Einführung in die Statistik für Biologen

Jörg Witte

1997

Inhaltsverzeichnis

Kapitel 1

Endliche Wahrscheinlichkeitstheorie

1.1 Grundbegriffe

Erste wahrscheinlichkeitstheoretische Untersuchungen sind von dem französischen Mathematiker P. S. de Laplace (1749 – 1827) überliefert. Er definierte die Wahrscheinlichkeit für das Eintreten eines Ereignisses als das Verhältnis der Zahl der günstigen Fälle zu der aller möglichen Fällen (siehe [5]). Diese genügt heute nicht mehr den modernen wissenschaftlichen Ansprüchen. Sie werden vielleicht an ein Glückspiel gedacht haben. Problematisch wird es, wenn sie nach der Wahrscheinlichkeit für eine Mädchen– oder Knabengeburt einer schwangeren Frau fragen, wenn das Geschlecht noch nicht feststeht. Sicherlich bereitet es Ihnen keine Schwierigkeiten, von der eventuell für ein Geschlecht diskriminierende Ausdrucksweise "günstig" zu abstrahieren. Solche Abstraktionen werden aber in der Mathematik methodisch durchgeführt. Dafür gibt es zwei Gründe. Zum einen führen Definitionen von Begriffen, die Anleihen aus den nicht immer eindeutigen Vorstellungen aus der alltäglichen Erfahrungswelt machen, zu Verwirrungen. In der Wissenschaftsgeschichte gibt es sogar Beispiele für daraus resultierende Widersprüche. In einer Wissenschaft sollte auch eine größt mögliche Klarheit über die Verwendungsweise der gebrauchten Begriffe hergestellt werden, denn diese gehört auch zum Wissen über den Gegenstand einer Wissenschaft. Zum anderen bedeutet Abstrahieren Absehen vom Unwesentlichen und Hervorheben des Wesentlichen. Dieses verspricht eine größere Palette von Anwendungsmöglichkeiten. Wenn Sie sich nun nochmal nach der Wahrscheinlichkeit für eine Mädchengeburt fragen, so wäre sie nach Laplace 1/2. Statistische Untersuchungen legen aber die Vermutung nahe, daß die Wahrscheinlichkeit geringfügig von 1/2 abweicht. Das ist mit der Laplaceschen Definition nicht so ohne weiteres zu vereinbaren. Völlig unvereinbar damit wäre die Frage nach der Wahrscheinlichkeit für eine gewisse Größe oder Gewicht des neugeborenen Babys. Diese Mängel werden heutzutage durch einen axiomatischen Aufbau der Wahrscheinlichkeitstheorie behoben.

Für eine axiomatische Vorgehensweise hat sich heute in allen Zweigen der Mathematik die Sprache der Mengenlehre als optimal erwiesen und sich daher durchgesetzt. G. Cantor (1845 – 1918) legte den Begriff der Menge folgendermaßen fest:

Definition 1.1.1 *Unter einer Menge versteht man jede Zusammenfassung von bestimmten, wohlunterschiedenen Objekten unsere Anschauung und unseres Denkens, welche Elemente der Menge genannt werden, zu einem Ganzen.*

Wie sie vielleicht unschwer bemerkt haben, entlehnt auch diese Definition Vorstellungen aus der alltäglichen Umgangsprache. Tatsächlich führt sie zu Widersprüchen. [1] Daher bemühten sich Mathematiker erfolgreich um eine Axiomatisierung der Mengenlehre. Es würde jedoch den Rahmen diese Buches sprengen, die Axiome der Mengenlehre zu erläutern. Daher werden wir die Mengenlehre im Cantorschen Sinn verwenden. Diese wird auch naive Mengenlehre genannt. In den meisten Fällen führt die naive Mengenlehre zu keinen Widersprüchen, insbesondere nicht, wenn man es mit endlichen Mengen zu tun hat. Für unsere

[1] Am bekanntesten ist die Russelsche Antinomie. Man bilde die Menge M aller Mengen, die sich nicht selbst als Element enthalten. Wenn nun M sich selbst als Element enthält, dann kann nach Definition M sich nicht selbst als Element enthalten. Und umgekehrt: wenn M sich nicht selbst als Element enthält, dann muß nach Definition M sich selbst als Element enthalten. Das erinnert an die Antinomie vom Dorfbarbier: der Dorfbarbier rasiert genau alle diejenigen, die sich nicht selbst rasieren. Wenn der Dorfbarbier sich selbst rasiert, dann rasiert er sich nach der Tätigkeitsbeschreibung seiner Berufsausübung nicht selbst und umgekehrt.

Zwecke reicht die naive Mengenlehre völlig aus. Es werden insbesondere in diesem Buch keine Begriffe gebraucht, aus denen sich Widersprüche abgeleitet werden können.

In der Wahrscheinlichkeitstheorie werden Aussagen über zufällige Ergebnisse von gewissen Vorgängen getroffen. Ein solcher Vorgang heißt ein Zufallsexperiment.[2] Z. B. kann man an einen Münzwurf denken. Die möglichen Ergebnisse dieses Zufallsexperimentes sind Kopf oder Zahl.

Definition 1.1.2 *Alle möglichen Ergebnisse eines Zufallsexperimentes nennen wir Elementarereignisse. Die Menge aller Elementarereignisse heißt die Ereignismenge. Sie wird i. A. mit* Ω *bezeichnet. Wenn* Ω *nur endlich viele Elemente enthält, dann nennen wir* Ω *eine endliche Ereignismenge. Die Teilmengen einer endlichen Ereignismenge heißen Ereignisse.* [3] *Wir sagen, ein Ereignis* $A \subset \Omega$ *tritt ein, wenn es eine Elementarereignis* $\omega \in A$ *gibt, welches eintritt, also welches als Ergebnis eines Zufallsexperimentes beobachtet wird.* Ω *heißt das sichere Ereignis und* $\emptyset$ *das unmögliche Ereignis.*

Die Ereignismenge ist nicht immer eindeutig bestimmt. Diese hängt davon ab, was wir als Ergebniss eines Zufallsexperiment definieren. So können wir beim Werfen eines Würfels die Augenzahl zu einer Ereignismenge

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$$

zusammenfassen oder wir können uns dafür interessieren, ob die Augenzahl gerade bzw. ungerade ist, d. h. wir erhalten eine Ereignismenge mit zwei Elementen:

$$\Omega = \{\omega_1, \omega_2\}.$$

Weiterhin werden wir jedem Ereignis eine Wahrscheinlichkeit P zuordnen. Im ersten Fall würden wir jedem Elementarereignis die Wahr-

[2] Es braucht sich dabei nicht um ein Experiment im klassischen Sinn zu handeln, also um einen künstlich hervorgerufenen Vorgang, sondern kann auch ein natürlicher Vorgang sein, wie etwa die Vererbung in der "freien" Natur

[3] Bei unendlichen Ereignismengen treten sehr tiefliegende mathematische Probleme auf, so daß es nicht zweckmäßig ist, jede Teilmenge als ein Ereignis zu definieren, da man dann nicht immer in einer konsistenten Art und Weise einem Ereignis eine Wahrscheinlichkeit zuordnen kann. In solchen Fällen ist es sinnvoll, sich ein besonders ausgezeichnetes Mengensystem als Ereignisse zugrundezulegen. Aus didaktischen Gründen werden wir das aber zunächst zurückstellen

scheinlichkeit 1/6 und im zweiten Fall die Wahrscheinlichkeit 1/2 zuordnen. [4] Wir können aber auch im ersten Fall die Wahrscheinlichkeit dafür bestimmen, daß eine gerade Augenzahl gewürfelt worden ist, nämlich:

$$P(\{\omega_2, \omega_4, \omega_6\}) = P(\{\omega_2\})+P(\{\omega_4\})+P(\{\omega_6\}) = 1/6+1/6+1/6 = 1/2.$$

Intuitiv leuchtet es also ein, daß es sinnvoll ist, die Wahrscheinlichkeit eines Ereignisses dadurch zu berechnen, indem man die Wahrscheinlichkeiten der Elementarereignisse dieses Ereignisses aufsummiert. Ebenso erscheint es sinnvoll, dem sicheren Ereignis die Wahrscheinlichkeit eins zuzuordnen und dem unmöglichen Ereignis die Wahrscheinlichkeit null, also $P(\Omega) = 1$ und $P(\emptyset) = 0$. solche Eigenschaften der Wahrscheinlichkeit werden wir axiomatisieren. Zunächst werden wir aber etwas über mengentheoretische Operationen auf den Ereignissen kennenlernen, auch weil dann die Axiomatisierung besser motiviert ist.

Sei im folgenden eine Ereignismenge Ω gegeben und zwei Ereignisse $A, B \subset \Omega$. Wir können die Vereinigung, den Durchschnitt von A und B, sowie das Komplement von A bilden. Diese mengentheoretischen Operationen besitzen gewisse Wahrscheinlichkeitstheoretische Interpretationen. Betrachten wir zunächst die Vereinigung. Sie ist wahrscheinlichkeitstheoretische wie folgt definiert:

$$A \cup B = \{\omega | \omega \in A \text{ oder } \omega \in B\},$$

sie besteht also aus allen Elementen, die in A oder B enthalten sind. Nach der Definition 1.1.2 tritt $A \cup B$ ein, wenn es ein Elementarereignis $\omega \in A \cup B$ gibt, welches eintritt. Da dann dieses Elementarereignis ω ein Element von A oder von B ist, heißt das, daß A oder B eintritt.

Wir erhalten eine analoge wahrscheinlichkeitstheoretische Interpretation für den Durchschnitt:

$$A \cap B = \{\omega | \omega \in A \text{ und } \omega \in B\}.$$

[4] Genaugenommen ordnen wir Wahrscheinlichkeiten Ereignissen, also Mengen zu. Da ein Elementarereignis keine Menge sondern ein Element einer Menge ist, ist es daher nicht sinnvoll, von der Wahrscheinlichkeit eines Elementarereignisses zu sprechen. Es führt aber zu keinen weiteren Konfusionen, wenn wir *in diesem Zusammenhang* ein Elementarereignis mit dem Ereignis identifizieren, das nur dieses Elementarereignis als Element enthält und kein anderes.

Der Durchschnitt tritt dann ein, wenn A und B eintritt, denn ein Elementarereignis $\omega \in A \cap B$ ist eine Element von A und von B.

Schließlich erhalten wir für das Komplement $\overline{A}$:

$$\overline{A} = \{\omega \in \Omega | \omega \notin A\}$$

die Interpretation, daß $\overline{A}$ eintritt, wenn A nicht eintritt. Wir sagen daß A nicht eintritt, wenn es kein Elementarereignis $\omega \in A$ gibt, das eintritt. Dann muß aber das eintretende Elementarereignis in dem Komplement $\overline{A}$ liegen. Bei dieser Argumentation geht wesentlich ein, daß Ω aus *allen* möglichen Ergebnissen eines Zufallsexperimentes besteht.

Schließlich ist es für die Wahrscheinlichkeitstheorie wesentlich, die Wahrscheinlichkeit für das Eintreten von A oder B zu bestimmen, wenn das Eintreten von A und B sich gegenseitig ausschließt, also $P(A \cup B)$ zu bestimmen, wenn $A \cap B = \emptyset$. Eine Vereinigung von zwei elementfremden Mengen wird in der Mengenlehre disjunkt genannt. Intuitiv ist es klar, daß man dann die Wahrscheinlichkeit nach der Formel $P(A \cup B) = P(A) + P(B)$ berechnen kann. Diese Eigenschaft der Wahrscheinlichkeit werden wir axiomatisch fordern. Als Beispiel kann man sich das Lotto vor Augen führen. Vielleicht stellen sie sich nicht nur die Frage, wie groß die Wahrscheinlichkeit ist, sechs Richtige zu haben (die ist ohnehin sehr gering), sondern genau fünf oder sechs Richtige angekreuzt zu haben. Da das Eintreten dieser beiden Ereignisse sich gegenseitig ausschließt, werden sie völlig richtig die Wahrscheinlichkeit für das Eintreten von sechs Richtigen und die Wahrscheinlichkeit für das Eintreten von fünf Richtigen ausrechnen und die zwei Ergebnisse addieren.

Die nun folgende Definition scheint nun sinnvoll für den Aufbau der endlichen Wahrscheinlichkeitstheorie zu sein:

Definition 1.1.3 *Ein endlicher Wahrscheinlichkeitsraum ist ein Paar* (Ω, P)*, wobei* Ω *eine endliche Ereignismenge ist und* P *eine Abbildung*

$$P : \mathcal{P}(\Omega) \longrightarrow [0, 1].$$

$\mathcal{P}(\Omega)$ *bezeichnet hier die Potenzmenge, d. i. die Menge aller Teilmengen von* Ω*.* P *ordnet also jeder Teilmenge aus* Ω *eine reelle Zahl aus dem Intervall* $[0, 1]$ *zu.* P *ist durch folgende Eigenschaften ausgezeichnet:*

1.

$$P(\Omega) = 1 \quad \textit{und} \quad P(\emptyset) = 0,$$

2. für disjunkte $A, B \subset \Omega$ gilt:

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

Diese Eigenschaft von P wird auch Additivität genannt.

Aus der Additivität der Wahrscheinlichkeit ergibt sich ohne weiteres, daß die Wahrscheinlichkeit P schon eindeutig bestimmt ist, wenn die Wahrscheinlichkeit für das Eintreten jedes Elementarereignisses feststeht. Beim Würfeln wäre diese etwa $1/6$ für jedes Elementarereignis, wenn Ω die Menge der mögliche Augenzahlen ist. Die Wahrscheinlichkeit für das Eintreten eines beliebigen Ereignisses $A \subset \Omega$ berechnet sich dann dadurch, daß man die Wahrscheinlichkeiten für das Eintreten aller Elementarereignisse, die in A enthalten sind, aufsummiert, also:

$$P(A) = \sum_{\omega \in A} P(\omega).$$

Bemerkung: Die Forderung $P(\emptyset) = 0$ ist eigentlich überflüssig, denn Ω ist die disjunkte Vereinigung von Ω und $\emptyset$, da $\Omega = \Omega \cup \emptyset$ und $\Omega \cap \emptyset = \emptyset$, woraus sich ergibt:

$$P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset).$$

Wenn wir nun von der rechten und der linken Seite $P(\Omega)$ abziehen, erhalten wir $P(\emptyset) = 0$.

Aus der Definition 1.1.3 können wir leicht folgern, wie man die Wahrscheinlichkeit für das Eintreten des Komplementes eines Ereignisse A ausrechnen kann, wenn $P(A)$ bekannt ist. Ω ist nämlich die disjunkte Vereinigung von A und $\overline{A}$. Damit haben wir:

$$1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A}),$$

woraus wir leicht die Formel

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

folgern. Es ist intuitiv auch klar, daß sie die Wahrscheinlichkeit dafür kennen, daß sie beim Lotto keine fünf oder sechs Richtigen angekreuzt haben, wenn sie die Wahrscheinlichkeit kennen, sechs oder fünf richtige angekreuzt zu haben.

Die Definition 1.1.3 legt eine einfache Relation zwischen der Wahrscheinlichkeit des Eintretens einer disjunkten Vereinigung von zwei Ereignissen A und B und der Wahrscheinlichkeit für das Eintreten von A und der für das Eintreten von B fest. Wir können aber auch eine komplizierter Beziehung folgern, die nützlich ist, wenn die Vereinigung von zwei Ereignissen nicht disjunkt ist. Dafür betrachten wir zunächst ein Beispiel, welches auch als Übung zur wahrscheinlichkeitstheoretischen Modellbildung geeignet ist. Angenommen, sie interessieren sich für die Wahrscheinlichkeit, daß beim einmaligen Würfeln die Augenzahl gerade oder durch drei teilbar ist. Als Ereignismenge wählen wir uns $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, als erstes Ereignis $A = \{\omega_2, \omega_4, \omega_6\}$ und als zweite Ereignis $B = \{\omega_3, \omega_6\}$. Wie man leicht sie, sind A und B nicht disjunkt, denn $A \cap B = \{\omega_6\}$. Wir können also nicht ohne weiteres die Additivität ausnützen. Überlegen wir uns aber trotzdem, was wir erhalten, wenn wir $P(A) + P(B)$ ausrechnen. Elementares rechnen ergibt: $1/2 + 1/3 = 5/6$. Andererseits bekommen wir $P(A \cup B) = 4/6 = 2/3$. Die Differenz erklärt sich dadurch, daß wir bei der ersten Rechnung die Wahrscheinlichkeit für das Eintreten des Durchschnittes $A \cap B$ doppelt berechnet haben. Das sieht man auch ein, wenn man die Vereinigung $A \cup B$ als disjunkte Vereinigung darstellen, nämlich z. B. $A \cup (B \setminus (A \cap B))$. Ähnlich wie oben die Formel $P(\overline{A}) = 1 - P(A)$ hergeleitet wurde, überlegen sie sich leicht, daß $P(B \setminus (A \cap B)) = P(B) - P(A \cap B)$. Damit erhalten sie:

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

Es gilt daher der Satz:

Satz 1.1.1 *Sei Ω ein endlicher Wahrscheinlichkeitsraum und $A, B \subset \Omega$ zwei Ereignisse, dann gilt:*

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

Diese Aussage stellt eine Verallgemeinerung der Additivität dar, da bei einer disjunkten Vereinigung $P(A \cap B) = 0$.

Dieser Satz soll noch durch ein weiteres Beispiel unter einer anderen Perspektive beleuchtet werden.

Beispiel 1.1.1 Nehmen wir an, eine Krankheit äußert sich durch mindestens eines von höchstens zwei Symptomen S_A und S_B. Das Ereignis, daß ein Patient das Symptom S_A hat, bezeichnen wir mit A, und daß er das Symptom S_B hat mit B. Weiterhin sei $P(A)$ und $P(B)$ bekannt. Uns interessiert, mit welcher Wahrscheinlichkeit beide Symptome gleichzeitig auftreten, also $P(A \cap B)$. Da es nur die beiden Symptome S_A und S_B gibt, haben wir $P(A \cup B) = 1$, und damit nach dem obigen Satz

$$1 = P(A) + P(B) - P(A \cap B),$$

woraus sich $P(A \cap B)$ berechnen läßt. Im ersten Teil von Beispiel 1.3.2 wird ein Zahlenbeispiel durchgerechnet. [5]

Wichtige und sehr instruktive Spezialfälle der endlichen Wahrscheinlichkeitsräume sind diese, bei denen jedes Elementarereignis die gleiche Wahrscheinlichkeit besitzt. Solche Räume heißen Laplace'sche Wahrscheinlichkeitsräume und Zufallsexperimente mit gleich wahrscheinlichen Elementarereignissen Laplaceexperimente. Sei (Ω, P) ein Laplace'scher Wahrscheinlichkeitsraum, dann berechnet sich die Wahrscheinlichkeit für das Eintreten eines Ereignisses $A \subset \Omega$ nach der Formel von Laplace:

$$P(A) = \frac{|A|}{|\Omega|}, \tag{1.1}$$

wobei wir für eine endliche Menge M die Anzahl der Elemente mit $|M|$ bezeichnen. Auch wenn sie vielleicht intuitiv einleuchtet, läßt sie sich

[5] Wir konnten unsere Überlegungen durchführen, ohne explizit eine Ereignismenge anzugeben. Da es nur die beiden Symptome S_A und S_B gibt, haben wir $\Omega = A \cup B$, und somit $P(A \cup B) = 1$, was für die weiteren Berechnungen ausreichend war. Es ist aber auch nicht allzu schwer, eine Ereignismenge zu konstruieren. Z. B.: $\Omega = \{\{S_A\}, \{S_B\}\{S_A, S_B\}\}$. Wir haben dann: $A = \{\{S_A\}, \{S_A, S_B\}\}$ und $B = \{\{S_B\}\{S_A, S_B\}\}$. Aus der Kenntnis von $P(A)$ sowie $P(B)$ läßt sich dann nicht nur $P(A \cap B)$ berechnen, sondern die Wahrscheinlichkeit für das Eintreten eines jeden Elementarereignisses. Wie? Aus den gegebenen Informationen können wir dann also den gesamten endlichen Wahrscheinlichkeitsraum bestimmen.

ohne Schwierigkeiten aus der Additivität herleiten:

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} \frac{1}{|\Omega|} = |A| \frac{1}{|\Omega|}.$$

Hier haben wir ausgenutzt, daß wenn jedes Elementarereignis die gleiche Wahrscheinlichkeit besitzt, diese dann $\frac{1}{|\Omega|}$ sein muß; und daß wir in der zweiten Summe $|A|$ gleiche Summanden haben.

Laplace'sche Wahrscheinlichkeitsräume können als Urnenmodelle vorgestellt werden. Wir beschränken uns zunächst der Einfachheit wegen auf Laplaceexperimente mit zwei mögliche Ausgängen. Dazu stellen wir uns eine Urne vor, die N_1 schwarze Kugeln und N_2 weiße Kugeln enthält. $N = N_1 + N_2$ ist dann die Gesamtzahl der Kugeln. Wird nun zufällig eine Kugel gezogen, nachdem sie gut durchmischt wurden sind, so beträgt die Wahrscheinlichkeit dafür, eine schwarze Kugel zu ziehen N/N_1, und dafür, eine weiße zu ziehen N/N_2. Es sei bemerkt, daß es sich um eine Modell handelt, und in Wirklichkeit es sich nicht um weiße oder schwarze Kugeln zu handeln braucht, sondern z. B. um Erfolg oder Mißerfolg, Mädchen– oder Knabengeburt oder irgendetwas anderes.

Beispiel 1.1.2 Wenn wir annehmen, daß Mädchen– und Knabengeburt nicht gleichwahrscheinlich sind, können wir die Geburt u. U. auch als Laplaceexperiment beschreiben, indem wir ein geeignetes Urnenmodell betrachten. Dazu stellen wir uns eine Urne mit weißen und schwarzen Kugeln vor, wobei das Verhältnis der Anzahl der weißen Kugeln zu der Anzahl der schwarzen Kugeln dem Verhältnis der Wahrscheinlichkeit für eine Mädchengeburt zu der einer Knabengeburt entspricht.

Interessant bei den Urnenmodellen ist nicht das einmalige Ziehen einer Kugel, sondern das mehrmalige Ziehen, und z. B. die Frage, mit welcher Wahrscheinlichkeit eine gewisse Anzahl von schwarzen oder weißen Kugeln (Erfolg oder Mißerfolg) gezogen wird. Dabei ist es wichtig, ob die Reihenfolge der gezogenen Kugeln berücksichtigt wird, und ob die Kugel nach jeder Ziehung wieder zurückgelegt wird oder nicht. Zunächst untersuchen wir das Urnenmodell mit Zurücklegen und Berücksichtigung der Reihenfolge. Die Berücksichtigung der Reihenfolge drücken wir dadurch aus, indem wir ein Elementarereignis als eine Sequenz von gezogenen Kugeln darstellen, nämlich $(k_1, \ldots, k_n)$ beim n–maligen Ziehen. Wenn wir also n mal ziehen und die Kugel jedesmal wieder zurücklegen, erhalten wir als Ereignismenge:

$$\Omega = \{(k_1, \ldots, k_n) | k_i \in K \text{ für } i = 1, \ldots n\},$$

wobei K die Menge der Kugeln bezeichne.

Da wir ein Laplace'schen Wahrscheinlichkeitsraum untersuchen, ist die Kenntnis der Anzahl der Elemente von Ω wichtig. Beschränken wir uns zunächst auf den Fall $n = 2$. Dann haben wir uns zu fragen, wieviele mögliche Paare (k_1, k_2) es bei N Kugeln gibt. Wählen wir uns dafür erst ein beliebige aber feste Kugel k_1, dann haben wir an der zweiten Stelle N Möglichkeiten, sie mit einer der N Kugeln zu besetzten. Dieses gilt für jede der zuerst beliebig gewählten Kugel k_1. Da wir für die erste Stelle aber auch N Möglichkeiten haben, sie mit einer beliebig gewählten Kugel zu besetzen, haben wir insgesamt N^2 Möglichkeiten.

Ebenso überlegt man sich die Anzahl der Möglichkeiten von Trippeln (k_1, k_2, k_3). Wenn man zunächst die erste Stelle festhält, dann haben wir nach den vorangegangenen Überlegungen N^2 Möglichkeiten, die letzten beiden Stellen mit jeweils einer der N Kugeln zu besetzten. Da wir für die erste Stelle N Möglichkeiten haben, bekommen wir $NN^2 = N^3$ Möglichkeiten.

Im allgemeinen Fall interessiert uns, wieviele Sequenzen $(k_1, \ldots, k_n)$ es bei N Kugeln gibt. An der ersten Stelle haben wir N Möglichkeiten, an der zweiten, dritten usw. ebenso. Wir erhalten dann N^n Möglichkeiten.

Beim Urnenmodell mit N Kugeln und n–maligen Ziehen beträgt die Wahrscheinlichkeit für ein Elementarereignis $\omega = (k_1, \ldots, k_n)$

$$P(\omega) = \frac{1}{N^n},$$

wenn die Reihenfolge berücksichtigt wird, und jede Kugel nach dem Ziehen wieder zurückgelegt wird.

Beispiel 1.1.3 Wir legen uns eine Urne mit N_1 schwarzen und N_2 weißen Kugeln zugrunde. Die Gesamtzahl der Kugeln ist dann $N = N_1 + N_2$. Wir ziehen zehn mal mit Zurücklegen und Berücksichtigung der Reihenfolge. Wie groß ist die Wahrscheinlichkeit dafür, daß die ersten fünf Kugeln schwarz und die restlichen fünf weiß sind? Die Ereignismenge ist dann

$$\Omega = \{(k_1, \ldots, k_{10}) | k_i \in K \ \text{ für } i = 1, \ldots n\}$$

und das interessierende Ereignis

$$A = \{(k_1, \ldots, k_{10}) \in \Omega | k_1, \ldots, k_5 \in S, k_6, \ldots, k_{10} \in W\},$$

wobei S die Menge der schwarzen Kugeln, W die Menge der weißen Kugeln und $K = S \cup W$ ist. Nach den obigen Überlegungen haben wir: $|\Omega| = N^{10}$. Nun haben wir noch $|A|$ zu bestimmen. Betrachten wir eine Sequenz der Länge zehn, also $(k_1, \ldots, k_{10})$ und überlegen uns, wieviele Möglichkeiten wir haben, die ersten fünf Stellen mit schwarzen und die restlichen fünf Stellen mit weißen Kugeln zu besetzen. Für die erste Stelle haben wir N_1 Möglichkeiten, ebenso für die nächste vier Stellen. Damit erhalten wir N_1^5 Möglichkeiten, die ersten fünf Stellen mit schwarzen Kugeln zu besetzen. Analog bekommen wir N_2^5 Möglichkeiten, die letzten fünf Stellen mit weißen Kugeln zu besetzen. Daraus ergibt sich $|A| = N_1^5 N_2^5$. Nach der Formel von Laplace (siehe (1.1)) haben wir dann:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{N_1^5 N_2^5}{N^{10}}.$$

Man macht sich leicht klar, daß die gleiche Wahrscheinlichkeit herauskommt, wenn nicht die ersten fünf Stellen mit schwarzen Kugeln besetzt werden sollen, sondern beliebige andere, sowie die restlichen mit weißen Kugeln. Man hat dann an jeder der fünf Stellen, die mit schwarzen Kugeln besetzt werden sollen, N_1 Möglichkeiten, und für die restlichen fünf Stellen N_2 Möglichkeiten, und man kann so wie oben verfahren.

Wie wir gesehen haben, läuft die Berechnung von Wahrscheinlichkeiten eines Laplaceexperimentes darauf hinaus, die Anzahl von Elementen gewisser Mengen zu bestimmen. Der Zweig der Mathematik, welcher sich damit beschäftigt, heißt Kombinatorik. Wir werden im folgenden alle Aussagen der Kombinatorik auf das sogenannte Grundabzählprinzip zurückführen, für das wir schon einen ersten Eindruck gewonnen haben.

Satz 1.1.2 *(Grundabzählprinzip) Seien $M_1, M_2, \dots, M_n$ n endliche Mengen und*

$$\Omega = \{(m_1, \dots, m_n) | m_1 \in M_1, \dots, m_n \in M_n\},$$

(Ω besteht also aus allen Sequenzen der Länge n, wobei jeweils die ersten Stelle mit einem Element aus M_1 besetzt wird, die zweite mit einem Element aus M_1 usw.) dann ist

$$|\Omega| = |M_1| \times |M_2| \times \cdots \times |M_n|.$$

(Wenn wir $|M_1|$ Möglichkeiten haben, die erste Stelle zu besetzten, $|M_2|$ Möglichkeiten haben, die zweite Stelle zu besetzten, usw, dann multiplizieren sich die Möglichkeiten.)

Beispiel 1.1.4 Nucleinsäuren RNS und DNS bestehen aus einer Folge von vier Basen, nämlich Adenin, Cytosin, Guanin und Uracil bzw. Thymin. Die DNS kann aus einer Folge von 10 000 bis 100 000 solcher Bausteine bestehen. Nach dem Grundabzählprinzip gibt es dann 4^{10000} bis 4^{100000} mögliche DNS.

Die Bedeutung des Grundabzählprinzips erschöpft sich nicht nur darin, die Anzahl der Elemente einer Menge zu bestimmen, deren Elemente gewisse Sequenzen sind, sondern auch andere Aussagen der Kombinatorik. Z. B. kann man mit dem Grundabzählprinzip die Menge aller Teilmengen mit einer gewissen Anzahl von Elementen einer endlichen Menge bestimmen. Solche Überlegungen spielen bei Urnenmodellen ohne Berücksichtung der Reihenfolge eine wichtige Rolle. Wir werden aber erst einige näherliegende Überlegungen anstellen.

Beispiel 1.1.5 Wir untersuchen ein Urnenmodell mit Berücksichtigung der Reihenfolge, aber ohne Zurücklegen. Es sei also eine Urne mit N Kugeln gegeben und es soll n mal gezogen werden, wobei die Kugel jedesmal nicht zurückgelegt wird. Dabei ist es sinnvoll, $n \leq N$ anzunehmen. Wenn K die Menge der Kugeln ist, dann hat die Ereignismenge folgende Gestalt:

$$\Omega = \{(k_1, k_2, \dots, k_n) | k_1 \in K, k_2 \in K \backslash \{k_1\}, \dots, k_n \in K \backslash \{k_1, \dots, k_{n-1}\}\}.$$

(Durch $k_2 \in K \setminus \{k_1\}$ wird ausgedrückt, daß nach dem ersten Ziehen die erste Kugel nicht mehr zurückgelegt worden ist, usw, und schließlich sind beim n–ten Ziehen die ersten $n-1$ Kugeln nicht mehr zurückgelegt worden, was durch $k_n \in K \setminus \{k_1, \dots, k_{n-1}\}\}$ ausgedrückt wird.) Nach dem Grundabzählprinzip haben wir:

$$|\Omega| = |K||K \setminus \{k_1\}| \cdots |K \setminus \{k_1, \dots, k_{n-1}\}| = N(N-1) \cdots (N-(n-1)).$$

Falls $N = n$ ist, also falls so oft eine Kugel ohne Zurücklegen gezogen wird, wie es Kugeln in der Urne gibt, erhalten wir:

$$n(n-1) \cdots 2\ 1 = n!\,.$$

Diese Größe heißt n–Fakultät, (Schreibweise: $n!$). Für sie gibt es noch eine andere Interpretation. Sie ist nämlich die Anzahl der Möglichkeiten, n Elemente einer Menge in einer Reihenfolge anzuordnen. Somit haben wir eine Folgerung eingesehen.

Folgerung 1.1.3 *Sei Ω eine Menge mit n Elementen, dann gibt es $n!$ Möglichkeiten, sie in einer Reihenfolge anzuordnen.*

Wir wollen uns nun überlegen, wieviele Möglichkeiten es gibt, aus einer n–elementigen Menge eine k–elementige Menge auszuwählen. Dazu ordnen wir zuerst hilfsweise die k Elemente in einer Reihenfolge an. [6] Nach dem Beispiel 1.1.5 gibt es dafür

$$n(n-1) \cdots n-k+1$$

Möglichkeiten. Da bei Mengen, insbesondere bei Teilmengen, die Reihenfolge nicht berücksichtigt wird, haben wir noch durch die Anzahl der möglichen Reihenfolgen zu dividieren. Nach Folgerung1.1.3 beträgt sie $k!$, also bekommen wir für die Anzahl der k–elementigen Mengen einer n–elementigen Menge:

$$\frac{n(n-1) \cdots n-k+1}{k!}.$$

Wenn dieser Bruch mit $(n-k)!$ erweitert wird, erhalten wir die elegantere Form $\frac{n!}{k!(n-k)!}$. Man bezeichnet den Ausdruck auch mit dem Symbol $\binom{n}{k}$, welcher auch Binomialkoeffizient genannt wird. Es gilt also der folgende Satz:

[6] Das können wir uns auch als Urnenmodell vorstellen, indem aus einer Urne mit n Kugeln k mal ohne Zurücklegen mit Berücksichtigung der Reihenfolge gezogen wird.

Satz 1.1.4 *Sei Ω eine Menge mit n Elementen, dann gibt es $\binom{n}{k}$ Möglichkeiten, eine Teilmenge mit k Elementen auszuwählen. Hierbei wird $\binom{n}{0} = 1$ gesetzt.* [7]

Beispiel 1.1.6 Vielleicht haben Sie sich schon einmal gefragt, wieviele Möglichkeiten es gibt, beim Lotto sechs Zahlen aus 49 anzukreuzen. Nach dem obigen Satz sind es $\binom{49}{6} = 13983816$ Möglichkeiten. Da das Lottospiel ein Laplaceexperiment ist, ist die Wahrscheinlichkeit, sechs Richtige anzukreuzen $1/13983816$.

Bemerkung: Es soll nicht verschwiegen werden, woher der Begriff Binomialkoeffizient stammt. Er ist nämlich kein Wort der Kombinatorik, sondern der Algebra, und bezeichnet die Koeffizienten von $(a+b)^n$, wenn man letzteren Term ausmultipliziert. Falls $n = 2$, dann haben wir nach einer bekannten binomischen Formel die Koeffizienten $1 = \binom{2}{2}$, $2 = \binom{2}{1}$, $1 = \binom{2}{0}$. Wenn wir $(a + b)^n = (a + b) \cdots (a + b)$ ausmultiplizieren, dann erhalten wir eine Summe mit allen möglichen Summanden der Form $a^k b^{n-k}$, wobei k alle natürlichen Zahlen zwischen null und n durchläuft. Fixieren wir ein beliebiges k und überlegen uns, wieviele Summanden $a^k b^{n-k}$ es gibt! Da $a^k b^{n-k}$ aus n einzelnen Faktoren, die entweder a oder b sind, besteht, und alle Kombinationen aus k a's vorkommt, haben wir uns zu überlegen, wieviele solche Faktoren es gibt, die aus genau k a's besteht, wenn wir n a's zur Auswahl haben. Das ist aber die Anzahl der k–elementigen Teilmengen aus einer n–elementigen Menge, also $\binom{n}{k}$. Somit erhalten wir die (verallgemeinerte) binomische Formel:

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{1.2}$$

Beispiel 1.1.7 Wir wollen die Anzahl aller Teilmengen einer endlichen Menge mit n Elementen bestimmt werden. Da die Anzahl aller Teilmengen mit k Elementen $\binom{n}{k}$ ist, haben wir über aller k zwischen null und n zu summieren. Wenn wir $\binom{n}{k} = \binom{n}{k} 1^k 1^{n-k}$ berücksichtigen

[7] Es gibt eine genau Möglichkeit, die leere Menge als Teilmenge einer beliebigen Menge auszuwählen.

und die binomische Formel anwenden, erhalten wir:

$$\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} 1^k 1^{n-k} = (1+1)^n = 2^n.$$

Dieses Ergebnis können wir auch direkt aus dem Grundabzählprinzip herleiten. Dazu ordnen wir jedem Element einer Menge mit n Elementen eine genau eine Stelle einer Sequenz der Länge n zu. Die einzelnen Stellen der Sequenz besetzen wir mit 1 oder 0, je nach dem, ob das dieser Stelle zugeordnete Element in einer bestimmten Teilmenge enthalten ist oder nicht. Dadurch erhalten wir eine eindeutige und umkehrbare Zuordnung zwischen den Teilmengen einer n–elementigen Menge und den Sequenzen der Länge n mit den Einträgen 1 und 0. Nach dem Grundabzählprinzip gibt es 2^n solche Sequenzen.

Wir haben gesehen, daß man jeder k–elementigen Teilmenge einer n–elementigen Menge nach obiger Vorschrift genau eine Sequenz der Länge n zuordnen können, bei der k Stellen mit Einsen und $n-k$ Stellen mit Nullen besetzt sind. Da es genau $\binom{n}{k}$ k–elementige Teilmengen einer n–elementigen Menge gibt, existieren ebensoviele Sequenzen der Länge n mit k Einsen und $n-k$ Nullen. Das Resultat können wir auch folgendermaßen interpretieren: es gibt genau $\binom{n}{k}$ Möglichkeiten, k weiße und $n-k$ schwarze Kugel auf eine Sequenz der Länge n zu verteilen. Dieser Sachverhalt ist für den folgenden Satz wichtig.

Satz 1.1.5 *Es sei eine Urne mit N_1 weißen und N_2 schwarzen Kugeln sowie insgesamt $N = N_1 + N_2$ Kugeln gegeben. Wenn n mal mit Zurücklegen und Berücksichtigung der Reihenfolge eine Kugel gezogen wird, dann ist die Wahrscheinlichkeit, k weiße und $n-k$ schwarze zu ziehen, gleich*

$$\binom{n}{k} \frac{N_1^k N_2^{n-k}}{N^n}.$$

Beweis: Die Ereignismenge ist:

$$\Omega = \{(k_1, \ldots, k_n) | k_1, \ldots, k_n \in K\},$$

wobei K die Menge der Kugeln bezeichnet. Das interessierende Ereignis ist:

$A = \{(k_1, \ldots, k_n) \in \Omega | k$ Stellen sind mit schwarzen Kugeln besetzt,

$$n - k \text{ mit weißen}\}.$$

Nach der Formel von Laplace haben wir $|\Omega|$ und $|A|$ zu bestimmen. Nach dem Grundabzählprinzip ist $|\Omega| = N^n$. Um $|A|$ zu bestimmen, überlegen wir uns zuerst, wieviele Möglichkeiten wir haben, die ersten k Stellen mit weißen Kugeln zu besetzten und die restlichen mit schwarzen Kugeln. Wir erhalten wieder nach dem Grundabzählprinzip: $N_1^k N_2^{n-k}$. Man macht sich leicht klar, daß man nach dem Grundabzählprinzip genausoviele Möglichkeiten erhält, wenn man nicht die ersten k Stellen mit weißen Kugeln besetzt, sondern sich beliebige andere k Stellen auswählt, weil man dann an jeder dieser k Stellen N_1 und an jeder der restlichen $n - k$ Stellen N_2 Möglichkeiten hat, sie zu besetzen. (siehe Beispiel 1.1.5) Da es $\binom{n}{k}$ Möglichkeiten gibt, aus n Stellen sich k Stellen auszuwählen, um sie mit weißen Kugeln zu besetzen und die restlichen Stellen mit schwarzen, bekommen wir:

$$|A| = \binom{n}{k} N_1^k N_2^{n-k}.$$

Nach der Formel von Laplace ist dann:

$$P(A) = \frac{|A|}{|\Omega|} = \binom{n}{k} \frac{N_1^k N_2^{n-k}}{N^n}.$$

Beispiel 1.1.8 In einer Population gibt es für ein gewisses Gen zwei Allele, A und a, wobei A zu 30 % und a zu 70 % vorkommt. Weiterhin sei A dominant. Mit welcher Wahrscheinlichkeit besitzt bei einer zufälligen Paarung der Nachkomme die Merkmalsausprägung, welche durch A vererbt wird? Es sei dabei vorausgesetzt, daß die zwei Allele bei beiden Geschlechtern gleichermaßen zu 30 % bzw 70 % vorkommen. (Präziser: in den Keimzellen beider Geschlechter beträgt das Verhältnis 3/7.)

Dieses läßt sich durch ein Urnenmodell beschreiben. Dazu stellen wir uns eine Urne mit 100 Kugeln vor, wobei 30 mit A und 70 mit a beschriftet sind. [8] Wir ziehen nun zweimal, wobei wir die Reihenfolge berücksichtigen. Das erste Ziehen entspricht dann der Wahl eines Weibchen und der zweite der Wahl eines Männchen. Da die beiden Allele

[8] Es würden auch 10 Kugeln genügen, wenn 3 mit A und 7 mit A beschriftet sind. Es mußen nur die relativen Häufigkeiten bewahrt werden

bei beiden Geschlechtern in der gleichen relativen Häufigkeit auftreten, können wir so tun, als ob wir die Kugel nach dem ersten Ziehen wieder zurücklegen, bevor wir die zweite Kugel ziehen. Nun ist die Frage zu beantworten, mit welcher Wahrscheinlichkeit mindesten eine Kugel mit der Aufschrift A gezogen wird. Dieses Ereignis kann in zwei disjunkter Ereignisse aufgespalten werden: es werden zwei Kugeln mit der Aufschrift A gezogen, und es wird genau eine Kugel mit der Aufschrift A gezogen. Für das erste Ereignis gibt es nach dem Grundabzählprinzip 30^2 Möglichkeiten. Daher ist die Wahrscheinlichkeit für das Eintreten dieses Ereignisses nach der Formel von Laplace $30^2/100^2$. Für das zweite Ereignis erhalten wir nach Satz1.1.5: $\binom{2}{1}30 * 100/100^2$. Da die beiden Ereignisse sich gegenseitig ausschließen, also disjunkt sind, können wir die zwei Wahrscheinlichkeiten addieren.

Nachdem Urnenmodelle mit Zurücklegen und Berücksichtigung der Reihenfolge untersucht worden sind, liegt es nahe, zwischen vier verschiedenen Typen von Urnenmodellen zu unterscheiden, nämlich:

1. Urnenmodelle mit Zurücklegen und Berücksichtigung der Reihenfolge,
2. Urnenmodelle mit Zurücklegen und ohne Berücksichtigung der Reihenfolge,
3. Urnenmodelle ohne Zurücklegen und Berücksichtigung der Reihenfolge,
4. Urnenmodelle ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge.

In der Praxis spielen nur der erste und der vierte Typ eine Rolle. Beim zweiten Typ hat man es noch nicht einmal mit Laplaceexperimenten zu tun, da nicht alle Elementarereignisse gleich wahrscheinlich sind. Wir wenden uns daher gleich dem vierten Typ zu.

Es sei eine Urne mit N Kugeln, wobei davon N_1 weiß und N_2 schwarz sind, gegeben. Ein interessierendes Ereignis ist, daß von n zufällig gezogenen Kugeln n_1 weiß und n_2 schwarz sind. Als Ereignismenge A wählen wir uns die Menge alle n–elementigen Teilmengen aus der Menge aller Kugeln K:

$$\Omega = \{\{k_1, \ldots, k_n\} \subset K\}.$$

Nach Satz1.1.4 haben wir:

$$|\Omega| = \binom{N}{n}.$$

Da das interessierende Ereignis A aus n_1 weißen und n_2 schwarzen Kugeln besteht, und wir

1. $\binom{N_1}{n_1}$ Möglichkeiten haben, aus N_1 weißen Kugeln uns n_1 auszuwählen,

2. $\binom{N_2}{n_2}$ Möglichkeiten haben, aus N_2 schwarzen Kugeln uns n_2 auszuwählen,

bekommen wir:

$$|A| = \binom{N_1}{n_1}\binom{N_2}{n_2}.$$

Damit erhalten wir nach der Formel von Laplace den folgenden Satz:

Satz 1.1.6 *Es sei eine Urne mit N_1 weißen und N_2 schwarzen Kugeln sowie $N = N_1 + N_2$ Kugeln gegeben. Wenn n Kugeln ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge gezogen werden, dann ist die Wahrscheinlichkeit für das Eintreten des Ereignisses A, daß n_1 weiße und n_2 schwarze zu ziehen, gleich:*

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}}{\binom{N}{n}}.$$

Beispiel 1.1.9 Das Lotto 6 aus 49 kann als Urnenmodell ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge beschrieben werde. Dazu stellen wir uns eine Urne mit 49 Kugeln vor, wobei 6 weiß und 43 schwarz sind. Nach Satz 1.1.6 erhalten wir als Wahrscheinlichkeit,

1. 6 Richtige zu haben,

$$\frac{\binom{6}{6}\binom{43}{0}}{\binom{49}{6}} = \frac{68757}{166474},$$

2. 5 Richtige zu haben,
$$\frac{\binom{6}{5}\binom{43}{1}}{\binom{49}{6}} = \frac{44075}{332948},$$

3. 4 Richtige zu haben,
$$\frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} = \frac{8815}{499422},$$

4. 3 Richtige zu haben,
$$\frac{\binom{6}{3}\binom{43}{3}}{\binom{49}{6}} = \frac{645}{665896},$$

5. 2 Richtige zu haben,
$$\frac{\binom{6}{2}\binom{43}{4}}{\binom{49}{6}} = \frac{43}{2330636},$$

6. eine Richtige zu haben,
$$\frac{\binom{6}{1}\binom{43}{5}}{\binom{49}{6}},$$

7. keine Richtige zu haben,
$$\frac{\binom{6}{0}\binom{43}{6}}{\binom{49}{6}} = \frac{1}{13983816}.$$

Beispiel 1.1.10 (Capture–Recapture–Methode:) Durch die hier vorgestellte Methode wird die Größe von wildlebenden oder anderen Populationen geschätzt. Stellen sie sich vor, ein Karpfenteichbesitzer will seinen Fischbestand schätzten, ohne den Teich leerzufischen. Er fängt lediglich 10 Karpfen, markiert sie mit einem roten Punkt, und läßt

=11cm recapt.eps

Abbildung 1.1: Die Wahrscheinlichkeit für das eingetretene Ereignis in Abhängigkeit von der Anzahl der Fische im Teich.

sie dann wieder frei. Nach dem sie sich gut durchmischt haben, fängt er wieder 10 Fische und stellt fest, daß von den gefangenen Fischen genau einer mit einem roten Punkt markiert ist. Intuitiv werden sie sicherlich daraus folgern, daß die markierten Karpfen so in etwa den zehnten Teil aller Karpfen ausmachen. Es gibt nun eine statistische Schätzmethode, die Maximum–Likelyhood–Schätzung, die zu diesem Resultat führt. Dazu stellen wir uns den Fischteich als eine Urne vor, wobei im Gegensatz zu den bisherigen Beispielen die Anzahl N aller Kugeln unbekannt ist. Die Wahrscheinlichkeit, neun unmarkierte und einen markierten Fisch zu fangen, hängt also von dem Parameter N ab. Nach Satz 1.1.6 beträgt sie:

$$\frac{\binom{10}{1}\binom{N-10}{9}}{\binom{N}{10}}.$$

æ Man kann nachrechnen, daß bei $N = 100$ tatsächlich ein Maximum vorliegt. An diesem Beispiel kann man auch einen ersten Eindruck von der Vorgehensweise der schießenden Statistik gewinnen. Es wir nämlich vorausgesetzt, daß gewisse Aussagen der Wahrscheinlichkeitstheorie bekannt sind, um auf tatsächlich vorhandene wahrscheinlichkeitstheoretische Beziehungen schließen zu können.

1.2 Zufallsgrößen und Verteilungsfunktionen

Es erscheint natürlich, beim Würfelexperiment den Elementarereignissen eine der Zahlen 1, 2, usw, 6 zuzuordnen. Es hat sich als sehr nützlich erwiesen, Elementarereignissen gewisse Zahlen zuzuordnen, da man dann z. B. die Arithmetik für wahrscheinlichkeitstheoretische Untersuchungen zur Verfügung hat, z. B. kann man ein arithmetisches Mittel bilden. Viele Aussagen können dann auch übersichtlicher formuliert werden.

Definition 1.2.1 *Sei ein endlicher Wahrscheinlichkeitsraum* (Ω, P) *vorgegeben. Eine Abbildung*

$$X : \Omega \longrightarrow \mathbb{R}$$

heißt eine Zufallsvariable.

Da die Argumente einer Zufallsvariablen zufällig sind, sind es auch ihren Werte. Daher existieren die Wahrscheinlichkeiten

$$P(X = x) \text{ für jedes beliebige aber festgewählte } x \in \mathbb{R}.$$

Dabei bilden wir die Menge

$$A = \{\omega \in \Omega | X(\omega) = x\}$$

und setzen $P(X = x) = P(A)$. D. h. wir bilden für ein $x \in \mathbb{R}$ die Menge aller $\omega \in \Omega$, für die $X(\omega) = x$, bezeichnen dieses Ereignis mit A und setzen $P(X = x)$ gleich $P(A)$. Wenn x von der Zufallsvariablen X nicht angenommen wird, dann ist A die leere Menge und wir haben $P(X = x) = 0$.

Beispiel 1.2.1 Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum und $A \subset \Omega$ ein Ereignis. Wenn wir uns nur dafür interessieren, mit welcher Wahrscheinlichkeit A eintritt, dann können wir die folgende Zufallsvariable bilden:

$$X(\omega) = \begin{cases} 1 & \text{wenn } \omega \in A \\ 0 & \text{wenn } \omega \notin A \end{cases} .$$

Wir haben dann die Wahrscheinlichkeiten:

$$P(X = 1) = P(A) \text{ und } \; P(X = 0) = P(\overline{A}) = 1 - P(A).$$

Einer oder mehreren Zufallsvariablen kann man eine Abbildung eine neue Zufallsvariable zuordnen. Auch ihre Werte sind dann zufällig. Seien $X_1, \ldots, X_n$ n Zufallsvariablen, die z. B. n identische Wiederholungen eines Zufallsexperimentes beschreiben, dann sind auch das arithmetische Mittel

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_1$$

sowie die mittlere quadratische Abweichung

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

Zufallsvariablen.

Wir können uns auch ein Zufallsexperiment mit zwei möglichen Ausgängen, z. B. Erfolg oder Mißerfolg, Kopf oder Zahl beim Münzwurf, vorstellen. Einem solchen Zufallsexperiment ordnen wir null oder eins zu, je nach dem, ob z. B. Erfolg oder Mißerfolg eintritt. Wenn wir dieses Zufallsexperiment n mal wiederholen, erhalten wir dann die Werte von n Zufallsvariablen $X_1, \ldots, X_n$. Die Zufallsvariable

$$Y = \sum_{i=1}^{n} X_i$$

liefert dann eine Information darüber, wie oft z. B. Erfolg oder Mißerfolg eintritt.

In vielen Fällen wird auch die Ordnungsrelation der reellen Zahlen ausgenutzt. So ist die Frage sinnvoll, mit welcher Wahrscheinlichkeit eine Zufallsvariable X einen Wert zwischen a und b einschließlich, also in dem Intervall $[a, b]$ annimmt? Dafür schreiben wir:

$$P(a \leq X \leq b).$$

Hierbei ist auch $a = -\infty$ und $b = \infty$ zugelassen. Durch die Ausnutzung der Ordnungsrelation kann eine Funktion definiert werden, die sämtliche Informationen über einen (endlichen) Wahrscheinlichkeitsraum besitzt.

Definition 1.2.2 *Sei X eine Zufallsvariable, dann heißt die Funktion*

$$F(x) = P(X \leq x)$$

die Verteilungsfunktion von X. (siehe Abbildung 1.2)

=10cm bern.eps

Abbildung 1.2: Graph der Verteilungsfunktion der Bernoulliverteilung $B(1, 0.5)$

æ

Die Bedeutung der Verteilungsfunktion liegt darin, daß die Wahrscheinlichkeiten für das Eintreten eines Ereignisses durch eine einzelne Funktion beschrieben werden kann. Daher gibt es auch Tafeln oder Tabellen mit den Werten der wichtigsten Verteilungsfunktionen. Es soll nun erläutert werden, wie die Wahrscheinlichkeiten $P(X = x_i)$ für $i = 1, \ldots, n$ aus der Verteilungsfunktion berechnet werden kann, wenn die Zufallsvariable nur die endlich vielen Werte $x_1, \ldots, x_n$ annimmt. [9] Wenn die möglichen Werte $x_1, \ldots, x_n$ der Größe nach angeordnet sind, dann gilt:

$$P(X = x_i) = P(X \leq x_i) - P(X \leq x_{i-1}) = F(x_i) - F(x_{i-1}),$$

wobei wir für x_0 irgend eine reelle Zahl, die kleiner als x_1 ist, wählen. Dann ist nämlich $F(x_0) = 0$.

Wir betrachten wieder eine Urne mit N Kugeln, von denen N_1 weiß und N_2 schwarz sind. Es soll n mal mit Zurücklegen und Berücksichtigung der Reihenfolge zufällig eine Kugel gezogen werden. Nach Satz 1.1.5 beträgt die Wahrscheinlichkeit, k weiße und $n-k$ schwarze zu ziehen,

$$\binom{n}{k} \frac{N_1^k N_2^{n-k}}{N^n}.$$

[9] Wenn eine Zufallsvariable unendlich viele Werte, wie z. B. Größe oder Gewicht, annehmen kann, dann laßen sich aus der Verteilungsfunktion auch die Wahrscheinlichkeiten für das Eintreten aller interessierenden Ereignisses berechnen. Die Verhältnisse sind aber dann sehr viel verwickelter, so daß es auch hier motiviert ist, zunächst endliche Wahrscheinlichkeitsräume zu betrachten.

Wenn wir $p = N_1/N$ setzen, dann ist $N_2/N = 1 - p$ (wegen $N_1 + N_2 = N$) und

$$\binom{n}{k} \frac{N_1^k N_2^{n-k}}{N^n} = \binom{n}{k} p^k (1-p)^{n-k}.$$

p ist hier die Wahrscheinlichkeit, eine weiße Kugel aus der Urne zu ziehen, und $1-p$ die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen. In dem Urnenmodell ist p immer eine rationale Zahl, welche die relative Häufigkeit der weißen Kugeln angibt. Davon kann man abstrahieren, und p sich als eine beliebige reelle Zahl aus dem Intervall $[0, 1]$ denken. [10] p beschreibt dann z. B. die Wahrscheinlichkeit für das Eintreten von Erfolg, und $1 - p$ die für das Eintreten von Mißerfolg.

Definition 1.2.3 *Eine Zufallsvariable X möge ausschließlich die Werte $0, 1, \ldots, n$ für ein $n \in \mathbb{N}$ annehmen. Falls die Wahrscheinlichkeit $P(X = k)$ für $k = 0, 1, \ldots, n$ durch die Formel*

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

berechnen läßt, heißt X binomialverteilt, genauer $B(n, p)$–verteilt. Die Verteilungsfunktion heißt Binomialverteilung.

Auch wenn wir aus der Urne ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge ziehen, können wir dieses Zufallsexperiment durch eine Zufallsvariable beschreiben, die die Anzahl der gezogenen weißen Kugeln angibt.

Definition 1.2.4 *Eine Zufallsvariable besitzt eine hypergeometrische Verteilung $H(N, M, m)$, wenn die Wahrscheinlichkeit $P(X = n)$ durch*

$$\frac{\binom{m}{n}\binom{M-m}{N-n}}{\binom{M}{N}}$$

gegeben ist.

[10] Da bei einem Laplaceexperiment p eine relative Häufigkeit ist, ist dann p rational. Wenn nun p irrational ist, handelt es sich also nicht um ein Laplaceexperiment. Auch hier hat man eine Beschränkung der Laplaceschen Definition der Wahrscheinlichkeit vor Augen.

1.3 Unabhängigkeit und bedingte Wahrscheinlichkeit

Einem Geschäftsreisenden, der einen Flug gebucht hatte, überfiel die Angst davor, eine Bombe könnte sich an Bord befinden. Daraufhin erkundigte er sich bei einer Versicherung, wie hoch das Risiko sei. Man erwiderte ihm, daß die Wahrscheinlichkeit dafür, daß sich eine Bombe an Bord befindet, 1/10000 sei. Diese schien dem Geschäftsreisenden zwar sehr niedrig, aber er hatte immer noch kein Vertrauen gefaßt. Er fragte dann, wie groß die Wahrscheinlichkeit dafür sei, daß sich zwei Bomben an Bord befinden. Er wurde darüber aufgeklärt, daß diese das Quadrat der Wahrscheinlichkeit dafür ist, daß sich eine Bombe an Bord befindet, also 1/100000000. Dieses Risiko schien ihm erträglich zu sein. Am nächsten Tag wurde der Geschäftsreisende mit einer Bombe im Gepäck auf dem Flughafen verhaftet. Er beteuerte, daß er nicht vorhatte, die Maschine in die Luft zu sprengen, sondern er wollte lediglich das Risiko minimieren, daß sich noch eine zweite Bombe an Bord befindet, da die Wahrscheinlichkeit für zwei Bomben 1/100000000 betrage, und die für eine Bombe nur 1/10000.

Die Pointe besteht darin, daß jedem klar ist, daß die Wahrscheinlichkeit für eine zusätzliche Bombe an Bord unabhängig davon ist, ob unser Geschäftsreisender eine im Gepäck hat. Es gibt aber auch Fälle, in denen die Wahrscheinlichkeit für das Eintreten eines gewissen Ereignisses sich ändert, wenn bekannt ist das ein bestimmtes anderes Ereignis bereits eingetreten ist. Es gibt z. B. Krankheiten, die vom weiblichen X–Chromosom vererbt wird. Daher ist die Wahrscheinlichkeit, an einer solchen Erbkrankheit zu erkranken, nicht unabhängig vom Geschlecht. Um sich mit der Thematik vertraut zu machen, werden wir noch zwei Beispiele erörtern, bevor wir zur mathematischen Handhabung schreiten werden.

Beispiel 1.3.1 Vor einiger Zeit trat in einer Fernsehshow eine indonesische Tänzerin auf, welche die intelligentste Frau auf der Welt sein soll. Sie wurde vor drei verschlossene Türen geführt. Hinter genau einer befand sich einer Ziege. Sie sollte erraten, hinter welcher sich die Ziege befand ohne diese zu öffnen. Nachdem dies geschehen war, öffnete der Showmaster eine andere Tür, die nicht die Ziege verbarg. Daraufhin änderte sie ihre Wahl und tippte auf die dritte noch verbliebene Tür, da sie meinte, die Wahrscheinlichkeit, daß sich die Ziege nun hinter dieser befinde, am größten sei. Das Vorgehen hat in der Öffentlichkeit zu heftigen Diskussionen geführt. Wir wollen nun erörtern, warum die indonesische Tänzerin recht hatte. Angenommen, sie tippt zuerst auf die mittlere Tür, dann kann man mehrere Fälle unterscheiden. Falls die Ziege sich nun wirklich hinter dieser Tür befindet, dann ist die Wahrscheinlichkeit dafür, daß der Showmaster die rechte oder die linke Tür öffnet jeweils 1/2. Wenn sie sich aber hinter der rechten Tür befindet, wird der Showmaster mit hundertprozentiger Sicherheit die linke öffnen. Nehmen wir nun einmal an, die Tänzerin tippt zuerst auf die mittlere Tür, und der Showmaster öffnet darauf die linke, dann beträgt für die Tänzerin die Wahrscheinlichkeit dafür, daß die Ziege sich hinter der mittleren Tür befindet 1/3, und dafür, daß sich sich hinter der rechten befindet, 2/3, da das Verhältnis der Wahrscheinlichkeiten eins zu zwei ist. Die anderen Fälle werden analog abgehandelt.

Das nächste Beispiel stellt eine Motivation für die Definition der sogenannten "bedingten Wahrscheinlichkeit" dar.

Beispiel 1.3.2 Bei einer Krankheit tritt immer zumindestens eins von zwei Symptomen auf. Das Ereignis des Auftretens des einen Symptoms S_A bezeichnen wir mit A, und das des Auftretens des anderen Symptoms S_B mit B. Bei 50 Patienten, die an dieser Krankheit leiden, haben 35 das Symptom S_A und 25 S_B. Wenn wir nun zufällig einen Patienten aus den 50 auswählen, haben wir eine Laplaceexperiment. Die Wahrscheinlichkeiten errechnen sich dann durch die relative Häufigkeiten. Wir erhalten also:

$$P(A) = \frac{35}{50} = 0,7 \quad \text{und} \quad P(B) = \frac{25}{50} = 0,5.$$

Wir überlegen uns zunächst, wieviele Patienten an beiden Symptomen leiden. Dazu berechnen wir uns $P(A \cap B)$ (siehe auch Beispiel 1.1.1). Da wir einen Laplaceraum haben, ist das dann die relative Häufigkeit der Patienten mit beiden Symptomen. Nach Satz 1.1.1 sowie Beispiel 1.1.1 erhalten wir:

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0,7 + 0,5 - 1 = 0,2.$$

Da $0,2$ die relative Häufigkeit der Patienten mit beiden Symptomen ist, haben 10 beide Symptome.

Wir fragen uns nun wie groß die Wahrscheinlichkeit dafür ist, daß ein Patient S_A hat, wenn wir wissen, daß er S_B hat. Wenn wir unter den 25 mit S_B zufällig einen auswählen, haben wir nach der Laplaceschen Formel:

$$\frac{10}{25} = \frac{10/50}{25/50} = \frac{P(A \cap B)}{P(B)}.$$

Hier haben wir mit der Anzahl 50 der Patienten erweitert, um die relativen Häufigkeiten der Patienten mit S_A und S_B und die der Patienten mit S_B zu erhalten, was nach Laplace gleich $P(A \cap B$ und $P(B)$ ist.

Analog erhalten wir für die Wahrscheinlichkeit, daß ein Patient S_B hat unter der Bedingung, daß er S_A hat:

$$\frac{P(B \cap A)}{P(A)} = \frac{0,2}{0,7} = 10/35.$$

Die bedingten Wahrscheinlichkeiten sind hier kleiner als die unbedingten. Vermutlich behindert das Auftreten eines Symptoms das Auftreten des anderen, aber sie schließen sich nicht aus, sonst wäre $P(A \cap B) = 0$.

Definition 1.3.1 *Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum und $A, B \subset \Omega$. Falls $P(B) \neq 0$, dann heißt $P(A|B)$ die Wahrscheinlichkeit für das Eintreten von A, wenn B eingetreten ist. Dabei setzen wir*

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

Die bedingte Wahrscheinlichkeit können wir auch mit Zufallsvariablen definieren. Seien X und Y zwei Zufallsvariablen, dann setzt man:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Beispiel 1.3.3 Markus, Matthäus und Lukas sind zum Tode verurteilt. Jedoch wird einer ausgelost und begnadigt, aber sein Name geheim gehalten. Markus wendet sich an den Wärter: "Einer von den anderen beiden wird hingerichtet. Du verrätst mir also nichts, wenn du mir sagst, wer das ist.” Der Wärter antwortete:” Matthäus wird hingerichtet.” Markus denkt sich nun, daß die Wahrscheinlichkeit für seine Begnadigung 1/2 beträgt, da nur ihm oder Lukas dieses Glück zuteil wird. Hat er recht?

Wir definieren uns die Ereignismenge folgendermaßen:

$$\Omega = \{\text{Markus, Matthäus, Lukas}\},$$

wobei wir den Eintritt eines Elementarereignisses als die Begnadigung der entsprechenden Person interpretieren. Die Wahrscheinlichkeit dafür sei jeweils 1/3. Nach dem der Wärter verraten hat, daß Matthäus hingerichtet wird, wissen wir, daß das Ereignis

$$A = \{\text{Markus, Lukas}\}$$

bereits eingetreten ist. Wir haben also die bedingte Wahrscheinlichkeit $P(\{\text{Markus}\}|A)$ auszurechnen:

$$P(\{\text{Markus}\}|A) = \frac{P(\{\text{Markus}\} \cap A)}{P(A)} = \frac{P(\{\text{Markus}\})}{P(A)} = \frac{1/3}{2/3} = \frac{1}{2}.$$

Beispiel 1.3.4 (Verallgemeinerung von Beispiel 1.3.3) Sei (Ω, P) ein Laplaceraum und $B \subset \Omega$ ein Ereignis, dann laßt sich für jedes Ereignis $A \subset B$ die bedingte Wahrscheinlichkeit $P(A|B)$ als die Wahrscheinlichkeit für das Eintreten von A eines Laplaceraumes mit der Ereignismenge B auffassen. Es gilt nämlich unter der Berücksichtignug der Laplaceschen Formel:

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

Also ist für $A \subset B$ die bedingte Wahrscheinlichkeit $P(A|B)$ die relative Häufigkeit der Elemente von A in B.

Beispiel 1.3.5 (Sterbewahrscheinlichkeit) Wir wollen uns fragen, mit welcher Wahrscheinlichkeit jemand, der gerade 20 Jahre alt geworden ist, auch 30 Jahre alt wird. Die Tabelle 1.1 (siehe [11]) verzeichnet

Tabelle 1.1: Relative Häufigkeiten der Sterbefälle in Berlin 1992 nach gewissen Altersklassen

Alter	rel. Häufigkeit
$0-9$	0.00659461
$10-19$	0.00273783
$20-29$	0.0128797
$30-39$	0.0217122
$40-49$	0.0394486
$50-59$	0.0921817
$60-69$	0.125464
$70-79$	0.221336
$80-89$	0.373822
$90 \leq$	0.103823

relativen Häufigkeiten der Sterbefälle 1992 in Berlin. Diese fassen wir als Sterbewahrscheinlichkeit auf.

Sei A das Ereignis, 20 Jahre, und B das Ereignis, 30 Jahre alt zu werden, dann haben wir die bedingte Wahrscheinlichkeit

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

zu bestimmen. Die Gleichheit der rechten Seite mit der linken Seite gilt wegen $B \subset A$, da jemand, der 30 Jahre alt wird, bereits 20 Jahre alt geworden sein muß. Nun rechnen wir uns $P(A)$ aus, welche die Wahrscheinlichkeit ist, bis zum 20. Lebensjahr nicht gestorben zu sein, also:

$$P(A) = 1 - 0.00659461 - 0.00273783 = 0,99066756.$$

Analog erhalten wir

$$P(B) = 1 - 0.00659461 - 0.00273783 - 0.0128797 = 0,97778786.$$

Somit bekommen wir für die bedingte Wahrscheinlichkeit, daß ein Zwanzigjähriger 30 Jahre alt wird:

$$\frac{P(B)}{P(A)} = 0,986998968655$$

Man würde vermuten, daß $P(B|B) = 1$. Es gilt in der Tat: $P(B|B) = P(B \cap B)/P(B) = P(B)/P(B) = 1$. Allgemeiner erhalten wir, wenn $B \subset A$:

$$P(A|B) = P(A \cap B)/P(B) = P(B)/P(B) = 1.$$

Dieses können wir so interpretieren: falls $B \subset A$, dann folgt aus dem Eintreten von B das von A.

Wenn A und B sich gegenseitig ausschließen, dann erwartet man $P(A|B) = 0$. Tatsächlich bekommen wir nach der Definition 1.3.1 $P(A|B) = P(A \cap B)/P(B) = 0/P(B) = 0$.

Wenn nun andererseits das Auftreten von B das von A nicht beinflußt, dann gilt: $P(A|B) = P(A)$, also:

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

Daraus erhalten wir:

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

Es ist nun naheliegend, die folgende Festlegung zu treffen:

Definition 1.3.2 *Zwei Ereignisse $A, B \subset \Omega$ eines endlichen Wahrscheinlichkeitsraumes (Ω, P) heißen unabhängig, wenn*

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

Zwei Zufallsvariablen X und Y heißen unabhängig, wenn für alle möglichen Werte x und y, die X und Y annehmen können,

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

gilt.

Beispiel 1.3.6 Wir fragen uns, wie groß die Wahrscheinlichkeit ist, zwei Sechsen zu Würfeln, wenn wir zwei Würfel werfen. Als Ereignismenge Ω definieren wir:

$$\Omega = \{(x_1, x_2)|x_1, x_2 \in \{1, \dots, 6\}\}.$$

Wenn wir die zwei Ereignisse:

$$A = \{(6, x_2) | x_2 \in \{1, \ldots, 6\}\}$$

und

$$B = \{(x_1, 6) | x_1 \in \{1, \ldots, 6\}\}$$

wählen, dann erhalten wir für das interessierende Ereignis: $A \cap B$. Nach Satz 1.3.2 bekommen wir:

$$P(A \cap B) = P(A)P(B) = 1/61/6 = 1/36.$$

(Selbstverständlich hätten wir das Ergebnis auch mit Hilfe des Grundabzählprinzips erhalten.)

Beispiel 1.3.7 Wie groß ist die Wahrscheinlichkeit, daß zwei zufällig und unabhängig ausgewählte Personen am gleichen Tag Geburtstag haben? Dabei nehmen wir an, daß die Geburtstage auf das Jahr gleich verteilt sind, und sehen von Schaltjahren ab. Wenn wir die Tage des Jahres mit den natürlichen Zahlen $1, \ldots, 365$ durchnummerieren, dann können wir den Sachverhalt mit zwei Zufallsvariablen X und Y beschreiben, wobei X den Geburtstag einer der zufällig ausgewählten Personen angibt, und Y den der zufällig gewählten anderen Person. Für ein $n \in \{1, \ldots, 365\}$ haben wir dann $P(X = n) = 1/365$ und $P(Y = n) = 1/365$. Damit ist die Wahrscheinlichkeit, daß die zwei Personen an einem ganz bestimmten Tag mit der Nummer n Geburtstag haben, gleich:

$$P(X = n, Y = n) = P(X = n)P(Y = n) = 1/365 * 1/365.$$

Da wir aber 365 verschiedene Möglichkeiten haben, uns ein bestimmtes $n \in \{1, \ldots, 365\}$ zu wählen, müßen wir das Ergebnis noch mit 365 multiplizieren, und erhalten $1/365$.

Nachdem wir die Unabhängigkeit von zwei Ereignissen bzw. zwei Zufallsvariablen definiert haben, können wir diese Definition in naheliegender Weise verallgemeinern.

Definition 1.3.3 *n Ereignisse $A_1, \dots, A_n \subset \Omega$ eines endlichen Wahrscheinlichkeitsraumes (Ω, P) heißen unabhängig, wenn*

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1) \cdots P(A_n).$$

n Zufallsvariablen $X_1, \dots, X_n$ nennen wir unabhängig, wenn für alle $x_1, \dots, x_n \in \mathbb{R}$,

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

gilt.

Beispiel 1.3.8 (Münzwurf) Einen Münzwuf können wir mit dem Wahrscheinlichkeitsraum (Ω, P) beschreiben, wobei $\Omega = \{\text{Kopf, Zahl}\}$ und $P(\text{Kopf}) = P(\text{Zahl}) = 1/2$. Wenn wir nun n mal hintereinander eine Münze werfen, läßt sich dieses Zufallsexperiment mit n Zufallsvariablen $X_1, \dots, X_n$ beschreiben. Die i–te Zufallsvariable erhält dabei den Wert eins, wenn beim i–ten Wurf Kopf erscheint, und null, wenn Zahl erscheint. Da die Würfe unabhängig voneinander sind, gilt:

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

So beträgt z. B. die Wahrscheinlichkeit n mal nur Kopf zu werfen, also daß alle n Zufallsvariablen den Wert eins erhalten, 2^{-n}.

Kehren wir nun nocheinmal zu den bedingten Wahrscheinlichkeiten zurück und zäumen das Pferd von hinten auf! in vielen Fällen ist es nämlich einfacher, die bedingten Wahrscheinlichkeiten zu bestimmen, und daraus auf die unbedingten zu schließen. So ist es z. B. in der Genetik einfacher, die Wahrscheinlichkeit für das Auftreten eines gewissen Genotypes eines Nachkommen zu bestimmen, wenn die Genotypen der Eltern bekannt sind. Wenn weiterhin bekannt ist, mit welcher Wahrscheinlichkeit ein zufällig ausgewähltes Individuum der Elterngeneration einen gewissen Genotyp aufweist, dann können wir auf die Wahrscheinlichkeit für einen gewissen Genotyp eines Nachkomme bei einer zufälligen Paarung schließen. Dazu ist der folgende Satz sehr hilfreich.

Satz 1.3.1 (Satz über totale Wahrscheinlichkeit) *Sei* (Ω, P) *eine endlicher Wahrscheinlichkeitsraum und* $A_1, \dots, A_n \subset \Omega$ *paarweise disjunkte Ereignisse. d. h.* $A_i \cap A_j = \emptyset$ *für* $i \neq j$, *mit der Eigenschaft* $\Omega = \bigcup_{i=1}^n A_i$, *dann gilt für ein Ereignis* $B \subset \Omega$:

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

Dabei setzen wir $P(B|A_i)P(A_i) = 0$, *falls* $P(A_i) = 0$.

Beweis: Zunächst setzen wir voraus, daß für alle $i = 1, \dots, n$ $P(A_i) \neq 0$, da $P(B|A_i)$ nur unter dieser Voraussetzung definiert ist. Es gilt dann:

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

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

Dabei haben wir ausgenutzt, daß die A_i's disjunkt sind, und ihre Vereinigung Ω ergibt.

Falls nicht für alle i $P(A_i) \neq 0$, dann nehmen wir an, daß bei einer geeigneten Nummerierung die ersten m Wahrscheinlichkeiten $P(A_i) \neq 0$ und die restlichen $n - m$ verschwinden. Wir haben dann:

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

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

Hier ist eingegangen, daß für $i = n - m, \dots, n$ mit $P(A_i) = 0$ auch $P(B \cap A_i) = 0$. Es gilt nämlich:

$$P(B \cap A_i) \leq P(A_i) = 0.$$

Beispiel 1.3.9 (Der Satz von Hardy und Weinberg) Unter den beiden Geschlechtern einer Population seien die drei Genotypen AA, Aa und aa mit den relativen Häufigkeiten u, $2v$ und w auf. Es gilt dann $u + 2v + w = 1$. (Daß wir für die relative Häufigkeit von Aa $2v$ und nicht etwa v wählen, hat rein rechentechnische Gründe.) Wir wollen uns ausrechnen, mit welcher Wahrscheinlichkeit AA bei einer zufälligen Paarung auftritt. Dabei sei vorausgesetzt, daß das Merkmal nur in den Varianten A und a vorkommt, und nach den Mendelschen Gesetzen vererbt wird.

Für die Eltern sind $3 * 3 = 9$ Kombinationen möglich, nämlich:

$\mathrm{K}_1 : AA \times AA$	$\mathrm{K}_4 : AA \times Aa$	$\mathrm{K}_7 : AA \times aa$
$\mathrm{K}_2 : Aa \times AA$	$\mathrm{K}_5 : Aa \times Aa$	$\mathrm{K}_8 : Aa \times aa$
$\mathrm{K}_3 : aa \times AA$	$\mathrm{K}_6 : aa \times Aa$	$\mathrm{K}_9 : aa \times aa$

Diese neun Kombinationen schließen sich gegenseitig aus und es sind alle Möglichkeiten für das Elternpaar berücksichtigt worden. Daher wissen wir, daß wir eine disjunkte Zerlegung der Ereignismenge Ω haben, ohne Ω näher zu bestimmen. Wir können daher den Satz 1.3.1 über die totale Wahrscheinlichkeit anwenden und erhalten:

$$P(AA) = \sum_{i=1}^{9} P(AA|K_i)P(K_i).$$

Bei K_3, K_6, K_7, K_8 und K_9 kann nach den Mendelschen Gesetzen kein Nachkomme vom Genotyp AA entstehen, d. h. $P(AA|K_i) = 0$. Die übrigen Wahrscheinlichkeiten sind nach Mendel:

$$P(AA|K_1) = 1,\ P(AA|K_2) = P(AA|K_4) = \frac{1}{2},\ P(AA|K_5) = \frac{1}{4}.$$

Die Wahrscheinlichkeiten K_i ergeben sich, wenn wir die zufällige Paarung so auffassen, als ob Vater und Mutter zufällig aus einer Urne gezogen werden, in der die drei Genotypen mit den relativen Häufigkeiten u, $2v$ und w vertreten sind. Es gilt dann:

$$P(K_1) = u^2,\ P(K_2) = P(K_4)2vu,\ P(K_5) = 4v^2 = (2v)^2.$$

Damit erhalten wir:

$$P(AA) = 1u^2 + \frac{1}{2}2vu + \frac{1}{2}2vu + \frac{1}{4}4v^2$$

$$= u^2 + 2vu + v^2 = (u+v)^2.$$

Analog bekommen wir durch Vertauschen von u mit w:

$$P(aa) = (w+v)^2.$$

Wegen $P(AA) + P(aa) + P(Aa) = 1$, ist

$$P(Aa) = 1 - (u+v)^2 - (w+v)^2$$

Da

$$1 = u + 2v + w = (u+v) + (v+w) = [(u+v) + (v+w)]^2,$$

ist

$$(u+v)^2 + (w+v)^2 = 1 - 2(u+v)(w+v).$$

Also erhalten wir:

$$P(Aa) = 2(u+v)(w+v).$$

Wenn sehr viele Nachkommen entstehen, können wir annehmen, daß die Wahrscheinlichkeiten die relativen Häufigkeiten wiederspiegeln. Damit ergeben sich als neue relative Häufigkeiten für die nächste Generation für die Genotypen AA, Aa und aa:

$$u_1 = (u+v)^2,\ \ 2v_1 = 2(u+v)(w+v),\ \ w_1 = (w+v)^2.$$

Für die zweite Generation folgen die relativen Häufigkeiten für die Genotypen AA, Aa und aa:

$$u_2 = (u_1+v_1)^2,\ \ 2v_2 = 2(u_1+v_1)(w_1+v_1),\ \ w_2 = (w_1+v_1)^2.$$

Wenn wir für u_1 und v_1 die obigen Ausdrücke einsetzen, dann erhalten wir:

$$u_2 = [(u+v)^2 + (u+v)(w+v)]^2 = (u+v)^2[(u+v) + (w+v)]^2$$

$$= (u+v)^2 = u_1.$$

Analog erhält man $w_2 = w_1$, so daß auch $v_2 = v_1$ sein muß. In der zweiten Generation ist also die Häufigkeitsverteilung der Genotypen

AA, Aa und aa unverändert gleich der Häufigkeitsverteilung in der ersten Generation. Das ist der berühmte Satz von Hardy und Weinberg.

Der Satz von Hardy und Weinberg gilt natürlich nur in dem Idealfall, in dem kein Genotyp gegenüber einem anderen in der Fortpflanzung und in den Überlebenschancen benachteiligt ist. Das dürfte z. B. bei den menschlichen Blutgruppen der Fall sein. Das dafür zuständige Gen besitzt zwar drei statt zwei Allele, nämlich A, B und 0, aber man kann zeigen, daß auch in dieser Situation der Satz von Hardy und Weinberg gilt. Daher dürften die relativen Häufigkeiten der menschlichen Blutgruppen über alle Generationen ziemlich gleich sein.

1.4 Bernoulli-Schema und Binomialverteilung

Wenn man n mal hintereinander würfelt und jedesmal darauf achtet, ob eine Sechs gewürfelt wurde oder nicht, sind wir überzeugt, daß sich die Ergebnisse nicht gegenseitig beeinflussen, und jedesmal die Wahrscheinlichkeit, eine Sechs zu würfeln, die gleiche bleibt. Der hirnlose Würfel besitzt weder ein Gedächnis noch eine Gefühl für ausgleichende Gerechtigkeit. Wenn ein Spieler zwanzig mal vergebens versucht hat, eine Sechs zu würfeln, wird beim 21. Wurf die Wahrscheinlichkeit nicht größer sein, eine Sechs zu würfeln, als bei jemanden, der bereits eine Sechs gewürfelt hat.

In der Statistik spielen Wiederholungen von Zufallsexperimenten eine wichtige Rolle. Man nennt diesen Vorgang "eine Stichprobe erheben". Dabei ist man daran interessiert, daß sich die einzelnen Stichprobenergebnisse sich nicht gegenseitig beeinflussen. Wir werden hier den Fall ausführlich untersuchen, in dem nur zwei Ergebnisse möglich sind. Dieser ist nicht nur der einfachste Fall, sondern das Wesentliche kommt in ihm auch am deutlichsten zu tragen. Insbesondere ist dieser Fall geeignet, einen grundlegenden Einblick in das Verhältnis von Statistik und Wahrscheinlichkeitstheorie zu gewinnen. Es sei noch erwähnt, daß man jedes Zufallsexperiment unter diesem Blickwinkel betrachten kann, wenn man sich dafür interessiert, ob ein gewisses Ereignis eintritt oder nicht (z. B. ob eine Sechs gewürfelt wurde oder nicht).

Definition 1.4.1 *Besitzt ein Zufallsexperiment nur zwei mögliche Ergebnisse, so spricht man von einem Bernoulli–Experiment (Jakob Bernoulli 1654–1705). Diese werden i. A. mit Zufallsvariablen beschrieben, die die Werte 0 oder 1 annehmen können. Wenn p die Wahrscheinlichkeit für das Ergebnis 1 ist, dann ist die Verteilungsfunktion durch den Parameter p schon eindeutig bestimmt. Diese heißt Bernoulliverteilung, Schreibweise* $B(1, p)$. [11]

Wird eine Bernoulliexperiment mehrmals wiederholt, ohne daß sich die Einzelexperimente gegenseitig beeinflussen, so spricht man von einem Bernoullischema.

So ist z. B. das Werfen einer Münze ein Bernoulliexperiment mit $p = 1/2$. Aber auch das Würfeln können wir als Bernoulliexperiment interpretieren, wenn wir uns nur dafür interessieren, ob eine Sechs gewürfelt wird oder nicht. Man kann jedes Zufallsexperiment als Bernoulliexperiment auffassen, wenn wir uns lediglich dafür interessieren, ob ein Ereignis A eines endlichen Wahrscheinlichkeitsraumes (Ω, P) eintritt oder nicht. Der Parameter p ist dann durch $p = P(A)$ gegeben. Dieses Zufallsexperiment wird zweckmäßigerweise mit der in Beispiel 1.2.1 diskutierten Zufallsgröße beschrieben.

Unser Ziel ist es nun, ein Bernoullischema durch einen geeigneten Wahrscheinlichkeitsraum zu beschreiben. Wenn wir ein Bernoulliexperiment n mal wiederholen und mit den unabhängigen und identisch $B(1, p)$ verteilten Zufallsvariablen $X_1, \ldots, X_n$ beschreiben, die die Werte 0 oder 1 annehmen können, dann ist ein Elementarereignis eine Sequenz der Länge n mit den Einträgen 0 oder 1. Also sieht die Ereignismenge folgendermaßen aus:

$$\Omega = \{(x_1, \ldots, x_n) | x_1, \ldots, x_n \in \{0, 1\}\}.$$

Um P anzugeben, genügt es, die Wahrscheinlichkeit für das Eintreten eines jeden Elementarereignisses zu bestimmen. Weil die Zufallsvariablen unabhängig sind, gilt nach Definition 1.3.3:

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n). \tag{1.3}$$

[11] In der Definition 1.2.3 lernten sie die Binomialverteilung $B(n, p)$ kennen. Die Bernoulliverteilung ist eine spezielle Binomialverteilung mit dem Parameter $n = 1$. Daher ist die Schreibweise $B(1, p)$ gerechtfertigt

Da die Zufallsvariablen alle identisch $B(1,p)$ verteilt sind, haben wir $P(X_i = 1) = p$ und $P(X_i = 0) = 1-p$ für $i = 1, \dots, n$. Sind z. B. in 1.3 k der n Zufallsvariablen gleich eins, und $n-k$ gleich null, so erhalten wir:

$$P(X_1 = x_1, \dots, X_n = x_n) = p^k(1-p)^{n-k}.$$

Dieses Ergebnis ist unabhängig von den Stellen, auf denen die Einsen und Nullen verteilt sind, richtig. Wir haben also das Resultat gewonnen, daß eine Sequenz der Länge n bestehend aus k Einsen und $n-k$ Nullen mit der Wahrscheinlichkeit $p^k(1-p)^{n-k}$ eintritt.

In der Statistik interessiert man sich oft dafür, wieviele Einsen oder Nullen man in einer Stichprobe erhalten hat, also z. B, wie oft Erfolg oder Mißerfolg eingetreten ist. Die Zufallsvariable $X = \sum X_i$ zählt die Einsen. Sie kann also die Werte $0, \dots, n$ annehmen. Wir wollen nun $P(X = k)$ für $0 \leq k \leq n$ bestimmen. Dazu überlegen wir uns, wieviele Sequenzen der Länge n mit genau k Einsen es gibt. Da genau $\binom{n}{k}$ Möglichkeiten existieren, sich aus den n Stellen der Sequenz k auszuwählen, um sie mit Einsen zu besetzen, gibt es ebensoviele Sequenzen der Länge n mit k Einsen und $n-k$ Nullen. Wenn wir berücksichtigen, daß die Elementarereignisse disjunkt sind, erhalten wir:

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

(vergleiche das Ergebnis mit der Definition 1.2.3.) Damit haben wir den folgenden Satz gezeigt:

Satz 1.4.1 *Seien $X_1, \dots, X_n$ unabhängige und identisch Verteilte Zufallsvariablen mit der Bernoulliverteilung $B(1,p)$, dann besitzt die Zufallsvariable*

$$X = \sum_{i=1}^{n} X_i$$

die Binomialverteilung $B(n,p)$.

Beispiel 1.4.1 Wir betrachten wieder eine Urne mit insgesamt N Kugeln, wobei N_1 weiß und N_2 schwarz sind. Das Ziehen einer Kugel ist ein Bernoulliexperiment mit $p = N_1/N$. Wenn wir n mal dieses Experiment wiederholen, dann bedeutet dieses, daß wir n mal mit Zurücklegen und unter Berücksichtigung der Reihenfolge eine Kugel aus der Urne ziehen. Die Wahrscheinlichkeit, k weiße und $n-k$ schwarze zu ziehen, ist nach Satz 1.4.1 gleich

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Wenn wir p durch N_1/N ersetzen, erhalten wir den Satz 1.1.5. Satz 1.4.1 ist somit eine Verallgemeinerung von Satz 1.1.5. Umgekehrt kann man ein Bernoullischema als ein Urnenexperiment mit Zurücklegen und Berücksichtigung der Reihenfolge interpretieren, falls p durch eine relative Häufigkeit gegeben ist.

Beispiel 1.4.2 Wir nehmen an, daß ein Genotyp Aa bei beiden Eltern vorhanden ist. Die Wahrscheinlichkeit, daß ein Nachkomme den Genotyp aa hat, ist nach Mendel $1/4$. Da mehrere Nachkommen ihre Genotypen nicht beeinflussen, wenn es sich nicht um Zwillinge oder Mehrlingsgeburten handelt, kann man das Entstehen von n Nachkommen als Bernoullischema mit n Einzelversuchen auffassen. Wenn z. B. $n = 5$ ist, und X die Anzahl der Nachkommen mit dem Genotyp aa angibt, dann haben wir folgende Wahrscheinlichkeiten:

$X = 0:$

$$P(X=0) = \binom{5}{0} \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^5 = 0,2373$$

$X = 1:$

$$P(X=1) = \binom{5}{1} \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^4 = 0,3955$$

$X = 2:$

$$P(X=2) = \binom{5}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3 = 0,2637$$

$X = 3:$

$$P(X=3) = \binom{5}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2 = 0,0879$$

$X = 4:$

$$P(X=4) = \binom{5}{0}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1 = 0,01465$$

$X = 5:$

$$P(X=5) = \binom{5}{5}\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^0 = 0,000976$$

Man sieht, daß die größten Wahrscheinlichkeiten sich in der Nähe von $p * 5 = 1/4 * 5$ befinden. Das legt die Vermutung nahe, daß p auch in diesem Fall irgendwie etwas mit den relativen Häufigkeiten zu tun hat. Man würde z. B. annehmen, daß, wenn wir statt $n = 5$ eine durch vier teilbare n Zahl gewählt hätten, die maximale Wahrscheinlichkeit bei $n/4$ liegt. Dieses gibt auch Anlaß für eine Maximum–Likelyhood–Schätzung. Diese wird im nächsten Beispiel besprochen. (siehe auch Beispiel 1.1.10)

Beispiel 1.4.3 In einem großen Wald werden 10 Jungvögel einer Art beringt. Wir nehmen an, daß der Wald so groß ist, daß die Vögel ihn nicht verlassen. Außerdem bilde diese Vogelart keine Schwärme sondern verteile sich zufällig im Wald. Diese beiden Voraussetzungen sind z. B. bei Greifvögeln annähernd erfüllt. Nachdem die Vögel flügge und ausgewachsen sind, beobachtet der Förster in größeren Abständen hintereinander zufällig 10 Exemplare dieser Art. Von diesen 10 Vögeln ist einer beringt. Der Förster vermutet, daß so in etwa ein zehntel dieser Vogelgeneration beringt ist, daß es also circa 100 Vögel dieser Art und Generation gibt. Hat der Förster recht?

Wenn wir das zufällige Beobachten der 10 Vögel in größeren Abständen hintereinander als Urnenexperiment mit Zurücklegen und Berücksichtigung der Reihenfolge auffassen, dann wird die Wahrscheinlichkeit für das Eintreten dieses Ereignisses durch eine Binomialverteilung $B(10,p)$ beschrieben, wobei der Parameter p die relative Häufigkeit der beringten Vögel ist. Die Maximum–Likelyhood–Schätzung besteht darin, den (oder einen) Wert für den Parameter p zu bestimmen, so daß das eingetreten Ereignis die (oder eine) maximale Wahrscheinlichkeit hat. Für $p = 1/10$ ist das tatsächlich der Fall, so daß der Förster recht behält. Dieser Sachverhalt wird im nächsten Satz bewiesen.

Satz 1.4.2 *Es werde n mal ein Bernoulliexperiment mit den Ergebnissen 0 oder 1 durchgeführt. Der Parameter p sei unbekannt. Wenn k mal eine 1 und $n-k$ mal eine 0 eingetreten ist, dann hat das eingetretene Ereignis für $p = k/n$ die maximale Wahrscheinlichkeit.*

Beweis: Seien die Voraussetzungen des Satzes gegeben, dann ist

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

zu maximieren. Dazu leiten wir $f(p)$ ab, und setzen die Ableitung null, um einen Extremwert zu bestimmen. Wir erhalten:

$$f'(p) = \binom{n}{k} p^k (1-p)^{n-k} \frac{k-np}{p(1-p)} = f(p)\frac{k-np}{p(1-p)},$$

wenn $p \neq 0$ oder $p \neq 1$. In genau diesen beiden Fällen verschwindet $f(p)$ ohnehin. Andernfalls besitzt $f'(p)$ genau eine Nullstelle bei $p = k/n$. Wenn wir diesen Wert in die zweite Ableitung

$$f''(p) = f(p)\frac{k^2 - k(2np - 2p + 1) + np^2(n-1)}{p^2(1-p)^2}$$

einsetzen, bekommen wir:

$$f''(\frac{k}{n}) = -f(\frac{k}{n})\frac{n^3}{n(n-k)}.$$

Da wir $f(\frac{k}{n}) > 0$ vorausgesetzt haben, ist $f''(\frac{k}{n}) < 0$. Somit liegt das Maximum von $f(p)$ nach elementarer Analysis bei $p = k/n$.

Es wurden in Beispiel 1.1.10 und in Beispiel 1.4.3 jeweils eine Maximum–Likelyhood–Schätzung erläutert. Sie unterscheiden sich darin, daß im ersten Fall ein Urnenmodell ohne Zurücklegen und ohne Reihenfolge und damit eine hypergeometrische Verteilung, während im letzten Fall ein Urnenmodell mit Zurücklegen und mit Reihenfolge und damit eine Binomialverteilung zugrundegelgt wird. In beiden Fällen besitzt das beobachtete Ereignis die maximale Wahrscheinlichkeit, wenn die relative Häufigkeit der Beobachtung die der Gesamtheit wiederspiegelt. Da die Binomialverteilung $B(n,p)$ einfacher zu handhaben ist, als die hypergeometrische Verteilung $H(N,M,m)$, letztere besitzt nämlich einen

Parameter mehr, ist es verlockend, sich bei Stichproben möglichst auf Urnenmodelle mit Zurücklegen und mit Reihenfolge zu beschränken. Dieses Vorgehen ist auch noch aus einem anderen Grund gerechtfertigt. Wenn nämlich die Grundgesamtheit sehr groß ist, dann ändern sich die relativen Häufigkeiten nur sehr geringfügig, wenn wir ein Exemplar entnehmen. Es ist dann nicht sehr ausschlaggebend, ob wir dieses Exemplar wieder zurücklegen oder nicht. Dieser Sachverhalt spiegelt sich auch in einem Satz wieder. Die hypergeometrische Verteilung kann man durch die Binomialverteilung approximieren.

Satz 1.4.3 *Die hypergeometrische Verteilung $H(N, M, m)$ konvergiert für $M \to \infty$ gegen die Binomialverteilung $B(N, p)$ mit $p = M/m$, wobei vorausgesetzt wird, daß die relativen Häufigkeit M/m beim Grenzübergang erhalten bleibt.*

Beweis: Wir zeigen, daß man die Hypergeometrische Verteilung in zwei Faktoren zerlegen kann: die Binomialverteilung und einen Faktor, der für $M \to \infty$ gegen eins geht. Dazu stellen wir uns eine Urne mit M Kugeln vor. Von denen sind W weiß und die restlichen S Kugeln schwarz. Wenn wir m mal ohne Berücksichtigung der Reihenfolge und ohne Zurücklegen ziehen, dann ist die Wahrscheinlichkeit, daß sich unter den m Kugeln s schwarze und w weiße befinden, gleich: [12]

$$\frac{\binom{S}{s}\binom{W}{w}}{\binom{M}{m}} = \frac{S!W!m!(M-m)!}{s!(S-s)!w!(W-w)!M!} =$$

$$\frac{M-m)!}{(S-s)!(W-w)!}\frac{S(S-1)\cdots(S-s+1)W(W-1)\cdots(W-w+1)}{M(M-1)\cdots(M-m+1)}.$$

Hier haben wir $S!$ durch $s!$, $W!$ durch $w!$ und $M!$ durch $m!$ dividiert sowie $\frac{M-m)!}{(S-s)!(W-w)!}$ ausgeklammert. Wegen $(S-s)+(W-w) = M-m$

[12] Dieses Vorgehen ist gerechtfertig, da jedes Zufallsexperiment, dessen Wahrscheinlichkeiten durch die hypergeometrische Verteilung $H(N, M, m)$ gegeben ist, durch ein Urnenmodell ohne Reihenfolge und ohne Zurücklegen vorgestellt werden kann. Dabei ist M die Anzahl der Kugeln, m die Anzahl der weißen Kugeln und N die Anzahl der gezogenen Kugeln. Wegen der Übersicht der Beweisführung haben wir hier eine andere Bezeichnungsweise gewählt, so daß wir in dieser eine $H(m, M, w)$ Verteilung haben.

gilt:

$$\frac{M-m)!}{(S-s)!(W-w)!} = \binom{M-m}{S-s} = \binom{M-m}{W-w}.$$

Dieser Ausdruck ist daher die Anzahl der Möglichkeiten, aus den restlichen $M-m$ Kugeln, sich die in der Urne verbleibenden $S-s$ schwarze und $W-w$ weiße Kugeln auszuwählen. Es gibt aber genausoviele Möglichkeiten, daß $S-s$ schwarze und $W-w$ weiße Kugeln nach der Ziehung in der Urne verbleiben, wie es Möglichkeiten gibt, von m Kugeln s schwarze und w weiße auszuwählen. Wenn man sich nämlich aus M Kugeln m auswählt, und von denen s schwarz und w weiß sind, dann bleiben $M-m$ übrig, von denen $S-s$ schwarz und $W-w$ weiß sind. Daher haben wir:

$$\frac{(M-m)!}{(S-s)!(W-w)!} = \binom{m}{s} = \binom{m}{w}.$$

Da der Parameter p und damit auch $1-p$ der Binomialverteilung $B(m,p)$ gleich der relativen Häufigkeiten S/M und W/M, wenn wir ein Urnenmodell mit Zurücklegen und Reihenfolge zugrundelegen, haben wir nur noch zu zeigen, daß

$$\frac{S(S-1)\cdots(S-s+1)W(W-1)\cdots(W-w+1)}{M(M-1)\cdots(M-m+1)} \tag{1.4}$$

gegen $(S/M)^s(W/M)^w$ konvergiert, wenn M gegen unendlich geht. Hier geht wesentlich ein, daß die relativen Häufigkeiten beim Grenzübergang erhalten bleiben. Dazu klammern wir aus dem Zähler von (1.4) S^s und W^w sowie aus dem Nenner M^m aus:

$$\frac{S^s 1(1-\frac{1}{S})\cdots(1-\frac{s+1}{S})W^w 1(1-\frac{1}{W})\cdots(1-\frac{w+1}{W})}{M^m 1(1-\frac{1}{M})\cdots(1-\frac{m+1}{M})}.$$

Weil

$$\frac{S^s W^w}{M^m} = \left(\frac{S}{M}\right)^s \left(\frac{W}{M}\right)^w,$$

ist noch zu begründen, warum

$$\frac{1(1-\frac{1}{S})\cdots(1-\frac{s+1}{S})1(1-\frac{1}{W})\cdots(1-\frac{w+1}{W})}{1(1-\frac{1}{M})\cdots(1-\frac{m+1}{M})}$$

gegen eins geht für $M \to \infty$. Das ist aber der Fall, da mit M auch W und S gegen unendlich geht.

Wir haben also gezeigt, daß

$$\frac{\binom{S}{s}\binom{W}{w}}{\binom{M}{m}} \to$$

$$\binom{m}{s}\left(\frac{S}{M}\right)^s\left(\frac{W}{M}\right)^w = \binom{m}{w}\left(\frac{S}{M}\right)^s\left(\frac{W}{M}\right)^w.$$

Die unterschiedliche Schreibweise des Binomialkoeffizienten rührt daher, daß man in einem Urnenmodell die Rolle der weißen Kugeln mit der der schwarzen Kugeln vertauschen kann. Dieser Sachverhalt hängt eng mit einer gewissen Symmetrie des Binomialkoeffizienten zusammem. Es gilt nämlich: $\binom{n}{k} = \binom{n}{n-k}$, da $\binom{n}{k}$ die Anzahl der k– elementigen Teilmengen einer n–elementigen Menge ist, und diese Anzahl gleich der $(n-k)$–elementigen Komplementärmengen ist.

1.5 Erwartungswert und Streuung

Wenn wir n mal mit einem fairen Würfel würfeln und die Zufallsvariable X_k die Augenzahl beim k–ten Wurf angibt, dann sind wir überzeugt, daß diese Zufallsvariablen unabhängig sind, sowie daß alle gleichverteilt sind. Es gilt also:

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

und

$$P(X_k = x_k) = \frac{1}{6}$$

für alle $k \in \{1, \ldots n\}$ und alle $x_k \in \{1, \ldots, 6\}$. Falls nun n sehr groß ist, erwarten wir, daß jede der Augenzahlen $1, \ldots, 6$ in etwa $1/6$ aller Würfe auftritt. Allgemeiner würden wir folgende Vermutung anstellen: Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen, welche die Werte $x_1, \ldots, x_k$ mit den Wahrscheinlichkeiten $p_1, \ldots, p_k$ annimmt, dann werden bei großem n die relativen Häufigkeiten $\frac{n_i}{n}$ ungefähr den

Wahrscheinlichkeiten p_i entsprechen. Hier ist n_i die Anzahl der Zufallsvariablen, die den Wert x_i bei einem Zufallsexperiment annehmen. Das arithmetische Mittel ist dann:

$$\frac{1}{n}(n_1x_1 + \cdots + n_kx_k) =$$

$$\sum_{i=1}^{k} x_i \frac{n_i}{n} \approx \sum_{i=1}^{k} x_i p_i.$$

Man würde auch erwarten, daß die Übereinstimmung mit der rechten Seite um so besser ist, je größer n ist. Das motiviert die Folgende Definition:

Definition 1.5.1 *Sei X eine Zufallsvariable, welche genau die Werte $x_1, \ldots, x_k$ mit den Wahrscheinlichkeiten $p_1, \ldots, p_k$ annehmen kann, dann ist der Erwartungswert $E(X)$ definiert durch:*

$$E(X) = \sum_{i=1}^{k} x_i p_i.$$

Beispiel 1.5.1 Sei X eine Zufallsvariable, die die Augenzahl beim Würfeln angibt, also die Zahlen $1, \ldots, 6$ mit gleicher Wahrscheinlichkeit annimmt, dann erhalten wir:

$$E(X) = \sum_{i=1}^{6} i\frac{1}{6} = \frac{21}{6} = 3,5.$$

Natürlich erwartet man nicht, daß $3,5$ gewürfelt wird, sondern daß der Durchschnitt von vielen Würfen etwa $3,5$ sein wird.

Beispiel 1.5.2 Stellen sie sich eine Mexikanische Würfelbude vor! Es wird mit zwei Würfeln gespielt. Wenn ein Pasch gewürfelt wird, dann bekommt der Spieler soviele Pesos, wie der Wurf Augen zeigt, also z. B. beim Sechserpasch 12 Pesos. Andernfalls behält der Budenbesitzer den Einsatz. Wie groß muß dieser sein, damit sich das Geschäft mit dem Glückspiel für den Budenbesitzer lohnt? Um den in Wahrscheinlichkeitstheorie unkundigen Budenbesitzer zu helfen, definieren wir uns eine Zufallsvariable X, die im Falle eines Paschs die Augenzahl, also den Gewinn des Spielers angibt, und wenn kein Pasch gewürfelt wird, den Verlust des Spielers, welcher gleich dem negativen Einsatz ist. Den Einsatz bezeichnen wir mit a. Die Zufallsvariable X kann also folgende Werte annehmen:

$$x_1 = 2,\ \ x_2 = 4,\ \ x_3 = 6,\ \ x_4 = 8,\ \ x_5 = 10,\ \ x_6 = 12,\ \ x_7 = -a.$$

Da die Wahrscheinlichkeit, einen Pasch zu Würfeln, 1/36 beträgt, bekommen wir:

$$P(X = x_i) = \frac{1}{36} \text{ für } i = 1, \ldots, 6.$$

Weiterhin gibt es von insgesamt 36 möglichen Würfen, 30 Möglichkeiten, keinen Pasch zu werfen. Nach der Formel von Laplace erhalten wir dann:

$$P(X = x_7) = \frac{30}{36} = \frac{5}{6}.$$

Nun können wir uns den Erwartungswert ausrechnen:

$$E(X) = \sum_{i=1}^{7} x_i P(X = x_i) =$$

$$2\frac{1}{36} + 4\frac{1}{36} + 6\frac{1}{36} + 8\frac{1}{36} + 10\frac{1}{36} + 12\frac{1}{36} - a\frac{5}{6} = \frac{7}{6} - a\frac{5}{6}.$$

Da dieser Erwartungswert der Gewinn des Spielers im langfristigen Mittel ist, muß er negativ sein, damit sich das Geschäft für den Würfelbudenbesitzer lohnt. Also muß gelten:

$$\frac{7}{6} < a\frac{5}{6}.$$

Das ist aber gleichbedeutend, daß der Einsatz a größer als 7/5 Pesos sein muß.

Beispiel 1.5.3 Nehmen wir an, daß die Wahrscheinlichkeit, bei einer U–Bahnfahrt kontrolliert zu werden, $0,05$ beträgt. Ein Fahrschein koste 3,50 DM und das Schwarzfahren 60,–DM. Lohnt sich das Schwarzfahren? Dazu definieren wir uns die Zufallsvariable X, welche die Werte 0 und 60 annehmen kann, wobei $P(X = 0) = 0,95$ und $P(X = 60) = 0,05$. Wir erhalten dann:

$$E(X) = 0 * 0,95 + 60 * 0,05 = 3.$$

Daraus können wir schließen, daß wir beim Schwarzfahren durchschnittlich 50 Pfennige sparen.

Es wurde bereits angedeutet, daß es einen Zusammenhang zwischen dem arithmetischen Mittel und dem Erwartungswert gibt. Dieser deutete sich auch in der Ausdrucksweise "durchschnittlich" an. Dahinter steckt eine tiefliegende mathematische Beziehung, die hier ohne Beweis mitgeteilt wird. Eine Plausibilitätsüberlegung wird auf Seite 57 angestellt.

Satz 1.5.1 (Gesetz der großen Zahlen) *Sei eine Folge* $(X_n)_{n\in\mathbb{N}}$ *von Zufallsvariablen, die identisch verteilt und unabhängig, gegeben. Dabei heißt eine unendliche Menge von Zufallsvariablen unabhängig, wenn jede endliche Teilmenge unabhängig ist (siehe Definition 1.3.3). Es gilt dann:*

$$\frac{1}{n}\sum_{i}^{n} X_i \to E(X_1) \text{ für } n \to \infty.$$

Dabei ist der Grenzwert folgendermaßen zu verstehen: Für jedes $i \in \mathbb{N}$ *ist die Zufallsvariable* X_i *eine Abbildung eines endlichen Wahrscheinlichkeitsraumes* (Ω, P) *in die reellen Zahlen, also:*

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

Es gilt dann für jedes $\omega_i \in \Omega$ *mit* $P(\omega_i) \neq 0$*:*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i}^{n} X_i(\omega_i) = E(X_1).$$

[13] *Man nennt diese Art der Konvergenz von Zufallsvariablen auch auch fast sichere Konvergenz.* [14]

[13]Da die Zufallsvariablen identisch verteilt sind, gilt natürlich: $E(X_1) = E(X_2) = \dots$.

[14]Fast sichere Konvergenz bedeutet in diesem Fall, daß die Folge der arithmetischen Mittel konvergiert, falls die arithmetischen Mittel von Werten gebildet werden, die mit positiver Wahrscheinlichkeit eintreten können. Da andere Werte in einer Erhebung nicht beobachtet werden können, reicht diese Konvergenz für statistische Zwecke völlig aus. In dem Kapitel über unendliche Wahrscheinlichkeitstheorie werden wir den Begriff der fast sicheren Konvergenz in einer größeren Allgemeingültigkeit definieren, nicht zu letzt deswegen, weil eine unendliche Folge von Ereignissen endlicher Wahrscheinlichkeitsräume als Ereignis eines unendlichen Wahrscheinlichkeitsraumes aufgefaßt werden kann.

Viele Menschen stellen sich unter Statistik so etwas wie das Bilden von arithmetischen Mittel vor. Das Gesetz der großen Zahlen drückt nun aus, daß dahinter auch eine ungeahnte Weisheit steckt. Wenn wir eine Stichprobe vom Umfang n erheben und diese mit n Zufallsvariablen beschreiben, dann können wir das arithmetische Mittel bilden. Ihr Wert wird gewissen zufälligen Schwankungen unterworfen sein, diese werden aber um so geringer ausfallen, je größer n ist. Lassen wir im Idealfall n gegen unendlich gehen, dann erhalten wir schließlich den Erwartungswert.

Nachdem wir die Binomialverteilung als eine der wichtigsten Wahrscheinlichkeitsverteilungen kennegelernt haben, wollen wir nun ihren Erwartungswert bestimmen.

Satz 1.5.2 (Erwartungswert der Binomialverteilung) *Die Binomialverteilung $B(n,p)$ besitzt den Erwartungswert np*

Beweis: Nach Definition des Erwartungswertes ist

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}. \tag{1.5}$$

Nach einer elementaren Rechnung können wir umformen:

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

Wenn wir dieses Ergebnis in (1.5) einsetzten und np ausklammern, erhalten wir:

$$np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}.$$

Es bleibt also zu zeigen, daß $\sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$ gleich eins ist. Dazu führen wir eine Indexverschiebung durch, indem wir k durch $k-1$ ersetzen, und erhalten dann:

$$\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k}) = (p + (1-p))^{n-1} = 1.$$

Im letzten Schritt haben wir die binomische Formel (1.2) angewendet.

Beispiel 1.5.4 (Erwartungswert der Bernoulliverteilung) Da die Bernoulliverteilung eine spezielle Binomialverteilung ist, erhalten wir nach Satz 1.5.2 für die Zufallsvariable X mit der $B(1,p)$–Verteilung: $E(X) = p$. Wir können den Erwartungswert aber auch sehr leicht direkt ausrechnen. Sei X $B(1,p)$–verteilt, d.h $P(X = 1) = p$ und $P(X = 0) = 1 - p$, dann haben wir:

$$E(X) = 1 * p + 0 * (1 - p) = p.$$

Der Erwartungswert besitzt zwei nützliche Eigenschaften, die das Rechnen mit ihm erleichtern.

Satz 1.5.3 (Eigenschaften des Erwartungswertes) *Seien* $X, X_1, \ldots, X_n$ *Zufallsvariablen und* $a \in \mathbb{R}$*, dann gilt:*

1.

$$E(aX) = aE(X)$$

2.

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n).$$

Beispiel 1.5.5 Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen, die alle die Bernoulliverteilung $B(1,p)$ besitzen. Nach Beispiel 1.5.4 gilt für alle $i = 1, \ldots, n$ $E(X) = p$. Die Zufallsvariable

$$X = X_1 + \cdots + X_n$$

hat nach Satz 1.4.1 die Binomialverteilung $B(n,p)$, deren Erwartungswert nach Satz 1.5.2 np ist. Mit dem Satz 1.5.3 können wir dieses Ergebnis leicht bestätigen:

$$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = nE(X_1) = np.$$

Wir haben gesehen, daß die Folge der arithmetischen Mittel einer Folge von unabhängigen und identisch verteilten Zufallsvariablen gegen den Erwartungswert konvergiert. Erlauben wir uns nun den Spaß, den Erwartungswert eines arithmetischen Mittels zu bestimmen!

Beispiel 1.5.6 $X_1, \ldots, X_n$ seien unabhängige und identisch verteilte Zufallsvariablen, dann gilt für das arithmetische Mittel:

$$E(\frac{1}{n}\sum_{i=1}^{n} X_i) = E(X_1).$$

Nach Satz 1.5.3 können wir wie folgt rechnen:

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

$$\frac{1}{n}\sum_{i=1}^{n} E(X_1) = \frac{1}{n} nE(X_1) = E(X_1).$$

Diese Eigenschaft des arithmetischen Mittels spielt in der schließende Statistik ein bedeutende Rolle. Wenn nämlich viele verschieden Personen unabhängig voneinander eine Stichprobe vom Umfang n erheben, diese durch Zufallsvariablen beschreiben, und das arithmetische Mittel bilden, dann werden ihre Ergebnisse nicht allzusehr voneinander abweichen.

In der Regel wird der Wert, den eine Zufallsvariable bei einem Zufallsexperiment annimmt, mehr oder weniger vom Erwartungswert abweichen. Ein Maß für die Streuung ist der Erwartungswert der quadratischen Abweichung vom Erwartungswert.

Definition 1.5.2 *Sei X eine Zufallsvariable und $E(X)$ ihr Erwartungswert, dann ist die Varianz $V(X)$ definiert durch:*

$$V(X) = E(X - E(X))^2 = \sum_x (x - E(X))^2 P(X = x).$$

Man bildet deswegen das Quadrat, damit sich Abweichungen, die kleiner als $E(X)$ sind, sich nicht von solchen, die größer sind, zu null addieren, bzw. zu einer sehr kleinen Summe addieren, obwohl die Abweichungen sehr groß sind. Der Leser wird jetzt vielleicht mit einem gewissen recht einwenden, warum man nicht etwa den Betrag statt das Quadrat bildet. Auch dadurch würden sich die unerwünschten Effekte vermeiden lassen. Darauf sei erwidert, daß der Erwartungswert und die Varianz Maßzahlen sind, ungefähr so, wie Länge und Winkel Maße sind, mit denen sich ein Seefahrer orientieren kann. Wir werden sehen, daß wir mit dem Erwartungswert und der Varianz sehr gut durch die Wahrscheinlichkeitstheorie und Statistik navigieren können.

In der Biologie ist die Varianz z. B. ein Maß für die Variabilität eines Merkmales, die genetisch oder umweltbedingt sein kann. Bei Züchtungen ist man an einer kleinen genetischen Variabilität interessiert, so daß die Varianz zunehmend kleiner wird.

Die Varianz besitzt eine Eigenschaft, die ihre Bedeutung plausibel erscheinen läßt. Da sie der Erwartungswert der Zufallsvariablen $(X - E(X))^2$, können wir das Gesetz der großen Zahlen anwenden (siehe Satz 1.5.1). Sei nämlich eine Folge $(X_n)_{n \in \mathbb{N}}$ von identisch verteilten und unabhängigen Zufallsvariablen gegeben, dann sind auch für alle $i \in \mathbb{N}$ die Zufallsvariablen $(X_i - E(X_i))^2$ identisch verteilt und unabhängig, was der Leser als Übung verifizieren möge. Nach dem Gesetz der großen Zahlen gilt dann:

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - E(X_i))^2 \to V(X) \text{ für } n \to \infty.$$

D. h. die mittlere quadratische Abweichung konvergiert gegen die Varianz. Ebenso konvergiert die empirische Varianz $\frac{1}{n-1} \sum_{i=1}^{n} (X_i - E(X_i))^2$ gegen die Varianz, da $\lim_{n \to \infty} \frac{n}{n-1} = 1$. Daß in der Statistik der empirischen Varianz gegenüber der mittleren quadratischen Abweichung der Vorzug gegeben wir, erscheint dem Leser zunächst sicher gekünstelt, wir aber in dem Kapitel über schließende Statistik deutlich werden.

Beispiel 1.5.7 (Varianz der Bernoulliverteilung) Sei X eine Zufallsvariable mit der Bernoulliverteilung $B(1, p)$. Der Erwartungswert ist dann nach Beispiel 1.5.4 gleich $E(X) = p$. Wir bilden nun die Zufallvariable $(X - p)^2$. Da X die Werte 1 und 0 mit Wahrscheinlichkeit p und $1 - p$ annimmt, haben wir:

$$P((X - p)^2 = (1 - p)^2) = p \text{ und } P((X - p)^2 = p^2) = 1 - p.$$

Dann erhalten wir:

$$E((X - p)^2) = (1 - p)^2) * p + p^2 * (1 - p) =$$

$$p(1 - p)[(1 - p) + p] = p(1 - p).$$

Der nun folgende Satz vereinfacht die Berechnung von Varianzen erheblich.

Satz 1.5.4 *Sei X eine Zufallsvariable und $\mu = E(X)$, dann gilt:*

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

Beweis: Unter Berücksichtigung von Satz 1.5.3:

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

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

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

Beispiel 1.5.8 (Varianz der Binomialverteilung) Wir werden den Satz 1.5.4 benutzen, um die Varianz der Binomialverteilung zu berechnen. Nach Satz 1.5.2 ist der Erwartungswert einer Zufallsvariablen X mit der Binomialverteilung $B(n,p)$ gleich $E(X) = np$. Wir haben also im wesentlichen nur noch $E(X^2)$ zu bestimmen. Dazu bedienen wir uns eines Tricks; wir berechnen nämlich $E(X^2 - X) = E(X^2) - E(X)$. Wir haben dann also nur noch np dazuzuaddieren, um $E(X^2)$ zu erhalten. Da die möglichen Werte, die X annehmen kann, $1, \ldots, n$ sind, haben wir:

$$E(X(X-1) = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k}. \tag{1.6}$$

So ähnlich wie im Beweis von Satz 1.5.2 zeigt man in einer kleinen Nebenrechnung, daß

$$k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}.$$

Wenn wir das in der rechten Seite von (1.6) substituieren und $n(n-1)p^2$ ausklammern, bekommen wir:

$$n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)}.$$

Wir zeigen nun, daß $\sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{(n-2)-(k-2)} = 1$. Wenn wir den Summationsindex k durch $k+2$ ersetzen, erhalten wir unter Berücksichtigung der binomischen Formel:

$$\sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{(n-2)-(k-2)} =$$

$$\sum_{k=0}^{n-2} \binom{n-2}{k} p^{k}(1-p)^{(n-2)-(k)} = (p+(1-p))^{n-2} = 1.$$

Damit haben wir:$E(X^2 - X) = E(X^2) - E(X) = n(n-1)p^2$. Die Varianz ist dann:

$$V(X) = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

Die Varianz V(X) einer Zufallsvariablen X besitzt noch einen kleinen Nachteil, sie hat nämlich eine andere Dimension als X und auch als $E(X)$. Wenn X z. B. eine Länge oder Temperatur bezeichnet, dann auch der Erwartungswert, aber nicht die Varianz, sie ist dann das entsprechende Quadrat. Aus diesem Grund hat sich die folgende Definition eingebürgert.

Definition 1.5.3 *Sei X eine Zufallsvariable, dann heißt die Quadratwurzel σ_X aus der Varianz $V(X)$ die Standartabweichung der Varianz $V(X)$.*

Beispiel 1.5.9 Ein Huhn einer bestimmten Sorte legt pro Tag mit der Wahrscheinlichkeit
$0,2$ kein Ei
$0,5$ ein Ei
$0,3$ zwei Eier .
Diesen Sachverhalt beschreiben wir mit der Zufallsvariable X, wobei $P(X=0) = 0,2$, $P(X=1) = 0,5$ und $P(X=2) = 0,3$ gilt. Wenn der Erlös für ein Ei 10 Pfennige beträgt, wollen wir wissen, wie hoch der erwartete Erlös und die Streuung ist. Zu diesem Zweck multiplizieren wir die Zufallvariable X mit 10. Nach Satz 1.5.3 gilt: $E(10X) = 10E(X)$ und $V(10X) = E(10X - E(10x))^2 = 10^2 E(X - E(X))^2$. Wir berechnen uns nun $E(X)$ und $V(X)$:

1.

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) =$$

$$0 \cdot 0,2 + 1 \cdot 0,5 + 2 \cdot 0,3 = 1,1.$$

2.

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

$$(0 - 1,1)^2 \cdot 0,2 + (1 - 1,1)^2 \cdot 0,5 + (2 - 1,1)^2 \cdot 0,3 = 0,49.$$

Damit ist der erwartete Erlös $E(10X) = 11$ Pfennige und die erwartete quadratische Abweichung $V(10X) = 100 \cdot 0,49 \text{Pfennige}^2 = 49 \text{Pfennige}^2$. Für die Streuung erhalten wir dann: $\sigma_X = \sqrt{49}$ Pfennige $= 7$ Pfennige.

Für den Hühnerhalter ist der erwartete Erlös und die Streuung bei einer größeren Anzahl von Tieren interessanter. Dazu ist der folgende Satz sehr nützlich.

Satz 1.5.5 *Seien X und Y zwei unabhängige Zufallsvariablen, dann gilt:*

$$V(X+Y) = V(X) + V(Y).$$

Beweis: Nach Satz 1.5.4 gilt:

$$V(X+Y) = E((X+Y)^2) - E(X+Y)^2 = E(X^2+2XY+Y^2) - (E(X)+E(Y))^2 =$$

$$E(X^2) + E(2XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y))^2 =$$

$$[E(X^2) - E(X)^2] + [E(Y^2) - E(Y))^2] + 2[E(XY) - E(X)E(Y)] =$$

$$V(X) + V(Y) + 2[E(XY) - E(X)E(Y)].$$

Es bleibt also noch zu zeigen, daß für unabhängige Zufallsvariablen X und Y gilt: $E(XY) = E(X)E(Y)$. Da diese Relation später noch benötigt wir, soll sie in einem Hilfssatz formuliert und bewiesen werden.

Hilfssatz 1.5.6 *Seien X und Y zwei unabhängige Zufallsvariablen, dann gilt:*

$$E(XY) = E(X)E(Y).$$

Beweis: Diese Relation zeigen wir durch eine Rechnung, wobei wir $P(XY = xy) = P(X = x)P(Y = y)$ für unabhängige Zufallsvariablen berücksichtigen:

$$E(XY) = \sum_{x,y} x \cdot yP(XY = xy) = \sum_{x,y} x \cdot yP(X = x)P(Y = y) =$$

$$\sum_x xP(X = x) \sum_y yP(Y = y) = E(X)E(Y).$$

Man kann aus Satz 1.5.5 folgern, daß für n unabhängige Zufallsvariablen $X_1, \ldots, X_n$ gilt:

$$V(X_1 + \cdots + X_n) = V(X_1 + \cdots + V(X_n).$$

Beispiel 1.5.10 (Fortsetzung von Beispiel 1.5.9) Wir nehmen an, daß der Hühnerhalter 100 Tiere besitzt. Diesen ordnen wir 100 Zufallsvariablen $X_1, \ldots, X_{100}$ zu. Entsprechend dem Beispiel 1.5.9 haben sie die Eigenschaft $P(X_i = 0) = 0,2$, $P(X_i = 1) = 0,5$ und $P(X_i = 2) = 0,3$ für $i = 1, \ldots, 100$. Sie sind also identisch verteilt. Wir setzen voraus, daß sie unabhängig sind, da die Annahme vernünftig erscheint, daß sich die Hühner in ihrem Legeverhalten nicht gegenseitig beeinflussen.

Für den erwarteten Erlös bekommen wir:

$$E(\sum_{i=1}^{100} 10X_i) = \sum_{i=1}^{100} E(10X_i) = 100 \cdot 10E(X_1) = 1100 \text{ Pfennige.}$$

Die Varianz berechnen wir uns nach Satz 1.5.5:

$$V(\sum_{i=1}^{100} 10X_i) = \sum_{i=1}^{100} V(10X_i) =$$

$$100V(10X_i) = 100 \cdot 100 \cdot 0,49 = 4900.$$

Als Streuung bekommen wir:$\sigma_{\sum_{i=1}^{100} 10X_i} = 70$ Pfennige.

Mit Hilfe der Standartabweichung können wir uns das Gesetz der großen Zahlen plausibel machen (siehe Satz 1.5.1). Sei $X_1, X_2, \ldots$ eine Folge von identisch verteilten und unabhängigen Zufallvariablen, dann betrachten wir die Folge der arithmetischen Mittel $\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)_{n\in\mathbb{N}}$, die gegen den Erwartungswert $E(X_1)$ konvergiert. Diesen Erwartungswert können wir als eine Zufallsvariable auffassen, die konstant ist. Daher ist ihre Streuung gleich null. Man würde nun erwarten, daß die Folge der Streuungen der arithmetischen Mittel $(\frac{1}{n}\sum_{i=1}^{n} X_i)_{n\in\mathbb{N}}$ gegen null geht. Das ist in der Tat der Fall. Nach Satz 1.5.5 erhalten wir nämlich:

$$V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} nV(X_1) \to 0.$$

Mit der Varianz geht aber auch die Standartabweichung gegen null.

æ

Kapitel 2

Schließende Statistik

2.1 Schätzmethoden

2.1.1 Punktschätzung

In den bisherigen Fällen war die Verteilungsfunktion von Zufallsvariablen bekannt. Sie folgte z. B. aus einem Modell wie etwa einem Urnenmodell oder dem Bernoullischema. Weil wir die Verteilung kannten, konnten wir Voraussagen über die Wahrscheinlichkeiten3n treffen, mit der eine Zufallsvariable Werte annehmen kann.

In der Praxis ist die Zielrichtung meistens umgekehrt. Es werden Werte von Zufallsvariablen beobachtet, um daraus Aussagen über die unbekannte Verteilungsfunktion zu gewinnen. Oft reichen die Beobachtungen bei weitem nicht aus, um die Verteilungsfunktion in ihrem gesamten Verlauf schätzen zu können, sondern nur einige Parameter wie z. B. Erwartungswert oder Varianz. Für viele Zwecke genügt das auch. Grundlegend hierfür ist eine Stichprobe. Sie besteht aus den Werten von n identisch verteilten und unabhängigen Zufallsvariablen $X_1, \ldots, X_n$, die auch Stichprobenvariablen genannt werden. n heißt der Umfang der Stichprobe.

Seien $X_1, \ldots, X_n$ unabhängig und nach einer unbekannten Wahrscheinlichkeitsverteilung identisch verteilt, dann kann z. B. der Erwartungswert $\mu = E(X_1)$ ein interessierender Parameter sein. Diesen kann man schätzen, ohne die Verteilung zu kennen. Es gilt nämlich (siehe Satz 1.5.1 und Beispiel 1.5.6):

1.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu,$$

2.

$$E(\overline{X}) = \mu.$$

Die beiden Eigenschaften rechtfertigen es, $\overline{X} = \overline{X}(X_1, \ldots, X_n)$ als Schätzfunktion des Erwartungswertes zu verwenden. Die zweite Eigenschaft bedeutet, daß die zufälligen Ergebnisse von verschiedenen Personen, die das arithmetische Mittel bilden, nicht allzu sehr voneinander abweichen. Diese Eigenschaft nennt man Erwartungstreue.

Die erste Eigenschaft besagt, daß die Schätzung um so genauer ist, je größer der Stichprobenumfang ist. Diese Eigenschaft nennt man Konsistenz.

Da diese Eigenschaften wünschenswert sind, ist die nun folgende Definition motiviert.

Definition 2.1.1 *Eine Funktion*

$$T(X_1, \ldots, X_n)$$

der Stichprobenvariablen $X_1, \ldots, X_n$ heißt Schätzfunktion für den Parameter Θ, wenn ihr Wert als Schätzung für Θ verwendet wird. Sie heißt erwartungstreue Schätzfunktion, wenn

$$E(T(X_1, \ldots, X_n) = \Theta.$$

Falls

$$\lim_{n\to\infty} T(X_1, \ldots, X_n) = \Theta,$$

dann nennt man T eine konsistente Schätzfunktion.

Da eine Schätzfunktion ebenfalls eine Zufallsvariable ist, ist es möglich, den Erwartungswert $E(T)$ zu bilden. Ebenso ist die Varianz $V(T)$ definiert. Man kann zeigen, daß

$$T(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

diejenige erwartungstreue Schätzfunktion für den Erwartungswert ist, welche die kleinste Varianz besitzt. Kleine Varianzen von Schätzfunktionen sind wünschenswert, da dann die Ergebnisse von Stichproben nicht so stark um den zu schätzenden Parameter streuen.

Wenn die Stichprobenvariablen konkrete Werte annehmen, kann der Wert der Schätzfunktion berechnet werden. Mit Hilfe der beobachteten Werte kann aber auch die Verteilungsfunktion näherungsweise bestimmt werden. In diesem Zusammenhang spielt die empirische Verteilungsfunktion eine wichtige Rolle. Sie erhält man, indem die beobachteten Werte $x_1, \ldots, x_n$ zu der Ereignismenge eines endlichen Wahrscheinlichkeitsraumes zusammenfaßt. Die Wahrscheinlichkeiten sind dann durch die relativen Häufigkeiten gegeben.

Definition 2.1.2 *Seien $x_1, \ldots, x_n$ die beobachteten Werte der Stichprobenvariablen $X_1, \ldots, X_n$. Wenn ein Wert y k–mal unter den beobachteten Werten vorkommt, dann heißt*

$$P_{emp}(X = y) = \frac{k}{n}$$

die empirische Wahrscheinlichkeit. X ist hier eine Zufallsvariable, die die beobachten Werte mit der Wahrscheinlichkeit, die gleich der relativen Häufigkeit ist, annimmt. Die zu diesen Wahrscheinlichkeiten konstruierte Verteilungsfunktion heißt die empirische Verteilungsfunktion:

$$F_{emp}(x) = P_{emp}(X \leq x).$$

Wenn wir uns z. B. für die Anzahl von Erfolgen bzw. Mißerfolgen interessieren, dann können wir die Stichprobe als eine Urne mit weißen (Erfolg) und schwarzen (Mißerfolg) Kugeln auffassen. Die empirischen Wahrscheinlichkeiten sind dann durch die relativen Häufigkeiten des Auftretens von Erfolg bzw. Mißerfolg gegeben.

In der schließenden Statistik macht man häufig von einem Schätzprinzip gebrauch. Dieses besagt, daß man einen Aspekt einer unbekannten Verteilungsfunktion durch den entsprechenden Aspekt der empirischen Verteilungsfunktion schätzten kann. Dieses Schätzprinzip liegt darin begründet, daß die empirische Verteilungsfunktion um so genauer der tatsächlichen entspricht, je größer der Stichprobenumfang ist. Es gilt nämlich der folgende Satz:

Satz 2.1.1 (Cantelli, Glivenko) *Seien $X_1, \ldots, X_n$ Stichprobenvariablen eines endlichen Wahrscheinlichkeitsraumes mit der Verteilungsfunktion $F : \mathbb{R} \to [0,1]$. In einer Stichprobe vom Umfang n nehmen sie gewisse Werte an, zu denen wir die empirische Verteilungsfunktion F_n bilden. Es gilt dann für alle $x \in \mathbb{R}$:*

$$\lim_{n\to\infty} F_n(x) = F(x).$$

1

Beweis: Wenn X eine Zufallsvariable mit der Verteilungsfunktion F ist, dann haben wir nach Definition:

$$F(x) = \sum_{y \leq x} P(X = y).$$

Da die empirischen Verteilungsfunktionen F_n mit den relativen Häufigkeiten gebildet werden, genügt es zu zeigen, daß die relative Häufigkeiten $S_n(y)$, mit denen y beobachtet wird, mit wachsendem n gegen $P(X = y)$ konvergieren. Dazu definieren wir uns neue Stichprobenvariablen $Y_1, \ldots, Y_n$ wie folgt:

$$Y_i = \begin{cases} 1 & \text{falls} X_i = y \\ 0 & \text{sonst} \end{cases}$$

für $i = 1, \ldots, n$. Diese Zufallsvariablen sind dann Bernoulliverteilt mit $p = P(X = y)$. Weiterhin haben wir:

$$S_n(y) = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

wobei für $i = 1, \ldots, n$ y_i die Werte von Y_i sind, die bei einer Stichprobe beobachtet wurden. Nach dem Gesetz der großen Zahlen (siehe Satz 1.5.1) konvergiert $\frac{1}{n} \sum_{i=1}^{n} Y_i$ fast sicher gegen $E(Y_1)$. Nun ist aber $E(Y_1) = p = P(X = y)$.

Da der Erwartungswert einer empirischen Verteilungsfunktion durch das arithmetische Mittel geben ist, genügt die Schätzung des Erwartungswertes einer unbekannten Verteilungsfunktion durch das arithmetische Mittel der beobachteten Werte dem obigen Schätzprinzip. Wir

[1]Der Satz von Cantelli und Glivenko kann auch für unendliche Wahrscheinlichkeitsräume bewiesen werden. Der Beweis ist dann aber komplizierter.

wollen uns nun die Varianz vorknüpfen. Seien $x_1, \ldots, x_n$ beobachtete Werte einer Stichprobe und $\overline{x}$ das arithmetische Mittel, dann ist nach dem Schätzprinzip

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

eine Schätzung für die unbekannte Varianz σ^2. Dieser Umstand legt es nahe, die Funktion

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

als Schätzfunktion für die Varianz zu nehmen. Nach dem Gesetz der großen Zahlen ist sie auch konsistent Sie ist jedoch nicht erwartungstreu, d. h. $E(\hat{S}^2) \neq V(X_1)$, was wie nun zeigen wollen.

Da die Stichprobenvariablen $X_1, \ldots, X_n$ identisch verteilt sind, haben wir

$$E(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2) = \frac{1}{n}\sum_{i=1}^{n}E((X_i - \overline{X})^2) = E((X_1 - \overline{X})^2).$$

Um die Rechnung zu vereinfachen, nehmen wir an, daß für $i = 1, \ldots n$ der Erwartungswert $E(X_i) = 0$ ist. Wenn dieses nicht der Fall sein sollte, betrachten wir statt dessen die Zufallsvariablen $X_i - E(X_i)$, die die gleiche Varianz besitzen, und auch identisch und unabhängig verteilt sind. Weiterhin verwenden wir den Hifssatz 1.5.6. Aus ihm folgt $E(x_1\overline{x}) = \frac{1}{n}E(X_1^2)$, denn:

$$E(X_1\frac{1}{n}\sum_{i=1}^{n}X_i) = \frac{1}{n}\sum_{i=1}^{n}E(X_1X_i) =$$

$$E(X_1^2) + \frac{1}{n}\sum_{i=2}^{n}E(X_1)E(X_i) = E(X_1^2).$$

(wegen $E(X_i) = 0$) Durch eine ähnliche Rechnung bekommen wir: $E(\overline{X_1}^2) = \frac{1}{n}E(X_1^2)$:

$$E(\frac{1}{n}\sum_{i=1}^{n}X_i\frac{1}{n}\sum_{j=1}^{n}X_j = \frac{1}{n^2}\sum_{i,j=1}^{n}E(X_iX_j) =$$

$$\frac{1}{n^2}(\sum_{i=1}^{n} E(X_i^2) + \frac{1}{n^2}\sum_{i\neq j} E(X_iX_j) =$$

$$\frac{1}{n^2}(\sum_{i=1}^{n} E(X_i^2) = \frac{1}{n^2} nE(X_1^2).$$

Nach dieser Vorarbeit können wir leicht $E((X_1 - \overline{X})^2)$ ausrechnen:

$$E((X_1 - \overline{X})^2) = E(X_1^2) - 2E(X_1\overline{X}) + E(\overline{X})^2)) =$$

$$E(X_1^2) - \frac{2}{n}E(X_1^2) + \frac{1}{n}E(X_1^2) =$$

$$\frac{n-1}{n}E(X_1^2) = \frac{n-1}{n}V(X_1),$$

da $E(X_1) = 0$. Wem die Transformation $X_1 \to X_1 - E(X_1)$ nicht ganz geheuer ist, möge sie nun wieder rückgängig machen, und in der letzten Zeile der Rechnung X_1 durch $X_1 - E(X_1)$ ersetzen. Man erhält dann:

$$\frac{n-1}{n}E((X_1^2 - E(X_1)) = \frac{n-1}{n}V(X_1).$$

Aus diesem Grund verwendet man die empirische Varianz

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - E(X_i)^2,$$

welche Erwartungstreu ist. Da $\lim \frac{n-1}{n} = 1$ ist, ist sie auch konsistent.

Aus Gründen der Übersicht fassen wir die bisher erhaltenen Ergebnisse in einem Satz zusammen.

Satz 2.1.2 *Seien $X_1, \ldots, X_n$ Stichprobenvariablen, dann gilt:*

1. *Das arithmetische Mittel*

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

ist eine erwartungstreue und konsistente Schätzfunktion für den Erwartungswert $E(X_1)$. Außerdem besitzt sie unter allen erwartungstreuen Schätzfunktionen für den Erwartungswert die kleinste Varianz.

2. Die empirische Varianz

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - E(X_i)^2$$

ist eine erwartungstreue und konsistente Schätzfunktion für die Varianz.

2.1.2 Intervallschätzungen

I. A. schätzt man mit einer Punktschätzung nicht exakt den wahren Wert, sondern er wird mehr oder wenig deneben liegen. Aus diesem Grund ist es sinnvoll, ein Intervall anzugeben, das den wahren Wert mit einer großen Wahrscheinlichkeit enthält. Wenn $x_1, \ldots, x_n$ die beobachteten Werte einer Stichprobe sind, dann sind wir daran interessiert, ein Intervall $I = [\underline{T}(x_1, \ldots, x_n), \overline{T}(x_1, \ldots, x_n)]$ zu bestimmen, welches den interessierenden Parameter Θ mit einer gewissen großen Wahrscheinlichkeit überdeckt. Wenn wir ein solches Intervall I bestimmt haben, dann ist entweder Θ in dem Intervall I enthalten oder nicht, da Θ ein fester Wert ist. Es ist also sinnlos, zu behaupten, daß mit einer gewissen Wahrscheinlichkeit $\Theta \in I$ gilt, sondern man kann das Intervall I so wählen, daß es mit einer vorgegebenen Wahrscheinlichkeit den festen Wert Θ trifft.

Bei dieser Vorgehensweise ist ein Kompromiß zwischen zwei sich ausschließenden, aber jeweils wünschenswerten Forderungen zu schließen. Zum einen soll das Intervall möglichst klein sein, um möglichst präzise Aussagen machen zu können, zum anderen ist es wünschenswert, die vorgegebene Wahrscheinlichkeit, mit der das Intervall den Parameter Θ enthält, sehr groß zu wählen. Wenn wir z. B. den Parameter p der Bernoulliverteilung schätzen wollen, dann ist sicherlich $p \in [0, 1]$ mit Wahrscheinlichkeit eins. Dieses Intervall ist aber so groß, daß es keine Aussagekraft mehr besitzt. Eine extrem kleines Intervall würden wir durch eine Punktschätzung bekommen, nämlich ein Intervall, das nur einen Punkt enthält. Die Wahrscheinlichkeit, daß dieser Punkt mit dem Parameter p übereinstimmt, ist aber sehr klein.

Definition 2.1.3 *Seien $x_1, \ldots, x_n$ die beobachteten Werte einer Stichprobe, dann ist ein Konfidenzintervall zum Vertrauensniveau γ ist ein*

Intervall

$$[\underline{T}(x_1,\ldots,x_n),\overline{T}(x_1,\ldots,x_n)],$$

das den wahren, aber unbekannten Parameter Θ *mit mindestens der vorgegebenen Wahrscheinlichkeit* γ *überdeckt.*

Man ist, wie oben bemerkt, daran interessiert, die Konfidenzintervalle möglichst klein zu wählen. Da die Berechnung der Funktionen $\underline{T}(x_1,\ldots,x_n)$ und $\overline{T}(x_1,\ldots,x_n)$ i. A. sehr aufwendig ist, gibt es umfangreiches Tabellenmaterial, in denen Konfidenzintervalle vertafelt sind.

Als Vertrauensniveau sind $0,95$ oder $0,99$ gängige Größen. In der Medizin oder Pharmazie werden auch größere Vertrauensniveaus verwendet, z. B. $0,999$. Für welche Größe man sich entscheidet, liegt an den praktischen Erfordernissen und ist weniger eine Frage der Mathematik. In den Tabellen wird statt dem Vertrauensniveau γ meistens $\alpha = 1-\gamma$ angegeben, da diese Größe auch für statistische Tests eine große Bedeutung besitzt. Je größer α ist um so kleiner wird das Konfidenzintervall ausfallen. Da Konfidenzintervalle aus den zufälligen Ergebnissen einer Stichprobe bestimmt werden, wird man durch verschiedene Stichproben auch unterschiedliche Konfidenzintervalle erhalten. Die Zahl α besagt nun, daß man im langfristigen Mittel in höchstens $\alpha\cdot 100\%$ der Fälle daneben liegen wird, während man in den restlichen Fällen ein Konfidenzintervall bestimmt hat, welches den wahren Wert enthält. α wird in der Literatur auch Irrtumswahrscheinlichkeit genannt.

Beispiel 2.1.1 (Fortsetzung von Beispiel 1.4.3) Im Beispiel 1.4.3 haben wir unter Berücksichtigung von Satz 1.4.2 durch eine Maximum–Likelyhood–Schätzung die Anzahl der Vögel geschätzt. Es stellte sich heraus, daß das eingetretenen Ereignis für $p = 1/10$ die maximale Wahrscheinlichkeit besitzt. Da p gleich der relativen Häufigkeit der beringten Vögel ist, schlossen wir, daß es in etwa 100 Vögel gibt. Die Aussage "in etwa 100 Vögel" können durch die Angabe eines Konfidenzintervalls präzisieren. Dazu wählen wir uns $\alpha = 0,05$. Aus der Tabelle (siehe Anhang) bestimmen wir nun ein Konfidenzintervall für den Parameter p der Binomialverteilung $B(n,p)$. Hier ist $n = 10$ und $X = 1$ und wir lesen das folgende Konfidenzintervall ab:

$$[0.01, 0.4].$$

Da die relative Häufigkeit $10/N$ mit der Irrtumswahrscheinlichkeit $\alpha = 0,05$ in diesem Intervall liegt, können wir eine Konfidenzintervall für die Anzahl N der Vögel zum Vertrauensniveau $0,95$ berechnen und bekommen für die untere Intervallgrenze $\underline{N} = 10/0,4 = 25$ sowie für die obere Intervallgrenze $\overline{N} = 10/0,01 = 1000$. Dieses Ergebnis ist so zu verstehen, daß im Mittel in 95% aller Fälle, in denen von zehn beobachten Vögel einer beringt ist, die Anzahl der Vögel in dem Intervall $[25, 1000]$ liegt. Das Intervall dürfte den Förster nicht sehr befriedigen, weil es ziemlich groß ist. Er würde dann zu recht auf die Idee kommen, die Anzahl der Beobachtungen zu erhöhen. Es ist auch eine Aufgabe der Statistik, den Stichprobenumfang so zu bestimmen, daß dabei brauchbare Ergebnisse herauskommen. Wir werden weiter unten ein Verfahren kennenlernen, mit dem man den Stichprobenumfang n so bestimmen kann, daß das Konfidenzintervall für den Parameter p der Binomialverteilung $B(n,p)$ eine gewisse Größe nicht überschreitet.

Die Konfidenzintervalle für die Binomialverteilung $B(n,p)$ sind i. A. bis zu $n = 30$ vertafelt. Das liegt nicht nur daran, daß es unmöglich ist, beliebig viele Tafeln zu erstellen, sondern man kann für großere n die Konfidenzintervalle sehr gut durch ein Verfahren näherungsweise berechnen. Dieses Verfahren beruht auf einer Approximation der Binomialverteilung. Es ist eine keinesweges abwegige Vorstellung, daß sich gewisse zufällig gewonnenen Werte sich in etwa so wie die berühmte Gaußsche Glockenkurve um den Mittelwert verteilen. Dieses ist bei der Binomialverteilung der Fall, wir die Graphic 2.1.2 andeutet.

æ

Die Approximation ist um so besser, je größer n ist. Da mit steigendem n der Erwartungswert np und die Varianz $np(1-p)$ der Binomialverteilung gegen unendlich geht, betrachtet man statt der binomialverteilten Zufallvariablen X_n [2] die sogenannte standarisierte Zufallsvariable

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}.$$

Sie besitzt unabhängig von n den Erwartungswert null und die Varianz

[2] Der Index n bedeutet, daß X_n die Binomialverteilung $B(n,p)$ besitzt.

=10cm binap.eps

Abbildung 2.1: Die Treppenfunktion beschreibt die Binomialverteilung $B(30, 0.5)$. Nichtganzzahlige Werte wurden zu der nächsten ganzen Zahl gerundet und dann die Wahrscheinlichkeit bestimmt, mit der eine $B(30, 0.5)$–verteilte Zufallsvariable diese ganze Zahl annimmt. Diese Wahrscheinlichkeit ist dann also gleich der Fläche zwischen der Abszisse und der "Treppenstufe". Diese Fläche ist aber näherungsweise die Fläche zwischen der "Glockenkurve" und der Abszisse in den entsprechenden Grenzen. Nach elementarer Analysis kann man die Fläche mit einem Integral berechnen (siehe Anhang....). Durch die Parametertransformation $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ bekommt man dann ein Integral der Form $\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

eins. Nach Satz 1.5.3 haben wir nämlich:

$$E(Z_n) = E(\frac{X_n - np}{\sqrt{np(1-p)}}) =$$

$$\frac{1}{\sqrt{np(1-p)}}(E(X_n) - np) = 0.$$

Unter Berücksichtigung von $V(aX) = a^2V(X)$ und dem Satz 1.5.5 bekommt man:

$$V(Z_n) = V(\frac{X_n - np}{\sqrt{np(1-p)}}) =$$

$$\frac{1}{np(1-p)}(V(X_n) - V(np)) = \frac{1}{np(1-p)}(np(1-p) - 0) = 1.$$

Es ist nicht nur der folgende Satz, weswegen die "Gaußsche Glockenkurve" eine solche Berühmtheit erlangt hat.

Satz 2.1.3 (Moivre, Laplace) *Sei X_n für jedes n eine $B(n,p)$–verteilte Zufallsvariable, dann gilt für $a \leq b$:*

$$\lim_{n\to\infty} P(a \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Nach dem Hauptsatz (siehe Anhang) können wir das Integral berechnen, wenn wir eine Stammfunktion $\Phi(x)$ mit der Eigenschaft $\frac{d\Phi(x)}{dx} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ kennen. Es gilt dann:

$$\lim P(a \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq b) \approx \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \Phi(a) - \Phi(b).$$

Da Stammfunktionen nicht eindeutig bestimmt sind, können wir noch zusätzliche Bedingungen an Φ stellen. Man kann z. B. zeigen, daß es eine eindeutig bestimmte Stammfunktion Φ gibt, die folgende Eigenschaften besitzt:

$$\lim_{x\to-\infty} \Phi(x) = 0 \text{ und } \lim_{x\to-\infty} \Phi(x) = 1.$$

Diese Funktion Φ ist die Verteilungsfunktion eines gewissen unendlichen Wahrscheinlichkeitsraumes (siehe ...). Sie heißt Standartnormalverteilung. Da ihre Werte nicht elementar berechnet werden können, sind sie mit guter Näherung vertafelt (siehe ...). Um nun Konfidenzintervalle zu bestimmen ist eine Begriffsbildung erforderlich.

Definition 2.1.4 *Sei X eine Zufallsvariable mit einer beliebigen Verteilung sowie $\alpha \in (0,1)$, dann heißt eine Zahl x_α mit*

$$P(X \leq x_\alpha) \leq \alpha \textit{ und } P(X \geq x_\alpha) \geq 1-\alpha$$

α–Quantil. Ein $1/2$–Quantil heißt auch Median.

Falls der Wert α von der Verteilungsfunktion F von X angenommen wird, dann kann man ein α–Quantil bestimmen, indem man einen Wert x_α mit $F(x_\alpha) = \alpha$ sucht. Ein α–Quantil braucht auch nicht eindeutig bestimmt zu sein. Als Beispiel kann man sich die Bernoulliverteilung $B(1,p)$ vor Augen führen. Wenn $\alpha = p$ gewählt ist, dann ist jedes $x_\alpha < 1$ ein α–Quantil, was die Leserin oder Leser zur Übung verifizieren möge.

Kehren wir nun zur approximativen Bestimmung von Konfidenzintervallen der Binomialverteilung $B(n,p)$ zurück! Dazu nehmen wir an, daß die $B(n,p)$–verteilte Stichprobenvariable X durch eine Stichprobe einen konkreten Wert x angenommen hat. Sei ein Vertrauensniveau $\gamma = 1-\alpha$ vorgegeben, dann bestimmen wir ein (in diesem Fall das) $\alpha/2$–Quantil $x_{\alpha/2}$ und $1-\alpha/2$–Quantil $x_{1-\alpha/2}$ der Standartnormalverteilung aus der Tabelle (siehe...). Aus Symmetriegründen ist übrigens $x_{1-\alpha/2} = -x_{\alpha/2}$. Wenn wir den Betrag dieser Zahlen gleich c setzen, haben wir nach Satz 2.1.3:

$$P(-c \leq \frac{x-np}{\sqrt{np(1-p)}} \leq c) \approx 1-\alpha.$$

Aus der Ungleichungskette $-c \leq \frac{x-np}{\sqrt{np(1-p)}} \leq c$ können wir uns das Konfidenzintervall berechnen, indem wir sie nach p auflösen. Man erhält dann das Konfidenzintervall:

$$\left[\frac{1-\frac{n}{c^2+n}+\frac{2\,x}{c^2+n}-\frac{c\sqrt{c^2\,n+4\,n\,x-4\,x^2}}{\sqrt{n}\,(c^2+n)}}{2}, \frac{1-\frac{n}{c^2+n}+\frac{2\,x}{c^2+n}+\frac{c\sqrt{c^2\,n+4\,n\,x-4\,x^2}}{\sqrt{n}\,(c^2+n)}}{2}\right] \tag{2.1}$$

Beispiel 2.1.2 (Mendelsche Kreuzungsexperiment) Mendel zählte 1865 bei seinen berühmten Kreuzungsexperiment 6022 gelbe Erbsen und 2001 grüne Erbsen (siehe [8]). Die Elterngeneration war dabei mischerbig. Wir wollen nun ein Konfidenzintervall mit der Irrtumswahrscheinlichkeit $\alpha = 0.05$ für die Wahrscheinlichkeit bestimmen, daß ein mischerbiges Erbsenpaar eine Pflanze mit grünen Erbsen als Nachkomme bekommt. Dazu bestimmen wir aus der Tabelle das 0.025–Quantil $x_{\alpha/2}$ der Standartnormalverteilung. Es lautet -1.96. Nach Formel (2.1) bekommen wir das Konfidenzintervall:

$$[0.240062, 0.258994].$$

Wir wollen uns nun die Länge des Intervalls ausrechnen. Indem wir die Intervallgrenzen von (2.1) voneinander subtrahieren, bekommen durch eine einfache Rechnung:

$$l = \frac{c\sqrt{c^2\, n + 4\, n\, x - 4\, x^2}}{\sqrt{n}\ (c^2 + n)}. \tag{2.2}$$

Unser Ziel ist es, eine untere Grenze für den Stichprobenumfang anzugeben, um ein Konfidenzintervall zu erhalten, das eine vorgegebene Länge nicht überschreitet. Wenn wir uns in Erinnerung rufen, daß $0 \le x \le n$ und somit $4nx - 4x^2 = 4x(n-x) \ge 0$ gilt, und die Monotonie der Wurzel berücksichtigen, bekommen wir die Abschätzung:

$$l \ge \frac{c\sqrt{c^2\, n}}{\sqrt{n}\ (c^2 + n)} = \frac{c^2}{c^2 + n}.$$

Damit erhalten wir für n die Abschätzung:

$$n \ge \frac{c^2}{l} - c^2.$$

Wenn wir z. B. ein Konfidenzintervall für den Parameter p der Bernoulliverteilung mit dem Vertrauensniveau 0.95 bestimmen wollen, so daß die Intervallänge nicht größer ist als 0.1, dann können wir uns nun den minimalen Stichprobenumfang berechnen. Dazu lesen wir aus der Tabelle das 0,025–Quantil der Standartnormalverteilung ab. Der Parameter c ist dann der Betrag davon, also 1,96. Damit haben wir:

$$n \ge \frac{1.96^2}{0.1} - 1.96^2 = 34.5744.$$

Der Stichprobenumfang muß also mindestens 35 betragen.

Beispiel 2.1.3 In einem Referendum ist die Bevölkerung aufgerufen, über die Annahme oder Ablehnung einer Gesetzesvorlage abzustimmen. Durch eine Befragung will man eine Wahlvorhersage treffen. Üblicherweise wird sie in Prozenten mit einer Stelle hinter dem Komma angegeben. Das bedeutet, daß wir für das Konfidenzintervall die Länge $l = 0.001$ wählen. Damit eine solche Vorhersage einigermaßen zutreffend ist, ist die Irrtumswahrscheinlichkeit entsprechend hoch anzusetzen, z. B. 0.001 . Für den Parameter c, also den $0,495$–Quantil der Standartnormalverteilung erhalten wir aus der Tabelle 2.58. Zur Planung einer Befragung gehört es, den Stichprobenumfang zu bestimmen. Dieser ist

$$n \geq \frac{2.58^2}{0,001} - 2.58^2 = 6649.74.$$

æ

2.2 Statistische Tests

Es wird hin und wieder behauptet, daß Mendel bei der Auswertung seiner Kreuzungsexperimente die Erbsen nicht ganz korrekt gezählt hat (was ihm hier allerdings nicht unterstellt werden soll). Wie ist das zu verstehen? Das kann man sich doch nur so denken, daß Mendel bereits eine Vorstellung darüber besaß, in welchen Zahlenverhältnissen sich gewisse Merkmale vererben. Da ein Kreuzungsexperiment ein Zufallsexperiment ist, werden die Zahlenverhältnisse, in denen sich gewisse Merkmale vererben, zufälligen Schwankungen unterworfen sein. Diese fallen um so geringer aus, wie wir bereits wissen, je größer der Stichprobenumfang ist. Die Versuchung ist nun groß, statt den Stichprobenumfang zu erhöhen (was mitunter mit praktischen Schwierigkeiten verbunden ist), die beobachteten Zahlenverhältnisse so zu verfälschen, wie sie nach der Überzeugung des Naturforschers im Idealfall sein müßten. In diesem Abschnitt werden wir Möglichkeiten kennenlernen, wie man mit einer Stichprobe Hypothesen testen kann (und zwar ohne zu schummeln, was sich ja von selbst versteht). Da die Ergebnisse eines Zufallsexperimentes immer zufällig sind, treten dabei gewisse Fehler auf. Ein besonderes

Augenmerk werden wir darauf richten, diese Fehler in den Griff zu bekommen. Zunächst soll jedoch ein Beispiel zur Einführung gegeben werden.

Beispiel 2.2.1 In Graz in Österreich wurden 1962 unter 3000 Geburten 1578 Jungen geboren (siehe [4]). Wir wollen die Hypothese testen, daß Jungen und Mädchen gleich häufig geboren werden. Wenn die Hypothese richtig ist, dann haben wir eine Binomialverteilung $B(3000, 0, 5)$. Die Hypothese $p = 0, 5$ nehmen wir an, wenn die beobachtete Anzahl von Jungen nicht allzusehr von 1500 abweicht. Dazu bestimmen wir eine Zahl c (wie das geht, werden wir weiter unten sehen), so daß

$$P(X > c)_{p=0,5} = \alpha,$$

wobei wir z. B. $\alpha = 0, 01$ wählen. Die Zufallsvariable X beschreibt die Anzahl der Knabengeburten. Wenn diese größer als c ist, lehnen wir die Hypothese ab. Da aber das Ereignis $X > c$ mit der Wahrscheinlichkeit $0, 01$ eintritt, wenn die Hypothese richtig ist, lehnen wir in in 1 % aller Fälle die Hypothese ab, obwohl sie richtig ist. In unserem Fall ist $c = 1564$, so daß wir die Hypothese verwerfen, und annehme, daß

$$p > 0, 5.$$

Ziel eines statistischen Tests ist es, eine Hypothese, die auch Nullhypothese H_0 genannt wird, abzulehnen oder anzunehmen. Die Alternative heißt Alternativhypothese und wir mit H_1 bezeichnet. Man sagt auch, daß man H_0 gegen H_1 testet. In dem Beispiel 2.2.1 ist

$$H_0 : p = 0, 5 \quad \text{und} \quad H_1 : p \neq 0, 5.$$

Dabei können zwei Fehler eintreten, nämlich daß H_0 abgelehnt wird, obwohl H_0 richtig ist. Dieser Fall tritt in dem Beispiel 2.2.1 mit der Wahrscheinlichkeit $\alpha = 0, 01$ ein. Es kann aber auch vorkommen, daß man H_0 annimmt, aber H_1 richtig ist. Die Wahrscheinlichkeit dafür, daß dieser Fall eintritt, ist ungleich schwieriger zu bestimmen.

Definition 2.2.1 *1. Falls die Nullhypothese H_0 verworfen wird, obwohl sie richtig ist, dann heißt dieser Fehler Fehler erster Art.*

2. *Wenn die Nullhypothese H_0 angenommen wird, obwohl die Alternativhypothese H_1 richtig ist, dann spricht man von einem Fehler zweiter Art.*

Vor einer falschen Interpretation der Fehler erster oder zweiter Art soll hier ausdrücklich gewarnt werden. Es ist ein großer Irrtum, zu behaupten, daß eine Hypothese mit einer gewissen Wahrscheinlichkeit richtig oder falsch ist, denn eine Hypothese trifft entweder zu oder nicht. Die Fehler erster und zweiter Art sind Aussagen über das Testverfahren (siehe die Definition unten) und nicht über den Wahrheitsgehalt einer Hypothese. In Beispiel 2.2.1 ist der Fehler erster Art so zu interpretieren, daß dieses Testverfahren in 99 % aller Fälle, in denen die Nullhypothese angenommen wird, diese auch zutrifft.

Wie wir an Hand von Beispiel 2.2.1 gesehen haben, kann man den Fehler erster Art vorgeben.

Definition 2.2.2 *Ein statistischer Test zum Niveau α für H_0 gegen H_1 ist ein Entscheidungsverfahren, mit dem die Hypothese H_0 angenommen oder zu verworfen wird. Dabei ist der Fehler erster Art kleiner oder gleich α.*

Es sollen nun einige wichtige statistisch Testverfahren vorgestellt werden.

Zweiseitiger Binomialtest

Seien $x_1, \ldots, x_n$ die Ergebnisse einer Stichprobe von Bernoulliverteilten Stichprobenvariablen (siehe Seite 59) und $x = \sum_{i=1}^{n} x_i$. Wir nehmen an, daß wir bereits eine Vorstellung über den Parameter p der Bernoulliverteilung gewonnen haben, und daß dieser p_0 betrage. Es ist also die Nullhypothese

$$H_0 : p = p_0$$

gegen

$$H_1 : p \neq p_0$$

zu testen. Dazu haben wir noch das Niveau α des Tests festzulegen. Nachdem dieses geschehen ist, bestimmen wir mit dem Ergebnis der Stichprobe ein Konfidenzintervall $[\underline{p}(x), \overline{p}(x)]$ zum Konfidenzniveau $\gamma =$

$1-\alpha$. Dieses Konfidenzintervall hat die Eigenschaft, daß es im langfristigen Mittel in mindestens $\gamma 100\%$ aller Fälle den wahren Wert enthält. Falls die Hypothese H_0 richtig ist, also p_0 der wahre Parameter der Bernoulliverteilung ist, dann enthält das Konfidenzintervall $[\underline{p}(x), \overline{p}(x)]$ in $\gamma 100\%$ aller Fälle p_0. In allerhöchsten $\alpha 100\%$ aller Fälle wird p_0 außerhalb liegen. Damit erhalten wir folgende Entscheidungsregel:

1. $H_0 : p = p_0$ wird abgelehnt, wenn $p_0 \notin [\underline{p}(x), \overline{p}(x)]$

2. $H_0 : p = p_0$ wird angenommen, wenn $p_0 \in [\underline{p}(x), \overline{p}(x)]$

An dieser Stelle sei noch einmal daran erinnert, daß entweder $p = p_0$ oder $p \neq p_0$, also entweder H_0 oder H_1 richtig ist. Die Werte α bzw γ sagen nur etwas über die Qualität des Testverfahrens aus, nämlich daß in höchstens $\alpha 100\%$ aller Fälle H_0 abgelehnt wird, obwohl die Nullhypothese richtig ist.

Man könnte nun einwenden, daß man statt p_0 einen anderen Wert p_1 hätte wählen können, der zufälligerweise auch in dem Intervall $[\underline{p}(x), \overline{p}(x)]$ liegt. Dann hätte man p_1 statt p_0 mit dem gleichen Fehler erster Art als wahren Wert akzeptiert. Es kann aber nur einer von beiden der wahre Wert sein. Dieser Sachverhalt ist folgendermaßen zu interpretieren: ist p_0 der richtige Parameter, dann überdeckt das Konfidenzintervall im langfristigen Mittel in $\gamma 100\%$ aller aller Fälle den Wert p_0 und $H_0 : p = p_0$ wird angenommen. Falls jedoch p_1 der richtige Wert ist, wir aber H_0 akzeptiert haben, dann ist uns ein Fehler zweiter Art unterlaufen. Da der Fehler zweiter Art i. A. sehr schwer zu kontrollieren ist, ist es wünschenswert, die Nullhypothese abzulehnen. In diesem Fall ist einem dann höchstens ein Fehler erster Art unterlaufen, den man ja sehr gut unter Kontrolle hat. So haben wir im Beispiel 2.2.1 keinen Fehler zweiter Art begangen, sondern nur einen erster Art mit $\alpha = 0,01$.

Beispiel 2.2.2 Statistische Tests treten häufig in der Pharmazie auf. Wir wollen die Hypothese testen, daß ein Medikament mit der Wahrscheinlichkeit $p = 0.5$ wirkt. Es ist also die Nullhypothese $H_0 : p = 0.5$ gegen die Alternativhypothese $H_1 : p \neq 0.5$. Als Niveau wählen wir $\alpha = 0.05$. Wir nehmen an, daß bei 20 Versuchspersonen 14 mal Erfolg eingetreten ist. Aus der Tabelle (siehe Anhang) entnehmen wir das Konfidenzintervall

$$[0.474, 0.855]$$

=10cm fe2a.eps

Abbildung 2.2: Fehler zweiter Art in Abhängigkeit vom $p \in [0,1]\backslash\{0.5\}$.

Die Nullhypothese wird also angenommen. Wir haben in diesem Fall also keinen Fehler erster Art begangen. Wie steht es aber mit dem Fehler zweiter Art? Dazu bestimmen wir die Anzahl k der Erfolge, für welche die Nullhypothese bestätigt wird. Das sind aber gerade diejenigen k's, dessen Konfidenzintervalle zum Vertrauensniveau $\gamma = 1 - \alpha = 0.95$ den Wert $p = 0.5$ enthalten. Aus der Tabelle (siehe Anhang) entnehmen wir die Zahlen $k = 6, 7, \ldots, 14$. Da wir eine Binomialverteilung $B(20, p)$ haben, ist die Wahrscheinlichkeit, k Erfolge zu bekommen, durch $\binom{20}{k} p^k (1-p)^{20-k}$ gegeben. Weil die Ereignisse, genau $6, 7, \ldots, 15$ Erfolge zu erzielen, disjunkt sind, können wir die Wahrscheinlichkeiten addieren. Wir definieren uns eine Funktion g durch

$$g(p) = \sum_{k=6}^{14} \binom{20}{k} p^k (1-p)^{20-k}.$$

Falls nun p nicht der Nullhypothese entspricht, dann liefert $g(p)$ die Wahrscheinlichkeit, einen Fehler zweiter Art zu begehen. Die Nullhypothese wird nämlich für die Werte $k = 6, \ldots, 14$ angenommen, obwohl nicht notwendigerweise $p = 0.5$. Die Graphik 2.2.2 zeigt den Verlauf des Graphen von $g(p)$. Man sieht sehr deutlich, daß der Fehler zweiter Art um so größer wird, je mehr sich p dem Wert 0.5, welcher der Nullhypothese entspricht, nähert.

æ

Eine Möglichkeit, die Wahrscheinlichkeit für einen Fehler zweiter Art zu minimieren, besteht darin, den Stichprobenumfang zu erhöhen. Dann erhält man nämlich ein kleineres Konfidenzintervall (siehe Seite 71), d. h. für weniger "Erfolge" wird die Nullhypothese bestätigt.

Eine andere Möglichkeit die Wahrscheinlichkeit zu minimieren, einen Fehler zweiter Art zu begehen, liegt darin, ihn zu vermeiden. Wie oben bereits erwähnt, ist das dann der Fall, wenn die Nullhypothese abgelehnt wird. Dann wissen wir, daß mit Wahrscheinlichkeit $\gamma = 1 - \alpha$ die Alternativhypothese richtig ist. Dazu soll nun ein Beispiel beschrieben werden. In ihm wird auch noch erläutert, wie man bei großen Stichprobenumfängen verfahren kann.

Beispiel 2.2.3 In Berlin wurden 1995 insgesamt 28561 Babys geboren. Davon waren 14715 männlich und 13846 weiblich (siehe [12]) Wir wollen die Hypothese testen, daß die Wahrscheinlichkeit für eine Knabengeburt 0.5 beträgt. Als Niveau des Tests wählen wir $\alpha = 0.05$. Da der Stichprobenumfang sehr groß ist, werden wir nach Satz 2.1.3 die Binomialverteilung durch die Normalverteilung approximieren. Nach der Formel 2.1.3 erhalten wir das Konfidenzintervall zum Vertrauensniveau $\gamma = 1 - \alpha = 0.95$:

$$I = [0.509415, 0.521007].$$

Da $0.5 \notin I$, lehnen wir die Hypothese ab. Damit haben wir auch keinen Fehler zweiter Art begangen.

Einseitiger Binomialtest

Nachdem wir mit dem zweiseitigen Binomialtest eine Nullhypothese der Form $H_0 : p = p_0$ gegen $H_1 : p \neq p_0$ getestet haben, wird beim einseitigen Binomialtest eine Nullhypothese der Form $H_0 : p \geq p_0$ gegen $H_1 : p < p_0$ bzw. $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$ getestet. Beim zweiseitigen Binomialtest sind zwei Werte zu bestimmen, nämlich die obere und untere Grenze des Konfidenzintervalls zum vorgegebenen Niveau. Beim einseitigen Binomialtest wird es dann ein Wert sein. Wenn wir $H_0 : p \geq p_0$ gegen $H_1 : p < p_0$ testen wollen, dann suchen wir die größte Konstante k mit der Eigenschaft

$$P_{p_0}(X \leq k) \leq \alpha.$$

P_{p_0} ist die Wahrscheinlichkeit der Binomialverteilung mit dem Parameter p_0. Dieser Parameter entspricht gerade noch der Nullhypothese H_0. Falls der beobachtete Wert x kleiner als k ist, wird H_0 abgelehnt, andernfalls wird H_0 nicht verworfen.

Beispiel 2.2.4 Die Beispiele 2.2.1 und 2.2.3 legen die Vermutung nahe, daß die Wahrscheinlichkeit für die Geburt eines Jungen größer als 0.5 ist. Wir wollen mit dem Datenmaterial aus Beispiel 2.2.3 die Hypothese $H_0 : p \geq 0.5$ gegen $H_1 : p < 0.5$ für die Wahrscheinlichkeit einer Knabengeburt testen. Das Niveau des Tests legen wir mit 0.05 fest Nach Satz 2.1.3 ist

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}} \tag{2.3}$$

mit guter Näherung standartnormalverteilt. Aus der Tabelle entnehmen wir -1.64 für das 0.05–Quantil der Standartnormalverteilung. Wenn wir diesen Wert für Z_n in die Gleichung (2.3) einsetzen, für $p = 0.5$ und für $n = 28561$ sowie nach X_n auflösen, dann bekommen wir:

$$X_{28561} = 14141.9\,.$$

Durch Runden erhalten wir dann:

$$P_{0.5}(X \leq 14142) \leq 0.05\,.$$

Da der beobachtete Wert 14715 größer als 14142 ist, wird die Nullhypothese angenommen.

Nun wollen wir uns überlegen, wir man $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$ zum Niveau α testet. In Analogie zum ersten Fall gelten die folgenden Regeln:

1. Bestimme das kleinste k mit
$$P_{p_0}(X \geq k) \leq \alpha.$$
2. H_0 wird abgelehnt, falls für den beobachteten Wert x gilt $x \geq k$.
3. H_0 wird nicht abgelehnt, falls $x < k$.

Beispiel 2.2.5 Da wir in Beispiel 2.2.4 die Nullhypothese angenommen haben, und somit eventuell einen Fehler zweiter Art begangen haben, wollen wir nun den Test etwas abwandeln, um einen Fehler zweiter Art möglicherweise zu vermeiden. Wenn wir $H_0 : p \leq 0.5$ gegen

$H_1 : p > 0.5$ testen, haben wir eine Chance, die Nullhypothese abzulehnen. Dabei gehen wir fast genauso vor, wir in Beispiel 2.2.4. Nur setzen wir in Gleichung (2.3) für Z_n statt des α–Quantils das $1 - \alpha$–Quantil ein. Es beträgt 1.64. Nach elementarer Rechnung erhalten wir für unsere Konstante k den Wert 14419. Da $14715 \geq 14419$, lehnen wir die Nullhypothese ab.

Beispiel 2.2.6 Ein Holzhändler will einen Wald kaufen, jedoch nur dann, wenn nicht mehr als 30 % der Bäume durch das Waldsterben geschädigt sind. Um den Anteil der kranken Bäume zu bestimmen, werden 30 Bäume gefällt. Von denen sind neun krank und 21 gesund. Soll der Holzhändler den Wald nun kaufen? Um ihm bei der Entscheidung behilflich zu sein, stellen wir die Nullhypothese auf, daß mehr als 30 % der Bäume beschädigt sind, und raten ihn zum Kauf, wenn diese Hypothese verworfen wird. Wir haben uns noch mit dem Holzhändler auf ein Niveau des statistischen Tests zu verständigen. Aus kaufmännischen Erwägungen schlägt der Holzhändler ein Niveau von 0.05 vor, da er es langfristig in 5 % aller Fälle verschmerzen kann, einen Wald zu kaufen, in dem mehr als 30 % aller Bäume krank sind. Einen Fehler zweiter Art ist bei dieser Vorgehensweise nicht zu befürchten. Für den Test ist jetzt im wesentlichen noch daß 0.05–Quantil der Binomialverteilung $B(30, 0.3)$ zu bestimmen. Da dieses gleich vier ist, wird die Hypothese angenommen. Mit anderen Worten: Wenn der Holzhändler den Wald nur dann kauft, wenn von 30 Bäumen nicht mehr als vier geschädigt sind, dann kann er davon ausgehen, daß langfristig in nur 5 % aller Fälle mehr als dreißig % aller Bäume krank sind. Es sei noch bemerkt, daß wir hier angenommen haben, daß die Anzahl der gefällten Bäume von dreißig im Vergleich zu der Anzahl aller Bäume im Wald verschwindend klein ist. Dann können wir nämlich die Hypergeometrische Verteilung, die wir eigentlich zu grunde legen müßten, durch die Binomialverteilung approximieren (siehe Satz 1.4.3). In der Ausdrucksweise der Urnenmodelle bedeutet das, daß wir ein Urnenmodell mit Zurücklegen betrachten, obwohl wir einen gefällten Baum nicht wieder einpflanzen können.

Literaturverzeichnis

[1] E. Batschelet *Einführung in die Mathematik für Biologen* Springer Verlag 1980

[2] J. Hainzl *Mathematik für Naturwissenschaftler* B. G. Teubner, Stuttgart 1985

[3] W.Köhler, G.Schachtel, P.Voleske *Biometrie – Einführung in die Statitstik für Biologen und Agrarwissenschaftler* Springer, Heidelberg, New York, Tokio 1984

[4] E. Kreyszig *Statistische Methoden und ihre Anwendungen* Vandenhoeck Ruprecht, Göttingen 1977

[5] P. S. de Laplace *Philosophischer Versuch über die Wahrscheinlichkeit* (Paris 1814), Frankfurt 1996

[6] J. Lehn, H. Wegmann, S. Rettig *Aufgabensammlung zur Einführung in die Statistik* B. G. Teubner, Stuttgart 1988

[7] R. J. Lorenz *Grungbegriffe der Biometrie* G. Fischer, Stuttgart, New York 1988

[8] G. Mendel *Versuche über Pflanzenhybriden* 1866/1870 zwei Abhandlungen, Hrsg. Erich von Tschernak–Seysenegg Akad. Verl.–Ges., 1901

[9] D. Rasch *Einführung in die Biostatistik* VEB Deutscher Landwirtschaftsverlag, Berlin 1988

[10] A. Riede *Mathenmatik für Biologen* Vieweg, Braunschweig, Wiesbaden 1993

[11] Statistisches Landesamt Berlin *Sterbefälle in Berlin 1992 nach Todesursachen* Sonderheft 459, 1994

[12] Statistisches Landesamt Berlin *Statistisches Jahrbuch 1996* Kulturbuchverlag Berlin 1996

[13] H. Vogt *Grundkurs Mathematik für Biologen* B. G. Teubner, Stuttgart 1994

[14] E. Weber *Grundriß der biologischen Statistik* G. Fischer Stuttgart 1980

æ

Index

æ