Detlef Dürr

Es gibt keinen Zufall
Eine Einführung in die Wahrscheinlichkeitstheorie

Einleitung und Kapitel 1 der Vorlesung „Einführung in die Stochastik“

5. November 2001

1. Einleitung

1.1 Namen

Wahrscheinlichkeitstheorie und *Stochastik* bedeuten das Gleiche, Letzteres ist griechischen Ursprungs und bedeutet das Anpeilen eines Zieles, das Vorhersagen eines Eintreffens unter wackligen Umständen. *Statistik* ist eine spezielle Anwendung der Wahrscheinlichkeitstheorie, in der aus empirischen Daten und *gewissen tieferen Einsichten* etwas über die zugrunde liegende Wahrscheinlichkeit vermutet wird. Die gewissen tieferen Einsichten haben mit dem Verständnis der zugrunde liegenden Ursachen zu tun und mit einem absolut klaren Verständnis von der Rolle der Wahrscheinlichkeit. Wahrscheinlichkeit steht also an erster Stelle.

1.2 Was ist Wahrscheinlichkeit?

Es sollte sich doch um ein mathematisches Objekt handeln, wie etwa der Begriff der Geraden in der Geometrie oder der Begriff der Zahl in der Analysis. Beiden Begriffen liegt ja etwas Intuitives (Anschauliches) zugrunde, und es ist diese Intuition, die uns ihrer (bis heute nicht erreichten und möglicherweise nie erreichbaren) wahren mathematischen Natur näher bringt: Wir haben zwar längst begriffen, daß Gerade und Zahl mathematisch doch ganz anders sind als zuerst empfunden (ich denke hier an nichteuklidische Geometrie und an Zahl als Cauchyfolge), aber ohne die erste Empfindung hätten wir: *nichts*. Unsere Intuition über Wahrscheinlichkeit ist anders, nicht faßbar, unklar als Begriff: In ihr steckt das, was wir Zufall[1] nennen, das Unvorhersagbare, das sich nicht sicher sein. Es ist das zu dem Gesetzmäßigen Gegensätzliche, es hat subjektiven Charakter und ist damit alles andere als mathematisch: Es ist nicht offenbarer Teil der Heraklitschen äußeren Welt oder später dann Teil der Platonischen Welt der existierenden Ideen, aus der die mathematischen Begriffe stammen, jene Welt, die für alle zugänglich ist und deren Begriffe für alle die gleiche Bedeutung haben. Es geht uns ja um das Begreifen des Gesetzmäßigen, um das ganze Gegenteil des subjektiven Empfindens also.

[1] wobei der Begriff des Zufalls nicht klarer ist, als der der Wahrscheinlichkeit — man gebraucht sie in gleicher Weise

Zufall erscheint da als krasser Gegensatz. Poincaré (in „Wissenschaft und Methode“, Teubner Verlag, Leipzig 1914, S.53) zitiert Bertrand: „Wie kann man wagen von den Gesetzen des Zufalls zu sprechen? Ist nicht der Zufall das Gegenteil aller Gesetzmäßigkeit?“

Ein naheliegender Gedanke ist es, Zufall als Ausdruck unserer Ignoranz zu sehen. Poincaré schreibt in „Wissenschaft und Hypothese“, (Teubner Verlag, Leipzig 1914, S. 190): „Wenn wir nicht unwissend wären, gäbe es keine Wahrscheinlichkeit“. Kann man daran denken, das persönliche Nichtwissen mathematisch zu fassen? Dann wäre Wahrscheinlichkeit ein Maß unserer Unkenntnis: „subjektive Wahrscheinlichkeit“. Sollte man sie in der Platonischen Welt der Ideen suchen, in der Hoffnung etwas Greifbares dort zu finden? Viele fanden das reizvoll; bekannt unter jenen ist Bayes und sogar Schrödinger. Er versuchte die Vermutungsstärke mathematisch zu fassen (Erwin Schrödinger: „The foundations of the theory of Probability I,II“, Proc. Royal Irish Academy, Vol. 51, 463-483, 1947).

Hier sind zwei Gründe, warum das attraktiv ist. (i) Der Versuch, eine objektive Wahrscheinlichkeit zu erkennen, ist fruchtlos: Der Münzwurf ist das Paradigma des Zufalls. Er legt nahe, Wahrscheinlichkeit als *relative Häufigkeit* zu setzen. In einer langen Münzwurfreihe kommt nämlich etwa gleichhäufig Kopf wie Zahl, also intuitiv mit objektiver *Wahrscheinlichkeit* 1/2 Kopf oder Zahl. Allerdings ist dieser Weg zu einer *Setzung* der Wahrscheinlichkeit nicht einsichtig, denn nur die wirklich *unendlichen* Folgen könnten bestenfalls eine Wahrscheinlichkeit *definieren*, aber in unserer endlichen Welt ist das keine einsehbare Setzung. Wenn wir uns aber auf endliche Folgen beschränken, dann brauchen wir eine *a priori Bewertung*, denn wie geht man mit langen „Durststrecken“ im Münzwurf um, in denen immer nur Kopf kommt? Wie sollen wir die deuten? (Das Gesetz der großen Zahlen, das wir ausführlich besprechen werden, liefert uns mit einer a priori Bewertung, dass und in welcher Form die relativen Häufigkeiten empirisch bedeutsam sind.) Also scheint eine direkte intuitive Setzung von objektiver Wahrscheinlichkeit kaum möglich. (ii) Im Alltäglichen redet man sowieso gerne von Wahrscheinlichkeiten von Einzelereignissen, die sich nur mit Anstrengung in eine Versuchsreihe von Ereignissen (also in die Sprache relativer Häufigkeiten) einbetten lassen: „Mit welcher Wahrscheinlichkeit scheint gleich die Sonne?“

Wahrscheinlichkeit als Vermutungsstärke ist darum ein naheliegender Ausweg. Aber nach einigem Nachdenken ist der Begriff „Vermutungsstärke“ doch viel zu komplex (genauso wie Information) und in einem fast greifbaren Widerspruch zur objektiven Welt der primitiven mathematischen Ideen, wie zum Beispiel der Idee eines Punktes oder einer Geraden.

Wir haben also das Problem, daß der Begriff des Zufalls nicht offenbar genug ist, d.h. nicht genügend selbsterklärend, um als primitive Größe gelten zu können.

Warum denken wir überhaupt, daß der Zufall etwas mit Mathematik zu tun hat? Weil es eine Gesetzmäßigkeit im Zufall gibt! Wir haben die Gesetzmäßigkeit bereits erwähnt und müssen sie nur richtig sagen: In einer Münzwurfreihe kommen *typischerweise* gleich oft Kopf und Zahl. Das ist eine neue Gesetzmäßigkeit, eine *typische* Gesetzmäßigkeit. Sie heißt „Gesetz der großen Zahlen", und die a priori Bewertung von der oben die Rede war, ist nichts anderes als die Festlegung von dem was „typisch" sein soll.

Wir werden uns also an Folgendes halten: „Typisch" ist objektiv und primär. Wenn wir zu einer gemeinsamen Idee von Wahrscheinlichkeit kommen wollen, müssen wir diese neue Art von Gesetzmäßigkeit in einer sonst regellosen Folge in primitiven *mathematischen Objekten* erkennen. Das würde uns der Sache näher bringen.

Das wird aber erst leicht, wenn man bereits Intuition über „Zahlen" entwickelt hat und zwar reelle Zahlen: Man greife blind eine Zahl $x \in [0,1]$. Was wird die typischerweise sein? Man denkt vielleicht $1/2$, weil die mittendrin liegt und weil man sie kennt. Aber das ist nicht gemeint. Gefragt ist nicht nach einer Zahl, die man gut kennt, sondern nach irgendeinem Bruch, einer der typisch ist, einer von der Sorte von Zahl, die am häufigsten vorkommt. Gefragt ist nach dem blinden Ziehen einer Zahl. Das ist schon eine schwierige Frage, denn wie kann es unter unendlich vielen Zahlen (überabzählbar unendlich viele Zahlen) welche geben, die häufiger sind als andere? Aber man weiß: Es gibt mehr irrationale Zahlen (enorm viele mehr) als rationale, aber dieses „mehr" ist natürlich schon eine Abstraktion und nicht mehr so primitiv wie der Begriff der Geraden. Aber nur auf dieser Ebene können wir weiterkommen: Schreibe die Zahl x in Dualdarstellung $x = 0,1001011110.......$ Dann sind alle $0,1-$ Folgen in eins zu eins Korrespondenz mit den Zahlen in $[0,1]$. Und nun ist die Frage schon klarer. Von welchem Typus wird die blind gezogene Zahl sein? Typischerweise wird es eine irreguläre Folge von Nullen und Einsen sein, allerdings mit einer Gesetzmäßigkeit. Die Anzahl von Nullen und Einsen werden im Mittel gleich sein; man nennt diese Art von Zahlen Normalzahlen. Dies ist intuitiv, aber es ist schwieriger zugänglich als euklidische Geometrie, denn die Intuition beruht auf Gelerntem über reelle Zahlen, und die reelle Zahl ist schon schwer genug zu begreifen. Aber man kann dieser Schwierigkeit nicht ausweichen, sie ist im *Wesen* der Sache.

Ich habe nun das Wort „typisch" betont: Was wird die Zahl „typischerweise" sein? Und ich sagte, dass das Wort „typisch" primitiv ist, dass es keiner weiteren Erklärung bedarf. Sollte man mehr dazu sagen müssen, kann man die Sache nur wie oben schon angedeutet sagen: Das, was am häufigsten vorkommt, ist typisch. Eine Analogie ist: Geometrie als Lehre von Ausdehnung. Ausdehnung ist primitiv, ein intuitiver Begriff, und den wollen wir mathematisch fassen. Dazu benutzen wir zuerst den Begriff von Geraden und Winkel und später den von linearer Unabhängigkeit von Vektoren. Wahrscheinlichkeit ist in diesem Sinne die Lehre vom typischen Verhalten. Wir brauchen dann eine Größe, die uns sagt, was typisch ist, die uns also sagt, was häufig und was weniger häufig ist. Diese mathematische Größe wird ein *Maß* sein, eine Abstraktion (eine Idee) eines Inhaltes bzw. eines Volumens, was wiederum eine Abstraktion von „Anzahl“ ist. Es ist eine andere Abstraktion von Anzahl als die Cantorsche Idee der Mächtigkeit einer Menge. Am Beispiel der Zahlen: Die typische Zahl gehört zu einer Menge von Zahlen, die einen großen Inhalt hat, wobei Inhalt so etwas wie die Länge eines Intervalls sein wird. Denn die Frage, mit welcher Wahrscheinlichkeit die blind gezogene Zahl in $[0,3;0,5]$ liegt, wird man intuitiv mit $0,2$ beantworten. Dieser Intuition müssen wir allemal gerecht werden. Die Schwierigkeit wird sein, dass diese Intuition über Wahrscheinlichkeit in eine über „typisch“ verwandelt werden muss.

Das ist ein grundlegender Teil der Wahrscheinlichkeitstheorie. Hinzu kommt eine ebenfalls ungewohnte, aber ungemein wichtige Struktur, die Wahrscheinlichkeitstheorie zu einer physikalischen Theorie macht und die uns gleichzeitig darin bestätigt, dass Wahrscheinlichkeit die Lehre des typischen Verhaltens ist: *Vergröberung.* Die Notwendigkeit zur Vergröberung sieht man sofort. Der Münzwurf ist ein physikalischer Prozess. Eine Münze ist ein Stück Materie, das durch die Luft gewirbelt wird. Eine komplizierte physikalische Bewegung findet statt. Aber was uns allein interessiert, ist Kopf oder Zahl. Das ist eine (subjektive) Vergröberung des tatsächlichen Geschehens. Und wir wissen schon, wie wir dieses vergröberte Bild zu beschreiben haben. Wir sagen einfach: Mit Wahrscheinlichkeit $1/2$ kommt jeweils Kopf oder Zahl. Aber die Frage ist, was das bedeutet. Da wir aber diese Aussage auf der Bildebene, also als (von uns) vergröberte Aussage machen, ist klar, daß wir auf dieser Ebene keinesfalls zu einer Einsicht in die Bedeutung kommen können. Wir brauchen die totale Einsicht in die Vergröberung, um zur Antwort auf die Frage zu kommen. Im Grunde wiederhole ich hier die Erkenntnis, dass die Vermutungsstärke ein zu komplexer Begriff ist, denn die Vermutungsstärke wird ja gerade auf vergröberte Situationen angewandt.

Wir müssen beim Münzwurf folgende Frage stellen: Warum liefert der physikalische Ablauf den gesetzmäßigen Zufall beim Münzwurf, der uns sagen läßt: Mit Wahrscheinlichkeit $1/2$ kommt Kopf oder Zahl? Dieses typische Verhalten muss bereits durch die physikalische Situation vorgegeben sein, und wenn wir etwas verstehen wollen, dann geht es nur auf der physikalischen Ebene. Dann muss es so sein, dass das fundamental Typische die Vergröberung mitmacht und auf der Bildebene als „Wahrscheinlichkeit $1/2$ “ erscheint. Wir müssen also mathematische Begriffe finden, die Vergröberungen und den Transport von „typisch“ zu der vergröberten Ebene hin erlauben. Ich habe den Begriff des „Typischen“ oben am Zahlenkontinuum erläutert, wobei man meinen könnte, dass das unnötig abgehoben sei: Warum nimmt man nicht einfach ein Gefäß mit tausend roten und einer weißen Kugel (eine Urne nennt man das in der Wahrscheinlichkeitstheorie), um zu erklären, dass man typischerweise ein rote Kugel zieht? Der Grund ist, dass das „Ziehen aus dem Kontinuum“ an der Wurzel von Wahrscheinlichkeit liegt, denn der physikalische zugrunde liegende Raum der möglichen Anfangsbedingungen eines physikalischen Verlaufs ist immer ein Kontinuum. Man muß das verstehen, denn die Konstruktion des Maßes auf dem Kontinuum, das uns das typische

Ziehen aus dem Kontinuum definieren soll, ist wie das Kontinuum selber keine simple Sache, es leidet wie alles „unendliche“ unter Zenos Nichtverstehbarem, und man muss hantieren und passend definieren, dass die Sache in Ordnung geht. Also muss man sich sicher sein, dass der Aufwand berechtigt ist.

Diese Begriffe: „Inhalt (synonym zu Maß) und Vergröberung“ bilden die Grundlage der sogenannten Axiomatik von Kolmogoroff. Wir werden uns die Einsicht erarbeiten müssen, um die Axiome als offenbar zu empfinden.

Natürlich werde ich an keiner Stelle Wahrscheinlichkeit als Begriff definieren. Ich werde keine mathematische Größe ausweisen, die Wahrscheinlichkeit an sich vertritt. Wahrscheinlichkeit wird ein Kürzel sein, ein Kürzel für eine vorhersagbare, weil typische Gesetzmäßigkeit in einer Regellosigkeit. Das bedeutet, dass die Vorhersage nicht mit Bestimmtheit, sondern nur typischerweise gilt. Und was das bedeutet werden wir ganz präzise sagen. Dennoch werde ich manchmal so reden, als gäbe es Wahrscheinlichkeit als mathematisches Objekt. Ich sage: „Mit Wahrscheinlichkeit 1/2 kommt im dritten Münzwurf Kopf.“ Gewohnheit ist das und ganz in Ordnung, wenn man verstanden hat, was gemeint ist.

Es gibt noch mehr zum Verhältnis zwischen Wahrscheinlichkeitstheorie und Physik zu sagen. Die physikalische Welt ist determiniert, d.h. eine mit Bestimmtheit ablaufende Welt. Wie kann der Zufall (im Sinne der prinizipiellen Regellosigkeit) darin Platz haben? Wie kann ein Ereignis einerseits festgelegt und andererseits zufällig sein? Darin wird noch einmal offenbar, dass der Zufall gar keine objektive Größe sein kann, sondern nur eine Erscheinung ist, die ganz sicher nicht mit unserem Unwissen verknüpft ist, aber die, weil wir unwissend sind, uns enorm hilft. Wir können unser weniges Wissen durch eine typische Gesetzmäßigkeit ausdrücken: Wüssten wir genau die Anfangsdaten des Münzwurfes (alle physikalischen Größen, die den Münzwurf determinieren), dann wüssten wir das typische Endergebnis mit Sicherheit. Wissen wir die nicht, dann erleben wir den typischen Ablauf dennoch und bekommen typische Endlagen.

Beispiel: Eine Normalzahl $x \in [0, 1]$ hat im Mittel genausoviele Nullen wie Einsen, egal ob wir diese Zahl nun kennen oder nicht. Also, gegeben die typische Zahl, dann ist nichts zufällig, dann wissen wir, welche Stellen kommen, aber dennoch erscheint die Abfolge von Nullen und Einsen zufällig. Zufall ist für uns vergröbernden Menschen eine phänomenologische Sache, und es ist nicht dramatisch — weil nicht im Wesen der Größe — dass es keinen primitiven mathematischen Vertreter gibt.

Damit sind wir ganz in der Denkweise von Laplace (1749-1827), dem Begründer der Wahrscheinlichkeitstheorie als mathematische Theorie. Für ihn war klar, daß kein Ding ohne erzeugende Ursache entstehen kann. Determinismus war Grundlage seines Denkens. Aber Wahrscheinlichkeitstheorie hat darin seinen berechtigten Platz. Laplace begann damit, das Maß für „typisch“ (Laplace-Wahrscheinlichkeit genannt) als Bruch von Anzahlen von Ereignissen zu setzen. Dagegen ist nur zu sagen, dass der *Begriff des Ereig-*

nisses unklar ist. Und nur deswegen wird die Setzung selber unklar, aber wenn man dem Gedanken Laplaces folgt und dabei zur fundamentalen Ebene der Beschreibung geht (wo Ereignisse eindeutig festgelegt sind), wird man zu den richtigen Einsichten geführt: in das Maß, das uns „typisch“ definiert und in die Vergröberungen, die uns die Ereignisse, die unseren groben Sinnen zugänglich sind, liefern.

Wahrscheinlichkeitstheorie ist die Theorie der *gesetzmäßigen Regellosigkeit*. Noch einmal: Als regellos bezeichne ich z.B. die Kopf-Zahl-Folge in einer Münzwurffolge, und mit gesetzmäßig meine ich die Tatsache, daß die relativen Häufigkeiten, mit denen Kopf und Zahl in einer solchen Folge auftreten, jeweils $\frac{1}{2}$ sind. Nun muss man Folgendes klar auseinander halten: Typisches Verhalten ist nicht notwendigerweise gesetzmäßig regellos. Man nehme einen Würfel, dessen Schwerpunkt nahe einer seiner Flächen ist, z.B. so, daß hauptsächlich 6 kommt. Das ist wenig regellos, und wir verstehen physikalisch, woran das liegt. Die Bewegung des Würfels ist nicht *instabil* genug. Die physikalische Grundlage der Regellosigkeit ist Instabilität. Smoluchowski erfand dafür das Kürzel „kleine Ursache — große Wirkung“, und besser als in seinem Aufsatz „Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik“ (Die Naturwissenschaften, Heft 17, 1918, Seite 253–263), kann man die Rolle der Instabilität für den Zufall nicht erklären.

Also muss man begreifen, dass „typisch“ und die Instabilität verschiedene Wahrheiten sind, die erst bei Zusammentreffen jene typische Gesetzmäßigkeit ergeben, die wir zufällig nennen.

Wenn man soweit gefolgt ist, wird man eine Frage in jedem Falle stellen wollen: Welche Einsicht legt eigentlich das grundlegende Maß fest, das uns „typisch“ definiert? Die Frage kann nur auf der grundlegenden Ebene, wo sie ja auch gestellt wird, beantwortet werden, und die Antwort ist einfach: Die Physik selbst muss dieses Maß bestimmen. Derart ist die Antwort von Ludwig Boltzmann und Willard Gibbs (ausgehendes 19. Jahrhundert), aber dennoch gibt es zu dieser Sache mehr (in der Tat vielmehr) zu sagen. Aber man muss schrittweise vorgehen, und wir beginnen naturgemäß mit dem ersten Schritt. Dennoch wünsche ich mir, dass diese Frage im Kopf bleibt und an vielen Stellen wiederholt wird, um das Ziel — zu verstehen, wie diese Frage durch die Physik beantwortet wird — nicht aus den Augen zu verlieren.

2. Jedermanns Wahrscheinlichkeit

Ich will erklären, was es mit dem Begriff Wahrscheinlichkeit auf sich hat. Dazu muß ich zunächst auf die üblichen Sprechweisen eingehen, damit die Sache nicht vom allgemeinen Verständnis entrückt dasteht, denn am Ende soll ja das allgemeine Verständnis nur auf den wahren Begriff fokussiert werden. Eine Erklärung des Begriffes, die nicht auf Einsicht gründen würde, wäre eine sehr nutzlose Übung. Aber dies bringt es mit sich, dass ich zunächst — insbesondere in diesem Kapitel — den Begriff der Wahrscheinlichkeit in einer noch unreflektierten Weise benutze — Jedermanns Wahrscheinlichkeit.

2.1 Laplace-Wahrscheinlichkeit

Jedermanns Wahrscheinlichkeit ist die Laplacesche Wahrscheinlichkeit, die für die Auswahl eines Objektes aus endlich vielen gleichen intuitiv angesetzt wird: Die Auswahl eines Objektes nennt man Elementarereignis und man setzt für die Wahrscheinlichkeit eines Elementarereignisses bei einer Gesamtzahl von N

$$W(\text{Elementarereignis}) = \frac{1}{N} = \frac{1}{\text{Anzahl aller Elementarereignisse}}. \qquad (2.1)$$

An dieser Setzung von Wahrscheinlichkeit, die man auch intutiv durch das offenbare Prinzip vom unzureichenden Grunde erläutern kann (ohne Grund wird keines der Objekte bevorzugt), ist zunächst alles in Ordnung und es ist jedem geraten, sich mit dieser Wahrscheinlichkeit vertraut zu machen. Sie ist eine praktikable vernünftige Setzung auf phänomenologischer Ebene. Ich greife jedoch schon vor, indem ich auf das *Problem* dieser Setzung hinweise. Es ist an ihr eines (und *nur* dieses) unklar: Der Begriff des Elementarereignisses! Der unterliegt einer Willkür, denn die Begriffe wie „Auswahl eines Objektes aus N Objekten“ , „ein Objekt aus N Objekten zu ziehen“ sind nicht primitiv. Jemand oder etwas, eine Maschine z.B. muß ziehen, und die Frage ist, ob nicht der Vorgang des Ziehens bei der Definition des Elementarereignisses berücksichtigt werden muß. Wenn aber der Begriff des Elementarereignisses dunkel ist, können wir unmöglich zu einer Einsicht über Wahrscheinlichkeit kommen. Darum müssen wir zuerst den Begriff des Elementarereignisses verstehen, um dann zu einer Abstraktion dieser Laplaceschen Setzung zu gelangen, aus der wir verstehen, was es mit Wahrscheinlichkeit auf sich hat, und warum und in welchem Sinne die obige Setzung funktioniert. Ich gebe ein einfaches Beispiel für eine Erscheinungsform dieser Problematik:

Beispiel 2.1.1. 1. Ein Würfel wird geworfen. Die Elementarereignisse — so wird man gewillt sein zu sagen — sind die Augenzahlen $\{1, 2, 3, 4, 5, 6\}$. Die Anzahl aller Elementarereignisse ist die Mächtigkeit dieser Menge $|\{1, 2, 3, 4, 5, 6\}| = 6$ und

$$W(\text{Augenzahl} = 5) = \frac{1}{6}$$

2. Zwei Würfel werden geworfen. Die Elementarereignisse sollen die Augensummen sein. Wenn ich das so sage, wird man das ersteinmal hinnehmen, warum nicht? Aber dann sieht man, daß

$$W(\text{Augensumme} = 3) = \frac{1}{11}$$

herauskommt und man zweifelt: Intuitiv sind doch die Elementarereignisse die *Paare* der Augenzahlen. Wie also geht man damit richtig um?

Um übrigens ausgehend von den Paaren die Wahrscheinlichkeit im 2. Fall des Beispiels auszurechnen, braucht man ein *Axiom*, denn man hat zwei Paare $(1, 2); (2, 1)$, die für die Augenzahl 3 günstig sind. Man addiert dann die Wahrscheinlichkeiten der Elementarereignisse (es gibt 36 Paare), um zur Wahrscheinlichkeit der Augensumme 3 zu kommen: $1/18$. Diese Additivität kann man mit relativen Häufigkeiten plausibel machen, wobei eine weitere Intuition, deren Hintergrund ich gar nicht ausführen will, hinzukommt: Das Gesetz vom Mittel. In einer langen Würfelreihe der Länge n komme n_5 mal die 5, wobei das Mittel (synonym für relative Häufigkeit)

$$\frac{n_5}{n} \approx \frac{1}{6} = W(5)$$

ist. Wenn nun 3 oder 5 interessant ist, dann kommen $n_3 + n_5$ mal diese Augenzahlen vor und dann sollte

$$W(3\,\text{oder}\,5) \approx \frac{n_3 + n_5}{n} \approx W(3) + W(5)$$

sein. Dies ist nur ein Versuch die Sache plausibel zu machen, das Gesetz vom Mittel wird am Ende ein beweisbarer Satz sein.

In jedem Falle setzt man die Additivität der Laplace-Wahrscheinlichkeit als Axiom an. Man kann das dann kurz so sagen: Ein Ereignis (z.B. „Augenzahl 3 oder 5“) ist eine Menge von Elementarereignissen und

$$W(\text{Ereignis}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}. \tag{2.2}$$

Es ist gedanklich und notationsmäßig günstig von der Menge Ω der Elementarereignisse zu sprechen, so daß die Ereignisse Teilmengen $M \subset \Omega$ sind. Damit wird (2.2) zu

$$W(M) = \frac{|M|}{|\Omega|},\ W(\Omega) = 1, \tag{2.3}$$

mit dem Symbol $|M|$ als Mächtigkeit der Menge M. Jedermanns Wahrscheinlichkeit reduziert sich also auf das Abzählen von Elementen von Mengen. Die Wahrscheinlichkeit wird in diesem Denkbehelf zu einer additiven Bewertung von Teilmengen: Falls $A \cap B = \emptyset$ dann ist

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

Dabei ergibt sich eine weitere Sprechweise für das Eintreten von Ereignissen, an die man sich gewöhnen muß. Wenn für $\omega \in \Omega$ das Elementarereignis $\{\omega\}$ eintritt, und $\omega \in A$ ist, dann tritt das Ereignis A ein. Wenn $B \subset A \subset \Omega$ und B eintritt, dann tritt A ein. Beispiel: Wenn beim Würfel $A =$ „Augensumme ist gerade“ $= \{2, 4, 6\}$ ist, und es wird 2 gewürfelt, dann tritt A ein. Diese Sprechweise ergibt sich aus der Bedeutung der Vereinigung von Mengen als das Eintreten von dem einen oder dem anderen, wobei das „oder“ das gemeinsame Eintreten beinhaltet, wohingegen Durchschnitte das logische „und“ vertreten. Mit dem Denkbehelf der Mengen kann man sich auch graphisch Durchschnitte und Vereinigungen, also „und“ und „oder“ Verknüpfungen anschauen und damit werden viele Wahrscheinlichkeitsregeln klar (Abb. **??**):

1. A^c ist das Komplement von A in Ω und heißt Gegenereignis zu A und $W(A) = 1 - W(A^c)$
2. $A \setminus B = A \cap B^c$ sind alle Ereignisse in A, die nicht in B sind, daher $W(A \setminus B) = W(A \cup B) - W(B)$
3. $W(A \cup B) = W(A) + W(B) - W(A \cap B)$. Dazu beachte man, dass $(A \setminus B) \cup A \cap B = A$ und $(A \setminus B) \cap A \cap B = \emptyset$, also ist $W(A) = W(A \setminus B) + W(A \cap B)$ und darauf wende (2) an.

Nummer (3) kann gewinnbringend verallgemeinert werden, indem die Wahrscheinlichkeit von „oder“ Ereignissen aus „und“ Ereignissen berechenbar wird:

$$\begin{aligned} W(A_1 \cup A_2 \cup A_3 ... \cup A_n) &= \sum_{i=1}^{n} W(A_i) \\ &- \sum_{i<j=2}^{n} W(A_i \cap A_j) \\ &+ \sum_{i<j<k=3}^{n} W(A_i \cap A_j \cap A_k) \\ &(-1)^{n-1}\, W(A_1 \cap ... \cap A_n). \end{aligned} \tag{2.4}$$

Diese Formel bekommt man sehr leicht auf viele verschiedene Weisen heraus, am direktesten mit Induktion, wobei man sich zunächst überlegt, wie man die Formel für drei Ereignisse zeigt (Das kann man noch gut graphisch begreifen). Das gibt einem dann den Weg wie man von $n-1$ auf n schließt.

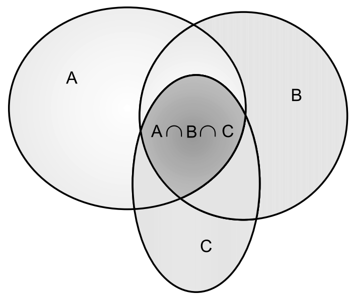

Abb. 2.1. symbolische Darstellung von Ereignissen als Mengen

Die Laplace-Wahrscheinlichkeit wird demnach zu einer Berechnung der Mächtigkeit endlicher Mengen, also zu einer Bestimmung der Anzahl der Elemente einer Menge. Das kann aufwendig sein. Oftmals sind es sogenannte kombinatorische Vorgehensweisen, die solche Anzahlen ergeben. Und es ist manchmal einfacher die Wahrscheinlichkeit von Gegenereignissen auszurechnen. Jedermanns Wahrscheinlichkeitsaufgaben sind häufig unbeliebt, weil sie meist nur umgangssprachlich und damit ziemlich unklar formuliert sind, so daß man sie umfomulieren muß, bis man erkennt, um welchen Typ von Aufgabe es sich handelt. Dabei versteht man zuweilen auch, daß die umgangssprachliche Formulierung nicht eindeutig ist. Das sollte man nicht überbewerten.

Bei Laplace Wahrscheinlichkeiten muß man die Anzahl von Elementen von Mengen berechnen. Die zugrunde liegenden Mengen (der Elementarereignisse) richten sich dabei nach der Art des Problems. Manchmal ist es hifreich, sich die Menge aus einem Urnenmodell zu besorgen, wobei die Urne ein mit gleichen numerierten Objekten (z.B. Kugeln) gefülltes Gefäß ist, und die Elementarereignisse einer Aufgabe werden als gezogene Objekte aus einer Urne interpretiert. Je nach Wesen der Aufgabe handelt es sich um verschiedene Vorgehensweisen des nacheinander Ziehens. Allerdings wird die Vorläufigkeit, die dem Begriff des Elementarereignisses noch zu eigen ist, auch in der Vielzahl von Möglichkeiten, mit der man eine gegebene Aufgabe behandeln kann, sich auch in Urnenmodellen niederschlagen.

Jedermanns Wahrscheinlichkeitsaufgaben sind in jedem Falle oft trickreich, man lernt schnell die Tricks, wenn man genügend Aufgaben gerechnet hat und man vergisst sie ebenso schnell wieder. Das macht diesen Teil der Wahrscheinlichkeitstheorie etwas unerfreulich.

2.2 Kombinatorik an Beispielen

Beispiel 2.2.1. Wenn N Leute in einem Raum sind, wie groß ist die Wahrscheinlichkeit dafür, daß mindestens zwei Leute am selben Tag Geburtstag haben? Das Jahr bietet 365 Tage als Geburtstage. Was sind die Elementarereignisse? Eine natürliche (wenn auch willkürliche[1]) Wahl sind alle gleichberechtigten Möglichkeiten von N-Tupeln von Tagen (den möglichen Geburtstagen), d.h.

$$\Omega = \{1, 2, ..., 365\}^N = \{\omega = (x_1, ...x_N), x_k \in \{1, 2, ..., 365\}\}.$$

Eine Grundaufgabe ist es also, die Anzahl aller N-Tupel anzugeben:

- 365 Möglichkeiten für x_1, jede dieser erlaubt weitere
- 365 Möglichkeiten für x_2 jede dieser erlaubt weitere
- 365 Möglichkeiten für x_n u.s.w. bis N

Insgesamt also multiplizieren sich die Möglichkeiten eines jeden Eintrages mit allen anderen. Damit ist die Anzahl aller N-Tupel, d.h. die Anzahl aller Möglichkeiten und $|\Omega| = 365^N$. Nun zum Ereignis G_N „mindestens zwei Leute haben am selben Tag Geburtstag“. Das ist die Menge aller N-Tupel mit mindestens zwei gleichen Einträgen. Also brauchen wir die Anzahl aller dieser N-Tupel. Leichter: Alle N-Tupel mit keinen gleichen Einträgen. Das ist das Gegenereignis $G_N^c =$„keine zwei Leute haben am selben Tag Geburtstag“. Denn dann können wir so vorgehen:

- 365 Möglichkeiten für x_1, jede dieser erlaubt weitere
- 364 Möglichkeiten für x_2 (denn x_2 darf ja nicht gleich x_1 sein) jede dieser erlaubt weitere
- bis $365 - N + 1$ Möglichkeiten für x_N .

Das ergibt $|G_N^c| = 365 \cdot 364 \cdot ...(365 - N + 1)$. Man könnte verwirrt meinen, daß an Platz 1 nach obiger Überlegung 365 Tage stehen können und an Platz 2 nur 364 Tage, was nicht stimmt. Alle Plätze werden mit gleichvielen Tagen besetzt. Die Auswahl von x_1 geschah ohne Einschränkung.

$$\text{Also} \quad W(\text{keine zwei Leute haben am selben Tag Geburtstag})$$

$$= \frac{365 \cdot 364 \cdot \ldots \cdot (365 - N + 1)}{365^N}.$$

Man sollte einmal seine a–priori Einschätzung zahlenmäßig überprüfen: Wie viele Leute müssen im Raum sein, damit $W \leq \frac{1}{2}$ wird, das heißt

$$W(\text{wenigstens zwei Leute haben am selben Tag Geburtstag}) =$$

$$1 - W(\text{keine zwei Leute haben am selben Tag Geburtstag}) \geq \frac{1}{2}?$$

[1] Ein Gebutstag setzt einen Zeugungstag voraus, der ist doch sicher elementarer.

(Es kommt $N = 23$ heraus.)

Wie groß ist die Wahrscheinlichkeit für mindestens einen gleichen Geburtstag, falls 366 Leute im Raum sind?

Beispiel 2.2.2. **m-maliges Ziehen aus n Objekten mit Zurücklegen mit Beachtung der Reihenfolge.**

In einem Urnenmodell sieht das so aus: Man hat eine Urne mit 365 Kugeln, numeriert, sonst aber gleich. Man holt blind aus der Urne eine Kugel, notiert die Nummer, legt sie zurück (schüttelt die Urne), und faßt erneut blind hinein. Das ganze N mal. Dieses N-malige blinde Ziehen von Kugeln soll als „Elementarereignisse" die N-Tupel von Geburtstagen der Leute 1 bis N liefern.

Nun allgemein: Man habe eine Urne, d.h. eine Menge $U(n)$ mit n Kugeln (Elementen). $m-$maliges Ziehen mit Zurücklegen liefert m-Tupel mit Einträgen aus $U(n)$. Die Menge aller solcher m-Tupel ist

$$T_m(n) = U(n)^m = \{(x_1, ..., x_m); x_i \in U(n)\}.$$

Die Mächtigkeit dieser Menge ist, wie wir bereits festgestellt haben,

$$|T_m(n)| = |U(n)^m| = n^m.$$

Und zur Beachtung der Reihenfolge ist zu sagen: Die Tupel $(1, 2, 3, 4, 5)$ und $(2, 1, 3, 4, 5)$ sind verschiedene Elementarereignisse. Die für die Geburtstagsaufgabe relevanten Tupel sind solche mit mindestens zwei gleichen Einträgen. Z.B. hat $(1, 1, 2, 3, 4)$ zwei gleiche Einträge. Die Tupel in U^m enthalten alle möglichen Wiederholungen, und man findet in der Kombinatorik auch hierfür den Begriff der *Variationen mit Wiederholungen*. Ich gebrauche lieber den Begriff der $m-$**Tupel**

Beispiel 2.2.3. **m-maliges Ziehen aus $n \geq m$ Objekten ohne Zurücklegen mit Beachtung der Reihenfolge**

Um die Geburtstagsaufgabe zu erledigen, haben wir auf jene Tupel geachtet, die keine Wiederholungen besitzen.

Allgemein sei nun $|U(n)| = n \geq m$, Dann liefert das Ziehen ohne Zurücklegen alle m-Tupel, die keine gleichen Einträge haben. Man nennt solche m-Tupel **Variationen** (ohne Wiederholungen):

$$V_m(n) = \{(x_1, ..., x_m); x_i \in U(n),\ x_i \neq x_j \text{ für alle } i \neq j\} \subset U(n)^m,$$

deren Mächtigkeit wir in Beipiel (**??**) bestimmt haben:

$$|V_m(n)| = n(n-1) \cdot \ldots \cdot (n-m+1). \tag{2.5}$$

Von speziellem Interesse ist der Fall $m = n$, dann erhalten wir als Variationen die **Permutationen** der n Objekte und ihre Anzahl ist

Anzahl aller Permutationen von n Objekten $= |V_n(n)| = n \cdot n-1 \cdot ... \cdot 1 =: n!$.

Damit ist (**??**) schreibbar als

$$|V_m(n)| = \frac{n!}{(n-m)!} \tag{2.6}$$

Die Gültigkeit solcher kombinatorischen Formeln für allgemeines n folgert man sehr leicht mit vollständiger Induktion.

Variationen kommen in Frage, wenn es um die Tätigkeit des Auswählens geht: Wieviele Möglichkeiten gibt es, m Objekte aus n gleichartigen auszuwählen. Vielleicht ist nicht klar was gemeint ist, darum ausführlicher. Reihe n Kugeln auf: $1, 2..., n$. Markiere m mit einem roten Punkt. Macht es einen Unterschied, ob ich erst Kugel 1 rot markiere und dann Kugel 3 oder umgekehrt? Macht die Reihenfolge einen Unterschied? Für mich als Markierer macht es sicher einen Unterschied, es sind verschiedene Möglichkeiten: Variationen.

Beispiel 2.2.4. Beim Lotto 6 aus 49 muß ein Spieler 6 Zahlen ankreuzen. Wieviele Möglichkeiten hat er, das zu tun? Da sind die Variationen gefragt: $49 \cdot 48 \cdot \ldots \cdot 44$. Mit welcher Wahrscheinlichkeit hat er 6 richtige? Es gibt 6! richtige Möglichkeiten, also mit (**??**)

$$W(6\text{richtige}) = \frac{6!43!}{49!}.$$

Beispiel 2.2.5. Ein Jedermanns Beispiel aus dem Zauberhut: n numerierte Karten werden nicht aufgedeckt ausgebreitet. Ein Spieler rät von jeder Karte die Nummer. Wie groß ist die Wahrscheinlichkeit für mindestens eine Übereinstimmung? Egal wie die Karten auf dem Tisch liegen, der Spieler errät eine Permutation von $1, 2, ..., n$. Gefragt ist also die Wahrscheinlichkeit für mindestens eine Übereinstimmung zwischen Permutationen.

Viele Jedermanns Aufgaben berühren wie hier folgende technische Frage: Bei n Objekten, gibt es $n!$ Permutationen von Anordnungen der Objekte. Wieviele Permutationen lassen dabei wenigstens ein Objekt unberührt? Keine einfache Frage! Aber mit einem Trick geht es schnell: Man numeriere die Objekte durchgehend. Sei $V_n(n) = \{\omega; \omega = \text{Permutation von}\,(1, 2, ..., n)\}$, und $A_j = \{\omega \in V_n(n); \omega_j = j\}$, wobei ω_j die $j-te$ Koordinate des $n-$Tupels ω ist. Das bedeutet, A_j enthält alle Permutationen die zumindest die Stelle j unverändert lassen. Dann sind die günstigen Ereignisse in $A_1 \cup A_2 \cup ... \cup A_n$ und gefragt ist nach der Mächtigkeit $|A_1 \cup A_2 \cup ... \cup A_n|$ dieser Menge. Indem wir durch $|V_n(n)| = n!$ dividieren können wir auch $W(A_1 \cup A_2 \cup ... \cup A_n)$ aufschreiben und dafür haben wir (2.4), und wir brauchen $W(A_j), W(A_i \cap A_j)$ usw.. Diese Terme sind aber einfach anzugeben. In A_j sind alle Permutationen, bei denen j unverändert bleibt, d.h. $n-1$ Objekte werden permutiert, also $|A_j| = (n-1)!$, für alle j. Daher

$$W(A_j) = \frac{(n-1)!}{n!} = \frac{1}{n}, \text{ für alle } j.$$

Genauso geht es weiter: In $A_i \cap A_j$ sind i und j fest, also werden $n-2$ Objekte permutiert, das gibt $(n-2)!$ Permutationen und

$$W(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \text{ für alle } i \neq j \ .$$

Die weiteren Terme sind damit klar und wir müssen uns nur noch überlegen *wieviele Summanden* die jeweiligen Summen in (2.4) summieren:

$$\sum_{i>j=1}^{n} 1 = ?, \quad \sum_{i>j>k=1}^{n} 1 = ?.$$

Man wird etwas nachdenken müssen, um zu sehen, daß hier nach etwas natürlichem Neuem gefragt wird, nämlich nach den $2-$, $3-, ..., k-$ elementigen Teilmengen von $\{1, 2, ...n\}$. Diese sind ebenfalls im Urnenmodell formulierbar:

Beispiel 2.2.6. **m-maliges Ziehen aus $n \geq m$ Objekten ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge.**

Wir wollen alle $m-$ Variationen (Tupel der Länge $m \leq n$), die untereinander durch Permutationen hervorgehen, als ein Elementarereignis werten. Wir können dies bildhaft ausdrücken, indem wir sagen, daß die Reihenfolge, in der gezogen wird, irrelevant ist.

Diese Zählung aller solcher permutierten $m-$Tupel als eins ist äquivalent zur Setzung eines Elementarereignisses als $m-$elementige Teilmenge von U, die man auch **Kombinationen** nennt:

$$K_m(n) =: \{m - \text{elementige Teilmengen} \subset U\}.$$

Deren Mächtigkeit ist leicht zu finden. Es gibt ja $\frac{n!}{(n-m)!}$ Variationen der Länge m. Aber eine Variation taucht darin $m!$ permutiert oft auf. Das müssen wir also rausdividieren, oder anders, für die Anzahl aller $m-$elementigen Teilmengen von n Objekten gilt

$$|K_m(n)| = |\{m - \text{elementige Teilmengen} \subset U\}| \cdot m! = \frac{n!}{(n-m)!},$$

d.h.

$$|K_m(n)| = \binom{n}{m} := \frac{n!}{m!(n-m)!}.$$

Definition 2.2.1.

$$\binom{n}{m} := \frac{n!}{m!(n-m)!}$$

mit

$$\binom{n}{0} := 1, \binom{n}{m} := 0, \; \textit{falls}\, m > n \; \textit{oder}\, m < 0$$

heißt Binomialkoeffizient.

Er ist von herausragender Bedeutung. Das werden wir nachher deutlich sehen. Weiter mit Beispiel **??**. Demnach ist

$$\sum_{i>j=1}^{n} 1 = \binom{n}{2}$$

$$\sum_{i>j>k=1}^{n} 1 = \binom{n}{3}$$

usw.. Damit wird

$$\begin{aligned} W(A_1 \cup A_2 \cup A_3 ... \cup A_n) &= \binom{n}{1} W(A_1) \\ &- \binom{n}{2} W(A_1 \cap A_2) \\ &+ \binom{n}{3} W(A_1 \cap A_2 \cap A_3) \\ &(-1)^{n-1} \binom{n}{n} W(A_1 \cap ... \cap A_n) \\ &= \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!}. \end{aligned}$$

An diesem Resultat wird üblicherweise das Verhalten für große n als interessant angesehen, denn

$$\lim_{n\to\infty} \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!} = 1 - e^{-1}.$$

Zurück zum Jedermanns Beispiel: Die Wahrscheinlichkeit des Gegenereignisses, daß nämlich keine richtige Karte benannt wird ist $\approx 1/e \approx 1/3$, wenn genügend viele Karten auf dem Tisch liegen. Das mag überraschen, denn man würde wohl eher an eine Wahrscheinlichkeit nahe 1 denken, dass bei sehr vielen Karten wenigstens einmal richtig geraten wird.

Allein um die Sache sprachlich zuende zu bringen, wird man nun noch nach folgendes anschauen wollen:

Beispiel 2.2.7. $m-$**maliges Ziehen aus** n **Objekten mit Zurücklegen und ohne Beachtung der Reihenfolge**

Also angeschaut werden $m-$Tupel mit Wiederholungen, wobei alle Permutationen eines $m-$Tupels als eins gewertet werden. Man nennt diese Dinger auch **Kombinationen mit Wiederholungen**, und deren Anzahl ist nun nicht mehr so einfach zu überlegen. Vor allem ist die Antwort nicht mehr suggestiv. Für $n = 2$ ist das noch leicht hinschreibbar, aber dann wird es etwas undurchsichtig. Beispiele sind $m = 4, n = 2$:

$$(1111), (1112), (1122), (1222), (2222).$$

Die Formel selber ist allerdings wieder leicht durch Induktion beweisbar. Dabei muß Summationsregeln für die Binomial-Koeffizienten benutzen, die letztendlich auch für die endgültige Formel verantwortlich sind.

Die Anzahl aller $m-$Kombinationen mit Wiederholungen aus n Objekten ist

$$\binom{n+m-1}{m.}$$

Für $m = 4, n = 2$ ist $\binom{2+4-1}{4} = 5$.

2.3 Der Binomialkoeffizient

Beispiel 2.3.1. Im Beipiel **??** hatten wir

$$W(6\text{richtige}) = \frac{6!43!}{49!} = \frac{1}{\binom{49}{6}}.$$

Ist dieses Ergebnis, daß nur die Anzahl von 6-elementigen Teilmengen relevant ist, direkt erklärbar? Ja, denn die Anzahl Mehrmöglichkeiten, die sich durch Änderung der Reihenfolgen beim Ankreuzen ergeben, werden durch die Normierung bei der Wahrscheinlichkeit wieder herausgekürzt. Das passiert in vielen analogen Fällen und man ignoriert oft die Tatsache, daß man grundlegend eigentlich Variationen betrachten muß, aber da die Endformel eine ist, in der dann nur noch die Anzahlen von Kombinationen eingehen, fokussiert man gleich auf Kombinationen. Manchmal kann man die Kombinationen schon als Elementarereignisse empfinden: Es kommt doch nur auf die Zahlenkombinationen an, nicht auf die Reihenfolge nach sie ausgefüllt wurden. Dem liegt natürlich die Problematik des Begriffes des Elementarereignisses zugrunde.

Beispiel 2.3.2. Man habe n Objekte (100 Schrauben in einem Karton), darunter k markierte (15 fehlerhafte). Man ziehe m (10 Schrauben entnehmen), wie groß ist die Wahrscheinlichkeit darunter j (2 fehlerhafte) markierte zu haben? Dies ist zunächst eine Aufgabe, die nach der Anzahl der Variationen fragt, denn jede Reihenfolge, nach der man die Objekte entnimmt, ist ja ein anderes Elementarereignis. Aber wir können gleich auf Kombinationen gehen, denn wie beim Lotto kürzen sich die Mehrmöglichkeiten, die sich durch Reihenfolgenänderungen ergeben heraus. Also ganz schnell: Es gibt $\binom{n}{m}$ Teilmengen mit m Elementen, die günstigen enthalten j markierte und $m-j$ nicht markierte. Es gibt $\binom{n-k}{m-j}$ Teilmengen der unmarkierten Objekte, jede dieser Teilmengen kann mit einer der $\binom{k}{j}$ Teilmengen der markierten zu $m-$elementigen Teilmengen komplettiert werden. Demnach

$$W(j \text{ markierte in}\, m) = \frac{\binom{n-k}{m-j}\binom{k}{j}}{\binom{n}{m}}$$

und für das Jedermanns Beispiel

$$W(2\,\text{fehlerhafte in}\, 10) = \frac{\binom{85}{8}\binom{15}{2}}{\binom{100}{10}}.$$

Beachte, daß sich folgende Summationsformel ergibt:

$$\sum_j \binom{n-k}{m-j}\binom{k}{j} = \binom{n}{m},$$

wobei sich die Summation nur über $0 \leq j \leq k$ erstreckt, weil alle überzähligen Terme null sind.

Eines der Leitbeispiele dieses Buches wird der Münzwurf sein. Eine typische Frage ist: Mit welcher Wahrscheinlichkeit kommt genau $m-$ mal Kopf in n Würfen? Das ist nun eine Kombinationsaufgabe: Gefragt ist nach der Anzahl der Kopf-Zahl-Folgen der Länge n (deren Gesamtzahl ist 2^n, denn es sind die $n-$Tupel aus 2 Elementen) in denen genau m Köpfe stehen. Die ist natürlich gleich der Anzahl aller 0-1-Folgen der Länge n, mit genau m Einsen. Nummeriere die Plätze, dann ist die Auswahl der Plätze, auf denen eine Eins steht, eine Teilmenge von $\{1, 2, ..., n\}$. Darum ist offenbar die Anzahl der gefragten Folgen gleich der Anzahl der $m-$elementigen Teilmengen von $\{1, 2, ..., n\}$, also $\binom{n}{m}$. Damit ist

$$W(\text{genau } m \text{ mal Kopf in } n \text{ Würfen}) = \frac{\binom{n}{m}}{2^n}. \tag{2.7}$$

Wir können von 0-1-Folgen auf Folgen mit mehreren Ziffern verallgemeinern: Betrachte Folgen der Länge n mit Einträgen aus $\{1, 2, ..., r\}$. Dann betrachte man jene Folgen mit n_1-mal die 1,...,n_r-mal die r und $n_1 + \ldots + n_r = n$. Diese Anzahl ist gleich der Anzahl der Anordnungen (ohne Beachtung der Reihenfolge) von n gleichen Objekte auf r Kästen mit n_k Objekten in Kasten k. Diese Anzahl ist leicht zu kriegen: Für 1 gibt es $\binom{n}{n_1}$ mögliche Plätze, und jede dieser Möglichkeiten erlaubt $\binom{n-n_1}{n_2}$ Möglichkeiten, und führt man das fort sieht man leicht den Multinomialkoeffizienten

$$\binom{n}{n_1, n_2, \ldots, n_r} := \frac{n!}{n_1! \ldots n_r!}$$

für die Anzahl der möglichen Aufteilungen der n Objekte in r *verschiedene* Gruppen der Größen $n_1, n_2, \ldots, n_r$ entstehen.

Dieselbe Kombinatorik steckt übrigends auch im Multinomialsatz:

$$(x_1 + \ldots + x_r)^n = \sum_{n_1+\ldots+n_r=n} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} \ldots x_r^{n_r}, \tag{2.8}$$

der für $r = 2$ der Binomialsatz ist:

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

Dieser Satz liefert uns noch einmal die Normierung

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

Übrigens ist die Anzahl der Summanden in der Summe in (**??**), also die Anzahl der Zerlegungen von n in Summen von $n_1, .., n_r \leq 0$ gerade die Anzahl von $n-$Kombinationen mit Wiederholungen aus r Objekten, also gemäß Beispiel **??**

$$\binom{r+n-1}{n}.$$

Anmerkung 2.3.1. Die Anzahl der $m-$ elementigen Teilmenge einer $n-$ elementigen Menge Ω ist $\binom{n}{m}$, damit ist $\sum_{k=0}^{n} \binom{n}{k}$ die Anzahl aller Teilmengen, also die Mächtigkeit der **Potenzmenge** $\mathcal{P}(\Omega)$, der Menge aller Teilmengen:

$$|\mathcal{P}(\Omega)| = 2^n.$$

Die Potenzmenge von $\mathbb{N}$ ist offenbar eins zu eins mit der Menge aller unendlichen 0-1-Folgen. Die ist wiederum eins zu eins mit den Zahlen des Intervalls $[0, 1]$, was selber eins zu eins mit $\mathbb{R}$ ist. Darum ist die Mächtigkeit von $\mathbb{R}$ einfach $|\mathbb{R}| = 2^{|\mathbb{N}|}$. Man nennt die Zahl $|\mathbb{N}| = \aleph_0$ (sprich aleph-null). Während $\mathbb{N}$ nach Definition eine abzählbare unendliche Menge ist, ist $\mathbb{R}$ nicht mehr abzählbar. Man nennt $\mathbb{R}$ das (eindimensionale) Kontinuum.

Die Binomialkoeffizienten kann man sich sehr schön am Kugellauf des Galtonbrettes Abb. **??** veranschaulichen, wobei die Links-Rechts-Abgänge bei insgesamt n Nagelreihen eins zu eins mit den Kopf-Zahl-Folgen der Länge n sind.

Wenn wir die Anzahl der Pfade aufschreiben, die an jedem Nagel des Galtonbrettes Abb. **??** passieren können ergibt sich bei vier Nagelreihen folgendes Schema:

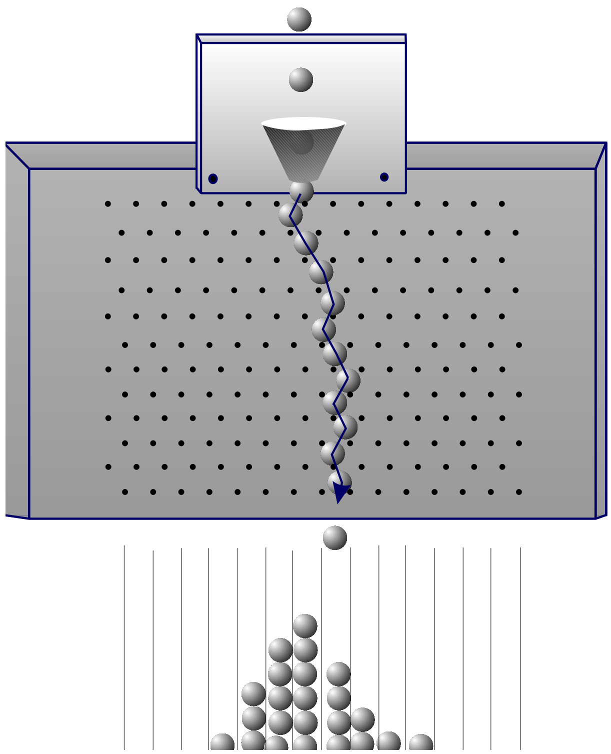

Abb. 2.2. Das Galtonsche Brett. Kugeln fallen (zentral) auf Nagelstifte und werden nach rechts oder links mit gleicher Wahrscheinlichkeit abgelenkt.

$$
\begin{array}{ccccccccccc}
 & & & & 0 & & & & & & n=0 \\
 & & & 1 & & 1 & & & & & n=1 \\
 & & 1 & & 2 & & 1 & & & & n=2 \\
 & 1 & & 3 & & 3 & & 1 & & & \cdot \\
1 & & 4 & & 6 & & 4 & & 1 & & \cdot
\end{array}
$$

Auf der Ebene n stehen von links ($k = 0$) nach rechts ($k = n$) die Zahlen $\binom{n}{k}$. Wir lesen ab, daß

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

Diese Beziehung kann auch als Definition für die Binomialkoeffizienten dienen.

Das Galtonbrett diente hier nur als intuitive Veranschaulichung, die man sofort gewillt ist, zu akzeptieren. Dabei gibt es zum Galtonbrett mehr zu sagen, als dieses Buch in der Lage ist, aufzunehmen. Eine Frage, die wir nur oberflächlich behandeln können ist die, in der Abbildung **??** angedeutete: Wie ist die Verteilung der Endplätze der Kugeln, wenn man viele Kugeln nacheinander durch das Nagelbrett fallen läßt? Intuitiv wird der Bruchteil der Kugeln die auf Platz k landen bei n Nagelreihen $\binom{n}{k}$ sein. Die Formalisierung dieser Antwort geschieht später unter **??**. Sie ist mit dem **Gesetz vom Mittel** verbunden, daß wir uns jetzt in der einfachsten Version anschauen.

2.4 Das Gesetz vom Mittel

Jedermanns Gefühl über die Regellosigkeit in einer sehr langen Münzwurfreihe (n sehr groß) ist, das ungefähr $\frac{n}{2}$ oft Zahl und ungefähr $\frac{n}{2}$ oft Kopf kommen sollten. Das können wir analytisch untersuchen, indem wir (**??**) für große n studieren. Das ist die klassische Aufgabe der Wahrscheinlichkeitstheorie schlechthin, die man einmal verstanden haben muß, um den Abstraktionen mit Einsicht begegnen zu können. An der Wurzel der Asymptotik für große n ist

Satz 2.4.1. Stirlingsche Formel

$$n! = \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

(Ich erinnere an die Bedeutung des Landauschen Ordnungs-Symbols $\mathcal{O}(x)$. Bei Division eines Terms der Ordnung $\mathcal{O}(x)$ durch x bleibt der Bruch für $x \to \infty$ beschränkt.) Der sehr interessante Beweis der Stirlingformel folgt später in Anmerkung **??**. Ganz grob soll man $n! = \left(\frac{n}{\mathrm{e}}\right)^n$ lesen (ganz unbedarft könnte man $n/2$ statt n/e vermuten).

Wir wenden das jetzt auf $\binom{n}{k}$ an, und dabei sei k ein Bruchteil von n, zum Beispiel $k = pn, 0 < p < 1$, so daß wir auch für k die Stirlingformel ansetzen können.

$$\begin{aligned}
\binom{n}{pn} &= \frac{n!}{(np)!(n-np)!} \\
&\approx \frac{n^n \mathrm{e}^{-n}\sqrt{2\pi n}}{(np)^{np}\mathrm{e}^{-np}\sqrt{2\pi np}(n(1-p))^{n(1-p)}\mathrm{e}^{-n(1-p)}\sqrt{2\pi n(1-p)}} \\
&= \frac{1}{\sqrt{2\pi n}\sqrt{p(1-p)}p^{np}(1-p)^{n(1-p)}}.
\end{aligned}$$

Dann speziell für $p = \frac{1}{2}$:

$$\binom{n}{\frac{n}{2}} \approx \frac{2 \cdot 2^n}{\sqrt{2\pi n}},$$

das heißt mit (??)

$$W\left(\text{genau } \frac{n}{2} \text{ Köpfe in } n\right) \approx \sqrt{\frac{2}{\pi n}} \sim \frac{1}{\sqrt{n}}. \tag{2.9}$$

Man wird auf den ersten Blick enttäuscht sein, dass nicht 1 herauskommt. Aber natürlich kann das gar nicht sein, denn wir haben ja *nur* solche Folgen erfasst, die genau $n/2$ Köpfe haben. Dabei ist unser Gefühl berechtigterweise nur über ungefähre Verhältnisse, d.h. nur asymptotisch sollten wir als relative Häufigkeit $1/2$ erwarten. Auch Folgen mit $\frac{n}{2} \pm n^{1-\epsilon}, \epsilon > 0$ Köpfen zeigen asymptotisch eine relative Häufigkeit von $1/2$ für Köpfe. Ganz grob können wir aber schon sagen, daß $\sqrt{n}$ Terme um $n/2$ herum (deren Wahrscheinlichkeitswerte ebenfalls von der Größenordnung $\frac{1}{\sqrt{n}}$ sein werden) bereits die Gesamtwahrscheinlichkeit 1 liefern! Insofern ist (??) bereits eine grobe Fassung des Gesetzes der großen Zahlen: Man kann daraus erwarten, daß nur $\sqrt{n}$ viele Terme in der Nachbarschaft von $k = \frac{n}{2}$ eine nennenswerte Wahrscheinlichkeit haben. Und das bedeutet, daß *typischerweise* die Anzahl von Köpfen in n Würfen im Bereich $\frac{n}{2} \pm \sqrt{n}$ ist. Relativ zu n geht also die Häufigkeit von Köpfen gegen $\frac{1}{2}$ mit relativer Schwankung $\pm\frac{1}{\sqrt{n}}$. (Nicht absolut: Fluktuationen $\sim \sqrt{n} \to \infty$ für $n \to \infty$, das heißt es kommen beliebig lange „Durststrecken“ vor, also Strecken in denen z.B. nur „Kopf“ kommt.) Damit bestätigt die Laplace-Wahrscheinlichkeit unser Gefühl für Zufall und die gesetzmäßige Regellosigkeit: Im Mittel (d.h. als relative Häufigkeit) kommt typischerweise $\frac{n}{2}$ mal Kopf und $\frac{n}{2}$ mal Zahl.

Wir können das etwas mehr ausarbeiten. Wir untersuchen für

$$p = \frac{1}{2} + \epsilon, \ \epsilon \text{ klein}$$

$$\frac{1}{2^n}\binom{n}{(\frac{1}{2}-\epsilon)n} \approx \frac{1}{2^n}\frac{1}{\sqrt{2\pi n}\sqrt{p(1-p)}p^{np}(1-p)^{n(1-p)}}.$$

Eine leichte Rechnung bringt

$$\frac{1}{2^n}\binom{n}{(\frac{1}{2}-\epsilon)n} \approx \frac{1}{\sqrt{2\pi n}}\frac{1}{\sqrt{\frac{1}{4}-\epsilon^2}}\frac{1}{(1-4\epsilon^2)^{n/2}}\left(\frac{1+2\epsilon}{1-2\epsilon}\right)^{n\epsilon}.$$

Beachte, daß

$$\frac{1}{(1-4\epsilon^2)^{n/2}}\left(\frac{1+2\epsilon}{1-2\epsilon}\right)^{n\epsilon}$$

$$= \exp - \left(\frac{n}{2}\ln\left(1-4\epsilon^2\right) + n\epsilon(\ln(1+2\epsilon) - \ln(1-2\epsilon))\right),$$

und $\ln(1+x) = x + \mathcal{O}(x^2)$, also entsteht am Ende

$$\frac{1}{2^n}\binom{n}{(\frac{1}{2}-\epsilon)n} \approx \frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{\frac{1}{4}-\epsilon^2}} \exp -2n\epsilon^2 \; . \tag{2.10}$$

Wir gehen nun etwas großzügig so weiter:

$$\begin{aligned} W\left(\left|\frac{\text{Anzahl Köpfe in } n}{n} - \frac{1}{2}\right| > \epsilon\right) &= \sum_{k<\frac{n}{2}-\epsilon n; k>\frac{n}{2}+\epsilon n} \frac{1}{2^n}\binom{n}{k} \\ &\leq n\frac{1}{2^n}\binom{n}{\frac{n}{2}+\epsilon n} \\ &\approx \frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{\frac{1}{4}-\epsilon^2}} \exp -2n\epsilon^2 \\ &\to 0 \quad \text{für} \quad n \to \infty \; . \end{aligned}$$

was wir als Gesetz vom Mittel ansehen.

Wir haben hier übrigends viel mehr an Rechnung investiert als notwendig ist: Wenn es nur um das Gesetz vom Mittel geht interessiert uns nicht, wie die Wahrscheinlichkeit der Abweichung klein wird, sondern nur, daß sie klein wird. Und das läßt sich abstrakt (und deswegen kommt es später) ganz einfach zeigen. Allerdings muß man die Abstraktion verstanden haben, man muß verstanden haben, daß da Strukturen abstrahiert werden, deren Gültigkeit für unsere Welt nur schwer nachweisbar sind. Damit wird die absolute Primitivität des Beweises des Gesetzes der großen Zahlen, das wir also im späteren Kapitel besprechen, relativiert werden. Im obigen Falle ergab sich die Mehrinformation (**??**) fast zwanglos, und es ist gut, die Sache einmal einfach ausgerechnet zu haben. (**??**) beinhaltet die Glockenkurve, die sich bei vielen Kugeln in den Auffangbehältern als Häufigkeitsverteilung ergibt.

Anmerkung 2.4.1. **Über die Bedeutung der Ergebnisse**
Das Gesetz vom Mittel besagt, daß für große Zahlen n die relative Häufigkeit von Köpfen in Folgen der Länge n typischerweise, d.h. für die **meisten** aller Folgen $\approx 1/2$ ist. Dieses Gesetz der großen Zahlen (und nur dieses Gesetz) macht Wahrscheinlichkeitstheorie experimentel überprüfbar.

Der Ausdruck „die meisten" ist dabei völlig klar. Man zählt einfach die Menge der Folgen mit den gefragten Eigenschaften ab, und unsere Rechnung oben ist nichts weiter als eine Abschätzung der Anzahlen. Später werden wir diesen Begriff des Typischen abstrahieren müssen, hin zu unendlich vielen (überabzählbar unendlich vielen) Objekten. Erst in einem solchen Rahmen können wir mit dem Problem der endgültigen Fassung des Begriffes des „Elementarereignisses" fertig werden. Mit der Festsetzung, daß die Phänomene typisches Verhalten (das was am meisten geschieht) wiederspiegeln , wird das Gesetz der großen Zahlen zu einer theoretischen Vorhersage über empirische Häufigkeiten, die man überprüfen kann.

Nun gibt es sicher Folgen, in denen die relative Häufigkeit von Köpfen nur sehr langsam gegen 1/2 geht, z.B. könnte ja die Anzahl von Köpfen in n Würfen bei $n/2 + n^{2/3}$ liegen, was immer noch eine relative Häufigkeit von 1/2 liefert. Aber unsere kleine Rechnung oben zeigt, daß solche Folgen untypisch sind, nur solche mit der Anzahl $n_K = n/2 \pm \delta\sqrt{n}$ von Köpfen in n Würfen sind im Bereich des Typischen. Dabei ist die Wahrscheinlichkeit, wenn wir unser Ergebnis etwas umformulieren

$$W\left(\frac{1}{\sqrt{n}}\left(n_K - n/2\right) \approx \delta\right) \approx \exp -2\delta^2$$

einer Gaußschen Glockenkurve folgend. Wir werden auch dieses Verhalten in der späteren Abstraktion im allgemeinen Falle als „zentralen Grenzwertsatz" beweisen. Alle diese Folgen gehören zum „Einzugsbereich" des Gesetzes der großen Zahlen. Die Binomialkoeffizienten $\binom{n}{k}$ sind also für große n um $n/2$ zentriert, und fallen nach beiden Seiten auf einer Breite von $\sqrt{n}$ gemäß einer Gaußkurve ab.

Anmerkung 2.4.2. Zum Beweis der Stirlingschen Formel (**??**) .
Die Stirlingsche Formel enthält zwei Aussagen, eine über den asymptotischen Wert und eine über den Fehler:

$$n! = \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

Der asymptotische Wert ist einfach zu kriegen, das geht mit der Laplaceschen Methode des stationären Punktes. Sehr viel schwieriger ist es, den Fehler

$$n!\left(\frac{e}{n}\right)^n - \sqrt{2\pi n} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$$

in dieser Güte zu bekommen. Der folgende Beweis gibt uns im ersten Schritt mit einem wunderbaren analytischen Argument, das ganz speziell auf diese Aufgabe zugeschnitten ist, den Fehler und das asymptotische Verhalten bis aus den Wert $\sqrt{2\pi}$, den wir uns in einem zweiten Schritt verschaffen. Dieser zweite Schritt besteht aus der fundamentalen Laplaceschen Methode der stationären Punkte. Sie gibt die gesamte Asymptotik, aber leider nicht die Güte des Fehlers.

Schritt 1: Hier ist die Idee. Warum ist $\frac{n}{\mathrm{e}}$ der Mittelwert der n Faktoren in $n!$? Wenn $n! \approx (n/\mathrm{e})^n$, dann ist das gleichbedeutend mit $\ln(n!) \approx n(\ln(n) - 1) = n\ln(n) - n = \int_1^n \ln(x)\mathrm{d}x$, und das gibt schon die Idee, denn $\ln(n!) = \sum_{k=1}^n \ln(k)$, und da der ln monoton steigt ist nunmehr rigoros:

$$\ln(k) \le \int_k^{k+1} \ln(x)\mathrm{d}x \le \ln(k+1). \tag{2.11}$$

Summation bis $(n-1)$ liefert

$$\ln((n-1)!) \le n\ln(n) - n + 1 \le \ln(n!), \tag{2.12}$$

und hieraus folgt leicht, unter Beachtung von $\ln((n-1)!) + \ln(n) = \ln(n!)$

$$n\ln(n) - n + 1 \le \ln(n!) \le n\ln(n) - n + 1 + \ln(n) \tag{2.13}$$

und damit

$$\left(\frac{n}{\mathrm{e}}\right)^n \mathrm{e} \le n! \le \left(\frac{n}{\mathrm{e}}\right)^n \mathrm{e}n. \tag{2.14}$$

Das ist schon einmal nicht schlecht, aber natürlich nicht gut genug. Was wir verbessern müssen, ist die Abschätzung (**??**), denn da haben wir ja wirklich ohne Mühen abgeschätzt, und gute Dinge brauche etwas Mühe. Wir müssen also das Integral $\int_k^{k+1} f(x)\mathrm{d}x (= f(\xi_k))$ (nach Mittelwertsatz) besser abschätzen, und das machen wir, indem wir den Mittelwert durch das arithmetische Mittel ersetzen:$f(\xi_k)) \approx (f(x_k) + f(x_{k+1}))/2$, und das wiederum führt ohne großes Nachdenken darauf, das Intervall $[k, k+1]$ zu halbieren, und auf jeder Hälfte die Stammfunktion $F(x)$ mit der Taylorformel bis zur zweiten Ordnung zu entwickeln:

$$F(x) = F(y) + f(y)(x-y) + (1/2)f'(\xi)(x-y)^2.$$

Wir entwickeln im folgenden um die untere Integralgrenze

$$\begin{aligned}
\int_k^{k+1} f(x)\mathrm{d}x &= \int_k^{k+1/2} f(x)\mathrm{d}x - \int_{k+1}^{k+1/2} f(x)\mathrm{d}x \\
&= \frac{1}{2}f(k) + \frac{1}{2}f'(\xi_k)\frac{1}{4} + \frac{1}{2}f(k+1) - \frac{1}{2}f'(\eta_k)\frac{1}{4} \\
&= \frac{1}{2}(f(k) + f(k+1)) + \frac{1}{8}(f'(\xi_k) - f'(\eta_k)),
\end{aligned}$$

mit $\xi_k \in (k, k+1/2), \eta_k \in (k+1/2, k+1)$. Summation über k bis $n-1$ liefert

$$\int_1^n f(x)\mathrm{d}x = -\frac{1}{2}(f(1) + f(n)) + \sum_{k=1}^{n} f(k) + \sum_{k=1}^{n-1} \frac{1}{8}(f'(\xi_k) - f'(\eta_k)).$$

Nun ist $f = \ln$ und daher $f'(x) = \frac{1}{x}$ und darum ist $\sum_{k=1}^{n-1} \frac{1}{8}\left(\frac{1}{\xi_k} - \frac{1}{\eta_k}\right)$ eine alternierende Reihe, die nach dem Leibnizschen Konvergenzsatz gegen einen Grenzwert s konvergiert. Und weil $\int_1^n \ln(x)\mathrm{d}x = n\ln(n) - n + 1$ ist, kommt

$$s - n\ln(n) + n - 1 - \frac{1}{2}\ln(n) + \ln(n!) = \sum_{k \ge n} \frac{1}{8}\left(\frac{1}{\xi_k} - \frac{1}{\eta_k}\right),$$

wobei nun, mit ganz leichter Überlegung

$$0 \le \sum_{k \ge n} \frac{1}{8}\left(\frac{1}{\xi_k} - \frac{1}{\eta_k}\right) \le \frac{1}{8n}$$

ist. Also existiert ein $\epsilon_n \in (0, \frac{1}{8n})$, so daß

$$s - n\ln(n) + n - 1 - \frac{1}{2}\ln(n) + \ln(n!) = \epsilon_n$$

ist, und damit erhalten wir

$$n! = \mathrm{e}^{\epsilon_n}\mathrm{e}^{-s+1}\left(\frac{n}{\mathrm{e}}\right)^n\sqrt{n}, \tag{2.15}$$

wobei man beachte, daß $\mathrm{e}^{\epsilon_n} = 1 + \mathcal{O}\left(\frac{1}{n}\right)$ ist.

2.Schritt: Damit bleibt die Frage nach dem Wert s, bzw. e^{-s+1}. Das muß ja $\sqrt{2\pi}$ sein. Woher kommen diese 2π? Von den vielen möglichen Antworten ist die folgende ziemlich nah an der Sache. Es kommt daher, daß

$$\int_{-\infty}^{\infty}\mathrm{e}^{-x^2}\mathrm{d}x = \sqrt{\pi}$$

ist. Warum dieses Gaußintegral eine Rolle spielt? Weil $\binom{n}{k}$ für große n gaußsche Fluktuationen hat. Wie kommt man nun an den Wert? Wir benutzen

$$\begin{aligned} n! = \Gamma(n+1) &= \int_0^{\infty}\mathrm{e}^{-x}x^n\mathrm{d}x \\ &= n^{n+1}\int_0^{\infty}\left(\mathrm{e}^{-x}x\right)^n\mathrm{d}x = n^n n\int_0^{\infty}\mathrm{e}^{-n(x-\ln(x))}\mathrm{d}x, \end{aligned}$$

wobei die Substitution sich deswegen rechtfertigt, weil man daran denken muß, dass im Limes $n \to \infty$ Werte $x \approx n$ relevant sein werden. Die Frage ist also, wie sich der Integrand für große n verhält. Um dafür ein Gefühl zu kriegen, betrachte man die Substitution

$$\int_0^{\infty}\mathrm{e}^{-nf(x)}\mathrm{d}x = \int_{f(0)}^{f(\infty)}\frac{1}{f'(x)}\mathrm{e}^{-ny}\mathrm{d}y$$

und wäre $\frac{1}{f'(x)}$ beschränkt, dann wäre das Integral mindestens $\sim \frac{1}{n}$. Das ist leicht. Also muß man auf Nullstellen von f' achten! Solche Nullstellen heißen kritische oder stationäre Punkte von f und sie sind für das Verhalten des Integrals von größter Bedeutung. Ganz schnell: Hat f einen kritischen Punkt ξ, dann entwickle man den Exponenten darum bis zur zweiten Ordnung, die erste liefert ja null, und da sieht man etwas quadratisches entstehen, das gibt, wenn alles günstig ist, ein Gaußsches Integral. Darum formuliere ich nun, was günstig ist.

Das Ganze heißt Laplace's Methode der „stationären Punkte“. Es gilt folgendes:

$$\int_a^b \phi(x)\mathrm{e}^{nh(x)}\mathrm{d}x \propto \phi(\xi)\mathrm{e}^{nh(\xi)}\sqrt{-\frac{2\pi}{nh''(\xi)}} \tag{2.16}$$

($a(n) \propto \frac{1}{n}$ heißt hier $\frac{a(n)}{\frac{1}{n}} \to 1$ für $n \to \infty$).

Günstige Bedingungen an h und ϕ sind:

1. h ist eine zweimal stetig differenzierbare Funktion die ihr Maximum bei ξ annimmt, d.h $h''(\xi) < 0$. Weiterhin sei außerhalb einer Umgebung von ξ das Supremum der Funktion kleiner als $h(\xi)$
2. ϕe^{ah} ist absolut integrierbar für $a > 0, \phi(\xi) \neq 0$ und ϕ stetig in ξ.

Wir wenden (**??**) an:

$$\phi = 1,\ \mathrm{e}^{-x}x = \mathrm{e}^{-x+\ln x}, \quad \text{das heißt} \quad h(x) = -x + \ln x.$$

$$h'(x) = -1 + \frac{1}{x} = 0 \ \Rightarrow\ \xi = 1,\ h''(\xi) = -1,\ h(1) = -1,$$

$$\text{das heißt} \quad n! \propto n^{n+1}\mathrm{e}^{-n}\sqrt{\frac{2\pi}{n}} = \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}.$$

Bemerkung: Wenn wir den Beweis uns näher anschauen würden, würden wir auch eine Kontrolle über den Fehler bekommen, aber wir würden die Qualität des obigen Fehlerterms $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$, nicht ohne weiteres erreichen.

Wir zeigen jetzt (**??**). Dazu zunächst:

1. Sei $k > 0$ und $a < \xi < b$. Dann gilt
$$\int_a^b \mathrm{e}^{-kn(x-\xi)^2}\mathrm{d}x \propto \sqrt{\frac{\pi}{kn}}$$
(substituiere $y = \sqrt{kn}(x-\xi)$ und verwende $\int \mathrm{e}^{-y^2}\mathrm{d}y = \sqrt{\pi}$).
2. Sei $-h''(\xi) > \epsilon > 0$. Wähle eine δ–Umgebung von ξ so klein, daß $U_\delta(\xi) \subset (a;b)$ ist, $h''(\xi) - \epsilon \leq h''(x) \leq h''(\xi) + \epsilon$ und $|\phi(x) - \phi(\xi)| < \epsilon$ für alle $x \in U_\delta(\xi)$.

Zu δ gibt es ein $C_\delta > 0$ mit $h(x) - h(\xi) < -C_\delta$ für $x \notin U_\delta(\xi)$.

Dann gilt, zunächst auf dem Komplement von $U_\delta(\xi)$

$$\begin{aligned}\int_{U_\delta^c(\xi)} \phi(x)\mathrm{e}^{n(h(x)-h(\xi))}\mathrm{d}x &= \int_{U_\delta^c(\xi)} \phi(x)\mathrm{e}^{\frac{n}{2}(h(x)-h(\xi))}\mathrm{e}^{\frac{n}{2}(h(x)-h(\xi))}\mathrm{d}x \\ &\leq \mathrm{e}^{-\frac{n}{2}C_\delta}\underbrace{\int_{U_\delta^c(\xi)} |\phi(x)|\mathrm{e}^{\frac{1}{2}(h(x)-h(\xi))}\mathrm{d}x}_{=:C} \\ &\leq C\mathrm{e}^{-\frac{n}{2}C_\delta}.\end{aligned}$$

Es bleibt

$$\int_{U_\delta(\xi)} \phi(x)\mathrm{e}^{n(h(x)-h(\xi))}\mathrm{d}x = \phi(\tilde{\xi_\delta})\int_{\xi-\delta}^{\xi+\delta} \mathrm{e}^{\frac{1}{2}n(x-\xi)^2h''(\xi_\delta)}\mathrm{d}x,$$

mit Zwischenstellen $\tilde{\xi_\delta}$ und ξ_δ, wobei wir

$$h(x) = h(\xi) + \underbrace{h'(\xi)(x-\xi)}_{=0} + \frac{1}{2}h''(\xi_\delta)(x-\xi)^2$$

entwickelt haben. Aber die rechte Seite liegt im Intervall

$$\left[(\phi(\xi)-\epsilon)\int_{\xi-\delta}^{\xi+\delta} \mathrm{e}^{\frac{n}{2}(h''(\xi)-\epsilon)(x-\xi)^2}\mathrm{d}x; (\phi(\xi)+\epsilon)\int_{\xi-\delta}^{\xi+\delta} \mathrm{e}^{\frac{n}{2}(h''(\xi)+\epsilon)(x+\xi)^2}\mathrm{d}x\right]$$

und diese Grenzen sind nach 1.

$$\propto (\phi(\xi) \pm \epsilon)\sqrt{-\frac{2\pi}{n(h''(\xi) \pm \epsilon)}}.$$

Da ϵ beliebig wählbar ist, haben wir (**??**).

2.5 Bedingte Wahrscheinlichkeit

Der Begriff der bedingten Wahrscheinlichkeit ist einfach und schwierig zugleich. Mit der bedingten Wahrscheinlichkeit wird das „Typische“ neu definiert, und ehemals Untypisches kann unter der bedingten Wahrscheinlichkeit typisch werden. Ich werde das später noch einmal aufgreifen. Nun zur sehr einfachen intuitiven Setzung. Wenn ein Ereignis A eingetreten ist, dann reduzieren sich gegebenenfalls die Möglichkeiten für das Eintreten des Ereignisses B. Das führt geradlinig auf die folgende Setzung für die „Wahrscheinlichkeit von B gegeben A“

$$W(B/A) = \frac{W(A \cap B)}{W(A)}.$$

Beispiel 2.5.1. Wenn bei zweimaligem Würfeln die Augensumme 7 gekommen ist, wie groß ist die Wahrscheinlichkeit, daß im ersten Wurf eine 4 kam? $A = (x_1, x_2); x_1 + x_2 = 7;\ |A| = 6,\ B = (4,3) \subset A$, also $W(B/A) = \frac{W(B)}{W(A)} = \frac{1/36}{6/36} = \frac{1}{6}$. Was ergäbe sich bei Augensumme 5?

Aus der Setzung der bedingten Wahrscheinlichkeit fließt sofort die Additivität für disjunkte Ereignisse B_k:

$$W\left(\bigcup B_j/A\right) = \sum_j W(B_j/A), \quad B_j \cap B_k = \emptyset\ k \neq j.$$

Technisch von großer Bedeutung ist aber folgende Zerlegungsformel: Sei $\bigcup A_k = \Omega$ eine disjunkte Zerlegung von Ω, dann gilt

$$W(B) = \sum_k W(B/A_k)W(A_k), \tag{2.17}$$

denn

$$\begin{aligned}\sum_k W(B/A_k)W(A_k) &= \sum_k \frac{W(B\cap A_k)}{W(A_k)}W(A_k)\\ &= \sum_k W(B\cap A_k) = W(B\cap \Omega)\\ &= W(B)\end{aligned}$$

Wir werden dazu noch Beispiele haben.

Wie sich die Wahrscheinlichkeiten durch Eintreten von Ereignissen ändern ist manchmal überraschend:

Beispiel 2.5.2. Betrachte ein Spiel mit 4 Karten, davon zwei unterschiedliche schwarze Karten, eine (s) mit Oberseite schwarz und eine (s_K) mit schwarzem Kreis, sowie zwei Karten mit weißer Oberseite w_1 und w_2; dazu 2 Spieler.

Der erste Spieler nimmt zwei Karten auf, der zweite Spieler kann verschieden fragen:

1. Möglichkeit: Hast du eine schwarze Karte?
   Antwort: Ja!
   Uns interessiert:

   $$W_1 = W(\text{beide Karten von 1 schwarz/1 hat eine schwarze Karte})$$

   Aber

   $$W_1 = \frac{W(\text{beide schwarz und eine schwarz})}{W(\text{eine schwarz})} = \frac{W(\text{beide schwarz})}{W(\text{eine schwarz})}.$$

   Es gibt $\binom{4}{2} = 6$ zweier Gruppen, davon eine mit (s,s) und eine mit (w,w). Also ist

   $$W_1 = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5}.$$

2. Möglichkeit der Frage: Hast du die Karte mit schwarzem Kreis?
   Antwort: Ja!
   Nun ist

   $$W_2 = W(\text{beide Karten schwarz/Karte schwarzem Kreis}) = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3},$$

   denn s_k kann in 3 Paaren vertreten sein, wohingegen im ersten Fall eine schwarze Karte in 5 Paaren vertreten ist. Also reduziert sich im zweiten Falle die Anzahl der Möglichkeiten, und $W_2 > W_1$: Wenn nach einer spezifischen schwarzen Karte gefragt wird, erhöht sich die Chance bei Antwort „ja“, daß der Spieler zwei schwarze Karten hat.

Beispiel 2.5.3. Ein weiteres Glücksspiel:

Drei Karten, davon eine beidseitig rot, eine beidseitig schwarz und eine mit einer roten und einer schwarzen Seite. Man zieht blind eine Karte und legt sie nieder. Wenn die sichtbare Seite rot ist, wie groß ist dann die Wahrscheinlichkeit dafür, daß die andere Seite schwarz ist? Die Antwort ist $\frac{1}{3}$, denn da eine rote Seite vorliegt, ist RR oder RS gezogen worden. Nun gibt es drei Möglichkeiten, R oben liegen zu haben, davon ist eine mit S unten. (Offenbar ist es also ganz egal, wieviele beidseitig schwarze Karten es gibt.)

Man kann allerdings auch darauf verfallen, daß die Kenntnis der roten Seite die schwarze Karte ausschließt, und es deswegen nur mehr zwei Möglichkeiten gibt, nämlich RR oder RS, also Wahrscheinlichkeit $\frac{1}{2}$ für RS, was aber falsch ist, denn die (relevanten) Elementarereignisse sind hier die Seiten der Karten. Darum könnte man hiermit ein „unfaires" Glücksspiel treiben, wenn das Gegenüber sich vom falschen Argument überzeugen ließe.

Wir werden im Beispiel **??** unten die technische Bedeutung der bedingten Wahrscheinlichkeit als Rechenhilfe beispielhaft vorführen. Hier will ich mit einer Bemerkung zur empirischen Bedeutung der bedingten Wahrscheinlichkeit abschließen. Wie in den Beispielen deutlich gemacht, treten Fakten ein, und dieser Eintritt verändert die Wahrscheinlichkeit anderer möglicher Fakten. Darin liegt ein starker Subjektivismus: *Meine Einschätzung* über das Eintreten von Ereignissen wird durch das Eingetreten sein eines anderen Ereignisses berührt. Dieser Subjektivismus ist sogar so stark, dass er an den Grundfesten einer objektiven Definition des „Typischen" zweifeln läßt. Es ist, als ob wir das „Typische" gar nicht so ernst nehmen, denn wenn die Fakten gegen das typische Verhalten sprechen, dann sind wir bereit — gegeben diese neuen Fakten — „typisch" neu zu definieren. Aber das bringt letztlich die Sache dann doch wieder in einen objektiven Rahmen. Wir verhalten uns nämlich so: Wenn in der Tat etwas „Untypisches" eingetreten ist, dann gehen wir davon aus, daß nunmehr bezogen auf die untypischen Elementarereignisse alles wieder typisch verläuft. Subjektiv hieran bleibt, dass wir glauben müssen, dass unsere Vorgehensweise der Natur der Dinge entspricht. Wir mussten natürlich auch schon vorher glauben, daß „typisches Verhalten" der Natur der Dinge entspricht. Das letztere kann man aber durchaus als physikalisches Gesetz empfinden, wohingegen das erstere sich als nicht mehr so ernst gemeint auszunehmen scheint. Ich will ein Beispiel geben: Ein Gas füllt typischerweise den Raum aus, indem es sich befindet. Die Herstellung einer Gasflasche ist in dem Sinne bereits ein untypischer Vorgang, denn da wird das Gas entgegen seiner typischen Natur in einen kleinen Behälter komprimiert. Dann aber, läßt man das Gas ausströmen, wird es sich — gegeben diesen untypischen komprimierten Zustand — typischerweise in dem ganzen zur Verfügung stehenden Raume ausbreiten. Das ist objektiv so, und wir tun gut daran, dass wir das so vohersagen. Man könnte auch jede Voraussage als unmöglich ablehnen, denn wenn schon ein untypischer Zustand des Gases vorliegt, warum nicht ein so untypischer, dass die zukünftige Entwicklung völlig irreal verläuft (z.B. weigert sich das Gas auszutreten)? Dabei bleibt eine Frage (auch in weiter Zukunft) unbeantwortet: Wie konnte es überhaupt zu diesem untypischen Vorgang der Komprimierung kommen? Ich will diese Frage hier nicht weiter ausführen, denn ihre Antwort führt unweigerlich zu einer Regression von weiteren Fragen, an deren Ende die Frage steht, warum unser Universum so ist, wie es ist. Das ist an dieser Stelle nicht diskutierbar. Aber ich will die Problematik des „Typischen“ im Untypischen mit Schrödinger's Gefangenen Geschichte noch einmal illustrieren, die ich mit meinen Worten erzähle:

Der Gefangene James ist dem Henker zum Tode ausgeliefert. Der gibt ihm eine Chance zu leben. James darf nämlich in einem Urnenspiel wählen zu raten. Rät er richtig, darf er als freier Mann das Gefängnis verlassen, rät er falsch, wird er enthauptet, zieht er es aber vor, nicht zu raten, bleibt er lebenslang in Haft. Das Spiel ist folgendes: Eine Urne enthält 100 Karten, 90 (blau,blau), 9 (blau,rot), 1 (rot,rot). James zieht blind eine Karte und legt sie nieder. Die Augenbinde wird ihm abgenommen und er sieht, dass rot oben liegt. Die Wahl ist nun entweder zu raten, welche Farbe die andere Seite hat, oder gar nicht zu raten. Und der gemeine Henker wirft noch ein, bevor James überhaupt etwas sagen kann: „Bedenke, mit dem gleichen Argument, mit dem du nun blau vermutest hättest du blau vermutet, hätte blau oben gelegen. Und dann wärest du enthauptet worden.“ Poor James. Er zog es vor, in lebenslanger Haft zu bleiben.

2.6 Unabhängigkeit

Als nächstes besprechen wir die häufig auftretende Situation, in der wir meinen, daß das Eintreten von A *unabhängig* vom Eintreten von B ist: Unabhängigkeit ist ein schwieriges Konzept und *keine* primitive Eigenschaft. Zunächst ist es nur die Produktregel:

A, B heißen **unabhängig**, genau dann wenn das Eintreten von A die Wahrscheinlichkeit des Eintreten von B nicht verändert: $W(B/A) = W(B)$, damit ist auch $W(A/B) = W(A)$, also

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

Die Unabhängigkeit ist in der Laplace-Wahrscheinlichkeit durch die Willkür, die dem Begriff des Elementarereignisses anhaftet oft enthalten:

Beispiel 2.6.1. Zweimal würfeln ergibt 36 Paare (x, y) von Augenzahlen, und

$$W(\{(i,j)\}) = \frac{1}{36} = \frac{1}{6}\frac{1}{6} = W(\{i\})W(\{j\}),$$

d.h. die Ereignisse A=„Augenzahl im ersten Wurf ist i“ und B=„Augenzahl im zweiten Wurf ist j “ sind unabhängig. Präziser:

$$A = \{(x,y); x = i\} \; |A| = 6, B = \{(x,y); y = j\} \; |B| = 6$$

und

$$A \cap B = \{(i,j)\},$$

deswegen

$$W(A \cap B) = \frac{1}{36} = W(A)W(B) = \frac{6}{36}\frac{6}{36}.$$

Diese Präzision ist keine Pedanterie. Das Wort „unabhängig“ suggeriert: A und B haben miteinander nichts zu tun, haben nichts gemeinsam und sie sollten deswegen *disjunkte* Ereignisse sein. *Falsch!* Als disjunkte Ereignisse wäre ja die Wahrscheinlichkeit des Durchschnittes null. Unabhängigkeit ist das ganze Gegenteil davon. A und B haben enorm viel miteinander zu tun. Es ist auf das Genaueste ausgetüftelt, wieviel sie miteinander zu tun haben müssen, damit Unabhängigkeit herrscht: Wir werfen eine Münze 10 mal und haben das Gefühl, daß die Ergebnisse genau so viel miteinander zu tun haben, daß vorherige Ergebnisse (sagen wir die ersten 5) die Wahrscheinlichkeit der nachfolgenden nicht beeinflußt. Wie aber kann die Natur soetwas Kompliziertes einrichten?

Wenn man also meint, daß beim Würfeln die Ereignisse

A : im Wurf kommt eine gerade Augenzahl und

B : im Wurf kommt eine ungerade Augenzahl

unabhängig sind, dann ist das schlichtweg falsch:

$$W(A \cap B) = 0 \neq W(A)W(B) = \frac{1}{4},$$

denn die Kenntnis des einen Ereignisses schließt das andere aus.

Beispiel 2.6.2. : Unabhängigkeit ist unwahrscheinlich. Verteile positive Zahlen — „Wahrscheinlichkeiten“ — auf zwei Ereignisse A und B, $A \cup B = \Omega$, so daß sie unabhängig sind. Aus den überabzählbar vielen Möglichkeit gibt es nur zwei Möglichkeiten (welche?), die Unabhängigleit realisieren. Würde man alle Möglichkeiten als Kugeln in einer Urne unterbringen, wäre die Wahrscheinlichkeit also null, Unabhängigkeit zu „ziehen“.

Beispiel 2.6.3. Konstruiere zwei unabhängige Würfelereignisse:A: Augenzahl 2 oder 3, d.h.

$$A = \{2,3\}$$

und B: Augenzahl ungerade, d.h.

$$B = \{1,3,5\}.$$

$$W(A \cap B) = W(3) = \frac{1}{6} = W(A)W(B) = \frac{2}{6} \cdot \frac{3}{6}.$$

Man macht sich dies auch mit den **reduzierten** Ereignismöglichkeiten, das heißt mit den bedingten Wahrscheinlichkeiten, klar:

$$W(A/B) = W(2 \text{ oder } 3 \text{ falls Augenzahl ungerade ist}),$$

also 2 kann nicht sein; für ungerade Augenzahl gibt es nur noch 3 Möglichkeiten, also hat 3 die Wahrscheinlichkeit $\frac{1}{3}$.

$$W(A/B) = \frac{W(A \cap B)}{W(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3} = W(A).$$

Bemerke auch:

$$W(B/A) = W(\text{Augenzahl ungerade, falls Augenzahl 2 oder 3 ist}) = \frac{1}{2} = W(B),$$

da die Kenntnis von Augenzahl 2 oder 3 nicht neues bringt.

Ganz anders ist die Sache bei

$$A': \quad \text{Augenzahl 2 oder 3 oder 4,}$$

$$B: \quad \text{Augenzahl ungerade.}$$

$$W(A' \text{ und } B) = \frac{1}{6} \neq W(A')W(B) = \frac{3}{6} \cdot \frac{3}{6} = \frac{1}{4}$$

und $W(A'/B) = \frac{1}{3}$ wie oben, aber $W(B/A') = \frac{1}{3}$, da die Kenntnis von Augenzahl 2 oder 3 oder 4 die ungerade Zahl 3 benachteiligt. Wenn wir nun

$$A'': \quad \text{Augenzahl 2 oder 3 oder 4 oder 5}$$

setzen, bekommen wir wieder Unabhängigkeit.

Wenn A und B sowie B und C unabhängig sind, dann folgt nicht, daß A und C unabhängig sind:

$$A = \text{Augenzahl ungerade,}$$

$$B = \text{Augenzahl 2 oder 3,}$$

$$C = \text{Augenzahl gerade.}$$

Unabhängigkeit von Ereignissen ist eine Familieneigenschaft: A,B,C heißen unabhängig, wenn

$$W(A \text{ und } B \text{ und } C) = W(A)W(B)W(C)$$

ist, und alle Paarungen A und B , C ; A , B; usw. unabhängig sind. Wenn A,B,C unabhängig sind, dann ist z.B. A unabhängig von allen Ereignissen, die sich aus B, C ergeben.

Übrigens gilt auch
A, B unabhängig $\Rightarrow$ $A, \neg B$ unabhängig, mit $\neg B =$ nicht B, denn

$$\begin{aligned} W(A) &= W(A \text{ und } (B \text{ oder } \neg B)) \\ &= W(A \text{ und } B) + W(A \text{ und } \neg B) \\ &= W(A)W(B) + W(A \text{ und } \neg B), \end{aligned}$$

also

$$W(A \text{ und } \neg B) = W(A) - W(A)W(B) = W(A)(1 - W(B)) = W(A)W(\neg B).$$

Seien nun A, B und B, C und A, C jeweils unabhängig. Dann folgt auch nicht die Unabhängigkeit von A, B, C. Dazu

Beispiel 2.6.4. Werfen zweier Würfel

$$A = \text{ Augensumme 7,}$$

$$B = \text{ 1. Würfel zeigt 4,}$$

$$C = \text{ 2. Würfel zeigt 3.}$$

$$W(A \text{ und } B) = W((4,3)) = \frac{1}{36} = W(A)W(B) = \frac{1}{6} \cdot \frac{1}{6}.$$

A, C genauso.

$$W(B \text{ und } C) = W((4,3)) = \frac{1}{36} = W(B)W(C) = \frac{1}{6} \cdot \frac{1}{6},$$

also sind A, B und B, C und A, C jeweils unabhängig, aber

$$W(A \text{ und } B \text{ und } C) = W((4,3)) = \frac{1}{36} \neq W(A)W(B)W(C) = \frac{1}{6^3}.$$

Solange man sich auf einer phänomenologischen Ebene befindet und „Modelle“ macht, läßt sich Unabhängigkeit leicht einstellen und noch leichter sagen. Aber Unabhängigkeit ist in Wahrheit ein schweres Konzept, und wenn man den Dingen auf den Grund geht, d.h. wenn man zu den wirklichen Elementarereignissen vorgedrungen ist, wird man mit großer Ehrfurcht vor der Tatsache stehen, daß es in der Tat unabhängige Ereignisse gibt — die existent sind wie reelle Zahlen. Diese Erkenntnis wird uns den Weg zum Verständnis von Wahrscheinlichkeit weisen.

2.7 Ende der Phänonmenologie

Man sagt leicht: Beim Münzwurf sind die Ergebnisse verschiedener Würfe unabhängig. Das bedeutet, dass man die Wahrscheinlichkeiten multiplizieren kann, aber die Wahrscheinlichkeiten wovon? Auf welchen Elementarereignissen sind die Wahrscheinlichkeiten definiert? Beim 10-fachen Münzwurf sind die Elementarereignisse 10−er Tupel mit K-Z-Einträgen und $A = \{(K, x_2, ..., x_10)\}, B = \{(x_1, K, x_3, ..., x_n)\}$ sind die Ereignisse, dass A : im ersten Wurf Kopf kommt und B : im zweiten Wurf Kopf kommt. $W(A \cap B) = W(A)W(B)$ ist in der Laplace-Wahrscheinlichkeit dann einfach eine Zahlengleichheit. Aber so geht man ja nicht vor. Die Intuition ist anders. Man weiß a priori, daß W(erster Wurf Kopf) $= 1/2$ und W(zweiter Wurf Kopf) $= 1/2$ unabhängig sind und multipliziert einfach die Wahrscheinlichkeiten. Denn wer weiß schon wie lange man wirklich würfeln wird. Muß man, wenn man über die Unabhängigkeit reden will, von vornherein die Anzahl der Würfe kennen? Ist es nicht so, dass egal wie lange ich würfeln werde, der erste und zweite Ausgang unabhängig sind? Man muss natürlich mindestens zweier Tupel betrachten, sonst kann man die Unabhängigkeit nicht mal formulieren. Aber das ist alles unerträglich undurchsichtig.

Anmerkung 2.7.1. Mehr noch: Das Konzept der Unabhängigkeit ist undurchsichtig, denn um von Unabhängigkeit reden zu können, müssen die Ereignisse von ein und derselben Sache handeln: Wenn ich eine Münze werfe und danach einen Würfel, dann — so würde man fühlen — sind die Ausgänge unabhängig. Aber *was ist ihnen gemeinsam*, dass sie soviel miteinander zu tun haben, wie es für Unabhängigkeit unbedingt notwendig ist, was ist das beiden gemeinsame Elementarereignis? (Die gemeinsame Darstellung auf einem Produktraum von Kopf,Zahl und Augenzahlen ist lächerlich!)

Weiter haben wir uns bei den bisherigen Beispielen immer auf endliche Ω beschränkt, sonst hätten wir die Laplace Wahrscheinlichkeit gar nicht setzen können. Die Additivität würde entweder null oder unendlich geben. Aber einige einfache Beispiele ergeben zwanglos einen unendlichen Ereignisraum, z.B. der Münzwurf bis zum ersten Kopf. Eine mögliche Wahl von „Elementarereignissen“ wären die Folgen bis zum ersten Kopf, z.B. $\omega_4 = (Z, Z, Z, K)$ und diese Folgen können im Prinzip unendlich lang sein. Wir können also keine Laplace-Wahrscheinlichkeit ansetzen, und intuitiv macht man einfach folgendes: Jeder Ausgang eines Wurfes bekommt Wahrscheinlichkeit 1/2 und dann multipliziert man diese Zahlen gemäß der Länge der Folge. Also $W(\{\omega_n\}) = 1/2^n$.

Um den Dingen auf den Grund zu gehen müssen wir den Münzwurf ernst nehmen und ihn von der phänomenologischen Ebene auf die Ebene befördern, auf der er vollständig beschreibbar ist. Das ist die Ebene des physikalischen Münzwurfes, in der z.B. die Münze als Festkörper durch seine Lage-Koordinaten im Raum und seine Geschwindigkeit sowie Drehgeschwindigkeit beschrieben wird. Das sind offenbar die Elementarereignisse im Rahmen einer Newtonschen Physik. Alle diese Größen sind aber Punkte eines Kontinuums, d.h. der Raum Ω ist in jedem Falle überabzählbar. Wie kann man da von typischen Punkten reden? Denn soviel haben wir ja schon verstanden: Wahrscheinlichkeit muss uns sagen, was typisch ist. Nun geht das nicht mehr, indem alle Elementarereignisse, d.h. alle Punkte in einem Kontinuum das gleiche Gewicht bekommen, das wäre notwendigerweise null (siehe auch Anmerkung **??** unten). Das heißt, wir müssen hier zu tieferem Verständnis kommen, bevor wir etwas über Wahrscheinlichkeit sagen können. Das folgende Beispiel liegt bereits auf der Grenze zwischen Jedermanns Können und Analyis und weist spielerisch in die richtige Richtung.

Beispiel 2.7.1. Die experimentelle Bestimmung von π durch Buffons Nadelexperiment.

Man hat ein liniertes Blatt mit N parallelen Linien im Abstand 1. Man lasse aus großer Entfernung auf das linierte Blatt eine Nadel der Länge 1 fallen. Die Nadel wird in ihrer Ruhelage möglicherweise eine Linie schneiden (siehe Abb. **??**).

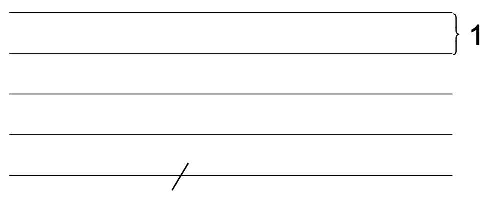

Abb. 2.3. Buffons Nadelexperiment.

Wie groß ist die Wahrscheinlichkeit dafür, daß die Nadel eine Linie schneidet? Auch hier steigen wir gleich auf einer groben Ebene ein, auf der wir nur auf die wesentlichen Merkmale fokussieren, die für eine schnelle pragmatische Bearbeitung vernünftig erscheinen. Dazu müssen wir zunächst uns überlegen, daß wir uns auf das Intervall zwischen zwei gegebenen Linien beschränken können, sagen wir Linien 0 und 1, denn

$$W(\text{Nadel schneidet eine Linie})$$

$$= \sum_{i=0}^{n} W(\text{Nadel schneidet und Nadelmitte liegt zwischen Linien}\, i, i+1)$$

$$= \sum_{i=0}^{N} W\,(\text{Nadel schneidet/ Mitte} \in [i, i+1])\, W(\text{Mitte} \in [i, i+1].$$

Es ist klar, daß alle Summanden gleich sind und

$$W(\text{Mitte zwischen}\, i, i+1) = 1/N$$

ist. Also haben wir nur

$$W(\text{Nadel schneidet/ Nadelmitte liegt zwischen Linien}\, 0,1)$$

zu behandeln. Das setzen wir nun intuitiv im folgenden Bild (Abb. **??**) um:

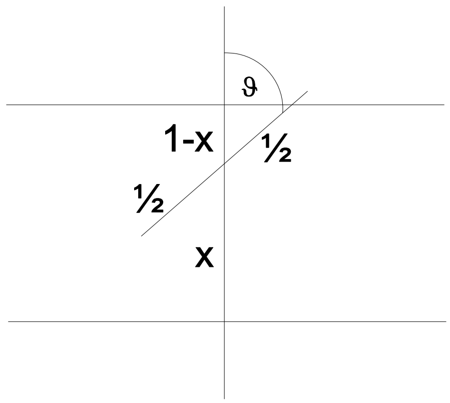

Abb. 2.4. Der Mittelpunkt und der Winkel sind zufällig in $[0,1]$ und $[-\pi/2, \pi/2]$ verteilt.

Also setzen wir folgendes an:

1. X ist gleichmäßig verteilt auf $[0;1]$, d.h. die Wahrscheinlichkeit, dass X im Bereich Δx liegt, hat Wahrscheinlichkeit Δx
2. Θ ist gleichmäßig verteilt auf $[-\frac{\pi}{2}; \frac{\pi}{2}]$, d.h. mit Wahrscheinlichkeit $\frac{1}{\pi}\mathrm{d}\vartheta$ liegt Θ im Bereich $\mathrm{d}\vartheta$

   Anmerkung 2.7.2. Man ziehe blind aus dem Einheitsintervall eine Zahl, Wie groß ist die Wahrscheinlichkeit, dass die gezogene Zahl einen vorgegebenen Wert hat (z.B. 0,3245)? Diese Wahrscheinlichkeit ist intuitiv null, denn es gibt unendliche viele andere Zahlen, von denen jede mit gleicher Berechtigung hätte gezogen werden können. Aber fragen wir nach der Wahrscheinlichkeit, dass die gezogene Zahl im Intervall $[0,4, 0,5]$ liegt, dann wird man nach kurzer Überlegung auf die Intervalllänge, nämlich $0,1$ kommen. Diese Wahrscheinlichkeit, gegeben durch die Intervalllänge, nennt man auch rein zufällige oder gleichmäßige Verteilung. X, Θ sind also rein zufällig verteilt.

3. X, Θ sind unabhängig.

S sei das Ereignis, daß die Nadel eine Gerade schneidet. Dann findet man ohne großes Bemühen

$$W(S) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} W(S \text{ und } \Theta = \vartheta) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{W(S|\Theta = \vartheta)}{\pi} \mathrm{d}\vartheta, \text{ denn zur Erinnerung}$$

$$W(S|\Theta = \vartheta) W(\Theta = \vartheta) = W(S \text{ und } \Theta = \vartheta).$$

Nach Anmerkung **??** hat $\Theta = \vartheta$ Wahrscheinlichkeit null und man kann sich mathematische Sorgen machen, was die Bedingung bedeutet: Wir denken an einen Grenzprozess, indem wir erst unter $\Delta\vartheta$ bedingen und dann $\Delta\vartheta \to 0$ betrachten, was unproblematisch ist. Wir drücken nun S als Menge von Werten von X und Θ aus: S bei gegebenem ϑ verlangt, daß $1 - x \leq \frac{1}{2} \cos \vartheta$ oder $\frac{1}{2} \cos \vartheta \geq x$ ist. Das definiert uns die Menge $S(\vartheta)$ (Abb. **??**). Also

Abb. 2.5. Die Menge $S(\vartheta)$.

Also kommt

$$W(S|\Theta = \vartheta) = W(X \in S(\vartheta)) = 2\frac{1}{2} \cos \vartheta = \cos \vartheta$$

und schließlich

$$W(S) = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos \vartheta \mathrm{d}\vartheta = \frac{2}{\pi}.$$

Damit kann π experimentell bestimmt werden: Mit dem Gesetz der großen Zahlen. Man mißt die relativen Häufigkeiten von Überschneidungen! Der Astronom Wolf in Zürich machte dies um 1850 und fand in 5000 Würfen den Wert 3,1596.