1 Zufallsvorgänge und Wahrscheinlichkeiten

MANFRED BOROVCNIK



Inhaltsverzeichnis

1.1 Zufallsvorgänge

- 1.2 Wahrscheinlichkeiten
- 1.3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

1.1 Zufallsvorgänge

Zufallsvorgänge und Ergebnismengen

Ereignisse und ihre Verknüpfung

Situationen mit Unsicherheit sind dadurch gekennzeichnet, dass man nicht weiß, welche von mehreren Möglichkeiten eintreten wird. Solche Situationen sind etwa Glücksspiele oder der Abschluss einer Versicherung. Wenn man ein Maß für die Unsicherheit einführt und für das Problem einen Wert, das ist eine Wahrscheinlichkeit, für das Auftreten bestimmter Möglichkeiten angeben kann, wird es möglich, Entscheidungen zu begründen und die Konsequenzen zu beurteilen.

B: Lotto

Entwicklung des Kurses einer Aktie

Abschluss einer Versicherung (Leben, Sache, andere Risiken)

Soll man einen Wettschein ausfüllen? Welche Zahlen soll man ankreuzen? Ist der Preis, den die staatliche Lottogesellschaft dafür verlangt, fair?

Wie entwickeln sich die Kurse von Aktien allgemein? Welche Aktien entwickeln sich besser? Welcher Preis ist beim Kauf zu erlegen? Welche Risiken geht man mit dem Kauf ein?

Soll man eine Kaskoversicherung für das neue, teure Auto abschließen, oder soll man darauf verzichten? Welchen Preis verlangen verschiedene Versicherungsgesellschaften dafür? Welche ist die günstigste? Was ist damit versichert? Lohnt es sich, die Versicherung abzuschließen? Unter welchen Bedingungen lohnt es sich?

Allen Situationen ist die Ungewissheit bezüglich dessen, was tatsächlich eintritt, gemeinsam. Ein erster Schritt in Richtung dieser Modelle besteht darin, wenigstens alle Möglichkeiten aufzulisten, die eintreten können. In einem zweiten Schritt werden diese Möglichkeiten dann mit Gewichten, den so genannten Wahrscheinlichkeiten bewertet.

Die Stochastik liefert Begriffe und Modelle, um vorausschauend eine – transparente – Basis für Entscheidungen zu liefern. Für die Entwicklung der Theorie greifen wir wechselweise auf relevante Anwendungen und den Bereich der Glücksspiele zu, aus dem viele Begriffe historisch stammen. Ohne diesen Bezug auf Glücksspiele sind die Begriffe oft nur schwer zu verstehen. Gleichzeitig liefern Glücksspiele einfache, überschaubare Situationen und konkrete Versionen für diese Begriffe.

1.1 Zufallsvorgänge





- Zufallsvorgänge und Ergebnismengen
- **Ereignisse und ihre Verknüpfung**

Situationen, in denen nicht bekannt ist, was tatsächlich zutrifft, können häufig doch mathematisch modelliert werden, damit man – in einem gewissen Sinne – rationale Entscheidungen trifft. Solche Situationen sind etwa Glücksspiele oder die klassische Versicherung (Leben, Auto etc.). Welcher Art solche Situationen sein müssen, wird hier geklärt. Daran knüpft eine erste mathematische Beschreibung an.

Def.: Zufallsvorgang ist ein Vorgang, bei dem

- alle Ergebnisse, die als möglich betrachtet werden, bekannt sind;
- das tatsächliche Ergebnis im Voraus nicht "ermittelbar" ist.

Zufallsexperiment: geplanter Zufallsvorgang; unter "gleichen Bedingungen wiederholbar".

Bem.: Für eine mathematische Modellierung von Zufallsvorgängen ist mindestens erforderlich, dass man prinzipiell über alle möglichen Ergebnisse Bescheid weiß. Beim Münzwerfen etwa wird nur Kopf oder Zahl anerkannt, andere Ergebnisse – die Münze bleibt z. B. an der Außenkante stehen – gelten damit vorweg nicht als Münzwurf. Wichtig ist auch, dass das Ergebnis mit anderen Methoden nicht "ermittelbar", also unbekannt ist. Wir sprechen von einem Experiment in Anlehnung an die Physik, wenn wir den Zufallsvorgang planen, sodass er *unter gleichen Bedingungen wiederholbar* wird. Allerdings, im Gegensatz zur Physik, wo durch genaues Herstellen der Rahmenbedingungen eines Experiments das Ergebnis durch die Theorie vorhersagbar wird, soll in der Stochastik das Ergebnis unkontrolliert und unkontrollierbar variieren.

B: Münzwurf im Sinne der *Physik*: Wir konstruieren einen Münzwurfapparat, der die Wurfverhältnisse so genau reproduziert, dass die Münze z. B. immer auf Kopf fällt.
 Münzwurf im Sinne der *Stochastik*: Wir werfen die Münze – ohne Manipulationsgedanken – möglichst hoch, ohne Ansehen der Seite, die obenauf liegt etc., sodass sich die Ergebnisse variieren. Genauer wird die Münze ideal gedacht als ein Experiment, das mit gleichen Chancen Kopf oder Zahl ergibt.

Die gleichen Bedingungen führen im physikalischen Experiment immer *zum selben Ergebnis*, im stochastischen Experiment immer *zur selben Chance* für Kopf; hier zur "Wahrscheinlichkeit" ½.

Def.: $\Omega = Ergebnismenge = Menge aller möglichen Ergebnisse eines Zufallsvorgangs <math>\omega \in \Omega$ *Ergebnis* des Zufallsvorgangs; $|\Omega| = Anzahl$ der Elemente

Bem.: Alle Ergebnisse, ob Zahlen oder nicht, werden zu einer Menge Ω zusammengefasst. Diese Ergebnismenge kann endlich oder unendlich sein.

B1: Ziehen eines Loses aus einer Trommel mit Losen, nummeriert von 1, 2, bis N: $\Omega = \{1, 2, ..., N\}$

B2: Würfeln: $\Omega = \{1, 2, ..., 6\}$

B4: 3x Münze werfen: KZK = (K, Z, K) als Tripel notiert $-\Omega = \{(K, K, K), (K, K, Z), ..., (Z, Z, Z)\}; |\Omega| = 2^3$

B6: Mensch-ärgere-dich-nicht: Ein Würfel wird solange geworfen, bis zum ersten Mal "6" erscheint: $\Omega = \{(x_1,...,x_{k-1},6) | x_i \in \{1,2,...,5\}, i=1,2,...,k-1; k \in \mathbb{N}\} \cup \{(x_1,x_2,...) | x_i \in \{1,2,...,5\}\}$ Die ersten Tupel beschreiben den ersten Sechser nach k Würfen; in der zweiten Menge sind alle Folgen enthalten, welche überhaupt keinen Sechser aufweisen; $|\Omega| = \infty$.

B7: Nettoertrag einer Investition in \in : $\Omega = \mathbb{R}$ oder ein geeignetes Intervall.

1.1 Zufallsvorgänge





Zufallsvorgänge und Ergebnismengen

Ereignisse und ihre Verknüpfung

In Wetten kann man auf verschiedenste Ereignisse gleichzeitig setzen. Genauso kann man sich bei Zufallsvorgängen für bestimmte "Kombinationen" von Ereignissen interessieren. Ziel wird sein, Wahrscheinlichkeiten für "einfache" Ereignisse zu modellieren und dann – durch Rechnung – auf komplexere Ereignisse zu übertragen. Den logischen Verknüpfungen von Aussagen entsprechen in der Beschreibung durch Mengen Mengenoperationen.

Def: Ereignisse A, B: Zusammenfassung von Ergebnissen, das sind Teilmengen von Ω .

Spezielle Ereignisse: $\{\omega\}$ *Elementarereignis* \emptyset *unmögliches* Ereignis – Ω *sicheres* Ereignis

Bem.: Sprechweise "Ereignis *tritt ein*": $\omega \in A$; "*bleibt aus*": $\omega \notin A$. Elementarereignis ist einfach ein einzelnes Ergebnis des Zufallsvorgangs. Das unmögliche Ereignis kann nicht eintreten, das sichere tritt immer ein. Bei beiden Ereignissen ist somit nichts mehr zufällig, es hat sich aber mathematisch bewährt, die beiden Extremfälle auch als Ereignis zu betrachten.

Bem.: Man kann mit jedem Ereignis auch eine "Wette" verknüpfen. Die Bank bietet eine Wette auf ein Ereignis A an; ein Spieler, der die Wette annimmt, der also auf A setzt, gewinnt einen vereinbarten Betrag, falls A eintritt, und er verliert einen (anderen) Betrag, falls A ausbleibt. Klar beeinflussen die Chancen, dass A eintritt, ob ein Spieler die Wette annimmt oder nicht. Ganz ähnlich ist die Situation bei einem Versicherungsvertrag.

B9: Lostrommel mit Losen 1, 2, ..., N für alle Studierenden. A: gezogenes Los "gehört" einer weiblichen Studierenden

B: ... hat ein Auto etc.

Bem.: Wie man Mengen durch Aussagen beschreibt, kann man auch Ereignisse durch "Aussagen" festhalten.

Bem.: Zufallsziehung aus einer endlichen Gesamtheit: Ordentlich mischen und ohne Manipulation ziehen.

B2 f: Würfeln mit $\Omega = \{1, 2, ..., 6\}$. Elementarereignisse: $\{1\}, \{2\}, ..., \{6\}$. Weitere Ereignisse etwa: $A = \{2, 4, 6\}$ "gerade Zahl"; $B = \{1, 3, 5\}$ "ungerade Zahl"

Durch Mengenverknüpfungen werden neue Ereignisse definiert.

Darstellung in Venn-Diagramm	Verknüpfung	Sprache	Tritt ein, wenn	
	Durchschnitt: $A \cap B$	A und B	beide eintreten	
	Vereinigung: $A \cup B$	A oder B	wenigstens eines eintritt	
	Differenz: $A - B$	A und nicht B	A eintritt und gleichzeitig B ausbleibt	
	Komplement: $\Omega - A = \overline{A}$	Nicht A	A ausbleibt	

Fig. 1

Def: Mit *n* Ereignissen: $A_1, A_2, ..., A_n$:

 $\bigcap_{i=1}^{n} A_i$ alle Ereignisse treten ein

 $\bigcup_{i=1}^{n} A_i$ wenigstens eines der Ereignisse tritt ein (das Komplement von : "alle bleiben aus")

Def: $A_1, A_2, ..., A_n$ ist eine vollständige Zerlegung von Ω , falls gilt:

 $A_i \neq \emptyset$ (nicht-trivial)

 $A_i \cap A_j = \emptyset$ für $i \neq j$ (paarweise disjunkt)

 $\bigcup_{i=1}^{n} A_{i} = \Omega \text{ (erschöpfend)}$

Bem.: Für Ereignisse gelten die Mengengesetze: Kommutativität, Assoziativität, Distributivität, de Morgan etc. De Morgan etwa: $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

Def: Potenzmenge (besteht aus allen Teilmengen einer Menge): $\wp(\Omega) = \{A \mid A \text{ ist Teilmenge von } \Omega\}$

Bem.: Die Potenzmenge einer endlichen Menge mit n Elementen hat 2^n Elemente. Bei endlichen Ergebnismengen kann man prinzipiell immer alle Teilmengen als Ereignisse zulassen; die Potenzmenge ist der "natürliche" Definitionsbereich von Wahrscheinlichkeit. Gilt jedoch $|\Omega| = \infty$ (genauer: ist Ω überabzählbar), dann muss man die zulässigen Teilmengen, die als Ereignisse in Betracht kommen, einschränken, dazu dient der Begriff *Ereignisalgebra*.

Def.: Ereignisalgebra ist eine echte Teilmenge A der Potenzmenge $\wp(\Omega)$, wenn zusätzlich gilt:

- (A1) $\Omega \in A$
- (A2) $A \in A \Rightarrow \overline{A} \in A$
- (A3) $A, B \in A \Rightarrow A \cup B \in A$

Bem.: Man darf mit Ereignissen "rechnen", ohne die Eigenschaft, ein Ereignis zu haben, zu verlieren. Alle üblichen Rechenoperationen führen dann nicht aus der Ereignisalgebra A hinaus.

B: Roulette mit $\Omega = \{0, 1, 2, ..., 36\}$. Die *kleinste* Ereignisalgebra, die einen Spieler, der nur auf Rouge oder Noir setzt, interessiert, lautet: $A = \{\emptyset, \Omega, R, N, Z, \overline{R}, \overline{N}, \overline{Z}\}$

Satz (Abgeschlossenheit gegenüber anderen Mengenoperationen):

$$A,B\in \ \mathbb{A} \Rightarrow A\cap B\in \ \mathbb{A}$$

$$A_1,\ A_2,...,A_n\in \ \mathbb{A} \Rightarrow \bigcup_{i=1}^n A_i\in \ \mathbb{A}$$

Bem.: Eine Ereignisalgebra ist nicht nur gegenüber der Vereinigung, sondern auch gegenüber dem Durchschnitt abgeschlossen. In den Axiomen wird nur ersteres angeführt, weil sich die zweite Beziehung ableiten lässt. Mathematiker streben möglichst redundanzfreie Axiomensysteme an; diese sind für Praktiker schwerer lesbar, weil sie sich um weitere Gesetze, die auch gelten, "selbst kümmern" müssen. Es ist unmöglich. die zweite Beziehung auf die Vereinigung *abzählbar vieler* Mengen auszuweiten und aus den Axiomen herzuleiten, daher fordert man in der Mathematik ein erweitertes Axiom anstelle von A3, nämlich:

(A3*)
$$A_1, A_2, ..., \in A \Rightarrow \bigcup_{i=1}^{\infty} A_i \in A$$



1 Zufallsvorgänge und Wahrscheinlichkeiten



Inhaltsverzeichnis

- 1.1 Zufallsvorgänge
- 1.2 Wahrscheinlichkeiten
- 1.3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

1.2 Wahrscheinlichkeiten

- **Definition von Wahrscheinlichkeit**
- Laplace-Experimente
- Kombinatorik

Ziel ist es, Ereignissen eine Zahl aus [0,1], zuzuordnen. Diese so genannte Wahrscheinlichkeit entspricht einem Maß für das Eintreten dieses Ereignisses. Damit wir für verschiedene Ereignisse *einen* Vergleichsmaßstab haben, normieren wir die Werte mit 1; 0,5 bedeutet etwa gleiche Chancen für Eintreten und Ausbleiben des Ereignisses und stellt in einer Wette Symmetrie zwischen Bank und Spieler her. Zahlen für diese Wahrscheinlichkeiten bekommt man einerseits durch Information über das Ereignis aus der Vergangenheit, den so genannten relativen Häufigkeiten des Eintretens bei einer schon durchgeführten Serie von Zufallsexperimenten. Oder man bezieht sich auf Zufallsexperimente besonderer Art, bei welchen die einzelnen Versuchsausgänge alle dieselbe Chance haben. Solche Laplace-Experimente zeichnen sich durch eine physikalische Symmetrie des Zufallsmechanismus aus. Münzen, Würfel, Lotto u. ä. fallen darunter. Ein Unterschied besteht zwischen den beiden Möglichkeiten, Werte für die Wahrscheinlichkeit zu bekommen.

- i) Eine Versicherung schätzt aus der Häufigkeit von Schadenszahlungen aus Kasko-Verträgen aus den letzten Jahren deren Wahrscheinlichkeit und bestimmt auf dieser Basis die Höhe der Prämie. Je nachdem, welchen Zeitraum man berücksichtigt, erhält man andere Schätzungen. Andere Versicherungen haben andere Information. Die Frage erhebt sich, wie genau man die Wahrscheinlichkeit dann eigentlich kennt. Weiters wäre zu klären, ob die Information auch auf die Zukunft zutrifft, oder ob sich an den Chancen etwas geändert hat. Oft rechnet man aber mit Schätzwerten so, als ob man die Wahrscheinlichkeit genau kennt.
- ii) In vielen Glücksspielen liegt es nahe, sie als Laplace-Experimente aufzufassen. So "hat" ein Würfel die Wahrscheinlichkeit 1 von 6 = 1/6 für den Sechser, eine Münze "hat" Wahrscheinlichkeit 1 von $2 = \frac{1}{2}$ für Kopf. Allerdings muss man diese Symmetrie erst rechtfertigen und in der Praxis überprüfen. Denn erst sie führt auf gleiche Chancen. Natürlich kann man relative Häufigkeiten einer Serie von Zufallsexperimenten mit den Laplace-Wahrscheinlichkeiten vergleichen, um zu prüfen, ob die gleichen Chancen erfüllt sind.

In Laplace-Experimenten muss man nur richtig zählen können: wie viele Ergebnisse gibt es, und wie viele Ergebnisse führen zum Ereignis. In komplexeren Situationen benötigt man dazu die Kombinatorik, das sind spezielle Techniken des Zählens.

Bevor wir konkrete Zahlen für Wahrscheinlichkeiten von Ereignissen bestimmen, sollen *Regeln* angegeben werden, welchen diese Zuordnung genügen muss. Mathematisch gesehen wird dabei die *Struktur* von Wahrscheinlichkeit festgelegt. Das hilft einerseits, die Grundlagen der Wahrscheinlichkeitsrechnung festzulegen und auf diesen Axiomen eine Theorie aufzubauen, welche wesentliche Fragen behandeln lässt. So etwa auch, wie empirische relative Häufigkeiten modelliert werden. Anderseits wird man persönliche Wahrscheinlichkeitsbewertungen immer mit diesen Regeln vergleichen müssen. Es zeigt sich, dass Menschen in ihren Bewertungen der Wahrscheinlichkeit im Vergleich zu diesen Regeln inkonsistent sind. Etwa werden den einzelnen Versuchsergebnissen Wahrscheinlichkeiten zugeordnet, die in der Summe mehr als 1 ergeben.

1.2 Wahrscheinlichkeiten





Definition von Wahrscheinlichkeit

Laplace-Experimente

Kombinatorik

Wahrscheinlichkeiten, die aus relativen Häufigkeiten geschätzt werden, haben strukturell ähnliche Eigenschaften wie Laplace-Wahrscheinlichkeiten, sieht man einmal davon ab, dass relative Häufigkeiten bei Wiederholung der Serie der Zufallsexperimente fluktuieren. Strukturelle Eigenschaften fassen Mathematiker in Axiomen zusammen. Die üblichen Axiome gehen auf Kolmogoroff 1933 zurück und können als grundlegende Spielregeln im Umgang mit Wahrscheinlichkeiten aufgefasst werden. Mit diesen Axiomen und dem Begriff der *Unabhängigkeit* (siehe 1.3) wird eine reichhaltige Theorie der Wahrscheinlichkeit ableitbar, was ihre spezielle Wahl entsprechend rechtfertigt.

Def. (Axiome für Wahrscheinlichkeit nach Kolmogoroff):

Ω Ergebnismenge, A die zugehörige Ereignisalgebra eines Zufallsvorgangs; dann ist die Zuordnung

P:
$$A \to \mathbf{R}$$
 mit $A \mapsto P(A)$

eine Wahrscheinlichkeit, wenn die Funktion folgende Bedingungen erfüllt:

K1 $P(A) \ge 0$ *Nichtnegativität*

K2 $P(\Omega) = 1$ Normierung

K3 $P(A \cup B) = P(A) + P(B)$, falls $A \cap B = \emptyset$ Additivität

Bem.: Aus den Axiomen kann man keine Zahlen für die in einem Anwendungsfall gesuchten Wahrscheinlichkeiten herleiten. Dazu bedarf es von Fall zu Fall zusätzlicher Festlegungen bei der Modellbildung. Wir werden später (Kapitel 2 und 3) Modelle vorstellen, die den Axiomen genügen; mit diesen Modellen kann man die meisten Anwendungen ausreichend gut erfassen. Die Axiome liefern aber grundsätzliche Muster und Deutungen von Wahrscheinlichkeit, welche eine Hilfe beim Verständnis von Begriffen und Ergebnissen der späteren mathematischen Sätze bieten:

Die statische Variante von Wahrscheinlichkeit ist direkt in den Axiomen enthalten, Wahrscheinlichkeit kann wie eine normierte Fläche (allgemein: ein normiertes Maß) gedeutet werden. Fläche als Zuordnung auf einer Grundmenge ordnet Teilmengen eine Fläche zu. Die Zuordnung ist additiv, das heißt, die Gesamtfläche zweier Mengen ist die Summe der beiden einzelnen Flächen, sofern die beiden Mengen disjunkt sind. Die dynamische Seite von Wahrscheinlichkeit, wie sich das Eintreten eines Ereignisses von Versuch zu Versuch ändert, wird über den Begriff der Unabhängigkeit erfasst werden.

Bem.: Auch hier muss man mathematisch noch eine Erweiterung verlangen, die so genannte σ -Additivität:

K3*
$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$
, falls $A_i \cap A_j = \emptyset$ für alle $i \neq j$ ("paarweise disjunkt").

Bem.: Wir brauchen 3 Komponenten, um einen Zufallsvorgang zu modellieren: Ergebnis*menge*, Ereignis*algebra*, und eine Wahrscheinlichkeit, die K1 bis K3 genügt. Für die Deutung von *Wahrscheinlichkeit als Fläche* ist die Additivität über (paarweise) disjunkte Mengen trivial. Man skizziert den Sachverhalt am einfachsten durch Venn-Diagramme und verwendet diese als Beweisstütze für weitere Sätze.

Satz (Eigenschaften von P):

$$P(\overline{A}) = 1 - P(A)$$
 Komplementärregel $P(\{\}) = 0$ $P(A) \le P(B)$, wenn $A \subset B$ Monotonie $0 \le P(A) \le 1$

Bew.: nur für die erste Beziehung: $A \cup \overline{A} = \Omega$ bildet eine Zerlegung, daher hat man:

$$= \begin{cases} P(A \cup \overline{A}) = P(A) + P(A) & K3 \\ P(\Omega) = 1 & K1 \end{cases}$$

Bem.: Personen verletzen diese Monotonie bei der subjektiven Bewertung von Wahrscheinlichkeiten.

B: Conjunction fallacy nach Tversky und Kahneman (1983)

Bill is 34 years old. He is intelligent, but unimaginative, compulsive and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Which statement is more probable?

- (a) Bill is an accountant who plays jazz for a hobby. Ereignis $A \cap B$
- (b) Bill plays jazz for a hobby. Ereignis B.

87% in der Untersuchung schätzen die Wahrscheinlichkeit der Konjunktion von den beiden Aussagen in (a) als höher ein als die Wahrscheinlichkeit des einzelnen Ereignisses in (b). In der Literatur ist dies unter der "conjunction fallacy" bekannt geworden.

Klarerweise ist aber $A \cap B \subset B$ und damit gilt – mit dem obigen Satz – $P(A \cap B) \leq P(B)$.

Bem.: Die Komplementärregel wird später mit Vorteil ausgenützt werden, wenn Wahrscheinlichkeiten von Ereignissen zu berechnen sind, wo das Komplement ganz einfach und überschaubar ist.

B.: Gesucht ist die Wahrscheinlichkeit, bei drei Würfen einer Münze wenigstens einmal Kopf – Ereignis \overline{A} – zu erhalten. Das komplementäre Ereignis A lautet: "niemals Kopf" = $\{(Z,Z,Z)\}$. Bestimmt man dessen Wahrscheinlichkeit mit $(\frac{1}{2})^3 = \frac{1}{8}$, so gilt: $P(\overline{A}) = 1 - P(A) = 1 - \frac{1}{8} = \frac{7}{8}$.

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

$$P(A_1 \cup A_2 \dots \cup A_n) = P(A_1) + P(A_2) + \dots P(A_n),$$
Erweiterte Additivität
$$falls \ A_i \cap A_j = \emptyset \text{ für alle } i \neq j$$

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

$$P(A \cup B \cup C) = \begin{cases} P(A \cap B) - P(A \cap C) - P(B \cap C) \\ P(A \cap B \cap C) \end{cases}$$

Bem.: Man kann Bilder mit Venn-Diagrammen statt eines Beweises verwenden. Für die allgemeine Vereinigung von 3 Mengen beachte man die Struktur der entstehenden Formel: Addiere die Wahrscheinlichkeiten der Mengen, subtrahiere davon die Wahrscheinlichkeiten der Schnitte von 2 Mengen, addiere die Wahrscheinlichkeiten der Schnitte von 3 Mengen. Es liegt nahe, wie eine Formel für *n* Ereignisse aussehen wird. Bewiesen wird diese dann mit vollständiger Induktion.



1.2 Wahrscheinlichkeiten





- Definition von Wahrscheinlichkeit
- Laplace-Experimente
- Kombinatorik

Mit der Zusatzvereinbarung, dass alle Elementarereignisse die gleichen Chancen haben, erfüllt man nicht nur die grundlegenden Spielregeln einer Wahrscheinlichkeit, wie sie in den Axiomen festgelegt sind. Man erhält gleichzeitig Zahlen für die Wahrscheinlichkeiten. Das ist von besonderem Vorteil, aber: man muss auch Bedingungen einhalten, damit die gleichen Chancen erfüllt sind. Experimente mit dieser Eigenschaft nennt man Laplace-Experimente. Aus den gleichen Chancen der Elementarereignisse ergeben sich durch die Additivität die Wahrscheinlichkeiten von Ereignissen als Anteil von für das Ereignis A günstigen zu möglichen Ergebnissen. Umgekehrt definiert man:

Def.: Zufallsexperiment heißt Laplace-Experiment, wenn

- $|\Omega| = n$ (endlich)
- $P(A) = \frac{|A|}{|\Omega|} = \frac{g \ddot{u} n s t \dot{t} g e}{m \ddot{o} g l \dot{c} h e}$

Bem.: Elementarereignisse haben in Laplace-Experimenten alle dieselbe Wahrscheinlichkeit:

$$P(\{\omega_i\}) = \frac{|\{\omega_i\}|}{n} = \frac{1}{n}.$$

Für die Anwendung des Laplaceschen Modells der Gleichverteilung wird i. A. ein *Symmetrieargument* über das Zufallsexperiment vorgebracht.

Bem.: Es gibt zwei Möglichkeiten zu prüfen, ob ein Zufallsexperiment als Laplace-Experiment gelten kann:

- Man prüft die Symmetrie des Experiments, man sucht, diese Symmetrie strikt einzuhalten.
- Man überprüft laufend die relativen Häufigkeiten der Elementarereignisse, ob diese in etwa gleich sind. Wie man genau vorgeht, wird später noch Thema sein (Kapitel 5 und 6).

B: Beim **Münzwurf** oder beim **Würfeln** wird argumentiert, dass das Zufallsgerät *physikalisch symmet- risch* ist und daher – Manipulation ausgeschlossen – alle Elementarergebnisse dieselbe Wahrscheinlichkeit haben sollten.

Beim **Ziehen** von nicht unterscheidbaren Kugeln **aus einer Urne** mit N Kugeln (nummeriert oder eingefärbt) wird sorgfältig durchgemischt. Danach greift man in die Urne, ohne hineinzusehen, und greift eine (oder mehrere) heraus. Hier sollten *alle möglichen Bevorzugungen* irgendwelcher Kugeln *ausgeschaltet* werden, sodass man auf diese Weise zu gleichen Chancen, Wahrscheinlichkeiten für die Elementarereignisse kommen sollte.

Satz: Die oben festgelegte Wahrscheinlichkeit nach Laplace erfüllt die Kolmogoroff-Axiome einer Wahrscheinlichkeit.

Bew.: nur für die Additivität. Für Anzahlen gilt die Additivität:

$$|E \cup F| = |E| + |F|$$
, falls $E \cap F = \emptyset$

Daher hat man:

$$P(E \cup F) = \frac{|E \cup F|}{n} = \frac{|E|}{n} + \frac{|F|}{n} = P(E) + P(F)$$

B2 f: Würfeln ist "fair", symmetrisch, daher ein Laplace-Experiment: $P(\{i\}) = \frac{1}{6}$; $P("ungerade") = \frac{3}{6} = \frac{1}{2}$

B3 f: Roulette als symmetrisches Zufallsexperiment liefert $P(Rouge) = \frac{18}{37} = 0,4865$.

Bem.: In "einstufigen" Zufallsexperimenten ist sowohl das Abzählen der möglichen als auch der für ein Ereignis günstigen Fälle meist einfach. Schon beim dreifachen Wurf eines Würfels ist schon nicht so klar, was zu tun ist. Es entsteht auch schnell ein "Streit", was als Möglichkeit zu gelten hat. So wie das bei den 3 Würfeln des Fürsten der Toscana (siehe unten) belegt ist. Ist die Reihenfolge, in der man die Ergebnisse notiert, wesentlich oder nicht? Für welche der Möglichkeiten, die Ergebnisse zu notieren, gilt ein Symmetrieargument und damit die Laplace-Wahrscheinlichkeit wirklich? Zur Illustration sei zuerst ein einfaches Beispiel angegeben, dann die 3 Würfel:

- **B:** Beim 2fachen Münzwurf kann man die Anzahl der Köpfe als wesentliches Merkmal betrachten. Für welche Auflistung der Möglichkeiten ist ein Symmetrieargument zutreffend?
 - i) *Ohne* Beachtung der *Reihenfolge*: $\Omega_1 = \{ZZ, KZ, ZZ\}$. E := "ein Kopf in zwei Würfen" entspricht das Elementarereignis $\{KZ\}$. Als Laplace-Experiment gilt $P(E) = \frac{1}{3}$
 - ii) *Mit* Beachtung der *Reihenfolge*: $\Omega_2 = \{(Z,Z), (Z,K), (K,Z), (Z,Z)\}$. Nun gilt: $E = \{(Z,K), (K,Z)\}$ und $P(E) = \frac{1}{4}$, falls wiederum alle Ergebnisse dieselbe Chance haben sollen.

Führt man nun den zweifachen Münzwurf mehrfach durch und zählt die relativen Häufigkeiten von E aus, so liegt diese in der Nähe von $\frac{1}{4}$, was für die Berücksichtigung der Reihenfolge spricht.

B: Die drei Würfel des Fürsten der Toscana: Soll man bei Werfen von 3 Würfeln auf die Augensumme 11 oder 12 setzen? Die Möglichkeiten **641** und **551** ergeben beide die Summe **11**. Hat **641** gleiches Gewicht wie **551**? Oder muss man die Würfel in der Notation unterscheiden, damit die Möglichkeiten gleiches Gewicht und somit gleiche Wahrscheinlichkeiten erhalten. Demnach hätte man

$$(6,4,1),\ (6,1,4),\ (4,6,1),\ (4,1,6),\ (1,6,4),\ (1,4,6)\ bzw.\ (5,5,1),\ (5,1,5),\ (1,5,5)\,,$$

womit der ersten Möglichkeit doppelt so viel Gewicht zukommt wie der zweiten. Für welche Art von Ergebnissen – mit oder ohne Unterscheidung der Würfel – gilt dann ein Symmetrieargument und damit die Regel günstige / mögliche?

Lässt man die Reihenfolge weg, so ergeben sich für die Summe 11 und 12 gleich viele Elementarereignisse und somit hat man P(11) = P(12). Berücksichtigt man die Reihenfolge, so ergeben sich für 12

"nur" 25, für 11 aber 27 günstige Fälle. Da es im zweiten Fall $216 = 6^3$ mögliche Fälle gibt, hat man

$$P(12) = \frac{25}{216} = 0,1157 < P(11) = \frac{27}{216} = 0,125 . Eine , Simulation des Zufallsexperiments spricht für die$$

Berücksichtigung der Reihenfolge.

Vielheit				Vielheit	
Summe 11	i) ohne Reihenfolge	ii) mit Reihenfolge	Summe 12	i) ohne Reihenfolge	ii) mit Reihenfolge
641	1	6	651	1	6
632	1	6	6 4 2	1	6
551	1	3	633	1	3
542	1	6	552	1	3
533	1	3	5 4 3	1	6
443	1	3	444	1	1
gesamt	6	27		6	25



1.2 Wahrscheinlichkeiten





Definition von Wahrscheinlichkeit

Laplace-Experimente

Kombinatorik

Bei Laplace-Experimenten hat jedes Elementarereignis dieselbe Wahrscheinlichkeit. Zur Bestimmung der Wahrscheinlichkeit eines Ereignisses braucht man nur die günstigen Ergebnisse (die zum Ereignis führen) abzählen. Die Wahrscheinlichkeit ist dann der Quotient **günstige** / **mögliche**. Speziell im Rahmen von Glücksspielen, aber nicht nur dort, werden die Mengen sehr unübersichtlich. Es bedarf dazu intelligenter Strategien des Abzählens (=Kombinatorik), die den Mengen eine Struktur aufprägen. Es gibt zwei Grundprobleme: Anordnungen und Auswahlen.

Def.: Anordnung von Objekten mit Nummern von 1 bis n = Permutation

Bem.: Die Objekte sind durch die Nummern unterscheidbar.

Satz: Es gibt $n! = n \cdot (n-1) \cdot ... \cdot 2 \cdot 1$ mögliche Anordnungen von n unterscheidbaren Elementen

Bem.: Der Beweis kann mit Induktion geführt werden, es gilt nämlich die Rekursion:

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

Bem.: 3 Objekte in einer speziellen Anordnung, das 4. kann dann an 4 Stellen eingefügt werden (vor dem ersten, ..., nach dem letzten). Die Zahl der Möglichkeiten wird daher beim Übergang von 3 auf 4 Objekte mit dem Faktor 4 vervielfacht.

Def: Hat man n Objekte, n_1 vom Typ 1, n_2 vom Typ 2, ... n_J vom Typ J (Objekte vom selben Typ sind untereinander nicht unterscheidbar), dann benennt man eine Anordnung dieser Objekte als *Permutation mit identischen Elementen*.

B: Wie viele Anordnungen (=Worte) gibt es mit den Symbolen 0 und 1, wenn man vom Typ 0 drei Elemente hat, vom Typ 1 zwei, und alle fünf Symbole benutzt werden sollen?

Aus dem Wort 0 0 0 1 1, in dem die Nullen vorne stehen, werden durch künstliche Unterscheidung der Nullen 6 neue Worte. Setzt man das Verfahren dann mit den Einsen fort, so erhält man $6 \cdot 2 = 12$ Worte (=Permutationen), wobei alle Symbole verschieden sind.

$$0_1$$
 0_2 0_3 1 1

$$0_1 \quad 0_3 \quad 0_2 \quad 1 \quad 1$$

$$0_2$$
 0_1 0_3 1 1

$$0_2$$
 0_3 0_1 1 1

$$0_3$$
 0_1 0_2 1 1

$$0_3$$
 0_2 0_1 1 1

0 0 0 1 1

Genauso verfährt man mit allen anderen Worten, in denen die Nullen an anderen Plätzen untergebracht sind. Je 12 Worten mit Unterscheidung von Nullen und Einsen untereinander entspricht ein neues Wort ohne Unterscheidung (der Nullen wie der Einsen untereinander). Wenn man jedoch alle Nullen und alle Einsen künstlich unterscheidet, hat man 5!=120 Worte (Permutationen). Ohne Unterscheidung

verbleiben
$$\frac{120}{12} = 10$$
 Worte.

Satz: Es gibt $\frac{n!}{n_1! \cdot n_2! \cdot ... \cdot n_J!}$ solche Permutationen mit identischen Elementen.

Bem.: Man nennt den Quotienten des Ergebnisses Multinomialkoeffizient.

Bew.: Nur Skizze des Gedankens. Seien n_1 vom Typ 1 und alle anderen $n-n_1$ verschieden und sei vorerst $n_1=3$. Versieht man die Elemente vom Typ 1 mit einem Zusatzlabel, damit sie unterscheidbar gemacht werden: a_1, a_2, a_3 , so hat man die Situation mit unterscheidbaren Objekten hergestellt und es gibt n! Permutationen. Wir illustrieren den Sachverhalt, dass je $n_1!=3!=6$ solche Anordnungen sich nur durch die Stellung der a_i unterscheide – wir notieren nur jene Fälle, bei denen die a_i vorne stehen.

Diese sind nach Weglassen der Zusatzlabels jedoch gleich. Also gibt es $\frac{n!}{n_1!}$ Anordnungen. Jetzt führt man

Elemente vom Typ 2 usw. ein und erhält sukzessive die Lösung. Exakt wendet man zum Beweis mehrfache vollständige Induktion an.

Bem.: Die Anordnung erfolgt nach dem Prinzip der lexikographischen Ordnung; das ist ein systematisches Vorgehen, welches den Überblick sichert, ob man alle Möglichkeiten erfasst hat und ob keine Doppelzählungen vorliegen.

Def.: Binomialkoeffizient $\binom{n}{n_1} = \frac{n!}{n_1! \cdot n_2!}$ entspricht der Anzahl der Anordnungen, wenn es 2 Typen von Elementen gibt mit je n_1 bzw. n_2 Elementen gibt ($n_2 = n - n_1$).

Bem.: Teilt man bei Auswahl einer Teilmenge aus einer Menge mit n Elementen die Elemente in 2 Typen: Jene, die in die Teilmenge aufgenommenen wurden, und jene, die nicht aufgenommen wurden, so gibt der Binomialkoeffizient die Anzahl der Teilmengen mit n_1 Elementen an, wenn die Grundmenge n Elemente enthält.

B: Für die Menge $\Omega = \{x_1, x_2, x_3, x_4, x_5\}$ kann man die Teilmenge $E = \{x_2, x_3, x_5\}$ durch das Quintupel $q_E = \{0, 1, 1, 0, 1\}$ charakterisieren.

Allen drei-elementigen Teilmengen von Ω entspricht nun ein (jeweils anderes) Quintupel mit drei 1'en und zwei 0'en, wobei die 1 anzeigt, dass das jeweilige Element zur Teilmenge gehört, die 0, dass es nicht dazu gehört.

Statt alle 3-Teilmengen einer 5-Menge zu zählen kann man auch Anordnungen von $n_1 = 3$ Elementen vom Typ 1 und $n_2 = 2$ Elementen vom Typ 0 zählen.

Davon gibt es:

$$\binom{n}{n_1} = \binom{5}{3} = \frac{5!}{3! \cdot 2!} = 10.$$

Bem.: In der Kombinatorik hilft häufig der Trick, ein Problem mit Auswahlen so umzuformulieren, dass ein neues Problem mit Anordnungen entsteht, welches dann leichter lösbar ist.

Satz (Binomische Formel):

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

Bem.: Der Beweis wird mit vollständiger Induktion geführt, wir lassen ihn weg. Intuitiv ergibt sich beim Ausmultiplizieren der *n* Klammern

 $(a+b)\cdot(a+b)\cdot..\cdot(a+b)$ immer eine gemischte Potenz von $a^k\cdot b^l$,

wobei die Hochzahlen davon abhängen, wie oft man aus einer Klammer beim Ausmultiplizieren a bzw. b wählt. Da man aus jeder Klammer auszuwählen hat, gilt k+l=n, oder l=n-k.

Jetzt braucht man nur zählen, wie oft der Term $a^k \cdot b^{n-k}$ auftritt.

Mit anderen Worten, wie oft kann man aus n Klammern k-mal das a auswählen – oder: wie viele k-Teilmengen aus einer n-Menge gibt es?

Die Antwort lautet: $\binom{n}{k}$. Das erklärt auch schon die Struktur der Formel.

Satz (Multinomische Formel):

$$(a_1 + a_2 + ... + a_J)^n = \sum_{n_1 + ... + n_J = n} {n \choose a_1 \ a_2 \ ... \ a_J} a_1^{n_1} \cdot a_2^{n_2} \cdot ... \cdot a_J^{n_J}$$

Bem.: Beim Ausmultiplizieren der n-ten Potenz des Multinoms $(a_1 + a_2 + ... + a_J)$ entstehen wieder gemischte Potenzen der Ordnung n. Jetzt hat man n_i Elemente vom Typ a_i .

B: Die Konstellationen für die n_i und die zugehörigen Summanden bei Ausrechnung von $(a+b+c)^3$:

n_1	n_2	n_3	"Multi"	Potenz	Summana
3	0	0	$\frac{3!}{3! \ 0! \ 0!}$	$a^3b^0c^0$	a^3
2	1	0	3! 2! 1! 0!	$a^2b^1c^0$	$3a^2b$
2	0	1			$3a^2c$
1	2	0			$3ab^2$
1	0	2			$3ac^2$
0	3	0			b^3
0	2	1			$3b^2c$
0	1	2			$3bc^2$
0	0	3			c^3

Def.: Bei der Auswahl von *k* Kugeln aus insgesamt *n* unterscheidbaren Kugeln (nummeriert) muss man hinsichtlich des Ergebnisses unterscheiden, ob die Reihenfolge der Ziehung relevant ist und ob gezogene Kugeln zurückgelegt werden oder nicht; je nach Beantwortung dieser Fragen handelt es sich um *Variationen* bzw. *Kombinationen*.

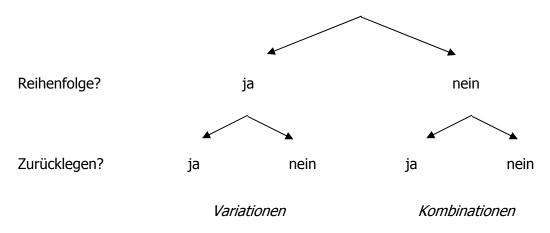


Abb. 1: Entscheidungsbaum bei Auswahlproblemen

Satz: Es gibt $n \cdot n \cdot ... \cdot n = n^k$ Variationen *mit* Zurücklegen.

Bew. Skizze: Die Zahl der Möglichkeiten wird auf jeder "Stufe" (mit jeder weiteren Auswahl) mit dem Faktor *n* vermehrt.

B 14: Man bildet aus den Symbolen 0 und 1 durch Aneinanderreihen "Worte" der Länge 10; man kann $2^{10} = 1024$ verschiedene Worte bilden.

Satz: Es gibt $n \cdot (n-1) \cdot ... \cdot (n-k+1) = \frac{n!}{(n-k)!}$ Variationen *ohne* Zurücklegen.

Bew. Skizze: Die Zahl der Möglichkeiten wird auf jeder weiteren "Stufe" (mit jeder weiteren Auswahl) mit einem Faktor vermehrt, der um 1 kleiner ist als zuvor; der letzte dieser Faktoren ist (n-k+1), denn es sind schon k-1 Elemente "verboten", also nur mehr n-(k-1) noch möglich.

B 15: Mit den Ziffern 1 bis 9 kann man $9.8.7.6 = \frac{9!}{5!}$ Zahlen bilden, bei denen alle Ziffern verschieden sind.

Satz: Es gibt $\frac{n!}{k! (n-k)!} = \binom{n}{k}$ Kombinationen *ohne* Zurücklegen.

Bew.: Skizze: Die Möglichkeiten entsprechen den Möglichkeiten, eine k-elementige Teilmenge aus einer n-Menge auszuwählen. Aus den Variationen ohne Wiederholung erhält man – nach Wegfall der Reihenfolge – aus je k! k-Tupeln dieselbe Menge von ausgewählten Elementen, also $\frac{n!}{(n-k)!}/k!$ Kombinationen ohne Zurücklegen.

Satz: Es gibt $\binom{n+k-1}{k}$ Kombinationen *mit* Zurücklegen.

Bew. Skizze: Zieht man k Elemente aus den n vorhandenen ohne Reihenfolge, aber mit Zurücklegen, so ist das Ergebnis eindeutig durch folgende Liste festgelegt:

Objekt Nr.	#1	#2	# n	Summe
Wie oft gezogen	m_1	m_2	m_n	$m_1 + \ldots + m_n = k$

Diese Belegung kann man durch * für jede Kugel und durch | für die "Behälter" symbolisieren; wir unterstellen n = 4 Elemente und k = 6 Ziehungen:

Objekt Nr.	#1	#2	#3	#4	Summe
Wie oft gezogen	2	1	0	3	6
	**	*		***	

Diese Anordnung aus den Elementen * und | fasst das Ergebnis zusammen. Allgemein hat man dafür k Symbole vom Typ * und n+1 Symbole vom Typ |, von denen aber die beiden Ränder nicht beweglich sind, weil ansonsten * außerhalb von Behältern gelangen können. Wir haben es daher mit einer Permutation mit identischen Elementen zu tun; es gibt davon:

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

B 17: 10 Kandidaten stehen zur Wahl; man kann 3 Stimmen auf Kandidaten häufen (alle drei auf einen Kandidaten oder etwa je eine auf drei verschiedene vergeben). Es gibt dann $\binom{10+3-1}{3} = 220$ Möglichkeiten zu wählen.





1 Zufallsvorgänge und Wahrscheinlichkeiten



Inhaltsverzeichnis

- 1.1 Zufallsvorgänge
- 1.2 Wahrscheinlichkeiten
- 1.3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

1.3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

- **Bedingte Wahrscheinlichkeit**
- ledownRechenregeln für bedingte Wahrscheinlichkeiten
- **Totale Wahrscheinlichkeit und Formel von Bayes**
- **Unabhängigkeit von Ereignissen**

Alle Wahrscheinlichkeit ist bedingte Wahrscheinlichkeit, nämlich bedingt auf vorhandene Information.

B: Beim Würfel etwa stützen wir uns auf die Information, dass er homogen ist und eine Symmetrie-Annahme daher sinnvoll sein mag.

Wenn wir Karten spielen, so können wir fragen, ob noch ein As im Talon liegen wird. Unsere Antwort wird auch davon abhängen, ob wir selbst schon 2 Asse auf der Hand haben oder etwa keines.

Eine Versicherung verlangt bei Abschluss einer Lebensversicherung auf 10 Jahre von einem Sechzigjährigen wesentlich mehr als von einem Dreißigjährigen. Klarerweise hat ein Dreißigjähriger eine viel höhere Wahrscheinlichkeit, dass er den Zeitraum von 10 Jahren überlebt, als ein doppelt so alter Mensch.

Ein wichtiger Bereich ist mit der Beurteilung von unsicheren Ereignissen aufgrund von Indizien umschrieben.

B – Medizinischer Diagnosekontext: Eine Person hat – im Bevölkerungsschnitt – eine Krankheit wie Tbc mit einer Wahrscheinlichkeit von, sagen wir, 1% (=0,01). Kommt diese Person zum Arzt, so wird dieser, bedingt durch Vorliegen einschlägiger Symptome, die Tbc-Wahrscheinlichkeit auf 7% aktualisieren, d. h. eine geänderte, eine bedingte Wahrscheinlichkeit für das Vorliegen von Tbc "errechnen" (wir werden lernen, wie man diese neue Wahrscheinlichkeit bestimmt). Diese Wahrscheinlichkeit "veranlasst" weitere Untersuchungen (Blut, Urin, Röntgen der Lunge etc.). Nach einem positiven Röntgenbefund wird die bedingte Wahrscheinlichkeit auf 60% steigen, bei einem negativen Befund zurück auf 3% sinken. Wenn nach weiteren Untersuchungen die bedingte Wahrscheinlichkeit groß genug ist, wird der Arzt die Behandlung veranlassen.

Immer geht es um neue Wahrscheinlichkeiten für ein Ereignis A, wenn ein anderes Ereignis B (als bekannt) unterstellt wird. Ein Sonderfall besteht darin, dass die bedingte Wahrscheinlichkeit von A unter B gleich der Wahrscheinlichkeit von A vorweg ist, dass also die Wahrscheinlichkeit von A unter B gar nicht von B abhängt. Wir sprechen dann von Unabhängigkeit.

B – Unabhängigkeit von Versuchen: Beim Münzwerfen sollte im zweiten Wurf die Wahrscheinlichkeit für Kopf nicht davon abhängen, ob Kopf oder Zahl beim ersten Mal war. Unser Zufallsexperiment soll ja unter denselben Bedingungen wiederholbar sein. Gerade diese Unabhängigkeit ist Kern von Fehlvorstellungen. So behaupten notorische Roulette-Spieler, dass nach (etwa) 13mal "rot" jetzt "schwarz" kommt und setzen hohe Beträge darauf. Sie "verletzen" damit die Unabhängigkeit und sind von ihrer Interpretation nicht abzubringen.

Für die mathematische Theorie ist diese Unabhängigkeit wichtige Voraussetzung und Bestandteil von wiederholbaren Experimenten. Sie ist also nicht beweisbar; empirische Untersuchungen zu Zufallsexperimenten (durch Analyse von tatsächlichen Ausspielungen im Casino) stützen aber diese Unabhängigkeit.

1.3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit





- **3 Bedingte Wahrscheinlichkeit**
- Rechenregeln für bedingte Wahrscheinlichkeiten
- Totale Wahrscheinlichkeit und Formel von Bayes
- **Unabhängigkeit von Ereignissen**

In vielen Situationen ist die neue, die bedingte Wahrscheinlichkeit, auch als zahlenmäßige Festlegung, klar. Zieht man aus einer Urne mit zwei weißen und zwei schwarzen Kugeln beim ersten Zug einen weiße (=Ereignis W_1 – "weiß beim ersten Zug"), so sind beim zweiten Zug nur mehr 3 Kugeln in der Urne (wenn man nicht zurücklegt), davon ist eine weiß. Die bedingte Wahrscheinlichkeit von "weiß beim zweiten Zug", W_2 , wenn W_1 zutrifft, ist 1/3. Die Wahrscheinlichkeit W_1 , bedingt durch W_2 , also dadurch, dass beim zweiten Zug eine weiße gezogen wird, macht nur einen Sinn, wenn wir die erste gezogene Kugel nicht gesehen haben. Um diese zu bestimmen, müssen wir den Begriff "bedingte Wahrscheinlichkeit" erst definieren und dann hoffentlich – unter Beachtung von Rechenregeln bestimmen.

Mathematisch gesehen ist das beste Argument, dass eine Zuordnung eine Wahrscheinlichkeit ist, die Erfüllung der Axiome von Kolmogoroff. Intuitiv aber ist wichtig, dass man versteht, wie eine Zuordnung entsteht; damit sei hier strukturell gemeint, nicht, wie eine Zahl als bedingte Wahrscheinlichkeit zugeordnet wird. Die folgende Definition ist mit der Deutung von Wahrscheinlichkeit als Fläche einfach zu verstehen.

Def.: Bedingte Wahrscheinlichkeit von A unter B:

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}$$
, falls $P(B) > 0$

Bem.: Wenn das Zufallsexperiment ein Ergebnis aus der Menge B ergibt, so bleibt "nur" mehr die Menge $A \cap B$ übrig; allerdings ist ausgeschlossen, dass ein Ergebnis außerhalb von B liegt; die bedingte Fläche von A unter B ist der Anteil von A, der in B liegt. Für $B = \Omega$ ergibt sich die Fläche von A, weil die Fläche von Ω mit 1 normiert ist.

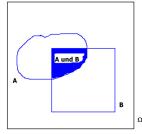


Abb. 2a: Schematische Darstellung der Fläche von A in der Menge B

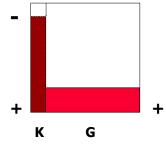


Abb. 2b: $A \equiv krank$; $B \equiv Testergebnis positiv$ +. Anteilige Fläche von A in B = Wahrscheinlichkeit von krank, wenn Test +

B: Bei der Diagnose einer Krankheit K verwendet man einen Blutbefund. Positiv (+) zeigt die Krankheit an, Negativ (-) deutet darauf, dass die Krankheit nicht vorliegt. Allerdings sind die Urteile nicht sicher. Einen Überblick gibt folgende Tabelle (etwa aus vergangenen Daten):

	K	G	Gesamt
+	9	99	108
-	1	891	892
gesamt	10	990	1000

Gibt man ein Los für jede Person in eine Urne und wählt aus dieser zufällig aus, so gilt:

$$P(+) = \frac{108}{1000}, \ P(K \cap +) = \frac{9}{1000} \ \text{und} \ P(K \mid +) = \frac{9}{108} = \frac{9/1000}{108/1000} = \frac{P(K \cap +)}{P(+)}$$

Satz (Eigenschaften bedingter Wahrscheinlichkeiten): Mit P(B) > 0 (B fix) gilt: die Zuordnung

$$P_B: A \to \mathbf{R} \text{ mit } A \mapsto P(A \mid B)$$

ist eine Wahrscheinlichkeit, denn die Funktion erfüllt folgende Bedingungen:

K1 $P(A|B) \geq 0$

 $\mathbf{K2} \quad \mathbf{P}(\Omega \,|\, B) \,=\, 1$

K3 $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$, falls $A_1 \cap A_2 = \emptyset$

Bew. nur für K3:

$$P(A_1 \cup A_2 \mid B) = \frac{P((A_1 \cup A_2) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} = ... = P(A_1 \mid B) + P(A_2 \mid B),$$

denn es ist $(A_1 \cap B) \cap (A_2 \cap B) = \emptyset$ und K3 gilt ja für P.

Bem.: Jede Funktion, welche die Axiome K1-K3 erfüllt, wird Wahrscheinlichkeit genannt.

Bem.: Bedingte Wahrscheinlichkeiten sind mit vielen Fehlvorstellungen verbunden. Es sind Wahrscheinlichkeiten für Ereignisse, die für den Fall gelten sollen, dass die Bedingung zutrifft. Im Beispiel mit der medizinischen Diagnose kann man einige Schwierigkeiten besser ansprechen: Während die Wahrscheinlichkeit, dass jemand positiven Befund hat, falls die "Krankheit" vorliegt, kausal interpretiert werden kann, ist die Wahrscheinlichkeit, die Krankheit zu haben, falls ein positiver Befund vorliegt, nur "indikativ", jedenfalls kann man nicht behaupten, dass das Vorliegen irgendeines Befundes eine Krankheit "verursacht". Es zeigt sich nun in der Praxis, dass Menschen "kausal interpretierbaren" bedingten Wahrscheinlichkeiten viel mehr Beachtung schenken, während "indikative" bedingte Wahrscheinlichkeiten übersehen und durch andere, "kausal interpretierbaren" ersetzt werden. Das erklärt das gar nicht so selten auftretende folgende Argumentationsmuster:

Da
$$P(+|K) = 0.9$$
, gilt auch $P(K|+) = 0.9$.

B.: Die Falk-Urne: In einer Urne sind 2 weiße und 2 schwarze Kugeln. Man zieht zweimal, ohne die gezogene Kugel zurückzulegen. Die erste wird verborgen, die zweite ist weiß. Wie groß ist die Wahrscheinlichkeit, dass die erste auch weiß ist?

Mit den Bezeichnungen der einleitenden Frage sucht man $P(W_1|W_2)$.

Aus der Definition der bedingten Wahrscheinlichkeit hat man:

$$P(W_1 | W_2) = \frac{P(W_1 \cap W_2)}{P(W_2)} = \frac{1}{6} / \frac{1}{2} = \frac{1}{3},$$

wie wir noch genauer rechnen werden.

Bem.: Wir werden die Wahrscheinlichkeiten in Zähler und Nenner noch berechnen lernen. Das Ergebnis klingt für viele Personen paradox, weil das Auftreten des zeitlich späteren W_2 die Wahrscheinlichkeit für W_1 von $\frac{1}{2}$ auf $\frac{1}{3}$ senkt. Viele "glauben", dass die erste Kugel "unabhängig" von der Farbe der zweiten mit $\frac{1}{2}$ auf weiß lauten sollte. Auch hier zeigt sich, dass bedinge Wahrscheinlichkeiten intuitiv ein schwieriges Konzept ausmachen. Ist die Bedingung B ohne kausalen Einfluss für ein Ereignis A, so muss das keineswegs gelten: P(A|B) = P(A). Man kann dennoch aus dem Vorliegen von B etwas für A lernen, wodurch man die bedingte Wahrscheinlichkeit neu berechnen muss. Für die Falk-Urne heißt das, eine weiße Kugel beim zweiten Zug ist *ein Indiz dafür*, dass die 2. Kugel aus einer Urne mit "mehr weißen" gezogen wurde, sodass beim 1. Zug eher schwarz gezogen wurde. Das spiegelt die bedingte Wahrscheinlichkeit: W_1 sinkt von $\frac{1}{2}$ auf $\frac{1}{3}$!



1.3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit





- lacksquare**Bedingte Wahrscheinlichkeit**
- Rechenregeln für bedingte Wahrscheinlichkeiten
- Totale Wahrscheinlichkeit und Formel von Bayes
- **Unabhängigkeit von Ereignissen**

Es ist ganz wesentlich, die folgenden Rechenregeln zu beachten, will man rational mit vorhandenen Wahrscheinlichkeiten umgehen. Die Erfahrung lehrt, dass intuitive Bewertungen von Wahrscheinlichkeiten diese Regeln missachten und zu anderen Wahrscheinlichkeiten führen. Bedingte Wahrscheinlichkeiten sind besonders davon betroffen, weil sie so stark von unseren Vorstellungen, die auch kausal geprägt sind, abweichen.

Satz: $P(A|\Omega) = P(A)$

P(A|B) = 1, wenn $B \subset A$

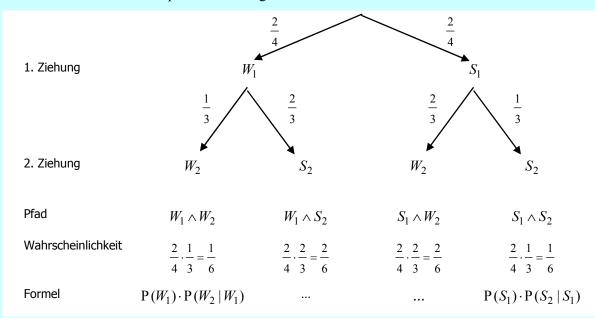
Multiplikationsregel:

 $P(A \cap B) = P(A \mid B) \cdot P(B)$, wenn P(B) > 0

 $P(A \cap B) = P(B \mid A) \cdot P(A)$, wenn P(A) > 0

Bew.: Umstellen der Definition der bedingten Wahrscheinlichkeit liefert die Multiplikationsregel.

B: Falk-Urne (Fortsetzung): Die Technik der *Baumdiagramme* erlaubt eine graphische Darstellung von kombinierten Zufallsversuchen, aus der Berechnungen leicht ablesbar werden. Es ist ein Baum (von oben nach unten oder von links nach rechts). Zuerst verzweigt man nach den möglichen Ergebnissen des 1. Versuchs. An jedem Knoten, der dadurch entsteht, verzweigt man nach den Möglichkeiten des 2. Versuchs. Jeder Pfad entspricht einem Ergebnis des kombinierten Versuchs.



Die einzelnen Kanten werden mit den bedingten Wahrscheinlichkeiten der Verzweigung beschriftet. Die Wahrscheinlichkeiten längs eines Pfades von oben nach unten ergeben sich nach der Multiplikationsregel als Produkt der bedingten Wahrscheinlichkeiten der Verzweigung. Man beachte: Es ist nicht egal, ob beim ersten Mal eine weiße oder eine schwarze Kugel gezogen wurde – die bedingten Wahr-

scheinlichkeiten für eine weiße Kugel beim 2. Zug sind $\frac{1}{3}$ bzw. $\frac{2}{3}$.

Multiplikationsregel für 3 Ereignisse:

$$P(A \cap B \cap C) = P(A \mid B \cap C) \cdot P(B \mid C) \cdot P(C)$$
, wenn $P(B \cap C) > 0$

Bew.: $P(A \cap (B \cap C)) = P(A | B \cap C) \cdot P(B \cap C)$ usf.

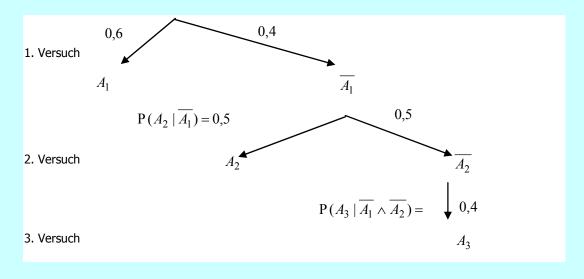
B 22: Bei 3 Versuchen für eine Prüfung bezeichne A_i "besteht beim *i*-ten Versuch".

Aus Daten wurde "geschätzt": $P(A_1) = 0.6$, $P(A_2 \mid \overline{A_1}) = 0.5$, $P(A_3 \mid \overline{A_1} \cap \overline{A_2}) = 0.4$.

Die Wahrscheinlichkeit, die Prüfung zu bestehen ist

 $P(A_1 \cup A_2 \cup A_3) = 1 - P(\overline{A_1} \cup A_2 \cup A_3) = 1 - P(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3})$ Komplementärregel & de Morgan $= 1 - P(\overline{A_3} \mid \overline{A_1} \cap \overline{A_2}) \cdot P(\overline{A_2} \mid \overline{A_1}) \cdot P(A_1) = 0.6 \cdot 0.5 \cdot 0.4 = 0.88$ Multiplikationsregel

Die Berechnungen kann man sich wieder am Baumdiagramm veranschaulichen:



Bem.: Bedingte Wahrscheinlichkeiten werden mit Vorteil für die Modellierung von Zufallsexperimenten mit mehreren Stufen verwendet. Im Beispiel mit der Prüfung aus Beispiel 22 bestehen die Stufen aus den einzelnen Versuchen, bei der Prüfung anzutreten, um sie zu bestehen. Kennt man die entsprechenden bedingten Wahrscheinlichkeiten, so erhält man – durch Rechnung – die Wahrscheinlichkeit beim Gesamtexperiment.

B 22f: Besteht ein Versuch aus 3 Stufen, so beschreibt man das Ergebnis durch ein Tripel (x_1, x_2, x_3) mit $x_i \in \Omega_i$, der Ergebnismenge der i-ten Stufe.

Der gemeinsame Ergebnisraum ist das kartesische Produkt der Ergebnisräume der einzelnen Stufen: $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$:

Manche Teilmengen betreffen nur den 1. Versuch. Beim dreifachen Münzwurf entspricht der Aussage "Kopf beim ersten Versuch" entspricht die Menge $A_1 = \{(1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$, wenn man Kopf mit 1 und das Gegenteil mit 0 codiert.

Manchmal, wie bei der Prüfung mit 3 Versuchen, ist der Ergebnisraum nur eine Teilmenge des kartesischen Produkts $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ (weil die Prüfung nach einem Erfolg beendet ist.)

Bem.: Manchmal rechnet man mit Vorteil mit den entsprechenden Aussagen statt mit den Mengen. Beim 1. und 2. Mal nicht zu bestehen wird durch $\overline{A_1} \wedge \overline{A_2}$ beschrieben. Dieser Konjunktion entspricht die Menge $\overline{A_1} \cap \overline{A_2}$. Man verwendet dieselben Buchstaben für die Aussage und die entsprechende Menge. Normalerweise entstehen dadurch keine Probleme. Man wechselt in jene Darstellung, in der die Berechnung leichter wird.



1.3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit





- Bedingte Wahrscheinlichkeit
- Rechenregeln für bedingte Wahrscheinlichkeiten
- Totale Wahrscheinlichkeit und Formel von Bayes
- Unabhängigkeit von Ereignissen

Die Totale Wahrscheinlichkeit ist am besten mit einer Fallunterscheidung zu vergleichen. Kennt man in den 2 (allgemein n) möglichen Fällen A_i jeweils die bedingte Wahrscheinlichkeit $P(B | A_i)$ eines Ereignisses B, so kann man die Wahrscheinlichkeit P(B) berechnen, sofern man nur auch die Wahrscheinlichkeiten $P(A_i)$ der Fälle kennt. Die Bayes-Formel verwendet dieses Resultat und ermöglicht, aus $P(B | A_i)$ die umgekehrte bedingte Wahrscheinlichkeit $P(A_i | B)$ zu berechnen. Die Bayes-Formel ist zentral für die Neubewertung von Wahrscheinlichkeiten aufgrund von Indizien, wie dies im medizinischen Diagnose-Kontext der Fall ist.

Satz (Totale Wahrscheinlichkeit):

 $A_1, A_2, ..., A_n$ ist eine vollständige Zerlegung von Ω ; $B \in A$. Dann gilt:

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

Bew. nur für n = 2 (der allgemeine Beweis erfolgt mit Induktion): Nach K2 hat man

$$P(B) = P(B \cap A_1) + P(B \cap A_2)$$

Das Ergebnis erhält man, wenn die Schnitte durch die Multiplikationsregel ersetzt:

$$P(B \cap A_i) = P(A_i) \cdot P(B \mid A_i)$$

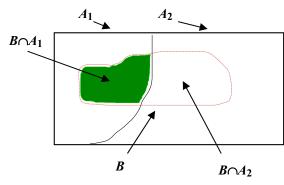


Abb. 3: Zerlegung einer Menge *B* in zwei Teilstücke

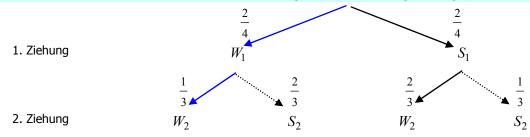
B Falk-Urne (Fortsetzung): Beim ersten Zug kann W_1 oder S_1 eintreten. Die beiden Fälle sind einander ausschließend und erschöpfend, d. h., sie bilden eine Fallunterscheidung. Für die bedingten Wahrscheinlichkeiten beim 2. Zug hat man:

 $P(W_2 \mid W_1) = \frac{1}{3}$ (eine weiße von drei verbleibenden) bzw. $P(W_2 \mid S_1) = \frac{2}{3}$ (beide weißen sind noch da).

Nach dem Satz von der totalen Wahrscheinlichkeit hat man dann:

$$P(W_2) = P(W_2 \mid W_1) \cdot P(W_1) + P(W_2 \mid S_1) \cdot P(S_1) = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{2}$$

Im Baumdiagramm entspricht der Regel der totalen Wahrscheinlichkeit das Aufsuchen aller Pfade, die zum genannten Ereignis – hier W_2 – führen und dem Summieren der Wahrscheinlichkeiten dieser Pfade. Man lässt die anderen Pfade, die nicht zum Ereignis führen, meist ganz weg (hier strichliert)



Satz (Formel von Bayes):

 $A_1, A_2, ..., A_n$ ist eine vollständige Zerlegung von Ω ; $B \in A$. Dann gilt:

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

Bew.:
$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)}$$

Jetzt ersetzt man den Zähler nach der Multiplikationsregel, den Nenner nach der totalen Wahrscheinlichkeit.

B Falk-Urne: Schließt man vom Ergebnis des zweiten Versuchs auf den ersten, und hat der zweite Versuch eine weiße Kugel ergeben W_2 (entspricht B in der Bayes-Formel), so sucht man $P(W_1 | W_2)$.

Die Fälle für den ersten Versuch sind W_1 oder S_1 (entsprechen den A_i in der Formel).

Im Nenner der Bayes-Formel hat man (nach der Totalen Wahrscheinlichkeit alle Pfade, die nach W_2 kommen):

$$P(W_2) = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{2}$$

Im Zähler hat man (Pfade über den Knoten W_1 , die zu einem Knoten mit W_2 führen):

$$P(W_2 | W_1) \cdot P(W_1) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$
.

Damit ist die gesuchte Wahrscheinlichkeit: $P(W_1 | W_2) = \frac{1/6}{1/2} = \frac{1}{3}$.

B 24: Labortest für die Krankheit K; man hat folgende Wahrscheinlichkeiten:

$$P(K) = \pi$$
; $P(+|K) = 0.99$; $P(+|\overline{K}) = 0.02$ "false positives".

Eine Person wird zufällig ausgewählt. Bestimmen Sie: a) P(+); b) P(Diagnose korrekt); c) P(K | +).

a) Nach der totalen Wahrscheinlichkeit gilt

$$P(+) = P(+|K) \cdot P(K) + P(+|\overline{K}) \cdot P(\overline{K}) = 0.99 \cdot \pi + 0.02(1-\pi) = 0.97 \cdot \pi + 0.02$$

b) korrekt =
$$(+ \cap K) \cup (- \cap \overline{K})$$
; $P((+ \cap K) \cup (- \cap \overline{K})) = \pi \cdot 0.99 + (1 - \pi) \cdot 0.98 = 0.01 \cdot \pi + 0.98$

c) nach a):
$$P(K \mid +) = \frac{P(K \cap +)}{P(+)} = \frac{\pi \cdot 0.98}{0.97 \cdot \pi + 0.02}$$

Bem.: Die Wahrscheinlichkeiten ohne Bedingung heißen auch *a priori*, mit Bedingung *a posteriori* (relativ zur Kenntnis des Befundes); die Neubewertung von Wahrscheinlichkeiten erfolgt aufgrund von "*Indizien*".

1.3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit



- - **Bedingte Wahrscheinlichkeit**
- Rechenregeln für bedingte Wahrscheinlichkeiten
- Totale Wahrscheinlichkeit und Formel von Bayes
- Unabhängigkeit von Ereignissen

Mit Hilfe der Axiome – anders gesprochen – den "Spielregeln" für Wahrscheinlichkeiten kann man viele solche Sätze (mathematisch) herleiten. Es entsteht eine interessante Theorie, die wichtige Eigenschaften von Wahrscheinlichkeiten erschließt; diese Eigenschaften kann und muss man dann auch wieder "intuitiv" interpretieren – damit man sie besser und direkter versteht und damit man sie sich sicher merkt und im Anwendungsfall beachtet.

Die Unabhängigkeit entspricht dem Fall, dass die bedingte Wahrscheinlichkeit zu keinem neuen Ergebnis führt. Die Bewertung von Ereignis A ist von Eintreten oder Ausbleiben von Ereignis B unabhängig. Dieser Sonderfall beschreibt am besten die Situation in einem grundlegenden Experiment in der Wahrscheinlichkeitstheorie: Wenn ein und dasselbe Experiment unabhängig voneinander wiederholt wird. Deswegen wird er in der weiteren Entwicklung der Theorie noch eine wichtige Rolle spielen.

```
Def.: Ereignisse A und B heißen
i) unabhängig: P(A \cap B) = P(A) \cdot P(B)
ii) abhängig: nicht unabhängig.
```

Bem.: Die Unabhängigkeit ist eine symmetrische Relation zwischen Ereignissen. Sind zwei Ereignisse unabhängig, dann ist (falls P(A), P(B) > 0) wechselweise keine Neubewertung nötig:

```
Aus P(A \mid B) = P(A) folgt P(B \mid A) = P(B) und umgekehrt.
```

Bem.: A, B unabhängig, so sind auch A, \overline{B} unabhängig usf.

```
B 25: 2x Werfen einer Münze als Laplace-Experiment \Omega = \{(Z,Z),...,(W,W)\}
Z_1 \equiv \text{Zahl beim 1. Wurf} \equiv \{(Z,Z),(Z,W)\}
W_2 \equiv \text{Wappen beim 2. Wurf}; W_{12} \equiv \text{Wappen bei beiden Würfen.}
Dann sind Z_1 und W_2 unabhängig; jedoch sind W_2 und W_{12} abhängig, wie man leicht nachrechnet.
```

Bem.: Unterscheidet man die Ergebnisse beim zweimaligen Münzwurf nach der Reihenfolge und modelliert man diesen Ergebnisraum als Laplace-Experiment, so erhält man eine interessante Eigenschaft: Alle Ereignisse, die den 1. Wurf betreffen wie Z_1 sind von allen Ergebnissen, die nur den 2. Wurf betreffen wie Z_2 oder W_2 unabhängig. Das gilt auch umgekehrt: was den 2. Wurf betrifft, ist von Ereignissen, die den 1. Wurf betreffen, unabhängig!

Sind die einzelnen Stufen eines Zufallsexperiment ein Laplace-Experiment und ist das aus mehreren Stufen zusammengesetzte Experiment ein Laplace-Experiment, so sind die Stufen voneinander unabhängig. Umgekehrt: Sie die einzelnen Stufen Laplace-Experimente und voneinander unabhängig, so wird das zusammengesetzte Zufallsexperiment ein symmetrisches Experiment mit gleichen Chancen für alle Elementarereignisse, sofern man die Ergebnisse als Tupel beschreibt, d. h., die Reihenfolge berücksichtigt.

B: Experiment 1: Drei Laplace-Würfel, die durch ihre Farbe unterscheidbar sind, werden auf einmal geworfen.

Experiment 2: Ein Laplace-Würfel wird drei Mal hintereinander geworfen. Und zwar so, dass jeder Wurf eine Wiederholung desselben Experiments darstellt.

Experiment 3: Drei nicht unterscheidbare, gleichartige, symmetrische Würfel werden auf einmal geworfen.

Bei allen drei Experimenten handelt es sich um dasselbe Experiment mit denselben Wahrscheinlichkeiten. Bei Experiment 3 kann man nicht alle Ereignisse beobachten, wie es in 1 und 2 möglich ist. Sofern man ein Ereignis in Experiment 3 beobachten kann, hat es dieselbe Wahrscheinlichkeit wie in den zwei anderen Experimenten.

Bem.: Experiment 1 und 2 heißen "äquivalente" Experimente.

```
Def.: n Ereignisse A_1, A_2, ..., A_n heißen paarweise unabhängig, wenn: P(A_i \cap A_j) = P(A_i) \cdot P(A_j) für i \neq j; vollständig unabhängig, wenn: P(A_{i_1} \cap ... \cap A_{i_m}) = P(A_{i_1}) \cdot ... \cdot P(A_{i_m}) für alle Auswahlen von Ereignissen.
```

Bem.: Für n = 3 bedeutet paarweise unabhängig, dass gilt:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2), \ P(A_1 \cap A_3) = P(A_1) \cdot P(A_3), \ P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$$

Für die vollständige Unabhängigkeit muss auch noch gelten:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$$

```
B 26: 2x Würfeln als Laplace-Experiment, also \frac{1}{36} - Wahrscheinlichkeit für jedes Paar. A_1 \equiv 1. Wurf ungerade; A_2 \equiv 2. Wurf ungerade; A_3 \equiv Augensumme ungerade. Dann sind A_1, A_2, A_3 paarweise unabhängig, aber nicht vollständig unabhängig.
```

```
Def.: Bernoulli-Experiment: Ein Zufallsvorgang, der nur auf "Erfolg' bzw. "Misserfolg' ausgehen kann. Bernoulli-Versuchsreihe: n unabhängige Wiederholungen davon, d. h.: Mit A_i "i-ter Versuch hat Erfolg' muss gelten: A_1, A_2, ..., A_n sind (vollständig) unabhängig und P(A_i) = p für alle i.
```

Bem: Jedes Glücksspiel ist eine Bernoulli-Versuchsreihe. Ferner jeder wiederholbare Vorgang, sofern man die beiden Forderungen gut "einlösen" kann,, d. h. dieselbe "Treffer"wahrscheinlichkeit bei allen Versuchen, unabhängig davon, ob man in anderen Versuchen einen Treffer hatte oder nicht.



2 Zufallsvariable und ihre Verteilungen

MANFRED BOROVCNIK

Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches. Druckfehler bitte rückmelden. Einige Links funktionieren schon.



2.1 Grundbegriffe

- 2.2 Verteilungsparameter
- 2.3 Spezielle diskrete Verteilungen
- 2.4 Spezielle stetige Verteilungen
- 2.5 Mathematische Ergänzungen

2.1 Grundbegriffe

- Zufallsvariable
- Verteilungsfunktionen
- Quantilfunktionen
- **Diskrete Zufallsvariable**
- Stetige Zufallsvariable
- Affin-lineare Transformation von Zufallsvariablen

Den Ergebnissen eines Zufallsvorgangs werden Werte zugeordnet. Es interessieren die Wahrscheinlichkeiten, mit denen diese Werte angenommen werden. Hier werden die wichtigsten Grundbegriffe erläutert.

2.1 Grundbegriffe





- Zufallsvariable
- **Verteilungsfunktionen**
- **Quantilfunktionen**
- **Diskrete Zufallsvariable**
- Stetige Zufallsvariable
- Affin-lineare Transformation von Zufallsvariablen

Die Zuordnung von Werten (= reelle Zahlen) zu Ergebnissen eines Zufallsvorgangs erlaubt es, die vorhandenen Wahrscheinlichkeiten auf dem Ergebnisraum auf die reellen Zahlen zu übertragen.

Def.: (Ω, A, P) Modell für einen Zufallsvorgang. *Zufallsvariable* ist eine Abbildung:

 $X: \Omega \to \mathbf{R} \text{ mit } \varpi \mapsto X(\omega)$

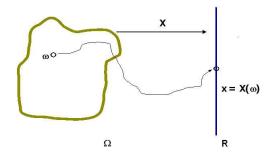


Abb. 1: Zufallsvariable als Funktion

Bem.: Zufallsvariable ist eine *Funktion*; $X(\omega) = x$ heißt *Wert* oder *Realisierung* der Zufallsvariablen; X bezeichnet die Funktion, x ihre Werte (Schreibung ist case sensitive)

Bem.: Mathematisch sind nicht alle Funktionen auf der Ereignisalgebra gleich auch Zufallsvariable. Es reicht jedoch zu fordern, dass die Urbilder Ereignisse sind

(*)
$$\{\omega \mid X(\omega) \le x\} = X^{-1}((-\infty, x]) \in A$$
,

d. h. ein Ereignis aus \mathbb{A} ist ($x \in \mathbb{R}$ beliebig). Solange man auf alle Teilmengen von Ω als Ereignisse zugreifen kann, ist (*) überhaupt keine Einschränkung.

B1: 3x Münzwerfen mit $\Omega = \{(K, K, K), (K, K, Z), ..., (Z, Z, Z)\}.$

 $X(\omega)$ ist dabei Anzahl der Köpfe in ω ;

z.B. gilt: X((K,K,Z)) = X((K,Z,K)) = X((Z,K,K)) = 2

B2: Roulette mit $\Omega = \{0, 1, 2, ..., 36\}$. Der Nettogewinn eines Spielers, der auf C1 = "1. Kolonne" =

$$\{1, 4, 7, ..., 31, 34\}$$
 setzt, lautet: $Y(\omega) = \begin{cases} 2 & \omega \in C_1 \\ -1 & \omega \in \overline{C_1} \end{cases}$

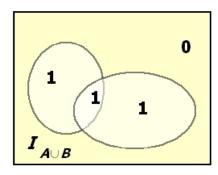
Def.: Indikatorvariable:
$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \in \overline{A} \end{cases}$$

Bem.: Man setzt aus den Indikatorvariablen komplexere Zufallsvariable zusammen; sie beschreiben z. B. Bernoulli-Versuchsreihen.

Satz:
$$I_{\overline{A}}(\omega) = 1 - I_{A}(\omega)$$

 $I_{A \cap B}(\omega) = I_{A}(\omega) \cdot I_{B}(\omega)$
 $I_{A \cup B}(\omega) = I_{A}(\omega) + I_{B}(\omega) - I_{A}(\omega) \cdot I_{B}(\omega)$

Bew.: Anhand der Skizze kann man sich die Gültigkeit der dritten Beziehung überlegen:



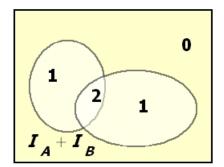


Abb.. 2: Rechnen mit Indikatorfunktionen

B2 f:
$$Y(\omega) = 2 \cdot I_C(\omega) - 1 \cdot I_{\overline{C}}(\omega) = ... = 3 \cdot I_C(\omega) - 1$$

Bem.: Für $\Omega \subset \mathbf{R}$ kann man auch die Zuordnung $X(\omega) = \omega$ als Zufallsvariable betrachten.



2.1 Grundbegriffe





- Zufallsvariable
- Verteilungsfunktionen
- Quantilfunktionen
- Diskrete Zufallsvariable
- Stetige Zufallsvariable
- Affin-lineare Transformation von Zufallsvariablen

Def.: X eine Zufallsvariable; F heißt *Verteilungsfunktion* von X, wenn:

$$F: \mathbf{R} \rightarrow [0,1]$$

$$x \mapsto F(x) := P(\{\omega \mid X(\omega) \le x\})$$

Bem.: Man schreibt kürzer: $\{\omega \mid X(\omega) \le x\} = \{X \le x\}$ und $P(\{\omega \mid X(\omega) \le x\}) = P(X \le x)$.

Bem.: $\{X \le x\}$ ist aus dem Ereignisfeld A und somit ein Ereignis und hat daher eine Wahrscheinlichkeit.

Satz (Eigenschaften von Verteilungsfunktionen):

Ist F Verteilungsfunktion einer Zufallsvariablen X, so hat F folgende Eigenschaften:

- 1. monoton wachsend: $F(x) \le F(y)$ für x < y
- 2. wächst von 0 bis 1: $\lim F(x) = 0$, $\lim F(x) = 1$
 - $x \rightarrow -\infty$ $x \rightarrow \infty$
- 3. rechtsseitig stetig: $\lim_{x \to 0} F(y) = F(x)$

 $y \rightarrow x+0$

Bem.: Eine Funktion F mit den genannten 3 Eigenschaften legt *eindeutig* eine Verteilungsfunktion einer Zufallsvariablen fest. Eine Verteilungsfunktion muss nicht stetig sein: sie kann Sprungstellen mit Sprunghöhen < 1 haben. Die Notation $y \rightarrow x + 0$ meint, dass der Grenzübergang gegen x nur von rechts her erfolgt.

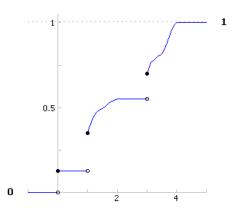


Abb. 3 a: Schematischer Verlauf einer Verteilungsfunktion

B: Die Zufallsvariable X - Anzahl der Köpfe bei 3x Münzwerfen – hat (als Laplace-Experiment) folgende Verteilungsfunktion:

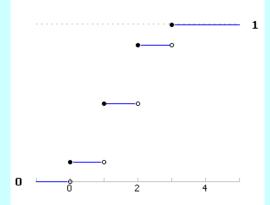


Abb. 3 b: Verteilungsfunktion beim dreifachen Münzwurf

Bem.:
$$F(2+0) - F(2) = P(X=2)$$
;

die Sprunghöhe gibt daher die Wahrscheinlichkeit der Sprungstelle an.

Bem.: Die Verteilung einer Zufallsvariablen wird durch ihre Verteilungsfunktion *vollständig* festgelegt. – Die Zuordnung selbst $X: \Omega \to R$ ist eher theoretisch interessant.

B4: Die Zufallsvariable *Lebensdauer* eines elektronischen Bauteils habe als Modell folgende Verteilungsfunktion (*Exponentialverteilung*)

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda \cdot x} & x \ge 0 \end{cases}$$

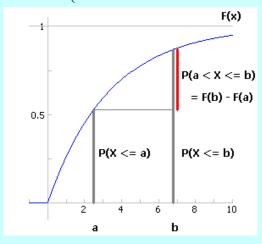


Abb. 3 c: Verteilungsfunktion einer Lebensdauer

Satz: Aus der Verteilungsfunktion F erhält man die Wahrscheinlichkeiten für Intervalle:

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

$$P(X = x) = \lim_{x \to 0} P(X \le x + h) - P(X \le x) = F(x + 0) - F(x)$$

$$P(a \le X \le b) = P(X \le b) - P(X < a) = F(b) - F(a - 0)$$

2.1 Grundbegriffe



- Zufallsvariable
- Verteilungsfunktionen
- Quantilfunktionen
- Diskrete Zufallsvariable
- Stetige Zufallsvariable
- Affin-lineare Transformation von Zufallsvariablen

Wie viel Prozent der Verteilung liegt unter bzw. über einem Punkt? Die Umkehrung dieser Frage lautet: Wo liegt der Punkt, unter welchem sich 100 p % der Verteilung befinden?

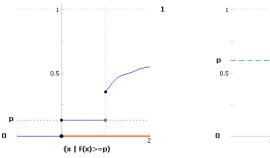
Def.: F sei *stetige* Verteilungsfunktion, 0 . $Die Lösung der Gleichung <math>P(X \le x) = p$ heißt p-Quantil und wird mit x_p bezeichnet. Die Zuordnung $Q: p \mapsto x_p$ heißt Quantilfunktion.

Bem.: Für stetige Verteilungsfunktionen ist die angegebene Gleichung eindeutig auflösbar.

B4 f: Die Lebensdauer X habe die Verteilungsfunktion $F(x) := 1 - e^{-\lambda \cdot x}$ (für x > 0) mit $\lambda > 0$ (Exponential-verteilung). Die Quantile bestimmt man sich aus der Gleichung $F(x) = 1 - e^{-\lambda \cdot x} = p$.

Nach kurzen Umformungen erhält man $x_p = \frac{1}{\lambda} \cdot \ln(\frac{1}{1-p})$

Bem.: Hat die Verteilungsfunktion Sprünge, so gibt es für manche *p* keine Lösung, die Verteilungsfunktion ist nicht umkehrbar. Hat die Verteilungsfunktion waagrechte Abschnitte, so ist die Lösung dort mehrdeutig.



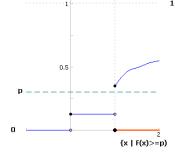


Abb. 4a: Keine Lösung für die Quantilgleichung

Abb. 4b: Mehrere Lösungen

Def.: X eine Zufallsvariable mit Verteilungsfunktion F. *Quantilfunktion* heißt die folgende Funktion: $Q:(0,1) \to R$ mit $p \mapsto Q(p) := \min\{x \mid F(x) \ge p\}$

Bem.: $x_p = Q(p)$ heißt p-Quantil und ist die kleinste Zahl, wo die Verteilungsfunktion das "Niveau" p erreicht oder überspringt. Nur für stetige Zufallsvariable gilt zusätzlich: das p-Quantil teilt die Fläche unter der Dichte links und rechts davon im Verhältnis p:(1-p). Die Quantilfunktion Q ist die Umkehrfunktion F^{-1} bis auf die konstanten Abschnitte von F.

Def.:Folgende Namen bezeichnenp0.50.250.750.1..0.9spezielle Quantile:Median1. Quartil3. QuartilDezile



2.1 Grundbegriffe

- Zufallsvariable
- **Verteilungsfunktionen**
- **Quantilfunktionen**
- Diskrete Zufallsvariable
- Stetige Zufallsvariable
- Affin-lineare Transformation von Zufallsvariablen

Def.: Eine Zufallsvariable X heißt *diskret*, wenn es entweder

- endlich viele $x_1, x_2, ..., x_J$ oder
- abzählbar viele $x_1, x_2,...$

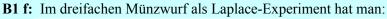
gibt, mit i)
$$p_j := P(X = x_j) > 0$$
 und ii) $\sum_j p_j = 1$

Träger von X: $T := \{x_i \mid P(X = x_i) > 0\}$

Bem.: Wahrscheinlichkeit ist auf den Träger konzentriert, der Rest hat Wahrscheinlichkeit 0.

Def.: X diskrete Zufallsvariable mit Träger T.

Wahrscheinlichkeitsfunktion (diskrete Dichte) von X ist die Funktion f mit: $f(x_j) := \begin{cases} p_j & x_j \in T \\ 0 & \text{sonst} \end{cases}$



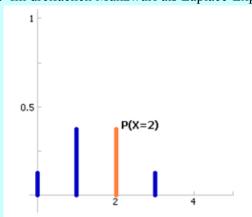


Abb.5 a: Wahrscheinlichkeitsfunktion

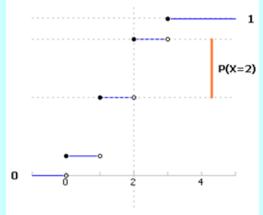


Abb.5 b: Verteilungsfunktion

Bem.: Die Wahrscheinlichkeitsfunktion zeichnet man analog zu Stabdiagrammen der deskriptiven Statistik.

Satz: Jede Zuordnung $x_j \mapsto p_j$ auf einer endlichen oder abzählbar unendlichen Menge $T = \{x_1, x_2, ...\}$

mit i)
$$p_j > 0$$
 und ii) $\sum_j p_j = 1$

definiert eindeutig eine Verteilungsfunktion einer Zufallsvariablen durch

$$F(x) := \sum_{j \mid x_j \le x} p_j = 1$$

Bem.: Zeichnen der Verteilung einer diskreten Zufallsvariablen X

$$p_j := P(X = x_j)$$

Träger

Diskrete Dichte

Stabhöhe

wo Stäbe gezeichnet werden

Verteilungsfunktion

Sprunghöhe

wo Verteilungsfunktion springt



2.1 Grundbegriffe





- Zufallsvariable
- Verteilungsfunktionen
- **Quantilfunktionen**
- **Diskrete Zufallsvariable**
- Stetige Zufallsvariable
- **Affin-lineare Transformation von Zufallsvariablen**

Def.: Eine Zufallsvariable heißt *stetig*, wenn ihre Verteilungsfunktion F eine Integraldarstellung hat, d. h. es gibt eine Funktion f mit:

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

Bem.: Die Verteilungsfunktion einer stetigen Zufallsvariablen ist stetig und bis auf "wenige" Ausnahmen differenzierbar, es gilt:

$$F'(x) = f(x)$$

Def.: X stetige Zufallsvariable mit Verteilungsfunktion F.

Die Funktion f heißt *Dichtefunktion* von X, wenn gilt: $F(x) = \int_{-\infty}^{x} f(t) dt$

Satz (Eigenschaften von Dichtefunktionen):

Ist f Dichtefunktion einer stetigen Zufallsvariablen, so gilt:

1.
$$f(x) \ge 0$$
 nicht-negativ

2.
$$\int_{-\infty}^{\infty} f(t) dt = 1$$
 normiert

Bem.: Für stetige Zufallsvariable haben Punkte keine Wahrscheinlichkeit: P(X = x) = 0. Eine Dichtefunktion kann man an einzelnen Stellen ändern, ohne dass sich dadurch die Wahrscheinlichkeiten ändern.

Satz: Formel integrierbar mit

1.
$$f(x) \ge 0$$

$$2. \int_{-\infty}^{\infty} f(t) dt = 1$$

legt eindeutig eine Verteilungsfunktion F einer stetigen Zufallsvariablen fest durch

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

Def.: $Tr\ddot{a}ger\ von\ X$: kleinste abgeschlossene Menge T mit $P(X \in T) = 1$

Bem.: Meist endliche, abzählbar unendliche Träger, Intervalle oder ganz R.

Satz: Die Quantilfunktion Q einer stetigen Zufallsvariablen nimmt folgende Gestalt an:

 $Q(p) = \min\{x \mid F(x) = p\} \text{ mit } 0$

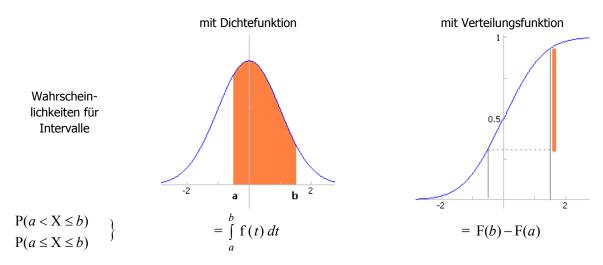


Abb. 6: Die Wahrscheinlichkeiten von Intervallen (a,b] bestimmt man durch Integration

Bem.: Die Unterscheidung von \leq und \leq kann man bei stetigen Verteilungen weglassen.

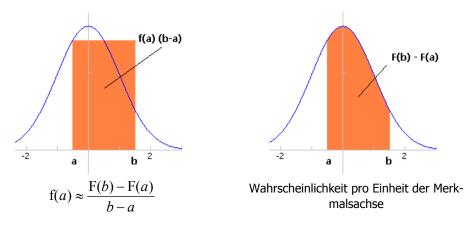


Abb. 7: Dichtefunktion als "Bevölkerungsdichte"

Also: Dichte ist ≈

B6: Drehen eines Glücksrads mit der Zufallsvariablen X als Winkel des Zeigers, wo dieser stehen bleibt. Mit der Begründung, dass eine Symmetrie dieselbe "Neigung" des Zeigers verursacht, eine Endposition zu erreichen, nimmt man folgende Dichtefunktion f als Modell – die so genannte *stetige Gleichverteilung* auf dem Intervall [0, 360]:

$$f(x) = \begin{cases} \frac{1}{360} & 0 < x \le 360 \\ 0 & \text{sonst} \end{cases} \qquad F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{360} \cdot x & 0 < x \le 360 \\ 1 & x \ge 30 \end{cases}$$

Die Verteilungsfunktion $F(x)$ ergibt sich durch Integration. Als Quantilfunktion erhält man: $Q(p) = 360 \cdot p$	

B4 f: Das Modell der *Exponentialverteilung* für eine Lebensdauer hat mit $\lambda = 1$ die Verteilungsfunktion:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \ge 0 \end{cases}$$

Die Dichtefunktion erhält man durch Differenzieren:

$$f(x) = F'(x) = -e^{-x}$$
 für $x \ge 0$.

Die Quantile ergeben sich durch Lösen der Gleichung $1 - e^{-x} = p$ zu: $x_p = \ln(\frac{1}{1-p})$



2.1 Grundbegriffe





- Zufallsvariable
- Verteilungsfunktionen
- **Quantilfunktionen**
- **Diskrete Zufallsvariable**
- **Stetige Zufallsvariable**
- Affin-lineare Transformation von Zufallsvariablen

Wie wirkt sich eine Mittelpunktsverschiebung bzw. eine Skalenänderung auf die Verteilung einer Zufallsvariablen aus?

Satz: X mit Verteilungsfunktion F, Wahrscheinlichkeitsfunktion f und Quantilfunktion x_p $Y := a + b \cdot X$ entsprechend mit G, g, und y_p . Dann gilt:

1.
$$G(y) = F(\frac{y-a}{b})$$

$$2. \quad y_p = a + b \cdot x_p$$

3.
$$g(y) = \frac{1}{b} \cdot f(\frac{y-a}{b})$$

Bew.: 1.
$$G(y) = P(Y \le y) = P(a + b \cdot X \le y) = P(X \le \frac{y - a}{b}) = F(\frac{y - a}{b})$$

1. wird benutzt, um die Gleichung für 2. auf die Variable Y zu übertragen und um für 3. die Dichte von Y durch Ableiten zu bestimmen.



2 Zufallsvariable und ihre Verteilungen

Inhaltsverzeichnis

2.1 Grundbegriffe

2.2 Verteilungsparameter

- 2.3 Spezielle diskrete Verteilungen
- 2.4 Spezielle stetige Verteilungen
- 2.5 Mathematische Ergänzungen

2.2 Verteilungsparameter

- Erwartungswert
- Varianz
- **Ungleichung von Tschebyscheff**
- Schiefe und Wölbung einer Verteilung

Es werden Größen zur Beschreibung der Lage und der Gestalt einer Verteilung definiert.

2.2 Verteilungsparameter



- Erwartungswert
- Varianz
- **Ungleichung von Tschebyscheff**
- Schiefe und Wölbung einer Verteilung

Der Erwartungswert ist eine Kennziffer der Verteilung, die ihre "Lage" angibt. Dynamisch wird er als Mittelwert einer längeren Serie von Versuchsdurchführungen zu interpretieren sein. Die Zufallsvariable als Glücksspiel mit Auszahlung X interpretiert lässt demnach "erwarten", dass bei einer längeren Serie von Spielen der mittlere Auszahlungsbetrag pro Spiel in etwa dem Erwartungswert entspricht. Genauer wird das beim Schätzen und Testen von Parametern noch zu klären sein.

Def.: Für die Zufallsvariable X wird der *Erwartungswert* wie folgt festgelegt, falls die Ausdrücke existieren:

$$\mu := E(X) := \begin{cases} \sum_{j} x_{j} \cdot p_{j} & X \text{ diskret} \\ \int_{\mathbf{R}} x f(x) \cdot dx & X \text{ stetig} \end{cases}$$

B2: Setzt man beim Roulette auf eine *Colonne* oder auf eine *einzelne Zahl*, so ergeben sich folgende Gewinne mit den entsprechenden Wahrscheinlichkeiten:

while thit den entsprechenden wantscheinhenkeit			
	Gewinn	Wahrscheinlichkeit	
X Colonne	2	<u>12</u> 37	
	-1	<u>25</u> 37	
	$E(X) = (-1) \cdot \frac{25}{37} + 2 \cdot \frac{12}{37} = -\frac{1}{37}$		
Y einzelne Zahl	35	$\frac{1}{37}$	
	-1	$\frac{36}{37}$	
	$E(Y) = (-1) \cdot \frac{36}{37} + 35 \cdot \frac{1}{37} = -\frac{1}{37}$		

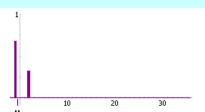
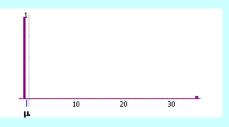


Abb. 8: Erwartungswerte zweier Glücksspiele



Bem.: Später wird eine Begründung für eine Deutung von Wahrscheinlichkeit P als relative Häufigkeit h gegeben: $P(A) \approx h(A)$. Je öfter man einen Zufallvorgang durchführt, desto "genauer" gilt diese Approximation. Für E(X) hätte man dann bei 3700 Spielen "zu erwarten", dass man 2500 Mal verliert und 1200 Mal gewinnt (wenn die Buchhaltung so genau eintritt, hätte man allerdings Anlass zu Bedenken wegen Manipulation). Für die Gewinnsumme hätte man daher im Voraus den Wert

$$(-1) \cdot 2500 + 2 \cdot 1200 = -100$$

anzusetzen, was im Durchschnitt, d. h. pro Spiel, einen Verlust von

$$\frac{-100}{3700} = \frac{1}{37} = 2,7\%$$

bedeutet. Der Erwartungswert gibt hier eine Vorausdeutung, wie hoch der durchschnittliche Gewinn sein wird. Abweichungen davon sind aber mehr als "normal".

B6:

Für die Zufallsvariable mit dem Glücksrad und der *Gleichverteilung* als Dichtefunktion ergibt sich für den Erwartungswert:

$$E(X) = \int_{-\infty}^{\infty} x f(x) \cdot dx = \int_{0}^{360} x \frac{1}{360} \cdot dx = \dots = 180$$

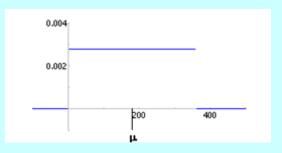


Abb. 9: Dichte einer stetigen Gleichverteilung

Def.: X eine Zufallsvariable (stetig oder diskret) und Y := g(X) eine reellwertige Funktion. Der Erwartungswert der Funktion g ist festgelegt durch

$$E(g(X)) := \begin{cases} \sum_{j} g(x_{j}) \cdot p_{j} & X \text{ diskret} \\ \int_{g} g(x) f(x) \cdot dx & X \text{ stetig} \end{cases}$$

Bem.: $g(X) = X^2$ oder $g(X) = a + b \cdot X$ etc. Die angegebene Prozedur kann man als Bestimmung des "mittleren" Funktionswertes deuten. Man wird später aus der Wahrscheinlichkeitsverteilung von X auch die von Y = g(X) berechnen können. Damit kommt man zum selben Erwartungswert von Y wie oben festgesetzt.

B2 f: Im Roulette hat man für X (*Colonne*) und Y (*Zahl*) mit der Funktion Formel $g(u) = u^2$ folgende Werte:

$$E(X^{2}) = (-1)^{2} \cdot \frac{25}{37} + 2^{2} \cdot \frac{12}{37} = \frac{73}{37} = 1,97$$

$$E(Y^{2}) = (-1)^{2} \cdot \frac{36}{37} + 35^{2} \cdot \frac{1}{37} = \frac{1261}{37} = 34,08$$

B6 f: Für die quadratische Funktion hat man beim Glücksrad folgenden Erwartungswert:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) \cdot dx = \int_{0}^{360} x^2 \frac{1}{360} \cdot dx = ... = 43200$$

Satz (Linearität des Erwartungswerts): $E(a+b\cdot X) = a+b\cdot E(X)$, falls E(X) existiert.

Bem.: Wird E als "Indikator" für den mittleren Gewinn (Gewinnerwartung) verwendet, so kann die Umrechnung auf andere Währungen die Gewinnerwartung nicht beeinflussen: die Gewinnerwartung wird nach demselben Kurs umgerechnet wie die Einzelwerte.

Bew. nur für stetige Zufallsvariable: zuerst wegen der Linearität des Erwartungswerts; dann ergibt das erstere der beiden letzten Integrale 1:

$$E(a+b\cdot X) = \int (a+b\cdot x) f(x) \cdot dx = a \cdot \int f(x) \cdot dx + b \cdot \int x f(x) \cdot dx = a+b \cdot E(X).$$

Def.: X Zufallsvariable mit E(X); der Erwartungswert existiert. Die Funktion Formel heißt *zentrierte Zufallsvariable*.

Bem.: Die zentrierte Zufallsvariable hat Erwartungswert 0.

Bem.: Eine Verteilung heißt symmetrisch, wenn Wahrscheinlichkeitsfunktion bzw. Dichtefunktion symmetrisch bezüglich eines Punktes c sind. Für solche Verteilungen gilt: E = c.

Bem.: Interpretiert man die Wahrscheinlichkeitsdichte als Massendichte im Sinne der Physik, so gibt der Erwartungswert die Lage des Schwerpunkts der Masse an.



2.2 Verteilungsparameter





Erwartungswert

Varianz

Ungleichung von Tschebyscheff

Schiefe und Wölbung einer Verteilung

Es geht um ein Maß für die Streuung um den Erwartungswert bzw. die Breite einer Verteilung.

Def.: X eine Zufallsvariable mit $\mu := E(X)$

Varianz von X: $V(X) := E(X - \mu)^2$, falls der Ausdruck existiert

Standardabweichung von X: $\sigma := \sigma(X) := \sqrt{V(X)}$

Bem.: Die Varianz ist der Erwartungswert der quadratischen Abweichung vom Erwartungswert, das ist von der Funktion $Y := g(X) := (X - \mu)^2$. Bei stetigen bzw. diskreten Zufallsvariablen wird wieder entsprechend ein Integral bzw. eine Summe ausgewertet.

Die Bedeutung der Standardabweichung als Maß für die Breite wird bei der Tschebyscheffschen Ungleichung klarer werden; jetzt sei nur festgehalten, dass ihre Dimension (im physikalischen Sinn) zu den Einzelwerten von X passt.

Satz (Verschiebungssatz): Mit der Abkürzung $\mu := E(X)$ gilt:

$$V(X) = E(X^2) - \mu^2$$

Bew.: Durch Ausquadrieren und Anwenden der Linearität des Erwartungswerts erhält man:

$$V(X) = E(X - \mu)^2 = E(X^2 - 2\mu \cdot X + \mu^2) = E(X^2) - 2\mu \cdot E(X) + E(\mu^2) = E(X^2) - \mu^2$$

B2 f: Gewinn bei Setzen auf *Colonne* bzw. eine *einzelne* Zahl beim Roulette: für die Varianzen hat man:

$$V(X) = E(X^2) - \mu_X^2 = \frac{73}{37} - (-\frac{1}{37})^2 = 1,97$$
 bzw. $\sigma_X = 1,40$

$$V(Y) = E(Y^2) - \mu_Y^2 = \frac{1261}{37} - (-\frac{1}{37})^2 = 34,08$$
 bzw. $\sigma_Y = 5,84$

Bem: Offenbar ist im Roulette die viel breitere Gewinnstreuung beim Setzen auf einzelne Zahlen im Vergleich zur Colonne durch den viel größeren Wert der Standardabweichung (5,84 gegen 1,40) abgebildet. Man hat einen viel höheren Gewinn in Aussicht, der aber wesentlich weniger wahrscheinlich ist als beim Setzen auf Colonne. Das Spiel ist "riskanter".

Satz (Auswirkung von Skalenänderungen)

Falls die Varianz existiert, gilt:

$$V(a+b\cdot X) = b^2 \cdot V(X)$$

$$\sigma(a+b\cdot X) = b\cdot \sigma(X)$$

Bem.: Wird der Gewinn in verschiedenen Währungen angegeben, so erfolgt die Umrechnung für die Standardabweichung nach demselben "Wechselkurs" wie für die Einzelwerte.



2.2 Verteilungsparameter





- Erwartungswert
- Varianz
- Ungleichung von Tschebyscheff
- Schiefe und Wölbung einer Verteilung

Obere Schranke für die Wahrscheinlichkeit von "extremen" Werten.

Satz (Tschebyscheffsche Ungleichung):

X Zufallsvariable mit $\sigma^2 := V(X) < \infty$, $\mu := E(X)$, $\varepsilon > 0$. Dann gilt:

$$P(|X-\mu| \ge \varepsilon) \le \frac{\sigma^2}{\varepsilon^2}$$

Bem.: Zentrale Werte von X sind Werte innerhalb eines " ε -Fensters" um μ : { $\mu - \varepsilon < X < \mu + \varepsilon$ }. Extreme Werte unterscheiden sich um mindestens ε von μ : { $|X - \mu| \ge \varepsilon$ }.

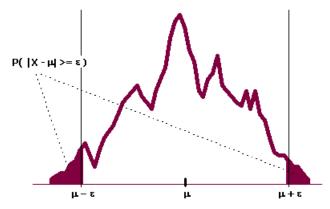
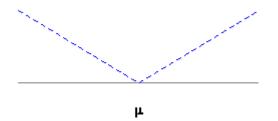


Abb. 10: Abschätzung für die extremen Werte einer Verteilung nach Tschebyscheff

Die angegebene Ungleichung besagt, dass die Schranke für die Wahrscheinlichkeit größer wird, wenn ε kleiner bzw. wenn die Varianz größer wird. Man beachte hierbei, dass die Abschätzung für alle (nur denkbar möglichen) Verteilungen gilt; ihr Vorzug liegt nicht in einer praktisch gut verwertbaren Schranke sondern eindeutig in späteren theoretischen Folgerungen daraus (siehe Gesetz der Großen Zahlen).

Bew.: Vergleicht man die folgenden beiden Funktionen,



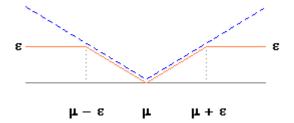


Abb.11 a: Absolute Abweichung:

$$g(\omega) := |X(\omega) - \mu|$$

Abb.11 b: Gestutzte absolute Abweichung:

$$Y(\omega) := \begin{cases} \varepsilon & |X(\omega) - \mu| \ge \varepsilon \\ 0 & \text{sonst} \end{cases}$$

so sieht man sofort, dass Y nicht größer als g ist; wir schließen weiters:

$$0 < Y(\omega) \le g(\omega) = |X(\omega) - \mu|$$

$$Y^2 \le (X - \mu)^2$$

$$E(Y^2) \le E(X - \mu)^2$$

$$\varepsilon^2 \cdot P(|X - \mu| \ge \varepsilon) \le \sigma^2$$

Jetzt braucht man nur noch durch ε^2 dividieren und erhält die behauptete Abschätzung für die Wahrscheinlichkeit " ε -extremer" Werte.

B11: X Verweildauer eines zufällig gewählten Patienten einer Klinik. μ = 10 und σ = 4 seien bekannt. Gesucht ist eine Abschätzung für die Wahrscheinlichkeit, dass der Patient zwischen 5 und 15 Tagen (ausschließlich der Grenzen) im Krankenhaus verbringt.

Formel mit Berechnung

Demnach verbringen mindestens 36% der Patienten zwischen 5 und 15 Tagen in der Klinik.

Bem.: Mit $\varepsilon = \lambda \cdot \sigma$ als Vielfaches der Standardabweichung ausgedrückt, liest man die Tschebyscheffsche Ungleichung so:

$$P(|X-\mu| \ge \lambda \cdot \sigma) \le \frac{1}{\lambda^2}$$

Das ergibt jetzt eine schöne Deutung für die Standardabweichung als Maß für die Breite einer Verteilung:

λ	innerhalb	P mindestens	außerhalb	P höchstens
2	$(\mu-2\cdot\sigma,\mu+2\cdot\sigma)$	$\frac{3}{4} = 0.75$	$ X-\mu \ge 2 \cdot \sigma$)	$\frac{1}{4} = 0.25$
3	$(\mu - 3 \cdot \sigma, \mu + 3 \cdot \sigma)$	$\frac{8}{9} = 0.89$	$ X-\mu \ge 3 \cdot \sigma$)	$\frac{1}{9} = 0.11$

2.2 Verteilungsparameter





Erwartungswert

Varianz

Ungleichung von Tschebyscheff

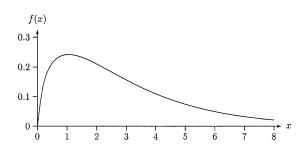
Schiefe und Wölbung einer Verteilung

Weitere Maßzahlen als Parameter, welche die Form einer Verteilung beschreiben.

Def.: *Schiefe* einer Zufallsvariablen X:

$$\gamma_1(X) := E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)$$
, falls der Erwartungswert $\mu := E(X)$ existiert.

Bem.: $\gamma_1(X) = 0$, wenn die Verteilung der Zufallsvariablen zu einem Punkt c symmetrisch ist



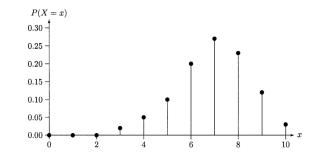


Abb. 12 a: $\gamma_1(X) > 0$

rechtsschief (skewed to the right) bzw. linkssteil

Abb. 12 b: $\gamma_1(X) < 0$ linksschief (skewed to the left) rechtssteil

Def.: Wölbung oder Kurtosis einer Zufallsvariablen X:

$$\gamma_2(X) := E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right)$$
, falls der Erwartungswert $\mu := E(X)$ existiert.

Bem.: Die *Normalverteilung* (siehe später) ist der Prototyp einer symmetrischen, eingipfeligen Verteilung. Sie hat Wölbung 3. Hat eine Verteilung dieses Typs eine Wölbung > 3, so ist sie spitzer im Zentrum und hat mehr Wahrscheinlichkeit in den Ausläufern. (Die Größe $\gamma_2 - 3$ wird auch als Exzess bezeichnet.) Das hat entsprechend Folgen für die Beurteilung der Wahrscheinlichkeit des Auftretens von "extremen" Werten.



2 Zufallsvariable und ihre Verteilungen

- **Inhaltsverzeichnis**
- 2.1 Grundbegriffe
- 2.2 Verteilungsparameter
- 2.3 Spezielle diskrete Verteilungen
- 2.4 Spezielle stetige Verteilungen
- 2.5 Mathematische Ergänzungen

2.3 Spezielle diskrete Verteilungen

- **Zentrale Experimente**
- **Binomialverteilung**
- **Poisson-Verteilung**
- **Geometrische Verteilung**
- Hypergeometrische Verteilung

Es werden wichtige Grundtypen von diskreten Verteilungen und ihre Anwendungsmöglichkeiten dargestellt. Alle Verteilungen werden aus dem Urnenmodell "abgeleitet".

2.3 Spezielle diskrete Verteilungen





- Zentrale Experimente
- Binomialverteilung
- **Poisson-Verteilung**
- Geometrische Verteilung
- Hypergeometrische Verteilung

Im Bernoulli-Experiment interessiert man sich nur für "Erfolg' oder "Misserfolg' eines Versuches. Eine Bernoulli-Versuchsreihe ist eine unabhängige Wiederholung davon, d. h., es muss für die $A_i \equiv$ "Erfolg im i-ten Versuch" gelten: i) $A_1, A_2, ..., A_n$ sind vollständig unabhängig und ii) $P(A_i) = \pi$ für alle i.

Def.: In einer Bernoulli-Versuchsreihe zeigt die Zufallsvariable

$$I_{A_i} = \begin{cases} 1 & \text{falls } A_i \text{ eintritt} \\ 0 & \text{falls } A_i \text{ ausbleibt} \end{cases}$$

an, ob im *i*-ten Versuch Erfolg oder Misserfolg zutrifft. $X := \sum_{i=1}^{n} I_{A_i}$ heißt dann *Anzahl der Erfolge*.

B13: Ein fairer Würfel wird n-mal geworfen. $A_i \equiv$ "beim i-ten Wurf ein Sechser". Hat man auf Ω ein Laplace-Modell (mit Gleichwahrscheinlichkeit für alle Elementarereignisse), so sind die A_i vollständig unabhängig und wir haben es mit einer Bernoulli-Versuchsreihe zu tun.

Bem.: Ziehen aus einer Urne mit M roten und N-M weißen Kugeln stellt den Prototyp eines Bernoulli-Experiments dar. In der bildlichen Vorstellung rührt man die Kugeln ordentlich um und zieht (ohne Manipulation) zufällig eine – wie beim Lotto. Bei mehrfacher Ziehung beschreibe $A_i \equiv$ "bei der i-ten Ziehung eine rote Kugel". n Ziehungen aus der Urne mit Zurücklegen liefern – bei sorgfältiger Mischung jeweils – vollständig unabhängige A_i und somit eine Bernoulli-Versuchsreihe. Bei Ziehungen ohne Zurücklegen sind die A_i nicht mehr unabhängig, z. B. ist:

$$P(A_2 | A_1) = \frac{M-1}{N-1} \neq \frac{M}{N};$$

d. h., wir erhalten keine Bernoulli-Versuchsreihe.

2.3 Spezielle diskrete Verteilungen





- **Zentrale Experimente**
- Binomialverteilung
- Poisson-Verteilung
- Geometrische Verteilung
- Hypergeometrische Verteilung

Def.: Ist $X := \sum_{i=1}^{n} I_{A_i}$ die Anzahl der Erfolge einer Bernoulli-Versuchsreihe der Länge n mit Erfolgswahrscheinlichkeit π , so heißt die Verteilung von X *Binomialverteilung*; wir schreiben $X \sim B(n, \pi)$

B14 : Man bestimme in einer Bernoulli-Versuchsreihe mit n = 4 alle Ergebnisse, bei denen genau 2 Erfolge eintreten, d. h. (X = 2).

$$A_1, A_2, \overline{A_3}, \overline{A_4}$$
 $A_1, \overline{A_2}, A_3, \overline{A_4}$ $A_1, \overline{A_2}, \overline{A_3}, A_4$ $\overline{A_1}, A_2, \overline{A_3}, A_4$ $\overline{A_1}, \overline{A_2}, \overline{A_3}, A_4$ $\overline{A_1}, \overline{A_2}, \overline{A_3}, A_4$

das sind 6 verschiedene Folgen. Kombinatorisch gezählt sind das $\binom{4}{2} = 6$, denn es sind 2 Positionen aus 4 auszuwählen, die Reihenfolge spielt keine Rolle, mehrfache Auswahl ist nicht gestattet, d. h., es handelt sich um Kombinationen ohne Zurücklegen.

Satz: Für eine Binomialverteilung $X \sim B(n, \pi)$ gilt:

$$P(X = k) = \binom{n}{k} \cdot \pi^k \cdot (1 - \pi)^{n-k}$$
 für $k = 0, 1, ..., n$ und $0 < \pi < 1$

Bew.: $T_X = \{0, 1, ..., n\}$; es können bei n Versuchen nur 0 bis maximal n Erfolge auftreten. Eine spezielle Folge der I_{A_i} mit genau k Erfolgen ist dann $\begin{pmatrix} 1, 1, 1, ..., 1, & 0, 0, ..., 0 \\ k - mal & n - k \end{pmatrix}$ und hat die Wahrscheinlichkeit $\pi^k \cdot (1 - \pi)^{n-k}$. Man kann die k Erfolge dann auf $\binom{n}{k}$ Arten auf die n Positionen aufteilen.

Bem.: $X \sim B(n, \pi)$ mit n = 1 heißt auch *Bernoulli-Verteilung*.

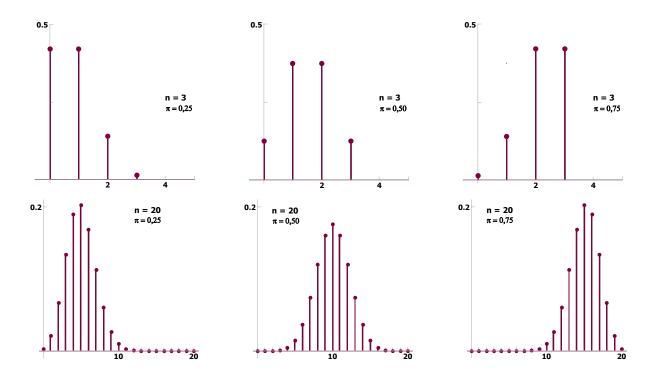


Abb. 13: Wahrscheinlichkeitsfunktion für einige Binomialverteilungen

B: Wie groß ist die Wahrscheinlichkeit, bei 6mal Würfeln genau einen Sechser, keinen Sechser, bzw. mehr als einen Sechser zu erhalten.

$$P(X = 1) = 0.4019$$
 $P(X = 0) = 0.3349$

$$P(X \ge 2) = P(X > 1) = 1 - P(X \le 1) = 1 - 0.7368 = 0.2632$$

Es ist doch recht wahrscheinlich (ca. 1/3), dass man beim Würfeln bei 6 Würfen keinen Sechser dabei hat. Die tatsächlichen Werte berechnet man sich mit einem Taschenrechner oder mit EXCEL oder aus Tabellen mit der Verteilungsfunktion der Binomialverteilung.

Satz (Erwartungswert und Varianz der Binomialverteilung):

a)
$$X \sim B(1, \pi)$$

$$E(X) = \pi$$

$$V(X) = \pi \cdot (1 - \pi)$$

b)
$$X \sim B(n, \pi)$$

$$E(X) = n \cdot \pi$$

$$V(X) = n \cdot \pi \cdot (1 - \pi)$$

Bew.: b) geht mit Hilfsmitteln von Kapitel 3 leicht.

a)
$$E(X) = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi$$

$$E(X^2) = 0^2 \cdot (1 - \pi) + 1^2 \cdot \pi = \pi$$
 $V(X) = E(X^2) - (E(X))^2 = \pi - \pi^2$

Bem.: Für $X \sim B(n, \pi)$ ist $Y := \frac{X}{n}$ die "mittlere Zahl" von Erfolgen bei n Versuchsdurchführungen und hat einen Erwartungswert von $E(Y) = \frac{1}{n} \cdot E(X) = \pi$, d. h., ihr Erwartungswert ist gleich der Erfolgswahrscheinlichkeit der Bernoulli-Versuchsreihe.

Bem.: Man rechnet die Wahrscheinlichkeiten aus Tabellen der Verteilungsfunktion gemäß

$$P(X \le k) = F(k)$$
 $P(a \le X \le b) = F(b) - F(a-1)$ oder besser mit Taschenrechner oder EXCEL.



2.3 Spezielle diskrete Verteilungen

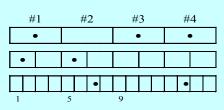




- Zentrale Experimente
- Binomialverteilung
- Poisson-Verteilung
- Geometrische Verteilung
- Hypergeometrische Verteilung

Jeder von n Semmeln wird mit Wahrscheinlichkeit π eine Rosine unabhängig voneinander beigemengt. **B2**: A_i zählt, ob es in der *i*-ten Semmel eine Rosine gibt ("Erfolg"). Die Anzahl aller verteilten Rosinen ist binomial verteilt: $X := \sum_{i=1}^{n} I_{A_i} \sim B(n, \pi)$. Halbiert man die Semmeln und gibt jeder "Halbsemmel" mit Wahrscheinichkeit $\frac{\pi}{2}$ eine Rosine dazu, so ist die Anzahl der verbrauchten Rosinen wieder binomial verteilt, nur die Parameter ändern sich: $X_2 := \sum_{i=1}^{2n} I_{B_i} \sim B(2n, \frac{\pi}{2})$.

Man beachte: der Erwartungswert ist in beiden Fällen $\mu = n \cdot \pi = (2n) \cdot (\frac{\pi}{2})$. Wie sieht die Verteilung von X_m (die verbrauchten Rosinen) aus, wenn die Teilung immer feiner wird?



Satz (Grenzwertsatz von Poisson):

$$X_n \sim B(n, \pi_n)$$
 $\lim_{n \to \infty} n \cdot \pi_n = \mu$ dann gilt: $\lim_{n \to \infty} P(X_n = k) = e^{-\mu} \frac{\mu^k}{k!}$, $k = 0, 1, ...$

Bem.: Man kann für großes n und kleines π_n (Faustregel $\pi \le 0,1$ und $n \ge 50$) die Wahrscheinlichkeiten der Binomialverteilung durch die Formel im Satz approximieren.

Def.: X mit Träger $T_X = \{0, 1, ...\}$ hat eine *Poisson-Verteilung* mit Parameter $\mu > 0$, wenn gilt:

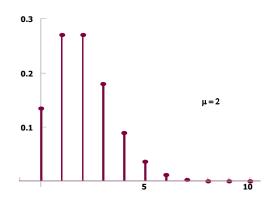
$$P(X = k) = e^{-\mu} \frac{\mu^k}{k!}, \quad k = 0, 1, ...; \text{ wir schreiben: } X \sim Po(\mu).$$

Bem.: Mit
$$p_k := P(X = k)$$
 gilt: i) $p_k > 0$ und ii) $\sum_k p_k = 1$,

die Zuordnung ist also eine diskrete Wahrscheinlichkeitsfunktion. Die Werte berechnet man sich aus entsprechenden Tabellen der Verteilungsfunktion bzw. mit EXCEL oder anderen Rechenhilfen. Die folgenden Stabdiagramme illustrieren, dass Poisson-Verteilungen bei kleinem μ rechtsschief sind, bei größeren Werten symmetrisiert sich die Verteilung zunehmend.

B17: Die Wahrscheinlichkeit, dass ein Säugling eine bestimmte Impfung nicht verträgt, betrage 0,05. Unverträglichkeit sei zwischen einzelnen Säuglingen unabhängig. Man berechne exakt mit Binomialverteilung und mit Poisson-Approximation die Wahrscheinlichkeit, dass bei 50 Impfungen 0, 1, 2 bzw. 3 und mehr Reaktionen auftreten.

	B($n = 50, \pi = 0.05$)	Po($\mu = 2.5$)
X = 0	0,0769	0,0821
X = 1	0,2025	0,2052
X = 2	0,2611	0,2565
$X \ge 3$	0,4595	0,4562



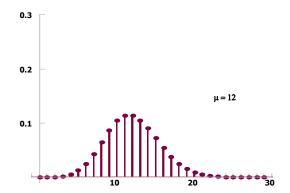


Abb. 14: Wahrscheinlichkeitsfunktion für einige Poisson-Verteilungen

Satz: Für Erwartungswert und Varianz einer Poisson-Verteilung $X \sim Po(\mu)$ gilt:

a)
$$E(X) = \mu$$
 b) $V(X) = \mu$

Bew.: Der Beweis sei nur für den Erwartungswert geführt, für die Varianz braucht man noch den Verschiebungssatz:

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



2.3 Spezielle diskrete Verteilungen



- **Zentrale Experimente**
- **Binomialverteilung**
- **Poisson-Verteilung**
- Geometrische Verteilung
- Hypergeometrische Verteilung

Def.: In einer Bernoulli-Versuchsreihe unbestimmter Länge mit Erfolgswahrscheinlichkeit π heißt die Verteilung der Zufallsvariablen $X := \min \left\{ n \left| \sum_{i=1}^n \mathbf{I}_{A_i} \right| = 1 \right\} \right\}$ geometrische Verteilung, man schreibt: $X \sim G(\pi)$

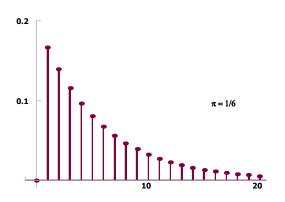
Bem.: Man nennt diese Zufallsvariable auch *Wartezeit auf den ersten Erfolg*; sie bezeichnet die Nummer desjenigen Versuchs, bei dem zum ersten Mal das Ereignis A_i eintritt.

Satz: Ist X geometrisch verteilt, $X \sim G(\pi)$, so gilt $P(X = k) = \pi \cdot (1 - \pi)^{k-1}$, k = 1, 2, ...

Bem.: Man rechnet mit der geometrischen Reihe leicht nach, dass es sich hierbei um eine diskrete Wahrscheinlichkeitsfunktion handelt. Die Werte für konkrete Intervalle berechnet man mit der Verteilungsfunktion. Diese ist etwa in EXCEL nicht extra angeführt, weil sie leicht zu bestimmen ist:

$$F(k) = P(X \le k) = \pi \cdot \sum_{i=1}^{k} (1 - \pi)^{i-1} = 1 - (1 - \pi)^{k}$$

Stabdiagramme illustrieren die Schiefe der Geometrischen Verteilungen.



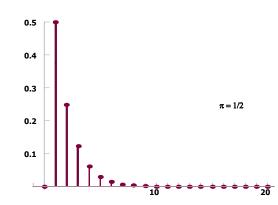


Abb. 15: Wahrscheinlichkeitsfunktion für zwei geometrische Verteilungen

B18: Aus einer Urne mit mit 4 roten und 6 weißen Kugeln wird zufällig *mit* Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass man

- a) bei der dritten Ziehung die erste rote Kugel erwischt,
- b) die erste rote Kugel frühestens bei der dritten Ziehung erwischt?

Die Anzahl X der benötigten Ziehungen bis zur ersten roten Kugel ist geometrisch verteilt: $X \sim G(\pi = 1/6)$. Aus EXCEL wurden die folgenden Werte berechnet (gesuchte Werte sind fett):

k	P(X=k)	$P(X \le k)$	$P(X \ge k)$
1	0,1667	0,1667	0,8333
2	0,1389	0,3056	0,6944
3	0,1157	0,4213	0,5787

Satz: Für Erwartungswert und Varianz einer Geometrischen Verteilung $X \sim G(\pi)$ gilt:

a)
$$E(X) = \frac{1}{\pi}$$

b)
$$V(X) = \frac{1-\pi}{\pi^2}$$

Bew.: Der Beweis erfolgt über die Auswertung der entsprechenden Reihen; bei der Varianz nutzt man wieder mit Vorteil den Verschiebungssatz.

B19: Für die Wartezeit X auf den ersten Sechser beim fortgesetztem Würfeln gilt: E(X) = 6. Man modelliert entsprechend der Bernoulli-Versuchsreihe mit einer Geometrischen Verteilung mit $\pi = \frac{1}{6}$.



2.3 Spezielle diskrete Verteilungen



- **Zentrale Experimente**
- Binomialverteilung
- **Poisson-Verteilung**
- **Geometrische Verteilung**
- Hypergeometrische Verteilung

Zieht man aus einer Urne mit roten und weißen Kugeln *mehrfach* hintereinander *ohne* Zurücklegen, so handelt es sich nicht mehr um eine Bernoulli-Versuchsreihe. Man kann aber dennoch die Verteilung X der Anzahl der Ereignisse bestimmen.

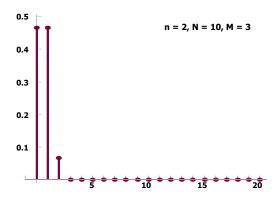
Def.: $X := \sum_{i=1}^{n} I_{A_i}$ ist die Anzahl der Erfolge beim *n*-maligen Ziehen aus einer Urne mit *M* roten und N-M weißen Kugeln; ferner ist $A_i =$ "bei der *i*-ten Ziehung eine rote Kugel". Man nennt dann X *Hypergeometrisch verteilt*: $X \sim H(n, N, M)$

Satz: Für eine Hypergeometrische Verteilung $X \sim H(n, N, M)$ gilt:

$$P(X = k) = \frac{\binom{M}{k} \cdot \binom{N - M}{n - k}}{\binom{N}{n}}$$
 für $k = 0, 1, ..., n$

Bem.: Wenn die angegebenen Binomialkoeffizienten $\binom{n}{k}$ nicht auswertbar sind, weil die Bedingung $0 \le k \le n$ verletzt ist, setze man diese gleich Null. Damit wird auch klar, dass der Träger i. A. eine echte Teilmenge von $\{0, 1, ..., n\}$ ist.

Bem.: Wegen der Nähe zur Binomialverteilung (Anzahl der Erfolge beim Ziehen *ohne* bzw. *mit* Zurücklegen) ist klar, dass die beiden Verteilungen ähnliche Werte liefern, wenn sich die Wahrscheinlichkeiten der Ereignisse A_i nur geringfügig ändern. Das ist der Fall, wenn $\frac{n}{N} \le 0.05$ (als Faustregel).



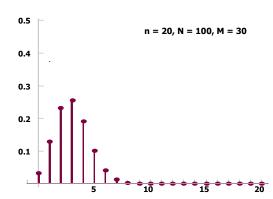


Abb. 16: Wahrscheinlichkeitsfunktion für zwei hypergeometrische Verteilungen

B20: Eine Warensendung enthält zehn elektronische Bauteile, von denen drei defekt sind. Aus der Sendung werden zwei Bauteile zufällig und ohne Zurücklegen entnommen und geprüft. Wie groß ist die Wahrscheinlichkeit, dass unter diesen mindestens ein defektes Bauteil ist? Die Anzahl X der defekten Bauteile ist hypergeometrisch verteilt: X ~ H(n = 2, N = 10, M = 3). Aus EXCEL wurden die folgenden Werte berechnet (gesuchter Wert ist fett):

k	P(X=k)	$P(X \le k)$	P(X > k)
0	0,4667	0,4667	0,5333
1	0,4667	0,9333	0,0667
2	0,0667	1,0000	0,0000
3	#ZAHL!		

Satz: Für Erwartungswert und Varianz einer Hypergeometrischen Verteilung $X \sim H(n, N, M)$ gilt:

a)
$$E(X) = n \cdot \frac{M}{N}$$
 b) $V(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \left(1 - \frac{n-1}{N-1}\right)$

Bem.: Wird zurückgelegt beim Ziehen, hat man eine Bernoulli-Versuchsreihe mit $\pi = \frac{M}{N}$. Im Vergleich dazu wird die Varianz kleiner, wenn nicht zurückgelegt wird. Der Faktor, um den die Varianz kleiner wird, ist ungefähr $1 - \frac{n}{N}$; dieser wird vernachlässigbar, wenn der *Auswahlsatz* $\frac{n}{N} << 1$ ("sehr klein") ist.



2 Zufallsvariable und ihre Verteilungen



- 2.1 Grundbegriffe
- 2.2 Verteilungsparameter
- 2.3 Spezielle diskrete Verteilungen
- 2.4 Spezielle stetige Verteilungen
- 2.5 Mathematische Ergänzungen

2.4 Spezielle stetige Verteilungen

- Stetige Gleichverteilung
- Exponentialverteilung
- **Pareto-Verteilung**
- Normalverteilung
- Lognormalverteilung

Die hier behandelten stetigen Verteilungen gehören zum Kernbestand der Wahrscheinlichkeitsrechnung und ihrer Anwendungen. Konnten bei den diskreten Verteilungen alle aus dem Urnenmodell entwickelt werden, so liegen stetigen Verteilungen die unterschiedlichsten Modellansätze zugrunde.

2.4 Spezielle stetige Verteilungen





Stetige Gleichverteilung

Exponentialverteilung

Pareto-Verteilung

Normalverteilung

Lognormalverteilung

Def.: Eine Zufallsvariable X hat eine *stetige Gleichverteilung* mit Parametern $\alpha, \beta \in \mathbb{R}$, wenn sie die Dichte

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \le x \le \beta \\ 0 & \text{sonst} \end{cases}$$
 besitzt; wir schreiben $X \sim S(\alpha, \beta)$.

Bem.: Die stetige Gleichverteilung ist dann ein sinnvolles Modell, wenn eine Zufallsvariable X keinen Wert aus dem Intervall $[\alpha, \beta]$ "bevorzugt"; Glücksräder haben i. A. annähernd diese Eigenschaft. Der Träger ist $T_X = [\alpha, \beta]$; es ist auch der Name Rechteckverteilung gebräuchlich.

B22: Wenn Straßenbahnen im Zehn- Minuten-Rhythmus verkehren und man rein zufällig zur Haltestelle geht, so ist die Wartezeit X (in Minuten) gut durch eine stetige Gleichverteilung zu modellieren, und zwar mit den Parametern 0 und 10.

Satz: Ist X stetig gleichverteilt mit α und β , so gilt für die Verteilungsfunktion:

$$F(x) = \begin{cases} 0 & x < \alpha \\ \frac{x - \alpha}{\beta - \alpha} & \alpha \le x \le \beta \\ 1 & x > \beta \end{cases}$$

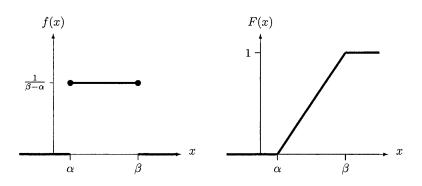


Abb. 17: Dichte und Verteilungsfunktion einer stetigen Gleichverteilung

Satz: Ist $Z \sim S(0,1)$, so ist die affin-linear Transformierte $X = \alpha + (\beta - \alpha) \cdot Z$ auch stetig gleichverteilt, und zwar mit den Parametern $\alpha_Y = \alpha$ und $\beta_Y = \beta$.

Bem.: $Z \sim S(0, 1)$ heißt auch *standard-gleichverteilt*.

Satz (Erwartungswert und Varianz einer stetigen Gleichverteilung):

$$X \sim S(\alpha, \beta)$$
, dann gilt: a) $E(X) = \frac{\alpha + \beta}{2}$ b) $V(X) = \frac{(\beta - \alpha)^2}{12}$

Bem: Der Erwartungswert ist wegen der Symmetrie der Verteilung gleich dem Symmetriezentrum $\frac{\alpha+\beta}{2}$ der Dichtefunktion.

Bew. (nur für die Varianz): Erst berechnet man für die Varianz einer standard-gleichverteilten Zufallsvariablen $Z \sim S(0, 1)$:

$$E(Z^2) = \int_{0}^{1} z^2 \cdot 1 \cdot dz = \frac{z^3}{3} \Big|_{0}^{1} = \frac{1}{3}$$

und dann mit dem Verschiebungssatz:

$$V(Z) = E(Z^2) - (E(Z))^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

Daraus bestimmt man leicht die Varianz der Transformierten $X := \alpha + (\beta - \alpha) \cdot Z$ als

$$V(X) = (\beta - \alpha)^{2} \cdot V(X) = \frac{(\beta - \alpha)^{2}}{12}$$

B23: Modell für die Dauer X eines bestimmten Arbeitsganges in der Produktion sei eine stetige Gleichverteilung mit den Parametern 7 und 12 [min].

- a) Berechnen Sie den Erwartungswert und die Standardabweichung der Dauer des Arbeitsganges X.
- b) Bestimmen Sie den 90%- Punkt dieser Verteilung und interpretieren Sie diesen.

a) Mit
$$\alpha = 7$$
, $\beta = 12$, gilt: $E(X) = \frac{\alpha + \beta}{2} = 9.5$ [min]; $\sigma(X) = \frac{(\beta - \alpha)}{\sqrt{12}} \approx 1.44$ [min].

b) Zu lösen ist die Gleichung F(x) = p, also $\frac{x - \alpha}{\beta - \alpha} = p$. Aus $\frac{x - \alpha}{\beta - \alpha} = 0.9$ ergibt sich nach einfachen Umformungen die Lösung $x = \alpha + (\beta - \alpha) \cdot 0.9 \approx 11.5$ [min]. Demnach ist der Arbeitsgang mit Wahrscheinlichkeit 0.9 nach 11.5 [min] beendet.

Satz (Zufallszahlen):

Sind F(x) eine beliebige Verteilungsfunktion, Q(p) die zugehörige Quantilfunktion und $Z \sim S(0, 1)$.

Die Zufallsvariable X ist definiert über die Quantilfunktion: X := Q(Z).

Dann hat X die Verteilungsfunktion F, d. h., $X \sim F$.

Bew.: Mit
$$Q(p) = \min\{x \mid F(x) \ge p\}$$
 gilt für jedes $x \in \mathbb{R}$
 $F_X(x) = P(X \le x) = P(Q(Z) \le x) = P(F(Q(Z)) \le F(x)) = P(Z \le F(x)) = F(x)$

Bem.: Wenn F eine stetige Funktion ist, gilt darüber hinaus $F = Q^{-1}$ und damit hat man:

$$F(X) = F(Q(Z)) = Q^{-1}(Q(Z)) = Z$$
, d. h., $F(X) \sim S(0, 1)$.

Jetzt deutet sich eine weitreichende Anwendung des Satzes zur Simulation beliebiger Zufallszahlen an: Man simuliert stetig gleichverteilte Zufallszahlen aus dem Intervall [0,1]: z_i . Dann berechnet man $x_i = F^{-1}(z_i)$ und erhält damit Zufallszahlen aus der Verteilung von F, sofern diese eine stetige Verteilungsfunktion ist.



2.4 Spezielle stetige Verteilungen





- Stetige Gleichverteilung
- Exponentialverteilung
- Pareto-Verteilung
- Normalverteilung
- Lognormalverteilung

Def.: Die Zufallsvariable X hat eine Exponentialverteilung mit Parameter $\lambda > 0$, falls ihre Dichte durch

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$
 gegeben ist; wir schreiben: $X \sim \text{Exp}(\lambda)$.

Bem.: Einige Dichten von Exponentialverteilungen in Abb. 18 zeigen den Einfluss des Parameters λ : Je größer dieser Wert, desto mehr Wahrscheinlichkeit befindet sich nahe bei 0 und die Dichtefunktion nimmt rasch mit wachsendem x ab.

Satz (Verteilungsfunktion der Exponentialverteilung):

$$X \sim \text{Exp}(\lambda)$$
 hat die Verteilungsfunktion: $F(x) = \begin{cases} 1 - e^{-\lambda \cdot x} & x \ge 0 \\ 0 & x < 0 \end{cases}$

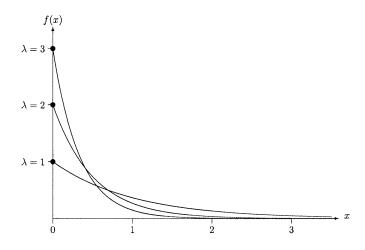
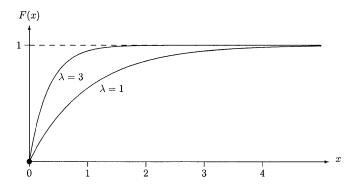


Abb. 18: Dichtefunktionen verschiedener Exponentialverteilungen



Satz (Quantile der Exponentialverteilung):

$$X \sim \text{Exp}(\lambda)$$
 hat die Quantilfunktion $x_p = \frac{1}{\lambda} \cdot \ln(\frac{1}{1-p})$

Bem.: Für den Median einer Exponentialverteilung ergibt sich: $x_{0,5} = \frac{1}{\lambda} \cdot \ln(2) = \frac{0,6931}{\lambda}$.

Satz (Erwartungswert, Varianz und Schiefe einer Exponentialverteilung):

$$X \sim \text{Exp}(\lambda)$$
, dann gilt: a) $E(X) = \frac{1}{\lambda}$ b) $V(X) = \frac{1}{\lambda^2}$ c) $\gamma_1(X) = 2$

Bem.: Je größer λ , desto kleiner Erwartungswert und Varianz. Die Schiefe hängt gar nicht von λ ab.

Bew.: Diese Parameter berechnet man mittels partieller Integration, wir tun dies nur für den Erwartungswert, um die Vorgangsweise zu illustrieren.

$$E(X) = \int_{0}^{\infty} x \cdot f(x) \cdot dx = \int_{0}^{\infty} x \cdot \left(\lambda e^{-\lambda x}\right) \cdot dx = x \cdot \left(-e^{-\lambda x}\right) \Big|_{0}^{\infty} - \int_{0}^{\infty} 1 \cdot \left(-e^{-\lambda x}\right) \cdot dx = \dots = \frac{1}{\lambda}$$

$$\int f \cdot g' \qquad \qquad f \cdot g \qquad \qquad \int f' \cdot g$$

Satz (Gedächtnislosigkeit der Exponentialverteilung):

$$X \sim \text{Exp}(\lambda)$$
, dann gilt: $P(X > t + s \mid X > t) = P(X > s)$

Bem.: Interpretiert man die Zufallsvariable als Lebensdauer, so ist

- P(X > s) die Wahrscheinlichkeit, den Zeitpunkt s zu überleben (von Beginn an);
- $P(X > t + s \mid X > t)$ die Wahrscheinlichkeit, mindestens s weitere Zeiteinheiten zu leben, wenn man schon t Einheiten "alt" ist (X > t).

Diese beiden sind für Exponentialverteilungen gleich. Das erreichte Alter spielt also keine Rolle für das weitere Überleben; es ist, als ob die Zeit von Neuem anfangen würde.

Bem.: Gilt für eine stetige Zufallsvariable die Gedächtnislosigkeit, so ist sie schon exponentialverteilt. Gilt für eine diskrete Zufallsvariable diese Eigenschaft, so ist sie geometrisch verteilt.

Bew.: Nach der Definition der bedingten Wahrscheinlichkeit und der Verteilungsfunktion einer Exponentialverteilung hat man:

$$P(X > t + s \mid X > t) = \frac{P(X > t + s)}{P(X > t)} = \frac{1 - F(t + s)}{1 - F(t)} = \frac{e^{-\lambda \cdot (t + s)}}{e^{-\lambda \cdot t}} = e^{-\lambda \cdot s} = 1 - F(s) = P(X > s)$$

- **B24:** Unterstellt man, dass die Lebensdauer X bestimmter Glühlampen (in Stunden) keinem Alterungsprozess unterliegt, so kann man mit Exponentialverteilung modellieren; für die erwartete Lebensdauer gelte 800 [h].
 - a) Wie groß ist die Wahrscheinlichkeit, dass eine Glühlampe länger als 1000 Stunden brennt?
 - b) Welche Brenndauer wird nur mit einer Wahrscheinlichkeit von 10% überschritten?

a) Die Glühlampe ist "zufällig ausgewählt". Wegen $E(X) = \frac{1}{\lambda} = 800$ ergibt sich $\lambda = \frac{1}{800}$ Die Überlebenswahrscheinlichkeit ist:

$$P(X > 1000) = 1 - F(1000) = e^{-\lambda \cdot 1000} = e^{-\frac{1}{800} \cdot 1000} = 0,29$$

b) Wieder muss man die Aussagen auf "zufällig gezogene" Lampen beziehen. Gesucht ist der 90%-Punkt der Verteilung: $x_{0,9} = \frac{1}{\lambda} \cdot \ln(\frac{1}{1 - 0.9}) = \dots = 1842.07$

Def.: Für eine Zufallsvariable X mit Verteilungsfunktion F(x) und Dichte f(x), die als Lebensdauer interpretiert wird, heißt

$$S(x) := 1 - F(x)$$
, $x \ge 0$ Überlebensfunktion (survivor function);

$$h(x) := \frac{f(x)}{1 - F(x)}, \quad x \ge 0$$
 Ausfallrate (hazard rate)

Bem.: Angenommen, die Lebensdauer sei größer als x. Die Frage ist, ob X – dadurch bedingt – nur mehr Δx Zeiteinheiten lebt. Für diese bedingte Wahrscheinlichkeit gilt:

$$P(x < X \le x + \Delta x \mid X > x) = \frac{P(x < X \le x + \Delta x)}{P(X > x)} = \frac{F(x + \Delta x) - F(x)}{S(x)} \approx \frac{f(x) \cdot \Delta x}{S(x)} = h(x) \cdot \Delta x$$

Für ein kleines Zeitintervall $\Delta x \rightarrow 0$ wird die angegebene Näherung gut. Man kann die Ausfallrate als "bedingte Wahrscheinlichkeit des Ausfalls pro Zeiteinheit zum Zeitpunkt x" auffassen, es gilt sogar:

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x < X \le x + \Delta x \mid X > x)}{\Delta x}$$

Satz: Für eine Lebensdauerverteilung X mit $T_X = [0, \infty)$ Verteilungsfunktion F(x), Überlebensfunktion S(x)und Ausfallrate h(x) gilt:

$$F(x) = 1 - S(x) = 1 - \exp\left(-\int_{0}^{x} h(t) \cdot dt\right)$$

Bem.: Aus der Ausfallrate kann man also durch Integration sowohl Verteilungsfunktion als auch Überlebensfunktion berechnen. Die Ausfallrate charakterisiert eine Lebensdauerverteilung.

Bew.: Es gilt $h(x) = -\frac{d}{dx}(\ln(S(x)))$, wie man leicht durch Ausdifferenzieren bestätigt. Daher ist $\ln(S(x))$ eine Stammfunktion von h(x). Setzt man noch die Grenzen 0 und x ein und wendet die Exponentialfunktion an, so erhält man das Ergebnis.

Satz (Ausfallrate der Exponentialverteilung):

$$X \sim \text{Exp}(\lambda)$$
, dann gilt: $h(x) = \lambda$

Bem.: Die konstante Ausfallrate der Exponentialverteilung entspricht genau der Gedächtnislosigkeit.



2.4 Spezielle stetige Verteilungen





Stetige Gleichverteilung

Exponentialverteilung

Pareto-Verteilung

Normalverteilung

Lognormalverteilung

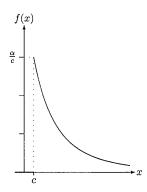
Def.: Eine Zufallsvariable X mit Träger $T_X = [c, \infty)$ heißt *Pareto-verteilt* mit den Parametern $\alpha > 0$ und c > 0, wenn die Verteilungsfunktion F(x) wie folgt festgelegt ist:

$$F(x) := \begin{cases} 1 - \left(\frac{c}{x}\right)^{\alpha} & x \ge c \\ 0 & x < c \end{cases}$$
; wir schreiben $X \sim Par(\alpha, c)$

Bem.: Man bezeichnet in einem allgemeinen Kontext die Überlebensfunktion S auch als Überschreitensfunktion. Für die Überschreitensfunktion einer Pareto-Verteilung gilt also (nach Logarithmieren) eine lineare Beziehung zwischen den logarithmischen Variablen:

$$S(x) = \left(\frac{c}{x}\right)^{\alpha}$$
 und damit: $\ln(S(x)) = \ln\left(\frac{c}{x}\right)^{\alpha} = \alpha \cdot \ln(c) - \alpha \cdot \ln(x)$

Bem.: Pareto hat empirisch diese logarithmische Beziehung für ökonomische Daten "bestätigen" können. Er untersuchte das Einkommen X in wohlhabenden Haushalten (die Beschränkung auf Einkommen > c).



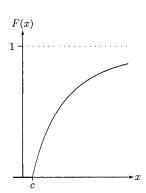


Abb. 20: Dichte und Verteilungsfunktion einer Pareto-Verteilung

Satz: Eine Pareto-verteilte Zufallvariable mit den Parametern $\alpha > 0$ und c > 0 hat eine Dichtefunktion f(x) mit:

$$f(x) := \begin{cases} \alpha \cdot \frac{c^{\alpha}}{x^{\alpha+1}} & x \ge c \\ 0 & x < c \end{cases}$$

B25: Studierende einer exklusiven Privatuniversität "haben" ein verfügbares Monatseinkommen von mindestens 1500 €. Auf Basis eines Pareto-Modells mit den Parametern α = 2,1 und c = 1500 löse man folgende Aufgaben:

- a) Berechnen Sie den Median der Einkommensverteilung.
- b) Wie viel Prozent der Studierenden haben mehr als 2500 € zur Verfügung?
- a) Zu lösen ist die Gleichung F(x) = 0.5, also

$$1 - \left(\frac{c}{x}\right)^{\alpha} = 0.5 \text{ oder } \alpha \cdot \ln\left(\frac{c}{x}\right) = \ln(0.5) \text{ bzw. } \alpha \cdot \ln(c) - \alpha \cdot \ln(x) = \ln(0.5)$$

Setzt man die Parameter ein, so erhält man als Median $x_{0,5} = 2086,6$.

b) Wir berechnen $S(x) = \left(\frac{c}{x}\right)^{\alpha} = 1 - F(x)$ durch Einsetzen der Parameter des Modells:

$$S(2500) = \left(\frac{1500}{2500}\right)^{2,1} = 0,342.$$

Satz (Erwartungswert und Varianz der Pareto-Verteilung):

$$X \sim Par(\lambda)$$
, dann gilt: a) $E(X) = \frac{\alpha \cdot c}{\alpha - 1}$, falls $\alpha > 1$ b) $V(X) = \frac{\alpha \cdot c^2}{(\alpha - 2)(\alpha - 1)^2}$, falls $\alpha > 2$

Bem.: Die Integrationen sind nicht schwer, wir lassen sie weg.



2.4 Spezielle stetige Verteilungen





Stetige Gleichverteilung

Exponentialverteilung

Pareto-Verteilung

Normalverteilung

Lognormalverteilung

Def.: Eine Zufallsvariable X hat eine *Normalverteilung* mit Parametern $\mu \in \mathbf{R}$ und σ^2 $(\sigma > 0)$, wenn sie die Dichte

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad x \in \mathbf{R} \quad \text{besitzt; wir schreiben } X \sim N(\mu, \sigma^2).$$

Def.: Die Zufallsvariable U hat eine Standard-Normalverteilung, wenn sie normalverteilt ist mit $\mu = 0$ und $\sigma^2 = 1$, d. h. $U \sim N(0, 1)$

Bem.: Die Standard-Normalverteilung hat also die Dichte $\varphi(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}u^2}$ $u \in \mathbb{R}$

Satz: Ist U standard-normal verteilt, so ist die affin-lineare Transformation $X := \mu + \sigma \cdot U \ (\sigma > 0)$ normal verteilt mit Parametern $\mu \in \mathbf{R}$ und σ^2

Bem.: Die Dichtefunktion einer allgemeinen Normalverteilung geht durch Strecken bzw. Stauchen und Verschieben aus der Dichte φ der Standard-Normalverteilung hervor.

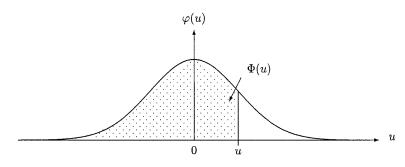
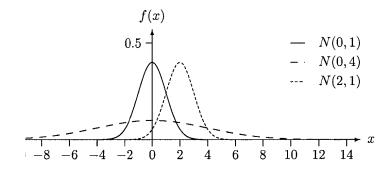


Abb. 21: Dichte der Standard-Normalverteilung



Satz (Eigenschaften der Dichtefunktionen der Normalverteilung):			
	$u = \frac{x - \mu}{\sigma}$	$\varphi(u)$	f(x)
	Maximum bei	u = 0	$x = \mu$
	Symmetrisch zu	u = 0	$x = \mu$
	Asymptoten	$\varphi(u) \to 0, u \to \pm \infty$	$f(x) \to 0, x \to \pm \infty$
	Wendepunkte bei	$u = \pm 1$	$x = \mu \pm 1 \cdot \sigma$

Def. (Verteilungsfunktion der Standard-Normalverteilung):

Die Verteilungsfunktion von $U \sim N(0,1)$ wird mit $\Phi(u)$ und bezeichnet und ist durch folgendes Integral gegeben:

$$\Phi(u) := \int_{-\infty}^{u} \varphi(t) \, dt$$

Bem.: Dieses Integral ist nicht durch die üblichen, elementaren Funktionen darstellbar, es muss numerisch ausgewertet werden – es gibt ausreichende Tabellen dazu oder man nimmt EXCEL oder eine andere Rechenhilfe, wo das Integral mittels Approximationsmethoden ausgewertet wird.

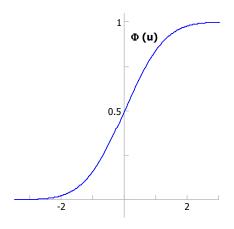


Abb. 23: Verteilungsfunktion der Standardnormalverteilung

Def. (Quantile der Standard-Normalverteilung):

Die eindeutige Lösung der Gleichung $\Phi(u) = p$, das so genannte p-Quantil, wird mit u_p bezeichnet.

Bem.: Als stetige Verteilung hat die angegebene Gleichung eine eindeutige Lösung; die Lösung muss aus Tabellen abgelesen werden, da die Verteilungsfunktion nicht durch einen elementaren Funktionsterm darstellbar wird. Heute liest man die Quantile aus EXCEL oder anderen Rechenhilfsmitteln ab.

Satz (Symmetrie von Verteilungsfunktion und Quantilen): $\,U \sim N(0,1)\,,\,so~gilt\,$

a)
$$\Phi(u) = 1 - \Phi(-u)$$

b)
$$u_p = -u_{1-p}$$

Bem.: Man kann sich das leicht aus der Symmetrie der Dichtefunktion graphisch herleiten.

Satz (Erwartungswert und Varianz der Standardnormalverteilung):

$$U \sim N(0, 1)$$
, dann gilt a) $E(U) = 0$ b) $V(U) = 1$

Bew.: a) ist eine unmittelbare Folge der schiefen Symmetrie des Integranden für den Erwartungswert: $u \cdot \varphi(u) = -(-u \cdot \varphi(-u))$.

b) Nach dem Verschiebungssatz und a) hat man für die Varianz:

$$V(U) = E(U^2) - (E(U))^2 = E(U^2)$$
.

Diesen Erwartungswert berechnet man mit partieller Integration:

$$E(U^{2}) = \int_{-\infty}^{\infty} u \cdot (u \cdot \varphi(u)) \cdot du = u \cdot (-\varphi(u)) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 1 \cdot (-\varphi(u)) \cdot du = 0 - (-1) = 1$$

$$\int f \cdot g' \qquad f \cdot g \Big| \qquad - \int f' \cdot g$$

Dazu verwendet man: i) $u \cdot \varphi(u)$ hat Stammfunktion $-\varphi(u)$ (durch Ausdifferenzieren leicht zu bestätigen),

ii)
$$\lim_{u\to\infty} u\cdot\varphi(u)=0$$
, ebenso für $u\to-\infty$ und iii) $\int\limits_{-\infty}^\infty \varphi(u)\cdot du=1$.

B27: Für $U \sim N(0, 1)$ bestimme man (mit EXCEL etc.):

- a) P(U > 0.5) sowie $P(-1 \le U \le 2)$
- b) den 20%- und 70%-Punkt der Verteilung von U
- c) den Wert von b, so dass $P(-b \le U \le b) = 0.9$
- a) $P(U > 0.5) = 1 P(U \le 0.5) = 1 \Phi(0.5) = 1 0.6915 = 0.3085$ $P(-1 \le U \le 2) = \Phi(2) - \Phi(-1) = \Phi(2) - (1 - \Phi(1)) = 0.9772 - (1 - 0.8413) = 0.8185$ Den letzten Zwischenschritt braucht man nur bei Tabellen, weil diese nur für u > 0 vorhanden sind.
- b) Das 0,2- bzw. 0,7-Quantil erhält man leicht in EXCEL: $u_{0,2} = -0.8416$ und $u_{0,7} = 0.5244$.

Arbeitet man mit Tabellen, muss man $u_{0,2} = -u_{0,8}$ nützen und allenfalls Tabellenwerte interpolieren.

Satz (Verteilungsfunktion und Quantile der allgemeinen Normalverteilung):

Zwischen den Zufallsvariablen $X \sim N(\mu, \sigma^2)$ und $U \sim N(0, 1)$ bestehen folgende Beziehungen:

	$U = \frac{X - \mu}{\sigma} \sim N(0, 1)$	$X = \mu + \sigma \cdot U \sim N(\mu, \sigma^2)$
Verteilungsfunktion	$\Phi(\frac{x-\mu}{\sigma})$	F(x)
Quantile	$u_p = \frac{x_p - \mu}{\sigma}$	$x_p = \mu + \sigma \cdot u_p$

Bew.: Nur für die Verteilungsfunktion; man hat

$$F(x) = P(X \le x) = P(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}) = P(U \le \frac{x - \mu}{\sigma}) = \Phi(\frac{x - \mu}{\sigma})$$

B28: Sei
$$X \sim N(\mu = 2, \sigma^2 = 25)$$
. Man bestimme

- a) P(X > 3) sowie $P(0 \le X \le 2)$
- b) den 80%- und 95%-Punkt der Verteilung von X
- c) den Wert von b, so dass $P(2-b \le X \le 2+b) = 0.9$

a)
$$P(X > 3) = 1 - P(X \le 3) = 1 - F(3)$$

$$= 1 - \Phi(\frac{3 - 2}{5}) = 1 - \Phi(0, 2) = 1 - 0,5793 = 0,4207$$

$$P(0 \le X \le 2) = F(2) - F(0)$$

$$= \Phi(\frac{2 - 2}{5}) - \Phi(\frac{0 - 2}{5}) = \Phi(0) - \Phi(-0, 4) = 0,5 - (1 - 0,6554) = 0,1554$$
b)

b)
$$x_{0,8} = \mu + \sigma \cdot u_{0,8} = 2 + 5 \cdot (0,8416) = 6,208$$
$$x_{0,95} = \mu + \sigma \cdot u_{0,95} = 2 + 5 \cdot (1,6449) = 10,224$$

c)

$$P(2-b \le X \le 2+b) = F(2+b) - F(2-b)$$

$$= \Phi(\frac{2+b-2}{5}) - \Phi(\frac{2-b-2}{5}) = \Phi(\frac{b}{5}) - (1-\Phi(\frac{b}{5})) = 2 \cdot \Phi(\frac{b}{5}) - 1 = 0,9$$

Daraus ergibt sich *b* als Lösung von $\Phi(\frac{b}{5}) = \frac{0.9 + 1}{2} = 0.95$, d. h., $\frac{b}{5} = u_{0.95} = 1.645$, und damit ist b = 8,224

Def.: Ist $\mu := E(X)$, so nennt man ein Intervall $[\mu - a, \mu + a]$ zentrales Schwankungsintervall der Breite a.

Bem.: Ist $X \sim N(\mu, \sigma^2)$, nennt man zentrale Schwankungsintervalle der Breite $\lambda \cdot \sigma$ auch λ -sigma-Bereiche.

B: Ist $X \sim N(\mu, \sigma^2)$, so haben die zentralen sigma-Bereiche folgende Wahrscheinlichkeiten im Vergleich zu den Abschätzungen aus der Tschebyscheff-Ungleichung:

$$P(\mu - \sigma \le X \le \mu - \sigma) = 0,683 \qquad --$$

$$P(\mu - 2 \cdot \sigma \le X \le \mu - 2 \cdot \sigma) = 0,954 \qquad \ge 0,75$$

$$P(\mu - 3 \cdot \sigma \le X \le \mu - 3 \cdot \sigma) = 0,997 \qquad \ge 0,89$$

Bem.: Die Abschätzungen aus der Tschebyscheff-Ungleichung sind für praktische Zwecke ziemlich wertlos, ihre Vorteile liegen in theoretischen Folgerungen daraus.

Bem.: Normalverteilungen sind dann als gute Modelle zu erwarten, wenn sich der Wert der tatsächlich zu messenden Größe als Summe von vielen kleinen Abweichungen zusammen setzt. (vgl. den Zentralen Grenzverteilungssatz). Insbesondere denkt man sich Fehler (bei Messungen in der Physik oder bei Abweichungen von Sollgrößen in der Produktion) als entstanden durch das Zusammenwirken vieler kleiner Einflussgrößen. Daher modelliert man Fehler gerne und erfolgreich mit der Normalverteilung. Fehler tauchen auch auf als

Abweichungen von Daten von einem Modell, das man Daten unterstellt. Auch hier ist es üblich, diese Fehler mit der Normalverteilung zu modellieren.



2.4 Spezielle stetige Verteilungen





- Stetige Gleichverteilung
- **Exponentialverteilung**
- **Pareto-Verteilung**
- Normalverteilung
- Lognormalverteilung

Def.: Die Zufallsvariable Y mit dem Träger $[0, \infty)$ heißt lognormalverteilt, wenn $X := \ln(Y)$ normalverteilt ist, und zwar mit denselben Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$; wir schreiben $Y \sim LN(\mu, \sigma^2)$

Bem.: Man kann auch umgekehrt feststellen, dass $Y := e^{X}$ lognormalverteilt ist, wenn X normalverteilt ist.

Satz (Verteilungsfunktion und Dichte der Lognormalverteilung): Für $Y \sim LN(\mu, \sigma^2)$ gilt

Verteilungsfunktion von Y

Dichtefunktion von Y

$$F_{Y}(y) := \begin{cases} \Phi\left(\frac{\ln y - \mu}{\sigma}\right) & y > 0 \\ 0 & y \le 0 \end{cases}$$

$$f_{Y}(y) := \begin{cases} \frac{1}{\sigma y} \cdot \varphi \left(\frac{\ln y - \mu}{\sigma} \right) & y > 0 \\ 0 & y \le 0 \end{cases}$$

Bew.: Wegen der Monotonie des Logarithmus hat man $(Y \le y) \equiv (\ln Y \le \ln y)$, d. h., die beiden Ereignisse haben dieselbe Wahrscheinlichkeit. Das erste Ereignis führt zur Verteilungsfunktion von Y, das zweite – nach entsprechender Standardisierung – zur Standardnormalverteilung. Die Dichtefunktion erhält man durch Differenzieren.

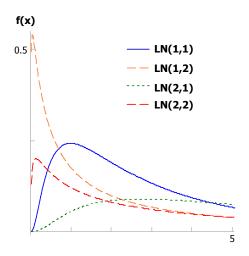


Abb. 24: Dichtefunktionen verschiedener Lognormalverteilungen

Satz (Erwartungswert und Varianz der Lognormalverteilung):

$$Y \sim LN(\mu, \sigma^2)$$
, dann gilt a) $E(Y) = e^{\mu + \sigma^2/2}$ b) $V(Y) = e^{2\mu + \sigma^2} \cdot (e^{\sigma^2} - 1)$ c) $y_{0.5} = e^{\mu}$

Bem.: Die Notation hinsichtlich der Lognormalverteilung ist nicht einheitlich. Manchmal findet man auch die Parameter $\xi = e^{\mu}$ und $\eta = e^{\sigma}$ in Verwendung. Einige Formeln gestalten sich dadurch einfacher. Der Zusammenhang mit der zugehörigen Normalverteilung geht aber verloren.

Bem.: Lognormalverteilungen haben immer eine positive Schiefe, sind also linkssteil. Interessante Anwendungen finden sich zur Analyse von Renditeverteilungen am Aktienmarkt.

- **B31:** Zur Analyse der verfügbaren Einkommen (in 1000 €) in einer Gesamtheit von Haushalten verwende man die Lognormalverteilung mit den Parametern $\mu = 0.9$ und $\sigma^2 = 0.1$. Man berechne
 - a) Erwartungswert, b) Standardabweichung und c) den Median dieser Verteilung.

Man setzt einfach in die Formeln ein und erhält:

a)
$$E(Y) = e^{0.9+0.1/2} = 2.586$$

a)
$$E(Y) = e^{0.9 + 0.1/2} = 2,586$$
 b) $\sigma(Y) = e^{0.9 + 0.1/2} \cdot \sqrt{e^{0.1} - 1} = 0,839$ c) $y_{0.5} = e^{0.9} = 2,460$

c)
$$y_{0,5} = e^{0,9} = 2,460$$



2 Zufallsvariable und ihre Verteilungen

- **Inhaltsverzeichnis**
- 2.1 Grundbegriffe
- 2.2 Verteilungsparameter
- 2.3 Spezielle diskrete Verteilungen
- 2.4 Spezielle stetige Verteilungen
- 2.5 Mathematische Ergänzungen

2.5 Mathematische Ergänzungen



↑

%%

2.5 Mathematische Ergänzungen



Borel-Mengen und Verteilung einer Zufallsvariablen

%%

Mit Hilfe der Verteilungsfunktion F einer Zufallsvariablen X kann man ganz einfach die Wahrscheinlichkeit dafür angeben, dass X in ein gegebenes Intervall fällt. Aus der Verteilungsfunktion lässt sich jedoch auch für allgemeinere Mengen B von reellen Zahlen die Wahrscheinlichkeit

$$P(X \in B) = P(\{\omega \in X(\omega) \in B\})$$

bestimmen. Beispielsweise, wenn B die Vereinigung von disjunkten Intervallen ist, ergibt sich $P(X \in B)$ als Summe der Wahrscheinlichkeiten der einzelnen Intervalle.

Satz: Man kann mathematisch beweisen, dass durch die Verteilungsfunktion von X die Wahrscheinlichkeiten $P(X \in B)$ für alle $B \in F$ bestimmt sind. Dabei ist F das Mengensystem der so genannten Borel-Mengen.

Def.: Das Mengensystem & der so genannten Borel-Mengen besteht aus gerade jenen Mengen, die sich durch wiederholte Ereignisoperationen aus Intervallen bilden lassen.

Satz: Die Borel-Mengen bilden eine Ereignisalgebra.

Die Borel-Mengen bilden sogar eine Ereignis- σ -Algebra.

D. h.: Nicht nur die Vereinigung von endliche vielen Mengen in \mathcal{E} ist wieder eine Menge in \mathcal{E} , auch wenn abzählbare viele Borel-Mengen vereinigt werden, führt das immer zu einer Borel-Menge.

Bem.: Alle für Anwendungen interessanten Mengen sind Borel-Mengen; es handelt sich zudem meist noch um recht einfache Borel-Mengen, nämlich im Wesentlichen alle Arten von endlichen oder unendlichen Intervallen und deren Vereinigung. Es ist sogar ausgesprochen schwierig, überhaupt eine Menge zu finden, die keine Borel- Menge ist; diese Mengen sind einer einfachen Vorstellung nicht zugänglich.

Def.: Die Funktion $B \mapsto P(X \in B), B \in \mathcal{B}$

heißt Wahrscheinlichkeitsverteilung oder einfach nur Verteilung von X.

Satz: Diese Verteilung von X ist durch ihre Verteilungsfunktion

$$x \mapsto F(X \in (-\infty, x]), x \in \mathbf{R}$$

bereits eindeutig festgelegt.

Bem: a) Die Verteilung einer diskreten Zufallsvariablen X ist bereits durch die Einzelwahrscheinlichkeiten $p_j = P(X = x_j), j = 1, 3, ...$ bestimmt.

b) Die Verteilung einer stetigen Zufallsvariablen X ist durch ihre Dichte f(x) eindeutig festgelegt.



3 Gemeinsame Verteilung und Grenzwertsätze

MANFRED BOROVCNIK

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



Inhaltsverzeichnis

- 3.1 Gemeinsame Verteilung von Zufallsvariablen
- 3.2 Grenzwertsätze
- 3.3 Ergänzungen

In Kapitel 2 wurden einzelne Zufallsvariable im Hinblick auf ihre Verteilung untersucht. In vielen Anwendungen untersucht man dagegen das gemeinsame Auftreten mehrerer Merkmale und ihre wechselweisen Abhängigkeiten. In der schließenden Statistik hat man es regelmäßig mit mehreren Zufallsvariablen zu tun. Dabei werden aus den beobachteten Realisationen wiederholter, gleichartiger Zufallsvorgänge Schlussfolgerungen gezogen. Speziell werden Summen von Zufallsvariablen untersucht, weil diese für die Verfahren in der schließenden Statistik besonders wichtig sind. Zuerst werden Erwartungswert und Varianz von Summen behandelt. Dann wird die Konvergenz von Summen behandelt (Schwaches Gesetz der Großen Zahlen) und die Verteilung näherungsweise bestimmt (über den Zentralen Grenzverteilungssatz). Ergänzt wird das Kapitel durch Einführungen in die multivariate Normalverteilung, den Poisson-Prozess und die Monte-Carlo-Simulation.

3.1 Gemeinsame Verteilung von Zufallsvariablen

- Gemeinsame Verteilung von zwei Zufallsvariablen
- **☑** Gemeinsame Verteilung von *n* Zufallsvariablen
- Summen von unabhängigen Zufallsvariablen

Kann man aus dem Alter $X(\omega)$ einer Person ω auf das vermutliche Abschneiden $Y(\omega)$ bei einer Prüfung schließen? Wie beschreibt man die gemeinsame Verteilung von zwei Zufallsvariablen? Wie modelliert man, dass zwei Zufallsvariable unabhängig sind, wie misst man "Grade" eines stochastischen Zusammenhangs zwischen den Zufallsvariablen. Wie überträgt man die Begriffe auf mehr als zwei Zufallsvariable? Wie bestimmt man den Erwartungswert und die Verteilung von Summen mehrerer Zufallsvariablen?

3.1 Gemeinsame Verteilung von Zufallsvariablen





Gemeinsame Verteilung von zwei Zufallsvariablen

Gemeinsame Verteilung von *n* **Zufallsvariablen**

Summen von unabhängigen Zufallsvariablen

Die genannten Begriffsbildungen werden mit zwei Zufallsvariablen entwickelt.

Def.: Die Zufallsvariablen $X_1, X_2, ..., X_n$ beziehen sich auf ein und denselben Zufallsvorgang; d. h. sie sind alle auf derselben Ergebnismenge Ω definiert:

$$X_i: \Omega \to \mathbf{R}$$
.

 $(X_1, X_2, ..., X_n)$ heißt dann Zufallsvektor.

Bem.: Wenn der Zufallsvorgang das Ergebnis $\omega \in \Omega$ hat, nehmen diese Zufallsvariablen gemeinsam die Werte $X_1(\omega), X_2(\omega), ..., X_n(\omega)$ an. Die gemeinsame Wahrscheinlichkeitsverteilung der n Zufallsvariablen soll untersucht werden. Wesentliche Einsichten bietet schon der Fall n = 2.

Def.: Für den Zufallsvektor (X, Y) heißt

$$F_{XY}:~\textbf{R}^2~\rightarrow~[0,1]$$

$$F_{XY}(x, y) = P(X \le x, Y \le y)$$

= $P(\{\omega \mid X(\omega) \le x \text{ und } Y(\omega) \le y\})$

die gemeinsame Verteilungsfunktion von X und Y.

Bem.: Wir schreiben F(x, y) statt $F_{XY}(x, y)$ und verwenden X und Y statt X_1, X_2 .

Satz (Eigenschaften einer zweidimensionalen Verteilungsfunktion):

1. F(x, y) wächst monoton in beiden Variablen x und y.

2.
$$P(a_1 \le X \le b_1, a_2 \le Y \le b_2) = F(b_1, b_2) + F(a_1, a_2) - (F(a_1, b_2) + F(a_2, b_1))$$

3. $F_X(x) = \lim_{y \to \infty} F(x, y)$ für alle $x \in \mathbb{R}$ und analog für $F_Y(y)$

Bem.: 2. gibt die Wahrscheinlichkeit des Rechtecks $(a_1, b_1] \times (a_2, b_2]$ an. Die folgende Abbildung zeigt, dass die Wahrscheinlichkeit durch die entsprechenden "Quadranten" berechnet werden kann. 3. zeigt, dass man aus der gemeinsamen Verteilungsfunktion von X und Y auch die der so genannten Randverteilungen von X bzw. Y isoliert zurück gewinnen kann.

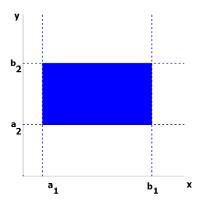


Abb. 1: Wahrscheinlichkeit, dass (X, Y) in ein achsenparalleles Rechteck fällt

Def.: Die Zufallsvariablen X und Y mit der gemeinsamen Verteilungsfunktion F(x, y) heißen *stochastisch unabhängig*, wenn gilt:

$$F(x, y) = F_X(x) \cdot F_Y(y)$$
 für alle $x, y \in \mathbb{R}$

Bem.: X und Y sind also genau dann stochastisch unabhängig, wenn sich ihre gemeinsame Verteilungsfunktion durch Multiplikation der beiden Randverteilungsfunktionen ergibt. Es ist keine zusätzliche Information notwendig darüber, wie die beiden Zufallsvariablen gemeinsam variieren. Man kann es auch so ausdrücken: alle Ereignisse, die nur X betreffen sind von allen Ereignissen, die nur Y betreffen, unabhängig: Die Wahrscheinlichkeit des gemeinsamen Auftretens von $(X \le x)$ und $(Y \le y)$ ergibt sich als Produkt der Einzelwahrscheinlichkeiten.

B1: Untersucht werden die Brenndauern zweier Glühlampen.

a) Die Zufallsvariablen X und Y bezeichnen die Brenndauern zweier Glühlampen aus zwei verschiedenen Produktionslosen. Die gemeinsame Verteilungsfunktion von X und Y sei durch

$$F(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x - y} & \text{falls} & x \ge 0, y \ge 0 \\ 0 & \text{sonst} \end{cases}$$

gegeben. Die Randverteilungsfunktionen sind beide exponentialverteilt mit $\lambda = 1$; sei nur die Verteilungsfunktion von X angegeben:

$$F_{X}(x) = \begin{cases} 1 - e^{-x} & \text{falls} \quad x \ge 0 \\ 0 & \text{sonst} \end{cases}$$

Offenbar gilt hier die Produkt"regel" $F(x, y) = F_X(x) \cdot F_Y(y)$; d.h., X und Y sind stochastisch unabhängig.

b) Die Zufallsvariablen Z und V mögen nun die Brenndauern von zwei Glühlampen aus demselben Produktionslos repräsentieren. Die gemeinsame Verteilungsfunktion von Z und V sei durch

Froduktions representerent. Die gemeinsame Verteilungsfunktion
$$F(z, v) = \begin{cases} (1 - e^{-z}) \cdot (1 - e^{-v}) \cdot (1 + \frac{1}{2}e^{-z - v}) & \text{falls} \quad z \ge 0, v \ge 0 \\ 0 & \text{sonst} \end{cases}$$

gegeben. Für die Randverteilungsfunktionen ergeben sich wiederum für Z und V Exponentialverteilungen mit $\lambda = 1$, wie man sich leicht überzeugt.

Z und V haben zwar die gleichen Randverteilungen wie X und Y in Teil a), die gemeinsamen Verteilungen sind jedoch verschieden. Im Gegensatz zu X und Y sind Z und V abhängig.

Bem.: Im Allgemeinen wird die gemeinsame Verteilungsfunktion durch die Randverteilungsfunktionen nicht eindeutig bestimmt, wie man aus dem obigen Beispiel sieht.

Satz: Zufallsvariable X und Y sind stochastisch unabhängig; g und h beliebige reellwertige Funktionen. Dann sind auch die beiden Zufallsvariablen g(X) und h(X) stochastisch unabhängig.

Bem.: Die Unabhängigkeit von Zufallsvariablen wird auf die Transformierten der Zufallsvariablen "vererbt".

B1 f: X und Y sind unabhängige Lebensdauern. Die neuen Zufallsvariablen X² und Y² sind nach dem Satz auch unabhängig.

Def.: Zwei Zufallsvariable X und Y heißen *gemeinsam diskret verteilt*, wenn es endlich viele (oder abzählbar unendlich viele) Werte $x_1, x_2, ...$ und $y_1, y_2, ...$ gibt, so dass

$$\sum_{j} \sum_{k} P(X = x_j, Y = y_k) = 1$$

Bem.: Man nennt die $p_{jk} := P(X = x_j, Y = y_k)$ auch diskrete Wahrscheinlichkeitsfunktion von (X, Y).

B: Die folgende Tabelle gibt eine diskrete Wahrscheinlichkeitsfunktion zweier Zufallsvariablen an. Aus der Tabelle liest man u. a. ab: P(X = 0, Y = 1) = 0,1. Die gemeinsame Verteilung kann man sich durch ein Stabdiagramm veranschaulichen.

Gemeinsame Verteilung		Y			Rand-
		0	1	2	verteilung X
X	0	0,1	0,1	0,2	0,4
	1	0,3	0,2	0,1	0,6
Rand- verteilung Y		0,4	0,3	0,3	1

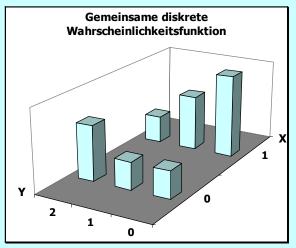
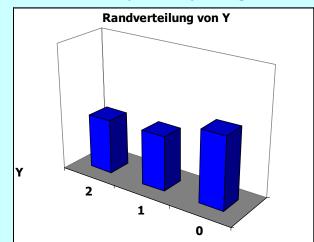


Abb. 2: Gemeinsame diskrete Wahrscheinlichkeitsfunktion als Stabdiagramm

Def.: X und Y sind gemeinsam diskret verteilt mit diskreter Wahrscheinlichkeitsfunktion $P(X = x_j, Y = y_k)$. Unter den *Randverteilungen von* X bzw. Y verstehen wir die Ausdrücke: $P(X = x_j) := \sum_k P(X = x_j, Y = y_k)$ bzw. $P(Y = x_k) := \sum_j P(X = x_j, Y = y_k)$

Bem.: Nach dem Satz von der totalen Wahrscheinlichkeit ergibt die erste Summe genau die Wahrscheinlichkeit, dass X den Wert x_j annimmt. Die Verteilungen von X und Y für sich (ohne die gemeinsame Verteilung) heißen auch Randverteilungen – sie erscheinen am Rand der Tabelle mit den Wahrscheinlichkeiten für das gemeinsame Auftreten von X und Y.

B f: Die Randverteilungen im obigen Beispiel veranschaulicht man auch durch Stabdiagramme.



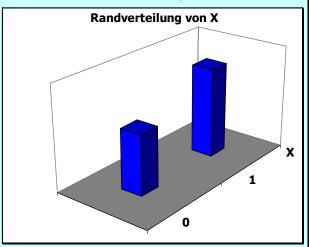


Abb. 3: Die Randverteilungen der einzelnen Zufallsvariablen – als Summen über Zeilen bzw. Spalten

- **B 2:** Wir betrachten eine Urne mit N Kugeln; davon seien M rot und N-M weiß.
 - a) Es wird zweimal mit Zurücklegen gezogen. Die Zufallsvariablen X und Y beschreiben die Farbe beim ersten bzw. beim zweiten Zug:
 - $X = \left\{ \begin{array}{ll} 1 & \text{falls die erste gezogene Kugel rot ist} \\ 0 & \text{falls die erste gezogene Kugel weiß ist} \end{array} \right. Y = \left\{ \begin{array}{ll} 1 & \text{falls die zweite gezogene Kugel rot ist} \\ 0 & \text{falls die zweite gezogene Kugel weiß ist} \end{array} \right.$

Die gemeinsamen Wahrscheinlichkeiten von X und Y sowie ihre Randwahrscheinlichkeiten sind in der folgenden Tabelle zusammengefasst:

	Y = 0	Y = 1	Zeilensumme
X = 0	$\left(\frac{N-M}{N}\right)^2$	$\frac{N-M}{N} \cdot \frac{M}{N}$	$\frac{N-M}{N}$
X = 1	$\frac{M}{N} \cdot \frac{N-M}{N}$	$\left(\frac{M}{N}\right)^2$	$\frac{M}{N}$
Spaltensumme	$\frac{N-M}{N}$	$\frac{M}{N}$	1

Für alle möglichen Ergebnisse bekommt man die Wahrscheinlichkeiten durch Multiplizieren der entsprechenden Randwahrscheinlichkeiten; man kann das für alle Einträge der Tabelle überprüfen – etwa:

$$P(X = 0, Y = 1) = \frac{N - M}{N} \cdot \frac{M}{N} = P(X = 0) \cdot P(Y = 1)$$
.

Die entsprechenden Ereignisse sind stochastisch unabhängig.

b) Es wird zweimal ohne Zurücklegen gezogen. Die Zufallsvariablen X und Y beschreiben die Farbe beim ersten bzw. beim zweiten Zug wie in a). Die gemeinsamen Wahrscheinlichkeiten von X und Y sowie ihre Randwahrscheinlichkeiten sind in der folgenden Tabelle zusammengefasst:

	Y = 0	Y = 1	Zeilensumme
X = 0	$\frac{N-M}{N} \cdot \frac{N-M-1}{N-1}$	$\frac{N-M}{N} \cdot \frac{M}{N-1}$	$\frac{N-M}{N}$
X = 1	$\frac{M}{N} \cdot \frac{N-M}{N-1}$	$\frac{M}{N} \cdot \frac{M-1}{N-1}$	$\frac{M}{N}$
Spaltensumme	$\frac{N-M}{N}$	$\frac{M}{N}$	1

Man kann wieder probieren, die Wahrscheinlichkeiten des gemeinsamen Auftretens von Werten durch Multiplizieren der entsprechenden Randwahrscheinlichkeiten zu erhalten; es gilt etwa:

$$P(X = 0, Y = 1) = \frac{N - M}{N} \cdot \frac{M}{N - 1} \neq \frac{N - M}{N} \cdot \frac{M}{N} = P(X = 0) \cdot P(Y = 1)$$
.

Die entsprechenden Ereignisse sind nicht stochastisch unabhängig.

B2 f: Wie ist die Verteilung von X (Ergebnis der ersten Ziehung), wenn man die zweite kennt, also etwa (Y = 1). Zu bestimmen ist nun – nur Ziehen ohne Zurücklegen in b):

$$P(X = 0 | Y = 1) = \frac{P(X = 0 \text{ und } Y = 1)}{P(Y = 1)} = \frac{\frac{N - M}{N} \cdot \frac{M}{N - 1}}{\frac{M}{N}} = \frac{N - M}{N - 1}$$

$$P(X=1 | Y=1) = \frac{P(X=1 \text{ und } Y=1)}{P(Y=1)} = \frac{\frac{M}{N} \cdot \frac{M-1}{N-1}}{\frac{M}{N}} = \frac{M-1}{N-1}$$

Die beiden bedingten Wahrscheinlichkeiten ergeben summiert 1; es handelt sich auch um eine Wahrscheinlichkeitsverteilung. Man spricht auch von der Verteilung in der 2. Spalte. Diese erhält man aus den gemeinsamen Wahrscheinlichkeiten $P(X = x_j, Y = 1)$, indem man diese durch die Spaltensumme = P(Y = 1) dividiert ("normiert"). Analog spricht man von bedingten Zeilenverteilungen.

Def: X und Y sind gemeinsam diskret verteilt mit diskreter Wahrscheinlichkeitsfunktion $P(X = x_j, Y = y_k)$. Die folgenden Ausdrücke nennt man *bedingte Verteilungen*, und zwar:

$$P(X = x_j | Y = y_k) := \frac{P(X = x_j, Y = y_k)}{P(Y = y_k)}; \quad j = 1, 2, ...$$

$$P(Y = y_k \mid X = x_j) := \frac{P(X = x_j, Y = y_k)}{P(X = x_j)}; k = 1, 2, ...$$

sofern die Wahrscheinlichkeit der Bedingung $(Y = y_k)$ bzw. $(X = x_j)$ größer als Null ist.

Bem.: Man spricht von bedingter Verteilung von X unter der Bedingung $(Y = y_k)$ und schreibt: $X | (Y = y_k)$.

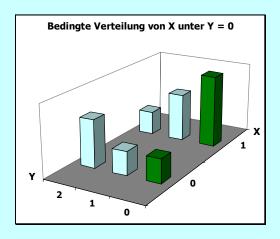
Satz: Die gemeinsam diskret verteilten Zufallsvariablen (X, Y) sind genau dann stochastisch unabhängig, wenn für alle x_i und y_k gilt:

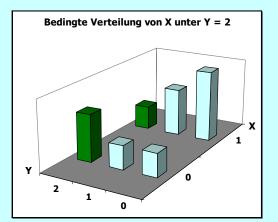
$$P(X = x_j, Y = y_k) = P(X = x_j) \cdot P(Y = y_k)$$

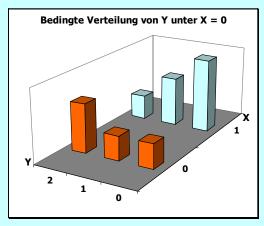
B f: Die bedingten Verteilungen im Beispiel mit der Tabelle für die gemeinsamen diskreten Wahrscheinlichkeiten gibt man auch in Tabellen an und veranschaulicht wieder durch Stabdiagramme:

Bedingte Verteilungen von Y			G		
		0	1	2	Summe
unter X	= 0	0,25	0,25	0,5	1
	= 1	0,5	0,33	0,17	1
= Zeilen- verteilung		= Wa keiten			

Bedingte Verteilungen von X		unter Y =			= Spalten-
		0	1	2	verteilung
X	= 0	0,25	0,33	0,67	=
	= 1	0,75	0,67	0,33	Wahrschein- lichkeiten
Summe		1	1	1	der Spalte auf 1 normiert







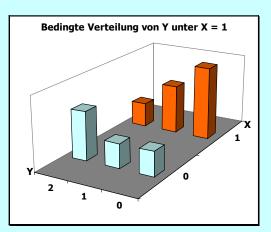


Abb. 4: Die bedingten Verteilungen von X bzw. Y – noch nicht auf Summe 1 normiert – wie sie in die gemeinsame Verteilung hineinpassen

Def.: Zwei Zufallsvariable X und Y heißen *gemeinsam stetig verteilt*, wenn die gemeinsame Verteilungsfunktion F(x, y) eine Integraldarstellung hat, d. h., es gibt eine so genannte gemeinsame Dichtefunkti-

on
$$f: \mathbf{R}^2 \to \mathbf{R}$$
 mit

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(z, v) \cdot dz \cdot dv \text{ für alle reellen } x \text{ und } y.$$

Satz: Die gemeinsame Dichte eines stetig verteilten Zufallsvektors (X, Y) hat folgende Eigenschaften:

i)
$$f(x, y) \ge 0$$
 und ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot dx \cdot dy = 1$. Es gilt ferner iii) $P(a_1 \le X \le b_1, a_2 \le Y \le b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y) dx dy$

iii)
$$P(a_1 \le X \le b_1, a_2 \le Y \le b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y) dx dy$$

Bem.: Die Wahrscheinlichkeiten von Rechtecken $(a_1, b_1] \times (a_2, b_2]$ ergeben sich durch mehrfache Integration. Ein Beispiel für die gemeinsame stetige Dichte ist die zweidimensionale Normalverteilung in folgender Abbildung.

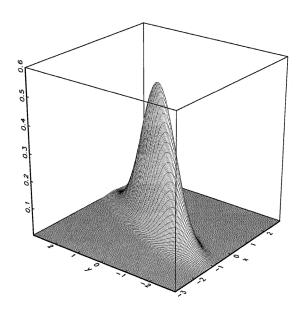


Abb.5: Gemeinsame Dichte zweier Zufallsvariablen X und Y (zweidimensionale Normalverteilungsdichte)

Satz: Aus der gemeinsamen stetigen Dichte f des Zufallsvektors (X, Y) erhält man

die Verteilungsfunktion der einzelnen Komponenten X durch Grenzübergang:

$$F_X(x) = F(x, \infty) = \lim_{y \to \infty} F(x, y) = \lim_{y \to \infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f(z, v) \cdot dz \cdot dv$$

b) die Dichten der Randverteilungen von X durch "Ausintegrieren":

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \cdot dy$$

Bem.: Analoge Ergebnisse gelten für die Randverteilung von Y.

Satz: Hat (X, Y) die gemeinsame Dichte f, so sind die beiden Zufallsvariablen stochastisch unabhängig, genau dann, wenn für alle x und $y \in \mathbf{R}$ gilt:

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

B1 f:

a) Die gemeinsame Verteilungsfunktion von X und Y ist durch

$$F(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & \text{falls} & x, y \ge 0 \\ 0 & \text{sonst} \end{cases}$$

Die gemeinsame Dichte erhält man durch zweimaliges Differenzieren nach y und nach x:

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \begin{cases} e^{-x-y} & \text{falls} \quad x, y \ge 0\\ 0 & \text{sonst} \end{cases}$$

Aus der gemeinsamen Dichte erhält man die Randdichten durch Integrieren:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \cdot dy = \int_{0}^{\infty} e^{-x-y} \cdot dy = e^{-x}$$
, falls $x \ge 0$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \cdot dx = \int_{0}^{\infty} e^{-x-y} \cdot dx = e^{-y}$$
, falls $y \ge 0$

Alternativ kann man auch die Randverteilungsfunktion differenzieren; etwa für X:

$$f_X(x) = \frac{d}{dx}F(x,\infty) = \frac{d}{dx}(1-e^{-x}) = e^{-x}$$
, falls $x \ge 0$

b) Die gemeinsame Verteilungsfunktion von Z und V ist

$$F(z, v) = \begin{cases} (1 - e^{-z}) \cdot (1 - e^{-v}) \cdot (1 + \frac{1}{2}e^{-z - v}) & \text{falls} \quad z \ge 0, v \ge 0 \\ 0 & \text{sonst} \end{cases}$$

Die gemeinsame Dichte erhält man durch zweimaliges Differenzieren:

$$f(z, v) = \frac{\partial^2}{\partial z \partial v} F(z, v) = \dots = \begin{cases} e^{-z-v} + \frac{1}{2} \cdot e^{-z-v} \cdot (2e^{-z} - 1)(2e^{-v} - 1) & z, v \ge 0 \\ 0 & \text{sonst} \end{cases}$$

Für die Randdichte von V etwa ergibt sich:

$$f_V(v) = \frac{d}{dv}F(\infty, v) = \frac{d}{dv}(1 - e^{-v}) = e^{-v}$$
, falls $v \ge 0$ (wieder 0 für $v < 0$).

Def.: Mit f(x, y) als gemeinsamer Dichte von (X, Y) und $f_X(x)$, $f_Y(y)$ als Randdichten von X bzw. Y definiert man als *bedingte Dichte von* X *unter der Bedingung* (Y = y) den folgenden Ausdruck:

$$f_{X|Y=y}(x) := \begin{cases} \frac{f(x,y)}{f_Y(y)} & \text{falls } f_Y(y) > 0 \\ 0 & \text{sonst} \end{cases}$$

Bem.: Analog ist der Ausdruck $f_{Y|X=x}(x)$ zu verstehen, x und y tauschen die Rollen.

Bem.: Der Ausdruck ist analog zur bedingten Verteilung von X unter der Bedingung, dass Y einen bestimmten Wert y_k einnimmt, gebildet. Jetzt wird die Schnittkurve des Graphen der gemeinsamen Dichte f(x, y) mit der Ebene Y = y durch die Normierung mit dem Nenner $f_Y(y)$ zu einer Dichtefunktion "aufgewertet".

B 1f: Für die Glühlampen in Teil b) soll die bedingte Dichte von Z unter der Bedingung V = 2 berechnet werden. Wir bilden den Quotienten aus gemeinsamer Dichte und Randdichte von V:

$$\mathbf{f}_{Z|V=2}(z) = \frac{\mathbf{f}(z,2)}{\mathbf{f}_{V}(2)} = \begin{cases} \frac{e^{-z-2} + \frac{1}{2} \cdot e^{-z-2} \cdot (2e^{-z} - 1) \cdot (2e^{-2} - 1)}{e^{-2}} = e^{-z} - 0.3647 \cdot e^{-z} \cdot (2e^{-z} - 1) & z \ge 0\\ 0 & \text{sonst} \end{cases}$$

Für die Bedingung V = 1,5 ergibt sich als bedingte Dichte ein anderes Ergebnis:

$$\mathbf{f}_{Z|V=2,5}(z) = \frac{\mathbf{f}(z,1,5)}{\mathbf{f}_V(1,5)} = \dots = \begin{cases} e^{-z} - 0.2769 \cdot e^{-z} \cdot (2e^{-z} - 1) & z \ge 0\\ 0 & \text{sonst} \end{cases}$$

Hingegen führt in Teil a) die Berechnung der bedingten Dichte immer zum selben Ergebnis, egal, welche Bedingung Y = y gestellt wird:

$$f_{X|Y=y}(x) = \dots = \begin{cases} e^{-x} & x \ge 0\\ 0 & \text{sonst} \end{cases}$$

Bem.: Sind X und Y gemeinsam stetig verteilt, so hat die Unabhängigkeit zwischen den beiden Zufallsvariablen verschiedene Ausprägungen:

- Es gilt die Produktformel: gemeinsame Dichte = Produkt der beiden Randdichten
- Die bedingten Dichten sind für alle Bedingungen immer gleich der Randdichte.

Def.: Für die Zufallsvariablen X und Y mit $\mu_X = E(X)$ und $\mu_Y = E(Y)$ heißt der Erwartungswert $\sigma_{XY} = Cov(X,Y) = E\{(X-\mu_X)\cdot (Y-\mu_Y)\}$ Kovarianz von X und Y.

Bem.: Die Auswertung der Kovarianz erfolgt über

$$\begin{array}{ll} \text{Doppelsummen} & \sum\limits_{j}\sum\limits_{k}\left(x_{j}-\mu_{\mathbf{X}}\right)\cdot\left(y_{k}-\mu_{\mathbf{Y}}\right)\cdot\mathbf{P}(\mathbf{X}=x_{j},\,\mathbf{Y}=y_{k}) & \text{falls}\left(\mathbf{X},\,\mathbf{Y}\right)\,\text{gemeinsam diskret} \\ \\ \text{Doppelintegrale} & \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left(x-\mu_{\mathbf{X}}\right)\cdot\left(y-\mu_{\mathbf{Y}}\right)\cdot\mathbf{f}(x,y)\cdot dx\cdot dy & \text{gemeinsam stetig} \\ \end{array}$$

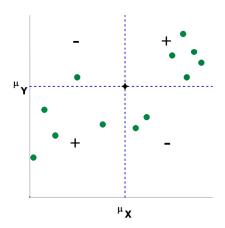


Abb. 6: Ein Koordinatenkreuz im Schwerpunkt der Verteilung erzeugt 4 Quadranten.

Bem.: Durch ein Koordinatenkreuz werden im "Schwerpunkt" (μ_X, μ_Y) der gemeinsamen Verteilung 4 Quadranten erzeugt. Das Vorzeichen von $(x - \mu_X) \cdot (y - \mu_Y)$ ist im 1. und 3. Quadranten positiv, im 2. und 4. Quadranten hingegen negativ. Wenn sehr viel Wahrscheinlichkeit im 1. und 3. Quadranten liegt und wenig in

den beiden anderen Quadranten, so ergibt das eine große Kovarianz. Liegt das Hauptgewicht der Verteilung hingegen im 2. und 4. Quadranten, so ergibt sich eine große negative Kovarianz. Ersteres entspricht einer "je größer X desto größer Y"-Beziehung – einem positiven Zusammenhang; Letzteres entspricht einer "je größer X desto kleiner Y"-Beziehung – einem negativen Zusammenhang. Die Kovarianz ist aber kein normiertes Maß für den "Zusammenhang" zweier Zufallsvariablen, sie kann beliebig große Werte annehmen.

Satz (Eigenschaften der Kovarianz):

1. Symmetrie: Cov(X, Y) = Cov(Y, X)

2. Fehlende Normierung: Cov(X, Y) kann beliebige Werte annehmen.

3. Linearität in beiden Argumenten: Cov(a + bX, Y) = a + bCov(X, Y) (für Y analog)

4. Verschiebungssatz: $Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$

5. X und Y stochastisch unabhängig, dann gilt: Cov(X, Y) = 0

Bew.: 4. – man kürzt die Erwartungswerte wie üblich durch μ_X , μ_Y ab. Der Beweis macht von der Additivität des Erwartungswerts Gebrauch (wird erst bewiesen) – danach gilt, egal, ob die Zufallsvariablen stochastisch abhängig oder unabhängig sind:

$$E(X+Y) = E(X) + E(Y)$$
 bzw. $E(aX+bY) = aE(X) + bE(Y)$

Nun hat man:

$$Cov(X,Y) = E\{(X-\mu_X)\cdot(Y-\mu_Y)\} = E\{X\cdot Y-\mu_X\cdot Y+X\cdot(-\mu_Y)\mu_X\cdot\mu_Y\} =$$

$$= E(X\cdot Y)-\mu_X\cdot E(Y)+E(X)\cdot(-\mu_Y)+\mu_X\cdot\mu_Y = E(X\cdot Y)-\mu_X\cdot\mu_Y$$

Bew. von 5. für unabhängige, gemeinsam diskret verteilte Zufallsvariable – für stetige Verteilungen analog:

$$E(X \cdot Y) = \sum_{j} \sum_{k} x_{j} \cdot y_{k} \cdot P(X = x_{j}) \cdot P(Y = y_{k})$$

$$= \sum_{j} x_{j} \cdot P(X = x_{j}) \sum_{k} y_{k} \cdot P(Y = y_{k})$$

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

Def.: Für die Zufallsvariablen X und Y mit $\sigma_X = \sqrt{V(X)}$ und $\sigma_Y = \sqrt{V(Y)}$ heißt der Erwartungswert

$$\rho_{XY} := Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Korrelation von X und Y; vorausgesetzt X und Y besitzen endliche positive Varianzen.

Satz (Eigenschaften der Korrelation):

1. Dimensionslosigkeit: ρ hat keine Benennung (keine Maßeinheit)

2. Normiertheit: $-1 \le \rho \le 1$; $\rho = \pm 1$ zeigt dabei einen vollständigen

linearen Zusammenhang an, d. h.: Y = a + bX

3. Linearität in beiden Argumenten (für Y analog): Corr(a+bX,Y) = |b|Corr(X,Y)

4. X und Y stochastisch unabhängig, dann gilt: $\rho_{XY} = 0$

Bem.: Der angesprochene lineare Zusammenhang in 2. gilt genau dann, wenn $\rho = \pm 1$.

In 4. hingegen gilt der Umkehrschluss nicht: Es gibt Verteilungen mit $\rho_{XY} = 0$, obwohl die beiden Zufallsvariablen X und Y stochastisch abhängig sind.

Bew.:

1. Setzt man in die Kovarianz die Zufallsvariablen ein und nutzt die Linearität des Erwartungswerts, so erhält man eine andere Darstellung für den Korrelationskoeffizienten als Erwartungswert des Produkts von standardisierten Zufallsvariablen $\frac{X-\mu_X}{\sigma_x}$:

$$Corr(X,Y) = E(\frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y}).$$

Standardisierte Zufallsvariable tragen keine Dimension – man kann sich vorstellen, dass die Einheiten herausgekürzt werden.

2. Nach der Cauchy-Schwarzschen Ungleichung gilt

$$|E((X-\mu_X)\cdot(Y-\mu_Y))| \leq \sqrt{E((X-\mu_X)^2)}\cdot\sqrt{E((Y-\mu_Y)^2)} = \sigma_X\cdot\sigma_Y$$

das ist auch schon der Beweis. Für Integrale liest sich diese Ungleichung so:

$$|\int f \cdot g| \le \sqrt{\int g^2} \cdot \sqrt{\int f^2}$$

- 3. und 4. werden direkt von der Kovarianz "geerbt".
- **B 3:** Stochastisch abhängige Zufallsvariable können Kovarianz und Korrelation 0 haben:

$$X \sim N(0, 1)$$
 und $Y := X^2$

Durch die funktionale Abhängigkeit steht zu vermuten, dass die beiden Zufallsvariablen auch stochastisch abhängig sind; zwei einfache Ereignisse bestätigen das:

$$P(X \ge 2, Y \ge 4) = P(X \ge 2) > P(X \ge 2) \cdot P(Y \ge 4)$$
, weil $P(Y \ge 4) < 1$.

Für die Kovarianz ergibt sich (mit Verschiebungssatz; Erwartungswert und Schiefe von X sind 0):

$$Cov(X, Y) = E(X \cdot X^{2}) - E(X) \cdot E(X^{2}) = E(X^{3}) = 0$$

B 2f: Für zweimaliges Ziehen aus der Urne mit N Kugeln, von denen M rot und N-M weiß sind, ergibt sich für die Korrelation zwischen der ersten und zweiten Ziehung 0, falls mit bzw. $-\frac{1}{N-1}$, falls ohne

Zurücklegen gezogen wird. Die beiden Ziehungen beeinflussen sich weniger, je größer die Anzahl *N* aller Kugeln ist. Es ist eine gute Übung, die Rechnungen selbst durchzugehen.

Satz (Erwartungswert und Varianz von Summen):

Sind X und Y zwei Zufallsvariable, dann gilt

- a) E(X+Y) = E(X) + E(Y)
- b) V(X+Y) = V(X) + V(Y) + 2 Cov(X, Y)
- c) Falls X und Y stochastisch unabhängig sind, gilt weiters:

$$V(X+Y) = V(X) + V(Y)$$

Bew. wird nur für gemeinsam diskrete verteilte Zufallsvariable (X, Y) erbracht:

a)

$$E(X+Y) = \sum_{j} \sum_{k} (x_{j} + y_{k}) \cdot P(X = x_{j}, Y = y_{k})$$

$$= \sum_{j} \sum_{k} \{x_{j} \cdot P(X = x_{j}, Y = y_{k}) + y_{k} \cdot P(X = x_{j}, Y = y_{k})\}$$

$$= \sum_{j} x_{j} \sum_{k} P(X = x_{j}, Y = y_{k}) + \sum_{k} y_{k} \sum_{j} P(X = x_{j}, Y = y_{k})$$

$$= \sum_{j} x_{j} P(X = x_{j}) + \sum_{k} y_{k} P(Y = y_{k})$$

$$= E(X) + E(Y) = \mu_{X} + \mu_{Y}$$

b) Berücksichtigt man die Erwartungswertgleichung in a) so hat man für die Varianz:

$$V(X+Y) = E\{(X-\mu_X + Y - \mu_Y)^2\}$$

$$= E\{(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2 \cdot (X-\mu_X) \cdot (Y-\mu_Y)\}$$

$$= E\{(X-\mu_X)^2\} + E\{(Y-\mu_Y)^2\} + 2 \cdot E\{(X-\mu_X) \cdot (Y-\mu_Y)\}$$

$$= V(X) + V(X) + 2 \cdot Cov(X, Y)$$

c) Aus der Unabhängigkeit folgt Cov(X, Y) = 0; dieser Summand fällt dann in b) weg.

Satz (Erwartungswert und Varianz von Linearkombinationen):

Sind X und Y zwei Zufallsvariable, dann gilt mit $a, b \in \mathbf{R}$

a)
$$E(aX+bY) = aE(X)+bE(Y)$$

b)
$$V(aX+bY) = a^2V(X) + b^2V(Y) + 2abCov(X, Y)$$

c) Falls X und Y stochastisch unabhängig sind, gilt weiters:

$$V(aX+bY) = a^2V(X) + b^2V(Y)$$

B 4: Seien R_A und R_B die Renditen der Aktien A und B [in %] bezogen auf ein Jahr. Ein Anleger investiert den Anteil a seines Vermögens in A, 1 - a in B. Die Gesamtrendite R_G ist daher eine Zufallsvariable $R_G = a \cdot R_A + (1 - a) \cdot R_B$

mit folgendem Erwartungswert und Varianz (die Indices geben die Zugehörigkeit zu den entsprechenden Zufallsvariablen an):

$$\mu_{G} = a \cdot \mu_{A} + (1 - a) \cdot \mu_{B},$$

$$\sigma_{G}^{2} = a^{2} \cdot \sigma_{A}^{2} + (1 - a)^{2} \cdot \sigma_{B}^{2} + 2 \cdot a \cdot (1 - a) \cdot \sigma_{A} \cdot \sigma_{B} \cdot \rho_{AB}.$$

Im Vergleich zur Varianz der Gesamtrendite bei Unabhängigkeit von A und B muss eine "Korrektur" nach oben erfolgen, wenn die Korrelation positiv ist; ist hingegen die Korrelation negativ, so erfolgt eine "Korrektur" nach unten.

- **B 4f:** Anlage B sei festverzinslich zu 5%; Anlage A habe eine erwartete Rendite von 15% bei einer Standardabweichung von 25%.
- a) Wie groß muss der Anteil *a* in Aktien mindestens sein, wenn die erwartete Gesamtrendite mindestens 10% betragen soll? Wie groß ist dabei die Standardabweichung der Gesamtrendite?
- b) Wie groß darf der in Aktien angelegte Anteil *a* höchstens sein, wenn die Standardabweichung der Gesamtrendite 5% nicht überschreiten darf?

- a) $\mu_{\rm G}=a\cdot\mu_{\rm A}+(1-a)\cdot\mu_{\rm B}\geq 10$ Dies führt auf: $a\geq 0,5$. Mindestens 50% sind demnach in Aktien anzulegen. Für die Standardabweichung beachte man zunächst: $\rho_{\rm AB}=0$, $\sigma_{\rm B}=0$, weil die festverzinsliche Rendite konstant ist. Damit folgert man weiter: $\sigma_{\rm G}=a\cdot\sigma_{\rm A}=a\cdot25\geq0,5\cdot25=12,5$.
- b) Nun muss gelten: $\sigma_G = a \cdot \sigma_A \le 5$; das führt auf $a \le 0,2$ höchstens 20% dürfen in Aktien investiert werden.



3.1 Gemeinsame Verteilung von Zufallsvariablen





Gemeinsame Verteilung von zwei Zufallsvariablen

-Gemeinsame Verteilung von n Zufallsvariablen

Summen von unabhängigen Zufallsvariablen

Def. Für den Zufallsvektor $(X_1, X_2, ..., X_n)$ heißt

$$F: \mathbf{R}^n \rightarrow [0,1]$$

$$F(x_1, x_2, ..., x_n) = P(X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n)$$

gemeinsame Verteilungsfunktion.

Bem.: Die Randverteilung einer Zufallsvariablen X_i erhält man durch Grenzwertbildung wie bei n=2Zufallsvariablen: $F_i(x_i) := F_{X_i}(x_i) = F(\infty, ..., \infty, x_i, \infty, ..., \infty)$.

Bem.: Zur übersichtlichen Notation verwendet man Vektoren und Matrizen.

Zufallsvariable X_i:

Zufallsvektor
$$X = (X_1, X_2, ..., X_n)$$

Erwartungswerte $\mu_i = E(X_i)$:

Erwartungswertvektor $\mu = (\mu_1, \mu_2, ..., \mu_n)$

Varianzen und Kovarianzen $\sigma_i^2 = V(X_i)$

bzw.
$$\sigma_{ij} = Cov(X_i, X_j)$$
 für $i \neq j$

Kovarianzmatrix:
$$Cov(\underline{X}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{2}^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{n}^2 \end{bmatrix}$$

Korrelationen
$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \cdot \sigma_j}$$
 für $i \neq j$ bzw. $\rho_{ij} = 1$

Korrelationsmatrix:

$$Corr(\underline{X}) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{pmatrix}$$

Bem.: Wie bei n = 2 Zufallsvariablen definiert man für Zufallsvektoren \underline{X} gemeinsam diskret bzw. gemeinsam stetig verteilt. Eine Darstellung allerdings wie im Fall n = 2 ist nicht mehr möglich (Tabelle der diskreten Wahrscheinlichkeitsfunktion bzw. Graph der gemeinsamen Dichtefunktion als Funktionengebirge).

Der Zufallsvektor \underline{X} heißt stochastisch unabhängig, wenn für alle $x_i \in \mathbf{R}$ gilt:

$$F(x_1, \dots, x_i, \dots, x_n) = F_1(x_1) \cdot F_2(x_i) \cdot \dots \cdot F_n(x_n)$$

Satz: Ist der Zufallsvektor stochastisch unabhängig, so gilt für gemeinsam stetige Verteilungen eine Produktformel analog zum Fall zweier Zufallsvariablen:

$$f(x_1, ..., x_i, ..., x_n) = f_1(x_1) \cdot f_2(x_i) \cdot ... \cdot f_n(x_n)$$

Bem.: Für gemeinsam diskrete Zufallsvektoren gilt eine analoge Produktformel.

Satz (Folgerungen aus der Unabhängigkeit):

Seien die Zufallsvariablen $(X_1, X_2, ..., X_n)$ stochastisch unabhängig. Dann gilt:

- a) Die Zufallsvariablen sind auch paarweise unabhängig
- b) Die Zufallsvariablen sind paarweise unkorreliert, d. h., $\rho_{ii} = 0$ für $i \neq j$.
- c) Die transformierten Zufallsvariablen $(g_1(X_1), g_2(X_2), ..., g_n(X_n))$ sind ebenfalls unabhängig.

Satz (Erwartungswert und Varianz einer Summe bzw. einer Linearkombination):

Seien $X_1, X_2, ..., X_n$ n Zufallsvariable und a_i reelle Zahlen, so gilt:

$$E\left\{\sum_{i=1}^{n} a_{i} X_{i}\right\} = \sum_{i=1}^{n} a_{i} E(X_{i})$$

$$V\{\sum_{i=1}^{n} a_i X_i\} = \sum_{i=1}^{n} a_i^2 V(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \ j \neq i}}^{n} a_i a_j Cov(X_i, X_j)$$

Bem.: In Vektorschreibweise liest man den Satz mit $\underline{a} = (a_1, a_2, ..., a_n)$ und $\underline{a'}$ dem transponierten, in eine Spalte umgeschriebenen Vektor so:

$$E(\underline{X}, \underline{a'}) = \mu \ \underline{a'}$$
 $V(\underline{X}, \underline{a'}) = \underline{a} \ Cov(\underline{X}) \ \underline{a'}$

Bem.: Die letzte Doppelsumme erstreckt sich über die Elemente einer $n \times n$ -Matrix, ausgenommen die Diagonale, wo i = j gilt. Man kann sie auch so notieren:

$$2 \cdot \sum_{i=1}^{n} \sum_{j=i+1}^{n} \cdots$$

weil die Matrixelemente wegen $Cov(X_i, X_j) = Cov(X_j, X_i)$ symmetrisch zur Diagonale sind.

Satz : Sind $X_1, X_2, ..., X_n$ paarweise unkorrelierte Zufallsvariable und a_i reelle Zahlen, so gilt:

$$V\{\sum_{i=1}^{n} a_i X_i\} = \sum_{i=1}^{n} a_i^2 V(X_i)$$

Bem.: Falls die Zufallsvariablen stochastisch unabhängig sind, sind sie auch paarweise unkorreliert. Somit gilt die Folgerung der "Additivität" auch dafür.

B 6: Ein Anleger besitzt 100 A-Aktien zum Kurs von 50€, 150 B-Aktien zum Kurs von 40€ sowie 50 C-Aktien zum Kurs von 20€. Die jeweiligen einjährigen Renditen [in %] seien mit R_A, R_B und R_C bezeichnet. Für Erwartungswerte, Standardabweichungen und Korrelationen der Renditen gelte:

$$\mu_{\rm A} = 10, \, \mu_{\rm B} = 5, \, \mu_{\rm C} = 5, \, \, \sigma_{\rm A} = 25, \, \sigma_{\rm B} = 25, \, \sigma_{\rm C} = 20 \,, \qquad \rho_{\rm AB} = 0.5, \, \rho_{\rm AC} = 0.55, \, \rho_{\rm BC} = 0 \,.$$

Man berechne den jetzigen Wert P_0 des Portfolios sowie Erwartungswert und Standardabweichung des Portfolios P_1 nach einem Jahr.

Der jetzige Wert ist einfach als Summe der Aktienstücke mal Kurs zu berechnen: $P_0 = 12~000$. Der Wert nach einem Jahr P_1 ist eine Zufallsvariable:

$$P_{1} \; = \; 100 \cdot 50 \cdot (1 + \tfrac{R_{A}}{100}) + 150 \cdot 40 \cdot (1 + \tfrac{R_{B}}{100}) + 50 \cdot 20 \cdot (1 + \tfrac{R_{C}}{100})$$

Erwartungswert und Standardabweichung dieser Zufallsvariablen auszurechnen stellt eine einfache Anwendung des Satzes über Linearkombinationen dar. Das Resultat lautet:

$$E(P_1) = 12850$$
; $\sigma(P_1) = 2444.9$.



3.1 Gemeinsame Verteilung von Zufallsvariablen





- Gemeinsame Verteilung von zwei Zufallsvariablen
- Gemeinsame Verteilung von n Zufallsvariablen
- Summen von unabhängigen Zufallsvariablen

Erwartungswert und Varianz einer Summe von Zufallsvariablen lassen sich aus diesen die entsprechenden Momente der Summanden herleiten. Für andere Parameter, welche die Dichtefunktion oder die Verteilungsfunktion kennzeichnen, ist der Zusammenhang komplizierter. In Spezialfällen geht auch das.

Def.: $X_1, X_2, ..., X_n$ seien n unabhängige Zufallsvariable, alle mit einer Verteilungsfunktion vom selben Typ: $X_i \sim F(\lambda)$. Hat die Zufallsvariable $S := \sum_{i=1}^n X_i$ denselben Verteilungstyp, d. h., es gilt $S \sim F(\lambda^*)$, so nennt man die Familie der Verteilungen vom Typ $F(\lambda)$ reproduktiv.

Satz: $X_1, X_2, ..., X_k$ seien k unabhängige, binomialverteilte Zufallsvariable mit identischem Parameter π , d. h. $X_i \sim B(n_i, \pi)$, so ist die Zufallsvariable $S := \sum_{i=1}^k X_i$ auch binomialverteilt, und zwar gilt $S := \sum_{i=1}^{k} X_i \sim B(\sum_{i=1}^{k} n_i, \pi).$

Bew. Skizze: Jedes X_i lässt sich als Zahl der Erfolge in einer Bernoulli-Versuchsreihe der Länge n_i auffassen; da die einzelnen X_i stochastisch unabhängig sind, kann man die Versuchsreihen einfach aneinander reihen und die Gesamtzahl aller Erfolge mit $S = \sum_{i=1}^{K} X_i$ zählen. Die Erfolgswahrscheinlichkeit ist über alle $n = \sum_{i=1}^{\kappa} n_i$ Versuche hinweg immer gleich π . Daher ist S binomialverteilt mit entsprechenden Parametern.

Satz: $X_1, X_2, ..., X_k$ seien k unabhängige, Poisson-verteilte Zufallsvariable: $X_i \sim \text{Po}(\mu_i)$, so ist die Zufallsvariable $S := \sum_{i=1}^{k} X_i$ auch Poisson-verteilt, und zwar gilt $S := \sum_{i=1}^{k} X_i \sim Po(\sum_{i=1}^{k} \mu_i)$

Eine Baufirma betreibt zwei kleine Baustellen und eine Großhaustelle. Die Anzahl der Arbeiter, die B 9: innerhalb eines Jahres einen schweren Unfall auf einer Baustelle erleiden, sei Poisson-verteilt und zwar mit $\mu = 2$ für die kleinen und mit $\mu = 10$ für die große Baustelle.

Die Anzahl der Arbeiter, die insgesamt einen Unfall haben, ist dann Poisson-verteilt mit $\mu = 10 + 2 + 2$.

Satz: $X_1, X_2, ..., X_k$ seien k unabhängige, normalverteilte Zufallsvariable: $X_i \sim N(\mu_i, \sigma_i^2)$, so ist die Zufallsvariable $S := \sum_{i=1}^{k} X_i$ auch normalverteilt, und zwar gilt $S := \sum_{i=1}^{k} X_i \sim N(\sum_{i=1}^{k} \mu_i, \sum_{i=1}^{k} \sigma_i^2)$

B 10: Ein Portfolio Y aus A-, B- und C-Aktien mit den Anteilen 40, 50 bzw. 10% ist normalverteilt, wenn die Renditen der einzelnen Aktien stochastisch *unabhängig* und normalverteilt sind. Für die folgenden Erwartungswerte und Standardabweichungen

$$\mu_{\rm A} = 10, \, \mu_{\rm B} = 12, \, \mu_{\rm C} = 8, \, \sigma_{\rm A} = 20, \, \sigma_{\rm B} = 25, \, \sigma_{\rm C} = 10,$$

ergibt sich für die Rendite R_G des Portfolios eine Normalverteilung mit den Parametern

$$\mu_{\rm G} = 0.4 \cdot 10 + 0.5 \cdot 12 + 0.1 \cdot 8 = 10.8 \quad \text{und} \quad \sigma_{\rm G} = 0.4^2 \cdot 20^2 + 0.5^2 \cdot 25^2 + 0.1^2 \cdot 10^2 = 221.25 \; .$$

Der Wert des Portfolios für ein Anfangskapital Y_0 zum Zeitpunkt t=1 ist eine Zufallsvariable Y_1 mit $Y_1=Y_0\cdot (1+\frac{R_G}{100})$.

Bei der gegebenen Modellierung ist auch diese Zufallsvariable normalverteilt, und zwar mit

$$\mu_{Y_1} = Y_0 \cdot (1 + \frac{\mu_G}{100}) \text{ und } \sigma_{Y_1}^{\ \ 2} = Y_0^{\ 2} \cdot (\frac{\sigma_G}{100})^2 \ .$$



3 Gemeinsame Verteilung und Grenzwertsätze

MANFRED BOROVCNIK

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



Inhaltsverzeichnis

- 3.1 Gemeinsame Verteilung von Zufallsvariablen
- 3.2 Grenzwertsätze
- 3.3 Ergänzungen

In Kapitel 2 wurden einzelne Zufallsvariable im Hinblick auf ihre Verteilung untersucht. In vielen Anwendungen untersucht man dagegen das gemeinsame Auftreten mehrerer Merkmale und ihre wechselweisen Abhängigkeiten. In der schließenden Statistik hat man es regelmäßig mit mehreren Zufallsvariablen zu tun. Dabei werden aus den beobachteten Realisationen wiederholter, gleichartiger Zufallsvorgänge Schlussfolgerungen gezogen. Speziell werden Summen von Zufallsvariablen untersucht, weil diese für die Verfahren in der schließenden Statistik besonders wichtig sind. Zuerst werden Erwartungswert und Varianz von Summen behandelt. Dann wird die Konvergenz von Summen analysiert (Schwaches Gesetz der Großen Zahlen) und die Verteilung näherungsweise bestimmt (über den Zentralen Grenzverteilungssatz). Ergänzt wird das Kapitel durch Einführungen in die multivariate Normalverteilung, den Poisson-Prozess und die Monte-Carlo-Simulation.

3.2 Grenzwertsätze

- Schwaches Gesetz der großen Zahlen
- Wahrscheinlichkeit und relative Häufigkeit Konvergenz
- **Zentraler Grenzwertsatz**

Für die schließende Statistik ist die Beurteilung von Mittelwerten und relativen Häufigkeiten aus Daten ein wesentlicher Punkt. Beides sind Werte von Zufallsvariablen, die sich im Wesentlichen als Summen von Zufallsvariablen darstellen lassen. Unter bestimmten Gesichtspunkten "konvergieren" diese Summen. Das wird hier untersucht.

3.2 Grenzwertsätze



Schwaches Gesetz der großen Zahlen

3

Wahrscheinlichkeit und relative Häufigkeit – Konvergenz

Zentraler Grenzwertsatz

Aus der Tschebyscheffschen Ungleichung heraus wird gezeigt, dass Mittelwerte von Zufallsvariablen "konvergieren".

Def.: Gegeben ist die Zufallsvariable X und ihre Verteilungsfunktion: $X \sim F$:

Unter Stichprobe von X verstehen wir eine Folge von Zufallsvariablen $X_1, X_2, ..., X_n$, wenn diese i) alle dieselbe Verteilung wie X haben, d. h., $X_i \sim F$ und ii) stochastisch unabhängig sind.

Von der Stichprobe abgeleitete Zufallsvariable heißen ganz allgemein Stichprobenfunktionen oder Statistiken; spezielle Statistiken sind:

arithmetisches Mittel: $\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$

 $Y_n := \sum_{i=1}^n X_i$ Variablensumme:

Bem.: Man nennt n den Umfang der Stichprobe. Man spricht auch von unabhängigen Wiederholungen derselben Zufallsvariablen. Wir betrachten Aussagen über arithmetisches Mittel bzw. die Variablensumme in Abhängigkeit von n; insbesondere auch den Fall, wie sich Erwartungswert oder Verteilung verhalten, wenn der Grenzübergang $n \to \infty$ betrachtet wird. D. h., wir betrachten die Folge von Zufallsvariablen $X_1, X_2, ...$ wobei je endliche viele stochastisch unabhängig sind und alle dieselbe Verteilung haben. Im Folgenden wird eine neue Art von Konvergenz für die Zufallsvariablen eingeführt: es sind ja nicht gewöhnliche reellen Zahlen a_1, a_2, \dots , die auf Konvergenz hin untersucht werden. Schließlich wird die Verteilung des arithmetischen Mittels untersucht; genauer wird aus einem Grenzwertprozess heraus die Verteilung für endliche Stichprobenumfänge *n* approximiert.

Satz (Erwartungswert und Varianz des arithmetischen Mittels):

 $X \sim F$ mit endlichem Erwartungswert μ und Varianz σ^2 .

 $X_1, X_2, ..., X_n$ eine Stichprobe vom Umfang n.

 $E\{\overline{X}_n\} = \mu \text{ und } V\{\overline{X}_n\} := \frac{\sigma^2}{n}.$

Bew.: Das arithmetische Mittel ist eine Linearkombination der Zufallsvariablen X_i mit den Gewichten $a_i = \frac{1}{n}$. Die Zufallsvariablen sind unabhängig und daher auch paarweise unkorreliert (Kovarianzen $\sigma_{ij} = 0$ für $i \neq j$). Es gilt damit nicht nur für den Mittelwert sondern auch für die Varianz die vereinfachte Addivität:

$$E\{\overline{X}_n\} = \frac{1}{n}(\mu + \mu + ... + \mu) = \mu \text{ und } V\{\overline{X}_n\} := \frac{1}{n^2} \cdot (\sigma^2 + \sigma^2 + ... + \sigma^2) = \frac{\sigma^2}{n}$$

Satz (Schwaches Gesetz der Großen Zahlen):

Sind $X_1, X_2, ...$ unabhängige Wiederholungen der Zufallsvariablen X; hat ferner X endlichen Erwartungswert μ und Varianz σ^2 , ist $\varepsilon > 0$, so gilt:

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| \ge \varepsilon\right) = 0$$

Bem.: Äquivalent zum Satz ist (über Komplementärereignisse) die Aussage:

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1$$

Bem.: Das Schwache Gesetz der großen Zahlen besagt, dass Abweichungen des arithmetischen Mittels vom Erwartungswert μ von mindestens ε eine Wahrscheinlichkeit haben, die für $n \to \infty$ gegen 0 strebt. Anschaulich bedeutet der Satz, dass sich die Verteilung von \overline{X}_n mit wachsendem n auf den Punkt μ zusammenzieht.

Bem.: Es gilt ein noch stärkeres Gesetz, das Starke Gesetz der großen Zahlen:

$$P\left(\left\{\omega \left| \lim_{n \to \infty} \overline{X}_n(\omega) = \mu\right.\right\}\right) = 1.$$

Danach führen fast alle ω zu arithmetischen Mittelwerten, die gegen μ konvergieren. Die Menge der ω , für die Konvergenz zutrifft, hat Wahrscheinlichkeit 1. Dies ist allerdings viel schwerer zu beweisen.

Bew.: Die Tschebyscheff-Ungleichung, angewendet auf das arithmetische Mittel, lautet:

$$P(|\overline{X}_n - E(\overline{X}_n)| \ge \varepsilon) \le \frac{V(\overline{X}_n)}{\varepsilon^2}$$

Beachtet man: $E\{\overline{X}_n\} = \mu$ und $V\{\overline{X}_n\} := \frac{\sigma^2}{n}$, so hat man:

$$P\left(\left|\overline{X}_n - \mu\right| \ge \varepsilon\right) \le \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n \cdot \varepsilon^2};$$

der Grenzübergang $n \to \infty$ ergibt die Aussage des Satzes.

Def.: Eine Folge von Zufallsvariablen Y_1, Y_2, \dots konvergiert stochastisch gegen die Zahl θ , wenn $\lim_{n \to \infty} P\left(\left|Y_n - \theta\right| < \varepsilon\right) = 1 \text{ für jedes } \varepsilon > 0 \text{ . Man schreibt: } p - \lim_{n \to \infty} Y_n = \theta$

Bem.: Hier ist nicht gefordert, dass die Y_i dieselbe Verteilung haben oder gar unabhängige Wiederholungen derselben Zufallsvariablen sind. Man spricht auch von *Konvergenz* der Zufallsvariablen *nach Wahrscheinlichkeit*. Nicht die Werte der Zufallsvariablen konvergieren gegen θ , sondern die Verteilungen ziehen sich um θ zusammen. Die Folge der arithmetischen Mittel konvergieren stochastisch gegen den Erwartungswert: $p-\lim \overline{X}_n=\mu$.

↑

3.2 Grenzwertsätze





- Schwaches Gesetz der großen Zahlen
- Wahrscheinlichkeit und relative Häufigkeit Konvergenz
- Zentraler Grenzwertsatz

Das Schwache Gesetz der großen Zahlen wird auf relative Häufigkeiten angewendet, um die stochastische Konvergenz der relativen Häufigkeiten zur Wahrscheinlichkeit abzuleiten. Damit wird die intuitive Verwendung von relativen Häufigkeiten als Schätzwerte für unbekannte Wahrscheinlichkeiten gerechtfertigt.

Def.: Wir betrachten eine Bernoulli-Versuchsreihe mit *n* unabhängigen Versuchen. Jeder Versuch führt zum Ereignis A mit Wahrscheinlichkeit π . Dann ist die relative Häufigkeit des Eintretens von A eine Zufallsvariable h_n mit

$$h_n := \overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$
 und $X_i := \begin{cases} 1 & \text{falls } A \text{ beim } i - \text{ten Versuch eintritt} \\ 0 & \text{falls } A \text{ beim } i - \text{ten Versuch ausbleibt} \end{cases}$

Satz: Die relative Häufigkeit h_n in einer Bernoulli-Versuchsreihe konvergiert stochastisch gegen die Wahrscheinlichkeit π : $p - \lim_{n \to \infty} h_n = \pi$.

Bew.: Schon die Definition der relativen Häufigkeit legt diese als speziellen Mittelwert fest. Die X_i sind unabhängige Wiederholungen von $X \sim B(1, \pi)$ mit Erwartungswert π und endlicher Varianz $\pi \cdot (1 - \pi)$. Für arithmetische Mittel hatten wir die stochastische Konvergenz schon bewiesen.

Bem: Man beachte, dass die relativen Häufigkeiten keine Zahlen sondern Zufallsvariable sind. Die Konvergenz ist eine Aussage über deren Verteilungen. Die Verteilungen ziehen sich um die Wahrscheinlichkeit π zusammen. In der Anwendung muss man sicher stellen, dass die Bedingungen einer Bernoulli-Versuchsreihe einhalten, was nicht immer leicht ist. Der Satz liefert eine theoretische Aussage und erzwingt nichts über das Verhalten relativer Häufigkeiten in der Wirklichkeit.

Def.: Für die unabhängigen Wiederholungen $X_1, X_2, ...$ der Zufallsvariablen X mit Verteilungsfunktion F liefert eine Stichprobe vom Umfang n das Ergebnis x_1, x_2, \dots, x_n .

Man definiert die empirische Verteilungsfunktion durch:

$$F_n(x) := \frac{1}{n} \cdot \left(\text{ Anzahl der } x_i \le x \right) \qquad \text{für } x \in \mathbf{R}$$

Bem.: Wahrscheinlichkeiten und relative Häufigkeiten stehen in engem Zusammenhang. Für unabhängige Wiederholungen desselben Experiments konvergieren relative Häufigkeiten stochastisch gegen die Wahrscheinlichkeit.

Empirische Verteilungsfunktion $F_n(x)$: Theoretische Verteilungsfunktion F(x):

Relative Häufigkeiten von Daten $(x_i \le x)$ Wahrscheinlichkeit von $(X \le x)$

Es ist zu erwarten, dass die empirische Verteilungsfunktion gegen die theoretische konvergiert, wir müssen nur noch $F_n(x)$ mit einer Bernoulli-Versuchsreihe in Verbindung bringen.

Satz:

a)
$$Y_i := \mathbf{1}_{(X_i \le x)} := \begin{cases} 1 & \text{falls } X_i \le x \\ 0 & \text{falls } X_i > x \end{cases}$$
, für ein festes x , sind unabhängig $B(1, \pi)$ -verteilt mit $\pi = F(x)$.

- b) $F_n(x) := \frac{1}{n} \cdot \sum_{i=1}^n Y_i$ ist das arithmetische Mittel der Y_i .
- c) $p \lim_{n \to \infty} F_n(x) = \pi = F(x)$

Bew.: Der Beweis ist trivial.

Bem.: Es gelten noch stärkere Sätze (von Glivenko-Cantelli) – die sind aber schwieriger zu beweisen:

- i) $p \lim_{n \to \infty} \left(\sup_{x \in \mathbf{R}} |F_n(x) F(x)| \right) = 0$, d. h., sogar das Supremum der Abstände zwischen empirischer
 - und theoretischer Verteilungsfunktion geht stochastisch gegen 0. Die Konvergenz ist gleichmäßig.
- ii) Zur schwachen gleichmäßigen Konvergenz gibt es auch hier wieder eine "starke" Variante, nämlich, dass die Menge der ω , auf der gleichmäßige Konvergenz stattfindet, Wahrscheinlichkeit 1 hat.

Bem.: Ist die Zahl der Daten x_i groß genug, so kann man die theoretische Verteilungsfunktion gleichmäßig über ihren ganzen Verlauf durch die empirische Verteilungsfunktion approximieren. Ein Ergebnis, das für die Praxis gar nicht hoch genug bewertet werden kann.



3.2 Grenzwertsätze



Schwaches Gesetz der großen Zahlen

Wahrscheinlichkeit und relative Häufigkeit – Konvergenz

Zentraler Grenzwertsatz

Die Verteilung der Summe und des arithmetischen Mittels der unabhängigen Wiederholung einer Zufallsvariablen kann näherungsweise durch die Normalverteilung beschrieben werden. Dieser zentrale Satz wird erläutert (nicht bewiesen); seine Konsequenzen werden dargestellt.

Def.: Sei X eine Zufallsvariable mit Erwartungswert μ und endlicher Varianz σ^2 . Unter der *standardisierten Zufallsvariablen* versteht man die Transformierte:

$$U := \frac{X - \mu}{\sigma}$$

Def.: Sind die Zufallsvariablen $X_1, X_2, ...$ unabhängige Wiederholungen der Zufallsvariablen X, so spricht man auch von *iid-Stichproben* von X: *iid* meint *independent identically distributed*.

Bem.: Falls die Zufallsvariablen $X_1, X_2, ...$ alle normalverteilt sind – jetzt nicht notwendigerweise mit denselben Parametern – so sind Summen, arithmetisches Mittel und standardisierte Summe immer auch normalverteilt, denn Normalverteilungen sind reproduktiv. Wir notieren die entsprechenden Zufallsvariablen für den im Folgenden wichtigen Fall der unabhängigen Wiederholung einer Zufallsvariablen:

Summe: $Y_n := \sum_{i=1}^n X_i$

Arithmetisches Mittel: $\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$

Standardisierte Summe: $U_n := \frac{\sum_{i=1}^n X_i - n \mu}{\sqrt{n\sigma^2}} = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}}$

Satz (Zentraler Grenzwertsatz):

Die Zufallsvariablen X_1, X_2, \ldots sind unabhängige Wiederholungen der Zufallsvariablen X mit Erwartungswert μ und endlicher Varianz σ^2 .

Für die standardisierte Summe U_n von $X_1, X_2, ..., X_n$ gilt:

 $\lim_{n \to \infty} P(U_n \le u) = \Phi(u) \qquad \text{für alle } u \in \mathbf{R}$

Bem.: Φ bezeichnet die Verteilungsfunktion der N(0,1). Nach dem zentralen Grenzwertsatz konvergieren beliebige, standardisierte Summen gegen die Normalverteilung, egal welcher Ausgangsverteilung, nur endliche Varianz vorausgesetzt. Der Satz gilt unter sehr allgemeinen Bedingungen: Man kann die identische Verteilung der Summanden weglassen; man kann leichte Formen der Abhängigkeit zwischen den Summanden erlauben (Ljapunov, Lindeberg).

Folgerung: Aus der Grenzwertaussage gewinnt man eine Approximation für endliche *n*:

$$U_n \sim N(0,1) \qquad \sum_{i=1}^n X_i \sim N(n \,\mu, n \,\sigma^2) \qquad \overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$$

Bem.: Es gibt eine Reihe von Untersuchungen über die Konvergenzgeschwindigkeit. Wenn die Zufallsvariable X nicht allzu "pathologisch" ist, so reichen eigentlich n = 30 Wiederholungen, um ausreichende Approximationsgüte zu bekommen; wenn die Verteilung sehr a-symmetrisch ist oder sehr viel Wahrscheinlichkeit in den Ausläufern hat, muss man aber 100 und mehr Wiederholungen in Kauf nehmen.

B 13: Der Rechnungsbetrag eines Gastes in € werde durch die Zufallsvariable X_i modelliert. Die Beträge seien untereinander unabhängig und identisch verteilt. Als Folge des zentralen Grenzwertsatzes gilt: Der Abendumsatz ist als Summe der Beträge über alle Gäste näherungsweise normalverteilt.

Haben die Beträge der einzelnen Gäste einen Erwartungswert von $\mu = 30$ und eine Standardabweichung $\sigma = 10$, so hat man für den Abendumsatz Y bei n = 60 Gästen:

$$\mu_{\rm Y} = 60 \cdot 30 = 1800 \text{ und } \sigma_{\rm Y} = \sqrt{60} \cdot 10 \approx 77.5.$$

Die Wahrscheinlichkeit, dass der Umsatz größer als 2000 beträgt, berechnet man wegen

$$P(Y \ge 2000) = P(\frac{Y - \mu_Y}{\sigma_Y} \ge \frac{2000 - 1800}{77,5}) = 1 - \Phi(2,58) = 1 - 0,9951 = 0,0049$$

B 14: Ist R_i die Tagesrendite einer Aktie, so ergibt sich der Kurs K_t nach t Tagen in Abhängigkeit von Anfangswert K_0 und den Tagesrenditen zu:

$$K_t = K_0 \cdot \prod_{i=1}^t (1 + \frac{R_i}{100}).$$

Durch Logarithmieren erhält man:

$$\ln K_t = \ln K_0 + \sum_{i=1}^t \ln \left(1 + \frac{R_i}{100}\right).$$

Modelliert man die Tagesrenditen R_i durch unabhängige Zufallsvariable, so ergibt sich die Entwicklung der Logarithmen der Tageskurse als Summe von (auch unabhängigen) Zufallsvariablen $\ln\left(1+\frac{R_i}{100}\right)$, welche nach dem zentralen Grenzwertsatz annähernd normalverteilt ist. Die Tageskurse sind daher annähernd logarithmisch normalverteilt.

Mit den Bezeichnungen E $\{\ln(1+\frac{R_i}{100})\}=\mu$ und $V\{\ln(1+\frac{R_i}{100})\}=\sigma^2$ hat man

$$E\{\ln K_t\} = \ln K_0 + t \cdot \mu \text{ und } V\{\ln K_t\} = t \cdot \sigma^2.$$

Schließlich ergibt sich für die Entwicklung des Tageskurses (nach langer Zeit t):

$$K_t \stackrel{approx}{\sim} LN(\ln K_0 + t \cdot \mu, \ t \cdot \sigma^2)$$

Satz: Ist
$$X \sim B(n, \pi)$$
, so gilt $X \sim N(n \cdot \pi, n \cdot \pi \cdot (1 - \pi))$, d. h.,
$$P(X \le k) \approx \Phi\left(\frac{k - n \cdot \pi}{\sqrt{n \cdot \pi \cdot (1 - \pi)}}\right) = \Phi\left(\widetilde{k}\right).$$

Bem: Man notiert zur Abkürzung die Standardisierung durch eine Tilde:

$$\widetilde{X} := \frac{X - \mu}{\sigma}$$
 und $\widetilde{k} := \frac{k - \mu}{\sigma}$,

wobei μ und σ^2 Erwartungswert und Varianz der Zufallsvariablen X bezeichnen.

Bew.: Eine binomialverteilte Zufallsvariable X kann als Summe der Indikatoren X_i in einer Bernoulli-Versuchsreihe dargestellt werden:

$$X = \sum_{i=1}^{n} X_{i} \text{ mit } X_{i} \overset{unabhängig}{\sim} B(1, \pi) \text{ und } E(X_{i}) = \pi, \ V(X_{i}) = \pi \cdot (1 - \pi).$$

Nach den Zentralen Grenzwertsatz ist daher X annähernd normalverteilt mit

$$\mu_{\rm X} = n \cdot \pi$$
 und $\sigma_{\rm X}^2 = n \cdot \pi \cdot (1 - \pi)$.

Bem.: Die Binomialverteilung ist diskret, die approximierende Normalverteilung stetig. Zwischen k und k+1 liegt bei der Binomialverteilung keine Wahrscheinlichkeit, wohl aber bei der Normalverteilung. Um den daraus resultierenden systematischen Fehler zu verringern, nimmt man eine Stetigkeitskorrektur vor (man beachte die \sim -Schreibweise für standardisierte Werte):

$$P(X \le k) \approx \Phi(k+0.5)$$

Für Intervalle erhält man die Stetigkeitskorrektur so:

$$P(k_1 \le X \le k_2) \approx \Phi(k_2 + 0.5) - \Phi(k_1 - 0.5),$$

denn:

$$P(X \ge k_1) = 1 - P(X \le k_1 - 1) \approx 1 - \Phi(k_1 - 1 + 0.5)$$
.

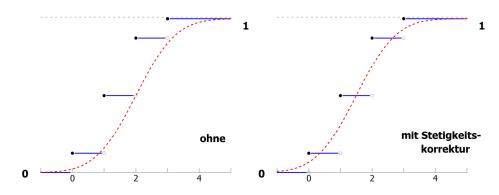


Abb. 7: Approximation der Verteilungsfunktion einer Binomialverteilung durch die entsprechende Normalverteilung – ohne und mit Stetigkeitskorrektur

B 15: Bei der Planung der Wartung von Getränkeautomaten will man den Arbeitsaufwand im Vorhinein abschätzen. Hat man 400 Automaten zu versorgen und modelliert man den Ausfall der einzelnen Automaten durch unabhängige Binomialverteilungen mit n=1 und $\pi=0.05$ (für die ganze Woche), so kann man beantworten, wie groß die Wahrscheinlichkeit ist, dass in einer Woche a) mindestens 15 Automaten ausfallen, b) nicht mehr als 30 Automaten ausfallen.

Die Automaten formieren zusammen eine Bernoulli-Versuchsreihe. Für die Gesamtzahl der Ausfälle X hat man daher eine Binomialverteilung mit $n=400\,$ und $\pi=0,05\,$. Nach dem zentralen Grenzwertsatz kann man die Wahrscheinlichkeiten durch die entsprechende Normalverteilung approximieren mit:

$$\mu_{\rm X} = n \cdot \pi = 400 \cdot 0.05 = 20 \text{ und } \sigma_{\rm X}^2 = n \cdot \pi \cdot (1 - \pi) = 400 \cdot 0.05 \cdot (1 - 0.05) = 19$$

a)
$$P(X \ge 15) = 1 - P(X \le 14) \approx 1 - \Phi\left(\frac{14 + 0.5 - 20}{\sqrt{19}}\right) = 1 - \Phi\left(-1.26\right) = \Phi\left(1.26\right) = 0.8962$$

b)
$$P(X \le 30) \approx \Phi\left(\frac{30+0.5-20}{\sqrt{19}}\right) = \Phi(2.41) = 0.9920$$

Mit einer hohen Wahrscheinlichkeit von fast 0,9 wird man demnach mit mindestens 15 Wartungen rechnen müssen. Es gibt aber nur ein "Risiko" von kleiner als 0,01, dass es zu mehr als 30 Ausfällen kommen wird.

Satz: Ist
$$X \sim Po(\mu)$$
, so gilt $X \sim N(\mu, \mu)$, d. h.,

$$P(X \le k) \approx \Phi(\frac{k - \mu}{\sqrt{\mu}}) = \Phi(\widetilde{k})$$

Bew.: Die Poisson-Verteilung ist reproduktiv. Daher kann man die Verteilung von X "zerlegen" in unabhängige, identisch verteilte Summanden:

$$X = \sum_{i=1}^{n} X_i$$
; dabei ist $X \sim Po(\mu)$ und $X_i \stackrel{iid}{\sim} Po(\frac{\mu}{n})$.

Wieder sind die Voraussetzungen für den zentralen Grenzwertsatz erfüllt.

B 16: Die mittlere Anzahl der Anrufe in der Rezeption einer Hotelanlage sei 15 pro Stunde. Modelliert man die Anzahl mit einer Poisson-Verteilung, so nimmt man $\mu = 15$. Die Wahrscheinlichkeit, dass zwischen 10 und 30 Anrufe einlangen, kann man aus der approximierenden Normalverteilung (mit Stetigkeitskorrektur) so ablesen:

$$P\left(10 \le X \le 30\right) \approx \Phi\left(\frac{30 + 0.5 - 15}{\sqrt{15}}\right) - \Phi\left(\frac{10 - 0.5 - 15}{\sqrt{15}}\right) = \Phi\left(4.0021\right) - \Phi\left(-1.4201\right) = \dots = 0.9222$$

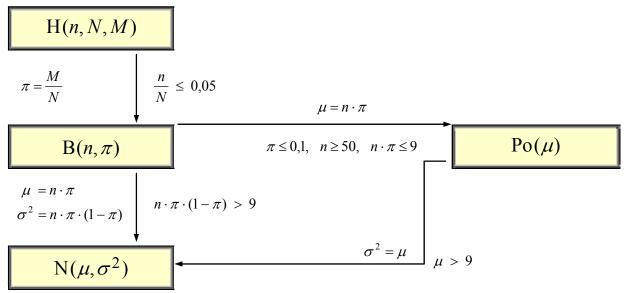


Abb. 8: Faustregeln, wann die Approximationen der Verteilungen untereinander brauchbar sind

3 Gemeinsame Verteilung und Grenzwertsätze

MANFRED BOROVCNIK

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



Inhaltsverzeichnis

- 3.1 Gemeinsame Verteilung von Zufallsvariablen
- 3.2 Grenzwertsätze
- 3.3 Ergänzungen

In Kapitel 2 wurden einzelne Zufallsvariable im Hinblick auf ihre Verteilung untersucht. In vielen Anwendungen untersucht man dagegen das gemeinsame Auftreten mehrerer Merkmale und ihre wechselweisen Abhängigkeiten. In der schließenden Statistik hat man es regelmäßig mit mehreren Zufallsvariablen zu tun. Dabei werden aus den beobachteten Realisationen wiederholter, gleichartiger Zufallsvorgänge Schlussfolgerungen gezogen. Speziell werden Summen von Zufallsvariablen untersucht, weil diese für die Verfahren in der schließenden Statistik besonders wichtig sind. Zuerst werden Erwartungswert und Varianz von Summen behandelt. Dann wird die Konvergenz von Summen behandelt (Schwaches Gesetz der Großen Zahlen) und die Verteilung näherungsweise bestimmt (über den Zentralen Grenzverteilungssatz). Ergänzt wird das Kapitel durch Einführungen in die multivariate Normalverteilung, den Poisson-Prozess und die Monte-Carlo-Simulation.

3.3 Ergänzungen

- Multivariate Normalverteilung
- **Poisson-Prozess**
- **☑** Monte-Carlo-Simulation

Für Theorie und Anwendungen wichtige Verteilungen und Modellierungen werden ergänzt. Auf die Problematik der Erzeugung von Zufallszahlen wird eingegangen.

3.3 Ergänzungen





Multivariate Normalverteilung

Poisson-Prozess

Monte-Carlo-Simulation

Die multivariate Normalverteilung zählt zu den wichtigsten Verteilungen, sowohl um mathematische Modelle zu entwickeln als auch um Anwendungen zu modellieren.

Def.: Seien $X_1, X_2, ..., X_n$ $n \ge 2$ Zufallsvariable. Man nennt sie gemeinsam multivariat normalverteilt, wenn jede Linearkombination $Y := \sum_{i=1}^{n} a_i X_i$ eine (gewöhnliche) Normalverteilung besitzt, unabhängig von der Wahl der reellen Zahlen a_i .

Satz: Sind $X_1, X_2, ..., X_n$ gemeinsam multivariat normalverteilt, so sind alle Randverteilungen gewöhnliche Normalverteilungen.

Bew.: Man wählt $a_i = 1$ und $a_j = 0$ für $j \neq i$; damit stellt man jedes X_i als Linearkombination aller Zufallsvariablen dar. Diese sind nach Definition normalverteilt.

Bem.: Wenn alle Randverteilungen normalverteilt sind, so muss die gemeinsame Verteilung nicht schon eine multivariate Normalverteilung sein.

Satz: Sind $X_1, X_2, ..., X_n$ multivariat normal verteilt, so ist diese Verteilung bereits eindeutig durch Erwartungswerte, Varianzen und Kovarianzen der X, bestimmt.

Bem.: Mit $\mu_i = E(X_i)$, $\sigma_i^2 = V(X_i)$ und den Kovarianzen $\sigma_{ij} = Cov(X_i, X_j) = \sigma_i \sigma_i \rho_{ij}$ schreibt man zunächst in Vektorschreibweise: $\mu := (\mu_i)$, $\Sigma := (\sigma_{ij})$ und notiert die Verteilung so: $\underline{X} \sim MN(\mu, \Sigma)$.

Satz: Sind $X_1, X_2, ..., X_n$ multivariat normal verteilt, so gilt Die X_i sind unabhängig, genau dann wenn sie paarweise unkorreliert sind, d. h., $\rho_{ij} = 0$ für $i \neq j$.

Satz: Sind $X_1, X_2, ..., X_n$ multivariat normal verteilt und ist $\Sigma := (\sigma_{ij})$ eine invertierbare Matrix, so hat \underline{X} eine Dichtefunktion, d. h., $X_1, X_2, ..., X_n$ sind gemeinsam stetig verteilt.

B 6f: Modelliert man die Renditen R_A, R_B und R_C als gemeinsam multivariat normalverteilt. Wie ist die Rendite des Portfolios verteilt? Wie groß ist die Wahrscheinlichkeit, dass die Rendite 8% übersteigt? Die Rendite des gesamten Portfolios ist ein gewichtetes Mittel der einzelnen Renditen: $R_P \ = \ \tfrac{100}{300} \cdot R_A + \tfrac{150}{300} \cdot R_B + \tfrac{50}{300} \cdot R_C \ .$

Als Linearkombination der multivariat normalverteilten Einzelrenditen ist R_P normalverteilt, und zwar mit Erwartungswert und Varianz:

$$E(R_P) = \frac{100}{300} \cdot E(R_A) + \frac{150}{300} \cdot E(R_B) + \frac{50}{300} \cdot E(R_C) = \frac{2}{6} \cdot 10 + \frac{3}{6} \cdot 5 + \frac{1}{6} \cdot 5 = \dots = 6,66$$

$$\begin{split} V(R_P) &= \left(\frac{2}{6}\right)^2 \cdot V(R_A) + \left(\frac{3}{6}\right)^2 \cdot V(R_B) + \left(\frac{1}{6}\right)^2 \cdot V(R_C) + \frac{2}{6} \cdot \frac{3}{6} \cdot \sigma_{AB} + \frac{2}{6} \cdot \frac{1}{6} \cdot \sigma_{AC} + \frac{3}{6} \cdot \frac{1}{6} \cdot \sigma_{BC} \\ &= \left(\frac{2}{6}\right)^2 \cdot 25^2 + \left(\frac{3}{6}\right)^2 \cdot 25^2 + \left(\frac{1}{6}\right)^2 \cdot 20^2 + \frac{2}{6} \cdot \frac{3}{6} \cdot 25 \cdot 25 \cdot 0, 5 + \frac{2}{6} \cdot \frac{1}{6} \cdot 25 \cdot 20 \cdot 0, 5 + \frac{3}{6} \cdot \frac{1}{6} \cdot 0 &= 302,78 \end{split}$$

Weiterhin ist

$$P(R_P \ge 8) \approx 1 - \Phi(\frac{8-6,66}{\sqrt{302,78}}) = 1 - \Phi(0,08) = ... = 0,4681$$
.

Die Rendite übertrifft 8% mit einer Wahrscheinlichkeit von ca. 47%.



3.3 Ergänzungen





Multivariate Normalverteilung



Poisson-Prozess



Monte-Carlo-Simulation

Die Poisson-Verteilung wird in einem mathematischen Modell mit einem Prozess des Entstehens von Ereignissen in Verbindung gebracht. Aus den Voraussetzungen dieses so genannten Poisson-Prozesses lernt man viel über die Eigenschaften der Poisson-Verteilung und wann man sie als Modell einsetzen kann. Wichtige Anwendungen der Poisson-Verteilung betreffen Prozesse des zufälligen Entstehens von Ereignissen in der Zeit. Man bezeichnet diese Ereignisse gelegentlich auch als Ankünfte.

B: Beispiele für Poisson-Ereignisse sind etwa:

- das Eintreten eines Kunden in ein Ladengeschäft,
- das Vorbeifahren eines Autos an einem Kontrollpunkt,
- das Eintreffen eines Telefonanrufes in einem Call-Center.

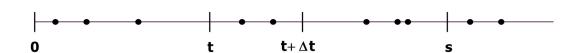


Abb. 9: Prozess des Entstehens von Ereignissen in der Zeit

Def.: Ereignisse, die den folgenden Bedingungen genügen, heißen Poisson-Ereignisse mit dem Parameter λ , der Parameter wird auch *Intensität* des Entstehens der Ereignisse genannt.

(P1) Die Wahrscheinlichkeit, dass Ereignisse im Intervall $(t, t + \Delta t)$ stattfinden, hängt nicht vom Beginn der Beobachtung t sondern nur von der Beobachtungsdauer Δt ab.

(P2) Die Wahrscheinlichkeit, genau ein Ereignis in einem kleinen Intervall der Länge Δt zu beobachten, ist in erster Näherung proportional zur Beobachtungsdauer Δt :

$$P(X_{\Delta t} = 1) \approx \lambda \cdot \Delta t$$

(P3) Zwei oder gar mehrere Ereignisse in einem kleinen Intervall der Länge Δt zu beobachten, ist in erster Näherung im Vergleich zu Δt vernachlässigbar, sofern Δt klein ist:

$$P(X_{\Delta t} \ge 2) \approx 0$$

(P4) Zufallsvariable, die in disjunkten Beobachtungsintervallen die Ereignisse zählen, sind stochastisch unabhängig.

Def.: Die Zufallsvariable $X_t, t \ge 0$ zähle die Anzahl der Ankünfte von Poisson-Ereignissen im Zeitraum [0, t]. Dann heißt die Familie von Zufallsvariablen $\{X_t, t \ge 0\}$ Poisson-Prozess mit dem Parameter λ .

Satz: Sei $\{X_t, t \ge 0\}$ ein Poisson-Prozess mit dem Parameter λ , dann gilt:

- i) $X_t \sim Po(\mu_t = \lambda \cdot t)$
- ii) $X_t X_s \sim Po(\lambda \cdot (t s))$

Def: In einem Prozess des Entstehens von Ereignissen in der Zeit heißt $W_i := \text{Zeit zwischen dem } (i-1)$ - ten und dem i - ten Ereignis die *Wartezeit auf das i* - *te Ereignis*.

Satz: In einem Poisson-Prozess mit dem Parameter λ gilt für die Wartezeiten:

- i) W₁, W₂, W₃, ... sind stochastisch unabhängig.
- ii) $W_i \sim \text{Ex}(\lambda)$, d. h., die Wartezeiten sind exponentialverteilt.
- **B 17:** Für eine kleine Versicherung sei der Prozess des Einlangens von Schadensmeldungen als Poisson-Prozess modelliert mit $\lambda = 4$ bezogen auf eine Stunde. Der Arbeitstag habe 8 Stunden.

Wie groß ist die Wahrscheinlichkeit, dass

- a) an einem Arbeitstag mindestens 40 Schadensmeldungen eintreffen?
- b) in jeder Stunde des Arbeitstages mindestens zwei Schadensmeldungen eintreffen?
- c) die erste Schadensmeldung schon eine halbe Stunde nach Arbeitsanfang eintrifft?
- d) zwischen den ersten 5 Meldungen nicht mehr als jeweils 20 Minuten vergehen?
- a) P(in 8 Stunden mindestens 40 Meldungen) = $P(X_8 \ge 40)$.

Nun ist $X_8 \sim Po(\mu_t = \lambda \cdot t = 4 \cdot 8 = 32)$, daraus ergibt sich etwa mit EXCEL: $P(X_8 \ge 40) = 0.0918$.

b) P(in einer Stunde mindestens 2 Meldungen) = $P(X_1 \ge 2)$.

$$X_1 \sim Po(\mu_t = \lambda \cdot t = 4 \cdot 1 = 4)$$
; EXCEL liefert: $P(X_1 \ge 2) = 0.9084$.

Jetzt beachtet man, dass in einem Poisson-Prozess Beobachtungen in disjunkten Intervallen stochastisch unabhängig sind, daher hat man:

P(in jeder Stunde mindestens 2 Meldungen) = $P(X_1 \ge 2)^8 = (0.9084)^8 = 0.4637$.

c) Da in einem Poisson-Prozess der Beginn der Beobachtung ohne Belang ist, kann man die gesuchte Wahrscheinlichkeit dadurch berechnen, dass man den Prozess eine halbe Stunde lang beobachtet und fragt, ob mindestens 1 Ereignis stattfindet, also

$$X_{1/2} \sim \text{Po}(\mu_t = \lambda \cdot t = 2)$$
 und $P(X_{1/2} \ge 1) = 0.8647$ - wieder mit einer Rechenhilfe.

Man kann aber auch über die Wartezeit auf das erste Ereignis rechnen:

$$W_1 \sim \text{Ex}(\lambda = 4) \text{ und } P(W_1 \le 0.5) = 1 - e^{-4.0.5} = 0.8647$$
.

d) Die Wartezeit auf das nächste Ereignis von 20 Minuten ist in Stunden ausgedrückt $\frac{1}{3}$:

$$W_i \sim \text{Ex}(\lambda = 4)$$
 und $P(W_i \le \frac{1}{3}) = 1 - e^{-4/3} = 0.7364$.

Wegen der Unabhängigkeit der Wartezeiten hat man für 5 Ereignisse:

$$P\left((W_1 \leq \tfrac{1}{3}) \cap (W_2 \leq \tfrac{1}{3}) \cap ... \cap (W_5 \leq \tfrac{1}{3}) \right) = P\left((W_1 \leq \tfrac{1}{3}) \right)^5 = (0,7364)^5 = 0,2166 \; .$$



3.3 Ergänzungen





Multivariate Normalverteilung

Poisson-Prozess

Monte-Carlo-Simulation

Wahrscheinlichkeiten stehen in engem Zusammenhang mit relativen Häufigkeiten. Oft sind die mathematischen Modelle zu kompliziert, um die nötigen Berechnungen geschlossen durchzuführen. Ein Lösungsansatz besteht dann darin, die dahinter stehenden Verteilungen zu simulieren, d. h., Zufallszahlen zu erzeugen, die diesen Verteilungen "genügen". Dann bekommt man Daten, aus denen man die zu berechnenden Wahrscheinlichkeiten über relative Häufigkeiten der entsprechenden Ereignisse schätzen kann. Erwartungswerte werden durch Mittelwerte der Daten geschätzt usw. Das nennt man Monte-Carlo-Simulation.

B 18: Wir betrachten den Output Y einer Produktion, deren Input X zufallsbehaftet ist, beispielsweise eine landwirtschaftliche Produktion. Der Ernteertrag sei über eine komplexe Funktion g von der Sonnenscheindauer X abhängig. Modelliert man X als Zufallsvariable, so ist auch die Transformierte Y = g(X) eine Zufallsvariable. Die Abhängigkeit über g sei so kompliziert, dass man die Verteilung von Y nicht bestimmen kann. Die Funktionswerte g(x) hingegen seien leicht auszuwerten.

Bem.: Die Verteilungsfunktion von Y kann man durch die empirische Verteilungsfunktion schätzen. Da Y eigentlich von der Zufallsvariablen X über die Funktion g abhängt, kann man einfach Daten zu X nutzen, um die Verteilung von Y zu schätzen:

$$x_1, x_2, ..., x_n, ...$$

Diese setzt man dann in die Funktion g ein und erhält Daten zur Verteilung von Y:

$$y_1 = g(x_1), y_2 = g(x_2), \dots y_n = g(x_n), \dots$$

Hat man keine echten Daten zu X, so kann man diese auf dem Computer künstlich erzeugen; diese sollen der Verteilung von X genügen und sich wie unabhängige Realisierungen von X verhalten.

Def.: Ein *Zufallszahlengenerator* ist eine Rechenvorschrift, die eine Folge von Zahlen $z_1, z_2, ..., z_n, ...$ liefert, die sich (fast) wie die Realisierungen von unabhängigen und identisch S(0,1) - verteilten Zufallsvariablen $Z_1, Z_2, ..., Z_n, ...$ verhalten.

Bem.: Solche Zahlen nennen wir (Pseudo-)Zufallszahlen. Viele Computerprogrammsysteme, so auch Excel oder SPSS, enthalten einen solchen Zufallszahlengenerator.

Satz: Die Verteilung von X ist gegeben durch ihre Verteilungsfunktion F mit der zugehörige Quantilfunktion Q. Man hat Pseudo-Zufallszahlen

 $z_1, z_2, ..., z_n, ...,$

die sich (fast) wie unabhängige Realisierungen von $Z \sim S(0,1)$ verhalten. Dann verhalten sich

$$x_1 = Q(z_1), x_2 = Q(z_2), ..., x_n = Q(z_n), ...$$

(fast so,) als ob sie echte unabhängige Realisierungen der Zufallsvariablen X wären.

Bew.: Es gilt $Z = F(X) \sim S(0, 1)$ und $X = Q(Z) \sim F$ (siehe den Abschnitt über die stetige Gleichverteilung).

Prozedur: Die künstliche Erzeugung von Zahlen, welche der Verteilung der Zufallsvariablen Y = g(X)genügen, wenn X die Verteilungsfunktion F mit Quantilfunktion Q hat, nennt man Monte-Carlo-Simulation:

- i) Erzeuge Zufallszahlen $z_1, z_2, ..., z_n, ...,$
 - (wie) unabhängige Realisierung von S(0, 1)
- ii) Berechne $x_1 = Q(z_1), x_2 = Q(z_2), ..., x_n = Q(z_n), ...,$
- (wie) Daten zur Verteilungsfunktion F von X
- iii) Bereche $y_1 = g(x_1), y_2 = g(x_2), ... y_n = g(x_n), ...,$
- (wie) Daten von Y.

Bem.: Aus den Daten von Y lassen sich die interessierenden Größen wie Erwartungswerte oder Varianzen schätzen. Bleibt noch die Frage, wie genau die Ergebnisse sind. Das wird in Kapitel 5 behandelt werden.

B 20: Man erzeuge *n* Zahlen, die als unabhängige Realisierungen einer Exponentialverteilung gelten können. Die Quantilfunktion einer $\operatorname{Ex}(\lambda)$ lautet $\operatorname{Q}(p) = \frac{1}{\lambda} \cdot \ln(\frac{1}{1-p})$.

Sind $z_1, z_2, ..., z_n, ...$ Zufallszahlen aus $Z \sim S(0, 1)$, so sind

 $x_i = \frac{1}{\lambda} \cdot \ln(\frac{1}{1-z_i})$ Zufallszahlen aus der Exponentialverteilung.

Satz: Die folgende Prozedur liefert *annähernd* standardnormalverteilte Zufallszahlen:

- i) Erzeuge k Zufallszahlen $z_1, z_2, ..., z_k$ aus S(0,1) und bilde $u_i := \frac{\sum\limits_{j=1}^{n} z_j \frac{k}{2}}{\left\lceil \frac{k}{k} \right\rceil}$
- ii) Wiederhole Schritt i)

Bem.: Für k = 12 vereinfacht sich die Formel: $u_i := \sum_{j=1}^{12} z_j - 6$; die Normal-Approximation ist ziemlich gut.

Bew.: Für die $Z_i \sim S(0,1)$ gilt: $E(Z_i) = \frac{1}{2}$ $V(Z_i) = \frac{1}{12}$.

Die Summe ist nach dem zentralen Grenzwertsatz approximativ normalverteilt mit $\mu = k \cdot \frac{1}{2}$ und $\sigma^2 = k \cdot \frac{1}{12}$.

 $U := \frac{\sum_{j=1}^{k} Z_j - \frac{k}{2}}{\sqrt{\frac{k}{k}}} \qquad approx \qquad N(0,1).$ Für die standardisierte Summe gilt:

Jetzt realisiert man die Zufallsvariablen mit Monte-Carlo-Simulation.

B 21: Erzeuge *n* Zahlen, die als unabhängige Realisierungen einer Standardnormalverteilung gelten können. Die Quantilfunktion der N(0,1) ist Φ^{-1} ; Φ^{-1} geeignet numerisch approximiert.

Sind $z_1, z_2, ..., z_n, ...$ Zufallszahlen aus $Z \sim S(0, 1)$, so sind

 $u_i = \Phi^{-1}(z_i)$ Zufallszahlen aus der Standardnormalverteilung.

(Zufallszahlen zu einer $N(\mu, \sigma^2)$ erhält man durch die Transformation $x_i = \mu + \sigma \cdot u_i$.)

Man kann nach dem Zentralen Grenzwertsatz auch so vorgehen, dass man die $u_1, u_2, ..., u_n, ...$ als standardisierte Summen von gewöhnlichen Zufallszahlen darstellt:

 $u_i := \sum_{j=1}^{12} z_j - 6$; dazu werden jeweils k = 12 Zufallszahlen $z_1, z_2, ..., z_{12}$ aus S(0, 1) erzeugt.

4 Stichproben und Stichprobenfunktionen

Manfred Borovcnik

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



4.1 Zufallsstichproben und statistisches Schließen

4.2 Stichprobenfunktionen (Statistiken)

4.3 Ergänzungen

Statistik ist die methodische Erhebung und Auswertung von Daten. Man unterteilt sie in beschreibende und schließende Statistik. Um zu erklären, was die schließende Statistik leistet, ist es ganz gut, zu sehen, was sie mehr leistet als die beschreibende Statistik: sie bietet die Möglichkeit, in Daten gefundene Ergebnisse zu verallgemeinern. Das wird Thema der Kapitel 5-7 sein. Und was sie an Mehr an Voraussetzungen verlangt: sie stellt Bedingungen daran, wie die Daten entstanden sind, aus denen man die Schlüsse zieht. Im Wesentlichen muss es sich bei den Daten um das Ergebnis einer Zufallsstichprobe handeln. Sind die Daten repräsentativ für eine Gesamtheit, so kann man die Schlüsse auf diese übertragen. Zufallsstichproben spielen eine große Rolle dabei, ob man Daten als repräsentativ ansehen kann. Das ist Thema des vorliegenden Kapitels.

4.1 Zufallsstichproben und statistisches Schließen

- Beschreibende Statistik
- Zufallsstichproben
- Statistisches Schließen

Die Zulässigkeit der Verallgemeinerung statistischer Ergebnisse ist gebunden an Voraussetzungen an die Entstehung der Daten. Es muss sich um eine Zufallsstichprobe handeln, will man die Ergebnisse auf größere Gesamtheiten übertragen. Der Begriff einer Zufallsstichprobe wird hier erläutert. Die mathematischen Konsequenzen führen noch einmal zurück in die Wahrscheinlichkeitsrechnung.

Beschreibende Statistik





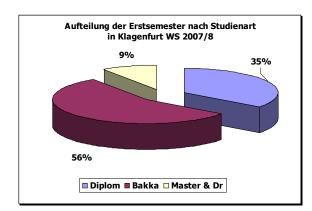
- Zufallsstichproben
- Statistisches Schließen

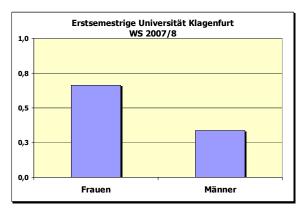
B: Wir befragen die Studierenden an der Universität Klagenfurt nach ihren persönlichen Daten und einigen Vorlieben. Wir erhalten damit eine Reihe von Merkmalen und eine Fülle von Daten.

Alter	Her- kunft	Ge- schlecht	Körper- größe	BMI	Anfahrt zur Uni	Anfahrtszeit zur Uni [min]	
19	D	m	179	23,1	PKW	90	
21	Ö	W	174	19,2	Bus	45	
18	Ö	W	172	22,4	Bahn+Bus	30	
18		W	168	26,8	PKW	80	

Mittels Methoden der Beschreibenden Statistik werden diese Informationen aufbereitet, in Tabellen zusammen gefasst, vergröbert, durch Diagramme bildlich dargestellt, in Kennzahlen charakterisiert. Etwa ergibt sich ein Anteil von Frauen unter den Erstsemestrigen an der Univ. Klagenfurt von 66% und ein mittlerer BMI (Bodymass-Index = $\frac{\text{K\"{o}}\text{rpergewicht}[kg]}{2}$) von 23,2. (Körpergröße [m])²

Bem.: Man kann die Klagenfurter Ergebnisse mit Österreich vergleichen: Der Frauenanteil an allen Studierenden in Österreich ist 48%; der BMI bei den Erwachsenen in Österreich (über 20 Jahre) liegt bei 25,1; die entsprechenden Diagramme zeigen deutliche Unterschiede.





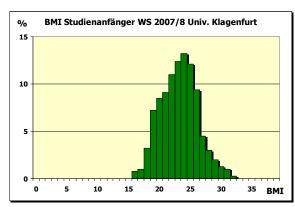
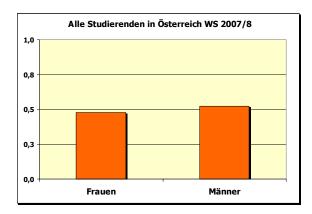
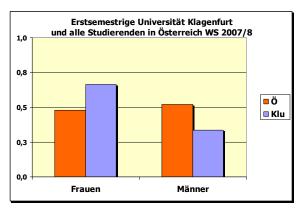
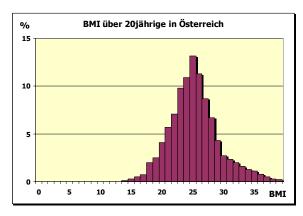


Abb. 1: Ergebnisse der Befragung von Klagenfurter Erstsemester WS 2007/8







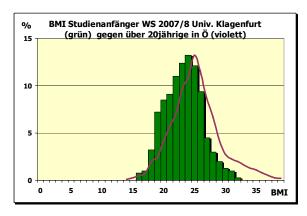


Abb. 2: Vergleiche der Ergebnisse von Klagenfurter Erstsemester mit Österreich

Bem.: Beschreibende Statistik bietet Methoden, mit denen man die Fülle an Information, die in Daten steckt, lesen und quantifizieren kann. Die Resultate beziehen sich dabei immer nur auf die vorliegenden Daten; ein Schluss auf größere Grundgesamtheiten dahinter ist unzulässig.



Beschreibende Statistik





Zufallsstichproben



Statistisches Schließen

Hier wird der Begriff der "Zufallsstichprobe" erläutert. Zufallsstichproben "garantieren", dass Daten repräsentativ sind für größere Gesamtheiten, aus denen sie stammen, und Schlüsse, die man aus den unvollständigen Daten zieht, verallgemeinern kann. Damit befasst sich die schließende Statistik.

B: Man hat die Klagenfurter Studierenden (Erstsemester) befragt und ihren Anteil an Frauen bestimmt. $x_1 = 0, x_2 = 1, x_3 = 1..., x_n = 1$

Es ergab sich n=1892 und 1256 der Daten waren 1; das entspricht 1256 Frauen oder $\frac{1256}{1892}=0,66$; das ist der Frauenanteil an Klagenfurter Erstsemestrigen. Diese Information ist unvollständig, wenn man Aussagen über den Frauenanteil an allen Studierenden in Österreich machen will. Eine naive Übertragung der Klagenfurter Verhältnisse führt zu einer Schätzung $\hat{\pi}=0,66$, was mit dem Wert $\pi=0,48$ von Statistik Austria nicht viel gemeinsam hat. Die Daten sind nicht repräsentativ für Österreich.

B: Man hat die Klagenfurter Erstsemester nach ihrem BMI befragt:

$$x_1 = 23.1, x_2 = 19.2, x_3 = 22.4.., x_n = 26.8.$$

Es ergab sich ein mittlerer BMI von $\bar{x}=23,2$. Kann man daraus den BMI für die österreichischen Erwachsenen (20+) schätzen? Auch hier sind die Daten nicht repräsentativ. Eine Schätzung $\hat{\mu}_A=23,2$ für Österreich ist nicht zulässig; aus Daten von Statistik Austria errechnet man $\mu_A=25,1$.

B: Man hat im Rahmen der Qualitätsüberwachung 25 Mehlpackungen zur Kontrolle gewogen:

 $x_1 = 1004,3, x_2 = 981,2, x_3 = 1021,6.., x_n = 1017,4.$

Als Mittelwert ergibt sich: $\bar{x} = 1004,5$. Kann man sagen, dass die Abfüllanlage gut eingestellt ist und daher die Mindestgewichtforderung von 985 g eingehalten wird? Das hängt davon ab, wie man die zur Kontrolle gewogenen Packungen ausgewählt hat und wovon im zeitlichen Verlauf es beeinflusst wird, dass Packungen unterschiedliches Gewicht eingefüllt bekommen.

B: Man würfelt mit einem gewöhnlichen Würfel und bekommt die Augenzahl als Gewinn ausbezahlt. Es erhebt sich die Frage nach dem Preis für dieses Spiel. Wir beobachten das Spiel eine Zeit lang und erhalten die Daten:

$$x_1 = 2$$
, $x_2 = 3$, $x_3 = 6$..., $x_n = 2$.

Kann man den Mittelwert $\bar{x} = 2.9$ einer Serie von 20 Würfen als Maßstab nehmen, um den Erwartungswert μ zu schätzen? Kann man damit beurteilen, ob der geforderte Preis gerechtfertigt ist?

Bem.: Man verfügt über Daten $x_1, x_2, ..., x_n$.

Diese stellen eine unvollständige Information über die Grundgesamtheit (Population) dar:

- Man hat nur eine Teilmenge einer endlichen Population befragt/ beobachtet. Oder:
- Man hat einen Prozess des Entstehens von Ereignissen nur eine Zeitlang beobachtet.

Welchen Bedingungen müssen die Daten in ihrer Entstehung genügen, damit wir die Information, die in ihnen steckt, als repräsentativ für die Grundgesamtheit (bzw. den Prozess) betrachten können und sie daher auf diese verallgemeinern?

Die Entstehung eines Datums x_i aus der Grundgesamtheit bzw. dem Prozess wird als ein Zufallsexperiment modelliert. Die entsprechende Zufallsvariable X hat eine – unbekannte – Verteilung F. Die einzelnen Beobachtungen entsprechen Zufallsvariablen X_i , die immer unter denselben Bedingungen realisiert werden.

Def.: Sei X eine Zufallsvariable. Unter *Zufallsstichprobe von* X verstehen wir die Zufallsvariablen

$$X_1$$
, X_2 , ..., X_n , wenn

i) alle dieselbe Verteilung wie X haben, d. h., $X_i \sim X$ und ii) die X_i stochastisch unabhängig sind. Die Zahl n heißt dabei Umfang der Stichprobe; die Werte $x_1, x_2, ..., x_n$ heißen Realisierung der Stichprobe.

Bem.: In der Statistik sind viele Verfahren der Stichprobengewinnung in Verwendung, welche die Unabhängigkeit der einzelnen Zufallsvariablen X_i verletzen. Die Auswirkungen sind dann meist sehr gering, weil der "Grad" der Abhängigkeit klein ist. Darüber hinaus werden andere Techniken der Stichprobengewinnung benützt, die nichts mit dem Zufall zu tun haben: etwa die Quotenstichprobe. Diese führen – Expertise vorausgesetzt – zu ähnlich guten Ergebnissen wie Zufallsstichproben. Von Zeit zu Zeit versucht man, diese Methoden extra zu validieren, d. h., mit Zufallsstichproben zu vergleichen.

Bem.: Die Realisierung der Stichprobe liefert konkrete Daten: $x_1, x_2, ..., x_n$ sind Zahlen – zum Unterschied von den Zufallsvariablen $X_1, X_2, ..., X_n$. Man drückt dies durch Kleinschreibung aus. Da meist Zufallsstichproben gemeint sind, wird auch kurz von Stichproben gesprochen.

Bem.: Zufallsstichproben "garantieren", dass die Ergebnisse für die dahinter steckende Grundgesamtheit repräsentativ werden.

Bem.: Welche Informationen über die Verteilung von X sollen nun mit Hilfe einer Zufallsstichprobe gewonnen werden? Dies hängt vom konkreten Problem ab. Oft stellt sich die Frage, wie groß sind

- Erwartungswert und Varianz von X,
- Anteile, etwa $F(x_0) = P(X \le x_0)$ an bestimmten Stellen x_0 ,
- bestimmte Quantile von X wie der Median, das obere oder das untere Quartil?

Def.: Empirische Verteilung eines Merkmals X über einer endlichen Gesamtheit $G = \{\omega_1, \omega_2, ..., \omega_N\}$ ist eine

Abbildung von $G \rightarrow \mathbf{R} : \omega \mapsto X(\omega)$.

Unter empirischer Verteilung von X versteht man die Funktion

 $F_e(x) = Anteil(X \le x)$

Satz: Wählt man aus einer endlichen Gesamtheit $G = \{\omega_1, \omega_2, ..., \omega_N\}$ zufällig aus (alle Elemente mit derselben Wahrscheinlichkeit) und bestimmt den Wert des Merkmals X, so erhält man eine Zufallsvariable X mit Verteilungsfunktion $F(x) = F_e(x)$.

Bew.: Wir setzen $\Omega = G$ und $P(\{\omega\}) = \frac{1}{N}$ für alle $\omega \in G$. Dann gilt: $P(X \le x) = Anteil(X \le x)$.

Folgerung: Bei zufälliger Auswahl aus einer endlichen Grundgesamtheit mit einem Merkmal X gilt:

$$E(X) = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i = \overline{x} \text{ und } V(X) = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i^2 - \overline{x}^2$$

Bem.: In diesem Fall sind Erwartungswert und Varianz der Wahrscheinlichkeitsverteilung von X nichts anderes als das arithmetische Mittel und die Varianz der empirischen Verteilung.

Bem.: Durch die zufällige Auswahl erhält man Anteile an der Grundgesamtheit als Wahrscheinlichkeiten. Wiederholt man das Zufallsexperiment (unabhängig), so spiegeln sich diese Wahrscheinlichkeiten in relativen Häufigkeiten in der Stichprobe, welche die Anteile in der Grundgesamtheit schätzen lassen. In gleicher Weise erhält man durch Mittelwerte in der Stichprobe Schätzungen für den Erwartungswert von F, welcher aber mit dem Mittelwert von Fe übereinstimmt, das ist der Mittelwert des Merkmals in der Grundgesamtheit.

B 1: Die Grundgesamtheit G besteht aus allen Haushalten in Deutschland am Jahresanfang 2001. Untersuchungsmerkmal ist das verfügbare monatliche Haushaltseinkommen (in $\mathfrak E$). Da die Gesamtzahl N| der Haushalte sehr groß ist, soll nur ein Teil der Einkommen erhoben werden. Wir wählen durch ein Zufallsexperiment mit gleichen Wahrscheinlichkeiten einen Haushalt ω_1 aus und bestimmen sein Einkommen $x_1 = X(\omega_1)$. Wir wiederholen den Vorgang der Auswahl aus der vollen Grundgesamtheit und ziehen unabhängig von der vorigen Ziehung einen Hauhalt ω_2, \ldots , bis wir genug Daten haben:

$$x_1, x_2, \dots, x_n$$

Wir schätzen das mittlere Einkommen aller Haushalte durch den Mittelwert \bar{x} der Daten.

Bem: In Beispiel 1 liegt die Grundgesamtheit G konkret vor. Praktisch realisieren wir die Auswahl mit Zufallszahlen.

- **B 3:** Ein Kosmetikhersteller entwickelt ein neues Herren-Parfum. Als Zielgruppe (= Grundgesamtheit G) definiert die Marketing-Abteilung alle deutschen Männer im Alter von 20 bis 45 Jahren mit höherer Bildung, gutem Einkommen und ausgeprägtem Modebewusstsein. Für jedes Individuum ω in G sei
 - $X(\omega) := \begin{cases} 1 & \text{falls } \omega \text{ die Duftnote positiv bewertet} \\ 0 & \text{falls } \omega \text{ die Duftnote nicht positiv bewertet} \end{cases}$

Wie schätzt man den Anteil der Männer, welche die Duftnote positiv bewerten? Dieser Anteil ist gleich $\pi = P(X = 1)$. Eine Liste der Grundgesamtheit G liegt nicht vor; es ist nicht klar, wer dazu gehört.

Bem.: Weil die Grundgesamtheit nirgends explizit aufgelistet ist, muss man andere Verfahren heranziehen, welche dem zufälligen Ziehen nahe kommen.

B 2: Wir betrachten die Tagesrendite X einer bestimmten Aktie (die relative Änderung ihres Kurses von einem Handelstag auf den anderen). Ein Anleger interessiert sich vor allem für Erwartungswert und Standardabweichung von X. Der Erwartungswert entspricht der mittleren Tagesrendite, die Standardabweichung ist ein Maß für ihr Risiko. Die Grundgesamtheit ist hypothetisch; sie besteht aus allen möglichen Entwicklungen des Aktienmarkts.

Bem.: In diesem Beispiel liegt die Grundgesamt nicht konkret vor. Es scheint auch schwierig, sich Daten auf dem Weg einer Stichprobe dafür zu beschaffen.

B 2f: Ein Anleger stellt an n aufeinander folgenden Tagen die Tagesrendite der Aktie fest und erhält die Daten $x_1, x_2, ..., x_n$. Kann man davon ausgehen, dass sie die Werte einer Zufallsstichprobe $X_1, X_2, ..., X_n$ aus X sind?

Zwischen aufeinander folgenden Tagen bestehen Abhängigkeiten, die ihre Ursache in allgemeinen Trends der Börse oder in speziellen Eigenheiten des zur Aktie gehörigen Unternehmens haben. Das kann Aussagen, die wir aus Stichprobenverfahren ableiten, sehr beeinträchtigen.

- Beschreibende Statistik
- Zufallsstichproben
- Statistisches Schließen

Die Aufgabe des Statistikers wird an weiteren Beispielen erläutert.

- **B 4:** Im Rahmen einer Steuerprüfung soll ermittelt werden, wie viel Mehrwertsteuer ein Kaufmann seinen Kunden insgesamt verrechnet hat. Die Rechnungen liegen als Belege vor. Aus Kostengründen wird die Prüfung auf einen Teil der Belege beschränkt. Folgende Fragen stellen sich:
 - Wie lässt sich aus den Ergebnissen der Teilprüfung die Gesamtsumme schätzen?
 - Wie wählt man die Rechnungen aus?
 - Wie viele Rechnungen muss man auswählen?

- B 5: Sei X die Temperatur in Köln morgen Mittag um zwölf Uhr. Wie lassen sich aus Aufzeichnungen der Temperatur Aussagen über X treffen?
 Die historischen Temperaturen am gleichen Datum früherer Jahre werden als Realisation einer Stichprobe aufgefasst. Ihr Mittelwert lässt sich als Schätzung für den Erwartungswert und damit als "Prognose" der morgigen Temperatur verwenden.
- **B 6:** Aus einer Totalerhebung über Garantiefälle weiß man: Die Pannenhäufigkeit für einen bestimmten PKW-Typ beträgt im ersten Jahr nach der Zulassung 7%. Man interessiert sich für die Pannenhäufigkeit π im zweiten Jahr. Bei n = 20 zufällig ausgewählten PKW beobachtet man im zweiten Jahr zwei Pannen. Es stellen sich folgende Fragen:
 - Ist $\pi > 7\%$?
 - Wie genau ist etwa die Schätzung $\hat{\pi} = 0.1$?
 - Wächst die Genauigkeit mit *n*?
 - Wie vertrauenswürdig ist die Angabe eines Intervalls, z.B. $8\% < \pi < 12\%$?
- **Def.:** Ist X eine auf der Grundgesamtheit Ω definierte Zufallsvariable. Ein *stochastisches Modell für* X ist eine Familie von Verteilungen $\mathcal{F} = \{F_{\lambda}, \lambda \in \Lambda\}$, die zur Modellierung von X herangezogen wird. λ heißt der Parameter der Verteilung.
- **B:** $\mathcal{F} = \{B(n, \pi), \ \pi \in (0, 1)\}$ ist die Familie der Binomialverteilungen mit festem n und variabler Erfolgswahrscheinlichkeit π . Das ist eine "natürliche" Modellfamilie für eine Bernoulli-Versuchsreihe. $\mathcal{F} = \{N(\mu, \sigma^2) \big| \ \mu \in \mathbf{R}\}$ ist die Familie der Normalverteilungen mit $\sigma^2 > 0$, das entweder als bekannt vorausgesetzt oder aus den Daten geschätzt wird.

Bem.: Beim statistischen Schließen setzt man häufig voraus, dass die Verteilung von X nicht völlig unbekannt ist. Oft kann man aufgrund von inhaltlichen Überlegungen davon ausgehen, dass X durch eine bestimmte Familie von Verteilungen modelliert werden kann, wobei der Wert gewisser Parameter offen ist. Durch die Beobachtung einer Stichprobe aus X werden dann im Rahmen der Verteilungsannahme Informationen über die unbekannten Parameter gewonnen.

B 7: Sei X die Temperatur in Köln morgen Mittag um zwölf Uhr. Die Grundgesamtheit ist hier wiederum hypothetisch und unendlich; sie besteht aus der Menge aller möglichen Wetterzustände morgen Mittag um zwölf Uhr. Aus physikalischen Gründen macht es Sinn, X mit der Familie der Normalverteilungen zu modellieren.

4 Stichproben und Stichprobenfunktionen

Manfred Borovcnik

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



4.1 Zufallsstichproben und statistisches Schließen

4.2 Stichprobenfunktionen (Statistiken)

4.3 Ergänzungen

Statistik ist die methodische Erhebung und Auswertung von Daten. Man unterteilt sie in beschreibende und schließende Statistik. Um zu erklären, was die schließende Statistik leistet, ist es ganz gut, zu sehen, was sie mehr leistet als die beschreibende Statistik: sie bietet die Möglichkeit, in Daten gefundene Ergebnisse zu verallgemeinern. Das wird Thema der Kapitel 5-7 sein. Und was sie an Mehr an Voraussetzungen verlangt: sie stellt Bedingungen daran, wie die Daten entstanden sind, aus denen man die Schlüsse zieht. Im Wesentlichen muss es sich bei den Daten um das Ergebnis einer Zufallsstichprobe handeln. Sind die Daten repräsentativ für eine Gesamtheit, so kann man die Schlüsse auf diese übertragen. Zufallsstichproben spielen eine große Rolle dabei, ob man Daten als repräsentativ ansehen kann. Das ist Thema des vorliegenden Kapitels.

4.2 Stichprobenfunktionen (Statistiken)

- Wichtige Statistiken
- Statistiken für Anzahlen und Anteile
- Normalverteilte Stichproben

Aus den Stichproben werden weitere Zufallsvariable berechnet: die Summe der Daten, das arithmetische Mittel, die Standardabweichung etc. Letztlich werden diese "Statistiken" dazu dienen, unbekannte Anteile, Mittelwerte etc. zu schätzen. Damit man verlässliche Schlüsse aus den Werten der Statistiken ziehen kann, muss man erst ihre Verteilung unter wohlbekannten Voraussetzungen ableiten. Das wird hier für zwei wesentliche Grundtypen von Problemstellungen getan, und zwar für Bernoulli-Versuchsreihen und für Messreihen, welche mit der Normalverteilung modelliert werden.

Wichtige Statistiken



- Statistiken für Anzahlen und Anteile
- Normalverteilte Stichproben

Def.: X_1 , X_2 , ..., X_n eine Stichprobe aus X. Die Funktion $g: \mathbb{R}^n \to \mathbb{R}$ vermittelt eine Zufallsvariable $Y := g(X_1, X_2, ..., X_n)$. Diese nennt man *Stichprobenfunktion* oder *Statistik*.

Def.: $X_1, X_2, ..., X_n$ eine Stichprobe aus X. Die Funktionen

$$\overline{X} := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$
 bzw. $S^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$

heißen Stichprobenmittel bzw. Stichprobenvarianz

Bem.: Die Funktion g in n Variablen bildet hier das arithmetische Mittel bzw. die Varianz (mit Nenner n) aus den Variablen $x_1, x_2, ..., x_n$. Diese Stichprobenfunktionen dienen dazu, Erwartungswert bzw. Varianz der Verteilung von X zu schätzen. Die Stichprobenvarianz wird über den Verschiebungssatz ausgewertet:

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i^2 - \overline{X}^2$$
.

Wichtige Statistiken





- Statistiken für Anzahlen und Anteile
- Normalverteilte Stichproben

Satz (Verteilung der Statistiken bei Bernoulli-Versuchsreihen):

 $X \sim B(1, \pi)$ und X_1 , X_2 , ..., X_n ist eine Stichprobe aus X. Dann gilt

a)
$$Y_n := \sum_{i=1}^n X_i \sim B(n, \pi)$$
 b) $S^2 = \overline{X} \cdot (1 - \overline{X})$

Bem.: Das arithmetische Mittel \overline{X} wird in diesem Fall zum Anteil der 1en in der Stichprobe.

Bew. nur von b): alle X_i nehmen nur die Werte 0 und 1 an; daher gilt $X_i = X_i^2$. Setzt man jetzt in die Formel nach dem Verschiebungssatz ein, so erhält man: $S^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i - \overline{X}^2$. Der Rest ist trivial.

Satz (Verteilung der Summe bei endlichen Gesamtheiten):

Eine Grundgesamtheit bestehe aus N Einheiten, von denen genau M eine bestimmte Eigenschaft aufweisen, wobei $0 \le M \le N$. Man zieht zufällig n Einheiten und bezeichnet

$$X_i := \begin{cases} 1 & \text{falls die } i - \text{te gezogene Einheit die Eigenschaft hat} \\ 0 & \text{falls nicht} \end{cases}$$

Stichprobenfunktion $Y_n = \sum_{i=1}^n X_i$ =Anzahl der gezogenen Einheiten mit dieser Eigenschaft.

a) mit Zurücklegen gezogen, dann gilt $Y_n \sim B(n, \pi = \frac{M}{N})$ Werden die Einheiten { b) ohne Zurücklegen gezogen, dann gilt $Y_n \sim H(n, N, M)$

Bew.: a) Beim Ziehen mit Zurücklegen bilden X_1 , X_2 , ..., X_n eine Stichprobe von $X \sim B(1, \pi = \frac{M}{N})$.

b) wurde bei der Einführung der hypergeometrischen Verteilung schon bewiesen.

Bem.: Beim Ziehen ohne Zurücklegen sind die X_i wohl alle verteilt wie $X \sim B(1, \pi = \frac{M}{N})$; es besteht aber eine Abhängigkeit zwischen den X_i , sodass X_1 , X_2 , ..., X_n keine Zufallsstichprobe darstellt. Weil nun die Auswahl durch einen eingeschränkten Zufall erfolgt – es wird aus allen, die noch nicht gezogen sind, zufällig mit derselben Wahrscheinlichkeit gewählt – spricht man auch hier von Zufallsstichproben (im weiteren Sinne). Darüber hinaus ist die Verteilung einer wesentlichen Statistik, nämlich der Summe, bekannt. Daraus kann man Aussagen für das arithmetische Mittel gewinnen, welches den Anteil $\pi = \frac{M}{N}$ schätzen lässt.

Bem.: In der Praxis wird fast immer ohne Zurücklegen gezogen. Wenn der Auswahlsatz $\frac{n}{N}$ nicht zu groß ist (Faustregel < 0,05), besteht wegen der Annäherung der hypergeometrischen Verteilung an die Binomialverteilung kein wesentlicher Unterschied zwischen den beiden Arten der Ziehung. Auch das rechtfertigt, trotz Verletzung der Unabhängigkeit zwischen den einzelnen Ziehungen von Zufallsstichproben zu sprechen.







Statistiken für Anzahlen und Anteile

■ Normalverteilte Stichproben

Def.: Sei $X \sim N(\mu, \sigma^2)$. Eine Stichprobe $X_1, X_2, ..., X_n$ aus X heißt normalverteilte Stichprobe.

Satz (Verteilung des Stichprobenmittels bei normalverteilten Stichproben):

Sei $X \sim N(\mu, \sigma^2)$ und X_1 , X_2 , ..., X_n eine Stichprobe aus X. Dann gilt:

- a) für das Stichprobenmittel $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$,
- b) für das standardisierte Stichprobenmittel $\frac{\overline{X} \mu}{\sigma} \sqrt{n} \sim N(0, 1)$.

Def.: $U_1, U_2, ..., U_{\nu}$ sind unabhängige N(0, 1)-verteilte Zufallsvariable.

$$Q := \sum_{i=1}^{\nu} U_i^2$$

Dann heißt die Verteilung von Q Chi-Quadrat-Verteilung mit v Freiheitsgraden; kurz: $Q \sim \chi^2(v)$

Satz (Chi-Quadrat-Verteilung): $Q \sim \chi^2(\nu)$. Dann gilt:

- a) Die Verteilung von Q ist stetig auf dem Träger $[0, \infty)$; die Dichten sind rechtsschief.
- b) Für Erwartungswert und Varianz gilt:

$$E(Q) = v$$
 und $V(Q) = 2v$

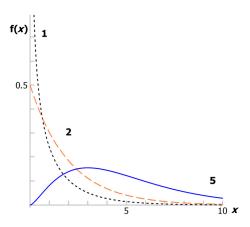


Abb. 3: Dichten der $\chi^2(\nu)$ -Verteilung mit verschiedenen Freiheitsgraden

Bem.: Das p-Quantil einer $\chi^2(v)$ -Verteilung wird mit $\chi^2(v)_p$ bezeichnet. Die wichtigsten Quantile findet man in Tabellen; man rechnet sie leicht mit EXCEL oder SPSS aus.

Satz (Quadratische Abweichungen bei normalverteilten Stichproben):

 $X \sim N(\mu, \sigma^2)$ und X_1 , X_2 , ..., X_n eine Stichprobe aus X. Für die quadratischen Abweichungen gilt:

a)
$$Q^* := \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

b)
$$Q := \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

Folgerung (Verteilung der Stichprobenvarianz): $X_1, X_2, ..., X_n$ normalverteilte Stichprobe, dann gilt:

$$\frac{n S^2}{\sigma^2} \sim \chi^2(n-1)$$

Bew.:
$$\frac{n S^2}{\sigma^2} = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2 = Q$$

Def.: Seien U und Q stochastisch unabhängig, U ~ N(0,1) und Q ~ $\chi^2(\nu)$.

$$W := \frac{U}{\sqrt{Q/\nu}}$$

Dann heißt die Verteilung von W Student t-Verteilung mit v Freiheitsgraden;

wir schreiben: $W \sim t(\nu)$

Satz (t-Verteilung): $W \sim t(\nu)$. Dann gilt:

- a) Die Verteilung von W ist stetig auf dem Träger $\boldsymbol{\mathsf{R}}$; die Dichten sind symmetrisch um 0.
- b) Für Erwartungswert und Varianz hat man:

$$E(W) = 0$$
, falls $v \ge 2$, und $V(W) = \frac{v}{v-2}$, falls $v \ge 3$.

c) $\lim_{v \to \infty} P(W \le w) = \Phi(w)$ für alle $w \in \mathbf{R}$.

Bem: Das p-Quantil von W wird mit $t(v)_p$ oder kurz t_p bezeichnet, wenn keine Gefahr besteht, auf die Freiheitsgrade zu vergessen. Die Quantile sind in Tabellen enthalten oder man berechnet sie mit Excel/SPSS.

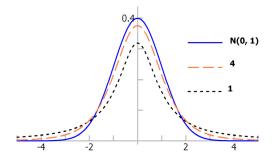


Abb. 4: Dichten der t-Verteilung mit verschiedenen Freiheitsgraden und Dichte der Standardnormalverteilung

Bem.: Die Dichte der t-Verteilung ähnelt der Dichte der Standard-Normalverteilung. An den Flanken weist die t-Dichte allerdings mehr Masse als die Normalverteilung auf. In der Mitte verläuft sie flacher. Aus der Konvergenz der t-Verteilungen mit wachsender Zahl von Freiheitsgraden gegen die Standard-Normalverteilung in c) gewinnt man eine Approximation, die für $\nu \ge 40$ (Faustregel) als ausreichend angesehen werden kann:

Quantile	1	t-Verteilung	3	N(0,1)	Relativer Fehler:
	v = 20 30 40		40	8	
p = 0.950	1,7247	1,6973	1,6839	1,6449	0,023
p = 0.995	2,8453 2,7500		2,7045	2,5758	0,048

Satz (Verteilung des Mittelwerts standardisiert mit S bei normalverteilten Stichproben):

Stichprobe X_1 , X_2 , ..., X_n aus $X \sim N(\mu, \sigma^2)$ und

$$U := \frac{\overline{X} - \mu}{\sigma} \sqrt{n} \text{ und } Q := \frac{n S^2}{\sigma^2}$$

Für standardisiertes Mittel U und die quadratische Abweichung Q gilt:

- a) U und Q sind stochastisch unabhängig.
- b) $U \sim N(0,1)$ und $Q \sim \chi^2(n-1)$.

c)
$$\frac{\mathrm{U}}{\sqrt{\mathrm{Q}/(n-1)}} \sim \mathrm{t}(n-1)$$

Bem.: Die angegebene Stichprobenfunktion in c) lässt sich vereinfachen zu $\frac{X-\mu}{S} \cdot \sqrt{n-1}$.

Bew.:

- a) ist schwierig.
- b) U als das standardisierte arithmetische Mittel und Q als Quadratische Abweichung sind bei normalverteilten Stichproben wie angegeben verteilt.
- c) ist nach Definition der t-Verteilung trivial.

Def.: Seien Q_n und Q_m unabhängige χ^2 -verteilte Zufallsvariable, $Q_n \sim \chi^2(n)$ und $Q_m \sim \chi^2(m)$.

$$Z := \frac{Q_n/n}{Q_m/m} \sim F(n, m)$$

Die Verteilung von Z heißt Fisher-Verteilung mit n und m Freiheitsgraden, in Zeichen: $Z \sim F(n, m)$.

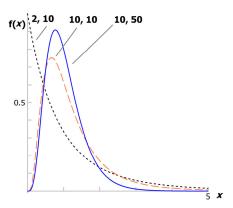


Abb. 5: Dichten einiger F-Verteilungen

Satz (Fisher-Verteilung): $Z \sim F(n, m)$. Dann gilt:

- a) Die Verteilung von Z ist stetig mit Träger $[0, \infty)$.
- b) Für Erwartungswert und Varianz gilt:

$$E(Z) = \frac{m}{m-2}$$
, falls $m \ge 3$, und $V(Z) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$, falls $m \ge 5$

c) Sei $Q_n \sim \chi^2(n)$ n fix. Dann gilt

$$\lim_{m \to \infty} P(Z \le z) = P(\frac{Q_n}{n} \le z) = P(Q_n \le nz)$$
 für alle $z \in \mathbf{R}$.

Bem.: Quantile der F-Verteilung findet man in geeigneten Tabellen oder man berechnet sie mit Excel/SPSS.

Bem.: Für große m lässt sich nach c) die F(n, m)-Verteilung durch die $\chi^2(n)$ -Verteilung approximieren.

Satz (Quotient quadratischer Abweichungen bei zwei normalverteilten Stichproben):

 X_1 , X_2 , ..., X_n eine Stichprobe aus $X \sim N(\mu_X, \sigma_X^2)$ und $Y_1, Y_2, ..., Y_n$ aus $Y \sim N(\mu_Y, \sigma_Y^2)$ Beide Stichproben sind unabhängig.

$$Q_X := \sum_{i=1}^n \left(\frac{X_i - \overline{X}}{\sigma_X}\right)^2$$
 und $Q_Y := \sum_{i=1}^n \left(\frac{X_i - \overline{Y}}{\sigma_Y}\right)^2$

Für die quadratischen Abweichung der X- bzw. Y-Stichprobe gilt:

- a) Q_X und Q_Y sind stochastisch unabhängig.
- b) $Q_X \sim \chi^2(n-1)$ und $Q_Y \sim \chi^2(m-1)$
- c) $\frac{Q_X/(n-1)}{Q_Y/(m-1)} \sim F(n-1, m-1)$

Bem: Die F-Verteilung wird zum Vergleich von Streuungen verwendet. Der Zähler Q_X schätzt die Varianz der X-Stichprobe, der Nenner die Varianz der Y-Stichprobe. Ist der Quotient nahe bei 1, so wird man annehmen können, dass die beiden Varianzen gleich sind. Falls nicht, so wird man den Unterschied "anerkennen".

4 STICHPROBEN UND STICHPROBENFUNKTIONEN

Manfred Borovcnik

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



- 4.1 Zufallsstichproben und statistisches Schließen
- 4.2 Stichprobenfunktionen (Statistiken)

4.3 Ergänzungen

Statistik ist die methodische Erhebung und Auswertung von Daten. Man unterteilt sie in beschreibende und schließende Statistik. Um zu erklären, was die schließende Statistik leistet, ist es ganz gut, zu sehen, was sie mehr leistet als die beschreibende Statistik: sie bietet die Möglichkeit, in Daten gefundene Ergebnisse zu verallgemeinern. Das wird Thema der Kapitel 5-7 sein. Und was sie an Mehr an Voraussetzungen verlangt: sie stellt Bedingungen daran, wie die Daten entstanden sind, aus denen man die Schlüsse zieht. Im Wesentlichen muss es sich bei den Daten um das Ergebnis einer Zufallsstichprobe handeln. Sind die Daten repräsentativ für eine Gesamtheit, so kann man die Schlüsse auf diese übertragen. Zufallsstichproben spielen eine große Rolle dabei, ob man Daten als repräsentativ ansehen kann. Das ist Thema des vorliegenden Kapitels.

4.3 Ergänzungen

- Zufallszahlen zur Zufallsauswahl
- Weitere Verfahren der Auswahl

Zufälliges Ziehen mit oder ohne Zurücklegen aus einer endlichen Gesamtheit wird auch als reine Stichprobenauswahl bezeichnet. Es gibt weitere Verfahren der Auswahl, bei denen der Zufall nicht immer eine tragende Rolle bei der konkreten Auswahl übernimmt.

Zufallszahlen zur Zufallsauswahl





Weitere Verfahren der Auswahl

Wie führt man eine Zufallsstichprobe aus einer endlichen Gesamtheit konkret durch. Man nutzt dazu Zufallszahlen. Viele Programme wie z.B. Excel enthalten einen so genannten Zufallszahlengenerator. Dieser liefert eine Folge von Zahlen $z_1, z_2, ..., z_n, ...$, die sich (fast) wie die Realisierungen von unabhängigen und identisch S(0,1) - verteilten Zufallsvariablen $Z_1, Z_2, ..., Z_n, ...$ verhalten. Mit Hilfe solcher Zufallszahlen lassen sich Beobachtungseinheiten aus einer gegebenen Grundgesamtheit auswählen.

Aus einer Population mit 750 Personen sollen fünf zufällig ausgewählt werden. Dazu nummerieren wir die Personen von 1 bis 750 und erzeugen mit einem Zufallszahlengenerator fünf Zufallszahlen, beispielsweise die Zahlen

0,357628, 0,984527, 0,756239, 0,623854, 0,423651.

Diese werden nun jeweils mit der Gesamtzahl 750 multipliziert und zur nächstgrößeren ganzen Zahl aufgerundet. Man erhält

269, 739. 568, 468, 318.

Entsprechend werden die Personen mit diesen Nummern ausgewählt.

Bem.: Mit dem beschriebenen Verfahren erhält man eine Auswahl mit Zurücklegen. Für eine Auswahl ohne Zurücklegen werden Zufallszahlen, welche die Nummer einer bereits ausgewählten Person ergeben, einfach ignoriert. Allenfalls muss man mehr Zufallszahlen erzeugen.







Weitere Verfahren der Auswahl

Def.: Die (endliche) Grundgesamtheit G wird in mehrere Teilgesamtheiten $S_1, S_2, ..., S_k$ zerlegt; d. h. insbesondere, dass die Teilgesamtheiten paarweise disjunkt sind. Die Zufallsauswahl wird in jeder Teilgesamtheit getrennt vorgenommen. Diese Methode der zufälligen Auswahl nennt man geschichtete Auswahl, die Teilgesamtheiten heißen Schichten.

Bem.: In den einzelnen Schichten kann man unterschiedliche Auswahlsätze anwenden. Die Schichtung kann nach mehreren Merkmalen vorgenommen werden, etwa nach Geschlecht und Altersklassen.

Bem.: Die geschichtete Auswahl ist dann der einfachen Auswahl vorzuziehen, wenn das interessierende Merkmal innerhalb der Schichten wenig streut, aber zwischen den Schichten ein größerer Unterschied besteht. Man reduziert damit die Streuung der Stichprobenfunktionen und verbessert so die Genauigkeit der ermöglichten Aussagen. Ferner kann man auch für die einzelnen Schichten separate Aussagen treffen.

B9: In einer Gesamtheit von Betrieben soll der durchschnittliche Gewinn geschätzt werden. Es liege eine Liste aller Betriebe der Gesamtheit und ihrer Größenklasse (Anzahl der Beschäftigten) vor. Dann ist eine Schichtung gemäß der Größenklasse sinnvoll.

Def.: Die (endliche) Grundgesamtheit G wird in mehrere disjunkte Teilgesamtheiten $S_1, S_2, ..., S_k$ zerlegt. Jetzt werden einige Teilgesamtheiten zufällig ausgewählt. Die ausgewählten Teilgesamtheiten werden vollständig erhoben. Die nicht-ausgewählten Teilgesamtheiten werden überhaupt nicht untersucht. Die Teilgesamtheiten werden hier Klumpen, die Auswahlmethode wird Klumpenauswahl genannt.

Bem.: Bei den Klumpen kann es sich etwa um Wahlkreise (oder andere räumliche Einheiten wie Stadtbezirke etc.), aber auch um andere Einheiten wie z.B. Unternehmen, Organisationen handeln. Die Klumpen werden in der Regel nach praktischen Gesichtspunkten zusammengestellt, um die Erhebung der Daten zu vereinfachen.

B 10: Die für die Bundestagswahl gebildeten Wahlbezirke stellen eine Zerlegung des Gebiets der Bundesrepublik dar. Eine Liste der Wahlberechtigten eines Bezirks ist bei der zuständigen Gemeinde erhältlich. In der Marktforschung werden häufig Wahlbezirke als Klumpen ausgewählt und eine Auswahl der dort Wahlberechtigten befragt.

Def.: Die (endliche) Grundgesamtheit G wird in mehrere disjunkte Teilgesamtheiten $S_1, S_2, ..., S_k$ zerlegt. Die Zerlegungsmerkmale sollen mit dem Zielmerkmal, das untersucht werden soll, eng korreliert sein. Dadurch entstehen Teilgesamtheiten, auf denen das Zielmerkmal wenig streut. Welches Element aus der Teilgesamtheit entnommen wird, ist dann fast egal. Man muss nur darauf achten, dass die Teilgesamtheiten in der Auswahl richtig repräsentiert sind: D. h., die Teilgesamtheiten müssen mit genau jenen *Quoten* in der Auswahl vertreten sein, mit der sie an der Grundgesamtheit teilhaben. Eine solche bewusste Auswahl nennt man *Quotenauswahl*.

Bem.: Für die Anwendung von Quotenverfahren ist viel Wissen über jene Merkmale nötig, die mit dem Zielmerkmal korrelieren. Diese Teilgesamtheiten haben Anteile (=Quoten) an der Grundgesamtheit. Über diese Quoten muss man Bescheid wissen. Man muss aber auch die Quoten erfüllen können. Die Auswahl wird nicht dem Zufall überlassen sondern bewusst getroffen. Einzige Einschränkung ist die "Erfüllung der Quoten". Man erzeugt damit eine Stichprobe, deren Quoten für die Teilgesamtheiten mit denen der Grundgesamtheit übereinstimmen. Bei sorgfältiger Anwendung liefern Quotenstichproben ziemlich gute Ergebnisse. Sie werden in der empirischen Sozialforschung häufig verwendet.

B 11: Bei Meinungsumfragen werden unter anderen die Hilfsmerkmale Geschlecht, Alter, Berufsgruppe und Wohnortgröße verwendet. Die gemeinsame Verteilung dieser Merkmale in der gesamten Bevölkerung ist bekannt. Auf ihrer Basis bildet man Schichten. Der Interviewer sucht in jeder Schicht aufs Geratewohl die Anzahl von Personen aus, die ihrem Anteil in der Bevölkerung entspricht.

Bem.: Bei der Schichtenauswahl und der Klumpenauswahl lässt sich die Wahrscheinlichkeit, dass eine gegebene Untersuchungseinheit in die Stichprobe kommt, quantifizieren. Auf dieser Basis kann man dann Schlüsse im Sinne der statistischen Inferenz ziehen. Für Quotenstichproben gilt dies nicht. Hier müssen andere Rechtfertigungen und Evaluierungen der Methode herangezogen werden. Für Einzelheiten dieser Verfahren sei auf das Lehrbuch Krug et al. (2001) verwiesen, ferner auf Bausch (1990) und StatBundesamt (1960).





5 Schätzverfahren für Parameter

Manfred Borovcnik

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



5.1 Punktschätzung

- 5.2 Konstruktionsprinzipien für Punktschätzer
- 5.3 Intervallschätzungen
- 5.4 Ergänzungen

Will man den unbekannten Mittelwert eines Merkmals in einer Grundgesamtheit kennen, so könnte man alle Werte erheben und das Mittel bestimmen. Normalerweise hat man aber nur die Information über einige wenige Merkmalswerte. Dann "schätzt" man mit dem Mittelwert der Daten den unbekannten Mittelwert der Grundgesamtheit. Die Vorgangsweise bekommt ihren Sinn, wenn die Daten als Realisation einer Stichprobe aufgefasst werden können. Dann nämlich kann man die Verallgemeinerung der partiellen Information von den wenigen Daten auf die unbekannte Grundgesamtheit rechtfertigen, indem man diesem Schätzverfahren "gute Eigenschaften" zusprechen kann. Man kann dann auch aus den Daten ein Intervall bestimmen, das einerseits den unbekannten Wert mit "hoher Sicherheit" einschließt und das anderseits die Genauigkeit der Information aus den Daten widerspiegelt. Welche Statistiken geeignet sind für bestimmte Parameter, wird mit Hilfe der Wahrscheinlichkeitsrechnung beurteilt.

5.1 Punktschätzung

- Kriterien für gute Schätzungen
- Schätzung von Erwartungswerten und Anteilen
- Schätzung anderer Parameter

Ist die Verteilung einer Zufallsvariablen X durch eine Familie von Verteilungen modelliert, der eigentliche Wert des Parameters, der die Verteilung festlegen würden, aber unbekannt, so bestimmt man aus einer Stichprobe der Zufallsvariablen eine Funktion, diese Statistik soll den unbekannten Wert schätzen. Man nennt die Vorgangsweise Punktschätzung.

Kriterien für gute Schätzungen



- Schätzung von Erwartungswerten und Anteilen
- Schätzung anderer Parameter

Jede Funktion der Stichprobe kann als Schätzer für unbekannte Parameter dienen. Man braucht Kriterien, die angeben, was einen "besseren" Schätzer von einem "schlechteren" unterscheidet.

B 1: Hängen Ausgaben von Studierenden vom Studienort ab? Um solche Fragen zu beantworten, untersucht man zunächst die monatlichen Ausgaben der Studierenden an *einer* Universität. Man möchte sich Information verschaffen über die durchschnittlichen Ausgaben, den Median und die beiden weiteren Quartile, sowie den Anteil der Studierenden mit monatlichen Ausgaben von mehr als 1000 €. Dazu wählt man zufällig eine Stichprobe von 50 Studierenden und erhebt deren monatliche Ausgaben:

309	661	236	1130	1288	1380	761	487	577	310
714	611	678	637	694	459	286	458	567	1010
398	107	951	1244	442	494	1082	827	727	252
234	561	797	1201	225	363	676	805	203	936
168	558	336	779	354	150	340	837	237	460

Den interessierenden Größen entsprechen offenbar die folgenden Parameter der Verteilung:

 $\theta_1 = \mu_X$ der Erwartungswert von X, $\theta_2 = x_{0.5}$ Median von X,

 $\theta_3 = x_{0.25}$ und $\theta_4 = x_{0.75}$ das untere und das obere Quartil von X,

 $\theta_5 = P(X > 1000)$, Wahrscheinlichkeit, einen Studierenden mit mehr als $1000 \in$ Ausgaben zu ziehen.

Def.: X_1 , X_2 , ..., X_n ist eine Stichprobe von X, θ ein Parameter der Verteilung von X. Eine Stichprobenfunktion $\hat{\theta}(X_1, X_2, ..., X_n)$ nennt man eine *Schätzfunktion* für den Parameter θ .

Bem.: Der Schätzer $\hat{\theta}(X_1, X_2, ..., X_n)$ ist eine Zufallsvariable (eine Stichprobenfunktion) und kann mit den Methoden der Wahrscheinlichkeitsrechnung beurteilt werden. Der Schätzwert $\hat{\theta}(x_1, x_2, ..., x_n)$ ist eine Zahl, ausgerechnet aus den Daten $x_1, x_2, ..., x_n$ einer Stichprobe. Trotz dieses Unterschieds schreibt man meist kurz $\hat{\theta}$ sowohl für den Schätzer als auch für den Schätzwert. Betrachtet man den Schätzer bei verschiedenen Stichprobenumfängen n, schreibt man $\hat{\theta}_n$. Es ist klar, dass ein Schätzer $\hat{\theta}$ den zu schätzenden Parameter θ in der Regel nicht exakt trifft. Abhängig von der konkreten Realisation der Stichprobe ist sein Wert größer oder kleiner als der gesuchte Wert θ . Da $\hat{\theta}$ eine Zufallsvariable ist, sind jedoch mittels der Wahrscheinlichkeitsrechnung Aussagen darüber möglich, in welcher Weise $\hat{\theta}$ um θ schwankt.

Def.: Ein Schätzer $\hat{\theta} = \hat{\theta}(X_1, X_2, ..., X_n)$ für θ heißt *unverzerrt* (*erwartungstreu*; unbiased), wenn $E\{\hat{\theta}\} = \theta$ für alle möglichen Werte von θ .

Die Differenz $b(\hat{\theta}) = E\{\hat{\theta} - \theta\}$ heißt *Verzerrung* (bias) von $\hat{\theta}$.

Die Größe $MSE(\hat{\theta}) = E\{(\hat{\theta} - \theta)^2\}$ heißt *mittlerer quadratischer Fehler* (Mean-Squared-Error).

Bem.: Der Erwartungswert des Schätzers hängt auch vom zutreffenden Wert des Parameters θ ab: $E\{\hat{\theta}\}=E_{\theta}\{\hat{\theta}\}$. Wenn ein Schätzer erwartungstreu ist, so ist sein Erwartungswert ist in jedem Fall gleich dem unbekannten Parameter. Es ist wünschenswert, dass ein Schätzer unverzerrt ist, d. h. einen Bias von 0 hat.

Satz: Ist $\hat{\theta}$ erwartungstreu, so ist $MSE(\hat{\theta}) = V(\theta)$

Bem.: Für erwartungstreue Schätzer ist der MSE gleich der Varianz des Schätzers.

Bew.: Wegen $E\{\hat{\theta}\} = \theta$ ist $MSE(\hat{\theta}) = E\{(\hat{\theta} - \theta)^2\} = E\{(\hat{\theta} - E(\hat{\theta}))^2\} = V(\theta)$.

Def.: Der Schätzer $\hat{\theta}_n = \hat{\theta}(X_1, X_2, ..., X_n)$ sei für jede Stichprobenlänge n definiert. Er heißt *konsistent* für θ , wenn er nach Wahrscheinlichkeit gegen θ konvergiert: $p - \lim_{n \to \infty} \hat{\theta}_n = \theta$.

Bem.: Die Konvergenz bedeutet ausführlicher: Für ein beliebig vorgegebenes $\varepsilon > 0$ gilt lim $P(|\hat{\theta}_n - \theta| \le \varepsilon) = 1$

Konsistenz besagt anschaulich, dass sich die Verteilung des Schätzers mit zunehmendem Stichprobenumfang n auf den zu schätzenden Wert θ zusammenzieht und dass die Wahrscheinlichkeit, den Parameter um mehr als einen beliebig kleinen Wert ε zu verfehlen, gegen Null strebt, wenn der Stichprobenumfang wächst.

Def.: Ein Schätzer, dessen Verzerrung für $n\to\infty$ gegen Null geht, d. h. $\lim_{n\to\infty}\mathbb{E}\{\hat{\theta}_n-\theta\}=0\,,$

dann ist er konsistent.

heißt asymptotisch unverzerrt oder asymptotisch erwartungstreu.

Satz: Ist $\hat{\theta}_n$ ein Schätzer für θ asymptotisch erwartungstreu und geht seine Varianz gegen Null, d. h., $\lim_{n \to \infty} \mathrm{E}\{\hat{\theta}_n\} = \theta \quad \text{und} \quad \lim_{n \to \infty} \mathrm{V}\{\hat{\theta}_n\} = 0 \;,$

Bem.: Die Bedingungen sind wohl hinreichend aber nicht notwendig dafür, dass ein Schätzer konsistent ist.











Schätzung von Erwartungswerten und Anteilen

Schätzung anderer Parameter

Satz: Sei X_1 , X_2 , ..., X_n eine Stichprobe aus X. Die Verteilung von X ist bis auf $V(X) < \infty$ beliebig. Der Schätzer $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ist erwartungstreu und konsistent für den Erwartungswert $\mu = E(X)$.

Bew.: Erwartungswert und Varianz für die Stichprobenfunktion \overline{X}_n wurden schon hergeleitet. Es gilt:

$$E(\overline{X}_n) = \mu$$
 und $V(\overline{X}_n) = \frac{V(X)}{n}$.

Die Varianz geht mit $n \to \infty$ gegen Null. Damit sind die hinreichenden Bedingungen für Konsistenz erfüllt.

Bem.: Ist die Varianz unbekannt, so muss man auch sie schätzen.

B 1f: \overline{X}_n ist erwartungstreu für μ ; aus den Daten ergibt sich der Schätzwert $\overline{x} = 599,94 \in$.

Satz: X_1 , X_2 , ..., X_n ist eine Stichprobe aus $X \sim B(1, \pi)$. Der Schätzer $\hat{\pi} = \overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ ist erwartungstreu und konsistent für π .

Bem.: Anteile sind auch Mittelwerte, wenn die einzelnen Daten nur die Werte 0 oder 1 sind. So ist es nicht verwunderlich, dass dieselbe Statistik wie bei der Schätzung von Erwartungswerten zum Zuge kommt.

Bew.: Die Statistik $\sum_{i=1}^{n} X_i$ ist binomialverteilt mit n und π ; für Erwartungswert und Varianz gilt:

$$\mathrm{E}\left(\sum_{i=1}^{n} \mathrm{X}_{i}\right) = n \cdot \pi \quad \mathrm{und} \quad \mathrm{V}\left(\sum_{i=1}^{n} \mathrm{X}_{i}\right) = n \cdot \pi \cdot (1-\pi)$$
.

Nach Eigenschaften von Erwartungswert und Varianz erhält man für den Schätzer $\hat{\pi}$:

$$E(\hat{\pi}) = \frac{1}{n} \cdot n \cdot \pi = \pi \quad \text{und} \quad V(\hat{\pi}) = \frac{1}{n^2} \cdot n \cdot \pi \cdot (1 - \pi) = \frac{\pi \cdot (1 - \pi)}{n}$$

Bem.: Diese Varianz muss man gegebenenfalls auch schätzen. Man kann auch Erfahrungswerte oder maximale Abschätzungen dafür einsetzen.

Bezeichne π den Anteil der Raucherinnen unter den weiblichen Studierenden der Universität zu Köln. Eine Studentin wird zufällig ausgewählt. Sei

$$X := \begin{cases} 1 & \text{falls die Studentin raucht} \\ 0 & \text{falls nicht} \end{cases}$$

dann ist $\pi = P(X=1)$. In einer Stichprobe vom Umfang n = 20 befinden sich vier Raucherinnen. Als Schätzwert erhalten wir demnach $\hat{\pi} = \frac{4}{20} = 0.2$.

B 1f: Hier bezeichnet X_i die Ausgaben des *i*-ten Studierenden in der Stichprobe. Um $\pi = P(X > 1000)$ zu schätzen, setzen wir

$$\mathbf{Y}_i := \begin{cases} 1 & \text{falls } \mathbf{X}_i > 1000 \\ 0 & \text{falls nicht} \end{cases}.$$

Aus den Daten ergibt sich für den Schätzer $\hat{\pi} = \overline{Y}$ der Wert 0,14.

Ebenso schätzt man etwa die Wahrscheinlichkeit $\pi_{400} = P(X < 400)$ durch $\hat{\pi}_{400} = 0.34$.

Bem.: Bei endlicher Grundgesamtheit liefert Ziehen *mit* Zurücklegen eine einfache Zufallsstichprobe $X_1, X_2, ..., X_n$ aus X. Beim Ziehen *ohne* Zurücklegen sind dagegen die $X_i, i=1,2,...,n$ abhängig (siehe Kapitel 4). Schätzt man in diesem Fall π mit $\hat{\pi}$, so ergibt sich als Varianz der Schätzfunktion

$$V(\hat{\pi}) = \frac{\pi \cdot (1-\pi)}{n} \cdot \frac{N-n}{N-1},$$

also ein kleinerer Wert im Vergleich zum Ziehen mit Zurücklegen, da der Korrekturfaktor kleiner als 1 ist.

Kriterien für gute Schätzungen

1

- Schätzung von Erwartungswerten und Anteilen
- Schätzung anderer Parameter

Satz: Sei $X_1, X_2, ..., X_n$ eine Stichprobe aus X. Die Verteilung von X ist bis auf $\sigma^2 = V(X) < \infty$ beliebig. Der Schätzer $S^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$ für σ^2 ist asymptotisch erwartungstreu und konsistent.

Bew.: Wir berechnen den Erwartungswert:

$$E\{S^2\} = E\{\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2\} = \frac{1}{n} \cdot \sum_{i=1}^{n} E\{X_i^2\} - E\{\overline{X}^2\}$$

Durch Umstellung im Verschiebungssatz für die Varianz von X_i bzw. \overline{X} erhält man

$$E\{X_i^2\} = \sigma^2 + \mu^2 \text{ und } E\{\overline{X}\} = \frac{\sigma^2}{n} + \mu^2$$

und damit

$$E\{S^{2}\} = \frac{1}{n} \cdot \sum_{i=1}^{n} (\sigma^{2} + \mu^{2}) - (\frac{\sigma^{2}}{n} + \mu^{2}) = \sigma^{2} - \frac{\sigma^{2}}{n} = \frac{n-1}{n} \sigma^{2}.$$

Der letzte Ausdruck strebt mit $n \to \infty$ gegen σ^2 .

Bem.: Lästiger gestaltet sich der Beweis, dass die Varianz dieses Schätzers gegen Null strebt.

Bem.: Offensichtlich ist die korrigierte Stichprobenvarianz

$$S^{*2} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{n}{n-1} \cdot S^2$$

erwartungstreu für σ^2 . Die Konsistenz bleibt durch die Skalierung erhalten.

B 1f: Aus den Daten ergibt sich der Schätzwert $s = \sqrt{s^2} = 322,60 \in$.

Satz: X_1 , X_2 , ..., X_n ist eine Stichprobe aus $X \sim B(1, \pi)$. Der Schätzer $S^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$ für $V(X) = \pi \cdot (1 - \pi)$ ist erwartungstreu und konsistent.

Bem.: Aufgrund der besonderen Werte 0 und 1 für die Zufallsvariablen kann man den Schätzer vereinfachen: $S^2 = \overline{X} \cdot (1 - \overline{X}) = \hat{\pi} \cdot (1 - \hat{\pi}).$

Def.: Sei X_1 , X_2 , ..., X_n eine Stichprobe aus X. Das kleinste der X_i bezeichnet man mit $X_{(1)}$, das zweitkleinste mit $X_{(2)}$ und so weiter bis zum Größten $X_{(n)}$.

Die Zufallsvariablen $X_{(1)}$, $X_{(2)}$, ..., $X_{(n)}$ heißen *Ordnungsstatistiken*. Sie sind aufsteigend geordnet: $X_{(1)}$, $\leq X_{(2)} \leq ... \leq X_{(n)}$.

Mithilfe der Ordnungsstatistiken definiert man für 0 das <math>p-Quantil der Stichprobe als

$$\hat{x}_p \coloneqq \begin{cases} \mathbf{X}_{(np)} & \text{falls } np \text{ ganzzahlig} \\ \mathbf{X}_{([np]+1)} & \text{sonst} \end{cases}$$

Bem.: [n p] ist der ganzzahlige Teil von n p. Diese Definition des Stichprobenquantils entspricht der des Ouantils in der beschreibenden Statistik.

Satz.: Sei X_1 , X_2 , ..., X_n eine Stichprobe aus X. Das p-Quantil der Stichprobe \hat{x}_p ist konsistenter Schätzer für das unbekannte Quantil x_p von X.

Bem.: Im Allgemeinen ist dieser Schätzer nicht unverzerrt.

B 1f: Aus den Daten sollen die Quartile von X geschätzt werden. Es ist $\hat{x}_{0,25} = 336$, $\hat{x}_{0,5} = 561$, $\hat{x}_{0,75} = 797$. Die Ausgaben, die den "mittleren" 50% der Studierenden entsprechen, liegen zwischen 336 und 797 €.

Satz.: Sei (X_1, Y_1) , (X_2, Y_2) , ..., (X_n, Y_n) einfache Stichprobe aus (X, Y). Den theoretischen Korrelationskoeffizienten ρ_{XY} schätzt man durch den Stichproben-Korrelationskoeffizienten

$$\hat{\rho}_{XY} = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{S_x \cdot S_y}.$$

Dieses Schätzer ist asymptotisch erwartungstreu und konsistent.

Bem.: Wir setzen voraus, dass jedes der n Paare (X_i, Y_i) , dieselbe gemeinsame bivariate Verteilung besitzt und für verschiedene Indizes i voneinander unabhängig sind. Insbesondere hat dann jedes der Paare den gleichen Korrelationskoeffizienten ρ_{XY} . Im Allgemeinen hat dieser Schätzer einen Bias, der von Null verschieden ist, der jedoch mit $n \to \infty$ verschwindet.





5 Schätzverfahren für Parameter

Manfred Borovcnik

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



5.1 Punktschätzung

5.2 Konstruktionsprinzipien für Punktschätzer

5.3 Intervallschätzungen

5.4 Ergänzungen

Will man den unbekannten Mittelwert eines Merkmals in einer Grundgesamtheit kennen, so könnte man alle Werte erheben und das Mittel bestimmen. Normalerweise hat man aber nur die Information über einige wenige Merkmalswerte. Dann "schätzt" man mit dem Mittelwert der Daten den unbekannten Mittelwert der Grundgesamtheit. Die Vorgangsweise bekommt ihren Sinn, wenn die Daten als Realisation einer Stichprobe aufgefasst werden können. Dann nämlich kann man die Verallgemeinerung der partiellen Information von den wenigen Daten auf die unbekannte Grundgesamtheit rechtfertigen, indem man diesem Schätzverfahren "gute Eigenschaften" zusprechen kann. Man kann dann auch aus den Daten ein Intervall bestimmen, das einerseits den unbekannten Wert mit "hoher Sicherheit" einschließt und das anderseits die Genauigkeit der Information aus den Daten widerspiegelt. Welche Statistiken geeignet sind für bestimmte Parameter, wird mit Hilfe der Wahrscheinlichkeitsrechnung beurteilt.

5.2 Konstruktionsprinzipien für Punktschätzer

- Momentenmethode
- Maximum-Likelihood (ML)-Methode
- Vergleich von Momenten- und ML-Methode

Bei manchen Schätzproblemen ist es besonders einfach, einen geeigneten Schätzer anzugeben. So liegt es nahe, den Erwartungswert einer Zufallsvariablen durch das Stichprobenmittel zu schätzen. Bei anderen Schätzproblemen ist das Auffinden eines geeigneten Schätzers schwieriger. Es gibt jedoch generelle Prinzipien, nach denen man Punktschätzer konstruieren kann. Zwei allgemeine Vorgehensweisen, die Momenten-Methode und die ML-Methode werden hier dargestellt. Das Bayes-Prinzip wird im 3. Abschnitt bei den Intervallschätzern eingeführt, die Kleinste-Quadrate-Methode wird in Kapitel 7 verwendet.

Momentenmethode



- Maximum-Likelihood (ML)-Prinzip
- Vergleich von Momenten- und ML-Methode

Def.: X eine Zufallsvariable mit $\mu = E(X)$. Unter Moment bzw. zentrierten Moment k-ter Ordnung von X versteht man die Erwartungswerte

$$\mu_k = E(X^k)$$
 bzw. $\widetilde{\mu}_k = E\{(X - \mu)^k\}$

Erwartungswert und Varianz sind demnach Moment 1. Ordnung bzw. zentriertes Moment 2. Ordnung. B: Schiefe und Wölbung sind zentrierte Momente 3. und 4. Ordnung von X.

Bem.: Die Tilde drückt hier immer die Zentrierung der Momente aus.

Def.: X_1 , X_2 , ..., X_n eine Stichprobe aus X, so heißt

$$\hat{\mu}_k = \frac{1}{n} \cdot \sum_{i=1}^n X_i^k$$
 Stichprobenmoment k-ter Ordnung.

B.:
$$\hat{\mu}_1 = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$
, $\hat{\mu}_2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \overline{X})^2$, $\hat{\mu}_3 = \frac{1}{n} \cdot \sum_{i=1}^n (\frac{X_i - \overline{X}}{S})^3$, $\hat{\mu}_4 = \frac{1}{n} \cdot \sum_{i=1}^n (\frac{X_i - \overline{X}}{S})^4$

sind der Reihe nach Stichprobenmomente 1. bis 4. Ordnung, mit Ausnahme des ersten sind alle zentriert.

Def. (Momentenmethode): X_1 , X_2 , ..., X_n eine Stichprobe aus X. Ist der zu schätzende Parameter θ

- i) ein Moment der Verteilung von X, so nimmt man das entsprechende Moment der Stichprobe: $\theta = \mu_k$, dann $\hat{\theta} = \hat{\mu}_k$
- ii) eine (umkehrbar eindeutige) Funktion $\theta = g(\mu_k)$ eines Stichprobenmoments, so nimmt man $\hat{\theta} = g(\hat{\mu}_k)$

Bem.: Im zweiten Fall schätzt man erst das Moment k-ter Ordnung, dann setzt man es in diese Funktion ein.

Bem.: Momentenschätzer sind in allen relevanten Fallen konsistent, jedoch im Allgemeinen nicht erwartungstreu.

Den Erwartungswert schätzt man nach dieser Methode durch das 1. Moment der Stichprobe, das ist das arithmetische Mittel; die Wölbung schätzt man durch das 4. zentrierte Moment der Stichprobe:

$$\hat{\mu} = \hat{\mu}_1$$
, $\hat{\gamma}_2 = \hat{\widetilde{\mu}}_4$.

B 3: Sei X exponential verteilt mit Parameter $\lambda > 0$, d. h. $X \sim \text{Exp}(\lambda)$. Weil $E(X) = \frac{1}{\lambda}$ gilt,

lässt sich λ als Funktion von E(X) darstellen:

$$\lambda = \frac{1}{\mathrm{E}(\mathrm{X})} = \frac{1}{\mu_{\mathrm{l}}} = \mathrm{g}(\mu_{\mathrm{l}}); \text{ und } \hat{\lambda} = \mathrm{g}(\hat{\mu}_{\mathrm{l}}) = \frac{1}{\hat{\mu}_{\mathrm{l}}} = \frac{1}{\overline{\mathrm{X}}} \text{ ist ein Momentenschätzer für } \lambda.$$

B 4: Bei einer Pareto-verteilten Zufallsvariablen $X \sim Par(\alpha, c)$ sei c gegeben. Der Parameter α soll geschätzt werden. Wir unterstellen $\alpha > 1$; dann ist der Erwartungswert endlich und gleich

$$E(X) = \frac{\alpha \cdot c}{\alpha - 1}.$$

Durch Umformung folgt

$$\alpha = \frac{E(X)}{E(X) - c} = g(\mu_1)$$

Mit dem Stichprobenmittel $\hat{\mu}_1 = \overline{X}$ konstruieren wir daraus einen Momentenschätzer für α :

$$\alpha = \frac{\hat{\mu}_1}{\hat{\mu}_1 - c} = \frac{\overline{X}}{\overline{X} - c}$$

B 5: Aufgrund einer Stichprobe X_1 , X_2 , ..., X_n aus $X \sim S(\alpha, \beta)$, $\alpha < \beta$, sollen die beiden Parameter geschätzt werden. Wegen

$$E(X) = \frac{\alpha + \beta}{2}$$
 und $V(X) = \frac{1}{12} \cdot (\beta - \alpha)^2$

gilt

$$\alpha = E(X) - \sqrt{3 \cdot V(X)}$$
, $\beta = E(X) + \sqrt{3 \cdot V(X)}$.

Daraus erhält man die Momentenschätzer: $\hat{\alpha} = \overline{X} - \sqrt{3 \cdot S^2}$, $\beta = \overline{X} + \sqrt{3 \cdot S^2}$.









☑ Vergleich von Momenten- und ML-Methode

Die Methode wird zuerst allgemein beschrieben, dann in einigen Beispielen angewendet. Schließlich werden Eigenschaften der so gewonnenen Schätzer besprochen.

Def.: Eine *parametrische Verteilungsannahme* ist: Die Verteilungsfunktion der Zufallsvariablen X gehört zur Klasse $\mathcal{F} = \{F_{\theta}(x) | \theta \in \Theta\}$.

Dabei bezeichnet θ einen (oder mehrere) unbekannte Parameter. Die Menge Θ der Werte, die für θ in Betracht gezogen werden, nennt man *Parameterraum*.

Bem.: Im parametrischen Modell wird der *Typ der Verteilung* festgelegt. Zur vollständigen Beschreibung der Verteilung von X fehlt nur die Kenntnis des Parameters.

B 6: Im Beispiel mit der Pannenhäufigkeit (vgl. Beispiel 6 in Kapitel 4) ist das Auftreten einer Panne bei einem zufällig ausgewählten Fahrzeug Bernoulli-verteilt mit Parameter π , $X \sim B(1, \pi)$, $0 < \pi < 1$. Dies stellt eine parametrische Verteilungsannahme dar. Hier ist $\theta = \pi$ und $\Theta = (0, 1)$.

B 3f: Hier haben wir als parametrische Verteilungsannahme $X \sim \text{Exp}(\lambda)$ vorausgesetzt; die Klasse der möglichen Verteilungen von X lautet explizit

$$\mathcal{F} = \{ F_{\lambda}(x) = 1 - e^{-\lambda x} \mid x \ge 0, \ \lambda \in (0, \infty) \}, \text{ mit } \theta = \lambda \text{ und } \Theta = (0, \infty)$$

Def.: X_1 , X_2 , ..., X_n eine Stichprobe aus X. Die gemeinsame Wahrscheinlichkeitsfunktion bzw. Dichte der Stichprobe als Funktion des Parameters θ bei gegebenen Beobachtungen x_1 , x_2 , ..., x_n

$$L(\theta) = f(x_1, x_2, \dots, x_n \mid \theta) = f(x_1 \mid \theta) \cdot f(x_2 \mid \theta) \cdot \dots \cdot f(x_n \mid \theta) \cdot \dots \cdot f(x_n \mid \theta)$$

heißt Likelihood-Funktion.

Einen Wert $\hat{\theta}$, der die Likelihood-Funktion bei gegebenen Beobachtungen $x_1, x_2, ..., x_n$ maximiert,

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta),$$

nennt man ML - Schätzwert für den Parameter θ . Der entsprechende Schätzer heißt ML - Schätzer.

Bem.: Anstatt die Likelihood-Funktion zu maximieren, ist es meist einfacher, ihren Logarithmus, das ist die Loglikelihood-Funktion

$$l(\theta) = \ln(L(\theta)) = \sum_{i=1}^{n} f(x_i | \theta),$$

zu maximieren. Die Stelle des Maximums der Likelihood-Funktion stimmt mit der der Loglikelihood-Funktion überein, denn der Logarithmus ist eine streng monoton wachsende Funktion.

B 6f: Es ist $X_i \sim B(1, \pi)$, $0 < \pi < 1$, für alle *i*. Das Ergebnis von n = 10 Beobachtungen lautet

$$x_1, x_2, \dots, x_n = (0,0,0,1,0,0,0,0,0,0)$$
.

Die Wahrscheinlichkeit für dieses Ergebnis beträgt, falls π zutrifft

$$f(x_1, x_2, ..., x_n | \pi) = \pi^1 \cdot (1 - \pi)^9$$
.

Nach der ML-Methode bestimmt man nun jenen Wert $\hat{\pi}$ als Schätzwert für diesen Parameter, bei dem diese Wahrscheinlichkeit maximal ist. Bei gegebenen Beobachtungen ist die genannte Wahrscheinlichkeit eine Funktion des unbekannten Parameters. Man maximiert die Loglikelihood-Funktion

$$l(\pi) = \ln(\pi^1 \cdot (1-\pi)^9) = \ln(\pi) + 9 \cdot \ln(1-\pi)$$

indem man ableitet und die Ableitung gleich Null setzt:

$$l'(\pi) = \frac{1}{\pi} - 9 \cdot \frac{1}{1 - \pi} = 0$$
.

Man erhält $\hat{\pi} = 0,1$. Da die zweite Ableitung überall negativ ist, bestimmt die Lösung ein globales Maximum.

Satz (Invarianz von ML-Schätzern gegen Transformationen): Ist $\hat{\theta}$ der ML-Schätzer für einen bestimmten Parameter θ , dann ist für jede streng monotone Funktion g() der ML-Schätzer für g(θ) durch g($\hat{\theta}$) gegeben.

Bew.: Die Lage des Maximums wird durch streng monotone Funktionen nicht verändert.

B: Wenn $\hat{\theta}$ der ML-Schätzer für die Varianz einer bestimmten Verteilung ist, dann ist $\sqrt{\hat{\theta}}$ der ML-Schätzer für ihre Standardabweichung.

Bem.: Um im konkreten Fall den ML-Schätzer zu bestimmen, muss ein Maximierungsproblem gelöst werden. In vielen einfachen Schätzproblemen – z.B. bei der Schätzung eines Erwartungswertes – lässt sich allerdings das globale Maximum der Loglikelihood-Funktion in Abhängigkeit von $x_1, x_2, ..., x_n$ explizit angeben.

Satz (ML für Bernoulli-Verteilungen):

Sei $X_i \sim B(1, \pi)$, $0 < \pi < 1$ und X_1 , X_2 , ..., X_n eine Stichprobe von X.

Dann ist $\hat{\pi} = \overline{X}$ ML-Schätzer von π .

Bem.: Der ML-Schätzer ist gleich der relativen Häufigkeit der "Erfolge" im Bernoulli-Experiment.

Bew.: Die gemeinsame Wahrscheinlichkeitsfunktion der Stichprobe lautet, falls π zutrifft:

$$L(\pi) = \prod_{i=1}^{n} \pi^{x_i} \cdot (1-\pi)^{1-x_i} = \pi^{\sum x_i} \cdot (1-\pi)^{\sum (1-x_i)} = \pi^z \cdot (1-\pi)^{n-z},$$

wobei z für die Summe der Daten x_i steht. Die Loglikelihood-Funktion ergibt sich als

$$l(\pi) = z \cdot \ln(\pi) + (n-z) \cdot \ln(1-\pi).$$

Ihre ersten beiden Ableitungen betragen

$$l'(\pi) = z \cdot \frac{1}{\pi} + (n-z) \cdot \frac{1}{1-\pi}$$
 sowie $l''(\pi) = -z \cdot \frac{1}{\pi^2} - (n-z) \cdot \frac{1}{(1-\pi)^2}$.

Um das Maximum zu bestimmen, wird die erste Ableitung gleich Null gesetzt, als Lösung ergibt sich: $\hat{\pi} = \overline{x}$.

Da die zweite Ableitung für alle π negativ ist, ist die Lösung ein lokales Maximum, und zwar das einzige im Inneren des Parameterraums (0,1). In beiden Randpunkten geht die Loglikelihood-Funktion gegen $-\infty$. Es folgt, dass die Lösung gleichzeitig das eindeutige globale Maximum liefert.

Satz (ML für Exponentialverteilungen):

Sei $X \sim \text{Exp}(\lambda)$, $\lambda > 0$ und X_1 , X_2 , ..., X_n eine Stichprobe von X.

Dann ist $\hat{\lambda} = \frac{1}{X}$ ML-Schätzer von λ .

Bew.: Die gemeinsame Dichte der Stichprobe in Abhängigkeit von λ lautet

$$L(\lambda) = (\lambda e^{-\lambda x_1}) \cdot (\lambda e^{-\lambda x_2}) \cdot \dots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n \exp(-\lambda \sum_{i=1}^n x_i).$$

Die Loglikelihood-Funktion ist daher

$$l(\lambda) = n \cdot \ln(\lambda) - (\lambda \sum_{i=1}^{n} x_i).$$

Wieder werden die beiden ersten Ableitungen bestimmt:

$$l'(\pi) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$
 und $l''(\pi) = -\frac{n}{\lambda^2}$.

Durch Null-Setzen der ersten Ableitung erhält man $\hat{\lambda} = \frac{1}{\bar{x}}$, was wegen der überall negativen zweiten Ableitung tatsächlich ein globales Maximum darstellt.

Satz (ML für Normalverteilungen):

Sei $X \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbf{R}$ und $\sigma^2 > 0$ und $X_1, X_2, ..., X_n$ eine Stichprobe von X. Dann ist $(\hat{\mu}, \hat{\sigma}^2) = (\overline{X}, S^2)$ der ML-Schätzer für (μ, σ^2) .

Bew.: Die gemeinsame Dichte der Stichprobe in Abhängigkeit von (μ, σ^2) lautet

$$L(\mu, \sigma^{2}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\frac{x_{1} - \mu}{\sigma})^{2}\right\} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\frac{x_{2} - \mu}{\sigma})^{2}\right\} \cdot ... \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\frac{x_{n} - \mu}{\sigma})^{2}\right\}$$
$$\frac{1}{\sigma^{n}(\sqrt{2\pi})^{n}} \exp\left\{-\frac{1}{2\sigma^{2}}\sum_{i=1}^{n}(x_{i} - \mu)^{2}\right\}$$

Logarithmieren ergibt:

$$l(\mu, \sigma^2) = -n \ln(\sigma) - n \ln(\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Die beiden partiellen Ableitungen sind:

$$\frac{\partial}{\partial \mu} l(\mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^n 2(x_i - \mu) \qquad \frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2$$

Null-Setzen der beiden Ableitungen liefert die Lösungen

$$\hat{\mu} = \overline{x} \text{ und } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 = s^2.$$

Eine Überprüfung der zweiten Ableitungen sowie der Ränder des Parameterraums ergibt, dass tatsächlich das globale Maximum vorliegt.









- Maximum-Likelihood (ML)-Methode
- **◄ Vergleich von Momenten- und ML-Methode**

Die folgende Tabelle enthält Momenten- und ML-Schätzer für die Parameter einiger Verteilungen. In jedem Fall liegt eine Stichprobe X_1 , X_2 , ..., X_n aus X zugrunde. In einigen Fällen stimmen die beiden Schätzer überein. Die ML-Schätzer werden in der Regel durch Differenzieren hergeleitet. Ausnahmen bilden die Rechteck- und Pareto-Verteilungen, weil deren Likelihood-Funktion nicht differenzierbar ist.

Verteilung	Momentenschätzer	ML-Schätzer				
$X \sim B(1, \pi)$	$\hat{\pi} = \overline{X}$					
$X \sim Po(\mu)$	$\hat{\mu} = \overline{X}$					
$X \sim Exp(\lambda)$	$\hat{\lambda} = \frac{1}{\overline{X}}$					
$X \sim N(\mu, \sigma^2)$	$\hat{\mu} = \overline{X}, \hat{\sigma}^2 = S^2, \hat{\sigma} = \sqrt{S^2}$					
$X \sim G(\pi)$	$\hat{\pi} = \frac{1}{\overline{X}}$					
$X \sim Par(\alpha, c)$	$\hat{\alpha} = \sqrt{\frac{\overline{X}^2}{S^2} + 1} + 1$ $\hat{c} = \frac{\hat{\alpha} - 1}{\hat{\alpha}} \cdot \overline{X}$	$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln \frac{X_i}{c}}$ $\hat{c} = \min \{X_1, X_2, \dots, X_n\}$				
$X \sim Par(\alpha, c)$, c be- kannt	$\hat{\alpha} = c \cdot \frac{\overline{X}}{\overline{X} - 1}$	$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln \frac{X_i}{c}}$				
$X \sim S(\alpha, \beta)$	$\hat{\alpha} = \overline{X} - \sqrt{3} S^2$ $\hat{\beta} = \overline{X} + \sqrt{3} S^2$	$\hat{\alpha} = \min \{X_1, X_2,, X_n\}$ $\hat{\beta} = \max \{X_1, X_2,, X_n\}$				
Eigenschaften	MM	ML				
in der Regel	konsistent					
häufig	nicht erwartungstreu, aber asymptotisch erwartungstreu					
für große Stichprobenumfange <i>n</i>	approximativ normalverteilt: $\hat{\theta}_n \stackrel{approx}{\sim} N(\theta, \frac{\sigma^2(\theta)}{n})$ $\frac{\sigma^2(\theta)}{n}$ näherungsweise die Varianz des Schätzers					
Varianz des Schätzers		kleiner				
Bestimmte Klassen		kleinste mögliche Varianz				
Verteilungsannahmen	keine	Parametrische Klasse				



5 Schätzverfahren für Parameter

Manfred Borovcnik

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



- 5.1 Punktschätzung
- 5.2 Konstruktionsprinzipien für Punktschätzer

5.3 Intervallschätzungen

5.4 Ergänzungen

Will man den unbekannten Mittelwert eines Merkmals in einer Grundgesamtheit kennen, so könnte man alle Werte erheben und das Mittel bestimmen. Normalerweise hat man aber nur die Information über einige wenige Merkmalswerte. Dann "schätzt" man mit dem Mittelwert der Daten den unbekannten Mittelwert der Grundgesamtheit. Die Vorgangsweise bekommt ihren Sinn, wenn die Daten als Realisation einer Stichprobe aufgefasst werden können. Dann nämlich kann man die Verallgemeinerung der partiellen Information von den wenigen Daten auf die unbekannte Grundgesamtheit rechtfertigen, indem man diesem Schätzverfahren "gute Eigenschaften" zusprechen kann. Man kann dann auch aus den Daten ein Intervall bestimmen, das einerseits den unbekannten Wert mit "hoher Sicherheit" einschließt und das anderseits die Genauigkeit der Information aus den Daten widerspiegelt. Welche Statistiken geeignet sind für bestimmte Parameter, wird mit Hilfe der Wahrscheinlichkeitsrechnung beurteilt.

5.3 Intervallschätzung

- Konfidenzintervalle für Erwartungswerte und Varianzen
- Konfidenzintervalle für Anteile
- Weitere Fragestellungen: Genauigkeit und Mindeststichprobenumfang

Bei der Punktschätzung wird ein einzelner unbekannter Parameter durch *eine* Zahl, den Schätzwert, geschätzt. Ermittelt man aus der Stichprobe zwei Statistiken, die dann ein *Intervall* bilden, das den unbekannten Parameter mit hoher Wahrscheinlichkeit enthält, so spricht man von Intervallschätzung. Der folgende Abschnitt behandelt die Grundbegriffe der Intervallschätzung und einige spezielle Methoden, um Intervallschätzer zu bestimmen.

Konfidenzintervalle für Erwartungswerte und Varianzen





- Konfidenzintervalle für Anteile
- Weitere Fragestellungen: Genauigkeit und Mindeststichprobenumfang

Sei X eine Zufallsvariable, deren Verteilung von θ abhängt. Aus den Daten einer Stichprobe von X soll man ein Intervall berechnen, das den unbekannten Wert mit einer vorgegebenen Wahrscheinlichkeit enthält. Zunächst sei der Parameter der Erwartungswert der Zufallsvariablen.

B 7: Um den unbekannten Erwartungswert einer Zufallsvariablen X zu schätzen, wird eine Stichprobe X_1 , X_2 , ..., X_n aus X beobachtet. \overline{X} ist ein Punktschätzer für $\mu = E(X)$. Da der Schätzer eine Zufallsvariable ist, stellt sich die Frage nach seiner Genauigkeit. Wie ändert sich der Wert, wenn man eine neue Stichprobe erhält? Innerhalb welcher Bandbreiten schwanken wiederholte Schätzungen? Wie genau ist die Schätzung? Zur Ergänzung ist es sinnvoll, ein von der Stichprobe abhängiges Intervall zu bestimmen, das μ mit einer bestimmten Wahrscheinlichkeit einschließt. Wenn X symmetrisch zu seinem Erwartungswert μ verteilt ist, liegt es nahe, ein Intervall der Form

$$[\overline{X} - \varepsilon, \overline{X} - \varepsilon]$$

zu wählen; vgl. Abb.1. Dies ist ein "zufälliges Intervall" der Länge 2ε ; ihm können daher Wahrscheinlichkeiten zugeschrieben werden. Wenn etwa die Einschlusswahrscheinlichkeit gleich 95% sein soll, ist die Länge des Intervalls so festzulegen, dass

$$P(\overline{X} - \varepsilon \le \mu \le \overline{X} - \varepsilon) = 0.95$$

gilt. Mit den realisierten Werten der Stichprobe ist der Punktschätzwert \bar{x} eine Zahl und das Schätzintervall[$\bar{x} - \varepsilon, \bar{x} - \varepsilon$] ein gewöhnliches Intervall.

Abb.1: Schätzintervall für μ in Beispiel 7

Def.: Sei X eine Zufallsvariable, deren Verteilung von θ abhängt, $X_1, X_2, ..., X_n$ eine Stichprobe aus X und $0 < \alpha < 1$. Ein von der Stichprobe abhängiges Intervall $[\hat{\theta}_u, \hat{\theta}_o]$ mit

$$P(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) = 1 - \alpha$$

heißt Konfidenzintervall für θ zum Konfidenzniveau 1-lpha .

Bem.: Das Intervall wird durch die Zufallsvariablen $\hat{\theta}_u$ und $\hat{\theta}_o$ begrenzt. Diese sind Stichprobenfunktionen:

$$\hat{\theta}_u(X_1, X_2, ..., X_n)$$
 und $\hat{\theta}_o(X_1, X_2, ..., X_n)$,

Ist x_1, x_2, \dots, x_n der Wert der Stichprobe X_1, X_2, \dots, X_n , so nennt man das Intervall

$$[\hat{\theta}_u(x_1, x_2, ..., x_n), \hat{\theta}_o(x_1, x_2, ..., x_n)]$$

den Wert des Konfidenzintervalles (oder auch das konkrete Konfidenzintervall).

Bem.: Für jede Realisation der Stichprobe erhält man ein konkretes Konfidenzintervall. Entweder enthält dieses den wahren Parameter, oder es enthält ihn nicht. Dafür, dass das konkrete Intervall den Parameter überdeckt, lässt sich keine Wahrscheinlichkeit angeben. Die Wahrscheinlichkeit $1-\alpha$ haftet nicht dem konkret aus den Daten bestimmten Intervall, sondern dem Schätzverfahren an. Sie lässt sich als relative Häufigkeit interpretieren: Verwendet man für eine große Anzahl von Konfidenzschätzungen Intervalle, die jeweils das Niveau $1-\alpha$ besitzen, so nähert sich die relative Häufigkeit, mit denen die konkreten Intervalle den Parameter überdecken, dem Wert $1-\alpha$.

Bem.: Zu vorgegebenem Niveau $1-\alpha$ soll ein gutes Konfidenzintervall "möglichst schmal" sein. Man wählt das Intervall in der Regel so, dass seine Länge bei gegebenem Konfidenzniveau minimal ist. Wenn X stetig verteilt ist, lässt sich für jedes $\alpha \in (0,1)$ ein Konfidenzintervall angeben, das die Wahrscheinlichkeit $1-\alpha$ exakt annimmt. Wenn X diskret verteilt ist, findet man dagegen meist nur ein Intervall, dessen Wahrscheinlichkeit ungefähr gleich $1-\alpha$ ist.

B 7f: Für die Berechnung eines Konfidenzintervalls für $\mu = E(X)$ muss man die Bandbreite ε aus der Bedingung $P(\overline{X} - \varepsilon \le \mu \le \overline{X} + \varepsilon) = 1 - \alpha$

berechnen. Dazu formt man das Ereignis äquivalent um, ohne seine Wahrscheinlichkeit zu ändern: $P(-\varepsilon \le \overline{X} - \mu \le +\varepsilon) = 1 - \alpha$

Dann ist eine Reihe von Fällen zu unterscheiden.

- i) Man kennt man die Verteilung von $\overline{X} \mu$, insbesondere gelte $X \sim N(\mu, \sigma^2)$; man kennt also auch die Varianz σ^2 . Die Verteilung von \overline{X} ist näherungsweise normal; für eine *approximative* Lösung reicht auch die Kenntnis der Varianz
- ii) Man kennt die Verteilung und zwar als $X \sim N(\mu, \sigma^2)$, kennt aber den Wert von σ^2 nicht.
- iii) Man kennt die Verteilung von X nicht, man kennt aber wenigstens die Varianz σ^2 .

Satz (i): Sei X_1 , X_2 , ..., X_n eine Stichprobe aus X, $X \sim N(\mu, \sigma^2)$ und sei σ^2 bekannt. Das Intervall

$$[\hat{\theta}_u = \overline{X} - \varepsilon, \hat{\theta}_o = \overline{X} + \varepsilon] \text{ mit } \varepsilon = u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

stellt ein Konfidenzintervall für μ zum Niveau $1-\alpha$ dar.

Zusatz: Das Konfidenzintervall gilt approximativ für $n \ge 40$, auch wenn die Normalverteilung für X nicht erfüllt ist.

Bem.: Da σ^2 bekannt ist, lässt sich das Intervall zu jeder Stichprobenrealisation konkret berechnen. Die Bedingung $n \ge 40$ ist als Faustregel zu verstehen, damit der Zentrale Grenzwertsatz für Mittelwerte gilt.

B 8: Der Verkaufspreis für 1 kg Erdbeeren (in €) bei den Einzelhändlern einer Großstadt sei N(μ, σ² = 0,09)-verteilt. Gesucht ist ein Konfidenzintervall für den Verkaufspreis μ zum Konfidenzniveau 95%. Dazu wird eine Stichprobenbefragung von 15 Einzelhändlern durchgeführt. Sei X_i der Verkaufspreis des *i*-ten Händlers. Man braucht nur die konkreten Daten in das allgemeine Intervall einsetzen: n = 15, α = 0,05, u_{1-α/2} = u_{0,975} = 1,96, σ² = 0,09,

das Intervall ergibt sich damit zu

$$[\overline{X}-1,96\cdot\frac{0,3}{\sqrt{15}}, \overline{X}+1,96\cdot\frac{0,3}{\sqrt{15}}] = [\overline{X}-0,15, \overline{X}+0,15]$$

Ergibt eine konkrete Stichprobe den Wert $x = 4,20 \in$, so ist die Realisierung des Konfidenzintervalls für den durchschnittlichen Verkaufspreis [4,05, 4,35].

Satz (ii): Sei X_1 , X_2 , ..., X_n eine Stichprobe aus X, $X \sim N(\mu, \sigma^2)$ und σ^2 unbekannt. Das Intervall

$$[\hat{\theta}_u = \overline{X} - \varepsilon, \quad \hat{\theta}_o = \overline{X} + \varepsilon] \text{ mit } \varepsilon = t(n-1)_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}$$

stellt ein Konfidenzintervall für μ zum Niveau $1-\alpha$ dar.

Zusatz: Für $n \ge 40$ ist ein approximatives Konfidenzintervall gegeben durch $\varepsilon = u_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}$

Bem.: Mit der korrigierten Standardabweichung erhält man $\varepsilon = t(n-1)_{1-\frac{\alpha}{2}} \cdot \frac{S^*}{\sqrt{n}}$.

B 9: Sei X normalverteilt, μ und σ^2 unbekannt. Für den Erwartungswert μ soll ein Konfidenzintervall zum Niveau $1-\alpha=0.99$ bestimmt werden. Dazu wird aus X eine Stichprobe vom Umfang n=12 gezogen, die für das Stichprobenmittel und die korrigierte Stichprobenvarianz die Werte $\bar{x}=20$ bzw. $s*^2=4$ ergibt. Das Konfidenzintervall für μ wird mit Hilfe der t-Verteilung bestimmt. Für 11 Freiheitsgrade und $1-\alpha=0.99$ lautet das $1-\frac{\alpha}{2}$ -Quantil der t-Verteilung: 3,106.

$$\varepsilon = 3{,}106 \cdot \frac{2}{\sqrt{12}} = 1{,}793$$
; mit $\bar{x} = 20$

erhält man daraus das konkrete Konfidenzintervall [18,21, 21,79].

Satz (iii): Sei X_1 , X_2 , ..., X_n eine Stichprobe aus X, die Verteilung ist nicht bekannt, aber σ^2 sei bekannt. Das Tschebyscheff-Intervall

$$[\hat{\theta}_u = \overline{X} - \varepsilon, \quad \hat{\theta}_o = \overline{X} + \varepsilon] \quad \text{mit } \varepsilon = \frac{\sigma}{\sqrt{n\alpha}}$$

stellt ein Konfidenzintervall für μ zum Mindest-Niveau $1-\alpha$ dar.

B 10: Speziell sei $\alpha = 0.05$, $\sigma = 2$, n = 100 und $\bar{x} = 7$.

a) Für das Tschebyscheff-Intervall erhalten wir

$$\varepsilon = \frac{\sigma}{\sqrt{n\,\alpha}} = \frac{2}{\sqrt{100 \cdot 0.05}} = 0.89$$

und als Konfidenzintervall zum Niveau von mindestens 0.95: [7-0.89, 7+0.89] = [6.11, 7.89].

b) Da $n = 100 \ge 40$ hinreichend groß ist, kann man auch das approximative Konfidenzintervall benutzen. Hierfür ergibt sich mit

$$\varepsilon = u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 1,96 \cdot \frac{2}{\sqrt{100}} = 0,39$$

das Intervall [7-0.39, 7+0.39] = [6.61, 7.39].

Bem.: Allgemein sind die Konfidenzintervall über den Zentralen Grenzwertsatz weit enger als über die Tschebyscheffsche Ungleichung. Letztere sind nur ein Anhaltspunkt.

Bew.: (i) Man kennt die Verteilung von $\overline{X} - \mu$, insbesondere gilt $X \sim N(\mu, \sigma^2)$. Um die Wahrscheinlichkeit $P(-\varepsilon \leq \overline{X} - \mu \leq \varepsilon)$.

zu bestimmen, standardisiert man mit $\frac{\sigma}{\sqrt{n}}$ und erhält damit eine N(0,1) -verteilte Größe:

$$P\left(-\frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}} \le \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right);$$

diese Wahrscheinlichkeit soll gleich $1-\alpha$ sein. Man bestimmt daher ε aus folgender Gleichung:

$$2 \cdot \Phi\left(\frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right) - 1 = 1 - \alpha$$

Also gilt: $\frac{\mathcal{E}}{\frac{\sigma}{\sqrt{n}}} = u_{1-\frac{\alpha}{2}}$, mit dem $1-\frac{\alpha}{2}$ -Quantil der Standardnormalverteilung; daraus berechnet man die Band-

breite zu:

$$\varepsilon = u_{1 - \frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Wenn n hinreichend groß ist, gilt nach dem Zentralen Grenzwertsatz

$$\overline{X} \stackrel{approx}{\sim} N(\mu, \frac{\sigma^2}{n})$$
, d. h. $\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \stackrel{approx}{\sim} N(0, 1)$

und die Berechnungen gelten approximativ, auch wenn die Normalverteilungsannahme für X nicht zutrifft.

ii) Man kennt die Verteilung und zwar als $X \sim N(\mu, \sigma^2)$, kennt aber den Wert von σ^2 nicht.

Mit der Standardisierung $\frac{S}{\sqrt{n}}$ erhält man nun für eine t(n-1)-verteilte Größe das Ereignis:

$$-\frac{\mathcal{E}}{\frac{\sigma}{\sqrt{n}}} \le \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \le \frac{\mathcal{E}}{\frac{\sigma}{\sqrt{n}}},$$

dessen Wahrscheinlichkeit $1-\alpha$ betragen soll. Man bestimmt daher ε aus folgender Gleichung:

$$2 \cdot F\left(\frac{\varepsilon}{\frac{S}{\sqrt{n}}}\right) - 1 = 1 - \alpha$$
 (F die Verteilungsfunktion der t-Verteilung)

Also gilt: $\frac{\varepsilon}{\frac{S}{\sqrt{n}}} = t(n-1)_{1-\frac{\alpha}{2}}$; daraus berechnet man die Bandbreite zu: $\varepsilon = t(n-1)_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$.

Für den Zusatz beachte man, dass nach dem Zentralen Grenzwertsatz die standardisierte Zufallsgröße auch mit S im Nenner näherungsweise normalverteilt ist.

iii) Man kennt wohl die Varianz σ^2 , nicht aber die Verteilung von X. Dann kann man die gesuchte Wahrscheinlichkeit mit Hilfe der Tschebyscheffschen Ungleichung abschätzen:

$$P(-\varepsilon \le \overline{X} - \mu \le \varepsilon) \le 1 - \frac{\frac{\sigma^2}{n}}{\varepsilon^2}.$$

Aus der Forderung

$$1 - \frac{\frac{\sigma^2}{n}}{\varepsilon^2} \le 1 - \alpha$$

leitet man nun eine Mindestbandbreite ab: $\frac{\sigma}{\sqrt{n \alpha}} = \varepsilon$.

Bem.: Wenn die Varianz σ^2 einer Verteilung nicht bekannt ist, so muss sie geschätzt werden. Auch dafür kann man für spezielle Verteilungsannahmen für X Intervallschätzungen angeben.

Satz: X_1 , X_2 , ..., X_n eine Stichprobe aus $X \sim N(\mu, \sigma^2)$, wobei μ und σ^2 unbekannt sind. Das Intervall

$$[\theta_u, \theta_o] = \left[\frac{n S^2}{\chi^2 (n-1)_{1-\frac{\alpha}{2}}}, \frac{n S^2}{\chi^2 (n-1)_{\frac{\alpha}{2}}}\right]$$

ist ein Konfidenzintervall für die Varianz σ^2 zum Niveau $1-\alpha$.

Bem.: Man beachte, dass dieses Konfidenzintervall asymmetrisch ist; es hat nicht den Punktschätzer S² für σ^2 zum Mittelpunkt. Symmetrisch ist das Intervall bezüglich der Wahrscheinlichkeiten: Mit Wahrscheinlichkeit $\frac{\alpha}{2}$ liegt das Konfidenzintervall links vom unbekannten Parameterwert (hier der Varianz), und mit gleicher Wahrscheinlichkeit $\frac{\alpha}{2}$ liegt es rechts davon. Für die Standardabweichung σ erhält man ein entsprechendes Konfidenzintervall, indem man statt der obigen Grenzen deren positive Quadratwurzeln einsetzt.

Bew.: Aus einem Satz über normalverteilte Stichproben weiß man

$$\frac{n S^2}{\sigma^2} \sim \chi^2(n-1).$$

Daraus erhält man die Wahrscheinlichkeitsaussage

$$P(\chi^{2}(n-1)_{\frac{\alpha}{2}} \le \frac{nS^{2}}{\sigma^{2}} \le \chi^{2}(n-1)_{1-\frac{\alpha}{2}}) = 1-\alpha$$

Durch Umformen der Ungleichungen erhält man ein äquivalentes Ereignis (mit derselben Wahrscheinlichkeit):

$$\frac{n \, \mathrm{S}^2}{\chi^2 (n-1)_{1-\frac{\alpha}{2}}} \, \, \leq \, \, \sigma^2 \, \leq \, \frac{n \, \mathrm{S}^2}{\chi^2 (n-1)_{\frac{\alpha}{2}}} \, \, .$$







Konfidenzintervalle für Anteile

Weitere Fragestellungen: Genauigkeit und Mindeststichprobenumfang

Um eine unbekannte Wahrscheinlichkeit oder einen Anteil π zu schätzen, verwenden wir eine Stichprobe X_1 , X_2 , ..., X_n aus $X \sim B(1,\pi)$. Der Anteil π wird dann durch den Mittelwert $\hat{\pi} = \overline{X}$ geschätzt. Alle Konfidenzintervalle für Mittelwerte müssen nur für die spezielle Verteilung von X spezifiziert werden. Von den Fällen kommen eigentlich nur (i) und (ii) in der Form der Approximation mittels des Zentralen Grenzwertsatzes in Frage. Die Varianz $\sigma^2 = V(X) = \pi \cdot (1-\pi)$ mit dem Parameter π ist immer auch unbekannt, man kann aber ihre Schätzung durch $S^2 = \hat{\pi} \cdot (1-\hat{\pi})$ aus den Daten umgehen.

Satz: Sei X_1 , X_2 , ..., X_n eine Stichprobe aus $X \sim B(1, \pi)$, $\hat{\pi} = \overline{X}$ eine Schätzung von π , $c = u_{1-\frac{\alpha}{2}}$.

Dann sind die folgenden Intervalle approximative Konfidenzintervalle für π zum Niveau $1-\alpha$:

i) Mit Schätzung der Varianz, falls $n \cdot \hat{\pi} \cdot (1 - \hat{\pi}) > 9$:

$$[\hat{\theta}_u = \hat{\pi} - \varepsilon, \quad \hat{\theta}_o = \hat{\pi} + \varepsilon] \quad \text{mit} \quad \varepsilon = c \cdot \frac{\sqrt{\hat{\pi} \cdot (1 - \hat{\pi})}}{\sqrt{n}}$$

ii) Ohne Schätzung der Varianz, falls $n \ge 20$:

$$[\hat{\theta}_u, \hat{\theta}_o] \text{ mit } \hat{\theta}_{u,o} = \frac{n}{n+c^2} \cdot (\hat{\pi} + \frac{c^2}{2n} \mp c \cdot \sqrt{\frac{\hat{\pi} \cdot (1-\hat{\pi})}{n} + \frac{c^2}{4n^2}}).$$

Bem.: Dabei ist $\hat{\pi}$ die relative Häufigkeit eines "Erfolges".

Bew.: (i) Gemäß dem Zentralen Grenzwertsatz ist die standardisierte relative Häufigkeit für $n \cdot \hat{\pi} \cdot (1 - \hat{\pi}) > 9$ approximativ normalverteilt

$$\widetilde{\pi} = \frac{\widehat{\pi} - \pi}{\frac{\sqrt{\widehat{\pi} \cdot (1 - \widehat{\pi})}}{\sqrt{n}}} = \frac{X - \mu}{\frac{S}{\sqrt{n}}} \stackrel{approx}{\sim} N(0, 1)$$

Es gilt daher

$$P(-c \le \widetilde{\pi} \le c) = 1 - \alpha$$
.

Die Auflösung in eine Doppelungleichung nach π erfolgt wie bei Erwartungswerten und liefert das angegebene Intervall.

(ii) Schon für
$$n \ge 20$$
 gilt: $\widetilde{\pi}' = \frac{\widehat{\pi} - \pi}{\frac{\sqrt{\pi \cdot (1-\pi)}}{\sqrt{n}}} = \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$.

Jetzt macht die Auflösung der Doppelungleichung nach π wegen π im Nenner weit mehr Schwierigkeiten

$$-c \leq \frac{\hat{\pi} - \pi}{\sqrt{\pi \cdot (1-\pi)}} \leq c.$$

Sie ist gleichbedeutend mit

$$\frac{\hat{\pi} - \pi}{\pi \cdot (1 - \pi)} n \le c^2 \quad \text{und} \quad \frac{n + c^2}{n} \cdot \pi^2 - (2\hat{\pi} + \frac{c^2}{n}) \cdot \pi + \hat{\pi}^2 \le 0.$$

Die Auflösung der quadratischen Gleichung nach π liefert zwei Lösungen, welche die angegebenen Grenzen $\hat{\theta}_u$ und $\hat{\theta}_o$ des Konfidenzintervalles darstellen.

Bem: Lösung (ii) kommt ohne Schätzung der Varianz aus den Daten aus. Sie gibt die Information besser wieder, ist aber umständlich auszurechnen. Allerdings ist die Lösung (i) für große *n* ziemlich gut.

B 11: Ein Hersteller lässt die Ausfallwahrscheinlichkeit eines technischen Gerätes bei erhöhter Belastung untersuchen. Um ein Konfidenzintervall für die unbekannte Wahrscheinlichkeit zu bestimmen, wird eine Versuchsreihe durchgeführt. Sei n die Ausfallwahrscheinlichkeit bei einem Versuch. Das Ergebnis der Versuchsreihe ist die Stichprobe π wobei

$$\mathbf{X}_{i} := \begin{cases} 1 & \text{falls } i - \text{ter Versuch ein Ausfall} \\ 0 & \text{falls nicht} \end{cases}$$

Dann ist X_1 , X_2 , ..., X_n eine Stichprobe aus $X \sim B(1, \pi)$. Gesucht ist ein Konfidenzintervall für π zum Niveau $1-\alpha=0.90$. Betrachtet werden die Ergebnisse folgender Versuchsreihen:

a) Bei n = 40 Versuchen traten insgesamt $\sum_{i=1}^{40} x_i = 8$ Ausfälle auf. Daraus folgt die Punktschätzung

$$\hat{\pi}_{40} = \frac{8}{40} = 0.2 \ .$$

Da die Faustregel für das vereinfachte Intervall wegen $40 \cdot 0.2 \cdot 0.8 = 6.4 < 9$ nicht erfüllt ist, muss man Prozedur ohne Schätzung für die Varianz anwenden. Es ist mit

$$c = u_{1-\frac{\alpha}{2}} = u_{0,95} = 1,645$$

$$\hat{\theta}_{u,o} = \frac{40}{40 + 1.645^2} \cdot \left(0.2 + \frac{1.645^2}{2 \cdot 40} \mp 1.645 \cdot \sqrt{\frac{0.2 \cdot 0.8}{40} + \frac{1.645^2}{4 \cdot 40^2}}\right)$$

Das daraus ermittelte Konfidenzintervall [0,1165, 0,3215] ist dem Hersteller zu breit; er lässt deshalb weitere 30 Versuche durchführen.

b) Gegeben seien 30 zusätzliche unabhängige Versuche mit 11 Ausfällen. Für den Gesamtversuch hat man daher 19=8+11 Ausfälle. Als Schätzwert erhält man $\hat{\pi}_{70} = \frac{19}{70} = 0,2714$.

Nun ist die Faustregel wegen $70 \cdot 0.27 \cdot 0.73 = 13.8 > 9$ erfüllt.

Die vereinfachte Formel für das Konfidenzintervall ergibt: [0,1841, 0,3587].

Das exakte Intervall enthält hier nicht viel mehr Information: [0,1937, 0,3661].









- **Weitere Fragestellungen: Genauigkeit und Mindeststichprobenumfang**

In vielen Anwendungen werden Konfidenzintervalle gesucht, deren Genauigkeit $\mp b$ bzw. deren Breite 2b vorgegeben ist. Da diese Breite der in den vorigen Abschnitten hergeleiteten Konfidenzintervalle mit wachsendem Stichprobenumfang n abnimmt, stellt sich die Frage, wie groß n mindestens sein muss, um eine vorgegebene Breite des Konfidenzintervalles einzuhalten, d.h. $\hat{\theta}_0 - \hat{\theta}_u \le 2b$.

Satz (Mindeststichprobenumfang): X_1 , X_2 , ..., X_n sei eine Stichprobe von $X \sim N(\mu, \sigma^2)$. Die Varianz

 σ^2 sei bekannt. Ist der Stichprobenumfang $n \ge n_{\min}$ mit

$$n_{\min} = (u_{1-\frac{\alpha}{2}})^2 \cdot \frac{\sigma^2}{h^2},$$

so ist die Länge des Konfidenzintervalls immer kleiner gleich 2b.

Bew.: Die Länge eines Konfidenzintervalls beträgt $2 \cdot u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$. Man setzt

$$2 \cdot b = 2 \cdot u_{1 - \frac{\alpha}{2}} \frac{\sigma}{\sqrt{n_{\min}}}$$

und löst nach n_{\min} auf. Wenn $n \ge n_{\min}$ ist, hat das Konfidenzintervall höchstens die Breite 2b.

Bem.: Ist die Verteilung der Zufallsvariablen X nicht bekannt, so muss man ein approximatives Konfidenzintervall berechnen. Dessen Länge ist aber gleich dem im Falle der Normalverteilung. Die Lösung ist daher gleich. Man muss nur zusätzlich beachten, dass die Approximation zulässig ist, das heißt, dass nach Faustregel $n \ge 40$ sein muss.

Bem.: Falls σ^2 unbekannt ist, kann man einen Erfahrungwert oder einen sicher zu hohen Wert dafür einsetzen. Man kann einen vorläufigen Schätzwert dafür auch aus einer Pilotstudie gewinnen.

- **B 12:** Für ein endverpacktes Produkt, etwa eine Gewürzmischung, gibt es eine Abnahmekontrolle, die das spezifische Füllgewicht untersucht. Dabei werden n Packungen durch eine Zufallsziehung ausgewählt. Es bezeichne X_i das Füllgewicht (in Gramm) der i-ten Packung der Stichprobe. Die Verteilungsannahme ist $X_i \sim N(\mu, \sigma^2 = 25)$.
 - a) Bei einer Stichprobe der Länge n=20 wird der Mittelwert $\bar{x}=99,35$ beobachtet. Man bestimme ein Konfidenzintervall für μ zum Niveau $1-\alpha=0,99$.
 - b) Wie viele Daten benötigt man, damit das Konfidenzintervall eine Breite von höchstens 1,0 hat.
 - a) Die Grenzen des gesuchten Konfidenzintervalles lauten

$$[\theta_u, \theta_o] = [\overline{X} - u_{1 - \frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + u_{1 - \frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}]$$

Durch Einsetzen von $u_{0.995} = 2,58$ und den Daten erhält man

$$[\theta_u, \theta_o] = [99.35 - 2.58 \cdot \frac{5}{\sqrt{20}}, 99.35 + 2.58 \cdot \frac{5}{\sqrt{20}}] = [99.35 - 2.884, 99.35 + 2.884]$$

Der Wert des Konfidenzintervalles beträgt [96,47, 102,23].

b) Die geforderte maximale Breite ist 2b = 1,0. Die untere Schranke für den Stichprobenumfang beträgt daher

$$n_{\min} = (u_{1-\frac{\alpha}{2}})^2 \cdot \frac{\sigma^2}{b^2} = 2,58^2 \cdot \frac{25}{0,5^2} = 665,6$$
.

Der Stichprobenumfang muss also größer gleich 666 sein.

Satz (Mindeststichprobenumfang bei Binomialverteilung):

 X_1 , X_2 , ..., X_n sei eine Stichprobe von $X \sim B(1, \pi)$. Die Varianz sei aus anderen Untersuchungen als $\sigma^2 = \pi^*(1-\pi^*)$ in der Größenordnung bekannt. Ist der Stichprobenumfang $n \ge n_{\min}$ mit

$$n_{\min} = \max \left\{ \frac{9}{\pi^* (1 - \pi^*)}, \ (u_{1 - \frac{\alpha}{2}})^2 \cdot \frac{\pi^* (1 - \pi^*)}{b^2} \right\},$$

so ist die Länge des Konfidenzintervalls immer kleiner gleich 2b.

Bew.: In der Lösung des allgemeinen Konfidenzintervalls setzt man $\sigma^2 = \pi^* (1 - \pi^*)$ ein. Der zweite Teil der Bedingung soll nur garantieren, dass die Normalapproximation zulässig ist, was nach Faustregel für

$$n \pi^* (1-\pi^*) \ge 9$$
, also für $n \ge \frac{9}{\pi^* (1-\pi^*)}$

der Fall ist. Wenn $n \ge n_{\min}$ ist, hat das Konfidenzintervall höchstens die Breite 2b.

Bem.: Wenn man π^* nicht wenigstens aus einer Pilotstudie kennt, so kann man wegen $\sigma^2 = \pi(1-\pi) \le \frac{1}{2} \cdot \frac{1}{2}$ (!) eine Mindestforderung an den Stichprobenumfang stellen, die auch für den "ungünstigsten" Fall $\pi = \frac{1}{2}$ die geforderte Breite erfüllt:

$$n_{\min} = (u_{1-\frac{\alpha}{2}})^2 \cdot \frac{1}{4h^2}$$





5 Schätzverfahren für Parameter

MANFRED BOROVCNIK

Zusammenfassung: Diese Kurzfassung ersetzt nicht das Studium des Lehrbuches.

Druckfehler bitte rückmelden. Einige Links funktionieren schon.



- 5.1 Punktschätzung
- 5.2 Konstruktionsprinzipien für Punktschätzer
- 5.3 Intervallschätzungen

5.4 Ergänzungen

Will man den unbekannten Mittelwert eines Merkmals in einer Grundgesamtheit kennen, so könnte man alle Werte erheben und das Mittel bestimmen. Normalerweise hat man aber nur die Information über einige wenige Merkmalswerte. Dann "schätzt" man mit dem Mittelwert der Daten den unbekannten Mittelwert der Grundgesamtheit. Die Vorgangsweise bekommt ihren Sinn, wenn die Daten als Realisation einer Stichprobe aufgefasst werden können. Dann nämlich kann man die Verallgemeinerung der partiellen Information von den wenigen Daten auf die unbekannte Grundgesamtheit rechtfertigen, indem man diesem Schätzverfahren "gute Eigenschaften" zusprechen kann. Man kann dann auch aus den Daten ein Intervall bestimmen, das einerseits den unbekannten Wert mit "hoher Sicherheit" einschließt und das anderseits die Genauigkeit der Information aus den Daten widerspiegelt. Welche Statistiken geeignet sind für bestimmte Parameter, wird mit Hilfe der Wahrscheinlichkeitsrechnung beurteilt.

5.4 Ergänzungen

- Anwendungen
- Weitere Prinzipien für Schätzfunktionen

An Finanzmarktdaten werden die Methoden von Punkt- und Intervallschätzungen illustriert. Schließlich werden weitere Prinzipien für "gute" Schätzfunktionen eingeführt.





Weitere Prinzipien für Schätzfunktionen

Satz: (X, Y) sei zweidimensional normalverteilt mit Korrelationskoeffizienten $\rho = \rho_{XY}$, sei (X_1, Y_1) , (X_2, Y_2) , ..., (X_n, Y_n) Stichprobe aus (X, Y); $\hat{\rho}$ der empirische Korrelationskoeffizient.

a) Die *Fishersche z-Transformation* $Z := \frac{1}{2} \ln \frac{1+\hat{\rho}}{1-\hat{\rho}}$ ist approximativ normalverteilt:

$$Z \stackrel{approx}{\sim} N(\frac{1}{2} \ln \frac{1+\rho}{1-\rho}, \frac{1}{n-3})$$

b) Mit
$$A := \ln \frac{1+\hat{\rho}}{1-\hat{\rho}} - 2 \cdot \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}$$
, $B := \ln \frac{1+\hat{\rho}}{1-\hat{\rho}} + 2 \cdot \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}$ ist

$$[\theta_u,\theta_o] = [\frac{e^A-1}{e^A+1}, \ \frac{e^B-1}{e^B+1}] \ \text{ ein n\"{a}herungsweises } 1-\alpha \text{-Konfidenzintervall f\"{u}r } \rho$$

Bem.: Das angeführte Quantil gehört zur Standardnormalverteilung.

B 13: Die Tagesrendite $X_t = \ln K_t - \ln K_{t-1} = \frac{\ln K_t}{\ln K_{t-1}}$ einer Aktie ist die börsentägliche Zuwachsrate ihres

Kurses in Prozent, bereinigt um Dividenden und ähnliche Zahlungen. Die stetige Rendite ist die stetige Wachstumsrate der Kurse. Die Tagesrenditen einer Aktie modellieren wir als Folge von unabhängigen, identisch verteilten Zufallsvariablen. Wir sehen die Tagesrenditen einer Aktie j als Realisation einer Stichprobe aus einer Zufallsvariablen X, an.

Die untersuchten Aktien sind im DAX enthalten: BASF, Hypobank, Bayer, BMW, Deutsche Bank, SAP, Siemens und VW. Der Erhebungszeitraum reicht vom 28.2.1992 bis zum 1.3.2002, das sind n =2611 Börsentage. Die Daten sind als Excel-Datei Aktienbeispiel.xls abrufbar.unter http://www.uni-koeln.de/wiso-fak/wisostatsem/buecher/wrechng schliessende/index.htm

Dabei werden die Punktschätzungen für die mittleren Renditen $\hat{\mu}_i = \overline{X}$ und die Standardabweichungen $\hat{\sigma}_j = S_j$ der einzelnen Aktien j für einzelne Tage auf ein Jahr umgerechnet. Für die Jahresrenditen wer-

den die durchschnittlichen Tagesrenditen mit 250 multipliziert (Börsenjahr = 250 Tage). Dies wird als Annualisierung bezeichnet. Durch Multiplikation der Standardabweichungen der Tagesrenditen mit $\sqrt{250}$ erhält man die Standardabweichungen der Jahresrenditen, das sind die Volatilitäten.

Aktie <i>j</i>	mittlere Jahresrendite der Aktie <i>j</i> (in %)	Volatilität der Aktie <i>j</i> (%)	
BASF	15,05	25,58	
HYPO	7,14	31,76	
BMW	17,08	32,08	
BAY	11,57	26,65	
DB	8,07	28,84	
SAP	32,22	46,42	
SIE	12,00	31,09	
VW	12,13	30,91	

Als nächstes geben wir Intervallschätzungen für die Erwartungswerte und Standardabweichungen der Renditen (pro Jahr). Als konkrete - zumindest approximative – 0,95-Konfidenzintervalle für die mittleren Renditen (in %) erhält man die ersten beiden Spalten der folgenden Tabelle. Wie man sieht, sind die 0,95-Konfidenzintervalle der mittleren Renditen recht breit. Wir schließen daraus, dass sich die mittlere Jahresrendite nur ungenau schätzen lässt, auch wenn man Daten für eine große Zahl von Börsentagen hat. Um Intervallschätzungen für die Standardabweichungen angeben zu können, benötigen wir zusätzlich eine Normalverteilungsannahme $X_j \sim N(\mu_j, \sigma_j^2)$. Die konkreten 0,95-Konfidenzintervalle für die Standardabweichungen (in %) sind in der folgenden Tabelle:

Aktie <i>j</i>	Konfidenzinterval Jahresrend		Konfidenzintervall für die Volatilität (in %)		
	$\hat{ heta}_u$ $\hat{ heta}_o$		$\hat{ heta}_u$	$\hat{ heta}_o$	
BASF	- 0,47	30,56	24,91	26,30	
HYPO	-12,13	26,40	30,93	32,65	
BMW	- 2,37	36,54	31,24	32,98	
BAY	- 4,60	27,74	25,96	27,41	
DB	- 9,42	25,57	28,09	29,66	
SAP	4,06	60,37	45,20	47,72	
SIE	- 6,86	30,86	30,27	31,96	
VW	- 6,62	30,87	30,10	31,78	

Die Tagesrenditen verschiedener Aktien sind im Allgemeinen nicht unabhängig, sondern miteinander korreliert. Für je zwei Aktien kann man den Korrelationskoeffizienten aus der vergangenen Entwicklung der Tagesrenditen schätzen. Unter der Annahme, dass alle Tagesrenditen im Zeitablauf nicht voneinander abhängen, erhält man hier für je zwei Tagesrenditen X_j und X_k , zum Beispiel BASF (j) und Bayer (k), eine einfache Stichprobe der Länge n=2610. Der empirische Korrelationskoeffizient r_{jk} der Aktien j und k stellt einen geeigneten Schätzwert für den theoretischen Korrelationskoeffizienten ρ_{jk} der beiden Aktien dar. In der folgenden Tabelle sind die empirischen Korrelationskoeffizienten der acht Aktien aufgeführt.

r_{jk}	BASF	HYPO	BMW	BAY	DB	SAP	SIE	VW
BASF	1,00							
HYPO	0,37	1,00						
BMW	0,47	0,36	1,00					
BAY	0,71	0,35	0,46	1,00				
DB	0,45	0,54	0,44	0,43	1,00			
SAP	0,27	0,22	0,25	0,26	0,34	1,00		
SIE	0,40	0,34	0,39	0,40	0,50	0,42	1,00	
VW	0,50	0,37	0,56	0,46	0,46	0,32	0,44	1,00

Wie man sieht, sind die Tagesrenditen sämtlicher Aktien positiv korreliert. Bei Unternehmen, die derselben Branche angehören, z.B. BASF und Bayer, ist die Korrelation besonders stark ausgeprägt. Unter der Annahme, dass die Tagesrenditen von je zwei Aktien bivariat normalverteilt sind, lassen sich Konfidenzintervalle für die Korrelationskoeffizienten bestimmen. Für die Tagesrenditen von BASF und HYPO ergibt als 0.95-Intervall für den Korrelationskoeffizienten $[0.37 \pm 0.03] = [0.34, 0.40]$. Für alle anderen Korrelationskoeffizienten ergibt sich ein fast gleicher "Aufschlag" von ± 0.03 zum Schätzwert aus der vorhergehenden Tabelle.

Bem.: Die Annahme der Unabhängigkeit der Tagesrenditen ist nicht unproblematisch: Es lässt sich bei vielen Aktien eine leichte Abhängigkeit der Tagesrenditen feststellen. Es sei darauf hingewiesen, dass die Normalverteilungsannahme bei Tagesrenditen (für die Intervallschätzung der Volatilität) problematisch ist, da diese in der Regel eine erheblich höhere Wölbung als 3 aufweisen und außerdem leicht asymmetrisch verteilt sind.









Weitere Prinzipien für Schätzfunktionen

Um den Erwartungswert μ einer Zufallsvariablen zu schätzen, gibt es viele mögliche Schätzer. Die Frage ist, welchen von zwei alternativ gegebenen Schätzern man vorziehen soll, oder ob es sogar einen – in noch zu präzisierendem Sinne – besten Schätzer gibt. Unter Alltagsbedingungen gibt es andere Kriterien, wie die der Robustheit, welche Schätzfunktionen auszeichnen. Bayes-Schätzer bauen auch Information über den unbekannten Parameter ein, die schon vor der Datenerhebung vorhanden sind.

B 14: $\hat{\mu} = \overline{X}$ ist erwartungstreu für μ bei beliebiger Verteilung F. Andererseits kann z. B. der erste beobachtete Wert ein Schätzwert für den unbekannten Erwartungswert sein. Es ist dann $\widetilde{\mu} = X_1$ Schätzer für μ . Da $E(X_i) = \mu$, ist auch $\widetilde{\mu}$ erwartungstreu. Die Varianz des zweiten Schätzers beträgt $V(X_1) = \sigma^2$, die des ersten beträgt $V(\overline{X}) = \frac{\sigma^2}{n}$. Der zweite Schätzer streut also starker als der erste; beide sind jedoch erwartungstreu.

Def.: Ein Schätzer $\hat{\mu}$ für μ heißt *linear*, wenn die Schätzfunktion linear ist, d.h. wenn

$$\hat{\mu} = \sum_{i=1}^{n} \alpha_i X_i$$

ist mit gewissen reellen Koeffizienten α_i .

Bem.: Offenbar ist $\hat{\mu} = \overline{X}$ ein linearer Schätzer für mit gleichen Gewichten $\alpha_i = \frac{1}{n}$. Aber auch der Schätzer $\widetilde{\mu} = X_1$ ist linear.

Satz: Ein linearer Schätzer $\hat{\mu}$ für μ ist genau dann unverzerrt, wenn $\sum_{i=1}^{n} \alpha_i = 1$.

Bew.: Es gilt nämlich

$$\mathbb{E}(\hat{\mu}) = \mathbb{E}(\sum_{i=1}^{n} \alpha_i \, \mathbf{X}_i) = \sum_{i=1}^{n} \alpha_i \, \mathbb{E}(\mathbf{X}_i) = \mu \sum_{i=1}^{n} \alpha_i$$

Satz (Gauß-Markoff-Theorem): Unter allen linearen erwartungstreuen Schätzern für μ hat $\hat{\mu} = \overline{X}$ die kleinste Varianz.

Bem.: Solche Schätzer werden BLUE – Best Linear Unbiased Estimator – genannt. Wenn man sich auf lineare Schätzer beschränkt, gibt es also nach dem Gauß-Markoff-Theorem keinen Schätzer, der besser wäre als das Stichprobenmittel.

Bew.: Wir berechnen die Varianz eines beliebigen linearen Schätzers und minimieren die Varianz bezüglich der Koeffizienten unter der Nebenbedingung, dass der Schätzer unverzerrt ist.

$$V(\hat{\mu}) = V(\sum_{i=1}^{n} \alpha_i X_i) = \sum_{i=1}^{n} \alpha_i^2 V(X_i) = \sigma^2 \sum_{i=1}^{n} \alpha_i^2 \rightarrow \min$$

Nebenbedingung:
$$\sum_{i=1}^{n} \alpha_i = 1$$

Die Lagrangre-Funktion lautet:

L
$$(\alpha_1, \alpha_2, \dots, \alpha_n, \lambda) = \sigma^2 \sum_{i=1}^n \alpha_i^2 + \lambda \cdot (\sum_{i=1}^n \alpha_i - 1)$$

Die Minimal-Lösung ergibt das arithmetische Mittel; man erhält sie durch Null-Setzen der partiellen Ableitungen.

Def.: Zwei Schätzer $\hat{\theta}_1$ und $\hat{\theta}_2$ für den gleichen Parameter θ seien (zumindest asymptotisch) unverzerrt. Der Quotient ihrer Varianzen

$$\mathrm{RE}\,(\hat{\theta}_1,\hat{\theta}_2) = \frac{\mathrm{V}(\hat{\theta}_2)}{\mathrm{V}(\hat{\theta}_1)}$$

heißt $\mathit{relative}$ $\mathit{Effizienz}$ von $\hat{\theta}_1$ gegenüber $\hat{\theta}_2$.

Ist RE $(\hat{\theta}_1, \hat{\theta}_2) \ge 1$ für alle θ und für mindestens einen Parameterwert echt größer als 1, so heißt $\hat{\theta}_1$ wirksamer als $\hat{\theta}_2$.

Man nennt einen Schätzer bezüglich einer gegebenen Klasse von Schätzern effizient, wenn er gegenüber allen Schätzern dieser Klasse minimale Varianz hat.

Bem.: Der mittlere quadratische Fehler eines Schätzers ist ein Maß für seine Präzision. Bei unverzerrten Schätzern stimmt er mit der Varianz überein. Es wird also – im Wesentlichen die relative Präzision des einen mit der des anderen Schätzer verglichen. Man beachte, dass sowohl die Varianzen und damit auch die relative Effizienz zweier Schätzer von dem unbekannten Wert des Parameters θ abhängen.

Bem.: Nach dem Satz von Gauß-Markoff ist das Stichprobenmittel \overline{X} bezüglich der Klasse aller linearen unverzerrten Schätzer für μ effizient.

B 15: Das Stichprobenmittel X̄ hat als Schätzer für μ eine Reihe von guten Eigenschaften. Dennoch kann seine Verwendung problematisch sein, wenn sich so genannte Ausreißer in der Stichprobe befinden. Ein Ausreißer ist ein extremer Wert der Zufallsvariablen X, d. h. ein Wert, der in einen Bereich sehr kleiner Wahrscheinlichkeit fällt. Ein Ausreißer kann ein fehlerhafter Wert sein, der durch einen Erhebungs-, Übertragungs- oder Aufbereitungsfehler in die Stichprobe gelangt ist; er kann aber auch einen korrekten, lediglich selten auftretenden Beobachtungswert darstellen. In beiden Fällen beeinflusst ein Ausreißer das Stichprobenmittel sehr stark, da sein extremer Wert mit dem gleichen Gewicht wie die übrigen Beobachtungen in die Bildung des Mittelwerts eingeht. Bei einem Verdacht auf Ausreißer ist es deshalb sinnvoll, den Mittelwert erst zu bilden, nachdem man die verdächtigen Werte entfernt hat.

Def.: Für $0 < \alpha < 1$ ist α -getrimmtes Stichprobenmittel ist der Mittelwert jener Daten, die übrig bleiben, wenn man die Stichprobe $X_1, X_2, ..., X_n$ ordnet und die $n\alpha$ größten und $n\alpha$ kleinsten Werte X_i weglässt.

Bem.: Der Wert von α ist dabei im Hinblick auf die konkrete Datensituation und den Anwendungszusammenhang festzulegen. Im Extremfall verwendet man lediglich den Stichprobenmedian.

Bem.: Getrimmte Stichprobenmittel werden robuste Schätzer genannt, da Beobachtungen, die extrem weit vom zu schätzenden Erwartungswert μ liegen, keinen Einfluss auf den Schätzwert haben. Entsprechende robuste Schätzer verwendet man für die Varianz und die höheren Momente.

B 6f: Im Beispiel "Pannenhäufigkeit" sei θ die unbekannte Wahrscheinlichkeit. Es sei angezeigt, das der Parameter zwischen 15 und 40% liegt, und zwar mit derselben Wahrscheinlichkeit(sdichte). Kann man diese Vorinformation über den Parameter bei der Konstruktion eines Schätzers berücksichtigen?

Def.: Unter *Vorinformation über* den *Parameter* θ ist eine Wahrscheinlichkeitsverteilung auf dem Parameterraum zu verstehen. Diese so genannte *a-priori-Verteilung* kann diskret oder stetig sein.

Bem.: Oft ist *vor* der Schätzung bekannt, dass in der konkreten Anwendung nicht jedes θ aus dem Parameterraum in gleicher Weise in Frage kommt. Eine gewisse Schwierigkeit mag darin liegen, eine geeignete Modellierung dieser Vorinformation vorzunehmen. Eine mögliche Form der Vorinformation besteht darin, dass θ auf eine bestimmte Teilmenge des Parameterraums beschränkt ist.

Satz und Def. (Bayes-Theorem in diskreter Version):

Der Parameter θ hat eine diskrete a priori-Verteilung $\pi_0(\theta_i)$ auf der Menge $\Theta = \{\theta_1, \theta_2, \dots\}$.

Die gemeinsame diskrete Dichte von X_1 , X_2 , ..., X_n hängt vom zutreffenden Parameter ab; ist es θ_j , so schreiben wir dafür: $f(x_1, x_2, ..., x_n | \theta_j)$

Die bedingte Verteilung von θ unter der Beobachtung x_1, x_2, \dots, x_n lautet dann:

$$\pi_{1}(\theta_{j} \mid x_{1}, x_{2}, \dots, x_{n}) = P(\theta = \theta_{j} \mid x_{1}, x_{2}, \dots, x_{n}) = \frac{f(x_{1}, x_{2}, \dots, x_{n} \mid \theta_{j}) \cdot \pi_{0}(\theta_{j})}{\sum_{k} f(x_{1}, x_{2}, \dots, x_{n} \mid \theta_{k}) \cdot \pi_{0}(\theta_{k})}$$

Diese Verteilung von $\pi_1(\theta_i)$ nennt man a posteriori-Verteilung.

Bem.: Hat Parameter θ hat eine stetige a priori-Verteilung $\pi_0(\theta)$ auf der Menge Θ und sind die Daten stetig verteilt mit dem Parameter mit Dichtefunktion $f(x_1, x_2, ..., x_n | \theta)$, so bestimmt man die Dichte der a posteriori-Verteilung mit einem Integral im Nenner statt der Summe, die Bayes-Formel lautet dann:

$$\pi_{1}(\theta \mid x_{1}, x_{2}, \dots, x_{n}) = \frac{f(x_{1}, x_{2}, \dots, x_{n} \mid \theta) \cdot \pi_{0}(\theta)}{\int_{\Theta} f(x_{1}, x_{2}, \dots, x_{n} \mid \theta) \cdot \pi_{0}(\theta) \cdot d\theta}$$

Bew.: Anwendung der Bayes-Formel aus Kapitel 1 liefert das Ergebnis.

Bem.: Dabei bezeichnet $f(x_1, x_2, ..., x_n | \theta_k)$ die Wahrscheinlichkeit, dass die Werte $x_1, x_2, ..., x_n$ beobachtet werden, wenn θ_k der wahre Parameter ist. Die Namen *a priori* und *a posteriori* beziehen sich auf die Verteilung für den Parameter θ vor und nach Einbeziehung der Daten.

Def.: Der Parameter hat eine diskrete a priori-Verteilung $\pi_0(\theta_i)$;

$$X_1$$
, X_2 , ..., X_n hat die diskrete Dichte $f(x_1, x_2, ..., x_n | \theta_i)$

die a posteriori-Verteilung lautet $\pi_1(\theta_i | x_1, x_2, ..., x_n)$.

$$\hat{\theta}(\pi_0) := \max_{\theta_i} \left\{ \pi_1(\theta_j \mid x_1, x_2, \dots, x_n) \right\}$$

heißt Bayes-Schätzung von θ unter Einbezug der Vorinformation $\pi_0(\theta_i)$ und den Daten.

Bem.: Die Bayes-Schätzung ist das Maximum der a posteriori-Verteilung des Parameters. Im stetigen Fall kann man das Maximum mit Differentialrechnung bestimmen:

$$\hat{\theta}(\pi_0) := \max_{\theta} \left\{ \pi_1(\theta \mid x_1, x_2, \dots, x_n) \right\}$$

Bem.: Offenbar genügt es, den Zähler des Bruchs im Bayes-Theorem zu maximieren, da der Nenner nicht von θ abhängt.

B 6f: Im Beispiel "Pannenhäufigkeit" sei θ die unbekannte Wahrscheinlichkeit.

a) Wir nehmen zunächst an, dass für θ jeder Wert aus dem Intervall [0, 1] gleichermaßen in Frage kommt Dies führt auf die a-priori-Verteilungsannahme $X \sim S(0, 1)$ mit der a-priori-Dichte

$$\pi_0(\theta) = \begin{cases} 1 & 0 \le \theta \le 1 \\ 0 & \text{sonst} \end{cases}.$$

Es wird $(x_1, x_2, ..., x_n) = (0, 0, 0, 1, 0, 0, 0, 0, 0, 0)$ beobachtet. Gemäß der Bayes-Formel berechnen wir den Zähler der a-posteriori-Dichte von θ

$$f(x_1, x_2, ..., x_n \mid \theta) \cdot \pi_0(\theta) = \begin{cases} \theta^1 \cdot (1 - \theta)^9 \cdot 1 & 0 \le \theta \le 1 \\ 0 & \text{sonst} \end{cases}.$$

Durch Differenzieren dieses Ausdrucks sieht man, dass die a-posteriori- Dichte für $0 < \theta < 0.1$ streng monoton wächst und für $0.1 < \theta < 1$ streng monoton fällt, bei 0.1 also ihr globales Maximum annimmt. Demnach ist bei dieser a-Priori-Verteilung $\hat{\theta} = 0.1$ der Bayes-Schätzer.

b) Wir gehen nun von der Vorinformation aus, dass der wahre Wert von θ zwischen 15% und 40% liegt und weiter nichts über θ bekannt ist. In diesem Fall ist $X \sim S(0,15,0,40)$ eine sinnvolle a-priori-Verteilungsannahme. Die a-priori-Dichte lautet nun

$$\pi_0(\theta) = \begin{cases} 4 & 0.15 \le \theta \le 0.4 \\ 0 & \text{sonst} \end{cases}$$

und der Ausdruck $\theta^1 \cdot (1-\theta)^9 \cdot 4$ ist für $0.15 \le \theta \le 0.4$ zu maximieren. Offenbar liegt das Maximum bei 0.15. Unter dieser Vorinformation ist $\hat{\theta} = 0.15$ der Bayes-Schätzer.



