wahrscheinlichkeit des Auftretens eines Ereignisses proportional zur Länge des Zeitintervalls und λ ist der Proportionalitätsfaktor.

3) Einfache Erklärung für $f_p(x=0)$:

Für $x=0$ Erfolge gilt, wenn das Einheitsintervall in n sich nicht überschneidende Teilintervalle gleicher Länge $1/n$ aufgeteilt wird, wobei in jedem Intervall das Ereignis mit der Wahrscheinlichkeit $1 - \pi = 1 - \lambda/n$ **nicht** eintritt, so tritt das Ereignis im ganzen Zeitraum mit der Wahrscheinlichkeit $(1 - \pi)^n$ nicht auf.

Das Zahlenergebnis ist ähnlich $f_p(0) = e^{-\lambda}$, weil $\lim_{n \rightarrow \infty} (1 - \lambda/n)^n = e^{-\lambda}$, wenn n hinreichend groß ist (Aufgabe 5.4.4 bis 6)

c) Zusammenhang mit der Exponentialverteilung

Ist die Wahrscheinlichkeit für das Nichteintreten eines Erfolges in jedem Teilintervall $1 - \pi$, so ist die Wahrscheinlichkeit des Nichtauftretens in xn Intervallen der Länge $1/n$ also in einem Gesamtintervall der Länge x , unter den Voraussetzungen des Poissonprozesses:

$(1 - \lambda/n)^{nx} = P(X \geq x)$ nach der GV*. Wegen $\lim_{n \rightarrow \infty} (1 - \lambda/n)^{nx} = e^{-\lambda x}$ ist die Funktion

$F(x) = 1 - P(X \geq x) = 1 - e^{-\lambda x}$, was aber die Verteilungsfunktion der $E(\lambda)$ -Verteilung ist.

Übersicht 5.6: Poissonverteilung $P(\lambda)$

Wahrscheinlichkeitsfunktion	$f_P(x) = \frac{\lambda^x}{x!} e^{-\lambda}, x \in \mathbb{N}$
Verteilungsfunktion	$F(x) = \sum_{v=0}^x f(v)$
Parameter	$\lambda > 0$
Urnenmodell	wie $B(n, \pi)$ (Grenzverteilung von $B(n, \pi)$ bei $n \rightarrow \infty, \pi \rightarrow 0$ und $\lambda = n\pi = \text{const.}$)
Reproduktivität	Die Summe unabhängig poissonverteilter ZVn mit $\lambda_1, \lambda_2, \dots, \lambda_k$ ist wieder poissonvert. mit $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_k$
Momente	$\mu = E(X) = \lambda, E^*(X, k) = \lambda^k, \sigma^2 = V(X) = \lambda, \gamma = \frac{1}{\sqrt{\lambda}} > 0$
erzeugende Funktionen	$W_x(t) = \exp[\lambda(t-1)], M_x(t) = \exp[\lambda(e^t-1)], \psi_x(t) = \exp[\lambda(e^{it}-1)]$

Übersicht 5.7: Anpassung einer Verteilung an eine andere Verteilung (verschiedene, in der Literatur erwähnte Faustregeln) Stochastische Konvergenz von Verteilungen

Verteilung	kann approximiert werden mit der	wenn
Binomialvert. (B)	Poissonvert. (P)	$n \rightarrow \infty, \pi \rightarrow 0, n\pi = \lambda = \text{const.};$ gute Approximation bereits für $n \geq 100, \pi \leq 0,05$ oder $n\pi \leq 5, n \geq 50, \pi < 0,1$
	Normalvert. (N)	$n\pi(1-\pi) \geq 9$
hypergeom. Vert.(H)	Binomialvert (B)	$N \geq 2000, \frac{n}{N} \leq 0,01$ oder $\pi, (1-\pi) < 0,05$
	Poissonvert. (P)	$\frac{n}{N} < 0,1, N$ groß, $\pi = \frac{M}{N}$ klein
	Normalvert. (N)	$n\pi \geq 4$
Poissonvert. (P)	Normalvert. (N)	$\lambda \geq 9$

Übersicht 5.8: Rekursionsformeln

a) Binomialverteilung	
$X \rightarrow X+1$	$f(x+1 n, \pi) = \frac{n-x}{x+1} \cdot \frac{\pi}{1-\pi} \cdot f(x n, \pi)$
b) Poissonverteilung	
$X \rightarrow X+1$	$f(x+1 \lambda) = \frac{\lambda}{x+1} \cdot f(x \lambda)$
c) Hypergeometrische Verteilung	
$X \rightarrow X+1$	$f(x+1 n, M, N) = \frac{n-x}{x+1} \cdot \frac{M-x}{(N-n)-(M-x)+1} \cdot f(x n, M, N)$
$M \rightarrow M+1$	$f(x n, M+1, N) = \frac{(N-M-n+x)(M+1)}{(N-M)(M+1-x)} \cdot f(x n, M, N)$
$N \rightarrow N+1$	$f(x n, M, N+1) = \frac{(N+1-M)(N+1-n)}{(N+1-M-n+x)(N+1)} \cdot f(x n, M, N)$
d) geometrische Verteilung	
$X \rightarrow X+1$	$f(x+1 \pi) = (1-\pi) \cdot f(x) = q \cdot f(x)$
e) negative Binomialverteilung	
$X \rightarrow X+1$	$f(x+1) = \frac{x+r}{x+1} \cdot q \cdot f(x)$

Aus den Rekursionsformeln folgt, dass alle Verteilungen eingipflig sind (wenn man davon absieht, dass zwei benachbarte Werte x und $x+1$ Modalwerte sein können), einige Verteilungen sind stets (die GV) andere unter bestimmten Bedingungen (etwa NB: $r \leq 1$) sofort (ab $X=0$) monoton fallend.

Übersicht 5.9: Weitere Zusammenhänge zwischen Verteilungen diskreter eindimensionaler Zufallsvariablen

a) Größenbeziehung zwischen Erwartungswert und Varianz und Schiefe

Verteilung	es gilt	Schiefe γ
geometr. (GV) u. negat. Binom. (NB).	$V(X) > E(X)$	$\gamma > 0$ immer linkssteil ^{*)}
Poissonverteilung (P)	$V(X) = E(X)$	$\gamma = \sqrt{\frac{1}{\lambda}} > 0$ immer linkssteil
Zweipunkt (Z), Binomial (B), u. hypergeometrische Verteilung (H)	$V(X) < E(X)$	$\gamma > 0$ wenn $\pi < (1-\pi)$ $\gamma < 0$ wenn $\pi > (1-\pi)$

^{*)} mit abnehmendem $p=\pi$ (zunehmendem $q=1-\pi$) und zunehmendem γ .

b) Reproduktivität (das Zeichen \sim bedeutet: ist verteilt. Die Zusammenhänge gelten auch bei der Summe von $n > 2$ ZVn.

	X_1, X_2 sind verteilt:	X_1+X_2	reproduktiv
Z	$X_1 \sim Z(\pi), X_2 \sim Z(\pi)$	$\sim B(2, \pi)$	nein
B	$X_1 \sim B(n_1, \pi), X_2 \sim B(n_2, \pi)$	$\sim B(n_1+n_2, \pi)$	ja
H	$X_1 \sim H, X_2 \sim H$		nein
GV	$X_1 \sim GV(\pi), X_2 \sim GV(\pi)$	$\sim NB(r=2, \pi)$	nein
P	$X_1 \sim P(\lambda_1), X_2 \sim P(\lambda_2)$	$\sim P(\lambda_1+\lambda_2)$	ja

c) Welche Werte kann die Zufallsvariable X annehmen?

Alle natürlichen Zahlen $X=0, 1, \dots$ bei GV, NB und P natürl. Zahlen bis zu einem Maximalwert bei B und H.

5. Weitere Verteilungen eindimensionaler diskreter Zufallsvariablen

Die **Polya-Verteilung** $PL(n, S, N, c)$ ist eine Verallgemeinerung der B- und H-Verteilung. Urnenmodell: S (oder M) schwarze Kugeln (Erfolg), R rote Kugeln, $N = S + R$. Nach Ziehung einer Kugel werden $c + 1$ Kugeln des gleichen Typs (schwarz oder rot) hinzugelegt.

6. Polytome Versuche (mehrdimensionale diskrete Verteilungen)

Urnenmodell: N Kugeln, davon $M_1 =$ schwarze, $M_2 =$ weiße, $M_3 =$ rote Kugeln ... usw., so dass $\sum M_j = N$, ($j = 1, 2, \dots, k$) und $\pi_j = M_j / N$. Wie man sieht liegt eine k -1-dimensionale Zufallsvariable vor: $P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$ ist die Wahrscheinlichkeit x_1 schwarze und x_2 weiße Kugeln usw. zu ziehen ($k-1$ Dimensionen, da die k -te Dimension durch $\pi_1 + \pi_2 + \dots + \pi_k = 1$ festliegt).

1) Multinomialverteilung (= Polynomialverteilung)

Ziehen **mit** Zurücklegen:

Wahrscheinlichkeit, bei n Zügen genau x_1, x_2, \dots, x_k Kugeln des Typs 1, 2, ..., k zu ziehen ($n = \sum x_i$), dargestellt durch den Vektor $\mathbf{x}' = [x_1 \dots x_k]$

$$f_{MV}(\mathbf{x}) = \left(\frac{n!}{x_1! x_2! \dots x_k!} \right) \pi_1^{x_1} \pi_2^{x_2} \dots \pi_k^{x_k}. \text{ Der Klammerausdruck ist der}$$

Multinomialkoeffizient.

Man erhält $f_{MV}(\mathbf{x})$ durch Expansion des Multinoms $(\pi_1 + \pi_2 + \dots + \pi_k)^n$.

2) Polyhypergeometrische Verteilung

Urnenmodell wie Multinomialverteilung, aber Ziehen **ohne** Zurücklegen

$$f(x) = \frac{\binom{M_1}{x_1} \binom{M_2}{x_2} \cdots \binom{M_k}{x_k}}{\binom{N}{n}}.$$

3) Multiple Poissonverteilung

$$f(x) = \frac{\lambda_1^{x_1} \lambda_2^{x_2} \cdots \lambda_k^{x_k}}{x_1! x_2! \cdots x_k!} \exp[-(\lambda_1 + \lambda_2 + \cdots + \lambda_k)]$$

als Grenzverteilung der Multinomialverteilung

Kapitel 6: Spezielle stetige Verteilungen

1. Rechteckverteilung und andere lineare Verteilungen	51
2. Normalverteilung (univariat)	52
3. Exponentialverteilung	54
4. Funktionen von normalverteilten Zufallsvariablen (χ^2 , t, F)	55
5. Modelle für Einkommensverteilungen	60
6. Exponentialfamilie, Zusammenhänge zwischen stetigen Vert.	61
7. Bivariate Normalverteilung	61

1. Rechteckverteilung und andere lineare Verteilungen

a) Rechteckverteilung $R(a, b)$ (stetige Gleichverteilung)

Eine Zufallsvariable X ist gleichmäßig (gleich-) verteilt über dem Intervall $[a, b]$, wenn $f_R(x)$ in diesem Intervall konstant $\frac{1}{b-a}$ ist. Für die zentralen Momente gilt dann

$$E[(X - \mu)^k] = \frac{1}{(k+1)(b-a)} \left[\left(\frac{b-a}{2} \right)^{k+1} - \left(\frac{a-b}{2} \right)^{k+1} \right], \text{ so dass}$$

$$E[(X - \mu)^k] = \begin{cases} \left(\frac{b-a}{2} \right)^k \frac{1}{k+1} & \text{falls } k \text{ gerade} \\ 0 & \text{falls } k \text{ ungerade} \end{cases}$$

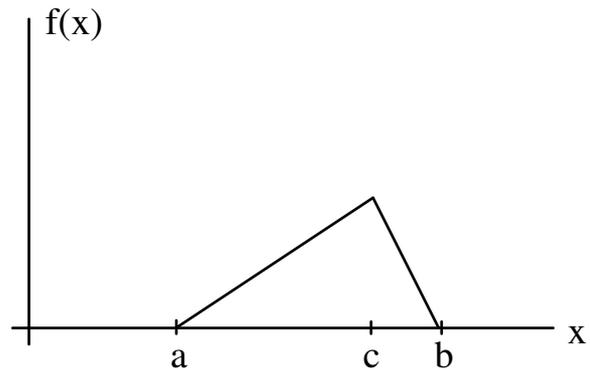
Spezialfall: Rechteckverteilung im Intervall $[0,1]$

Dieser Spezialfall der Rechteckverteilung $R(0,1)$, also $a = 0$, $b = 1$, ist aus folgendem Grunde von besonderem Interesse: Bildet man eine Zufallsvariable Z , so dass $z = F(x)$, so besitzt Z eine $R(0,1)$ Verteilung, denn $0 \leq Z \leq 1$ wegen $0 \leq F(x) \leq 1$ bei beliebiger stetiger Verteilung

von X und nach Gl. (4.26) gilt: $f^*(z) = f[F^{-1}(z)] \left| \frac{dF^{-1}(z)}{dz} \right| = f[F^{-1}(z)] \left| \frac{dx}{dz} \right|.$

b) Abschnittsweise lineare Dichtefunktion; Beispiel: Dreiecksverteilung

$$f(x) = \begin{cases} -\frac{2a}{m} + \frac{2}{m}x & \text{für } a \leq x \leq c \\ \frac{2b}{k} - \frac{2}{k}x & \text{für } c \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$



bei $m = (b-a) \cdot (c-a)$ und $k = (b-a) \cdot (b-c)$

Übersicht 6.1: Rechteckverteilung $R(a, b)$

Wahrscheinlichkeitsfunktion (Dichte)	$f_R(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$
Verteilungsfunktion	$F_R(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 1 & \text{für } x > b \end{cases}$
Parameter	$a, b \in \mathbb{R}$
Momente	$\mu = E(X) = \frac{b+a}{2}, \sigma^2 = V(X) = \frac{(b-a)^2}{12}$
erzeugende Funktionen	$M_x(t) = \frac{e^{bt} - e^{at}}{b-a}, \Psi_x(t) = \frac{e^{ibt} - e^{iat}}{(b-a)it}$
Bedeutung	Erzeugung von Zufallszahlen; eine Zufallsvariable $Z=F(x)$ wobei X beliebig stetig verteilt sein kann ist $R[0,1]$ -verteilt

2. Normalverteilung

a) Allgemeine Normalverteilung $N(\mu, \sigma^2)$

Die Dichtefunktion hat zwei Parameter μ und σ^2

$$(6.1) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (-\infty < x < \infty)$$

Sie ist symmetrisch um $\mu = E(X)$, eingipflig (Modus: μ) und hat zwei Wendepunkte bei $x = \mu - \sigma$ und $x = \mu + \sigma$. Für die Verteilungsfunktion $F(x)$ gilt:

$$(6.2) \quad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \exp\left[-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2\right] dv, \quad F(\mu) = \frac{1}{2} \text{ und}$$

für jedes c : $F(\mu - c) = 1 - F(\mu + c)$, so dass es ausreicht, sie für die Werte $x \geq c$ zu tabellieren.

Reproduktivität (Additionstheorem)

Die Summe von stochastisch unabhängigen normalverteilten Zufallsvariablen $N(\mu_i, \sigma_i^2)$ ist wiederum normalverteilt mit $N(\Sigma\mu_i, \Sigma\sigma_i^2)$. Die Linearkombination $b_1x_1 + \dots + b_nx_n$ ist $N(\Sigma b_i \mu_i, \Sigma b_i^2 \sigma_i^2)$ verteilt.

b) Standardnormalverteilung $N(0,1)$

Jede Zufallsvariable X mit endlichem Mittelwert $E(X) = \mu$ und endlicher Varianz $E(X - \mu)^2 = \sigma^2$ kann durch Transformation $Z = \frac{X - \mu}{\sigma}$ in eine Zufallsgröße Z mit $E(Z) = 0$ und $\text{Var}(Z) = E(Z^2) = 1$ überführt werden.

Wichtige Signifikanzschranken und Wahrscheinlichkeiten

a) z-Werte für gegebene Wkt. $1 - \alpha$

P=1- α	einseitig F(z)	zweiseitig $\Phi(z)$
90%	1,2816	$\pm 1,6449$
95%	1,6449	$\pm 1,9600$
99%	2,3263	$\pm 2,5758$
99,9%	3,0902	$\pm 3,2910$

b) Wkt. für gegebenes z

z	F(z)	$\phi(z)$
0	0,5000	0,0000
1	0,8413	0,6827
2	0,9772	0,9545
3	0,9987	0,9973

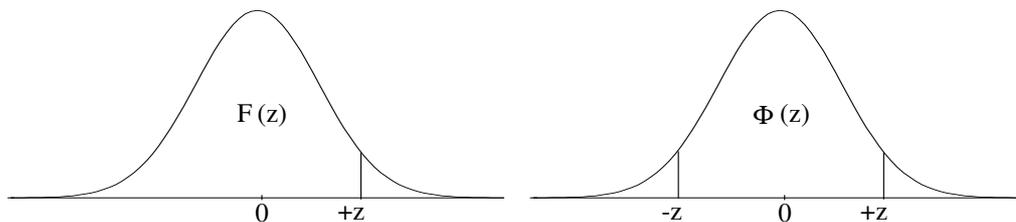
Übersicht 6.2: Normalverteilung $N(\mu, \sigma^2)$

Wahrscheinlichkeitsfunktion (Dichte)	$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$ $-\infty < x < +\infty$
Verteilungsfunktion (Tabellierung)	wenn $X \sim N(\mu, \sigma^2)$ dann ist $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ (Standardnormalverteilung). Die Verteilungsfunktion der $N(0,1)$ -Verteilung ist tabelliert (Tab. auf Seite T-2)
Parameter	μ (zugleich $E(X)$) und σ^2 (zugleich $V(X)$); $\mu \in \mathbb{R}$, $\sigma > 0$
Momente	$E(X) = \mu$ (zugl. Median, Modus), $V(X) = \sigma^2$, $\gamma = 0$
erzeugende Funktionen	$M_x(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$, $\Psi_x(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$
Bedeutung	Grenzverteilung (asymptotische Verteilung) der Summe (und des Durchschnitts) von ZVn.

Dichtefunktion:

$$(6.3) \quad f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (-\infty < z < \infty)$$

Verteilungsfunktion $F(z) = P(Z \leq z)$ und symmetrische Intervallwahrscheinlichkeiten
 $\Phi(z) = P(-z \leq Z \leq z)$



$$\text{Es gilt } F(-z) = 1 - F(z) \text{ und } \Phi(z) = \begin{cases} 2F(z) - 1 & (z > 0) \\ 1 - 2F(z) & (z < 0) \end{cases}$$

Die Tabelle $\Phi(z)$ ist streng genommen überflüssig. Sie kann aber eine Erleichterung der konkreten Berechnung bieten.

Approximation einer diskreten Verteilung durch $N(0,1)$

Die Zufallsvariable X etwa $x = 0, 1, 2, \dots$ werden Intervalle $\left[x - \frac{1}{2}; x + \frac{1}{2}\right]$ zugeordnet und für $a \leq X \leq b$ erhält man bei $E(X) = \mu$, $V(X) = \sigma^2$ die folgenden z -Werte

$$z_a = \frac{a - \mu - \frac{1}{2}}{\sigma} \quad \text{und} \quad z_b = \frac{b - \mu + \frac{1}{2}}{\sigma}$$

Momente

Aus der momenterzeugenden Funktion $M_x(t) = \exp\left(\frac{1}{2}t^2\right) = 1 + \frac{t^2}{2!} + \frac{t^4}{4 \cdot 2!} + \frac{t^6}{8 \cdot 3!} + \dots$ folgt,

dass alle ungeraden Momente der Standardnormalverteilung verschwinden $E(X) = E(X^3) = E(X^5) = \dots = 0$ und für die geraden Momente gilt

$$E(X^k) = (k-1) E(X^{k-2}) \text{ also } E(X^2) = 1, E(X^4) = 3, E(X^6) = 5 \cdot 3, E(X^8) = 7 \cdot 5 \cdot 3 \text{ usw.}$$

3. Exponentialverteilung $E(\lambda)$ und Verallgemeinerungen der $E(\lambda)$ -Verteilung

Die Exponentialverteilung $E(\lambda)$ beschreibt Wartezeiten zwischen Signalen, Ausfällen von Maschinen usw., die einem Poisson-Prozess folgen. Sie ist das stetige Analogon zur geometrischen Verteilung. Die Summe von $E(\lambda)$ verteilten Zufallsvariablen ist Erlang-verteilt $[EL(n, \lambda)]$. Die Verteilungen $E(\lambda)$ und $EL(n, \lambda)$ kann man auch als Spezialfälle der Gamma-Verteilung betrachten (vgl. Übers. 6.4 und 6.10 für Zusammenhänge zwischen der Poissonverteilung und den Verteilungen GV und E einerseits und NB und EL andererseits). Eine andere Verallgemeinerung der $E(\lambda)$ -Vert. ist die Weibull-Verteilung $[W(n, \lambda)]$. Bei bestimmten Annahmen über die Sterbewahrscheinlichkeit bzw. über die Ausfallwahrscheinlichkeit $a(t)$ eines Werkstücks erhält man für die Wahrscheinlichkeit $F(t)$ bis t auszufallen (und damit für die Abgangsordnung $1-F(t)$ einer Sterbetafel) entweder die $E(\lambda)$ -Vert. ($a(t)$ als Exponentialfunktion) oder die Weibull-Verteilung. Die $W(n, \lambda)$ -Vert. sollte nicht verwechselt werden mit der Weibull-Gamma-Verteilung (WG-Vert.). Ein Sonderfall der WG-Vert. ist die Pareto-Verteilung.

Übersicht 6.3: Exponentialverteilung $E(\lambda)$

Wahrscheinlichkeitsfunktion Dichte	$f_E(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$
Verteilungsfunktion	$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$
Parameter	λ mit $\lambda > 0$
Eigenschaften	nicht reproduktiv \rightarrow EL-Verteilung
Momente	$\mu = E(X) = \frac{1}{\lambda}$, $E(X^r) = \frac{r!}{\lambda^r}$, Varianz $\sigma^2 = V(X) = \frac{1}{\lambda^2}$ Schiefe $\gamma = 2$, Median $\tilde{\mu} = \frac{\ln 2}{\lambda}$
andere Verteilungen	Erlang-Verteilung EL: $E(\lambda) = EL(1, \lambda)$, Spezialfall Gamma-Verteilung $\alpha = 1$, $\beta = 1/\lambda$ Spezialfall der Weibull-Vert. $E(\lambda) = W(1, \lambda)$
erzeugende Funktionen	$M_x(t) = \frac{\lambda}{\lambda - t}$, $t < \lambda$, $\Psi_x(t) = \frac{\lambda}{\lambda - it}$
Bedeutung	Wartezeiten, Lebensdauer, Abgangsordnung (Sterbetafel)

Übersicht 6.4: Erlangverteilung $EL(n, \lambda)$

Wahrscheinlichkeitsfunktion (Dichte)	$f(x) = x^{n-1} \lambda^n \frac{e^{-\lambda}}{(n-1)!}$ für $x > 0$
Parameter	λ , n (mit $n \in \mathbb{N}$ und $\lambda > 0$)
Momente	Erwartungswert $E(X) = \frac{n}{\lambda}$, Varianz $V(X) = \frac{n}{\lambda^2}$
andere Verteilungen	Verallgemeinerung der Exponentialverteilung $E(\lambda) = EL(1, \lambda)$, Spezialfall der Gammaverteilung $EL(n, \lambda) = G\left(\alpha = n, \beta = \frac{1}{\lambda}\right)$
erzeugende Funktionen $\psi_x(t)$	$\Psi_x(t) = \left(\frac{\lambda}{\lambda - it}\right)^n$ d.h. die Summe unabhängig identisch $E(\lambda)$ verteilter Variablen X_1, \dots, X_n ist EL verteilt
Bedeutung	Wartezeiten, Bedienungstheorie

4. Funktionen von normalverteilten Zufallsvariablen

Eine lineare Transformation einer $N(0, 1)$ verteilten ZV ist ebenfalls normalverteilt. Anders verhält es sich bei nichtlinearen Funktionen einer $[Z \rightarrow X]$ oder mehrerer [etwa $Z_1^2 + Z_2^2 \rightarrow X$]

standardnormalverteilter Zufallsvariablen¹⁾. Einiger solcher Funktionen spielen für die Schätz- und Testtheorie eine große Rolle. Sie werden in diesem Abschnitt behandelt.

a) Die χ^2 -Verteilung

Sind n Zufallsvariablen Z_1, Z_2, \dots, Z_n unabhängig identisch standardnormalverteilt, so ist die ZV $X = Z_1^2 + Z_2^2 + \dots + Z_n^2 = \sum Z_i^2$ Chi-Quadrat-verteilt mit n "Freiheitsgraden" (d.f. = degrees of freedom) $X \sim \chi_n^2$.

Die χ^2 -Verteilung ist stetig, nur für $x > 0$ definiert, für kleines n linkssteil und sie strebt mit wachsendem n gegen die Normalverteilung mit $\mu = n$ und $\sigma^2 = 2n$. Sie ist reproduktiv (Theorem von Cochran): Sind $X_i \sim \chi_{n_i}^2$, dann ist $\sum X_i \sim \chi_n^2$ (mit $n = \sum n_i$ $i = 1, 2, \dots, m$).

Anwendungen

1. Bei Stichproben aus normalverteilter Grundgesamtheit $X_i \sim N(\mu, \sigma^2)$ ($i = 1, 2, \dots, n$) ist

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \sum Z_i^2 = \frac{1}{\sigma^2} \sum (X_i - \mu)^2 \sim \chi_n^2.$$

2. Wird μ durch \bar{x} ersetzt, ($i = 1, 2, \dots, n$) so gilt $\sum \left(\frac{X_i - \bar{x}}{\sigma} \right)^2 \sim \chi_{n-1}^2$.

3. Bei einfacher linearer Regression ist die geschätzte Störgröße $\hat{u}_i = y_i - \hat{y}_i$, $i = 1, \dots, n$ und $U \sim N(0, \sigma^2)$, so dass gilt, weil zwei Parameter α, β zu schätzen sind (Absolutglied α und Steigung β der Regressionsgerade): $\sum \frac{\hat{u}_i^2}{\sigma^2} = \frac{1}{\sigma^2} \sum \hat{u}_i^2 \sim \chi_{n-2}^2$;

und entsprechend bei multipler linearer Regression mit $p-1$ Regressoren X_1, X_2, \dots, X_{p-1} und einem Absolutglied (dummy regressor X_0) $\sum \frac{\hat{u}_i^2}{\sigma^2} = \frac{1}{\sigma^2} \sum \hat{u}_i^2 \sim \chi_{n-p}^2$,

während für die wahren Residuen gilt $\sum_{i=1}^n \frac{u_i^2}{\sigma^2} \sim \chi_n^2$.

b) Die t-Verteilung

Wenn $X \sim N(0, 1)$ und $Y \sim \chi_n^2$ wobei X und Y unabhängig sind, dann ist $t = X / \sqrt{Y/n} \sim t_n$ (t-verteilt mit n Freiheitsgraden).

Die t-Verteilung (Student-Verteilung) ist stetig, hat einen Parameter n (Anzahl der Freiheitsgrade) und hat eine ähnliche Gestalt wie die Normalverteilung mit um so größerer Streuung (wegen des Nenners $\sqrt{Y/n}$), je kleiner n ist. Mit $n > 30$, zumindest ab $n \approx 40$ ist die t-Verteilung durch $N(0, 1)$ bereits gut approximierbar.

¹⁾ Eine andere nichtlineare Transformation von z ist $x = e^z$ (also $z = \ln(x)$). Vgl. Logarithmische Normalverteilung. Oder: Wenn Z_1 und Z_2 normalverteilt sind $N(0, \sigma^2)$, dann ist $X = Z_1/Z_2$ Cauchy-verteilt $C(\sigma, 0)$.

Anwendungen

1. Wenn $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$ dann ist die "quasistandardisierte" Variable

$$Z^* = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1} \text{ mit } s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \text{ und } \hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 .$$

2. Eine entsprechende "Quasistandardisierung" liegt bei der Schätzung eines Regressionskoeffizienten β durch $\hat{\beta} = b$ vor (z.B. bei einfacher Regression) mit $\sigma_b^2 = \sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2$ als Standardabweichung der Stichprobenverteilung (Kap. 7, 8) des Regressionskoeffizienten b (Steigung der Regressionsgeraden).
3. Sind \bar{x}_1, \bar{x}_2 und s_1^2, s_2^2 Mittelwerte und Varianzen von Stichproben des Umfangs n_1 und n_2 aus normalverteilten Grundgesamtheiten (mit identisch $N(\mu, \sigma^2)$), dann ist

$$(6.4) \quad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t_{n_1 + n_2 - 2} \quad (\text{vgl. Gl. 9.16})$$

c) Die F-Verteilung

Sind $X \sim \chi_m^2$ und $Y \sim \chi_n^2$ unabhängig verteilt, so ist die Zufallsvariable $F = (X/m) / (Y/n)$ F-verteilt mit m und n Freiheitsgraden.

Die F-Verteilung ist eine stetige Verteilung mit zwei Parametern m und n und sie ist wichtig für den Test auf Gleichheit zweier Varianzen und zum Testen von Korrelations- und Regressionskoeffizienten sowie für die Varianzanalyse.

Anwendungen

1. Sind die Zufallsvariablen X_1 und X_2 unabhängig normalverteilt $N(\mu_i, \sigma^2)$ ($i = 1, 2$), so ist der Quotient der Stichprobenvarianzen (mit den Mittelwerten \bar{x}_1 und \bar{x}_2 und den Stichprobenumfängen $n_1 = m + 1$ und $n_2 = n + 1$)

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{n_1 s_1^2}{n_2 s_2^2} \frac{n_2 - 1}{n_1 - 1} \sim F_{m,n} \quad (\text{mit } s_1^2 > s_2^2) .$$

2. In einer linearen Regression mit p Regressoren (Absolutglied mitgezählt) wenn der Regressionskoeffizient β durch b geschätzt wird, ist $t = \frac{b - \beta}{\sigma_b} \sim t_{n-p}$ (mit σ_b als Standardabweichung der Stichprobenverteilung von $b = \hat{\beta}$, in Abhängigkeit von σ^2 , der wahren Varianz der Störgröße), dann ist $t^2 \sim F_{p, n-p}$, wenn in σ_b die Varianz σ^2 ersetzt ist durch $\hat{\sigma}^2 = \sum \hat{u}^2 / (n - p)$ (Varianzschätzer).

3. Ist $X \sim F(m,n)$ dann ist $Z = \frac{\alpha X}{\beta(1-X)} \sim B_1(\alpha, \beta)$ (Beta-Verteilung 1. Art) mit $\alpha = m/2$ und $\beta = n/2$

Übersicht 6.5: χ^2 -Verteilung (χ_n^2)

Wahrscheinlichkeitsfunktion (Dichtefunktion) $x > 0$	$f(x) = \begin{cases} \frac{1}{2^m \Gamma(m)} X^{m-1} e^{-\frac{x}{2}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$ <p>bei $m = n/2$ und $\Gamma(p)$, $p > 0$ die Gammafunktion (vgl. Kap.2)</p>
Parameter	$n = \text{Anzahl der Freiheitsgrade, } n \in \mathbb{IN}$
Momente	$E(X) = \mu = n \quad V(X) = \sigma^2 = 2n \quad \gamma = \frac{2\sqrt{2}}{\sqrt{n}}$ (linkssteil) $E(X^k) = 2^k \left[\frac{\Gamma\left(k + \frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \right]$; Modus: $n - 2$ wenn $n > 2$
andere Verteilungen Grenzübergang zu $N(n, 2n)$	Spezialfall: Gamma-Verteilung $G(\alpha, \beta) = G\left(\frac{n}{2}, 2\right)$
erzeugende Funktionen	$\Psi_x(t) = (1 - 2it)^{-\frac{n}{2}}$
Bedeutung	Konfidenzintervall und Test von Varianzen, Regressions- und Korrelationsanalyse, Anpassungs- und Unabhängigkeitstests
Reproduktivität	wenn $X_1, X_2, \dots \sim \chi_{n_i}^2$ dann $\sum X_i \sim \chi_n^2$ mit $n = \sum n_i$

Übersicht 6.6: t -Verteilung (t_n)

Wahrscheinlichkeitsfunktion (Dichtefunktion) $-\infty < x < +\infty$	$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}$
Parameter	$n = \text{Anzahl der Freiheitsgrade, } n \in \mathbb{IN}$
Momente	$E(X) = 0$ (alle ungeraden Momente: 0) wenn $n \geq 2$ $V(X) = \frac{n}{n-2}$ wenn $n \geq 3$; $\gamma = 0$ (symmetrisch), wenn $n \geq 4$ $E(X^k) = \frac{(2k-1)!! (n-2k-2)!!}{(n-2)!!}$ (gerade Anfangsmomente; $2k \leq n-1$), Modus: $n - 2$ wenn $n > 2$
andere Verteilungen	bei $n = 1$ Cauchy-Verteilung mit $n \rightarrow \infty$ Übergang zu $N(0, 1)$
Bedeutung	Konfidenzintervalle und Tests bei Mittel- und Anteilswerten sowie bei Regressionskoeffizienten

Übersicht 6.7: F-Verteilung $F(m, n)$

Wahrscheinlichkeitsfunktion (Dichtefunktion) $x > 0$	$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right) \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right) \left(1 + \frac{m}{n}x\right)^{\frac{(m+n)}{2}}}$
Momente	$E(X) = \frac{n}{n-2} \quad (\text{wenn } n > 2)$ $V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (\text{wenn } n > 4)$ $E(X^k) = \frac{\Gamma\left(\frac{m-k}{2}\right) \Gamma\left(\frac{n-k}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)}; \text{ Modus } \frac{n(m-2)}{m(n+2)} \text{ wenn } m \geq 2$
Grenzübergänge	$n \rightarrow \infty$ dann $F \rightarrow \chi_m^2$
Bedeutung	Vergleich von Varianzen, Linearitätstest bei Regressionsanalyse, Varianz- und Kovarianzanalyse

Übersicht 6.8: Zusammenhänge zwischen χ^2 , t und F -Verteilung und Anwendungen der Verteilungen

a) Zusammenhänge

wenn	dann ist	mit den Parametern	verteilt
$F \sim F_{m, n}$	\sqrt{F}	$m = n = 1$ $m = 1$ $m = 1, n \rightarrow \infty$	$C(1, 0)^*$ t_n $N(0, 1)$
	F	$n \rightarrow \infty$	χ_m^2
$t \sim t_n$	t^2	$m = 1$	$F(1, n)$
	t	$n \rightarrow \infty$	$N(0, 1)$

*) Cauchy-Verteilung

b) Anwendung der Verteilungen bei Konfidenzintervallen und Tests

Substitution einer Varianz σ^2 durch die Stichprobenvarianz $\hat{\sigma}^2$ bedeutet:

	wenn die ZV mit σ^2	dann ist die ZV mit $\hat{\sigma}^2$	Anwendung: Test und Konfidenzintervalle für
1.	normalverteilt ist	t-verteilt	einen Parameter
2.	C^2 -verteilt ist*)	F-verteilt	mehrere Parameter**)

*) Ist $X \sim \chi_n^2$ dann ist $X/n \sim C^2$ verteilt (modifizierte χ^2 -Verteilung).

***) Anwendung von Nr. 2 auch: Konfidenzellipse bei linearer Regression. Die Verteilungen in Spalte 1 sind Spezialfälle der Verteilungen von Spalte 2:

wenn $n \rightarrow \infty$ dann $t_n \rightarrow N(0, 1)$; wenn $n \rightarrow \infty$ dann $F_{m, n} \rightarrow C_m^2$

5. Modelle für Einkommensverteilungen

In diesem Abschnitt werden zwei Verteilungen für nichtnegative ZVn (also $X > 0$) dargestellt, die zwar formal kaum in Beziehung zueinanderstehen, aber für Ökonomen (u.a. als Modelle für die Einkommensverteilung) von Interesse sind.

a) Logarithmische Normalverteilung $L(\mu, \sigma^2)$

Sie spielt für Ökonomen eine wichtige Rolle als Modell für die Einkommensverteilung. Ist die Summe von Einflussgrößen asymptotisch normalverteilt, so gilt dies bei der L-Verteilung für ein Produkt. Die L-Verteilung ist stets linkssteil und zwar um so mehr, je größer σ^2 ist. Ist X (etwa das Einkommen) $L(\mu, \sigma^2)$ verteilt, dann ist Gini's Dispersionsmaß die Fläche unter der Standardnormalverteilung im Intervall $\left[-\frac{\sigma}{\sqrt{2}} \leq x \leq +\frac{\sigma}{\sqrt{2}}\right]$.

b) Pareto-Verteilung

$$f(x) = \begin{cases} \frac{a}{b} \left(\frac{b}{x}\right)^{a+1} & \text{für } x \geq b \text{ (} a > 0, 0 < b < x \text{) (streng monoton fallend)} \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = 1 - \left(\frac{b}{x}\right)^a, \quad \mu = E(X) = b \frac{a}{a-1}, \quad V(X) = \frac{ab^2}{(a-2)(a-1)^2}$$

$$\text{Median } \tilde{\mu} = b \cdot (0,5)^{\frac{1}{a}} \text{ also } \tilde{\mu} < b < \mu.$$

Übersicht 6.9: Logarithmische Normalverteilung $L(\mu, \sigma^2)$

Wahrscheinlichkeitsfunktion, (Dichte) für $x > 0$	$f_L(x) = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]$
Parameter	$\mu \in \mathbb{R}, \sigma > 0$
(Zusammenhang mit $N(\mu, \sigma^2)$ -verteilung)	wenn $\ln X = Y \sim N(\mu_y, \sigma_y^2)$ dann $X \sim L(\mu_x, \sigma_x^2)$
Reproduktivität	wenn $X_1 \sim L(\mu_1, \sigma_1^2)$ und $X_2 \sim L(\mu_2, \sigma_2^2)$ dann $X_1 \cdot X_2 \sim L(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ und $\frac{X_1}{X_2} \sim L(\mu_1 - \mu_2; \sigma_1^2 + \sigma_2^2)$
Momente	$\mu = E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad E(X^r) = \exp\left(r\mu + \frac{1}{2}r^2\sigma^2\right)$ $\sigma^2 = V(X) = \exp\left[2\mu + \sigma^2(S-1)\right]$ mit $S = e^{\sigma^2}, \gamma = \sqrt{S-1}(S+2)$ Median $\tilde{\mu} = \exp(\mu)$, Modus $\exp(\mu - \sigma^2)$
Bedeutung	Einkommensverteilung, Disparitätsmessung, Lebensdauer, Festigkeit

6. Die Exponentialfamilie von Verteilungen und Zusammenhänge zwischen stetigen Verteilungen

Eine umfangreiche Klasse von diskreten und stetigen Verteilungen,

$$(6.5) \quad f(x|\theta) = A(\theta) p(x) \exp [B(\theta) \cdot q(x)]$$

die bei entsprechender Spezifizierung der Funktionen A, B, p, q wichtige Wahrscheinlichkeitsverteilungen als Spezialfälle enthält, heißt Exponentialfamilie.

Spezialfälle (Beispiele)

a) diskret

Binomialverteilung $A(\theta) = (1 - \pi)^n$, $p(x) = \binom{n}{x}$,

$$B(\theta) = \ln \left[\frac{\pi}{(1 - \pi)} \right], \quad q(x) = x$$

ferner: Zweipunkt-, Poisson-, geometrische-, negative Binomialverteilung

b) stetig

Exponentialverteilung $A(\theta) = \lambda$, $p(x) = 1$, $B(\theta) = -\lambda$, $q(x) = x$

ferner: N (μ , σ^2), N (0, 1), Gamma-[G (α , β)], B_1 -, B_2 -, χ^2 -, Rayleigh- und Maxwell-Verteilung als Spezialfälle zur G-, B_1 - und Cauchy-Verteilung [C (λ , μ)].

7. Bivariate Normalverteilung $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Der Zufallsvektor $\mathbf{x}' = [X_1 \ X_2]$ besitzt eine N_2 -Vert. wenn die Dichtefunktion lautet

$$(6.6) \quad f_{N_2}(\mathbf{x}) = \left[(2\pi) |\boldsymbol{\Sigma}|^{\frac{1}{2}} \right]^{-1} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

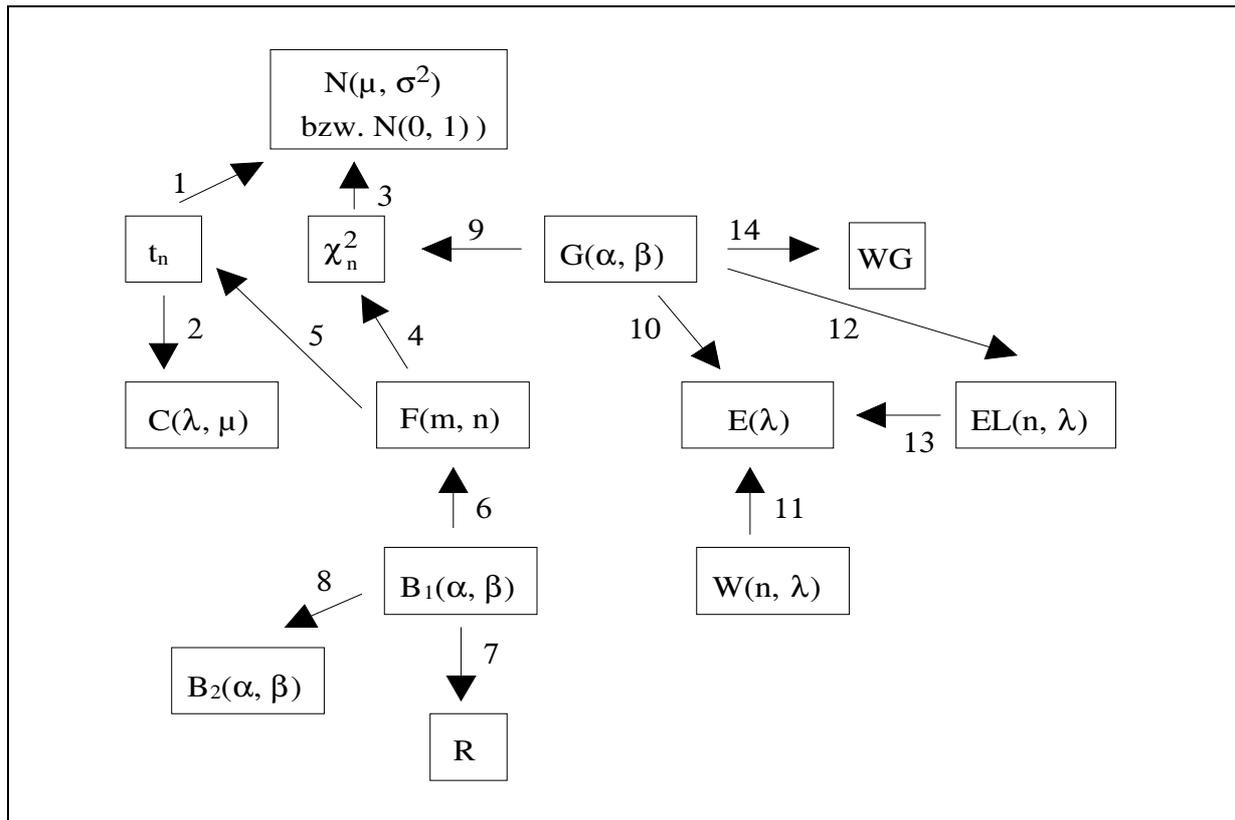
mit $\boldsymbol{\mu}' = [\mu_1, \mu_2]$ und $\boldsymbol{\Sigma}$ als Varianz-Kovarianzmatrix $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. Mit standardisierten

Variablen $Z_i = \frac{X_i - \mu_i}{\sigma_i}$ und der Korrelationsmatrix $\mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

(ρ = Korrelationskoeffizient) erhält man

$$(6.7) \quad f_{N_2}(\mathbf{z}) = (2\pi)^{-1} |\mathbf{R}|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{z}' \mathbf{R}^{-1} \mathbf{z} \right)$$

Übersicht 6.10: Zusammenhänge zwischen eindimensionalen stetigen Verteilungen



Verteilungen

$B_1(\alpha, \beta)$	Beta Verteilung 1. Art	$N(\mu, \sigma^2)$	Normalverteilung
B_2	Beta Verteilung 2. Art	R	Rechteckverteilung ($0 \leq x \leq 1$)
$C(\lambda, \mu)$	Cauchy Verteilung	t_n	t - Verteilung
$E(\lambda)$	Exponentialverteilung	$W(n, \lambda)$	Weibullverteilung
$EL(n, \lambda)$	Erlangverteilung	WG	Weibull-Gamma-Verteilung
$F(m, n)$	F-Verteilung	χ_n^2	χ^2 -Verteilung
$G(\alpha, \beta)$	Gamma - Verteilung		

Übergänge bzw. Spezifizierung von Parameter und Transformationen ($x \rightarrow y$)

1. $n \rightarrow \infty$	8. $y = \frac{x}{1+x}$
2. $n = 1, \lambda = 1, \mu = 0$ also $C(1, 0) = t_1$	9. $\alpha = \frac{n}{2}, \beta = 2$
3. $n = 1, Y = +\sqrt{X}$ ist $X^2 \sim \chi_1^2$ dann $Y = X \sim N(0, 1)$	10. $\alpha = 1, \beta = \frac{1}{\lambda}$
4. ist $X \sim F(m, n)$ dann $Y = \frac{X}{m} \sim \chi_m^2$ $n \rightarrow \infty$	11. $n = 1$
5. wenn $t^2 \sim F(1, n)$ dann $t \sim t_n$	12. $\alpha = n, \beta = \frac{1}{\lambda}$
6. $\alpha = \frac{m}{2}, \beta = \frac{n}{2}, y = \frac{x}{1-x} \frac{n}{m}$ $X \sim B_1(\alpha, \beta)$ dann $Y \sim F(m, n)$	13. $n = 1$
7. $\alpha = \beta = 1$	14. Übergang Weibull-Gamma-Vert.; ein Spezialfall von WG ist die Pareto-Verteilung

Horizontale (parallel zur x_1, x_2 - Ebene) Schnitte $f(x) = \text{const} = c$ durch die Dichte sind Ellipsen im x_1, x_2 - Koordinatensystem $z_1^2 - 2\rho z_1 z_2 + z_2^2 = c$ und im Grenzfall $\rho = 0$ Kreise. Für die Regressionslinien $E(X_2 | X_1)$ in Abhängigkeit von X_1 [und $E(X_1 | X_2)$ von X_2] erhält man Geraden (die Regressionsgeraden)

$$E(X_2 | X_1 = x_1) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1) \quad \text{und} \quad E(X_1 | X_2 = x_2) \quad \text{analog} .$$

Die momenterzeugende Funktion von N_2 ist

$$M_{\mathbf{x}}(t_1, t_2) = \exp \left[\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2} (\sigma_1^2 t_1^2 + 2\sigma_{12} t_1 t_2 + \sigma_2^2 t_2^2) \right]$$

Wie die Gestalt von $M(0, t_2)$ und $M(t_1, 0)$ zeigt sind die Randverteilungen von N_2 wieder Normalverteilungen. Die Momente der Randverteilungen erhält man durch mehrmaliges Differenzieren nach t_1 bzw. nach t_2 . Das Produktmoment $E(X_1 X_2)$ erhält man mit der Ableitung $\frac{\partial^2 M(t_1, t_2)}{\partial t_1 \partial t_2}$ an der Stelle $t_1 = t_2 = 0$.

Kapitel 7: Grenzwertsätze, Stichprobenverteilung

1. Stochastische Konvergenz und Konvergenz von Verteilungen	63
2. Stichproben und Schätzproblem, Stichprobenverteilung	66
3. Abschätzungen von Wahrscheinlichkeiten	68
4. Gesetze der großen Zahlen	69
5. Grenzwertsätze	70

1. Stochastische Konvergenz und Konvergenz von Verteilungen

a) Konvergenzbegriffe und Grenzwertsätze

▪ *Nichtstochastische Konzepte*

1. Man sagt eine von n abhängige Zahlenfolge $\{z_1, z_2, \dots, z_n\}$ konvergiere gegen einen festen Wert c , wenn gilt: $\lim_{n \rightarrow \infty} z_n = c$.
2. Eine Folge von Funktionen $\{f_1(x), f_2(x), \dots, f_n(x)\}$ mit dem gemeinsamen Definitionsbereich D ($x \in D$) konvergiert gegen eine Grenzfunktion $f(x)$, wenn gilt $\lim_{n \rightarrow \infty} f_n(x) = f(x)$. (Man unterscheidet gleichmäßige und punktweise Konvergenz).

▪ *Stochastische Konzepte*

Diese Konzepte für nicht zufällige Größen bzw. Funktionen werden übertragen auf die Wahrscheinlichkeit bzw. Wahrscheinlichkeitsverteilung zufälliger Größen. Die sog. Gesetze der großen Zahl beziehen sich auf die stochastische Konvergenz (Def. 7.1), die man als Spezialfall der Konvergenz von Verteilungen (Def. 7.2) auffassen kann (Konvergenz gegen eine Einpunktverteilung). Konvergenz im zweiten Sinne ist Gegenstand von Grenzwertsätzen. In lokalen Grenz-

wertsätzen wird die Grenzwertverteilung $f(x)$ von diskreten Wahrscheinlichkeitsfunktion bzw. Dichtefunktion und in globalen Sätzen die Konvergenz von Verteilungsfunktionen nach $F(x)$, der Grenzwertverteilungsfunktion (asymptotische Verteilungsfunktion), untersucht.

b) Stochastische Konvergenz

Def. 7.1: Stochastische Konvergenz

Eine vom Parameter n abhängende Zufallsgröße X_n strebt für $n \rightarrow \infty$ stochastisch gegen eine Konstante c , wenn für jedes beliebige $\varepsilon > 0$ gilt [man sagt auch: X_n konvergiert "mit (oder: "in der") Wahrscheinlichkeit" gegen c]:

$$(7.1) \quad \lim_{n \rightarrow \infty} P(|X_n - c| < \varepsilon) = 1$$

oder: $\text{plim } X_n = c$ (d.h. der "**Wahrscheinlichkeitslimes**" von X_n ist c).

Bemerkungen zu Def. 7.1:

1. Gl. 7.1 besagt:
die Wahrscheinlichkeit dafür, dass X_n in einer ε -Umgebung um c liegt strebt gegen 1,
nicht aber:
die Größe X_n strebt gegen den festen (nicht von n abhängigen) Wert c .

2. Man kann auch $c = 0$ setzen und erhält mit

$$(7.2) \quad \lim_{n \rightarrow \infty} P(|X_n| \geq \varepsilon) = 0$$

die Aussage: X_n konvergiert stochastisch gegen Null.

Zwei Folgerungen:

- 1) Eine Folge $\{X_n\}$ von Zufallsvariablen strebt gegen eine Grenz-Zufallsvariable X oder Konstante c (Gl. 7.1), wenn die Folge $\{X_n - X\}$ oder $\{X_n - c\}$ stochastisch gegen Null konvergiert.
- 2) Gl. 7.2 gilt dann und nur dann, wenn die Folge $F_n(x)$ der Verteilungsfunktion von X_n im gewöhnlichen Sinne gegen die Verteilungsfunktion der Einpunktverteilung

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 & \text{für } x > 0 \end{cases}$$

in jeder Stetigkeitsstelle von $F(x)$ (alle Werte $x \neq 0$) konvergiert.

3. Eine stärkere Forderung als die stochastische Konvergenz (Konvergenz mit Wahrscheinlichkeit) ist die Konvergenz im Mittel (vgl. Def. 7.4).

c) Konvergenz von Verteilungsfunktionen

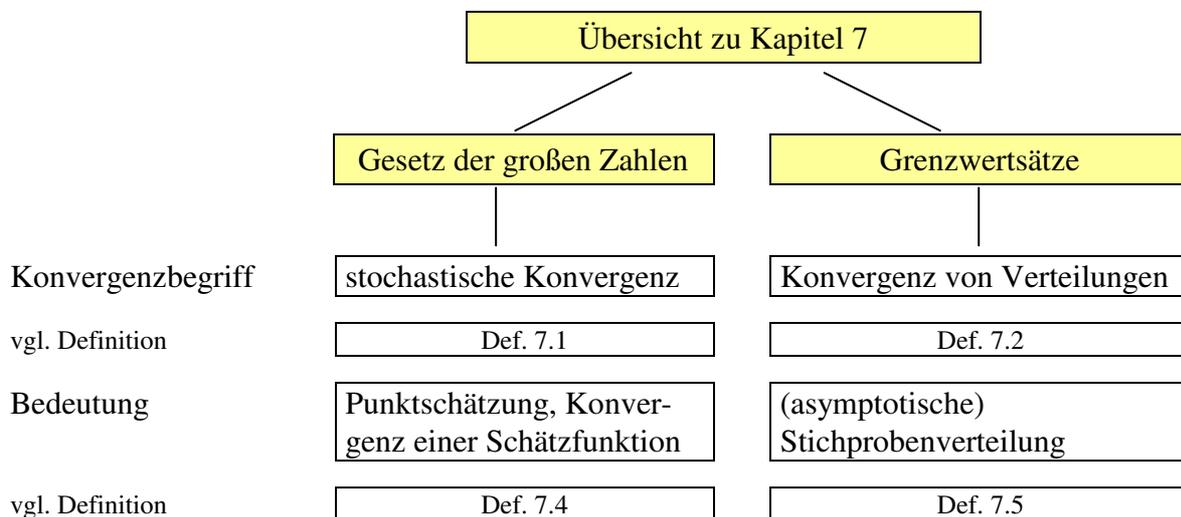
Def. 7.2: Konvergenz von Verteilungsfunktionen

Die Folge der Verteilungsfunktionen $F_n(x)$ einer Folge von Zufallsvariablen X_n heißt konvergent, wenn eine Verteilungsfunktion $F(x)$, die Grenzverteilungsfunktion, existiert, so dass für jede Stetigkeitsstelle von $F(x)$ gilt:

$$(7.3) \quad \lim_{n \rightarrow \infty} F_n(x) = F(x)$$

Bemerkungen zu Def. 7.2: vgl. nächste Seite

Übersicht 7.1: Grenzwertsätze und Gesetz der großen Zahlen



	homograd	heterograd
Grundgesamtheit	Zweipunktverteilung $E(X_i) = \pi, V(X_i) = \pi(1 - \pi) \quad \forall i$	X diskret oder stetig; beliebig verteilt mit $E(X_i) = \mu, V(X_i) = \sigma^2 \quad \forall i$
Stichprobe	X_1, X_2, \dots, X_n also n unabhängige*) Zufallsvariablen aus identischen Grundgesamtheiten; nach Ziehung der Stichprobe Realisationen x_1, x_2, \dots, x_n	
Kennzahlen einer Stichprobe vom Umfang n	$X = \sum X_i$ Anzahl der Erfolge $P = \frac{1}{n} X$ Anteil der Erfolge	$Y = \sum X_i$ Merkmalssumme $\bar{X} = \frac{1}{n} \sum X_i$ Mittelwert
(schwaches) Gesetz der großen Zahl(en)	Theorem von Bernoulli $\text{plim}(P) = \pi$	Satz von Ljapunoff $\text{plim}(\bar{X}) = \mu$
Grenzwertsätze (Aussagen über Grenzverteilungen)	a) die Wahrscheinlichkeitsvert. von X (Binomialverteilung) bzw. P (relat. Binomialvert.) strebt gegen die Normalvert. (de Moivre-Laplace). b) das gilt auch für die Verteilungsfunktion $F(z), F(p)$	a) asymptotische Verteilung von Y und \bar{X} ist jeweils die Normalvert. $Y \sim N(n\mu, n\sigma^2)$ $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ b) als globaler Grenzwertsatz: Satz von Lindeberg-Lévy **)
a) lokal $f_n(x) \rightarrow f(x)$		
b) global $F_n(x) \rightarrow F(x)$		

*) Diese Annahme kann auch gelockert werden (Normalverteilung auch Grenzverteilung bei abhängigen ZVn, vgl. Satz von Markoff).

***) Verallgemeinerung (nicht identische Verteilungen): **Zentraler** Grenzwertsatz von **Ljapunoff**

Bemerkungen zu Def. 7.2: vgl. nächste Seite

1. Es kann vorkommen, dass eine Folge von Verteilungsfunktionen gegen eine Funktion konvergiert, die keine Verteilungsfunktion ist.
2. Man beachte, dass hier Konvergenz im üblichen (nichtstochastischen) Sinne (punktweise Konvergenz einer Folge von Funktionen [s. o.]) gemeint ist, nicht stochastische Konvergenz.
3. Im allgemeinen folgt aus der Konvergenz von Verteilungsfunktionen $F_n(x)$ nicht, dass auch ein lokaler Grenzwertsatz gilt, d.h. die Wahrscheinlichkeits-, bzw. Dichtefunktionen $f_n(x)$ konvergieren.
4. Konvergiert $F_n(x)$ gegen $F(x)$ und sind a, b ($a < b$) zwei beliebige Stetigkeitspunkte der Grenzverteilung $F(x)$ dann gilt

$$(7.4) \quad \lim_{n \rightarrow \infty} P(a \leq X_n \leq b) = F(b) - F(a)$$

2. Stichproben und Schätzproblem, Stichprobenverteilung

Im Zusammenhang mit Stichproben (Def. 7.3) tritt das Problem auf, "Parameter" der (unbekannten) Grundgesamtheit (GG) mit Hilfe von beobachteten Werten einer Stichprobe (genauer: einer Funktion dieser Werte, der "Schätzfunktion", Def. 7.4) zu schätzen.

Def. 7.3: Stichprobe

1. Eine durch Zufallsauswahl genommene endliche Menge von Beobachtungswerten x_1, x_2, \dots, x_n aus einer endlichen oder unendlichen Grundgesamtheit heißt Stichprobe. Die Zahl n ist der Stichprobenumfang. Die Grundgesamtheit hat, wenn sie endlich ist, N Elemente (den Umfang N).
2. Hat jedes Element einer endlichen Grundgesamtheit die gleiche a priori (vor Ziehung der Stichprobe) bekannte von Null verschiedene Wahrscheinlichkeit in die Stichprobe zu gelangen, so spricht man von uneingeschränkter Zufallsauswahl.
3. Erfolgen die Ziehungen der n Einheiten im Rahmen einer uneingeschränkten (reinen) Zufallsauswahl unabhängig voneinander, so spricht man von einer einfachen Stichprobe.

Bei einer einfachen Stichprobe geht man davon aus, dass jeder Stichprobenwert x_i ($i = 1, 2, \dots$) die Realisation einer Zufallsvariable X_i ist, wobei alle Zufallsvariablen X_1, X_2, \dots, X_n identisch verteilt sind. In diesem Fall sind die ZVn auch unabhängig identisch verteilt (independently, identically distributed [i.i.d.]), so dass die gemeinsame Verteilungsfunktion $F_G(x_1, x_2, \dots, x_n)$ als Produkt $F_1(x_1) F_2(x_2) \dots F_n(x_n)$ darstellbar ist.

Def. 7.4: Schätzwert, Schätzfunktion, Mean Square Error

1. Bestimmte kennzeichnende Größen, etwa das arithmetische Mittel, die Varianz, der Anteilswert einer Grundgesamtheit werden Parameter (θ) genannt. Sie sind i.d.R. Funktionen der Zufallsvariablen X_1, X_2, \dots, X_N (bei endlicher Grundgesamtheit).
2. Eine Funktion der Stichprobenwerte, wie $\hat{\theta} = g(X_1, X_2, \dots, X_n)$, die **vor** Ziehung der Stichprobe eine Zufallsvariable ist, heißt Schätzfunktion (engl. estimator oder statistic) oder Stichprobenfunktion. Ein konkreter Funktionswert $\hat{\theta} = g(x_1, x_2, \dots, x_n)$ (**nach** Ziehung der Stichprobe errechnet) heißt Schätzwert. $\hat{\theta}$ dient der Schätzung von θ .

3. Die Größe

$$(7.5) \quad \text{MSE} = E[(\hat{\theta} - \theta)^2]$$

heißt mean square error oder mittlerer quadratischer Fehler.

Bemerkungen zu Def. 7.4:

1. Unter den Parametern und den hierzu korrespondierenden Schätzfunktionen spielen vor allem zwei lineare Funktionen eine besondere Rolle, nämlich:
 - a) bei einer metrisch skalierten Zufallsvariable $\theta_1 = \mu$, der Mittelwert (bei einer endlichen Grundgesamtheit) bzw. der Erwartungswert (bei einer Wahrscheinlichkeitsverteilung) und
 - b) bei dichotomen (zweipunktverteilten) Variablen $\theta_2 = \pi$, der Anteil der Erfolge (bei endlicher Grundgesamtheit) bzw. die Wahrscheinlichkeit eines Erfolges.
 Man spricht im ersten Fall von heterograder, im zweiten von homograder Theorie (vgl. Übers. 7.1).
2. Der MSE läßt sich wie folgt zerlegen (Varianzzerlegung):

$$(7.5a) \text{MSE} = E[\{\hat{\theta} - E(\hat{\theta})\}^2] + [E(\hat{\theta}) - \theta]^2 = V(\hat{\theta}) + [B(\hat{\theta})]^2$$

3. Daraus folgt: Der MSE ist genau dann Null, wenn **Varianz** $V(\hat{\theta})$ und **Bias** $B(\hat{\theta})$ Null sind. Die "Verzerrung" $B(\hat{\theta})$ ist Null, wenn $E(\hat{\theta}) = \theta$, also $\hat{\theta}$ ein erwartungstreuer (unverzerrter, biasfreier) Schätzer für θ ist.
4. Da $\hat{\theta}$ als Stichprobenfunktion vom Umfang n der Stichprobe abhängig ist, kann man auch $\hat{\theta} = \hat{\theta}_n$ schreiben. Mit

$$(7.6) \quad \lim_{n \rightarrow \infty} E |\hat{\theta}_n - \theta|^2 = \lim_{n \rightarrow \infty} \{ \theta - E(\hat{\theta}_n) \}^2 = 0$$

ist die Konvergenz im Mittel (Konvergenz im quadratischen Mittel) der Zufallsvariable $\hat{\theta}_n$ gegen die ZV oder Konstante θ definiert. Sie impliziert Konvergenz mit Wahrscheinlichkeit (stochastische Konvergenz), aber nicht umgekehrt. Es ist üblich, die Standardabweichung $\sqrt{\text{MSE}}$ als Stichprobenfehler zu bezeichnen (vgl. Abschn. 8.2).

Def. 7.5: Stichprobenverteilung

Die Verteilung $f(\hat{\theta})$ der Zufallsvariable $\hat{\theta}$ als Schätzfunktion $\hat{\theta}$ für θ heißt Stichprobenverteilung von $\hat{\theta}$.

Erklärung zu Def. 7.5:

Aus einer endlichen GG des Umfangs N sind $\binom{N}{n}$ verschiedene Stichproben des Umfangs n ohne Zurücklegen zu ziehen, die bei uneingeschränkter Zufallsauswahl alle gleich wahrscheinlich sind^{*)}. Jede dieser Stichproben liefert eine Häufigkeitsverteilung des Merkmals X und z.B. ein arithmetisches Mittel \bar{x} . Auch für die endliche, bzw. unendliche Grundgesamt-

^{*)} Bei Ziehen mit Zurücklegen gibt es N^n mögliche und damit gleichwahrscheinliche Stichproben, die aber nicht alle verschieden (unterschiedliche Elemente enthaltend) sind.

heit gibt es eine Häufigkeits-, bzw. Wahrscheinlichkeitsverteilung von X mit dem Mittelwert bzw. Erwartungswert μ .

Die Wahrscheinlichkeitsverteilung aller $\binom{N}{n}$ Stichprobenmittelwerte \bar{x} , die nicht notwendig alle verschiedene Werte annehmen müssen, ist die Stichprobenverteilung des arithmetischen Mittels.

3. Abschätzungen von Wahrscheinlichkeiten

a) Ungleichung von Markoff

Bei nicht negativen Zufallsvariablen X mit einem endlichen Erwartungswert $E(X)$ gilt für jede positive reelle Zahl c

$$(7.7) \quad P(X \geq c) \leq \frac{E(X)}{c}$$

Daraus folgt mit $X = [Y - E(Y)]^2$, der Standardabweichung $\sigma_y = \sigma_x = \sigma$ und $c = (t\sigma)^2$

$$P(X \geq c) = P\{[Y - E(Y)]^2 \geq (t\sigma)^2\} \leq \frac{\sigma^2}{t^2\sigma^2} = \frac{1}{t^2}, \text{ was äquivalent ist mit Gl.7.8.}$$

b) Tschebyscheffsche Ungleichung*)

X sei eine Zufallsgröße, c eine gegebene Konstante und $\varepsilon > 0$. Dann ist:

$$(7.8) \quad P\{|X - c| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} E(X - c)^2$$

wenn über die Verteilung $f(x)$ nur bekannt ist, dass $E(X)$ und die Varianz $V(X)$ endlich sind. Setzt man $c = \mu$, dann ist $E(X - c)^2 = \sigma^2$ und

$$(7.9) \quad P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}. \text{ Wird schließlich } \varepsilon = t\sigma \text{ gesetzt, so erhält man}$$

$$(7.9a) \quad P\{|X - \mu| \geq t\sigma\} \leq \frac{1}{t^2} \quad \text{und für die Gegenwahrscheinlichkeit:}$$

$$(7.10) \quad P\{|X - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2} \quad \text{und} \quad (7.10a) \quad P\{|X - \mu| \geq t\sigma\} \leq 1 - \frac{1}{t^2}$$

Man kann in der Klasse der Zufallsvariablen, deren Verteilung nicht bekannt ist und deren Momente zweiter Ordnung existieren, keine bessere Abschätzung erzielen.

c) weitere Ungleichungen

1. Satz von Cantelli (einseitige Tschebyscheffsche Ungleichung)

$$(7.11) \quad P(X \geq \mu + t\sigma) = P(x \leq \mu - t\sigma) \leq \frac{1}{1 + t^2}$$

2. Tschebyscheffsche Ungleichung bei zweidimensionaler Zufallsvariable

*) oder Ungleichung von Bienaymé-Tschebyscheff.

Gegeben sei die Zufallsvariable (X, Y) mit den Erwartungswerten μ_x, μ_y und den Standardabweichungen σ_1, σ_2 sowie dem Korrelationskoeffizienten ρ , dann gilt

$$(7.12) \quad P(|X - \mu_x| \geq t\sigma_x \text{ oder } |Y - \mu_y| \geq t\sigma_y) \leq \frac{1}{t^2} - \frac{\sqrt{1-\rho^2}}{t^2} .$$

4. Gesetze der großen Zahlen

Mehr aus historischen Gründen werden die Gesetze der großen Zahlen oft getrennt von den später gefundenen Grenzwertsätzen (Grenzverteilungssätzen) behandelt, aus denen sie als Folgerungen abgeleitet werden können. Das Gesetz der großen Zahl (oder "der großen Zahlen") in seinen verschiedenen Varianten ist eine Aussage über das Konvergenzverhalten von Summen und Durchschnitten.

a) Konvergenzverhalten von Summen

Erwartungswert und Varianz von (gewogenen) Summen und damit auch von Mittelwerten lassen sich herleiten aus den Sätzen zur Lineartransformation von Zufallsvariablen. Mit $E(X_i) = \mu_i$ und $V(X_i) = \sigma_i^2$ ($i = 1, \dots, n$) erhält man für den Erwartungswert der ungewogenen Summe $Y = X_1 + X_2 + \dots + X_n$

$E(Y) = \mu_1 + \mu_2 + \dots + \mu_n$ und für die Varianz $V(Y) = \sum \sigma_i^2$ bei Unabhängigkeit der ZVn.

Entsprechend gilt für das arithmetische Mittel $\bar{X} = \frac{1}{n} X$ erhält man $E(\bar{X}) = \frac{\sum \mu_i}{n}$ und bei

Unabhängigkeit $V(\bar{X}) = \frac{\sum \sigma_i^2}{n^2}$.

Bei unabhängig identisch verteilten Zufallsvariablen X_i (also $\mu_i = \mu$ und $\sigma_i = \sigma$ für alle $i = 1, 2, \dots, n$) gilt dann:

$$(7.13) \quad E(\bar{X}) = \mu \quad \text{und} \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

Die entsprechenden Formeln für den homograden Fall sind hieraus als Spezialfall zu entwickeln (Übers. 7.1).

b) Bernoullis Gesetz der großen Zahlen (homograde Theorie)

Die Zufallsvariablen seien identisch unabhängig zweipunktverteilt mit dem Erwartungswert π und der Varianz $\pi(1-\pi)$. Dann strebt bei n wiederholten unabhängigen Versuchen die relative Häufigkeit $P = \frac{X}{n}$ der Erfolge gegen die Wahrscheinlichkeit π wegen Gl. 7.10.:

$$(7.14) \quad P\left\{\left|\frac{X}{n} - \pi\right| < \varepsilon\right\} \geq 1 - \frac{\pi(1-\pi)}{n\varepsilon^2} = 1 - \frac{\sigma^2}{n\varepsilon^2} .$$

Man bezeichnet als schwaches*) Gesetz der großen Zahl (oder Theorem von Bernoulli) die Folgerung

*) Das starke Gesetz (Konvergenz mit Wahrscheinlichkeit 1, "fast sichere Konvergenz"), ist demgegenüber die Aussage

$$(7.15) \quad \lim_{n \rightarrow \infty} P \left\{ \left| \frac{X}{n} - \pi \right| < \varepsilon \right\} = 1$$

bzw. weil X/n die relative Häufigkeit P ist: $\text{plim}(P) = \pi$.

Voraussetzungen:

- 1) Unabhängige Beobachtungen
- 2) aus identisch verteilten Grundgesamtheiten
- 3) mit endlichen Momenten μ und σ^2 (in diesem Fall $\mu = \pi$ und $\sigma^2 = \pi(1-\pi)$).

Gilt 1 nicht: Satz von Markoff.

Gilt 2 nicht: Satz von Poisson (unterschiedliche Erwartungswerte π_1, \dots, π_2) dann ist

$$\text{plim} \frac{X}{n} = \frac{1}{n} \sum \pi_i \quad (i = 1, \dots, n).$$

c) Satz von Ljapunoff (heterograde Theorie)

Sind X_1, X_2, \dots, X_n unabhängig identisch verteilte Zufallsvariablen mit $\mu_i = \mu$ und $\sigma_i^2 = \sigma^2$, dann gilt für den Mittelwert der Stichprobe (Stichprobenmittel)

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{mit } E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n} \quad \text{gem. Gl. 7.13.}$$

Aus der Tschebyscheffschen Ungleichung

$$(7.10^*) \quad P\left\{ \left| \bar{X}_n - \mu \right| < \varepsilon \right\} \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \quad \text{folgt dann als Satz von Ljapunoff:}$$

$$(7.17) \quad \lim_{n \rightarrow \infty} P \left\{ \left| \bar{X}_n - \mu \right| < \varepsilon \right\} = 1, \quad \text{also also } \text{plim}(\bar{X}_n) = \mu.$$

Gl. 7.15 kann man als Spezialfall von 7.17 auffassen. Hinsichtlich der drei Voraussetzungen gelten die Bemerkungen unter b (Theorem von Bernoulli). Aus Gl. 7.17 folgt zwar, dass \bar{X} stochastisch konvergiert gegen μ , man weiß aber nicht bei welchem Stichprobenumfang n das Stichprobenmittel \bar{X} einen genügend guten Näherungswert für μ darstellt. Hierfür liefert die Tschebyscheffsche Ungleichung (Gl. 7.10*) eine großzügige Abschätzung und der Satz von Lindeberg-Lévy (Gl. 7.18) eine wesentlich bessere Abschätzung. Das gilt entsprechend im homograden Fall (Gl. 7.14 und 7.18a).

5. Grenzwertsätze

Bei hinreichend großem n (als Faustregel $n > 30$) ist die asymptotische Verteilung der Schätzfunktion (vgl. Übersicht 7.1, 7.2, Abb. 7.1 bis 7.3)

- im homograden Fall (Satz von de Moivre-Laplace)

$$X = \sum X_i \quad (\text{Anzahl der Erfolge, } i = 1, 2, \dots, n)$$

$$(7.16) \quad P \left\{ \lim_{n \rightarrow \infty} \left(\frac{X}{n} = \pi \right) \right\} = 1 .$$

$$P = Y/n \quad (\text{Anteil der Erfolge})$$

- im heterograden Fall (Satz von Lindeberg-Lévy)

$$Y = \sum X_i \quad (\text{Merkmalssumme})$$

$$\bar{X} = Z/n \quad (\text{Stichprobenmittelwert})$$

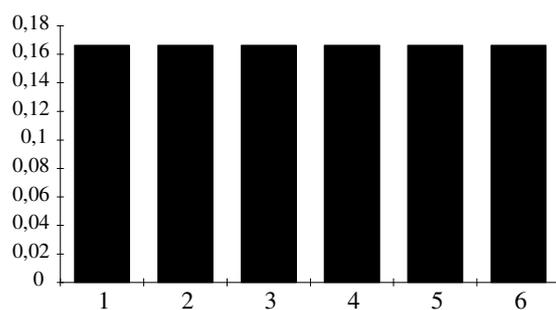
jeweils eine Normalverteilung (unter den drei Voraussetzungen des Theorems von Bernoulli und ZmZ) mit den in Übersicht 7.2 genannten Parametern. Mit der z-Transformation erhält man entsprechend jeweils eine $N(0,1)$ -verteilte ZV.

Die Normalverteilttheit der Summe Y (auch wenn die X_i nicht normalverteilt sind) läßt sich leicht veranschaulichen am Beispiel der Augenzahl beim Würfeln (vgl. Abb. 7.1 bis 3).

Die Grundgesamtheit (Verteilung von X_1, X_2, \dots usw. allgemein X_i): ist jetzt eine (diskrete) Gleichverteilung vgl. Abb. 7.1. An der Verteilung der Augensumme und der durchschnittlichen Augenzahl bei unterschiedlich vielen Würfeln ist die asymptotische Normalverteilung schnell zu erkennen (vgl. Abb. 7.2 und 7.3).

Abb. 7.1 bis 7.3: Veranschaulichung des Grenzwertsatzes von Lindeberg-Lévy

Abb. 7.1: Verteilung der Grundgesamtheit



Parameter

$n = 1$ (Grundgesamtheit)

$\mu = 3,5 \quad \sigma^2 = 2,167$

Abb. 7.2: Augensumme bei $n = 1, 2, \dots$ mal Würfeln (Darstellung

Wahrscheinlichkeitsverteilungen als Polgonzug $X = \sum X_i \sim N(n\mu, n\sigma^2)$)

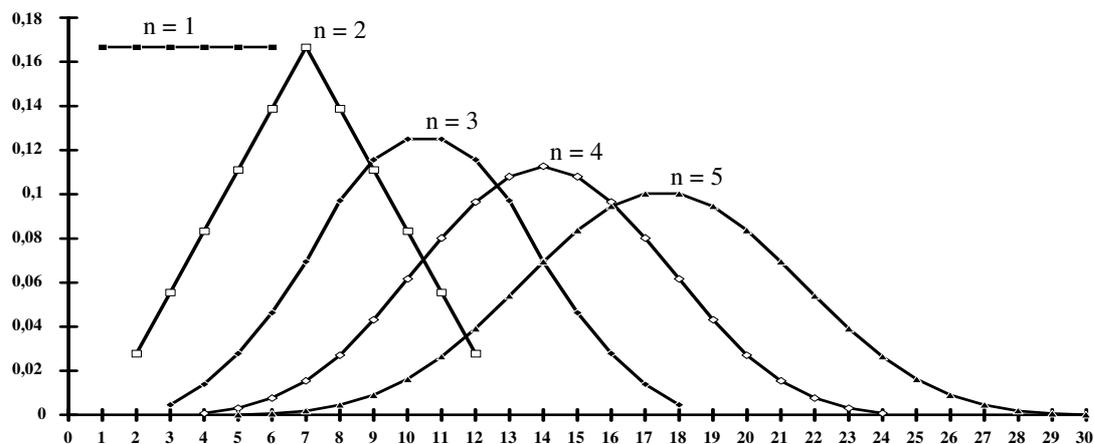
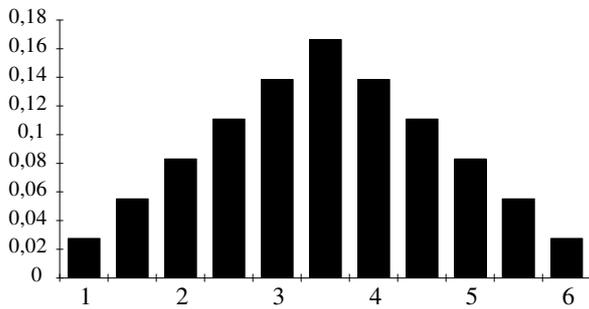


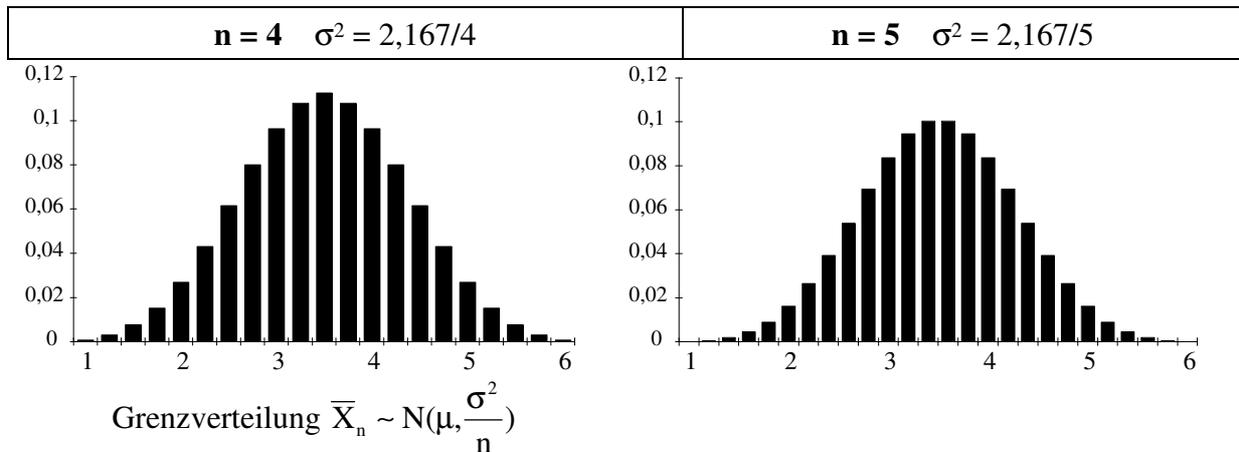
Abb. 7.3: **Mittlere** Augenzahl \bar{X}_n beim n maligen Werfen eines Würfels.
(Wahrscheinlichkeitsverteilungen als Balkendiagramme)



Durchschnittliche Augenzahl
beim zweimaligen Werfen eines
Würfels

$$n = 2 \quad \sigma^2 = 2,167/2$$

Dreiecksgestalt der Verteilung
(Simpson-Verteilung)



Der Satz von Lindeberg-Lévy setzt unabhängig identische ZV'n voraus, d.h. für alle $i = 1, 2, \dots, n$ gilt, dass X_i (beliebig) verteilt ist mit Erwartungswert μ und Varianz σ^2 . Die ZV'n X_i müssen nicht normalverteilt sein (vgl. das Beispiel der Augensumme beim Werfen von n Würfeln [vgl. Abb. 7.2 auf der vorangegangenen Seite]).

Sind die ZV'en normalverteilt, dann ist für jedes n (nicht erst für $n \rightarrow \infty$) die Summe $Y = \sum X_i$ bzw. das Mittel \bar{X} normalverteilt, weil die Normalverteilung reproduktiv ist.

Übersicht 7.2: Grenzwertsätze

Problem	homograd (Anzahl X und Anteil P der "Erfolge")
Grenzwertsatz von	de Moivre-Laplace

Schätzfunktion $\hat{\theta}$	Erwartungswert $\mu_{\hat{\theta}}$	Varianz $\sigma_{\hat{\theta}}^2$	N(0,1) verteilt ist*
$X = \sum X_i$ $P = X/n$	$n\pi$ π	$n\pi(1-\pi)$ $\pi(1-\pi)/n$	$Z = \frac{P - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$

* Das ist die asymptotische Verteilung (Grenzverteilung) nach dem entsprechenden Grenzwertsatz von...

noch Übers. 7.2

Problem	heterograd (Mittel- und Erwartungswerte)
Grenzwertsatz von	Lindeberg-Lévy

Schätzfunktion $\hat{\theta}$	Erwartungswert $\mu_{\hat{\theta}}$	Varianz $\sigma_{\hat{\theta}}^2$	N(0,1) verteilt ist
$Y = \sum X_i$ $\bar{X} = X/n$	$n\mu$ μ	$n\sigma^2$ σ^2/n	$Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$

Verallgemeinerung ohne die Voraussetzung identischer Verteilungen: **Zentraler Grenzwertsatz (von Ljapunoff)**

Unter sehr allgemeinen, praktisch immer erfüllten Bedingungen sind Summen und Durchschnitte von unabhängigen ZVn für große n angenähert normalverteilt.

Bedeutung der Grenzwertsätze

Bei kleinem N und n kann man die exakte Stichprobenverteilung einer Schätzfunktion (wie \bar{X} oder S^2 usw.) durch Auflisten aller möglichen Stichproben gewinnen. Die Anzahl möglicher Stichproben ist jedoch schnell enorm und die N Elemente der Grundgesamtheit sind i.d.R. nicht bekannt, so dass dieser Weg in der Praxis nicht gangbar ist. Durch die Grenzwertsätze ist jedoch die Gestalt der Stichprobenverteilung asymptotisch zumindest dann bekannt, wenn n hinreichend groß ist. Das rechtfertigt insbesondere die Benutzung der Normalverteilung in Kap. 8 und 9.

Kapitel 8: Schätztheorie

1. Einführung in die schließende Statistik (Induktive Statistik)	73
2. Punktschätzung	77
3. Intervallschätzung: Mittel- und Anteilswerte, eine Stichprobe	82
4. Intervallschätzung: andere Stichprobenfunktionen, eine Stichprobe	83
5. Intervallschätzung: Differenz von Mittel- und Anteilswerten, zwei Stichproben	85
6. Weitere Intervallschätzungsprobleme	87

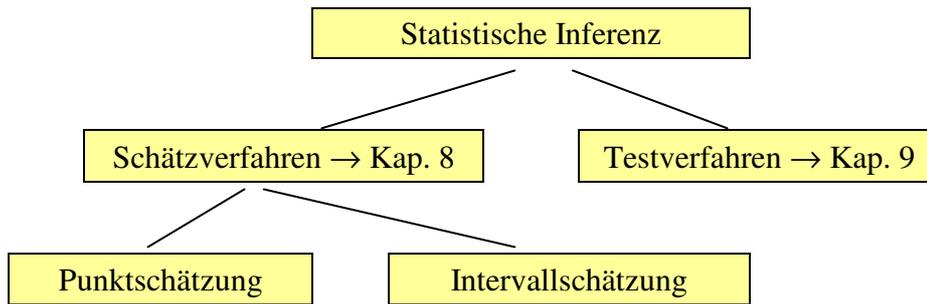
Gegenstand der Induktiven Statistik sind Schlüsse von den mit einer Stichprobe gegebenen Daten auf die der Stichprobe zugrundeliegende Grundgesamtheit. Deshalb ist einleitend die Art des Schließens und die Besonderheit der Stichprobe als Teilerhebung darzustellen. Vom praktischen Standpunkt gesehen sind die Kapitel 8 ff. das eigentliche Ziel der Beschäftigung mit Statistik II, und die Kapitel 1 bis 7 haben demgegenüber einen einleitenden, theoretisch fundierenden Charakter.

1. Einführung in die schließende (induktive) Statistik

a) Schlussweisen: Terminologie

Bei dem Schluss von einer Stichprobe auf die Grundgesamtheit kann man zwei Arten von Schlussweisen unterscheiden: Schätzen und Testen (vgl. Übers. 8.1, Def. 8.1)

Übersicht 8.1:



Terminologie zur Intervallschätzung

	direkter Schluss	indirekter Schluss
Name des Intervalls	Schwankungs- ^{a)} , Prognose- oder Toleranzintervall ^{b)}	Konfidenz-, Vertrauens- oder Mutungsintervall
$1 - \alpha$	Sicherheitswahrscheinlichkeit oder Prognosewahrscheinlichkeit	Sicherheitswahrscheinl.t oder Vertrauens- oder Konfidenzniveau (oder -grad)
α	Irrtumswahrscheinlichkeit oder beim Testen: Signifikanzniveau	

^{a)} Der Begriff wird auch benutzt für ein Intervall auf der x -Achse (meist symmetrisch um μ) der $N(\mu, \sigma^2)$ -Verteilung statt auf der \bar{x} -Achse bei der Stichprobenverteilung von \bar{x} , also der Verteilung $N(\mu, \sigma^2/n)$.

^{b)} Dieser Begriff wird auch anders gebraucht.

Def. 8.1: Statistische Inferenz

- Der Schluss von der bekannten Stichprobe auf die Grundgesamtheit heißt indirekter Schluss, Rück- oder Repräsentationsschluss. Der umgekehrte Schluss von der bekannten oder hypothetisch angenommenen Grundgesamtheit auf eine zu ziehende Stichprobe heißt direkter Schluss oder Inklusionsschluss.
- Aufgabe einer Schätzung ist die (näherungsweise) Bestimmung von Parametern der Grundgesamtheit oder der Verteilung der Variable(n) in der Grundgesamtheit. Die Bestimmung eines einzelnen Schätzwertes (vgl. Def. 7.4) nennt man Punktschätzung. Die Intervallschätzung liefert ein Intervall, in dem die zu schätzende Größe mit einer bestimmten Wahrscheinlichkeit $1 - \alpha$ erwartet wird.

Das Konfidenzintervall (indirekter Schluss) für den unbekannt Parameter θ kann einseitig

$(\theta_u \leq \theta < \infty$ oder $-\infty < \theta \leq \theta_o)$ oder zweiseitig $\theta_u \leq \theta \leq \theta_o$ und dann meist symmetrisch $(\theta_o - \theta = \theta - \theta_u = e)$ sein. Die Unter- (θ_u) und Obergrenze (θ_o) heißen Konfidenz- oder Vertrauensgrenzen. Für den direkten Schluss gilt dies entsprechend hinsichtlich $\hat{\theta}_u$ und $\hat{\theta}_o$.

- Ein Verfahren mit dem über die Annahme oder Ablehnung einer Hypothese aufgrund der Stichprobenverteilung einer Prüfgröße entschieden wird, heißt Testverfahren. Hypothesen beziehen sich auf Parameter oder Verteilungen der Grundgesamtheit.

Hypothesen sind Annahmen über die Wahrscheinlichkeitsverteilung, die dem Stichprobenbefund zugrunde liegt. Man unterscheidet Parameter- und Verteilungshypothesen und danach Parameter- und Anpassungstests. Beziehen sich die Hypothesen auf einen festgelegten Verteilungstyp, so spricht man von parametrischen Tests.

Übersicht 8.2:

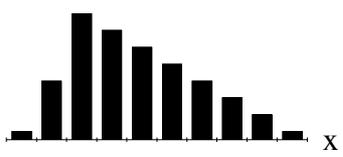
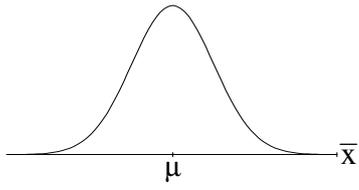
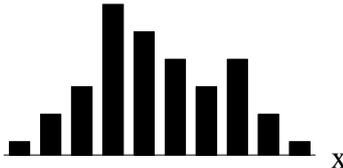
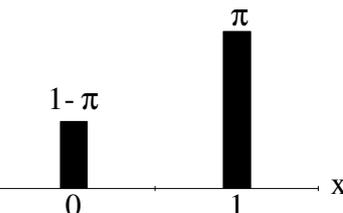
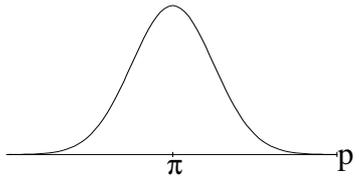
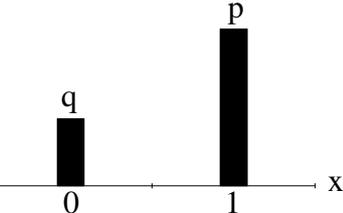
	Grundgesamtheit ^{a)}	Stichprobe(n) ^{b)}
allgemeine Terminologie	Parameter θ	Kennzahl (Schätzer) $\hat{\theta}$
Beispiele	Mittelwert μ (endl. GG) oder Erwartungswert $\mu = E(X)$	\bar{x} (Stichprobenmittelwert)
1. Mittelwert (heterograd)		
2. Anteilswert (homograd)	π Zweipunktverteilung $Z(\pi)$	p (Anteil in der Stichprobe)
3. Varianz	σ^2 (bzw. homog. $\sigma^2 = \pi(1-\pi)$)	$\hat{\sigma}^2$ oder s^2
4. Mittelwertdifferenz (heterograd)	$\mu_1 - \mu_2$ (Mittel- oder Erwartungswertdifferenz der GG)	$\bar{x}_1 - \bar{x}_2$ (Mittelwerte der Stichpr., Umfänge n_1 und n_2)
5. Korrelation	ρ (Korrel. zwischen X und Y)	r (Stichprobenkorrelation)

a) endliche (Umfang N) oder unendliche Grundgesamtheit

b) **vor** Ziehung einer Stichprobe einer Zufallsvariable (Schätzfunktion vgl. Def. 7.4) $\hat{\theta} = f(X_1, X_2, \dots, X_n)$; **nach** Ziehung eine Realisation dieser Zufallsvariable (ein Funktionswert der Stichprobenfunktion) $\hat{\theta} = f(x_1, x_2, \dots, x_n)$ (entsprechend \bar{X} und \bar{x} usw.).

Übersicht 8.3:

*Grundgesamtheit, Stichprobe und Stichprobenverteilung
(Stichproben vom Umfang n mit Zurücklegen)*

Grundgesamtheit	Stichprobenverteilung	eine konkrete Stichprobe
<p>a) heterograd</p>  <p>N Elemente, wenn GG endl., Mittelwert μ</p>	 <p>Normalverteilung nach Grenzwertsätzen (Lindeberg-Lévy); alle möglichen Stichproben</p>	<p>Häufigkeitsverteilung</p>  <p>n Elemente; Mittelwert $\bar{x} = \hat{\mu}$</p>
<p>b) homograd</p>  <p>Zweipunktverteilung</p>	 <p>relativierte Binomialverteilung bzw. asymptotische Normalverteilung</p>	 <p>bei $X = x$ Erfolge ein Anteil von $\hat{\pi} = p = x/n$</p>

Übersicht 8.4:
Zur Interpretation der Intervallschätzung

	Schwankungsintervall	Konfidenzintervall
Intervall um	$\hat{\theta}$ (Beispiel: \bar{x})	θ (Beispiel μ)
Wahrscheinlichkeit	$P(\hat{\theta}_u \leq \hat{\theta} \leq \hat{\theta}_o) = 1 - \alpha$	$P(\theta_u \leq \theta \leq \theta_o) = 1 - \alpha$
Zufallsvariable (n) ist (sind)	die Variable $\hat{\theta}$	die Grenzen θ_u und θ_o des Intervalls (abhängig von $\hat{\theta}$) ^{a)}
feste Größen (nicht zufällig) ist (sind)	$\hat{\theta}_u$ und $\hat{\theta}_o$ (Unter- / Obergrenze als Funktionen u.a. von θ)	θ "liegt" nicht zwischen θ_u und θ_o sondern ist ein fester ^{b)} Wert
"Mit Wahrscheinlichkeit $1 - \alpha$ "	liegt der in der Stichpr. zu erwartende Wert $\hat{\theta}$ zwisch. $\hat{\theta}_u$ u. $\hat{\theta}_o$ "	überdeckt das Intervall $[\theta_u, \theta_o]$ den wahren Parameter θ "
$1 - \alpha$ (Sicherheitswahrscheinl.)	Wahrscheinlichkeit des Enthaltenseins von $\hat{\theta}$ im Intervall	Wahrscheinlichkeit der Überdeckungs des Parameters θ durch das Intervall

- a) Das Konfidenzintervall ist ein Paar von Zufallsvariablen (die Konfidenzgrenzen) und ein (aufgrund der Stichprobenwerte) errechnetes konkretes Konfidenzintervall ist eine Realisation des Zufallsintervalls.
b) aber unbekannter

b) Stichprobenfehler und Stichprobenplan

Kennzeichnend für eine Stichprobe ist die Zufallsauswahl. Ihr ist es zu verdanken, dass die in Abschnitt a dargestellten Schlussweisen wahrscheinlichkeitstheoretisch fundiert sind. Im Mittelpunkt steht dabei die i.d.R. aus Grenzwertsätzen hergeleitete Stichprobenverteilung.

Def. 8.2: Stichprobenfehler

Die Standardabweichung $\sigma_{\hat{\theta}}$ der Stichprobenverteilung von $\hat{\theta}$ als Schätzer für θ ist der Stichprobenfehler (oder auch Standardfehler). Er ist in Verbindung mit dem von der Sicherheitswahrscheinlichkeit $1 - \alpha$ abhängigen Multiplikator z_{α} ein Maß für die Genauigkeit von Stichproben (vgl. Kap. 10) und bestimmt die Breite eines Konfidenz- oder Schwankungsintervalls.

Def. 8.3: Stichprobenplan, einfache Stichprobe

- a) Die Art der Entnahme von Elementen der Grundgesamtheit (GG) heißt Stichprobenplan.
b) Der einfachste, zunächst ausschließlich behandelte Stichprobenplan, ist die einfache (oder: reine) Zufallsauswahl (Stichprobe) bei gleicher Auswahlwahrscheinlichkeit und unabhängiger Ziehung (vgl. Bem. 3)

Bemerkungen zu Def. 8.3:

- Die Größe der Stichprobenfehler hängt u.a. vom Stichprobenplan ab. In den Kapiteln 8 und 9 wird allein die einfache Stichprobe behandelt. Im Kap. 10 werden auch Stichprobenpläne bei Ausnutzung von Kenntnissen über die Beschaffenheit der GG behandelt (vgl. Übers. 8.5).

2. Beispiel der Abhängigkeit des Stichprobenfehlers $\sigma_{\bar{x}}$ des arithmetischen Mittels vom Stichprobenplan. Er ist bei

a) uneingeschränkter Zufallsauswahl und

1. Ziehen ohne Zurücklegen von $n < N$ Elementen aus einer Grundgesamtheit des Umfangs N

$$(8.1) \quad \sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}}, \text{ bzw.}$$

2. Ziehen mit Zurücklegen oder $\frac{N-n}{N-1} \approx 1 - \frac{n}{N} \approx 1$ (8.2) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

b) andere Stichprobenpläne:

1. Geschichtete Stichprobe vgl. Gl. 10.12f
2. Klumpenstichprobe vgl. Gl. 10.18 .

3. Die Ziehung eines Elements (einer Einheit) der GG stellt ein Zufallsexperiment dar. Die Ziehung der Einheit i oder der Einheit j sind die Elementarereignisse. Sind diese gleichwahrscheinlich (Laplace-Annahme), so spricht man von uneingeschränkter Zufallsauswahl. Sind darüber hinaus die Ziehungen (Wiederholungen des Zufallsexperiments) unabhängig, wie bei Ziehen mit Zurücklegen, so spricht man von einfacher Zufallsauswahl. Die Stichprobenwerte x_1, x_2, \dots, x_n sind dann Realisationen von unabhängig identisch verteilten Zufallsvariablen X_1, X_2, \dots, X_n .

2. Punktschätzung

a) Eigenschaften von Schätzfunktionen (Gütekriterien)

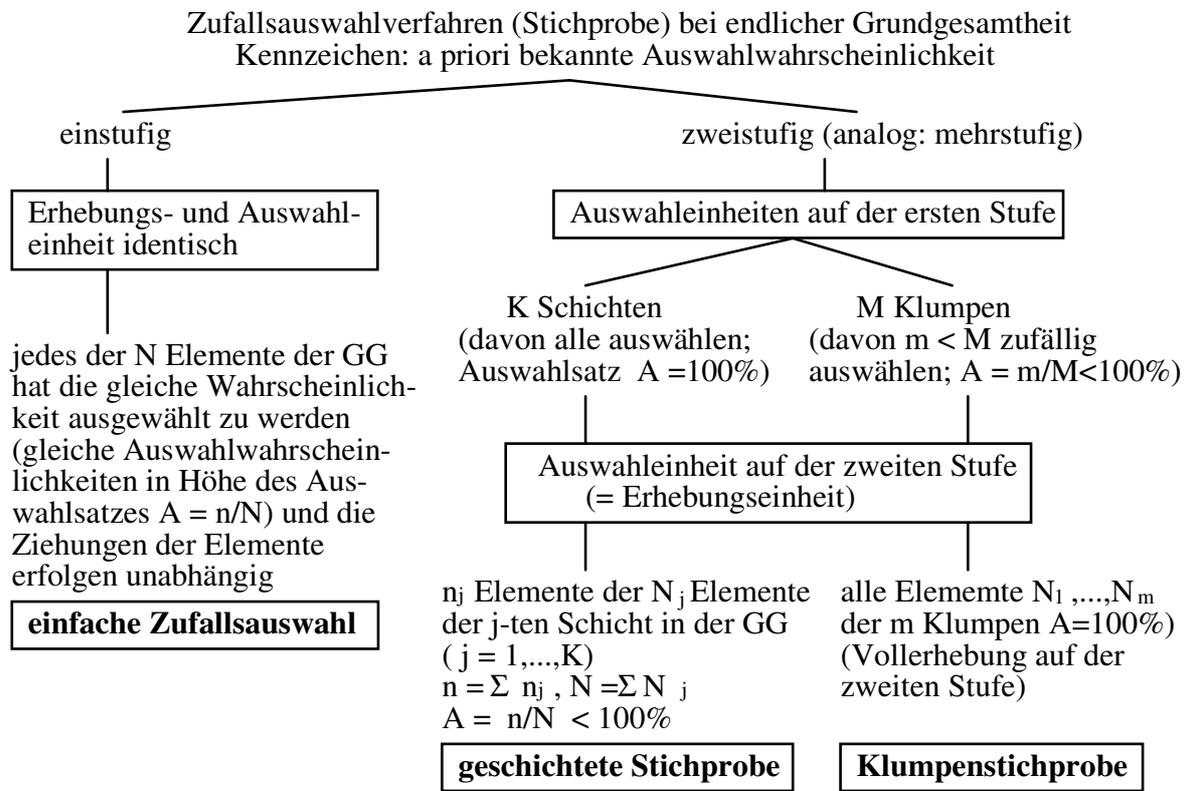
Die Schätzfunktion $\hat{\theta}$ bzw. ihre Stichprobenverteilung, die u.a. vom Stichprobenumfang n abhängig ist, hat bestimmte Eigenschaften (vgl. Übers. 8.6).

So ist z.B. die Stichprobenfunktion $\hat{\mu} = \bar{X} = \frac{1}{n} \sum X_i$ ($i = 1, 2, \dots, n$) zur Schätzung des

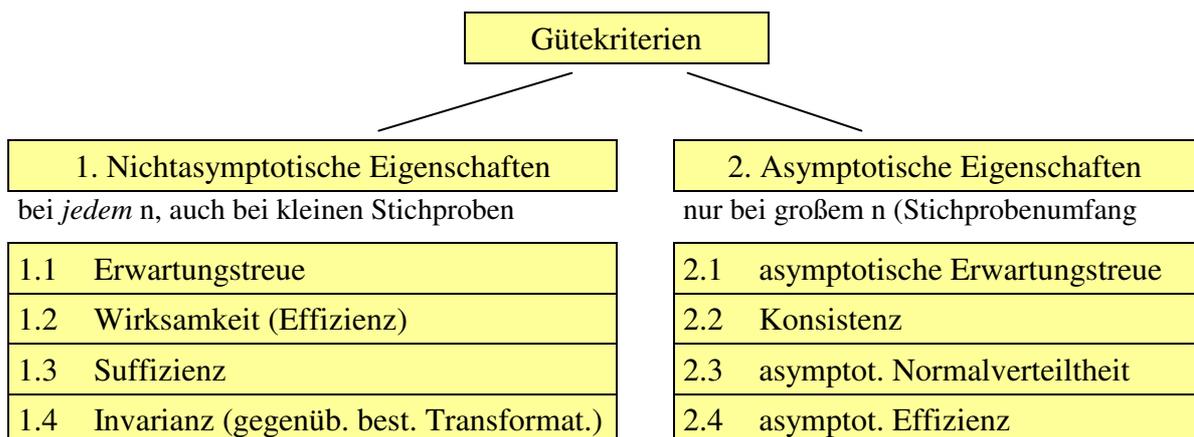
Mittelwertes μ nach dem zentralen Grenzwertsatz (asymptotisch) $N(\mu, \frac{\sigma^2}{n})$ verteilt ($n > 30$)

so dass z.B. auch der Erwartungswert $E(\bar{X}) = \mu$, also μ durch \bar{x} "erwartungstreu" geschätzt wird.

Übersicht 8.5:
Einige einfache Stichprobenpläne



Übersicht 8.6:
Einige Eigenschaften (Gütekriterien) von Stichprobenfunktionen



1. Nichtasymptotische Eigenschaften (bei jedem n)

1.1 Der Schätzer $\hat{\theta}$ ist **erwartungstreu** (biasfrei, unbiased, unverzerrt) für θ , falls $E(\hat{\theta}) = \theta$ auch bei kleinem n. Obgleich ein einzelner Schätzwert $\hat{\theta}$ für eine bestimmte Stichprobe durchaus von θ abweichen kann, weicht $\hat{\theta}$ "im Mittel" nicht von θ ab. Die Abweichung $E(\hat{\theta}) - \theta$ heißt Bias. Wie oben gesagt, ist z.B. das arithmetische Mittel

erwartungstreu $E(\bar{X}) = \mu$, nicht aber die Stichprobenvarianz S^2 , denn $E(S^2) = \frac{n-1}{n}\sigma^2$, weshalb man auch $\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$ als Schätzer für σ^2 verwendet: $E(\hat{\sigma}^2) = \sigma^2$.

1.2 Wirksamkeit (Effizienz)

Bei zwei erwartungstreuen Schätzfunktionen g_θ und g_θ^* (einer Klasse von Schätzfunktionen) für θ ist g_θ^* wirksamer, falls bei gleichem Stichprobenumfang n gilt $V(g_\theta^*) = E(\hat{\theta} - \theta)^2 < V(g_\theta) = E(\hat{\theta} - \theta)^2$; die wirksamste (oder "beste") Schätzfunktion ist die Minimum-Varianz-Schätzung (mit minimalem MSE [Def. 7.6]).

So ist z.B. das arithmetische Mittel der Stichprobe \bar{X} ein besserer Schätzer als der ebenfalls erwartungstreue Median $\tilde{X}_{0,5}$ der Stichprobe weil $V(\tilde{X}_{0,5}) = \frac{\pi}{2} \cdot \frac{\sigma^2}{n} = \frac{\pi}{2} V(\bar{X}) > V(\bar{X})$.

2. asymptotische Eigenschaften (wenn $n \rightarrow \infty$)

2.1 asymptotische Erwartungstreue: $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$

die Stichprobenvarianz S^2 ist nicht erwartungstreu, wohl aber asymptotisch erwartungstreu, weil $\lim_{n \rightarrow \infty} \left(\frac{n-1}{n} \right) = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n} \right) = 1$.

2.2 Konsistenz: $\hat{\theta}$ konvergiert stochastisch gegen θ , also $\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \leq \varepsilon) = 1$ bzw. $\text{plim } \hat{\theta} = \theta$, d.h. die Punktschätzung gehorcht dem Gesetz der großen Zahl.

Dem entspricht die intuitiv überzeugende Forderung, dass mit einer Vergrößerung des Stichprobenumfangs auch eine zuverlässigere (weniger streuende) Schätzung zu erreichen ist. Aus der Zerlegung des MSE (vgl. Def. 7.6)

$$(8.3) \quad E(\hat{\theta} - \theta)^2 = [B(\hat{\theta})]^2 + V(\hat{\theta})$$

mit $B = E(\hat{\theta}) - \theta$, der Bias

und $V(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2$

folgt, dass Konsistenz, d.h. Verschwinden des MSE, also $\lim_{n \rightarrow \infty} E(\hat{\theta} - \theta)^2 = 0$ nicht gelten kann, wenn nicht wenigstens $\lim_{n \rightarrow \infty} [B(\hat{\theta})]^2 = 0$ (asymptotische Erwartungstreue). Konsistenz setzt somit zwar nicht Erwartungstreue voraus, impliziert aber asymptotische Erwartungstreue.

Eine konsistente Schätzfunktion ist stets auch mindestens asymptotisch erwartungstreu. Die Umkehrung des Satzes gilt nicht.

Zur Herleitung der Konsistenz kann auch die Tschebyscheffsche Ungleichung herangezogen werden.

2.3 Asymptotische Normalverteiltheit

Nach dem zentralen Grenzwertsatz ist z.B. der Mittelwert \bar{x} asymptotisch normalverteilt. Es gibt aber auch Schätzfunktionen, bei denen selbst bei großem n die Stichprobenverteilung nicht die Normalverteilung ist, etwa bei r als Schätzer für die Korrelation ρ in der Grundgesamtheit wenn $\rho \neq 0$.

2.4 Asymptotische Effizienz wenn $\lim_{n \rightarrow \infty} \frac{E(\hat{\theta}^* - \theta)^2}{E(\hat{\theta} - \theta)^2} = 1$ und $\hat{\theta}$ die beste Schätzfunktion

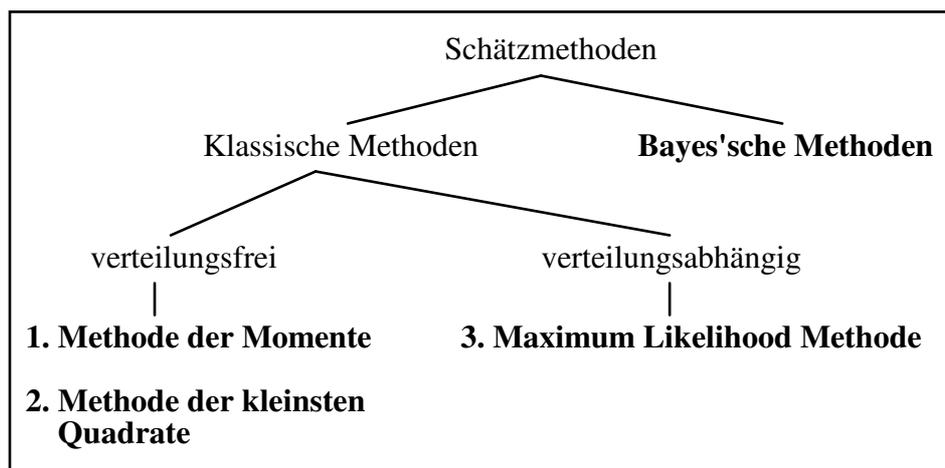
ist, heißt $\hat{\theta}^*$ asymptotisch effektiv.

Man beachte, dass alle Kriterien (ähnlich wie das Konzept des Stichprobenfehlers) Eigenschaften der Stichprobenverteilung, also die Gesamtheit **aller** möglichen Stichproben betreffen. Es gibt kein Kriterium um die Qualität einer Schätzfunktion bei einer einzelnen Stichprobe zu beurteilen. Es ist deshalb auch Unsinn, zu sagen, eine Stichprobe sei "repräsentativ" und die andere nicht (oder weniger)

b) Schätzmethoden

In Übersicht 8.7 sind die wichtigsten Verfahren zur Gewinnung eines Punktschätzers aufgeführt. Auf Bayes'sche Methoden soll hier nicht eingegangen werden.

Übersicht 8.7:



1. Methode der Momente

Bei der Momentenmethode werden Momente wie \bar{x} , s^2 , ... der empirischen Häufigkeitsverteilung der Stichprobe herangezogen als Schätzer für die unbekannt entsprechenden Momente μ , σ^2 , ... der theoretischen Verteilung (als Modell für die Verteilung der Grundgesamtheit).

2. Methode der kleinsten Quadrate (KQ)

Bei Schätzung des Mittel- oder Erwartungswertes μ der GG mit dem Kriterium der kleinsten Quadrate wird ein Schätzwert $\hat{\mu}$ so bestimmt, dass die Summe der Quadrate der

Abweichungen minimal wird. Aus $\frac{d}{d\hat{\mu}} \sum (x_i - \hat{\mu})^2 = 0$ folgt $\hat{\mu} = \bar{x}$, also Identität von KQ-

und Momentenschätzer. Die KQ-Methode spielt v.a. in der Regressionsanalyse eine Rolle.

3. Maximum Likelihood Methode (ML-Methode)

Die Wahrscheinlichkeit, eine bestimmte Stichprobe x_1, x_2, \dots, x_n zu ziehen und damit einen bestimmten Schätzwert etwa $\bar{x} = \sum x_i/n$ zu erhalten ist abhängig von Parametern $\theta_1, \theta_2, \dots$ der Grundgesamtheit, deren Verteilung dem Typ nach (d.h. bis auf eine numerische Spezifizierung der Parameter) als bekannt vorausgesetzt wird. Es liegt dann nahe diese Parameter $\theta_1, \theta_2, \dots$ so zu bestimmen, dass diese Wahrscheinlichkeit maximal ist.

Die Abhängigkeit der Wahrscheinlichkeit $P(x_1, x_2, \dots, x_n)$ von der Gestalt $f(\cdot)$ der Verteilung der Grundgesamtheit und ihren Parametern wird ausgedrückt durch

$$P(x_1, x_2, \dots, x_n) = L(x_1, x_2, \dots, x_n \mid \theta_1, \theta_2, \dots, \theta_p) = L(\mathbf{x} \mid \boldsymbol{\theta})$$

was sich bei unabhängigen (einfachen) Stichproben zu

$$(8.4) \quad L(\mathbf{x} \mid \boldsymbol{\theta}) = f(x_1 \mid \boldsymbol{\theta}) f(x_2 \mid \boldsymbol{\theta}) \dots f(x_n \mid \boldsymbol{\theta})$$

vereinfacht. Diese Funktion heißt **Likelihoodfunktion**.

Der Schätzer $\hat{\theta}_j$ ist eine Maximum-Likelihood-Schätzung von θ_j ($j = 1, 2, \dots, p$), wenn die Likelihood-Funktion Gl. 8.4 ein Maximum annimmt bzw. die logarithmierte Likelihood-Funktion (da die In-Transformation monoton ist, d.h. dass sich die Lage des Maximums nicht verändert). ML-Schätzer $\hat{\theta}_1, \dots, \hat{\theta}_p$ sind diejenigen Werte für $\theta_1, \dots, \theta_p$ in

$$(8.6) \quad \ln L = \sum \ln f(x_i | \theta_1, \theta_2, \dots, \theta_p) \text{ bei denen } \ln L \text{ maximal ist. Der Quotient}$$

$$(8.7) \quad \lambda = \frac{L^*}{L} \text{ heißt Likelihood-Verhältnis (likelihood ratio).}$$

c) Schätzfunktionen für Mittel- und Anteilswert und die Varianz

Im folgenden soll unterschieden werden zwischen homograde und heterograde Fragestellung. Dabei interessiert u.a. die Punktschätzung des Mittel- bzw. Erwartungswerts μ der Grundgesamtheit, bzw. im Falle einer dichotomen (zweipunktverteilten) Grundgesamtheit der Anteilswert bzw. die Erfolgswahrscheinlichkeit π . In Übers. 8.8 sind die Schätzfunktionen (Stichprobenfunktionen) und ihre Eigenschaften zusammengestellt. Die Funktion PQ ist als Schätzfunktion für $\pi(1 - \pi)$ ein ML-Schätzer. Offenbar sind $\hat{\sigma}^2$ und der entsprechende Wert $n PQ/(n-1)$ keine Momentenschätzer: Die Stichprobenvarianzen wären nämlich $S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$ und im homograden Fall PQ. Das sind Schätzer, die aber nicht erwartungstreu sind.

Übersicht 8.8:

Punktschätzung von Mittel- bzw. Anteilswert und Varianz

a) Mittel- bzw. Anteilswert, Stichprobe Ziehen **mit Zurücklegen (ZmZ)**

	heterograd	homograd
Schätzfunktion für μ bzw. π	$\hat{\mu} = \bar{X} = \frac{1}{n} \sum X_i$	$\hat{\pi} = P = \frac{X}{n} \quad (X = \sum X_i)$
Eigenschaften der Schätzfunktion	$E(\bar{X}) = \mu, V(\bar{X}) = \frac{\sigma^2}{n}$ $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$	$E(P) = \pi, V(P) = \frac{\pi(1-\pi)}{n}$ $P \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$

b) Varianz

Schätzfunktion für die Varianz	$\hat{\sigma}^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$	$\frac{n}{n-1} P Q \quad \text{mit } Q = 1 - P$
Eigenschaften der Schätzfunktion	$E(\hat{\sigma}^2) = \sigma^2, \text{plim } \hat{\sigma}^2 = \sigma^2$	Erwartungstreu, Konsistenz wie im heterograden Fall*

c) Stichproben aus endlicher Grundgesamtheit **ohne Zurücklegen (ZoZ)**;
Schätzfunktionen \bar{X} und P wie unter a)

Varianz der Schätzfunktion \bar{X} bzw. P für μ bzw. π	$V(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1} < \frac{\sigma^2}{n}$	$V(P) = \frac{\pi(1-\pi)}{n} \frac{N-n}{N-1}$
--	--	---

* $E(PQ) = \frac{n-1}{n} \pi(1-\pi)$, so dass die Schätzfunktion (Stichprobenfunktion) $\frac{n}{n-1} PQ$ (statt PQ) erwartungstreu ist.

3. Intervallschätzung: Mittel- und Anteilswerte, eine Stichprobe

In diesem Abschnitt werden Intervallschätzungen bei einer Stichprobe von bestimmten in der Regel normalverteilten Stichprobenfunktionen behandelt und zwar insbesondere von Mittelwerten. Zu weiteren Intervallschätzungsproblemen vgl. Abschn. 4 bis 6.

a) Allgemeine Formulierung

Die Stichprobenfunktion $\hat{\theta}$ für den Parameter θ habe die Stichprobenverteilung $N(\theta, \sigma_{\hat{\theta}}^2)$ [Normalverteilung als Stichprobenverteilung], dann gilt für ein (zweiseitiges) zentrales Schwankungsintervall (direkter Schluss):

$$(8.8) \quad P(\theta - z \sigma_{\hat{\theta}} \leq \hat{\theta} \leq \theta + z \sigma_{\hat{\theta}}) = 1 - \alpha \quad \text{Schwankungsintervall}$$

Hierbei ist $\sigma_{\hat{\theta}}$ die Standardabweichung der Stichprobenverteilung: (Standardfehler der Schätzung von $\hat{\theta}$) und z ist ein der Sicherheitswahrscheinlichkeit von $1 - \alpha$ zugeordneter Abzissenwert der Standardnormalverteilung. Entsprechend erhält man für ein zweiseitiges symmetrisches Konfidenzintervall

$$(8.9) \quad P(\hat{\theta} - z \sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z \sigma_{\hat{\theta}}) = 1 - \alpha \quad \text{Konfidenzintervall}$$

Analog erhält man für ein einseitiges nur unten (z_u) oder nur oben (z_o) begrenztes Konfidenzintervall

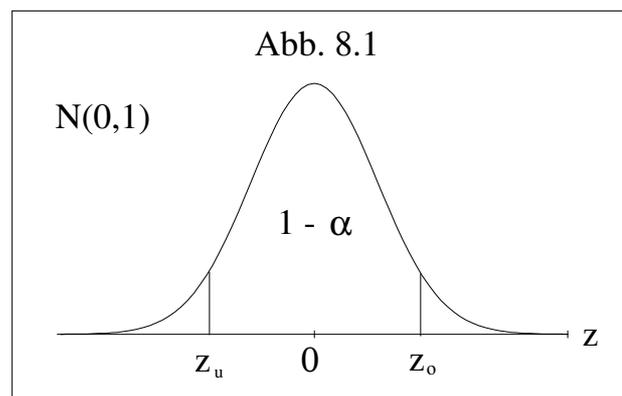
$$P(\hat{\theta} - z_u \sigma_{\hat{\theta}} \leq \theta < \infty) = P(\theta > \hat{\theta} - z_u \sigma_{\hat{\theta}}) = 1 - \alpha \quad \text{oder}$$

$$P(-\infty < \theta \leq \hat{\theta} + z_o \sigma_{\hat{\theta}}) = P(\theta < \hat{\theta} + z_o \sigma_{\hat{\theta}}) = 1 - \alpha$$

Dies folgt aus $\hat{\theta} \sim N(\theta, \sigma_{\hat{\theta}}^2)$, weshalb die standardisierte Variable $Z \sim N(0,1)$ also

$$(8.10) \quad P\left(z_u \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_o\right) = 1 - \alpha.$$

Aus einer Umformung dieser beiden Ungleichungen ergibt sich Gl. 8.8 und 8.9. Zur Veranschaulichung nebenstehende Abbildung:



Im konkreten Fall von Mittel- und Anteilswerten ist der Stichprobenfehler mit einem Abfrageschema (Übers. 8.9) festzustellen und anstelle von $\sigma_{\hat{\theta}}$ die in Übers. 8.9 Teil b angegebene Größe einzusetzen.

Beispiel: Schwankungsintervall für \bar{x} (heterograd) σ bekannt, N sehr groß bzw. ZmZ (Fall 1 in Übers. 8.8), dann ist $\sigma_{\hat{\theta}} = \frac{\sigma}{\sqrt{n}}$ nach Gl. 8.2 in Gl. 8.8 einzusetzen und man erhält

$$P\left(\mu - z_u \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + z_o \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Häufig ist der Stichprobenfehler $\sigma_{\hat{\theta}}$ aus Stichprobenwerten zu schätzen, etwa in $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ die unbekannte Größe σ der Verteilung der Grundgesamtheit durch eine Stichprobenfunktion (etwa $\hat{\sigma}$) zu ersetzen und mit dem geschätzten Stichprobenfehler $\hat{\sigma}_{\hat{\theta}}$ zu rechnen. Damit ist i.d.R. verbunden, dass die Stichprobenverteilung keine Normalverteilung mehr ist.

b) Intervallschätzung von Mittel- und Anteilswerten

Auch in diesem Teil ist wie für die Punktschätzung (Übers. 8.8) zwischen unabhängigen Stichproben aus unendlicher Grundgesamtheit bzw. bei Ziehen mit Zurücklegen (ZmZ) und dem Ziehen ohne Zurücklegen (ZoZ) zu unterscheiden. Die dabei auftretende

Endlichkeitskorrektur (finite multiplier) $\frac{N-n}{N-1} \approx 1 - \frac{n}{N}$ wird i.d.R. bei Auswahlätzen $\frac{n}{N} <$

0,05 vernachlässigt. Die Korrektur führt wegen $\frac{N-n}{N-1} \leq 1$ zu schmaleren Intervallen (besseren

Abschätzungen).

Einzelheiten vgl. Übers. 8.9 (nächste Seite)

4. Intervallschätzung: andere Stichprobenfunktionen, eine Stichprobe

Auf weitere Intervallschätzungsprobleme wird auch in Kap. 9 im Zusammenhang mit dem entsprechenden Parameter-Test hingewiesen. Einige wichtige Schätzungen im Ein- und Zwei-Stichprobenfall sowie Ergänzungen zur Betrachtung von Abschn. 3 werden bereits hier behandelt⁴.

a) Varianz

Die Stichprobenfunktion (Schätzfunktion)

$$(8.11) \quad \frac{(n-1)\hat{\sigma}^2}{\sigma^2} = \frac{nS^2}{\sigma^2} = \sum \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 = c$$

ist χ^2 verteilt mit $n - 1$ Freiheitsgraden (mit n Freiheitsgraden, wenn μ bekannt wäre und nicht durch \bar{x} zu ersetzen wäre). Dem Intervall $c_u \leq c \leq c_o$ entspricht eine

Sicherheitswahrscheinlichkeit von $1 - \alpha$ und da $c_u < c_o$ ist, erhält man eine untere und eine obere Grenze des Konfidenzintervalls

für σ^2 durch Umformung der Ungleichung $c_u \leq \frac{nS^2}{\sigma^2} \leq c_o$ zu $\frac{nS^2}{c_o} \leq \sigma^2 \leq \frac{nS^2}{c_u}$ bzw.

$$(8.11a) \quad \frac{(n-1)\hat{\sigma}^2}{c_o} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{c_u}$$

Das Konfidenzintervall ist nicht symmetrisch um S^2 (bzw. $\hat{\sigma}^2$) weil die χ^2 -Verteilung nicht symmetrisch ist, also $c - c_u \neq c_o - c$.

⁴ Im folgenden werden nur einige ausgewählte Schätzfunktionen behandelt. Eine vertiefere Betrachtung sollte z.B. auch den Median als Schätzfunktion umfassen.

Übersicht 8.9:

*Abfrageschema für Intervallschätzungen von Mittelwerten
(heterograd) und Anteilswerten (homograd)*

a) Fallunterscheidung

Grundgesamtheit: heterograd $X \sim N(\mu, \sigma^2)$
 homograd $X \sim Z(\pi)$ (Zweipunktvert. mit Varianz $V(X)=\pi(1-\pi) \leq 1/4$)

Stichprobe: Umfang n ; Ergebnis heterograd $\bar{x}, \hat{\sigma}^2$; homograd $p, q = 1-p$, im heterograden Fall ist Varianz der GG bekannt

Stichprobe	direkter Schluss		indirekter Schluss	
	heterograd $\hat{\theta} = \bar{x}$	homograd $\hat{\theta} = p$	heterograd $\theta = \mu$	homograd $\theta = \pi$
ZmZ	1	2	3	(4)*
ZoZ	5	6	7	(8)

wenn Varianz der GG nicht bekannt, dann ist $\sigma_{\hat{\theta}}$ durch $\hat{\sigma}_{\hat{\theta}}$ zu schätzen

	indirekter Schluss	
ZmZ	9	10
ZoZ	11	12

*) weniger relevante Fälle sind eingeklammert

b) Stichprobenfehler $\sigma_{\hat{\theta}}$, Fälle nach Teil a

Stichprobenverteilung *): Normalverteilung mit der folgenden Standardabweichung

1	$\frac{\sigma}{\sqrt{n}}$	5	$\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
2	$\sqrt{\frac{\pi(1-\pi)}{n}}$	6	$\sqrt{\frac{\pi(1-\pi)}{n} \cdot \frac{N-n}{N-1}}$
3	wie 1	7	wie 5
4	wie 2	8	wie 6

geschätzter Stichprobenfehler $\hat{\sigma}_{\hat{\theta}}$, Fälle nach Teil a (Stichprobenvert.: t_{n-1} -Verteilung)

9	$\frac{\hat{\sigma}}{\sqrt{n}}$	11	$\frac{\hat{\sigma}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N}}$
10	$\sqrt{\frac{pq}{n-1}} \leq \frac{1}{2\sqrt{n-1}}$	12	$\sqrt{\frac{pq}{n-1} \cdot \frac{N-n}{N}}$

*) Abkürzungen: N = Normalverteilung, t_m = t-Verteilung mit m Freiheitsgraden. Bei $n > 30$ kann von $N(0,1)$ ausgegangen werden.

b) Merkmalssummen und absolute Häufigkeiten

Bei endlicher Grundgesamtheit (Umfang N) ist im heterograden Fall die **Merkmalssumme** (Gesamtmerkmalsbetrag) gegeben mit

$N \cdot \mu$ in der Grundgesamtheit und
 $n \bar{x}$ in der Stichprobe.

Entsprechend erhält man die **Anzahl** (absolute Häufigkeiten oder Bestand) im homograden Fall aus dem Anteil mit

$N\pi$ in der Grundgesamtheit und
 $n\pi$ in der Stichprobe.

Es handelt sich mithin um Lineartransformationen von \bar{x} bzw. p .

Im Falle einer Punktschätzung spricht man auch von **Hochrechnung**⁵ (vgl. Kap. 10); das ist der Schluss $\bar{x} \rightarrow \mu \rightarrow N \cdot \mu$ bzw. $p \rightarrow \pi \rightarrow N\pi$. Für die Intervallschätzung dabei gilt:

Die Grenzen des Intervalls sind

- bei Schwankungsintervallen mit n
 - bei Konfidenzintervallen mit N
- zu multiplizieren.

5. Intervallschätzung: Differenz von Mittel- und Anteilswerten, zwei unabhängige Stichproben

a) heterograder Fall

GG: $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$.

Stichproben vom Umfang n_1 und n_2 ergeben die Mittelwerte \bar{x}_1 und \bar{x}_2 . Die

Stichprobenfunktion $\hat{\Delta} = \bar{x}_1 - \bar{x}_2$ als Schätzer für $\Delta = \mu_1 - \mu_2$ ist normalverteilt $N(\Delta, \sigma_{\hat{\Delta}}^2)$,

wobei für den Fall mit Zurücklegen gilt:

$$(8.17) \quad \sigma_{\hat{\Delta}}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Weitere Fallunterscheidung

	X_1 und X_2 sind unabhängig verteilt mit μ_1 und μ_2 sowie σ_1^2 und σ_2^2 und zwar	
σ_1 und σ_2 sind	normalverteilt	nicht normalverteilt
bekannt	Fall A	Fall C
unbekannt ^{*)}	Fall B	Fall D

^{*)} und $n_1 > 30$ und $n_2 > 30$.

Fall A

1. ungleiche Varianzen:

$\hat{\Delta}$ ist normalverteilt mit $E(\hat{\Delta}) = \Delta$ und Varianz $\sigma_1^2/n_1 + \sigma_2^2/n_2$. Folglich ist

⁵ Der Begriff ist an sich überflüssig und er existiert auch nicht in der englischen Fachliteratur, weil es sich im Prinzip nur um eine Variante der Punktschätzung handelt.

$$(8.18) \quad z = \frac{\hat{\Delta} - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \quad \text{und das Konfidenzintervall lautet}$$

$$(8.19) \quad P\left(\hat{\Delta} - z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \Delta \leq \hat{\Delta} + z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

2. gleiche Varianzen (homogene Varianzen)

Die Varianz der Stichprobenvert. von $\hat{\Delta}$ vereinfacht sich zu $\sigma_{\hat{\Delta}}^2 = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$. Folglich ist

$$(8.20) \quad z = \frac{\hat{\Delta} - \Delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1), \text{ so dass}$$

$$(8.21) \quad P\left(\hat{\Delta} - z_\alpha \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \Delta \leq \hat{\Delta} + z_\alpha \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha$$

das gesuchte Konfidenzintervall ist.

Fall B

σ_1^2 und σ_2^2 sind in Gl. 8.18 durch die Stichproben-Varianzschätzer $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$ zu ersetzen und bei homogenen Varianzen $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (zu überprüfen mit F-Test) ist zu bestimmen

$$(8.22) \quad \hat{\sigma}^2 = \frac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n-2} \quad (n = n_1 + n_2) \quad [\text{pooled sample variance}], \text{ so dass}$$

$$(8.23) \quad t = \frac{\hat{\Delta} - \Delta}{\hat{\sigma}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{\hat{\Delta} - \Delta}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n-2}$$

bzw. die Größe t^2 ist F-verteilt mit 1 und $n - 2$ Freiheitsgraden ($t_2 \sim F_{1, n-2}$).

Fälle C und D: Die Aussagen über den Verteilungstyp der Stichprobenverteilung von z bzw. t gelten nicht mehr bei kleinem n .

b) Homograde Fall

Die Stichprobenverteilung der Schätzfunktion $\hat{\Delta} = P_1 - P_2$ für $\Delta = \pi_1 - \pi_2$ ist asymptotisch normalverteilt (wenn $n_j \pi_j (1 - \pi_j) \geq 9$ bei $j = 1, 2$) mit $E(\hat{\Delta}) = \Delta$ und

$$(8.24) \quad \sigma_{\hat{\Delta}}^2 = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$$

wobei π_j durch p_j ($j = 1, 2$) geschätzt werden kann (dann t-Verteilung mit $n - 2$ Freiheitsgraden)

6. Weitere Intervallschätzungsprobleme

Für das Varianzverhältnis bei zwei unabhängigen Stichproben gilt

$$(8.25) \quad P\left(z_o \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq z_u \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}\right) = 1 - \alpha, \text{ wobei } z_u \text{ und } z_o \text{ das } \alpha/2 \text{ und } 1 - \alpha/2 \text{ Quantil}$$

(Fraktile) der F_{n_1-1, n_2-1} Verteilung sind. Äquivalent ist mit $F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}$

$$(8.25a) \quad P\left(\frac{1}{z_u^*} F \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1}{z_o^*} F\right) = 1 - \alpha, \text{ wenn } z_u^* \text{ (bzw. } z_o^*) \text{ das } \alpha/2\text{- (bzw. } 1 - \alpha/2) \text{ Fraktile}$$

der F_{n_2-1, n_1-1} Verteilung ist.

Gl. 7.10a liefert eine Abschätzung für das Schwankungsintervall eines einzelnen Stichprobenwerts X mit $P(\mu - t\sigma < X < \mu + t\sigma) \geq 1 - \frac{1}{t^2}$. Entsprechend erhält man aus Gl. 7.10*

$P\{|\bar{X}_n - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$ mit $\varepsilon = t\sigma$ das folgende Konfidenzintervall (Schwankungsintervall und Intervalle für Anteilswerte analog)

$$(8.26) \quad P(\bar{X} - t\sigma < \mu < \bar{X} + t\sigma) \geq 1 - \frac{1}{nt^2}.$$

Kapitel 9: Statistische Testverfahren

1.	Allgemeine Einführung	87
2.	Ein-Stichproben-Test	92
3.	Zwei-Stichproben-Test	92
4.	Varianten des χ^2 -Tests	97
5.	Mehr als zwei Stichproben	99

Nach einer allgemein gehaltenen Einführung werden Ein- und Zwei-Stichprobentests für Mittel- (heterograd) und Anteilswerte (homograd) sowie Varianzen behandelt. Die Abschnitte 4 und 5 behandeln auch bivariate Daten mit zwei und mehr Stichproben.

1. Allgemeine Einführung

Def. 9.1: Hypothese, Test

- Eine (statistische) Hypothese ist eine Annahme/Vermutung über die Verteilung der Grundgesamtheit. Als Modell für diese Verteilung dient i.d.R. eine Wahrscheinlichkeitsverteilung $f(x | \theta_1, \dots, \theta_p)$ mit den Parametern $\theta_1, \dots, \theta_p$.
- Ein (statistischer) Test (Hypothesentest) ist ein Verfahren, mit dem auf der Basis einer Prüfgröße (Testgröße) T , die eine Stichprobenfunktion (vgl. Def. 7.4) ist, mit einer vorgegebenen Irrtumswahrscheinlichkeit α (Signifikanzniveau) über die Verwerfung (Ablehnung) oder Annahme (besser: Nichtverwerfung) einer Hypothese entschieden werden kann. Grundlage der Entscheidung ist die bedingte (bei Geltung der Hypothese H_0) Stichprobenverteilung $g(t | H_0)$ der Prüfgröße T , deren Realisationen t genannt werden.

Bemerkungen zu Def. 9.1

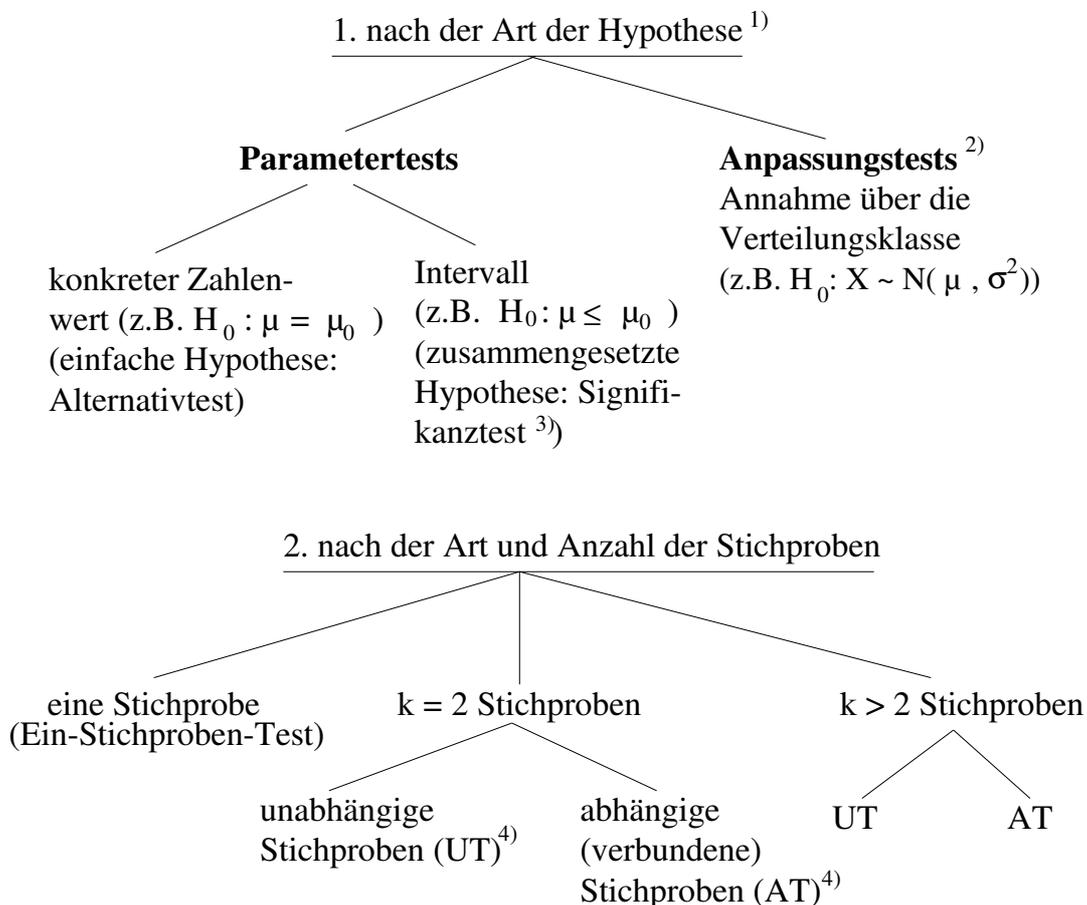
1. Nach Art der Hypothese werden verschiedene Testarten unterschieden (Übers. 9.1)
2. Logik eines Tests:

Man tut so, als ob H_0 richtig ist und prüft, ob dann der Stichprobenbefund noch "im Rahmen der Wahrscheinlichkeit" (der mit Stichproben verbundenen Zufälligkeit) ist oder so wenig wahrscheinlich (weniger wahrscheinlich als α) oder "überzufällig" (oder "signifikant") ist, dass H_0 abgelehnt werden sollte.

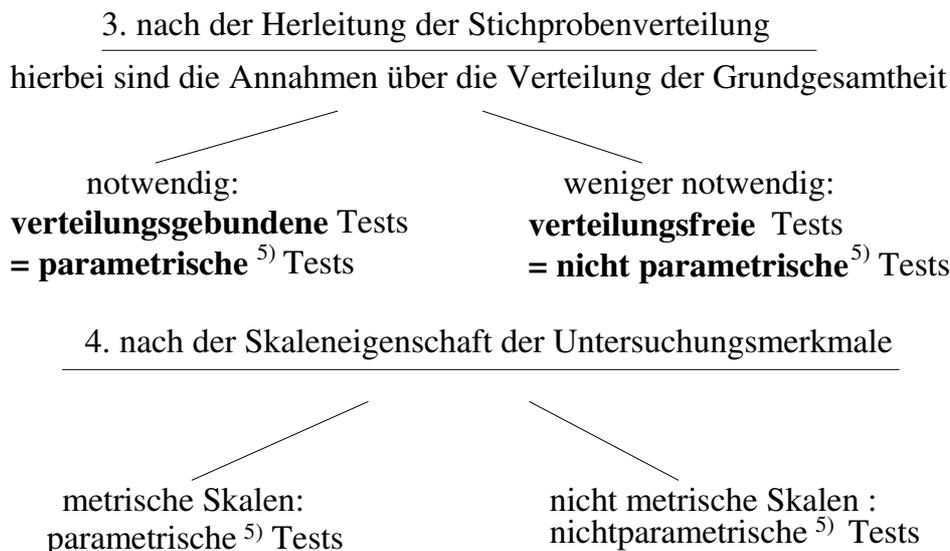
Man beachte, dass wegen der Abhängigkeit der Stichprobenverteilung $g(t | H_0)$ von n die Verwerfung, also ein "signifikantes Ergebnis", auch mit entsprechend vergrößertem Stichprobenumfang zu erzielen ist.

3. Eine Testgröße T ist eine speziell für die Entscheidung geeignete Stichprobenfunktion und damit eine ZV (meist eine standardisierte ZV, die direkt mit den Quantilen einer Wahrscheinlichkeitsverteilung verglichen werden kann).
4. Die möglichst zu verwerfende Hypothese wird Nullhypothese H_0 genannt (Begründung vgl. Gl. 9.2), eine ihr entgegengesetzte Hypothese Alternativhypothese H_1 . Die Funktion $g(t | H_0)$ ist die Stichprobenverteilung der Prüfstatistik T bei Geltung von H_0 .
5. Der Annahme (bzw. Ablehnung) von H_0 äquivalent ist die Situation, dass der Wert (die Realisation) der Stichprobenfunktion $\hat{\theta}$ innerhalb (bzw. außerhalb) eines bei Geltung von H_0 (also im Sinne eines hypothetischen direkten Schlusses) konstruierten Schwankungsintervalls liegt. Während jedoch $\hat{\theta}$ von der Maßeinheit von X abhängt, gilt für dies die spezielle Stichprobenfunktion T nicht.

Übersicht 9.1: Arten von statistischen Tests



noch Übers. 9.1



- 1) In der Literatur werden auch Homogenitäts- (H) und Unabhängigkeitstests (U) unterschieden. Man kann H (stammen mehrere Verteilungen aus der gleichen GG?) und U (statistische Unabhängigkeit?) auch als spezielle Anpassungstests auffassen (Anpassung an Modelle der GG der folgenden Art: $F_1(x) = F_2(x) = \dots = F(x)$ oder $f(x,y) = f_1(x) f_2(y)$ [$f_1 =$ Randverteilung]).
- 2) Auch goodness of fit Test.
- 3) Der Begriff wird auch im allgemeinen Sinne für alle Arten von Tests gebraucht oder auch im Gegensatz zu "Gegenhypotheseentests" .
- 4) Zur Unterscheidung zwischen abhängigen und unabhängigen Stichproben vgl. Def. 9.4.
- 5) Terminologie ist nicht einheitlich.

Def. 9.2: Kritischer Bereich, Entscheidungsregel

- a) Mit der Vorgabe des Signifikanzniveaus α und der Stichprobenverteilung $g(t | H_0)$ ist eine kritische Region (Ablehnungsbereich, Verwerfungsbereich) K_α gegeben mit

$$(9.1) \quad P(t \in K_\alpha | H_0) \leq \alpha .$$

Das Komplement von K_α heißt Annahmebereich (oder Verträglichkeitsbereich, besser: Nicht-Ablehnungsbereich). Das Intervall K_α ist je nach Art von H_1 einseitig oder zweiseitig eingeschränkt. Die Grenzen c_1 und/oder c_2 von K_α heißen kritische Werte.

- b) Mit t_α als das dem vorgegebenen α entsprechende Quantil der Stichprobenverteilung und t als dem mit den Stichprobenwerten errechneten Zahlenwert der Prüfgröße gilt als Entscheidungsregel:

wenn $t \in K_\alpha$, dann H_0 verwerfen,

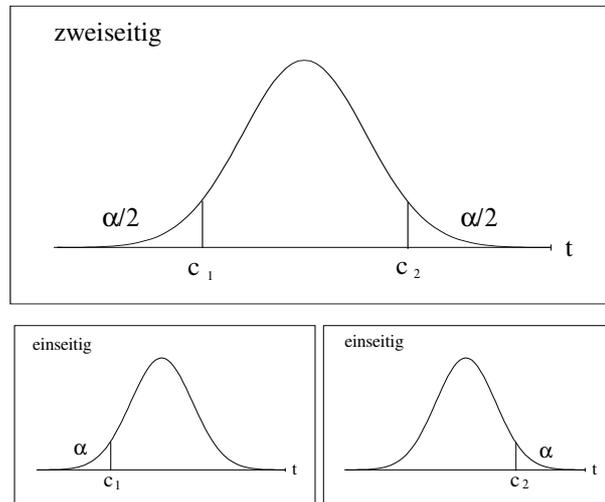
wenn $t \notin K_\alpha$ (also im Annahmebereich), dann H_0 annehmen.

- c) Mit der Annahme von H_0 ist nicht deren Richtigkeit bewiesen. Bei der Testentscheidung können Fehler auftreten (vergl. Übers. 9.2).

Bemerkungen zu Def. 9.2

- Der kritische Bereich K_α ist so konstruiert, dass die Realisation t der Prüfgröße T unter H_0 mit einer Wahrscheinlichkeit von höchstens α in diesen Bereich fällt. "Höchstens" gilt, weil bei einer diskreten Stichprobenverteilung $g(t|H_0)$, etwa der Binominalverteilung als Stichprobenverteilung der Wert von genau α evtl. nicht eingehalten werden kann.
- Ist die Alternativhypothese zu $H_0: \mu = \mu_0$
 $H_1: \mu \neq \mu_0$ (also $\mu > \mu_0$ oder $\mu < \mu_0$) so liegt ein **zweiseitiger** Test vor mit zwei kritischen Bereichen $t < c_1 = t_{\alpha/2}$ und $t > c_2 = t_{1-\alpha/2}$ (Abb. 9.1)
 $H_1: \mu < \mu_0$ **einseitig** "nach unten" (linksseitig) $t < c_1 = t_\alpha$
 $H_1: \mu > \mu_0$ **einseitig** "nach oben" (rechtsseitig) $t > c_2 = t_{1-\alpha}$.
- Der Wert von t_α hängt von der Gestalt der Stichprobenverteilung $g(t|H_0)$ ab. Ist diese die Standard-NV, so ist $t_\alpha = z_\alpha$. Sie kann auch die t -, χ^2 -, F- oder Binomialverteilung sein. Dann ist t_α aus der entsprechenden Tabelle zu bestimmen.
- Fehlerarten** (vgl. Übers. 9.2, nächste Seite)
- Zwei-Stichproben-Test**

Abb. 9.1: Ein- und zweiseitiger Tests



Zu prüfen ist i.d.R., bei zwei Stichproben, ob die beiden Stichproben $x_{11}, x_{12}, \dots, x_{1n_1}$ und $x_{21}, x_{22}, \dots, x_{2n_2}$ aus der gleichen GG stammen. Im heterograden Fall ist die Frage, ob $\bar{x}_1 - \bar{x}_2 \neq 0$ verträglich ist mit der Hypothese "kein Unterschied", also mit

$$(9.2) \quad H_0: \mu_1 - \mu_2 = \Delta = 0 \quad \text{oder} \quad \mu_1 = \mu_2$$

oder ob H_0 zugunsten einer Alternativhypothese

$$(9.3) \quad H_1: \mu_1 - \mu_2 = \Delta_1 \text{ zu verwerfen ist.}$$

Annahme (Ablehnung) von H_0 ist übrigens äquivalent der Überschneidung (Nichtüberschneidung) von Schwankungsintervallen für \bar{X}_1 und \bar{X}_2 bzw. der Überdeckung (Nichtüberdeckung) des Wertes 0 durch das Schwankungsintervall für $\bar{X}_1 - \bar{X}_2$.

Übersicht 9.2: Fehlerarten beim Hypothesentest

Testentscheidung (action)	wirklicher Zustand (state of nature)	
	H_0 ist wahr	H_0 ist falsch
H_0 ablehnen	Fehler 1. Art $P(t \in K_\alpha H_0) \leq \alpha$	kein Fehler $P(t \in K_\alpha H_1) \geq 1 - \beta$
H_0 annehmen	kein Fehler $P(t \notin K_\alpha H_0) \geq 1 - \alpha$	Fehler 2. Art $P(t \notin K_\alpha H_1) \leq \beta$

Fehler 1. Art: H_0 wird verworfen, obwohl die Hypothese richtig ist. Die Wahrscheinlichkeit des Fehlers ist höchstens α .

Fehler 2.Art: H_0 wird nicht verworfen, obgleich H_0 falsch ist. Die Wahrscheinlichkeit dieses Fehlers kann nur bestimmt werden, wenn eine Alternativhypothese (H_1) spezifiziert ist (Gegenhypthesen- oder Alternativtest statt Signifikanztest).

In der Statistischen Entscheidungstheorie werden die Wahrscheinlichkeiten α und β aufgrund der Konsequenzen der entsprechenden Fehlentscheidungen festgelegt.

Def. 9.3: Macht, Gütefunktion

a) Die Wahrscheinlichkeit $1 - \beta$ mit der ein tatsächlich bestehender Unterschied $\theta_1 - \theta_2 = \Delta_1$ (oder z. B. bei Mittelwerten $\mu_1 - \mu_2$) auch erkannt wird, d.h. dass die richtige Hypothese H_1 angenommen wird, heißt Macht (Trennschärfe, power, power efficiency, Güte) eines Tests.

b) Der Zahlenwert für die Wahrscheinlichkeit $1 - \beta$ ist abhängig vom konkreten Wert θ_1

$$H_1: \theta = \theta_1 \quad (\text{Ein-Stichproben-Problem})$$

bzw. von Δ_1

$$H_1: \theta_1 - \theta_2 = \Delta_1 \quad (\text{Zwei-Stichproben-Problem})$$

bei gegebenem α sowie n , bzw. n_1 und n_2 und evtl. von weiteren Grundgesamtheitsparametern.

Der Graph der Funktion

$$(9.4) \quad 1 - \beta = f_G(H_1) = P(t \in K_\alpha \mid H_0)$$

heißt Gütefunktion.

Für die gleiche Fragestellung (das gleiche Auswertungsproblem) und für die gleichen Daten können verschiedene Tests geeignet sein. Es sollte dann der mächtigste Test angewandt werden.

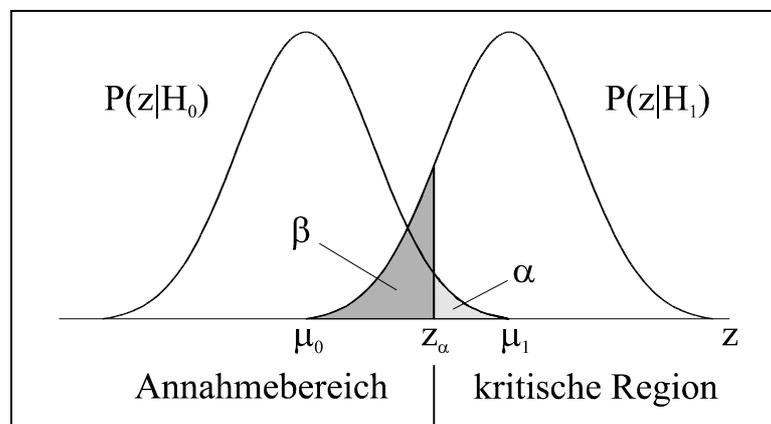
Entscheidend für β und damit für $1 - \beta$ ist es, welchen Wert man für μ_1 annimmt, weil die Stichprobenverteilung von T (bzw. Z) wenn H_1 gilt $g(t \mid H_1)$, also die rechte Kurve in Abb. 9.2, von μ_1 (und anderen Parametern wie z.B. hier σ_1^2) abhängt (vgl. hierzu Abb. 9.2).

Abb. 9.2: Veranschaulichung von α und β

Beispiel: einseitiger Test über den Mittelwert μ (eine Stichprobe)

$$H_0: \mu = \mu_0; \quad H_1: \mu = \mu_1 > \mu_0$$

Prüfgröße T ist $N(0,1)$ - verteilt daher z statt t



2. Ein-Stichproben-Tests (Übers. 9.3)

Übers. 9.3 (nächste Seite) enthält die relevanten Informationen über die Prüfgrößen (jeweils T genannt) und deren Stichprobenverteilung bei Geltung von H_0 zur Durchführung von Tests über

- den Mittelwert μ (heterograd)
- den Anteilswert π (homograd)
- die Varianz σ^2 (heterograd)

der GG aufgrund einer Stichprobe. In diesem Abschnitt werden ausschließlich univariate Ein-Stichproben-Tests behandelt. Tests über Regressions- und Korrelationskoeffizienten (bivariate Daten) können aus Platzgründen nicht behandelt werden. Ein bivariater Datensatz (eine Stichprobe) liegt auch vor bei einem Test auf Unabhängigkeit (Abschn. 4).

3. Zwei-Stichproben-Tests (Übers. 9.4)

Def. 9.4: unabhängige Stichproben

- Wird über die Zugehörigkeit eines Elements der Grundgesamtheit zur Stichprobe 1 oder 2 nach dem Zufallsprinzip entschieden, so liegen unabhängige Stichproben vor.
- Zwei Stichproben gleichen Umfanges $n_1 = n_2$ heißen abhängig, wenn sie die gleichen Elemente enthalten oder strukturell gleich zusammengesetzt sind.

Bedeutung der Zwei-Stichprobenfragestellung

- Für eindeutige Kausalaussagen (z.B. Wirksamkeit eines Medikaments) sind Kontrollgruppenexperimente nötig. Man kann z.B. durch Münzwurf entscheiden, ob eine Einheit in die Experimentgruppe E oder in die Kontrollgruppe K gelangt. E und K sollen sich allein dadurch unterscheiden, dass z.B. die Einheiten von E das zu testende Medikament erhalten, diejenigen von K aber nicht. Die Stichprobenumfänge $n_1 = n_E$ und $n_2 = n_K$ sind i.d.R. nicht gleich.
- Abhängigkeit liegt vor, wenn je ein Element von Stichprobe 1 (E) und von Stichprobe 2 (K) hinsichtlich der für das Experiment relevanten Eigenschaften gleich sind (matched pairs), bzw. (schwächer) wenn die Verteilungen dieser Eigenschaften in E und K gleich sind. Individuelles matching liegt z.B. bei "repeated observations" der gleichen Einheit vor.

Zwei-Stichprobenproblem: Stichprobenverteilung

Stammen die Zufallsvariablen $X_{11}, X_{12}, \dots, X_{1n_1}$ und $X_{21}, X_{22}, \dots, X_{2n_2}$ aus Grundgesamtheiten mit μ_1, σ_1^2 und μ_2, σ_2^2 , so ist die Varianz der Linearkombination

$$\hat{\Delta} = \bar{X}_1 - \bar{X}_2 = \frac{1}{n_1} \sum_i X_{1i} + \left(-\frac{1}{n_2}\right) \sum_j X_{2j} \quad (i = 1, 2, \dots, n_1 \text{ und } j = 1, 2, \dots, n_2) :$$

$$(9.12) \quad V(\hat{\Delta}) = V(\bar{X}_1) + V(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

wenn zwei **unabhängige** Stichproben vorliegen. Bei zwei abhängigen Stichproben ($n_1 = n_2 = n$) ist dagegen auch die Kovarianz $C(\bar{X}_1, \bar{X}_2)$ zwischen \bar{X}_1 und \bar{X}_2 zu berücksichtigen.⁶ Es gilt

$$(9.13) \quad V(\hat{\Delta}) = V(\bar{X}_1) + V(\bar{X}_2) - 2C(\bar{X}_1, \bar{X}_2) = (\sigma_1^2 + \sigma_2^2)/n - ((n-1)/n) \sigma_{12}$$

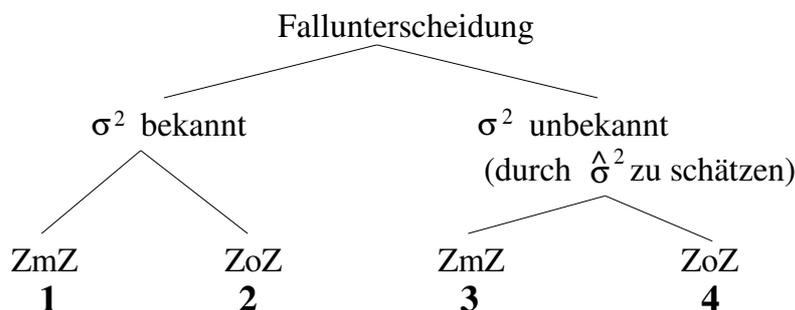
⁶ Das ist die Kovarianz der zweidimensionalen (gemeinsamen) Stichprobenverteilung von \bar{X}_1 und \bar{X}_2 . Sie

beträgt $\frac{1}{n^2} \binom{n}{2} E(X_1 X_2)$ mit $E(X_1 X_2) = \sigma_{12}$

Übers. 9.3: Prüfgrößen im Ein-Stichproben-Fall

a) Heterograd (Test über Mittelwert/Erwartungswert)

$H_0: \mu = \mu_0, X_i \sim N(\mu_0, \sigma^2)$ sonst $n \geq 100$



Fall	Prüfgröße T	Stichprobenverteilung von T
1	(9.5) $\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$	N(0,1)
2	(9.6) $\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}}$	wie 1
3	(9.7) $\frac{\bar{X} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$	t_{n-1} asymptot. N(0,1) wenn $n > 30$
4	(9.8) $\frac{\bar{X} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}} \sqrt{\frac{N-n}{N}}}$	t_{n-1} asymptot. N(0,1) wenn $n > 50$

b) Homograd (Test über Anteilswert/Wahrscheinlichkeit)

$H_0: \pi = \pi_0, X_i \sim Z(\pi)$, Anteilswert $P = X/n, X = \sum X_i$ mit π_0 ist auch die Varianz der zwei-punktverteilten GG hypothetisch angenommen $V(X) = \pi_0(1 - \pi_0)$

Fallunterscheidung: nur Fälle 3 und 4 oben [σ^2 hypothetisch angenommen]

Fall	Prüfgröße T	Stichprobenverteilung von T
ZmZ (3)	(9.9) $\frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{p - \pi_0}{\sigma_P}$	asymptotisch N(0,1) wenn $npq > 9$; sonst Binomialtest ^{*)}
ZoZ (4)	(9.10) $\frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n} \cdot \frac{N-n}{N-1}}}$	$t_{n-1} \rightarrow N(0,1)$ sonst hypergeometrisch verteilt

^{*)} Ist die Stichprobenverteilung die Binomialverteilung (diskret) so empfiehlt sich die Kontinuitätskorrektur. Die errechneten Prüfgrößen sind dann $t_{1/2} = (p - \pi_0 \pm 1/2n)/\sigma_p$ wobei + gilt für den unteren (linken) und - für den oberen (rechten) kritischen Bereich, also + für den Vergleich mit $z_{\alpha/2}$ und - mit $z_{1-\alpha/2}$.

noch Übersicht 9.3:

c) Test über die Varianz

der GG $H_0: \sigma^2 = \sigma_0^2, X_i \sim N(\mu, \sigma^2)$

Prüfgröße T	Verteilung
(9.11) $T = \frac{(n-1)\hat{\sigma}^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2}$	χ_{n-1}^2 (Normalverteilung sehr wichtig)

Durchführung der Zweistichproben-Tests

Übers. 9.4 enthält alle Informationen zur Durchführung der Tests auf Unterschiedlichkeit von

- Mittel- bzw. Erwartungswerten, beurteilt aufgrund von Mittelwertdifferenzen (heterograd, Teil a der Übersicht)
- Anteilswerten bzw. Wahrscheinlichkeiten, beurteilt aufgrund der Differenz zwischen Anteilswerten der Stichproben (homograd, Teil b der Übersicht), und
- der Unterschiedlichkeit der Varianzen zweier unabhängiger Stichproben, beurteilt aufgrund des Quotienten der beiden Varianzen (Teil c der Übersicht 9.4)

In Übers. 9.5 stellt die entsprechenden Tests bei zwei abhängigen Stichproben dar.

Bemerkung zu t-Test für Paardifferenzen im heterograden Fall (d.h. für abhängige Stichproben; Übers. 9.5):

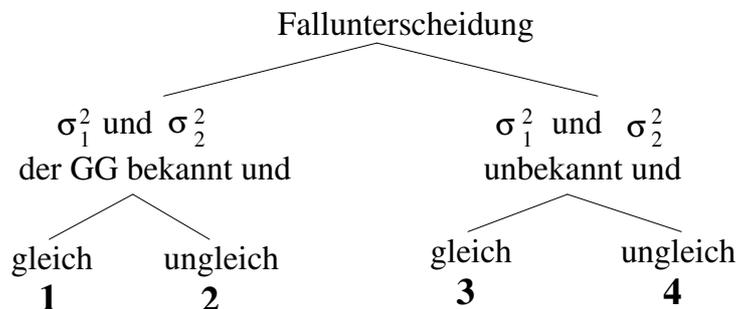
Für eine "verkürzte" Version (leichter rechenbar) ist die Verteilung der Prüfgröße

$A = \sum D_i^2 / (\sum D_i)^2$ tabelliert (A-Test). Da offensichtlich $\bar{D} = \bar{X}_1 - \bar{X}_2$, kann der Zähler von T in Gl. 9.19 auch lauten $(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)$, vorausgesetzt $X_{1i} \sim N(\mu_1, \sigma_1^2)$ und $X_{2i} \sim N(\mu_2, \sigma_2^2)$. Bei unabhängigen Stichproben stellt T eine standardisierte Differenz zweier Mittelwerte dar, bei abhängigen Stichproben ist T ein (mit $\hat{\sigma}_{\bar{D}}$) standardisierter Mittelwert von Differenzen.

Übersicht 9.4: Prüfgrößen im Zwei-Stichproben-Fall (zwei unabhängige Stichproben)

a) **Heterograd (Test über Mittel- bzw. Erwartungswertdifferenz)** auch Sigma-Differenz-Verfahren genannt, große Stichproben

$H_0: \mu_1 - \mu_2 = 0$, $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$, Stichprobenwerte: $\bar{x}_1, \bar{x}_2, n_1, n_2$



Auf die Fälle ZoZ soll hier verzichtet werden. Ist $H_0: \mu_1 - \mu_2 = \Delta \neq 0$, so ist in den Formeln $\hat{\Delta}$ durch $\hat{\Delta} - \Delta$ zu ersetzen.

Fall	Prüfgröße T	Verteilung von T
1	homogene Varianzen $\sigma_1^2 = \sigma_2^2 = \sigma^2$ $\Delta = \mu_1 - \mu_2 = 0$, $\hat{\Delta} = \bar{X}_1 - \bar{X}_2$ (9.14) $T = \frac{\hat{\Delta}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\hat{\Delta}}{\sigma} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$	$N(0,1)$
2	(9.15) $T = \frac{\hat{\Delta}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$N(0,1)$
3	$\sigma_1^2 = \sigma_2^2$ aber unbekannt a) zuerst $\hat{\sigma}^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n - 2}$ bestimmen (pooled variance), dann analog Fall 1 b) (9.16) $T = \frac{\hat{\Delta}}{\hat{\sigma}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$	t_{n-2} ($n = n_1 + n_2$) approximativ $N(0,1)$, wenn $n_1 > 30$ und $n_2 > 30$
4	σ_1^2 und σ_2^2 ungleich und unbekannt (9.17) $T = \frac{\hat{\Delta}}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}} = \frac{\hat{\Delta}}{\sqrt{N}} = \frac{\hat{\Delta}}{\sqrt{K_1 + K_2}}$	angenähert t-verteilt mit ν Freiheitsgr. ^{c)} , bzw. $N(0,1)$ wenn $n_1 > 30$ und $n_2 > 30$

a) Damit ist durchaus $\hat{\sigma}_1^2 \neq \hat{\sigma}_2^2$ oder $s_1^2 \neq s_2^2$ verträglich. Die Homogenität der Varianzen $\sigma_1^2 = \sigma_2^2$ wird mit dem F-Test (Gl. 9.19) getestet.

b) die Wurzel nimmt ihren maximalen Wert $\sqrt{n/2}$ an, wenn $n_1 = n_2$ ist.

c) Bestimmung der Anzahl ν der Freiheitsgrade

$$\nu = \frac{N^2}{\frac{K_1^2}{n_1 - 1} + \frac{K_2^2}{n_2 - 1}} \quad \text{Wenn } \nu \text{ nicht ganzzahlig, nehme man die nächste ganze Zahl.}$$

b) Homograd: zwei gleiche Anteilswerte/Wahrscheinlichkeiten

$H_0: \pi_1 = \pi_2$, X_1 und X_2 identisch zweipunktverteilt mit π

Prüfgröße T	Verteilung
Zuerst bestimmen $P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$ dann $(P_1 - P_2 = \hat{\Delta})$, dann (9.18) $T = \frac{P_1 - P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	$n \pi(1 - \pi) \geq 9$ dann $N(0,1)$ Test ist identisch mit χ^2 -Test einer Vierfeldertafel \rightarrow Gl. 9.21

Ist die Hypothese $\pi_1 - \pi_2 = \Delta \neq 0$ zu testen, dann ist mit $\hat{\Delta} = P_1 - P_2$ die Prüfgröße

(9.18b)
$$T = \frac{\hat{\Delta} - \Delta}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \sim N(0,1) \quad (Q_i = 1 - P_i, i=1,2).$$

c) Test über die Gleichheit zweier Varianzen

$H_0: \sigma_1^2 = \sigma_2^2$ oder $\sigma_1^2/\sigma_2^2 = 1$, Bedingung $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$

Prüfgröße F	Verteilung
(9.19) $F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\frac{n_1 S_1^2}{n_1 - 1}}{\frac{n_2 S_2^2}{n_2 - 1}} \quad \text{mit } \hat{\sigma}_1^2 > \hat{\sigma}_2^2$	F_{n_1-1, n_2-1}

Statt des Quotienten zweier Varianzen könnte man auch die Differenz zweier Standardabweichungen testen, wobei die Statistik $S_1 - S_2$ bei großen Stichproben normalverteilt ist mit Erwartungswert 0 und Varianz $S_1^2/2n_1 + S_2^2/2n_2$.

Übersicht 9.5: Tests bei zwei abhängigen (verbundenen) Stichproben

Hypothese	Prüfgröße	Verteilung
$H_0: \mu_1 - \mu_2 = 0$ (t-Test für Paardifferenzen) $n_1 = n_2 = n$ (heterograd)	$\Delta_i = X_{1i} - X_{2i}$ oder $D_i = X_i - Y_i$ ($i = 1, 2, \dots, n$) (9.19) $T = \frac{\bar{D}}{\hat{\sigma}_{\bar{D}}} \quad \text{mit } \hat{\sigma}_{\bar{D}}^2 = \frac{\hat{\sigma}_D^2}{n}$ $\hat{\sigma}_D^2 = \frac{1}{n-1} \sum (D_i - \bar{D})^2 \approx s_D^2 = \frac{1}{n} \sum D_i^2 - \bar{D}^2$	$T \sim t_{n-1}$
$H_0: \pi_1 = \pi_2$ (homograd)	vgl. McNemar - Test (Gl. 9.22)	
$H_0: \sigma_1^2 = \sigma_2^2$	(9.20) $T = \frac{(\hat{\sigma}_1^2 - \hat{\sigma}_2^2)\sqrt{n-2}}{2\hat{\sigma}_1\hat{\sigma}_2\sqrt{1-r_{12}^2}}$ r_{12} = Korrelation von x_{1i} mit x_{2i} in der Stichprobe	$T \sim t_{n-2}$

4. Varianten des χ^2 -Tests

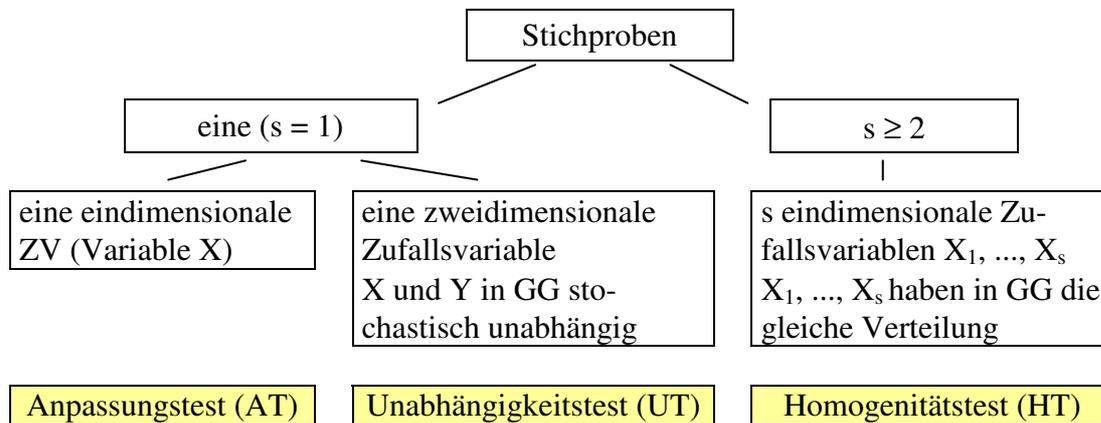
Die Prüfgröße χ^2 ist i.d.R. konstruiert aufgrund beobachteter absoluter Häufigkeiten n_i ($i = 1, 2, \dots, r$) bzw. bei einer zweidimensionalen Häufigkeitstabelle n_{ij} ($j = 1, 2, \dots, s$) und "theoretischer", d.h. bei Geltung von H_0 zu erwartender absoluter Häufigkeiten in der Stichprobe e_i bzw. e_{ij} .

$$(9.20) \quad \chi^2 = \sum_{i=1}^r \frac{(n_i - e_i)^2}{e_i} = \sum_{i=1}^r \frac{n_i^2}{e_i} - n \quad \text{bzw. (9.20a)} \quad \chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Die je nach Art des Tests unterschiedliche H_0 wird verworfen, wenn der stets nicht-negative nach Gl. 9.20 bzw. 9.20a errechnete Wert der Prüfgröße χ^2 größer ist als der α entsprechende Wert der χ^2 -Verteilung. Die theoretischen Häufigkeiten e_i bzw. e_{ij} sollten für alle $i = 1, 2, \dots, r$ und $j = 1, 2, \dots, s$ nicht kleiner als 5 sein.

Übersicht 9.6: Varianten des χ^2 Tests

a) Fragestellung und Daten



b) Hypothesen und Prüfgrößen

Test	Hypothese	e_i bzw. e_{ij} in Gl. 9.20/9.20a	Verteilung
AT	Verteilung von X in der GG hat eine durch H_0 spezifizierte Gestalt mit p Parametern $H_0: f(x) = f_0(x \theta_1, \dots, \theta_p)$ a) etwa $f_0(x \theta_1, \theta_2) = N(\mu, \sigma^2)$	wenn X diskret b) $e_i = np_i = n f_0(x_i)$ (die diskrete Variable kann auch nur nominalskaliert sein)	$\chi^2 \sim \chi^2_\nu$ $\nu = r-p-1$ Freiheitsgrade
UT	$H_0: f(x,y) = f_x(x) f_y(y)$ $f_x(x)$ und $f_y(y)$ sind die Randverteilungen	$e_{ij} = \frac{n_i \cdot n_j}{n}$, Stichprobenkontingenztafel bei Unabhängigkeit	χ^2_ν mit $\nu = (r-1)(s-1)$
HT c)	$H_0: f_1(x) = \dots = f_s(x) = f_0(x)$ s Stichproben aus homogenen Grundgesamtheiten, $j = 1, 2, \dots, s$ mit $n_j =$ umfang der j-ten Stichprobe	$p_i = \frac{\sum_{j=1}^s n_{ij}}{\sum_{j=1}^s \sum_{i=1}^r n_{ij}} = \frac{n_{i.}}{n}$; $e_{ij} = n_j p_i$	χ^2_ν mit $\nu = (r-1)(s-1)$

a) mit p aus der Stichprobe zu schätzenden Parametern.

- b) wenn anstelle von x_i die i -te Klasse mit der Untergrenze x_{iu} und der Obergrenze x_{io} tritt (stetiger Fall), gilt: $p_i = F_0(x_{io}) - F_0(x_{iu})$. Für diesen Fall stehen jedoch bessere Tests zur Verfügung.
- c) Wie an der Konstruktion der Prüfgröße erkennbar ist, sind UT und HT **formal** identische Tests. Die Forderung $f_j(x) = f_0(x)$ für alle $j = 1, 2, \dots, s$ kann als Gleichheit der bedingten Verteilungen (und damit Unabhängigkeit) aufgefaßt werden.

Vierfeldertafel (UT/HT mit $r = s = 2$)

Häufigkeiten

	Stichprobe* (Variable Y)	
	1	2
$x = 1 (+)$	$n_{11} = a$	$n_{12} = b$
$x = 0 (-)$	$n_{21} = c$	$n_{22} = d$
Summe	n_1	n_2

*) Zwei unabhängige Stichproben bei HT, bzw. die zweite dichotome Variable Y (mit $Y = 1$ und $Y = 0$) bei UT:

$$\left. \begin{aligned} P_1 &= \frac{a}{a+c} = \frac{a}{n_1} \\ P_2 &= \frac{b}{b+d} = \frac{b}{n_2} \end{aligned} \right\} P = \frac{a+b}{n} \quad \text{mit } n = n_1 + n_2$$

$$PQ = \frac{a+b}{n} \cdot \frac{c+d}{n}$$

Prüfgröße nach Gl. 9.18: $T = \sqrt{n(ad - bc)} / \sqrt{(a+b)(c+d)(a+c)(b+d)}$ und beim χ^2 -Test:

$$(9.21) \quad \chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} = T^2$$

Wenn $T \sim N(0,1)$, dann ist $\chi^2 = T^2 \sim \chi_1^2$. Der χ^2 -Test ist also äquivalent dem t-Test (Gl.9.18).

Zwei abhängige Stichproben (Mc Nemar-Test)

typische vorher (1) - nachher (2) - Befragung der gleichen Personen $n = a + b + c + d$, $n^* = b + c$. Für den Test interessiert nur die reduzierte Zahl n^* der Beobachtungen.

Stich- probe 1	Stichprobe 2	
	y = 1	y = 0
x = 1	a	b
x = 0	c	d

$H_0: P(x = 1, y = 0) = P(x = 0, y = 1) = \pi = 0,5$. Die Prüfgröße ist dann

$$T = \frac{b - n^* \cdot \pi}{\sqrt{n^* \cdot \pi(1-\pi)}} = \frac{-(c - n^* \cdot \pi)}{\sqrt{n^* \cdot \pi(1-\pi)}} = \frac{b - c}{\sqrt{b+c}} \sim t_{n-1} \text{ [bzw. } N(0,1)\text{]}$$

für den Test bei $n^* > 20$ (bei $n^* \leq 20$ Binomialtest). Zusammenhang mit χ^2 -(Ein-Stichproben)-AT: Die zu erwartenden Häufigkeiten sind bei H_0 jeweils $e_i = \frac{1}{2} n^* = \frac{1}{2} (b + c)$

Häufigkeiten		
empirisch (n_i)	b	c
theoretisch (e_i)	$\frac{1}{2} n^*$	$\frac{1}{2} n^*$

$$(9.22) \quad \chi^2 = \frac{1}{n^*} (2b^2 + 2c^2) - n^* = \frac{(b-c)^2}{b+c}; = T^2 \sim \chi_1^2.$$

Bei kleinem n sind in Gl. 9.21 und 9.22 auch Kontinuitäts (Stetigkeits)-Korrekturen üblich. Es gilt dann im zweiseitigen Test:

$$(9.21a) \quad \chi^2 = \frac{n \left(\left| ad - bc \right| - \frac{n}{2} \right)^2}{(a+b)(c+d)(a+c)(b+d)} \quad \text{und} \quad (9.22a) \quad \chi^2 = \frac{(|b-c|-1)^2}{b+c}.$$

5. Mehr als zwei Stichproben

Die Verallgemeinerung des t-Tests für zwei unabhängige Stichproben im heterograden Fall bei Varianzhomogenität (Gl. 9.16) für $k > 2$ Stichproben heißt auch **Varianzanalyse** (genauer: einfache univariate Varianzanalyse [ANOVA]; "einfach" weil nach nur einem [evtl. nur nominalskalierten] Merkmal k Klassen [k Stichproben] gebildet werden und "univariat" weil Mittelwerte und Streuungen hinsichtlich nur einer Variable X betrachtet werden). Man kann die Varianzanalyse als Homogenitätstest begreifen (in diesem Sinne: Verallgemeinerung des t-Tests) oder aber auch im Sinne eines Unabhängigkeitstests, wenn die Daten statt k unabhängige Stichproben k Ausprägungen (oder Klassen) eines (nicht notwendig mehr als nur nominalskalierten) Merkmals Y sind. Die Varianzanalyse zeigt, ob eine Abhängigkeit der metrisch skalierten Variable X von der Variable Y besteht.

Y (unabhängig)	X (abhängig)	
	nominalskaliert N	metrisch skaliert M
N (nominal)	χ^2 -Test *)	Varianzanalyse
M (metrisch)	-	Regressionsanalyse

*) auf Unabhängigkeit (UT)

Übers. 9.7: Tests bei $k \geq 2$ unabhängige Stichproben

Test (Hypothese)	Prüfgröße T	Verteilung
Test auf Gleichheit von k Mittelwerten $H_0: \mu_1 = \mu_2 = \dots = \mu_k = \mu$ bei Normalverteilung und Streuungsgleichheit ^{a)} $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ F-Test, Varianzanalyse (Prüfgröße ist (9.23))	$\bar{X}_i =$ Mittelwert der i -ten Stichprobe $i = 1, 2, \dots, k \quad \bar{X} = (\sum \bar{X}_i n_i) / n$ Gesamtmittelwert erklärte SAQ ^{b)} : $SAQ_E = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$ Residual SAQ: $SAQ_R = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$ $(9.23) \quad T = \frac{SAQ_E}{\frac{SAQ_R}{n-k}}$	$T \sim F_{v_1, v_2}$ mit $v_1 = k-1$ $v_2 = n-k$ Freiheitsgraden
Gleichheit von k Anteilswerten $\pi_1 = \dots = \pi_k = \pi$	χ^2 -Homogenitätstest: gleiche Zweipunktverteilung $Z(\pi)$ bei k Stichproben	χ^2_{k-1}
auf Gleichheit von k Varianzen (BartlettTest) $H_0: \sigma_1^2 = \dots = \sigma_k^2 = \sigma^2$ $H_1: \sigma_i^2 \neq \sigma^2$ für best. i	Mit $\hat{\sigma}_i^2 = \sum_j (x_{ij} - \bar{x}_i)^2 / (n_i - 1), j = 1, \dots, n_i$ und $\hat{\sigma}^2 = \sum_i (n_i - 1) \hat{\sigma}_i^2 / (n - k), i = 1, \dots, k$ erhält man (924)	$T \sim \chi^2_{k-1}$

a) Varianzhomogenität; Kenntnis der Varianzen ist aber nicht erforderlich.

b) Summe der Abweichungsquadrate SAQ_E ist ein Maß der externen und SAQ_R der internen Streuung.

$$(9.24) \quad T = -\frac{\sum_i (n_i - 1) \ln \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2}}{N} \quad \text{und } N \text{ in (9.24) ist } N = 1 - \frac{\sum_i \frac{1}{n_i - 1} - \frac{1}{n - k}}{3(k - 1)}$$

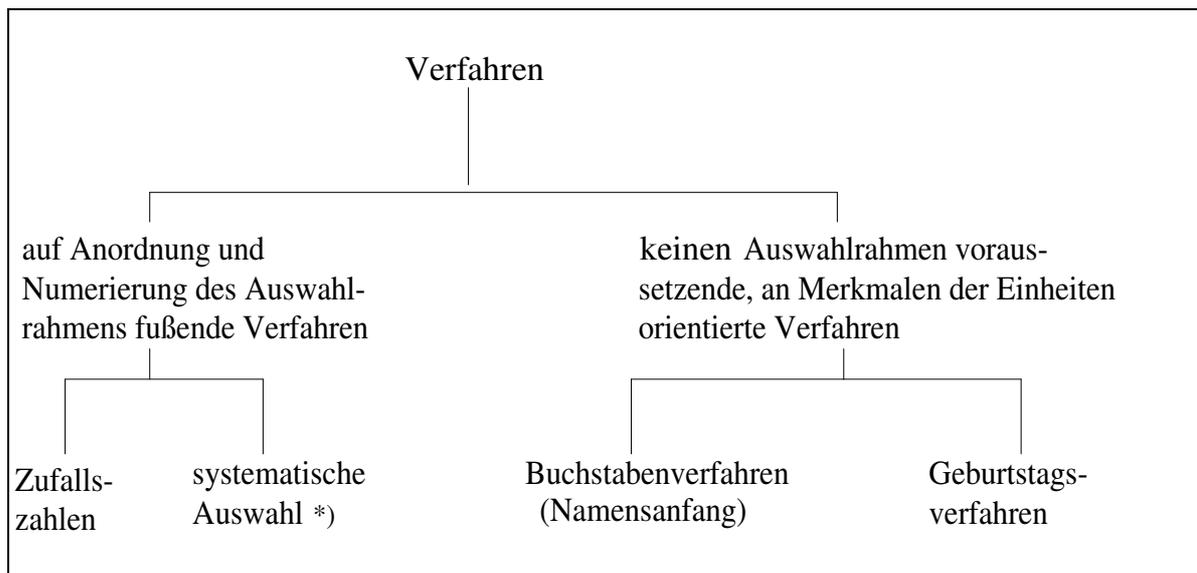
Kapitel 10: Stichprobentheorie

1. Durchführung von Stichprobenerhebungen	100
2. Geschichtete Stichprobe	103
3. Klumpenstichprobe und zweistufige Auswahl	106
4. Weitere Stichprobenpläne	108

1. Durchführung von Stichprobenerhebungen

a) Techniken der Zufallsauswahl und Stichprobenpläne

Übersicht 10.1: Techniken der Zufallsauswahl



*) Jede k-te Karteikarte nach der i-ten (Zufallsstart i). Der Abstand k (als Anzahl der Karteikarten oder als Breite des Kartenstapels) wird durch den Auswahlsatz n/N definiert. Oder: Karteikarten deren fortlaufende Nummer auf i lautet (Schlußziffernverfahren).

b) Notwendiger Stichprobenumfang

Der für eine Stichprobe von geforderter Genauigkeit und Sicherheit mindestens erforderliche Stichprobenumfang n^* ergibt sich aus einer Umformung der Formeln für das Konfidenzintervall. Die Größe n^* hängt ab von:

- Genauigkeit
- Sicherheit
- Homogenität der GG.

1. **Genauigkeit** ist definiert als **absoluter Fehler** e die halbe Länge des (symmetrischen zweiseitigen) Schwankungsintervalls (direkter Schluss) gem. Übers. 8.8 also:

- im Fall 1 (heterograd):
$$e = z_\alpha \frac{\sigma}{\sqrt{n}} \quad \text{und}$$
- im Fall 2 (homograd):
$$e = z_\alpha \sqrt{\frac{\pi(1-\pi)}{n}}.$$

Auflösung dieser Gleichungen nach n liefert die in Übers. 10.2 zusammengestellten Formeln für den notwendigen Stichprobenumfang.

relativer Fehler: $e^* = \frac{e}{\mu}$, bzw. $\frac{e}{\pi}$, allgemein $\frac{e}{\theta}$, also e in Einheiten von θ .

2. **Sicherheit** ist die Wahrscheinlichkeit $1 - \alpha$, der ein bestimmter Wert z_α zugeordnet ist. Genauigkeit und Sicherheit sind konkurrierende Forderungen.

3. **Homogenität der Grundgesamtheit** σ^2 bzw. $\pi(1-\pi)$. Der Stichprobenfehler $\sigma_{\bar{x}}$ bzw. σ_p ist direkt proportional zu σ bzw. $\sqrt{\pi(1-\pi)}$. Häufig ist die Varianz der GG nicht bekannt. Mit σ^{*2} bzw. $\pi^*(1-\pi^*)$ soll angedeutet werden, dass diese Größen geschätzt sind. Eine konservative Schätzung des notwendigen Stichprobenumfangs erhält man im homograden Fall mit $\pi^*(1-\pi^*) = 1/4$.

Übers. 10.2 enthält die Abschätzungen des bei gewünschter Sicherheit erforderlichen Stichprobenumfangs aufgrund der Formeln für die Intervallschätzung (Übers. 8.8) und damit aufgrund der **Grenzwertsätze**.

Würde man demgegenüber den für die Intervallschätzung von μ für eine Sicherheit von $1 - \alpha$ erforderlichen Stichprobenumfang bei einem absoluten Fehler in Höhe von $e = \varepsilon$ mit der **Tschebyscheffschen Ungleichung** (Gl. 7. 10*) abschätzen, so erhielte man:

$$(10.1) \quad n^* \geq \frac{\sigma^{*2}}{e^2 \alpha},$$

was natürlich erheblich größer ist als n^* gem. Übers. 10.2 weil $1/\alpha > z^2$.

$1 - \alpha$	z	z^2	$\frac{1}{\alpha}$	$\frac{1}{\alpha} : z^2$
0,9	1,6449	2,7057	10	3,696
0,95	1,9600	3,8416	20	5,206
0,99	3,2910	10,8307	100	9,233

Bei einer geforderten Sicherheit von 90% (99%) wäre der danach erforderliche Stichprobenumfang 3,7 - mal (9,2 - mal) so groß wie gem. Übers. 10.2.

Es gibt auch Formeln für den erforderlichen Stichprobenumfang um z.B.

- eine vorgegebene Genauigkeit in bestimmten Teilgesamtheiten (z.B. Bundesländern) zu garantieren (das ist v.a. in der amtlichen Statistik ein wichtiges Kriterium),
- eine Varianz mit vorgegebener Genauigkeit und Sicherheit abschätzen zu können oder
- um im Zwei-Stichproben-Fall einen hypothetischen Unterschied (etwa $\mu_1 - \mu_2 = \Delta$) mit einer bestimmten Irrtumswahrscheinlichkeit in den Stichproben zu erkennen.

*Übersicht 10.2: Notwendiger Stichprobenumfang bei einfacher Zufallsauswahl^{a)}
(Fallunterscheidung wie in Übers. 8.8)*

	heterograd	homograd
	Fall 1	Fall 2
ohne Endlichkeitskorrektur	$n^* \geq \frac{z^2 \sigma^{*2}}{e^2} = \frac{z^2 V^2}{e^{*2}}$ (mit $V = \sigma^*/\mu$ Variationskoeffizient)	$n^* \geq \frac{z^2 \pi^* (1 - \pi^*)}{e^2} = \frac{z^2 (1 - \pi^*)}{e^{*2} \pi^*}$ $n_{\max}^* = \frac{z^2}{4e^2}$
	Fall 5 ^{b)}	Fall 6 ^{c)}
mit Endlichkeitskorrektur	$n^* \geq \frac{K}{e^2 + \frac{K}{N}}$ mit $K = z^2 \sigma^{*2}$ und mit entsprechender Formel unter Verwendung von e^*	$n^* \geq \frac{K'}{e^2 + \frac{K'}{N}}$ mit $K' = z^2 \pi^* (1 - \pi^*)$ ^{d)}

a) Zur geschichteten Stichprobe vgl. Gl. 10.14

b) Wenn $N - 1 \approx N$ sonst, $n^* \geq NK / [e^2(N - 1) + K]$

c) Wenn $N - 1 \approx N$ sonst, $n^* \geq NK' / [e^2(N - 1) + K']$

d) Maximaler Wert $n_{\max}^* = \frac{z^2}{4e^2 + \frac{z^2}{N}}$

c) Hochrechnung

In Kap. 8, Abschn. 4b wurde als Hochrechnung^{*)} das Problem bezeichnet, von einem Punktschätzer $\bar{x} = \hat{\mu}$ bzw. $p = \hat{\pi}$, also von einem Mittelwert bzw. Anteilswert auf eine Merkmalssumme $N\mu$ oder eine Gesamthäufigkeit (einen Bestand) $N\pi$ zu schließen. Es ist plausibel anzunehmen, dass im heterograden Fall gilt (und im homograden Fall entsprechend):

$$\frac{\bar{x}}{\mu} = \frac{N \bar{x}}{N \mu},$$

*) In der "Alltagssprache", z.B. bei Fernsehsendungen am Wahlabend wird "Hochrechnung" im Sinne der "Punktschätzung" gebraucht, nicht in den oben definierten Sinne einer "Rückvergrößerung" eines Bildes von den kleineren Dimensionen einer Stichprobe auf die größeren Dimensionen der Grundgesamtheit.

dass also die Merkmalssumme evtl. im gleichen Maße über- oder unterschätzt wird wie der Mittelwert. Man kann dann mit dem reziproken Auswahlatz hochrechnen (**freie Hochrechnung**):

$$(10.2a) \quad N\mu \cong \frac{N}{n} (n \bar{x}) = N \bar{x}$$

$$(10.2b) \quad N\pi \cong \frac{N}{n} (n p) = N p$$

(\cong soll heißen: wird geschätzt mit)

Es gibt Fälle, in denen der Rückgriff auf andere Stichprobeninformationen (z.B. über andere Merkmale y, z neben x) zu bessern Ergebnissen führt als die freie Hochrechnung. Man nennt solche Verfahren auch **gebundene Hochrechnung**.

2. Geschichtete Stichprobe (stratified sample)

Notation

Aufteilung der Grundgesamtheit in K Schichten mit den Umfängen N_k mit

$$(10.3) \quad N = N_1 + N_2 + \dots + N_K = \sum_{k=1}^K N_k$$

und Aufteilung der Stichprobe des Umfangs n auf K Stichproben

$$(10.4) \quad n = n_1 + n_2 + \dots + n_K = \sum_{k=1}^K n_k.$$

Schichtenbildung aufgrund eines Schichtungsmerkmals oder einer Kombination von Schichtungsmerkmalen.

Mittel- und Anteilswertschätzung

Bei Schätzfunktionen $\hat{\theta}$ für θ , wie etwa P für π (homograde Fall) oder \bar{X} für μ (heterograde Fall) gilt die Aggregationsformel

$$(10.5) \quad \hat{\theta} = \sum_{k=1}^K \frac{N_k}{N} \hat{\theta}_k \quad \text{mit}$$

$$(10.6) \quad E(\hat{\theta}) = \sum_{k=1}^K \frac{N_k}{N} E(\hat{\theta}_k) \quad \text{und} \quad (10.7) \quad V(\hat{\theta}) = \sum_{k=1}^K \left(\frac{N_k}{N} \right)^2 V(\hat{\theta}_k),$$

denn eine geschichtete Stichprobe bedeutet K unabhängige Stichproben aus K Schichten zu ziehen. Je nach Art der Stichprobenfunktion $\hat{\theta}$ und der Stichprobenziehung erhält man anstelle von Gl. 10.5 bis 10.7 spezielle Formeln. Etwa im **heterograden Fall** (Mittelwert):

	Gl. (10.6): $E(\hat{\theta})$	Gl. (10.7): $V(\hat{\theta})$
ZmZ	$E(\bar{X}) = \sum \frac{N_k}{N} E(\bar{X}_k) = \mu$	(10.8) $V(\bar{X}) = \sum \frac{N_k^2}{N^2} \frac{\sigma_k^2}{n_k}$
ZoZ	wie ZmZ	(10.9) $V(\bar{X}) = \sum \frac{N_k^2}{N^2} \frac{\sigma_k^2}{n_k} \frac{N_k - n_k}{N_k - 1}$

und **im homograden Fall** analoge Formeln mit $\sigma_k^2 = \pi_k(1 - \pi_k)^*$. Damit sind Intervallschätzungen und Tests (bezüglich μ bzw. π) durchführbar.

Mit $\hat{\sigma}_k^2$ anstelle von σ_k^2 in Gl. 10.9 lauten die K finite multipliers $(N_k - n_k)/N_k$.

Aus Gl. 10.9 folgt, dass der Standardfehler $\sigma_{\bar{x}} = \sqrt{V(\bar{X})}$ um so kleiner ist und damit die Schätzung um so besser ist, je homogener die K Schichten (je kleiner die K Varianzen σ_k^2) sind.

Aufteilung (Allokation) der Stichprobe

a) Proportionale Aufteilung:

$$(10.10) \quad \frac{n_k}{n} = \frac{N_k}{N}, \quad \text{für alle } k = 1, 2, \dots, K$$

Konsequenz: $\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_k}{N_k}$ (gleiche Auswahlätze) und $\bar{x} = \frac{1}{n} \sum n_k \bar{x}_k$.

b) Nichtproportionale Aufteilung

Darunter eine Möglichkeit: **optimale Aufteilung**. (J. Neyman). Minimierung von $V(\hat{\theta})$, etwa $V(\bar{X})$ nach Gl. 10.8 unter der Nebenbedingung $n = \sum n_k$ liefert:

$$(10.11) \quad \frac{n_k}{n} = \frac{N_k \sigma_k}{\sum N_k \sigma_k}$$

Man kann eine optimale Aufteilung auch so bestimmen, dass die Varianz unter Einhaltung vorgegebener Kosten der Erhebung minimiert wird. Die Kosten müssen dabei linear abhängen von c_k einem (innerhalb der k-ten Schicht gleichen) konstanten Kostenbetrag je Einheit. Sind alle σ_k gleich ist wegen $\sum N_k = N$ die proportionale Aufteilung zugleich die optimale. Gl. 10.10 und 10.11 eingesetzt in Gl. 10.8 ergibt:

$$(10.12) \quad V(\bar{X})_{\text{opt}} = \frac{1}{n} \left(\sum \frac{N_k}{N} \sigma_k \right)^2 = \frac{1}{n} \left[\sum \frac{n_k}{n} \left(\frac{\sum N_k \sigma_k}{N} \right) \right]^2,$$

was offenbar nicht größer ist als¹⁾

$$(10.13) \quad V(\bar{X})_{\text{prop}} = \frac{1}{n} \sum \frac{N_k}{N} \sigma_k^2 = \frac{1}{n} \sum \frac{n_k}{n} \sigma_k^2.$$

*) Wird z.B. der Parameter π durch p geschätzt, so ist bei ZoZ $V(p) = \sum \frac{N_k^2}{N^2} \frac{p_k q_k}{n_k - 1} \frac{N_k - n_k}{N_k}$.

1) Gleichheit wenn alle K Standardabweichungen σ_k gleich sind. Bei den behaupteten Größenvergleich liegt der gleiche Zusammenhang vor wie bei $E(X^2) > [E(X)]^2$

Der Klammerausdruck in Gl. 10.12 (zweiter Teil, runde Klammer) kann als mittlere Schicht-Standardabweichung $\bar{\sigma}$ gedeutet werden, so dass gilt:

$$V(\bar{X})_{\text{prop}} - V(\bar{X})_{\text{opt}} = \frac{1}{n} \sum \frac{N_k}{N} - \frac{1}{n} \bar{\sigma}^2 = \frac{1}{n} \sum \frac{N_k}{N} (\sigma_k - \bar{\sigma})^2 \geq 0.$$

Auf die entsprechenden Formeln im Fall ZoZ soll hier verzichtet werden. Eine andere i.d.R. nicht proportionale Aufteilung von n in n_k^* ($n = \sum n_k^*$) wäre die Aufteilung mit vorgegebener (gewünschter) Genauigkeit e_k in jeder Schicht (Gl. 10.14a).

Schichtungseffekt

Bekanntlich gilt bei einfacher Stichprobe (ZmZ)

$$\sigma_{\bar{X}}^2 = V(\bar{X})_{\text{einf}} = \frac{\sigma^2}{n},$$

wobei für die Varianz σ^2 der Grundgesamtheit die Varianz wie folgt in externe und interne Varianz

$$\sigma^2 = \sum_{k=1}^K \frac{N_k}{N} (\mu_k - \mu)^2 + \sum_{k=1}^K \frac{N_k}{N} \sigma_k^2 = V_{\text{ext}} + V_{\text{int}} = nV(\bar{X})_{\text{einfach}}.$$

zu zerlegen ist. Im Falle einer geschichteten Stichprobe mit proportionaler Aufteilung gilt demgegenüber wegen Gl. 10.13:

$$nV(\bar{X})_{\text{prop}} = \sum \frac{N_k}{N} \sigma_k^2 = V_{\text{int}} \leq \sigma^2 = V_{\text{ext}} + V_{\text{int}}$$

Sobald eine externe Varianz auftritt (zwischen den Schichten große Unterschiede sind) ist der Standardfehler der Schätzung bei Schichtung kleiner als bei einfacher Stichprobe. Der Schichtungseffekt ist um so größer je homogener (je kleiner V_{int} ist) die Schichten sind.

Schichtung ist auch eine Möglichkeit zu verhindern, dass zufällig kein Element einer bestimmten Schicht in einer Stichprobe vertreten ist.

Notwendiger Stichprobenumfang

Für den absoluten Fehler e im heterograden Fall ZmZ und bei proportionaler Aufteilung gilt nach Gl. 10.8 $e^2 = z^2 V(\bar{X})_{\text{prop}}$, was nach Gl. 10.13 und aufgelöst nach n den folgenden

Mindeststichprobenumfang n^* ergibt:

$$(10.14) \quad n^* \geq \frac{z^2}{e^2} \cdot V_{\text{int}} = \frac{z^2}{e^2} \sum \frac{N_k}{N} \sigma_k^2,$$

im Unterschied zur einfachen Stichprobe (Übers. 10.2):

$$n^* \geq \frac{z^2}{e^2} \cdot \sigma^2 = \frac{z^2}{e^2} (V_{\text{int}} + V_{\text{ext}})$$

Entsprechende Formeln erhält man im Fall ZoZ und bei optimaler Aufteilung.

Es ist auch eine i.d.R. nicht-proportionale Aufteilung der Stichprobe in der Weise möglich, dass für jede Schicht eine vorgegebene Genauigkeit (gemessen am absoluten Fehler e_k) eingehalten wird, was bei:

$$(10.14a) \quad n_k^* \geq \frac{z^2}{e_k^2} \sigma_k^2 \quad \forall k \text{ gewährleistet ist.}$$

3. Klumpenstichprobe (cluster sample) und zweistufige Auswahl

Notation, Stichprobenpläne

Aufteilung der GG in M Klumpen (cluster) mit den Umfängen N_i ($i = 1, 2, \dots, M$) in der GG. Von den M Klumpen werden m zufällig ausgewählt und jeweils mit

- **allen** ihren Einheiten (**einstufiges** Auswahlverfahren^{*)}) ausgezählt, also mit N_j Einheiten, wenn der j -te Klumpen in die Auswahl gelangt ($j = 1, 2, \dots, m$),
- einer Zufallsauswahl von n_j Einheiten beim ausgewählten j -ten Klumpen (Auswahlsätze (n_j/N_j) $100 \leq 100\%$) untersucht (**zweistufige** Klumpenauswahl).

Ein Klumpen ist eine natürliche (vorgefundene) Ansammlung von Untersuchungseinheiten, die in sich möglichst heterogen sein sollte (eine verkleinerte GG) und die Klumpen sollten untereinander möglichst homogen sein. Hinsichtlich Klumpen und Schichten werden also gegensätzliche Forderungen aufgestellt. Der in der Praxis wichtigste und aus Kostengründen besonders beliebte spezielle Fall einer Klumpenstichprobe ist die **Flächenstichprobe (area sample)** mit Regionen (z.B. Gemeinden) als Klumpen.

Mittelwertschätzung bei einstufiger Klumpenauswahl und $N_i = \bar{N}$

a) Punktschätzung \bar{X}

$$(10.15) \quad \bar{x} = \hat{\mu} = \frac{\sum N_j \mu_j}{\sum N_j} = \frac{\sum_j \sum_k X_{jk}}{m \bar{N}} = \frac{\sum \mu_j}{m},$$

($j = 1, 2, \dots, m$; Stichprobenumfang $n = m \bar{N}$, $i = 1, \dots, M$ und $k = 1, 2, \dots, N_i = \bar{N}$)

Wegen der Vollerhebung innerhalb eines Klumpens ist das Klumpenmittel μ_j ohne Stichprobenfehler zu schätzen (allerdings nur bei den Klumpen, die in die Auswahl gelangen, $j = 1, 2, \dots, m$). Der Schätzwert $\hat{\mu}$ beruht auf den m ausgewählten Klumpen im Unterschied zum wahren Mittelwert μ der

endlichen GG $\mu = \frac{\sum N_i \mu_i}{\sum N_i}$ ($j=1,2,\dots,m$, $i = 1, 2, \dots, M$ und mit $N = \sum N_i = M \cdot \bar{N}$).

b) Varianz von \bar{X}

Die Varianz von \bar{X} bei einer Auswahl von m aus M Klumpen (ZoZ) ist

$$(10.16) \quad V(\bar{X}) = \sigma_{\bar{x}}^2 = \frac{\sigma_b^2}{m} \frac{M-m}{M-1}$$

mit $\sigma_b^2 =$ Varianz zwischen (between) den Klumpen. Nach Definition ist das

^{*)} Klumpenauswahl im engeren Sinne, in der Art, wie in Übers. 8.5. In Gegenüberstellung zur geschichteten Stichprobe beschrieben; Auswahlsatz auf der zweiten Stufe 100%. Der Stichprobenumfang n liegt dann mit der Anzahl m der ausgewählten Klumpen fest mit $n = \sum N_j$ ($j = 1, 2, \dots, m$).

$$\begin{aligned}\sigma_b^2 &= \frac{1}{N} \sum N_i (\mu_i - \mu)^2 = \frac{1}{M} \sum (\mu_i - \mu)^2 = \frac{1}{M} \sum_{i=1}^M \left(\frac{1}{\bar{N}} \sum_{k=1}^{\bar{N}} x_{ik} - \mu \right)^2 \\ &= \frac{1}{M\bar{N}^2} \sum_{i=1}^M \left(\sum_{k=1}^{\bar{N}} x_{ik} - \bar{N}\mu \right)^2 = \frac{1}{M\bar{N}^2} \sum_i \left[\sum_k (x_{ik} - \mu) \right]^2\end{aligned}$$

Diese Summe der quadrierten Summen lässt sich zerlegen in

- $M\bar{N} = N$ Größen $(x_{ik} - \mu)^2$, die in ihrer Summe die N -fache ($M\bar{N}$ -fache) Gesamtvarianz (σ^2) der Variable X in der (endlichen) Grundgesamtheit ergeben und
- $M\bar{N}(\bar{N} - 1)$ Glieder der Art $(x_{ik} - \mu)(x_{il} - \mu)$ mit $k \neq l$, und $k, l = 1, 2, \dots, N_i = \bar{N}$.

Das Mittel dieser Produkte $\frac{1}{M\bar{N}(\bar{N} - 1)} \sum_{i=1}^M \left[\sum_{k=1}^{\bar{N}} \sum_{l=1}^{\bar{N}} (x_{ik} - \mu)(x_{il} - \mu) \right] = \sigma_{kl} = \rho\sigma^2$

(die Doppelsumme in den eckigen Klammern hat $\bar{N}(\bar{N} - 1)$ Summanden) ist die durchschnittliche Kovarianz der Betrachtungen innerhalb der Klumpen und ρ heißt **Intraklasskorrelationskoeffizient**. Er ist ein Maß der Homogenität der Klumpen.

Offenbar ist $M\bar{N}^2 = M\bar{N} + M\bar{N}(\bar{N} - 1)$, so dass gilt:

$$(10.17) \quad \sigma_b^2 = \frac{\sigma^2}{N} + \frac{M\bar{N}(\bar{N} - 1)}{M\bar{N}^2} \rho\sigma^2 = \frac{\sigma^2}{N} [1 + (\bar{N} - 1)\rho] = \frac{\sigma^2}{N} V_{BL}.$$

Der Faktor in den eckigen Klammern wird auch **Varianzaufblähungsfaktor** (V_{BL}) genannt. Ist (was die Regel ist) $V_{BL} > 1$, so ist die Klumpenstichprobe nicht so wirksam wie die einfache Stichprobe, denn Gl. 10.17 eingesetzt in Gl. 10.16 liefert:

$$(10.18) \quad V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \frac{M - m}{M - 1} [1 + (\bar{N} - 1)\rho] \approx \frac{\sigma^2}{n} \left(1 - \frac{m}{M}\right) V_{BL}.$$

(wegen $n = m\bar{N}$) im Vergleich zur Varianz der Stichprobenverteilung

$$(10.19) \quad \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1} = \frac{\sigma^2}{n} \cdot \frac{M\bar{N} - m\bar{N}}{M\bar{N} - 1} = \frac{\sigma^2}{n} \cdot \frac{M - m}{M - 1/\bar{N}} \approx \frac{\sigma^2}{n} \left(1 - \frac{m}{M}\right)$$

bei einfacher Zufallsauswahl*).

Zweistufige (Klumpen-)Auswahl

Die Varianz $V(\bar{X})$ lässt sich zerlegen in eine von σ_b^2 und eine von σ_w^2 (Varianz innerhalb [within] der Klumpen) abhängige Komponente. Anders als in Gl. 10.15 sind die Klumpenmittelwerte μ_i durch \bar{x}_j zu schätzen. Man erhält wieder mit $N_j = \bar{N}$ als Punktschätzer für μ :

*) Wird in Gl. 10.18 und 10.19 jeweils σ^2 durch $\hat{\sigma}^2$ geschätzt, so ist im Nenner der Endlichkeitskorrektur M (bzw. N) statt $M-1$ (bzw. $N-1$) zu schreiben.

$\bar{x} = \sum x_j / m$. Der Stichprobenumfang ist jetzt $n = \sum n_j \leq \sum N_j = m\bar{N}$, weil auf der zweiten Stufe ausgewählt wird $n_j / N_j \leq 1$. Für die Varianz von \bar{x} erhält man mit $\hat{\sigma}_b^2$ und $\hat{\sigma}_w^2$

$$(10.20) \quad V(\bar{X}) = \frac{1}{N^2} \left[M^2 \left(\frac{1}{m} - \frac{1}{M} \right) \hat{\sigma}_b^2 + \frac{M}{m} \sum N_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i} \right) \hat{\sigma}_w^2 \right]$$

4. Weitere Stichprobenpläne

Ungleiche Auswahlwahrscheinlichkeiten: PPS-Verfahren

Prinzip: Berücksichtigung der Größe (des Merkmalsbetrags x_i) der Einheiten ($i = 1, 2, \dots, N$) einer endlichen GG bei Auswahl und Hochrechnung (Auswahl mit der Wahrscheinlichkeit w_i proportional zur Größe x_i der Einheit i [probability proportional to size PPS] statt mit $w_i = 1/N$ für alle i bei einfacher Stichprobe).

Die Schätzfunktion für μ lautet mit w_i gem. Gl. 10.22:

$$(10.21) \quad \bar{X} = \frac{1}{n} \sum_{j=1}^n \frac{x_j}{Nw_j} \quad (j=1, 2, \dots, n)$$

$$(10.22) \quad w_i = \frac{x_i}{\sum x_i} = \frac{x_i}{N\mu} \quad (i=1, 2, \dots, N)$$

Sie erlaubt eine exakte "Schätzung" von μ (sogar mit $n=1$) **wenn** $w_i = p_i$, also w_i identisch ist mit dem Anteil von x_i am Gesamtmerkmalsbetrag $\sum x_i$. Meist wird w_i jedoch mit einem anderen Merkmal Y (etwa Y : Fläche bei der Schätzung von X : Ernteertrag) geschätzt:

$$(10.23) \quad w_i^* = \frac{y_i}{\sum y_i}. \text{ Es gilt bei Stichproben ZmZ}$$

$$(10.24) \quad \sigma_{\bar{x}}^2 = V(\bar{X}) = V\left(\frac{1}{n} \sum_{j=1}^n \frac{X_j}{Nw_j}\right) = \left(\frac{1}{n}\right)^2 \sum_{j=1}^n \sum_{i=1}^N p_i \left(\frac{x_i}{Nw_i} - \mu\right)^2 = \frac{1}{n} \left[\frac{1}{N^2} \sum_{i=1}^N \frac{x_i^2}{w_i^2} p_i - \mu^2 \right].$$

Für $w_i = 1/N$ erhält man die bekannten Ergebnisse für die einfache Stichprobe $E(\bar{X}) = \mu$ und $V(\bar{X}) = \sigma^2 / n$. $V(\bar{X})$ nach Gleichung 10.24 ist erwartungstreu zu schätzen mit:

$$\hat{\sigma}_{\bar{x}}^2 = \frac{1}{n} \sum_{j=1}^n \left(\frac{x_j}{Nw_j} - \bar{x} \right)^2 w_j [1 - (n-1)w_j]. \text{ Die Differenz zwischen der Varianz der Stichprobenverteilung von } \bar{X} \text{ bei einfacher Stichprobe und bei PPS ist}$$

$$\Delta V = \frac{1}{n} \left[\sum_{i=1}^N \left(x_i^2 - \frac{x_i^2}{N^2 w_i^2} \right) p_i \right].$$

Mit $w_i = 1/N$ gibt es keinen Genauigkeitsgewinn ($\Delta V = 0$) und die Varianz $V(\bar{X})$ bei PPS ist dann kleiner ($\Delta V > 0$) als bei einfacher Stichprobe, wenn der Ausdruck in den eckigen Klammern positiv ist. Das ist z.B. der Fall, wenn für w_i Gl. 10.22 gilt, denn dann ist dieser Ausdruck $\bar{x}^2 - \sum \mu^2 p_i = \bar{x}^2 - \mu^2 > 0$, wobei \bar{x}^2 das zweite und μ das erste Anfangsmoment der endlichen GG ist.

Der Abschnitt über **mehrphasige Stichproben** ist hier gestrichen. Eine weitere, ausführlichere Darstellung der Stichprobentheorie findet sich in <http://www.von-der-lippe.org/downloads1.php>

Tabelle 1: Binomialverteilung $x \sim B(n, \pi)$; $n-x \sim B(n, 1-\pi)$

n	x	$\pi = 0,1$		$\pi = 0,2$		$\pi = 0,3$		$\pi = 0,4$		$\pi = 0,5$	
		f(x)	F(x)								
1	0	0,9000	0,9000	0,8000	0,8000	0,7000	0,7000	0,6000	0,6000	0,5000	0,5000
	1	0,1000	1,0000	0,2000	1,0000	0,3000	1,0000	0,4000	1,0000	0,5000	1,0000
2	0	0,8100	0,8100	0,6400	0,6400	0,4900	0,4900	0,3600	0,3600	0,2500	0,2500
	1	0,1800	0,9900	0,3200	0,9600	0,4200	0,9100	0,4800	0,8400	0,5000	0,7500
	2	0,0100	1,0000	0,0400	1,0000	0,0900	1,0000	0,1600	1,0000	0,2500	1,0000
3	0	0,7290	0,7290	0,5120	0,5120	0,3430	0,3430	0,2160	0,2160	0,1250	0,1250
	1	0,2430	0,9720	0,3840	0,8960	0,4410	0,7840	0,4320	0,6480	0,3750	0,5000
	2	0,0270	0,9990	0,0960	0,9920	0,1890	0,9730	0,2880	0,9360	0,3750	0,8750
	3	0,0010	1,0000	0,0080	1,0000	0,0270	1,0000	0,0640	1,0000	0,1250	1,0000
4	0	0,6561	0,6561	0,4096	0,4096	0,2401	0,2401	0,1296	0,1296	0,0625	0,0625
	1	0,2916	0,9477	0,4096	0,8192	0,4116	0,6517	0,3456	0,4752	0,2500	0,3125
	2	0,0486	0,9963	0,1536	0,9728	0,2646	0,9163	0,3456	0,8208	0,3750	0,6875
	3	0,0036	0,9999	0,0256	0,9984	0,0756	0,9919	0,1536	0,9744	0,2500	0,9375
	4	0,0001	1,0000	0,0016	1,0000	0,0081	1,0000	0,0256	1,0000	0,0625	1,0000
5	0	0,5905	0,5905	0,3277	0,3277	0,1681	0,1681	0,0778	0,0778	0,0313	0,0313
	1	0,3281	0,9185	0,4096	0,7373	0,3602	0,5282	0,2592	0,3370	0,1563	0,1875
	2	0,0729	0,9914	0,2048	0,9421	0,3087	0,8369	0,3456	0,6826	0,3125	0,5000
	3	0,0081	0,9995	0,0512	0,9933	0,1323	0,9692	0,2304	0,9130	0,3125	0,8125
	4	0,0005	1,0000	0,0064	0,9997	0,0284	0,9976	0,0768	0,9898	0,1563	0,9688
	5	0,0000	1,0000	0,0003	1,0000	0,0024	1,0000	0,0102	1,0000	0,0313	1,0000
6	0	0,5314	0,5314	0,2621	0,2621	0,1176	0,1176	0,0467	0,0467	0,0156	0,0156
	1	0,3543	0,8857	0,3932	0,6554	0,3025	0,4202	0,1866	0,2333	0,0938	0,1094
	2	0,0984	0,9841	0,2458	0,9011	0,3241	0,7443	0,3110	0,5443	0,2344	0,3438
	3	0,0146	0,9987	0,0819	0,9830	0,1852	0,9295	0,2765	0,8208	0,3125	0,6563
	4	0,0012	0,9999	0,0154	0,9984	0,0595	0,9891	0,1382	0,9590	0,2344	0,8906
	5	0,0001	1,0000	0,0015	0,9999	0,0102	0,9993	0,0369	0,9959	0,0938	0,9844
	6	0,0000	1,0000	0,0001	1,0000	0,0007	1,0000	0,0041	1,0000	0,0156	1,0000
7	0	0,4783	0,4783	0,2097	0,2097	0,0824	0,0824	0,0280	0,0280	0,0078	0,0078
	1	0,3720	0,8503	0,3670	0,5767	0,2471	0,3294	0,1306	0,1586	0,0547	0,0625
	2	0,1240	0,9743	0,2753	0,8520	0,3177	0,6471	0,2613	0,4199	0,1641	0,2266
	3	0,0230	0,9973	0,1147	0,9667	0,2269	0,8740	0,2903	0,7102	0,2734	0,5000
	4	0,0026	0,9998	0,0287	0,9953	0,0972	0,9712	0,1935	0,9037	0,2734	0,7734
	5	0,0002	1,0000	0,0043	0,9996	0,0250	0,9962	0,0774	0,9812	0,1641	0,9375
	6	0,0000	1,0000	0,0004	1,0000	0,0036	0,9998	0,0172	0,9984	0,0547	0,9922
	7	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0016	1,0000	0,0078	1,0000
8	0	0,4305	0,4305	0,1678	0,1678	0,0576	0,0576	0,0168	0,0168	0,0039	0,0039
	1	0,3826	0,8131	0,3355	0,5033	0,1977	0,2553	0,0896	0,1064	0,0313	0,0352
	2	0,1488	0,9619	0,2936	0,7969	0,2965	0,5518	0,2090	0,3154	0,1094	0,1445
	3	0,0331	0,9950	0,1468	0,9437	0,2541	0,8059	0,2787	0,5941	0,2188	0,3633
	4	0,0046	0,9996	0,0459	0,9896	0,1361	0,9420	0,2322	0,8263	0,2734	0,6367
	5	0,0004	1,0000	0,0092	0,9988	0,0467	0,9887	0,1239	0,9502	0,2188	0,8555
	6	0,0000	1,0000	0,0011	0,9999	0,0100	0,9987	0,0413	0,9915	0,1094	0,9648
	7	0,0000	1,0000	0,0001	1,0000	0,0012	0,9999	0,0079	0,9993	0,0313	0,9961
	8	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0007	1,0000	0,0039	1,0000

Tabelle 2: Wahrscheinlichkeitsfunktion $f(x)$ und Verteilungsfunktion $F(x)$ der Poisson-Verteilung

x	$\lambda=0.1$		$\lambda=0.2$		$\lambda=0.3$		$\lambda=0.4$		$\lambda=0.5$	
	f(x)	F(x)								
0	0.9048	0.9048	0.8187	0.8187	0.7408	0.7408	0.6703	0.6703	0.6065	0.6065
1	0.0905	0.9953	0.1637	0.9825	0.2222	0.9631	0.2681	0.9384	0.3033	0.9098
2	0.0045	0.9998	0.0164	0.9989	0.0333	0.9964	0.0536	0.9921	0.0758	0.9856
3	0.0002	1.0000	0.0011	0.9999	0.0033	0.9997	0.0072	0.9992	0.0126	0.9982
4	0.0000	1.0000	0.0001	1.0000	0.0003	1.0000	0.0007	0.9999	0.0016	0.9998
5							0.0001	1.0000	0.0002	1.0000

x	$\lambda=0.6$		$\lambda=0.7$		$\lambda=0.8$		$\lambda=0.9$		$\lambda=1$	
	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0.5488	0.5488	0.4966	0.4966	0.4493	0.4493	0.4066	0.4066	0.3679	0.3679
1	0.3293	0.8781	0.3476	0.8442	0.3595	0.8088	0.3659	0.7725	0.3679	0.7358
2	0.0988	0.9767	0.1217	0.9659	0.1438	0.9526	0.1647	0.9371	0.1839	0.9197
3	0.0198	0.9966	0.0284	0.9942	0.0383	0.9909	0.0494	0.9865	0.0613	0.9810
4	0.0030	0.9996	0.0050	0.9992	0.0077	0.9986	0.0111	0.9977	0.0153	0.9963
5	0.0004	1.0000	0.0007	0.9999	0.0012	0.9998	0.0020	0.9997	0.0031	0.9994
6			0.0001	1.0000	0.0002	1.0000	0.0003	1.0000	0.0005	0.9999
7									0.0001	1.0000

x	$\lambda=1.5$		$\lambda=2$		$\lambda=3$		$\lambda=4$		$\lambda=5$	
	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0.2231	0.2231	0.1353	0.1353	0.0498	0.0498	0.0183	0.0183	0.0067	0.0067
1	0.3347	0.5578	0.2707	0.4060	0.1494	0.1991	0.0733	0.0916	0.0337	0.0404
2	0.2510	0.8088	0.2707	0.6767	0.2240	0.4232	0.1465	0.2381	0.0842	0.1247
3	0.1255	0.9344	0.1804	0.8571	0.2240	0.6472	0.1954	0.4335	0.1404	0.2650
4	0.0471	0.9814	0.0902	0.9473	0.1680	0.8153	0.1954	0.6288	0.1755	0.4405
5	0.141	0.9955	0.0361	0.9834	0.1008	0.9161	0.1563	0.7851	0.1755	0.6160
6	0.0035	0.9991	0.0120	0.9955	0.0504	0.9665	0.1042	0.8893	0.1462	0.7622
7	0.0008	0.9998	0.0034	0.9989	0.0216	0.9881	0.095	0.9489	0.1044	0.8666
8	0.0001	1.0000	0.0009	0.9998	0.0081	0.9962	0.0298	0.9786	0.0653	0.9319
9			0.0002	1.0000	0.0027	0.9989	0.0132	0.9919	0.0363	0.9682
10					0.0008	0.9997	0.0053	0.9972	0.0181	0.9863
11					0.0002	0.9999	0.0019	0.9991	0.0082	0.9945
12					0.0001	1.0000	0.0006	0.9997	0.0034	0.9980
13							0.0002	0.9999	0.0013	0.9993
14							0.0001	1.0000	0.0005	0.9998
15									0.0002	0.9999
16									0.0000	1.0000

Tabelle 3: Standardnormalverteilung $N(0,1)$

Zur Erklärung der Funktionen $F(z)$ und $\Phi(z)$ vgl. Seite 6.4:

	Dichtefunktion	Verteilungsfunktion	Symmetrische Intervall- wahrscheinlichkeit
z	$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$	$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du$	$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-\frac{u^2}{2}} du$
0	0.3989	0.5000	0.0000
0.1	0.3970	0.5398	0.0797
0.2	0.3910	0.5793	0.1585
0.3	0.3814	0.6179	0.2358
0.4	0.3683	0.6554	0.3108
0.5	0.3521	0.6915	0.3829
0.6	0.3332	0.7257	0.4515
0.7	0.3122	0.7580	0.5161
0.8	0.2897	0.7881	0.5763
0.9	0.2661	0.8159	0.6319
1.0	0.2420	0.8413	0.6827
1.1	0.2178	0.8649	0.7287
1.2	0.1942	0.8849	0.7699
1.3	0.1714	0.9032	0.8064
1.4	0.1497	0.9192	0.8385
1.5	0.1295	0.9332	0.8664
1.6	0.1109	0.9452	0.8904
1.7	0.0940	0.9594	0.9109
1.8	0.0790	0.9641	0.9281
1.9	0.0656	0.9713	0.9426
2.0	0.0540	0.9772	0.9545
2.1	0.0440	0.9821	0.9643
2.2	0.0355	0.9861	0.9722
2.3	0.0283	0.9893	0.9786
2.4	0.0224	0.9918	0.9836
2.5	0.0175	0.9938	0.9876
2.6	0.0136	0.9953	0.9907
2.7	0.0104	0.9963	0.9931
2.8	0.0079	0.9974	0.9949
2.9	0.0060	0.9981	0.9963
3.0	0.0044	0.9987	0.9973

Wichtige Signifikanzschranken und Wahrscheinlichkeiten

a) z-Werte für gegebene Wkt. $1 - \alpha$

$P=1-\alpha$	einseitig $F(z)$	zweiseitig $\phi(z)$
90%	1,2816	$\pm 1,6449$
95%	1,6449	$\pm 1,9600$
99%	2,3263	$\pm 2,5758$
99,9%	3,0902	$\pm 3,2910$

b) Wkt. für gegebenes z

z	$F(z)$	$\phi(z)$
0	0,5000	0,0000
1	0,8413	0,6827
2	0,9772	0,9545
3	0,9987	0,9973

Table 4: Die t-Verteilung (Student-Verteilung) für r Freiheitsgrade linksseitige Konfidenzintervalle $t_r < t_{r,p}$ ($p = 1 - \alpha$), Werte für $t_{r,p}$

r	p	.90	.95	.975	.99	.995
1		3.08	6.31	12.71	31.82	63.66
2		1.89	2.92	4.30	6.96	9.92
3		1.64	2.35	3.18	4.54	5.84
4		1.53	2.13	2.78	3.75	4.60
5		1.48	2.02	2.57	3.36	4.03
6		1.44	1.94	2.45	3.14	3.71
7		1.41	1.89	2.36	3.00	3.50
8		1.40	1.86	2.31	2.90	3.36
9		1.38	1.83	2.26	2.82	3.25
10		1.37	1.81	2.23	2.76	3.17
11		1.36	1.80	2.20	2.72	3.11
12		1.36	1.78	2.18	2.68	3.05
13		1.35	1.77	2.16	2.65	3.01
14		1.35	1.76	2.14	2.62	2.98
15		1.34	1.75	2.13	2.60	2.95
16		1.34	1.75	2.12	2.58	2.92
17		1.33	1.74	2.11	2.57	2.90
18		1.33	1.73	2.10	2.55	2.88
19		1.33	1.73	2.09	2.54	2.86
20		1.33	1.72	2.09	2.53	2.85
21		1.32	1.72	2.08	2.52	2.83
22		1.32	1.72	2.07	2.51	2.82
23		1.32	1.71	2.07	2.50	2.81
24		1.32	1.71	2.06	2.49	2.80
25		1.32	1.71	2.06	2.49	2.79
26		1.32	1.71	2.06	2.48	2.78
27		1.31	1.70	2.05	2.47	2.77
28		1.31	1.70	2.05	2.47	2.76
29		1.31	1.70	2.05	2.46	2.76
30		1.31	1.70	2.04	2.46	2.75
40		1.30	1.68	2.02	2.42	2.70
60		1.30	1.67	2.00	2.39	2.66
80		1.29	1.66	1.99	2.37	2.64
100		1.29	1.66	1.98	2.36	2.63
200		1.29	1.65	1.97	2.35	2.60
500		1.28	1.65	1.96	2.33	2.59
∞		1.282	1.645	1.960	2.326	2.576

Ablesebeispiel für symmetrisches, zweiseitiges Intervall bei $r = 5$,
 $p = 1 - \alpha = 0.9 \Rightarrow t_{r,p} = 2.02$.

Tabelle 5: χ^2 -Verteilung

r \ p	0.005	0.01	.025	.05	.95	.975	.99	.995
1	.00	.00	.00	.00	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34
30	16.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	90.53	95.02	100.43	104.22
80	51.17	53.54	57.15	60.39	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	113.15	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	124.34	129.56	135.81	140.17