Stochastik

Barth, Friedrich

München, [20]03

14. 7. Die Ungleichung von Bienaymé-Tschebyschow für binomial verteilte Zufallsgrößen und das Gesetz der großen Zahlen

Figur 247.1 zeigt, wie gut diese Näherung ist.
Wir haben bereits oben (Seite 244) gezeigt, daß σ bei festem n für $p = \frac{1}{2}$ am größten wird. Also muß $\mathrm{M}(n; p)$ bei festem n für $p = \frac{1}{2}$ bezüglich p am kleinsten werden, was Beobachtung **5.** entspricht. Andererseits wächst σ bei festem p mit n echt monoton; also nimmt $\mathrm{M}(n; p)$ echt monoton ab (Beobachtung **10.**).
Anschaulich ist dies alles klar: Da die Histogramme immer breiter werden, ihre Flächeninhalte aber konstant den Wert 1 haben, sollte das höchste Rechteck des Histogramms immer niedriger werden.

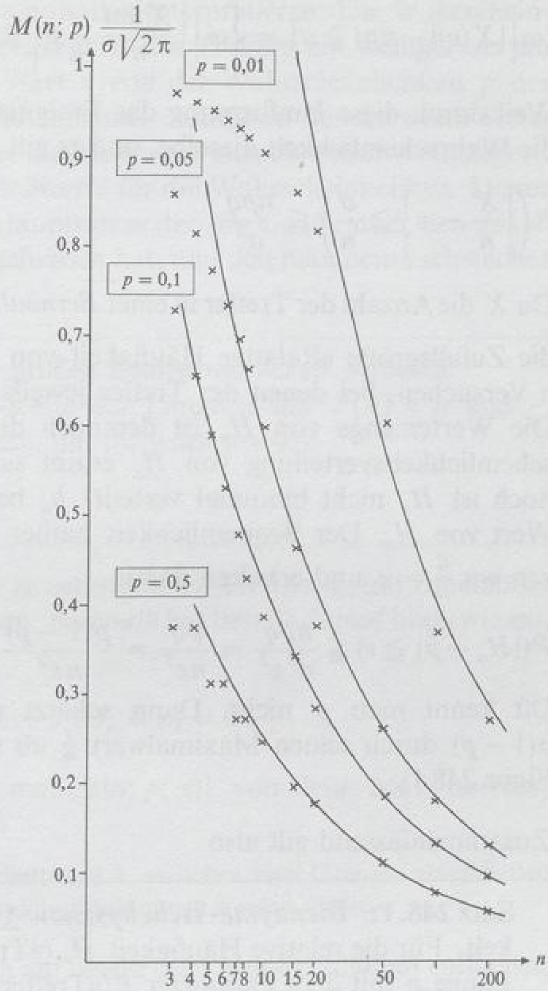

Fig. 247.1 Güte der Näherungsformel für die Maxima von Binomialverteilungen
Einzelpunkte: Maximalwerte $M(n; p)$ der Binomialverteilungen $\mathrm{B}(n; p)$.
Durchgezogene Kurven: zugehörige Näherungen $(\sigma\sqrt{2\pi})^{-1}$.
Beachte: Auf der n-Achse logarithmischer Maßstab!

14.7. Die Ungleichung von *Bienaymé-Tschebyschow* für binomial verteilte Zufallsgrößen und das Gesetz der großen Zahlen

Wenden wir die Ungleichung von *Bienaymé-Tschebyschow*, nämlich

$$P(|X - \mu| \geqq a) \leqq \frac{\mathrm{Var}\, X}{a^2},$$

auf binomial nach $\mathrm{B}(n; p)$ verteilte Zufallsgrößen X an, dann lassen sich μ und $\mathrm{Var}\, X$ durch np bzw. npq ersetzen, und wir erhalten

$$P(|X - np| \geqq a) \leqq \frac{npq}{a^2}.$$

Die Ungleichung $|X - np| \geqq a$ beschreibt kurz das Ereignis $\{\omega \,|\, |X(\omega) - np| \geqq a\}$. Dividiert man die in der Mengenklammer stehende Ungleichung durch n, so wird weiterhin dasselbe Ereignis beschrieben, also

$$\{\omega \,|\, |X(\omega) - np| \geqq a\} = \left\{\omega \,\middle|\, \left|\frac{X(\omega)}{n} - p\right| \geqq \frac{a}{n}\right\}.$$

Weil durch diese Umformung das Ereignis nicht verändert wurde, bleibt auch die Wahrscheinlichkeit dieselbe, und es gilt

$$P\left(\left|\frac{X}{n} - p\right| \geqq \frac{a}{n}\right) \leqq \frac{npq}{a^2}.$$

Da X die Anzahl der Treffer in einer *Bernoulli*-Kette der Länge n ist, stellt $\frac{X}{n} =: H_n$ die Zufallsgröße »Relative Häufigkeit von Treffer in einer *Bernoulli*-Kette von n Versuchen, bei denen der Treffer jeweils die Wahrscheinlichkeit p hat« dar. Die Wertemenge von H_n ist demnach die Menge $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$, die Wahrscheinlichkeitsverteilung von H_n ergibt sich zu $P(H_n = \frac{k}{n}) = \mathrm{B}(n; p; k)$. Dennoch ist H_n nicht binomial verteilt! h_n bezeichne weiterhin einen bestimmten Wert von H_n. Der Bequemlichkeit halber setzen wir $\frac{a}{n} =: \varepsilon$ und erhalten damit

$$P(|H_n - p| \geqq \varepsilon) \leqq \frac{npq}{n^2 \varepsilon^2} = \frac{pq}{n\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

Oft kennt man p nicht. Dann schätzt man $p(1-p)$ durch seinen Maximalwert $\frac{1}{4}$ ab (vgl. Figur 248.1).

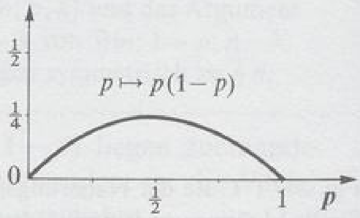

Fig. 248.1 Graph der Funktion $p \mapsto p(1-p)$

Zusammenfassend gilt also

> **Satz 248.1: *Bienaymé-Tschebyschow*-Ungleichung für die relative Häufigkeit.** Für die relative Häufigkeit H_n (»Treffer«) in einer *Bernoulli*-Kette der Länge n mit dem Parameter $P(\text{»Treffer«}) = p$ gilt:
>
> $$P(|H_n - p| \geqq \varepsilon) \leqq \frac{pq}{n\varepsilon^2} \leqq \frac{1}{4n\varepsilon^2}$$

Bemerkung: Das *Tschebyschow*-Risiko $r_T = \frac{\operatorname{Var} X}{a^2}$ wird hier zu $r_T = \frac{pq}{n\varepsilon^2}$ und beträgt höchstens $\frac{1}{4n\varepsilon^2}$.

Sowohl in der Interpretationsregel für Wahrscheinlichkeiten (**5.2.**) wie auch beim Versuch der Definition der Wahrscheinlichkeit eines Ereignisses durch *v. Mises* wird ein intuitiver Zusammenhang zwischen relativer Häufigkeit und Wahrscheinlichkeit sichtbar. Satz 248.1 gibt uns nun die Möglichkeit, diesen Zusammenhang zu erkennen. Dazu schreiben wir die *Tschebyschow*-Ungleichung von Satz 248.1 für das Gegenereignis auf, also

$$P(|H_n - p| < \varepsilon) \geqq 1 - \frac{pq}{n\varepsilon^2}.$$

Diese Ungleichung können wir folgendermaßen interpretieren: Die Wahrscheinlichkeit dafür, daß sich die relative Häufigkeit des Treffers um weniger als ein beliebig kleiner, aber fest gewählter Wert ε von der Wahrscheinlichkeit p des Treffers unterscheidet, wächst mit zunehmender Länge n der *Bernoulli*-Kette und kommt dem Wert 1 beliebig nahe. Damit erweist sich die relative Häufigkeit für hinreichend großes n als guter »Meßwert« für die Wahrscheinlichkeit. Dieser Sachverhalt ist die Aussage des sog. Hauptsatzes der *Ars Conjectandi*, den *Jakob Bernoulli* (1655–1705) wohl um 1685 gefunden hat, und den man heute schwaches Gesetz der großen Zahlen nennt.*

Satz 249.1: Schwaches Gesetz der großen Zahlen von *Jakob Bernoulli*.
Ist A der Treffer einer *Bernoulli*-Kette der Länge n mit $P(A) = p$ und $H_n(A)$ seine relative Häufigkeit, dann gilt für jedes $\varepsilon > 0$:

$$\lim_{n \to \infty} P(|H_n - p| < \varepsilon) = 1$$

Man könnte nun versucht sein, $\varepsilon = 0$ zu setzen, in der Hoffnung, mit zunehmendem n schließlich p exakt zu bestimmen. *Bernoulli* hat bereits darauf hingewiesen,

»daß sich dann das Gegenteil ergäbe«,

nämlich $$\lim_{n \to \infty} P(|H_n - p| = 0) = \lim_{n \to \infty} P(H_n = p) = 0,$$

was mit unserer Beobachtung über $\max\{B(n; p; x)\}$ von Seite 246f. übereinstimmt, und daß wir den Wert von p

»nur mit einer bestimmten Annäherung erhalten, d.h. zwischen zwei Grenzen einschließen können, welche aber beliebig nahe beieinander angenommen werden dürfen«.

Der scheinbare Widerspruch klärt sich auf, wenn man bedenkt, daß im endlichen Intervall $]p - \varepsilon; p + \varepsilon[$ für großes n sehr viel mögliche Werte von H_n liegen, die alle im Abstand $\frac{1}{n}$ aufeinanderfolgen. Es gibt also ungefähr $\frac{2\varepsilon}{\frac{1}{n}} = 2n\varepsilon$ Werte für H_n in diesem Intervall, von denen jeder zwar eine verschwindend kleine Wahrscheinlichkeit hat, die Summe all dieser Wahrscheinlichkeiten aber nahezu 1 ergibt.

Was besagt im Sinne der Analysis eigentlich $\lim_{n \to \infty} P(|H_n - p| < \varepsilon) = 1$? Diese Gleichung drückt doch aus, daß sich bei fest vorgegebenem positiven ε zu jeder beliebigen Schranke $\eta > 0$ eine Länge n_0 für *Bernoulli*-Ketten des Parameters p

* *Bernoulli* hat, wie er selbst in der *Ars Conjectandi* (ed. 1713) wohl um 1703/4 schreibt, dieses Problem schon 20 Jahre mit sich herumgetragen. Wie stolz er auf diesen Satz war, zeigen seine Worte am Schluß des Beweises in seinen Tagebüchern:

»Hoc inventum pluris facio quam si ipsam circuli quadraturam dedissem, quod si maxime reperiretur, exigui usus esset.«

»Diese Entdeckung gilt mir mehr, als wenn ich gar die Quadratur des Kreises geliefert hätte; denn wenn diese auch gänzlich gefunden würde, so wäre sie doch sehr wenig nütz.«

Der Name *Gesetz der großen Zahlen* stammt von *Siméon-Denis Poisson* (1781–1840), der 1837 einen allgemeinen Satz veröffentlichte, den er *la loi des grands nombres* nannte, und von dem das *Bernoulli*sche Gesetz der großen Zahlen ein Spezialfall ist.

finden läßt, so daß für alle $n \geqq n_0$ die Wahrscheinlichkeit dafür, daß sich die relative Trefferhäufigkeit um weniger als ε von der Wahrscheinlichkeit p für einen Treffer unterscheidet, mindestens $1-\eta$ wird, daß also $P(|H_n - p| < \varepsilon) \geqq 1-\eta$ gilt. Nehmen wir z.B. $\eta = \frac{1}{10}$, so bedeutet $P(|H_n - p| < \varepsilon) \geqq 90\%$ nach der Interpretationsregel für Wahrscheinlichkeiten: Bestimmt man sehr oft die relative Häufigkeit H_n des Treffers in *Bernoulli*-Ketten einer Länge $n \geqq n_0$ zum selben Parameter p, so erhält man in ungefähr mindestens 90% aller Fälle Werte h_n, die in das Intervall $]p-\varepsilon; p+\varepsilon[$ fallen. Diesen Sachverhalt drückt man dadurch aus, daß man sagt, $\boldsymbol{H_n}$ **konvergiere in Wahrscheinlichkeit nach** $\boldsymbol{p}$, oder auch, $\boldsymbol{H_n}$ **konvergiere stochastisch nach** $\boldsymbol{p}$. Figur 250.1 veranschaulicht diese Art von Konvergenz.

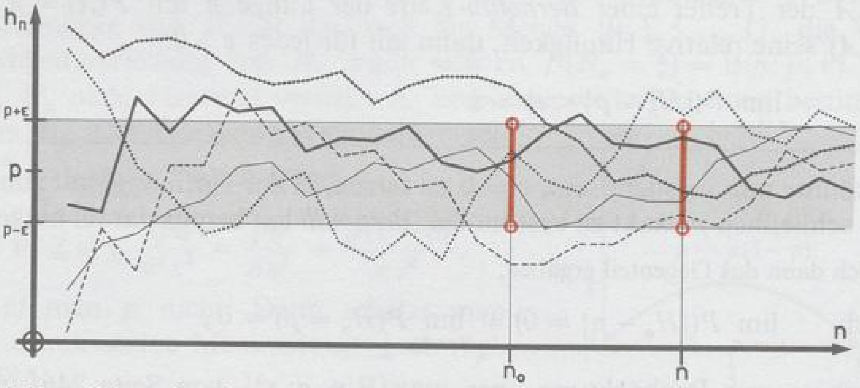

Fig. 250.1 Zum Schwachen Gesetz der großen Zahlen: Es gibt ein n_0, so daß für alle $n \geqq n_0$ die Wahrscheinlichkeit dafür, daß die Werte h_n der relativen Häufigkeit H_n in das Intervall $]p-\varepsilon; p+\varepsilon[$ fallen, mindestens $1-\eta$ beträgt. – Anschaulich: Der Anteil der Schlangen, die durch das 2ε-Tor um p hindurchgehen, ist für $n \geqq n_0$ etwa $1-\eta$.*

Aus der stochastischen Konvergenz von H_n darf auf keinen Fall geschlossen werden, daß von dem gefundenen n_0 ab die relative Häufigkeit für noch größere Längen in dem Intervall $]p-\varepsilon; p+\varepsilon[$ bleibt, d.h., daß etwa $\lim\limits_{n\to\infty} H_n(A) = P(A)$ gelte! Eine etwas schwächere Behauptung als diese hat im Jahre 1909 *Émile Borel* (1871–1956) für $p = \frac{1}{2}$ gefunden. Sie wurde 1917 von *Francesco Paolo Cantelli* (1875–1966) für $0 < p < 1$ verallgemeinert und heißt

Das starke Gesetz der großen Zahlen:

$$P(\lim_{n\to\infty} H_n = p) = 1$$

Es besagt, daß die relative Häufigkeit **fast sicher** gegen die zugehörige Wahrscheinlichkeit **konvergiert**.

Wir verzichten auf den Beweis, da wir dazu unendliche Ergebnisräume benötigten.

* Jede gezeichnete Schlange ist folgendermaßen entstanden: Zu jedem n werden n unabhängige Versuche gemacht, und dann h_n bestimmt. Z. B.: Um h_{100} zu bestimmen, müssen 100 unabhängige Versuche gemacht werden. Um dann eine Schlange bis $n = 100$ zeichnen zu können, müssen $1 + 2 + \ldots + 100 = 5050$ Versuche ausgeführt werden! Man darf die Schlangen von Figur 250.1 nicht mit denen der Figuren 31.1, 33.1, 34.1 und 71.1 verwechseln, die die Entwicklung von h_n darstellen. So ist z.B. in Figur 31.1 die Entstehung von h_{800}(»Adler«) $= \frac{1}{2}$ dargestellt; die Schlange gibt also die Entwicklung für diesen einen Wert an.

Das schwache Gesetz der großen Zahlen rechtfertigt unsere Interpretationsregel für Wahrscheinlichkeiten, d.h. die statistische Bestimmung von Wahrscheinlichkeiten. Um mit *Jakob Bernoulli* zu sprechen: Wir können die Wahrscheinlichkeit

»a posteriori fast ebenso genau finden, als wenn sie uns a priori bekannt«

wäre. Es liefert uns also gewissermaßen eine Meßvorschrift für die Wahrscheinlichkeit von solchen Ereignissen, die unter gleichen Bedingungen beliebig oft wiederholbar sind. Die Wahrscheinlichkeit solcher Ereignisse läßt sich damit wie eine physikalische Konstante messen!

Bei flüchtiger Betrachtungsweise könnte man meinen, daß im Gesetz der großen Zahlen ein Zirkelschluß vorliegt, da es eine Aussage über einen Zusammenhang zwischen der relativen Häufigkeit eines Ereignisses und seiner Wahrscheinlichkeit macht, den man über die Interpretationsregel schon zur Grundlage der Definition der Wahrscheinlichkeit gemacht hat. Ein solcher circulus vitiosus liegt aber nicht vor, weil wir als Grundlage der mathematischen Theorie der Wahrscheinlichkeit die Wahrscheinlichkeit eines Ereignisses im Axiomensystem von *Kolmogorow* völlig unabhängig vom Begriff der relativen Häufigkeit definiert haben. Das Gesetz der großen Zahlen zeigt nun, daß diese abstrakte Definition der Wahrscheinlichkeit genau den realen Hintergrund erfaßt, für dessen Beschreibung man die Wahrscheinlichkeitstheorie geschaffen hatte. Wir können nun auch noch verstehen, warum wir das Empirische Gesetz der großen Zahlen, die Stabilisierung der relativen Häufigkeit um einen festen Wert, nicht präzise formulieren konnten. Wir benötigen zu diesem Zweck nämlich den Begriff der Wahrscheinlichkeit. Das schwache Gesetz der großen Zahlen drückt diese Stabilisierung aus; es besagt ja gerade, daß große Abweichungen der relativen Häufigkeit von diesem festen Wert nach einer sehr langen Versuchsreihe sehr unwahrscheinlich sind.

Die Aussage des schwachen Gesetzes der großen Zahlen wird von vielen Leuten mißverstanden. So neigen manche Lottospieler wie einst *d'Alembert* (1717–1783) dazu, gerade diejenigen Zahlen zu tippen, die bei den bis dahin erfolgten Ausspielungen sehr selten erschienen sind. Sie meinen nämlich, das schwache Gesetz der großen Zahlen arbeite wie ein Buchhalter, der darauf achtet, daß alle Zahlen gleich oft gezogen werden. Das schwache Gesetz der großen Zahlen arbeitet aber anders, nämlich gewissermaßen durch Überschwemmung*: Defizite oder Überschüsse, die sich bei den *absoluten* Häufigkeiten im Laufe der Zeit ergeben, werden in der *relativen* Häufigkeit dadurch ausgebügelt, daß sie als Differenzen im Zähler bei sehr großem Nenner keine Rolle mehr spielen. So hat z.B. die Zahl 13, wie die Tabelle zu Aufgabe 38/7 zeigt, nach 1225 Ziehungen ein Defizit von 29 gegenüber dem Sollwert von 150. Das bedeutet für die relative Häufigkeit ein Defizit von $\frac{29}{1225} < 2{,}4\%$. Dasselbe Defizit von 29 würde bei 10000 Ziehungen in der relativen Häufigkeit nur mehr 0,29% ausmachen; nach 1 Million Ziehungen spielt dieses Defizit mit 0,0029% aber keine Rolle mehr.

Analog sorgt beim *Galton*brett das schwache Gesetz der großen Zahlen dafür, daß auf lange Sicht, wenn immer mehr Kugeln durch den Nagelwald laufen, die Fächer immer genauer nach $B(n; \frac{1}{2})$ gefüllt werden. Dabei ist es offensichtlich

* swamping effect – *L.H.C. Tippett* prägte 1943 diesen Begriff.

unsinnig anzunehmen, daß eine startende Kugel weiß, in welchem Fach gerade Defizit herrscht, um bevorzugt dorthin zu springen.
Unterstellt man dem schwachen Gesetz der großen Zahlen also einen Buchhaltercharakter, so müßte man wider alle Vernunft annehmen, daß stochastische Geräte Gewissen und Gedächtnis hätten, wie es *Joseph Bertrand* (1822–1900) einmal treffend formulierte*. Wäre dem so, entgegnete 1785 *Leonhard Euler* (1707 bis 1783) in seinen *Opuscula Analytica*** der Auffassung *d'Alemberts*,

»dann müßte jeder nach einem Jahr, ja nach einem Jahrhundert stattfindende Zug vom Ergebnis aller Züge abhängen, die seit undenklichen Zeiten an irgendwelchen Orten dieser Erde stattgefunden haben; Absurderes kann sicherlich kaum gedacht werden.«

14.8. Anwendungen der Ungleichung von *Bienaymé-Tschebyschow*

Die Ungleichung von *Bienaymé-Tschebyschow* kann, je nach Bedarf, unterschiedlich formuliert werden. Wir stellen die drei häufigsten Formulierungen der *Bienaymé-Tschebyschow*-Ungleichung in der Form, in der sie sich am leichtesten merken lassen, zusammen:

1) Ist X eine Zufallsgröße mit $\mathscr{E}X = \mu$ und ist $a > 0$, dann gilt

$$P(|X-\mu| \geqq a) \leqq \frac{\operatorname{Var} X}{a^2}. \qquad \text{(Satz 184.1)}$$

2) Ist H_n die relative Häufigkeit eines Ereignisses mit der Wahrscheinlichkeit p in einer *Bernoulli*-Kette der Länge n und ist $\varepsilon > 0$, dann gilt

$$P(|H_n - p| \geqq \varepsilon) \leqq \frac{pq}{n\varepsilon^2} \leqq \frac{1}{4n\varepsilon^2}. \qquad \text{(Satz 248.1)}$$

3) Ist $\bar{X}_n$ das arithmetische Mittel n gleichverteilter, paarweise unabhängiger Zufallsgrößen X_i mit $\mathscr{E}X_i = \mu$ und $\operatorname{Var} X_i = \sigma^2$ und ist $a > 0$, dann gilt

$$P(|\bar{X}_n - \mu| \geqq a) \leqq \frac{\sigma^2}{na^2}. \qquad \text{(Aufgabe 271/71)}$$

Viele Aufgaben der Wahrscheinlichkeitsrechnung handeln davon, daß das wahre Risiko, d.h., daß die Wahrscheinlichkeit dafür, daß die Werte einer Zufallsgröße X von ihrem Erwartungswert μ um mindestens a abweichen, eine gewisse Schranke η nicht überschreiten soll, kurz, daß

$$P(|X-\mu| \geqq a) \leqq \eta \tag{1}$$

sein soll. Anders ausgedrückt: Die Wahrscheinlichkeit, daß die Werte von X sich um weniger als a von μ unterscheiden, soll einen gewissen Mindestwert besitzen, d.h.,

* »On fait trop d'honneur à la roulette: elle n'a ni conscience ni mémoire.« (*Calcul des Probabilités*, p. XXII, 1889)
** Die Abhandlung lautet *Solutio quarundam quaestionum difficiliorum in Calculo Probabilium*. – *Friedrich II.* bat *Euler* 1749 und 1763 um Rat bezüglich der Errichtung von Lotterien, um die Finanznot seines Staates zu beheben. Aus der Beschäftigung mit diesem Problem entstanden *Eulers* wahrscheinlichkeitstheoretische Arbeiten.