Zur Geschichte der „analytischen“ Wahrscheinlichkeitstheorie: vom Haupt- zum Nieschenthema

Hans Fischer

Klärung des Rahmens

Unter „analytischer“ Wahrscheinlichkeitstheorie wollen wir in Nachfolge von Laplace eine Wahrscheinlichkeitsrechnung verstehen, die sich der Methoden der (endlichdimensionalen) reellen und der komplexen Analysis bedient ohne wesentlichen Rückgriff auf maßtheoretische Konzepte und Methoden.

Hauptobjekte dieser Theorie sind (im Eindimensionalen, auf das wir uns im folgenden beschränken, X bezeichnet eine Zufallsgröße):

- Die Verteilungsfunktion $V(x) = P(X \leq x)$, monoton steigend,

$$V(-\infty) = 0, \quad V(\infty) = 1$$

- Das Riemann-Stieltjes-Integral $\int_{\mathbb{R}} f(x)dV(x)$
- Charakteristische Funktionen $\varphi_X(t) = \mathrm{E}e^{itX} = \int_{\mathbb{R}} e^{itx}dV(x)$
- Momente $\mathrm{E}X^n := \int_{\mathbb{R}} x^n dV(x)$

Die Verwendung dieses Begriffsinventars geht wesentlich auf Richard von Mises [1919] zurück.

Das Laplacesche Paradigma

Der „eigentliche“ Beginn der mathematischen Wahrscheinlichkeitstheorie im Sinne eines einigermaßen geschlossenen Bereichs aus spezifischen Problemen und mathematischen Methoden kann mit dem Erscheinen von Laplaces *Théorie analytique des probabilités* [1812] angesetzt werden. Der Titel ist dabei alles andere als zufällig oder irgendeiner Konvention geschuldet: Laplace sieht seine „Theorie“ gerade wegen der hier dargelegten analytischen Methoden (Fouriermethoden, Approximationsmethoden, Berechnung bestimmter Integrale transzendenter Funktionen, rudimentäre komplexe Analysis) als bedeutsam für die gesamte Mathematik an. Mit Laplace beginnt somit auch eine bis zum Ende des 19. Jahrhunderts andauernde Tradition, in der Probleme der Wahrscheinlichkeitsrechnung nicht nur wegen ihres anwendungsbezogenen Charakters, sondern auch – und teilweise sogar vorrangig – zur Illustration spezieller analytischer Verfahren dienten und deren Weiterentwicklung motivierten.

Besonders bedeutend für die Laplacesche Wahrscheinlichkeitsrechnung war der zentrale Grenzwertsatz, der um 1810 im Rahmen der Fehlerrechnung und in späteren Arbeiten immer wieder im Sinne einer auf den jeweiligen Anwendungsfall bezogenen Approximation von neuem hergeleitet wurde. In moderner, allgemeiner Form handelt es sich um folgenden Sachverhalt: $X_1, X_2, \ldots, X_n$ seien identisch verteilte (mit Dichte bzw. gitterverteilt), unabhängige Zufallsgrößen. Dann besteht mit den Abkürzungen $\mu := \mathrm{E}X_1$, $\sigma^2 := \mathrm{E}(X_1-\mu)^2$ für „sehr großes" n die Näherung

$$P\left(n\mu + r_1\sqrt{n} \leq \sum X_k \leq n\mu + r_2\sqrt{n}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{r_1}^{r_2} \frac{1}{\sigma} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt.$$

Aus heutiger Sicht läßt sich das in die Grenzwertsaussage übertragen:

$$P\left(\frac{\sum(X_k-\mu)}{\sigma\sqrt{n}} \leq r\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r} \exp\left(-\frac{t^2}{2}\right) dt = \Phi(r).$$

Diese Approximation eröffnete für Laplace eine große Vielfalt von stochastischen Anwendungen, etwa in der Fehlerrechnung, im Testen von Hypothesen, oder auch in einem frühen Versuch zu einer Theorie des Risikos. Damit bezog die Laplacesche Wahrscheinlichkeitsrechnung hohe Relevanz aus der Summe ihrer verschiedenen, auf (fast) alle Lebensbereiche bezogenen Anwendungen. Daneben stand aber die bereits oben dargelegte analytische Durchdringung stochastischer Probleme. Besonders wichtig für Laplaces Herleitung des zentralen Grenzwertsatzes war die Verwendung charakteristischer Funktion in diskreter Rohform, besonders hinsichtlich der Grundeigenschaft

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) \tag{1}$$

für unabhängige Zufallsgrößen X, Y. Aber auch Laplaces Approximationstechnik für Integrale der Form $\int_a^b g(x)[f(x)]^n dx$ bei „sehr großem" n, wenn $f > 0$ sein Maximum in $]a, b[$ annimmt, spielte eine entscheidende Rolle.

Um ein Beispiel für Laplaces analytische Innovationen zu geben, wollen wir seine für die Approximation durch die Normalverteilung benötigte Herleitung des Integrals

$$\int_{-\infty}^{\infty} e^{izx} e^{-x^2} dx = \int_{-\infty}^{\infty} \cos(zx) e^{-x^2} dx \tag{2}$$

geben (s. z.B. [Laplace 1812, 97]). Man führt zuerst für den Exponenten in

$$\int_{-\infty}^{\infty} e^{izx} e^{-x^2} dx = \int_{-\infty}^{\infty} e^{-x^2+izx} dx$$

die quadratische Ergänzung

$$-x^2 + izx = -(x^2 - izx - z^2/4) - z^2/4 = -(x - iz/2)^2 - z^2/4$$

durch. Mit der Substitution $x - iz/2 =: u$ ergibt sich

$$\int_{-\infty}^{\infty} e^{izx} e^{-x^2} dx = e^{-z^2/4} \int_{-\infty-iz/2}^{\infty-iz/2} e^{-u^2} du.$$

Wenn man nun „wie üblich“ beachtet, daß das Unendliche zuzüglich einer endlichen Größe (in unserem Falle $\pm iz/2$) immer noch Unendlich ist, so wird letzteres Integral zu

$$e^{-z^2/4} \int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi} e^{-z^2/4}.$$

Die letzte Gleichung hat Laplace übrigens nicht mit dem heute üblichen Vorgehen über das Doppelintegral bestritten, sondern durch kompliziertere Betrachtungen (s. [Laplace 1812, 95 f.]). Laplace sah die eben beschriebene Vorgehensweise als heuristisch und unstreng; er beschrieb auch noch alternative Herleitungen für (2) (etwa vermöge einer Differentialgleichung bezüglich der Variablen z [Laplace 1812, 98]). Sein „Umweg“ über das Komplexe war jedoch eine der Problemstellungen, die Cauchy (z.B. [1825]) zu einer Theorie der Integration im Komplexen anregten.

Weiterentwicklung nach Laplace

Die Weiterentwicklung der Stochastik nach Laplace ist durch die folgenden wesentlichen Punkte charakterisierbar (s. [Daston 1988]; [Schneider 1987]; [Fischer 2011]):

- Es ist ein „decline“ der Anwendungen im Bereich individueller menschlicher Entscheidungen, beispielsweise im Rahmen von Gerichtsprozessen feststellbar. Statt dessen gewinnt die stochastische Durchdringung von Massenerscheinungen an wesentlichem Gewicht

- Die in letzterem Bereich wichtigen Grenzwertsätze (bzw. Approximationen bei großen Zahlen) erfahren eine Verfeinerung der analytischen Methoden und Anpassung an sich verändernde analytische Standards

- Die Wahrscheinlichkeitsrechnung wird gar als „Appendix“ der Analysis, insbesondere der Integralrechnung empfunden, etwa bei Dirichlet und insbesondere Chebyshev

- Besonders Summen unabhängiger Zufallsgrößen rücken in den Mittelpunkt der wahrscheinlichkeitstheoretischen Bemühungen.

Die Chebyshev-Markovsche Momententheorie ab ca. 1875 ist ein (besonders prominentes) Beispiel für die analytische Entwicklung im 19. Jahrhundert, von der besonders die Grenzwertsätze der Wahrscheinlichkeitsrechnung betroffen waren. Knapp dargestellt, ist das Grundproblem bezüglich von Momenten das folgende: $f(x)$ sei eine nichtnegative „Dichtefunktion“ mit Träger $[a, b]$. Bekannt seien $m_1 = \int_a^b x f(x) dx$,

$m_2 = \int_a^b x^2 f(x)dx$, ..., $m_n = \int_a^b x^n f(x)dx$, die sogenannten Momente zu der Dichtefunktion. Was kann man über die Verteilungsfunktion $\int_a^x f(t)dt$ aussagen? Chebyshev entwickelte ab 1874 Ungleichungen zur Lösung dieses Problems, und er versuchte auch den zentralen Grenzwertsatz in diesen Problemkreis einzubetten. Es gelang aber erst Chebyshevs Schüler Andrei Markov [1898] zu zeigen, daß die Konvergenz der Momente gegen die der Normalverteilung die Verteilungskonvergenz gegen die Normalverteilung impliziert. Damit gelang ein strenger Beweis des zentralen Grenzwertsatzes [Fischer 2011, Chapt. 4].

Analytische Wahrscheinlichkeitstheorie bis ca. 1940

Nachdem sich die „Post-Weierstraß-Analysis" als allgemeiner Standard im letzten Drittel des 19. Jahrhunderts durchgesetzt hatte, hielt die im wesentlichen auf Laplace zurückgehende „analytische" Ausrichtung der „klassischen" Bestandteile der Wahrscheinlichkeitstheorie an. Bis in die 1920er Jahre bestand ein gewisser Wettbewerb zwischen Momentenmethoden einerseits und Fouriermethoden (im wesentlichen charakteristischen Funktionen) andererseits, der schließlich zugunsten der charakteristischen Funktionen endete. Letztere setzen überhaupt nicht die Existenz irgendwelcher Momente voraus und eignen sich hervorragend zur Charakterisierung von Verteilungstypen, die als Grenzverteilungen von Summen vorkommen, also von stabilen und unbegrenzt teilbaren Verteilungen. Eine erste Übersicht über die vielfältigen Anwendungen charakteristischer Funktionen gab Lévy in seinem Buch [1925]. In den 1930ern kam noch als analytische Methode die Charakterisierung von Verteilungen aufgrund parabolischer Differentialgleichungen hinzu, wie sie etwa in dem Büchlein von Khinchin [1933] dargestellt ist.

Zwei „Highlights" aus der Theorie der charakteristischen Funktionen seien zur Charakterisierung des analytischen Charakters dieser Problemstellungen kurz angesprochen: Die „stetige" Entsprechung der Konvergenz von charakteristischen Funktionen und von Verteilungen und das Zerlegungsproblem für die Normalverteilung:

Konvergenz charakteristischer Funktionen impliziert Verteilungskonvergenz, [Lévy 1922; 1925]: *Sei (V_n) eine Folge von Verteilungsfunktionen und seien für $t \in \mathbb{R}$ die zugehörigen charakteristischen Funktionen durch $\varphi_n(t) := \int_{-\infty}^{\infty} e^{itx} dV_n(x)$ bestimmt. Wenn für alle beschränkten Intervalle I und alle $t \in I$ gilt, daß*

$$\lim_{n\to\infty} \varphi_n(t) = \varphi(t),$$

wobei φ die charakteristische Funktion zu einer Verteilungsfunktion V ist, so gilt $\lim_{n\to\infty} V_n(x) = V(x)$ für jedes $x \in \mathbb{R}$, in dem V stetig ist.
Dieser Konvergenzsatz wurde nochmals von Cramér [1937] erheblich verallgemeinert, indem nur noch gefordert werden mußte, daß es eine beliebige Grenzfunktion

$$\varphi(t) = \lim_{n\to\infty} \varphi_n(t)$$

gibt, die in $t = 0$ stetig ist.

Zerlegungssatz für die Normalverteilung, [Cramér 1936; Lévy 1937]: *Sei φ die charakteristische Funktion einer Normalverteilung und gelte für alle $t \in \mathbb{R}$ daß $\varphi(t) = \varphi_1(t) \cdot \varphi_2(t)$, wobei φ_1 und φ_2 nicht konstante charakteristische Funktionen sind. Dann sind sowohl φ_1 wie auch φ_2 die charakteristischen Funktionen von Normalverteilungen.*

Berücksichtigt man (1), so stellt man fest, daß diese Aussage äquivalent dazu ist, daß jede normalverteilte Zufallsgröße sich nur so in zwei unabhängige Summanden mit nicht-entarteten Verteilungen zerlegen läßt, daß beide Summanden ihrerseits normalverteilt sind. Cramérs Beweis wurde von Lévy noch einmal deutlich vereinfacht.

Bis ca. 1940 konnten die wesentlichen Fragen der Charakterisierung spezifischer Verteilungen (stabiler und – allgemeiner – unbegrenzt teilbarer Verteilungen) und der Konvergenz gegen diese Verteilungen, insbesondere bei Summen unabhängiger und schwach abhängiger Zufallsgrößen, als gelöst angesehen werden (s. [Gnedenko & Kolmogorov 1949]). Diese Probleme spielen natürlich auch heute noch eine wichtige Rolle in Anwendungen und sind daher als Hintergrundwissen unverzichtbar, in der aktuellen Forschung haben sie aber nur noch den Status eines Nischenthemas, beispielsweise, wenn es um solche charakteristische Funktionen geht, die typgleich zur Dichtefunktion der zugrundeliegenden Verteilung sind.

Erste Zusammenfassung

Auch wenn sich die Anforderungen an "mathematische Strenge" in Argumentationen wie auch Begriffsfestlegungen von Laplace bis zur Mitte des 20. Jahrhunderts drastisch geändert haben, so blieben doch in dieser Zeit wesentliche Charakteristika einer analytisch geprägten Wahrscheinlichkeitsrechnung weitgehend gleich: Einerseits bestand ein Inventar an analytischen Methoden (Fouriermethoden, Approximationsmethoden, Lösungen von Differentialgleichungen, ...), die vielfältiger inner- wie außermathematischer Anwendungen fähig sind und durch diese modifiziert und erweitert werden. Darunter fallen auch stochastische Anwendungen, die man grob so beschreiben kann, daß Wahrscheinlichkeiten dafür berechnet werden sollen, daß eine (ein- oder mehrdimensionale) Zufallsgröße innerhalb eines (ein- oder mehrdimensionalen) Intervalls liegt. Die benötigten wahrscheinlichkeitstheoretischen Begriffe sind elementar und überschaubar. Die mathematische Substanz ist analytisch geprägt. Die Leitideen dieser Phase sind Zahl (z.B. als Wahrscheinlichkeiten, Momente) und Funktion (z.B. als Verteilungsfunktionen, charakteristische Funktionen).

Der Weg zur maßtheoretisch geprägten Stochastik

Die reelle Analysis erlebte ab ca. 1870 eine rasante Entwicklung. Bis ca. 1890 etablierte sich die „Weierstraßsche Analysis", also die mit ε und δ hantierende „klassische" Analysis, im Lehr- und Forschungsbetrieb. Diese Entwicklung ging jedoch, zunächst motiviert durch das Problem der „Ausnahmemengen“ bezüglich der Konvergenz von Fourierreihen nahtlos über in weitergehende Untersuchungen, die in die „reelle Funktionentheorie“ und hier insbesondere in die Maß- und Integrationstheorie moderner Prägung mündeten, wie sie gegen 1930 als weitgehend gefestigt angesehen werden konnte (s. z.B. [Pier 1994]). Mit der berühmten Arbeit von Émile Borel [1909] setzte die maßtheoretische Erfassung „nichttrivialer Mengen“ (in diesem Falle im Sinne von starken Gesetzen der großen Zahlen) auch in der Wahrscheinlichkeitstheorie ein. Ab 1920 befaßte sich Norbert Wiener – zunächst noch recht isoliert von der übrigen mathematischen Welt – mit Wahrscheinlichkeiten in Funktionenräumen, entsprechend der Untersuchung von stochastischen Prozessen, wie etwa der Brownschen Bewegung (s. z.B. [Fischer 2011, 333–336]). 1933 erschienen die „Grundbegriffe“ von Kolmogorov, fast schon als zwangsweises Produkt dieser beginnenden Vernetzung der Maßtheorie und neueren Problemen der Wahrscheinlichkeitstheorie. Im Grunde war aber, wie diese kurze Darstellung zeigt, auch die „moderne Wahrscheinlichkeitstheorie“ eine mit der Analysis, wenngleich in ihrer modernen Ausprägung, eng verbundene Theorie.

Bezüglich der Verbindungen zwischen Wahrscheinlichkeitstheorie und der modernen Theorie reeller Funktionen bestehen jedoch erhebliche Unterschiede im Vergleich zu den oben dargelegten Zusammenhängen im Rahmen der „klassischen“ Theorien. Die Maß- und Integrationstheorie stellt das Handwerkszeug zur Verfügung, welches dazu dient, Wahrscheinlichkeiten für „beliebige“, abstrakte Mengen (also nicht nur für endlichdimensionale Intervalle, wie in der klassischen Theorie) zu erfassen. Die in diesem Rahmen entwickelten stochastischen Konzepte und spezifischen Methoden haben – mindestens teilweise – autonomen Charakter. Man denke in diesem Zusammenhang etwa an die Theorie der Martingale, wie sie ab den 1930ern entwickelt wurde. Diese eigenständigen Inhalte der modernen Theorie wirkten freilich weniger auf die allgemeine Analysis zurück, sondern führten ab den 1950ern zu einer spezifischen „stochastischen“ Analysis, in der analytische Fragestellungen mit Methoden der Wahrscheinlichkeitstheorie bearbeitet wurden. Als „Musterbeispiel“ kann hier die Untersuchung potentialtheoretischer Probleme mit Hilfe der Brownschen Bewegung dienen (s. z.B. [Doob 1954]).

Die Leitideen in dieser zweiten, „modernen“ Phase lassen sich durch die Schlagworte Menge (im Sinne komplexer Ereignisse), Maß (im Sinne „abstrakter“ Wahrscheinlichkeitsmaße) und Struktur (etwa im Sinne der betrachteten Maßräume oder der topologischen Räume) beschreiben.

Fazit

In der bisherigen Geschichte der mathematisch geprägten Wahrscheinlichkeitstheorie haben wir es mit zwei unterschiedlichen Beziehungsmustern zwischen Stochastik und Analysis zu tun. In der „klassischen“, durch Laplace wesentlich geprägten Phase bestanden, bei sehr sparsamem stochastischen Begriffsinventar, eine relativ klare Trennung zwischen analytischen Handwerkszeugen einerseits und damit zu bewältigenden Problemen der Wahrscheinlichkeitsrechnung andererseits. Diese Trennung wird auch deutlich in den beiden entsprechend ausgerichteten Büchern der *Théorie analytique des probabilités* von Laplace. Stochastische Probleme, insbesondere solche, die auch allgemein mathematischen Charakter haben, wie Grenzwertsätze, dienen im Verlauf des 19. Jahrhunderts oft als Illustration spezifischer analytischer Techniken, die auch heute unverzichtbar sind, aber nicht mehr im Zentrum der aktuellen Forschung stehen. In der zweiten, „modernen" Phase beeinflussen wesentlich neue Leitideen sowohl die Analysis als auch die Wahrscheinlichkeitstheorie. Letztere erlangt eigenständig mathematische Relevanz, die sich teilweise unabhängig von der analytischen Basis entwickelt und ihrerseits zu neuen, stochastisch geprägten, Untersuchungsobjekten in der Analysis Anlaß gibt.

Natürlich ist man heute geneigt, gleich allen Bestandteilen der Wahrscheinlichkeitstheorie das vollständige Begriffsinventar überzustülpen. Doch dadurch wird möglicherweise nicht nur die Darstellung unnötig verkompliziert, sondern es werden auch die ursächlich analytischen Bezüge – in der einen wie der anderen Ausprägung – vernebelt. Ab und zu täte der Disziplin eine Rückbesinnung auf ihre klassisch analytischen Wurzeln gut.

Literatur

Borel, É. 1909. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del circolo matematico di Palermo* **27**, 247–271.

Cauchy, A.L. 1825. *Sur les intégrales définies prises entre des limites imaginaires.* De Bure, Paris. Wiederabdruck in *Œuvres complètes (2)* **15**, 41–89, Gauthier–Villars, Paris.

Tchebycheff (Chebyshev), P. L. 1887. Sur deux théorèmes relatifs aux probabilités. *Acta mathematica* **14** (1890/91) 305–315. Ursprünglich in russisch in *Zapiski Akademii Nauk* **55** (1887).

Cramér, H. 1936. Über eine Eigenschaft der normalen Verteilungsfunktion. *Mathematische Zeitschrift* **41**, 405–414.

Cramér, H. 1937. *Random Variables and Probability Distributions.* University Press, Cambridge.

Daston, L. 1988. *Classical Probability in the Enlightenment.* University Press, Princeton.

Doob, J.L. 1954. Semimartingales and Subharmonic Functions. *Transactions of the American Mathematical Society* **77**, 86–121.

Gnedenko, B.V. & Kolmogorov, A.N. 1949. *Predelnye raspredelniya dlya summ nezamivisimykh sluchainykh velichin.* Moskva-Leningrad.

Fischer, H. 2011. *A History of the Central Limit Theorem, from Classical to Modern Probability Theory.* Springer, New York.

Khintchine (Khinchin), A. 1933. *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung.* Springer, Ergebnisse der Mathematik und Ihrer Grenzgebiete **2**, Heft 4, Berlin.

Kolmogoroff (Kolmogorov), A. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Ergebnisse der Mathematik und ihrer Grenzgebiete **2**, Heft 3, Berlin.

Laplace, P.-S. 1812. *Théorie analytique des probabilités.* 1. Aufl. 1812, 2. Aufl. 1814, 3. erweiterte Aufl. 1820. Courcier, Paris. Alle Seitenangaben gemäß Wiederabdruck der 3. Aufl. in *Œuvres complètes de Laplace* **VII**, Gauthier–Villars, Paris 1886.

Lévy, P. 1922. Sur la détermination des lois de probabilité par leurs fonctions charactéristiques. *Comptes rendus hebdomadaires de l'Académie des Sciences de Paris* **175**, 854–856.

Lévy, P. 1925. *Calcul des probabilités.* Gauthier–Villars, Paris.

Lévy, P. 1937. *Théorie de l'addition des variables aléatoires*, 1. Aufl. Gauthier–Villars, Collection de monographies sur la théorie des probabilités Fasc. 1, Paris.

Markov, A. A. 1998. Sur les racines de l'équation $e^{x^2}\frac{d^m e^{-x^2}}{dx^m} = 0$. *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg (5)* **9**, 435–446.

Pier, J.-P. 1994. Intégration et mesure 1900–1950. In: Pier, J.-P.: *Development of Mathematics 1900–1950.* Birkhäuser, Basel–Boston–Berlin, S. 517–564.

Schneider, I. 1987. Laplace and Thereafter: The Status of Probability Calculus in the Nineteenth Century. In L. Krüger & L. Daston & M. Heidelberger, M. (Hsg.) *The Probabilistic Revolution*, Bd. 1, Cambridge (Mass.)-London, MIT Press, S. 191–214.

Von Mises, R. 1919. Fundamentalsätze der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift* **4**, 1–97.

Hans Fischer
Katholische Universität Eichstätt-Ingolstadt
Mathematisch-Geographische Fakultät
hans.fischer@ku.de