
Bernoullis Gesetz der Großen Zahlen

Erwin Bolthausen

1 Einführung

„Obwohl aber seltsamerweise durch einen sonderbaren Naturinstinkt auch jeder Dummste ohne irgend eine vorherige Unterweisung weiß, dass je mehr Beobachtungen gemacht werden, umso weniger die Gefahr besteht, dass man das Ziel verfehlt, ist es doch ganz und gar nicht Sache einer Laienuntersuchung, dieses genau und geometrisch zu beweisen.“ Aus einem Brief Jakob Bernoullis an Gottfried Wilhelm Leibniz, 3. Oktober 1703 [3].

Die Briefstelle bezieht sich auf das Gesetz der großen Zahlen. Bernoulli bringt zum Ausdruck, dass eine gewisse Vorstellung davon gedankliches Allgemeingut und keineswegs neu ist. Sehr stolz ist er jedoch darauf, dieses Gesetz erstmals mathematisch präzise mit genauen Schranken bewiesen zu haben, sowie noch mehr auf die Erkenntnis, dass es die Basis für eine – wie man heute sagen würde – Schätztheorie abgibt, mit der die Wahrscheinlichkeitstheorie weit über den Bereich von Spielen hinaus angewendet werden kann. Das Gesetz wird in Teil IV der *Ars conjectandi* hergeleitet, welche 1713, acht Jahre nach Bernoullis Tod, von seinem Neffen Nikolas Bernoulli publiziert wurde. Der Text ist in Ostwalds Klassikern übersetzt herausgegeben worden [1].

Die Schweizer Mathematiker halten das Gesetz offenbar für den bedeutendsten Beitrag eines Schweizer zur Mathematik. Jedenfalls wurde ihm die Ehre zuteil, anlässlich des Weltkongresses der Mathematiker 1994 in Zürich auf einer Briefmarke der Schweizerischen Post dargestellt zu werden.



Die Graphik für die arithmetischen Mittel über eine längere Zeit wurde jedoch offensichtlich nicht durch Zufall erzeugt, denn der Verlauf wechselt gegenüber dem Erwartungswert praktisch permanent das Vorzeichen, was – wie sich Bernoulli ausdrücken würde – „moralisch sicher“ nicht auftreten würde. Die Graphik spiegelt eine weitverbreitete Vorstellung wider, nämlich dass in einem Zufallsexperiment, wie dem Werfen einer Münze oder beim Roulette, eine quasi zielgerichtete Tendenz zum Ausgleich vorhanden ist. Wenn also eine Münze bis zu einem Zeitpunkt in mehr als der Hälfte der Fälle auf „Kopf“ gefallen ist, so habe sie die Tendenz, das in Zukunft auszugleichen. Tatsächlich sind mathematische Laien nicht imstande, eine Folge für das Werfen einer Münze aus dem Kopf zu produzieren, die halbwegs zufällig ist, einfach weil eine durch menschliche Vorstellung erzeugte „zufällige“ Folge eine stark zielgerichtete Tendenz zum Ausgleich besitzt, an der man die „Nichtzufälligkeit“ wie im oben dargestellten Verlauf sofort erkennt. Eine bemerkenswerte Tatsache ist, dass Bernoullis Gesetz der Zahlen gilt, ohne dass eine derartige Zielgerichtetheit angenommen werden muss.

Hier ist das Gesetz in moderner Formulierung: Es sei X_1, X_2, \dots eine Folge von unkorrelierten Zufallsvariablen, die alle dieselbe Verteilung haben und deren Erwartungswert $\mu = EX_i$ existiert. Sei ferner $S_N = \sum_{i=1}^N X_i$. Dann gilt für jedes $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P \left(\left| \frac{S_N}{N} - \mu \right| \geq \varepsilon \right) = 0.$$

$P(\cdot)$ steht für die Wahrscheinlichkeit eines Ereignisses. Bernoulli hat das Gesetz für das Bernoulli-Experiment bewiesen, d.h. für den Fall, wo die X_i unabhängige $\{0, 1\}$ -wertige Zufallsvariablen mit $p = P(X_i = 1)$ sind. In diesem Fall ist $\mu = p$. Auf seinen Beweis werde ich unten eingehen. In der obigen allgemeinen Formulierung stammt es von Khinchin [2].

In Anfängervorlesungen wird es in der Regel unter der Annahme der Existenz der Varianz (d.h. der L_2 -Integrierbarkeit der Zufallsvariablen) bewiesen. Dann folgt nämlich nach der Tschebyscheff-Ungleichung

$$\begin{aligned} P \left(\left| \frac{S_N}{N} - \mu \right| \geq \varepsilon \right) &\leq \frac{1}{\varepsilon^2} \operatorname{var} \left(\frac{S_N}{N} \right) \\ &= \frac{1}{\varepsilon^2} \frac{1}{N^2} \operatorname{var} (S_N) \\ &= \frac{1}{\varepsilon^2} \frac{1}{N^2} N \operatorname{var} (X_1) = \frac{\operatorname{var} (X_1)}{\varepsilon^2 N}. \end{aligned} \tag{1.1}$$

Hier wurde verwendet, dass für unkorrelierte Zufallsvariablen die Varianz einer Summe gleich der Summe der Varianzen ist, und da angenommen wird, dass die Verteilungen der Zufallsvariablen alle gleich sind, folgt $\operatorname{var}(X_i) = \operatorname{var}(X_1)$.

Bernoullis Gesetz nennt man heute auch das „schwache Gesetz der großen Zahlen“. Das starke Gesetz besagt (unter stärkeren Annahmen):

$$P \left(\lim_{N \rightarrow \infty} \frac{S_N}{N} = \mu \right) = 1.$$

Diese Formulierung setzt voraus, dass eine unendliche Folge der Zufallsvariablen, etwa für einen Münzwurf, auf einem festen Wahrscheinlichkeitsraum definiert ist. Die Existenz einer derartigen Folge ist erst zu Beginn des 20. Jh. mathematisch präzise bewiesen worden. Im Gegensatz dazu benötigt das Bernoulli-Gesetz bloß die Wahrscheinlichkeiten für endliche Folgen. Für das Bernoulli-Experiment der Länge N mit $P(X_i = 1) = p$ sind diese durch die sogenannte Binomialverteilung

$$P(S_N = k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad 0 \leq k \leq N,$$

gegeben.

(1.1) ist für den Bernoulli-Fall übrigens eine sehr schwache Abschätzung. Eine wesentlich bessere kann für unabhängige Zufallsvariablen X_i , die ein exponentielles Moment besitzen, mit einer exponentiellen Markov-Ungleichung gewonnen werden. Im binomialverteilten Fall ergibt sich für $1 \geq t > p$

$$P\left(\frac{S_N}{N} \geq t\right) \leq \exp[-NI_p(t)] \quad (1.2)$$

mit der sogenannten „Ratenfunktion“

$$I_p(t) = t \log \frac{t}{p} + (1-t) \log \frac{1-t}{1-p}.$$

Die gleiche Ungleichung gilt für $0 \leq t < p$ und $P(S_N/N \leq t)$. Solche Sätze, die einen exponentiellen Abfall von Wahrscheinlichkeiten beschreiben, nennt man heute „Prinzipien großer Abweichungen“. Diese wurden in den vergangenen 30 Jahren intensiv untersucht, natürlich in sehr viel allgemeineren Situationen. Die Abschätzung (1.2) ist, was den exponentiellen Abfall betrifft, scharf. Sie kann jedoch mit einem Vorfaktor der Größenordnung $N^{-1/2}$ noch verbessert werden. Bemerkenswerterweise hat Bernoulli bereits ein derartiges Prinzip hergeleitet, mit einer Ratenfunktion, die etwa um einen Faktor 2 zu klein ist, aber die wesentlichen Aspekte korrekt beschreibt.

2 Bernoullis Beweis

Bernoulli führt den Beweis für eine etwas eingeschränkte Situation. Er betrachtet eine Urne mit $t = r+s$ Kugeln, von denen r rot und s schwarz sind. Die Erfolgswahrscheinlichkeit für das Ziehen einer roten Kugel ist also $p = r/t$. Für ε nimmt er $1/t$, und er untersucht $N = nt$ Ziehungen (mit Zurücklegen) mit $n \rightarrow \infty$. So liefert etwa $r = 3$, $s = 2$ ein Bernoulli-Experiment mit $p = 0.6$ und $\varepsilon = 0.2$. Für dasselbe Bernoulli-Experiment kann er auch Vielfache von 3 bzw. 2 nehmen, was entsprechend kleinere Wert für ε ergibt. Seine Einschränkung führt zu etwas einfacheren Rechnungen; ich vermute aber, dass Bernoulli klar war, dass sie für den Kern seines Beweises belanglos ist.

Der Einfachheit halber beschränke ich mich auf die Darstellung seines Beweises im symmetrischen Fall $p = 1/2$ mit N Ziehungen, belasse es jedoch bei einem allgemeinen

$\varepsilon > 0$. Bezeichnet also S_N die Anzahl der Kopfwürfe in einem symmetrischen Bernoulli-Experiment, so untersuchen wir für $0 < \varepsilon < 1/2$

$$P\left(\left|S_N - \frac{N}{2}\right| \geq N\varepsilon\right) = 2P\left(S_N \geq \left(\frac{1}{2} + \varepsilon\right)N\right).$$

Wir schreiben

$$b_N(k) := P(S_N = k) = \binom{N}{k} 2^{-N}.$$

Aus den Quotienten

$$\frac{b_N(k)}{b_N(k+1)} = \frac{N!(k+1)!(N-k-1)!}{N!k!(N-k)!} = \frac{k+1}{N-k} \quad (2.1)$$

schließt Bernoulli, dass $b_N(\lceil \frac{N}{2} \rceil) = \max_k b_N(k)$ ist. Der Einfachheit halber nehme ich an, dass N gerade ist, sodass das Maximum exakt bei $N/2$ angenommen wird. Nach dem de Moivre'schen Zentralen Grenzwertsatz folgt, dass $b_N(N/2)$ sich asymptotisch wie $\sqrt{2}/\sqrt{\pi N}$ verhält, aber Bernoulli wusste das noch nicht. Auch Stirlings Formel wurde erst ein Vierteljahrhundert später entdeckt. Er untersucht nun für $r \geq 0$

$$\frac{b_N(N/2+r)}{b_N(N/2+\lceil N\varepsilon \rceil+r)} = \frac{N/2+r+1}{N/2-r} \frac{N/2+r+2}{N/2-r-1} \cdots \frac{N/2+r+\lceil N\varepsilon \rceil}{N/2-r-\lceil N\varepsilon \rceil+1}. \quad (2.2)$$

Diese Darstellung impliziert einerseits (mit $r = 0$)

$$\lim_{N \rightarrow \infty} \frac{b_N(N/2)}{b_N(N/2+\lceil N\varepsilon \rceil)} = \infty$$

und andererseits, dass die Quotienten auf der linken Seite von (2.2) monoton in r ansteigen. Somit konvergieren diese *gleichmäßig in r* gegen ∞ . Daraus folgt

$$\lim_{N \rightarrow \infty} \frac{\sum_{j=0}^{\lceil N\varepsilon \rceil-1} b_N(N/2+j)}{\sum_{j=0}^{\lceil N\varepsilon \rceil-1} b_N(N/2+\lceil N\varepsilon \rceil+j)} = \infty,$$

und verwendet man nochmals die Monotonie, so ergibt sich

$$\lim_{N \rightarrow \infty} \frac{\sum_{j=0}^{\lceil N\varepsilon \rceil-1} b_N(N/2+j)}{\sum_{j=0}^{\lceil N\varepsilon \rceil-1} b_N(N/2+k\lceil N\varepsilon \rceil+j)} = \infty, \quad 1 \leq k. \quad (2.3)$$

Da $\lceil 1/2\varepsilon \rceil$ Intervalle der Form $\{N/2+k\lceil N\varepsilon \rceil, \dots, N/2+(k+1)\lceil N\varepsilon \rceil-1\}$ genügen, die Menge $\{N/2, \dots, N\}$ zu überdecken, folgt

$$\lim_{N \rightarrow \infty} \frac{\sum_{j=0}^{\lceil N\varepsilon \rceil-1} b_N(N/2+j)}{\sum_{j=\lceil N\varepsilon \rceil}^N b_N(N/2+j)} = \infty, \quad (2.4)$$

woraus sich

$$\lim_{N \rightarrow \infty} P(S_N \geq N/2 + \lceil N\varepsilon \rceil) = 0$$

ergibt. Wegen der Symmetrie erhält man die gleiche Aussage auch für die Abweichung nach unten und mithin

$$\lim_{N \rightarrow \infty} P \left(\left| \frac{S_N}{N} - \frac{1}{2} \right| \geq \varepsilon \right) = 0. \quad (2.5)$$

An dieser Stelle ist Bernoullis Text sehr amüsant. Das Konzept der gleichmäßigen Konvergenz war zu dieser Zeit offenbar formal nicht präzise formuliert worden. Bernoulli schreibt hier (in der Übersetzung von R. Haussner in Ostwalds Klassikern): „Gegen [dieses Argument] könnte von denen, welche sich nicht mit Unendlichkeitsbetrachtungen befreundet haben, der folgende Einwurf gemacht werden...“. Dann führt er etwas verklausuliert aus, dass aus

$$\lim_{N \rightarrow \infty} \frac{b_N(N/2 + j)}{b_N(N/2 + \lceil N\varepsilon \rceil + j)} = \infty$$

für jedes j , nicht auf (2.3) geschlossen werden kann. Statt nun einfach auf die Monotonie und daraus auf die Gleichmäßigkeit der Konvergenz hinzuweisen, was ihm offensichtlich völlig klar war, schreibt er: „Diesen Bedenken kann ich nicht besser entgegentreten, als dass ich die Berechnungen für einen endlichen Wert von N wirklich durchführe.“ Mit anderen Worten: Er leitet eine konkrete Abschätzung der Quotienten (2.4) her und schreibt: „Ist diese aber gezeigt, so muss der Einwand notwendiger Weise in sich zusammenfallen.“

Die Abschätzung, die er durchführt, ist sehr interessant, denn sie führt auf exponentielle Konvergenz in (2.5), d.h. auf eine Abschätzung vom Typus (1.2). Bevor ich Bernoullis Herleitung vorstelle, sollen die wesentlichen Aspekte von (1.2) kurz diskutiert werden. Im symmetrischen Fall ist

$$I_{1/2}(t) = t \log 2t + (1-t) \log[2(1-t)],$$

und (1.2) impliziert für $\varepsilon > 0$:

$$P \left(\left| \frac{S_N}{N} - \frac{1}{2} \right| \geq \varepsilon \right) \leq 2 \exp \left[-N I_{1/2} \left(\frac{1}{2} + \varepsilon \right) \right].$$

Zwei Fakten sind bemerkenswert: Erstens fallen die Wahrscheinlichkeiten für festes $\varepsilon > 0$ exponentiell in N ab, denn $I_{1/2}(t) \neq 0$ für $t \neq 1/2$. Das ist quasi „good news“. Andererseits hat man für kleine ε folgende Rate des exponentiellen Abfalls:

$$I_{1/2} \left(\frac{1}{2} + \varepsilon \right) \sim 2\varepsilon^2.$$

Wenn man also nach einer festen Schranke für die Wahrscheinlichkeiten der Abweichungen trachtet, z.B. $1/1000$, so muss N proportional zum Inversen des Quadrates von ε gewählt werden. Das liegt nicht daran, dass die Ungleichung ungenau wäre, sondern an der Tatsache, dass die Standardabweichung von S_N/N nur mit $1/\sqrt{N}$ gegen 0 fällt.

Die Ungleichung hängt eng mit dem de Moivreschen Satz zusammen: Da sie für jedes ε und N gilt, können wir $\varepsilon = x/\sqrt{N}$ einsetzen, $x > 0$. In diesem Fall ist

$$I_{1/2} \left(\frac{1}{2} + \frac{x}{\sqrt{N}} \right) = \frac{2x^2}{N} + O(N^{-2}),$$

woraus sich

$$P\left(\left|\frac{S_N}{N} - \frac{1}{2}\right| \geq \frac{x}{\sqrt{N}}\right) \leq 2(1 + O(N^{-1})) \exp[-2x^2]$$

in Übereinstimmung mit dem de Moivreschen Gesetz ergibt, denn $\text{var}(X_i)$ ist $1/4$. Allerdings ist der Vorfaktor zu groß.

Bernoulli leitet eine Abschätzung her, welche diese wesentlichen Fakten schon widerspiegelt. Allerdings ist seine Rate um etwa einen Faktor 2 kleiner als die optimale.

Hier ist sein Argument, das ausschließlich auf der Auswertung der Quotienten (2.1) basiert. In dem Ausdruck auf der rechten Seite von (2.2) mit $r = 0$ strebt der erste Faktor für $N \rightarrow \infty$ gegen 1 und der letzte Faktor gegen $(1 + 2\varepsilon)/(1 - 2\varepsilon)$. Betrachtet man nun irgendeine reelle Zahl λ mit

$$1 < \lambda < \frac{1 + 2\varepsilon}{1 - 2\varepsilon}, \quad (2.6)$$

so ist die Anzahl der Faktoren, die λ übersteigen, proportional zu N . Da alle Faktoren mindestens 1 sind, lässt sich so das Produkt nach unten abschätzen. Bernoulli macht diesen einen Schnitt und schätzt also die Faktoren links vom Schnitt durch 1 und rechts davon durch λ ab. Durch diesen einen Schnitt erhält er eine Abschätzung des Ausdrucks nach unten, die exponentiell in N wächst, mit einer Rate des exponentiellen Wachstums, die vom gewählten λ und von ε abhängt. Natürlich hätte er auch zwei, drei oder mehr Schnitte machen können und hätte damit seine Abschätzung verbessert. Das ist so naheliegend, dass Bernoulli es gesehen haben muss, aber als nicht des Aufwands wert erachtet hat.

Bernoulli wählt $\lambda = 1 + 2\varepsilon$. Für festes ε ist das nicht ganz die optimale Wahl; es ist jedoch asymptotisch optimal für $\varepsilon \rightarrow 0$. Ich führe die Rechnungen ohne Berücksichtigung der kleinen (irrelevanten) Rundungsfehler durch: Der Schnitt ist also bei dem Faktor auf der rechten Seite von (2.2), der (bis auf eine Rundung) gleich $\lambda = 1 + 2\varepsilon$ ist. Es geht also um das j mit

$$\frac{N/2 + j}{N/2 - j + 1} \approx 1 + 2\varepsilon, \quad \text{d.h.} \quad j \approx \frac{\varepsilon N}{2(1 + \varepsilon)}.$$

Das Produkt hat daher

$$\varepsilon N - j \approx \left(\varepsilon - \frac{\varepsilon}{2(1 + \varepsilon)}\right) N = \frac{\varepsilon + 2\varepsilon^2}{2(1 + \varepsilon)} N$$

Faktoren, die mindestens $1 + 2\varepsilon$ sind. Somit erhalten wir

$$\frac{b_N(N/2)}{b_N(N/2 + \lceil N\varepsilon \rceil)} \geq \exp\left[\frac{\varepsilon + 2\varepsilon^2}{2(1 + \varepsilon)} \log(1 + 2\varepsilon)N\right],$$

mit einem zusätzlichen Faktor für die rechte Seite infolge der Rundungen, der aber für kleine ε und $N \rightarrow \infty$ bei 1 liegt. Mit den schon oben diskutierten Monotonieeigenschaften erhält man

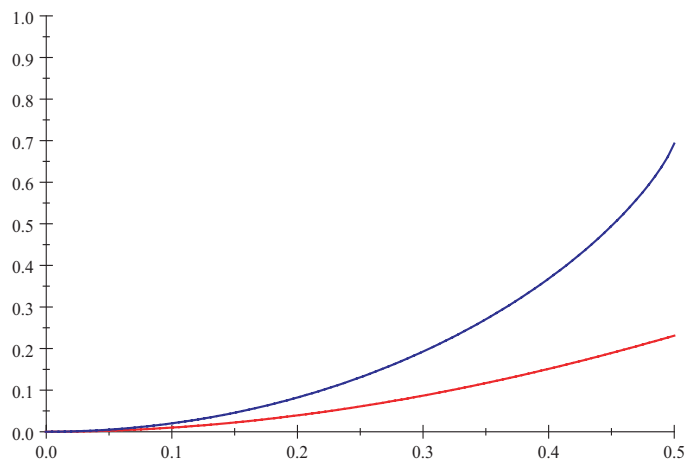
$$P\left(\left|\frac{S_N}{N} - \frac{1}{2}\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon} \exp\left[-\frac{\varepsilon + 2\varepsilon^2}{2(1 + \varepsilon)} \log(1 + 2\varepsilon)N\right]. \quad (2.7)$$

Bernoullis „Ratenfunktion“ ist also

$$\frac{\varepsilon + 2\varepsilon^2}{2(1 + \varepsilon)} \log(1 + 2\varepsilon),$$

was sich für kleine ε wie ε^2 verhält. Sie ist also um rund einen Faktor 2 kleiner als die korrekte Ratenfunktion. Interessant ist jedoch, dass Bernoulli sowohl den exponentiellen Abfall in N wie auch die quadratische Abhängigkeit der Ratenfunktion von ε für $\varepsilon \sim 0$ erhält.

Zum Vergleich sind unten (rot) diese Ratenfunktion und (blau) die korrekte Ratenfunktion $I_{1/2}(1/2 + \varepsilon)$ eingezeichnet.



Die Rechnungen, die hier im Hinblick auf die moderne Sichtweise der großen Abweichungen durchgeführt wurden, sind bei Bernoulli jedoch etwas versteckt. Aus seinem Text ist nicht ersichtlich, dass ihm die quadratische Abhängigkeit in ε aufgefallen ist. In der Tat wird diese durch seine Wahl $\varepsilon = 1/(r + s)$ etwas verschleiert. Auf den exponentiellen Abfall in N ist er jedoch sehr stolz, und betont dies auch sehr ausdrücklich.

Am Schluss seiner Untersuchung rechnet er ein Beispiel durch. Er nimmt $p = 0.6$, also nicht symmetrisch, $\varepsilon = 1/50$ und will eine Wahrscheinlichkeit von unter $1/1000$ für die Abweichung erhalten. Nach seiner Abschätzung benötigt man dafür 25 550 Repetitionen. Die korrekte Anzahl ist etwa 6 520. Die Ratenfunktion ist zwar nur um einen Faktor 2 zu klein, jedoch macht der schlechte Vorfaktor in (2.7) seine Abschätzung noch etwas schwächer.

Stephen Stigler hat in seinem Buch [4] angedeutet, dass der für die Praxis natürlich sehr große Wert von über 25 000 Bernoulli möglicherweise enttäuscht habe und er deshalb zu Lebzeiten von einer Veröffentlichung abgesehen hat. Auch der korrekte Wert ist noch sehr groß. Der Grund liegt einfach in der sehr kleinen Wahl von ε in seinem Beispiel und dem Fluktuationsverhalten des Bernoulli-Experiments, das Bernoulli vielleicht nicht klar geworden ist, obwohl es, wie oben ausgeführt, sich implizite in seinen Rechnungen verbirgt. Ob Bernoulli mit seiner Abschätzung etwas „unzufrieden“ war, lässt sich heute natürlich nicht mehr feststellen. Der einzige Hinweis dafür liegt in der Nichtpublikation der *Ars*

conjectandi zu seinen Lebzeiten. Nach Lektüre seines Textes sehe ich jedoch dafür keinen Anhaltspunkt. Er hätte leicht eine bessere Abschätzung erhalten können, indem er zusätzliche Schnitte – wie oben ausgeführt – in der Abschätzung des Produktes (2.2) eingeführt hätte. Offensichtlich sah er dafür keine Notwendigkeit.

3 Bernoulli als Begründer der Statistik

Der Grund für die verzögerte Publikation der *Ars conjectandi* wird im eingangs zitierten Brief an Leibniz angeführt. Bernoulli schreibt, dass ihm infolge seiner angeschlagenen Gesundheit das Schreiben schwer falle, dass er aber den größten Teil des Buches vollendet habe, fährt dann aber fort: „Es fehlt noch der wichtigste Teil, in welchem ich zeige, wie sich die Grundlagen der Mutmaßungskunst aufs Bürgerliche, Sittliche und Wirtschaftliche anwenden lassen.“ Bernoulli führt dann kurz aus, worum es geht, nämlich unbekannte Wahrscheinlichkeiten über Stichproben zu approximieren. Als Beispiel nennt er die Wahrscheinlichkeit, mit der ein Greis von 60 Jahren einen Jüngling von 20 Jahren überlebt. Er schreibt: „Von daher begann ich zu überlegen, ob vielleicht das, was uns a priori nicht bekannt ist, uns wenigstens a posteriori bekannt werden könnte aus dem Ausgang in großer Zahl beobachteter ähnlicher Beispiele.“

Was er im Sinne hat, ist offenbar eine wissenschaftlich fundierte Durchdringung weiterer Bereiche des öffentlichen Lebens mit Hilfe statistischer Methoden. Wie bekannt, haben die Grundlagen dafür zu großen Kontroversen noch im 20. Jh. geführt. Bernoulli schwebt eine frequentistische Interpretation vor: Unbekannte Wahrscheinlichkeiten werden mit einer Vielzahl wiederholter Messungen statistisch bestimmt, wobei sich die wahre Wahrscheinlichkeit im Limes unendlich vieler Beobachtungen ergibt. Er vergleicht das mit der Berechnung der Kreiszahl π über stets feinere Approximationen.

Leibniz scheint von den Ideen zunächst nicht sehr angetan gewesen zu sein und äußert gegenüber Bernoulli Vorbehalte. Im Wesentlichen sieht er die Schwierigkeit, für etwas so Komplexes wie Krankheiten Wahrscheinlichkeiten, wie für das Ziehen von Steinchen aus einer Urne, festzulegen. Er sieht auch das Problem, dass anders als bei Urnenproblemen die zeitliche Konstanz der Situation nicht garantiert werden kann. Seine Vorbehalte erinnern zum Teil an Einwände gegen die frequentistische Interpretation von Wahrscheinlichkeiten, z.B. durch de Finetti im 20. Jahrhundert.

Bernoulli geht in seinem Text kurz auf die „Einwände [...], welche einige Gelehrte erhoben haben“ ein. Er nennt keine Namen, aber es bezieht sich wohl hauptsächlich auf Leibniz. Bernoullis Diskussion zu diesen statistischen Fragen ist sehr kurz, und es ist naheliegend, dass er diese Aspekte breiter darstellen wollte, dass ihm jedoch die Zeit dafür gefehlt hat.

Einige Stellen in seinem Text lassen darauf schließen, dass er eine Art Konfidenzintervalle im Auge hatte, denn er schreibt, dass eine genaue Bestimmung der Wahrscheinlichkeiten unmöglich sei und man sich mit Approximationen begnügen müsse. Im Gegensatz zu der Situation mit der Zahl π können für die Erfolgswahrscheinlichkeiten p im Bernoulli-Experiment selbstverständlich keine Intervalle gefunden werden, in denen p sicher liegt. Das ist Bernoulli bewusst gewesen. Andererseits gibt es keinen Hinweis darauf, dass er Konfidenzintervalle begrifflich ganz klar konzipiert hat. In der modernen Statistik erfolgt das über die Testtheorie.

4 Spätere Entwicklungen

Das Gesetz der großen Zahlen ist eine der ganz großen Leistungen der Mathematik der Neuzeit. Genauso wichtig wie das rein mathematische Resultat ist der Gedanke Bernoullis, mit seiner Hilfe den Anwendungsbereich der Wahrscheinlichkeitstheorie auf Bereiche außerhalb von Glücksspielen auszudehnen.

Wie vital das Gesetz in der Mathematik nach wie vor ist, erkennt man an vielen neuen Entwicklungen, die auf dessen Grundgedanken aufbauen. Ich will auf einige Beispiele aus der Wahrscheinlichkeitstheorie kurz eingehen.

Das Gesetz ist ein Spezialfall eines sehr allgemeinen Maßkonzentrations-Phänomens, das in den letzten Jahrzehnten von vielen Mathematikern intensiv und tiefgehend untersucht wurde, z.B. von Talagrand. Ich erläutere es kurz für den einfachsten Fall des symmetrischen Bernoulli-Experiments der Länge N . Ist A_N das Ereignis, dass höchstens $N/2$ Kopfwürfe auftreten, so hat dieses Ereignis aus Symmetriegründen eine Wahrscheinlichkeit $\geq 1/2$, exakt $1/2$, wenn N ungerade ist. Bernoullis Gesetz lässt sich wie folgt formulieren: Sei $\varepsilon > 0$. Eine Sequenz von Kopf-Zahl-Würfeln, die mehr als $N/2 + \varepsilon N$ Kopfwürfe enthält, wird dadurch charakterisiert, dass es *keine* Sequenz in A_N gibt, die sich von ihr an weniger als εN Stellen unterscheidet. Bernoulli weist nach, dass die Wahrscheinlichkeit dafür exponentiell in N gegen 0 fällt. Interessant ist das Faktum, dass eine entsprechende Aussage für völlig *beliebige* Folgen A_N von Ereignissen mit $\inf_N P(A_N) > 0$ gilt. Anders ausgedrückt: Haben Ereignisse A_N makroskopisch große Wahrscheinlichkeiten für $N \rightarrow \infty$, so haben εN -Umgebungen bezüglich des Hamming-Abstandes Wahrscheinlichkeiten, die exponentiell nahe bei 1 liegen. Solche Maßkonzentrations-Sätze sind für sehr allgemeine Situationen bewiesen worden und haben viele Anwendungen gefunden, z.B. für den Nachweis von „self-averaging“ in der Theorie ungeordneter Medien.

Ein anderes weites und sehr aktives Feld ist natürlich die Ergodentheorie. Birkhoffs berühmter Ergodensatz besagt, dass für eine *beliebige* stationäre Folge X_1, X_2, \dots von integrierbaren Zufallsgrößen die Folge der arithmetischen Mittel S_N/N fast sicher – und als Konsequenz auch im „schwächeren“ Sinn von Bernoulli – konvergiert, allerdings im allgemeinen nicht mit Bernoullis exponentiellen Abschätzungen für Abweichungen. Stationarität bedeutet, dass (X_1, X_2, \dots) dieselbe Verteilung wie (X_2, X_3, \dots) hat. Diese Bedingung ist für das Bernoulli-Experiment erfüllt. Birkhoffs Ergodensatz war der Ausgangspunkt für eine rasante Entwicklung. Ich erwähne nur die Verallgemeinerung durch Kingmans subadditiven Ergodensatz, der aus der modernen Wahrscheinlichkeitstheorie nicht wegzudenken ist.

Die präzise Beschreibung des exponentiellen Abfalls für Abweichungen hat eine lange Geschichte in der Wahrscheinlichkeitstheorie. Von großer Bedeutung sind die Sätze von Cramér über Summen von unabhängigen Zufallsvektoren und von Sanov über empirische Maße. Seit Mitte der 70er Jahre des letzten Jahrhunderts hat die Theorie durch Arbeiten von M.D. Donsker und S.R.S. Varadhan eine enorme Erweiterung erfahren, vor allem mit sehr allgemeinen Resultaten vom Sanovschen Typus für stochastische Prozesse und deren konzeptionelle Verknüpfung mit Variationsproblemen. Varadhan wurde 2007 unter anderem für diese Beiträge mit dem Abel-Preis ausgezeichnet.

Verdankung

Ich bedanke mich bei Stephen Stigler für einige Anregungen, insbesondere für den Hinweis auf den Briefwechsel mit Leibniz.

Literatur

- [1] Bernoulli, J.: *Wahrscheinlichkeitsrechnung (Ars conjectandi)*. Ostwalds Klassiker der exakten Wissenschaften, W. Engelmann in Leipzig 1899.
- [2] Khinchin, A.: Sur la loi des grands nombres. *Comptes rendus de l'Académie des Sciences* 189 (1929), 477–479.
- [3] Kohli, K.: Aus dem Briefwechsel zwischen Leibniz und Jakob Bernoulli. In: *Die Werke von Jakob Bernoulli*. Vol. 3, Basel 1975, 509–513.
- [4] Stigler, S.M.: *The History of Statistics: The Measurement of Uncertainty before 1900*. The Belknap Press of the Harvard University Press, Cambridge, Massachusetts 1986.

Erwin Bolthausen
Universität Zürich
Institut für Mathematik
Winterthurerstrasse 190
CH-8057 Zürich, Schweiz
e-mail: eb@math.uzh.ch