II.2. Wahrscheinlichkeitsverteilungen und Systeme von Zufallsereignissen

II.2.1. Wahrscheinlichkeitstheorie und Beobachtungsfehler

Nach einer ersten Entwicklungsphase beginnt im 18. Jahrhundert ein ungeheurer Aufschwung in den Anwendungen der Wahrscheinlichkeitstheorie. "Toward the middle of the eigtheenth century probability theory began to be applied more and more in various areas, first and foremost in demography, insurance, estimation of observational errors, organization of lotteries etc." (Maistrov, 1974, S.101) Des weiteren muß hier die Anwendung auf Zeugenaussagen und auf Entscheidungen vor Gericht genannt werden, ein Bereich der Wahrscheinlichkeitstheorie, welchem heute keine Bedeutung mehr zukommt, der jedoch lange Zeit eine wichtige Rolle spielte. Hand in Hand mit der zunehmenden Ausweitung der Anwendungsbereiche wächst auch die Zahl der an der Wahrscheinlichkeitstheorie interessierten Mathematiker:"A progressively larger circle of mathematicians became involved in the development of this scientific discipline." (Maistrov, 1974, S.101)

Wir werden uns nun in diesem Abschnitt auf die Entwicklungen der Wahrscheinlichkeitstheorie im Zusammenhang mit den Anwendungen auf die Beobachtungsfehler konzentrieren, einem Bereich, in dem sich genügend komplizierte neue Probleme als Anstoß für die Weiterentwicklung ergaben, die jedoch gleichzeitig "ideal genug" waren, um die noch relativ jungen wahrscheinlichkeitstheoretischen Methoden entscheidend verbessern zu können. E. Nagel umreißt die Bedeutung der Theorie der Beobachtungsfehler für die Wahrscheinlichkeitstheorie so:"One of the earliest and most successful of these applications (of probability) was to the systematization of measurements and observations in the experimental sciences. Astronomy was the first to employ the theory of probability for this purpose. Justly regarded for a long time as the most exact science of measurements, it nevertheless was patent to everyone that the measurements actually performed did not yield identical numerical values for what was presumably the same magnitude, however carefully gross disturbing factors were eliminated. In consequence, the measurable predictions calculated from astronomical theory were not in precise agreement with the numbers obtained by direct measurement. Given the climate of opinion within which astronomical theory was developed, it was congenial to interpret these fluctuations as deviations or 'errors' from the 'true values' of magnitudes, and to attribute the 'inexactitude' of actual measurements to human failing. Nevertheless, there was a pressing need for techniques to estimate the 'true values' from the actual measurements and to measure the degree to which the latter 'approximate' to the former." (Nagel, 1939, S.9/10)

II.2.1.1 Grundprobleme einer wahrscheinlichkeitstheoretischen Meßtheorie

Probleme des Messens gehören zu den ältesten überhaupt (vgl. Sheynin, 1973). Mit den in Handwerk und Technik wachsenden Anforderungen nach größerer Meßgenauigkeit bei einer zunehmenden Anzahl vielfältigster Meßprobleme insgesamt wurden im 16. und 17. Jahrhundert theoretische Methoden zur Beurteilung von Meßergebnissen immer dringlicher. Ein schon altes Verfahren war das sogenannte "Prinzip des arithmetischen Mittels":"When several 'equally good' measurements of a single quantity were available, the Principle of the Arithmetic Mean stated that the 'best' value to take was their arithmetic mean. The arithmetic mean a of a set of measurements Y_1, Y_2 ... Y_n is the solution of the equation

$$\sum_1^n (Y_i - a) = 0.$$

that is the value determined by the condition of zero sum of residuals. This principle seems to have originated in western Europe sometime in the latter half of the 16th century A.D. ..." (Eisenhart, 1964, S.25)

Dieses Prinzip wurde theoretisch nicht begründet; die Erfahrung hatte seine Wirksamkeit in vielen Situationen erwiesen. Daneben benutzte man jedoch auch vereinzelt andere praktisch orientierte Verfahren; so wurden z.B. sehr stark abweichende Messungen einfach vernachlässigt, oder man begnügte sich auch überhaupt nur mit einer einzigen Messung, die man mit besonderer Vorsicht und Genauigkeit glaubte ausgeführt zu haben (vgl. Adams, 1974, S. 33/34).

Diese stark an praktischen Bedingungen ausgerichteten Überlegungen verdeutlichen, daß das arithmetische Mittel

$$\frac{1}{n} \sum_{j=1}^{n} Y_j$$

der gewonnenen Beobachtungen Y_j in ähnlicher Weise begründet und gehandhabt wurde, wie anfangs wohl auch die relative Häufigkeit: Wenn nur genügend viele Beobachtungen vorgenommen worden sind, so wird sich das arithmetische Mittel bzw. die relative Häufigkeit dem "wahren" Wert bzw. der Wahrscheinlichkeit beliebig genau annähern. (vgl. Sheynin, 1970, S.231/32) Diese, man könnte sagen, "empirische Konvergenz" wurde zunächst nicht weiter begründet. In den ersten Abhandlungen zur Fehlertheorie wurden vor allem Methoden zur Kombination von Beobachtungsfehlern vorgeschlagen, die den praktischen Erfordernissen nach einfach berechenbarer und eindeutiger Lösung genügten. (Hier sind besonders Cotes, Euler, Mayer und Boscovich zu nennen; vgl. Merriman, 1877, S.154ff.) Diese Untersuchungen bezogen bis auf das empirische "Prinzip des arithmetischen Mittels" keine wahrscheinlichkeitstheoretischen Überlegungen ein.

Den Übergang zu den wahrscheinlichkeitstheoretischen Begründungen der Fehlertheorie für die Astronomie samt der sich damit ändernden Auffassung vom Untersuchungsgegenstand beschreibt Pannekoek in seiner Geschichte zur Astronomie folgendermaßen:"In former centuries (before Gauß), the astronomer selected from among his observations those that seemed the best; this made him liable to bias or inclined to select such data as showed a possibly unreal agreement. It could not seem unreasonable that what agreed in the end was deemed the best. ... . In the seventeenth century scientists like Huygens and Picard realized that the average of a number of equivalent

measurements would be better than one of a couple selected from them, and in the eighteenth century this averaging came more and more into use, all the more so since the concept of chance or probability of errors as a quantitative character had gradually become clearer. The notion of 'laws of chance', already applied by Huygens ('plays of luck'), by Jan de Witt, and by Halley ('mortality tables'), acquired its pure theoretical form a century later in the theory of errors developed by Laplace and Legendre and in Gauß's quadratic-exponential law of errors. It afforded to the computers a method of dealing with a series of observed data according to rules which excluded any arbitrariness.
Thereby a new attitude was brought into being, typical of the nineteenth-century scientist towards his material: it was no longer a mass of data from which he selected what he wanted, but it was the protocol of an examination of nature, a document of facts to which he had to defer." (Pannekoek, 1961, S.339/40)

Thomas Simpson war wohl der erste, der die Wahrscheinlichkeitstheorie in die Diskussion um die Beobachtungsfehler einbezog. Ausgehend von der Suche nach einer theoretischen Begründung des bis dahin ausschließlich praktisch benutzten "Prinzips des arithmetischen Mittels" blieb ihm, so kann man feststellen, wohl auch im Grunde nichts anderes übrig, als sich auf wahrscheinlichkeitstheoretische Überlegungen zu stützen. Das Problem besteht nämlich in folgendem: Der <u>Fehler</u> einer Messung bzw. Beobachtung Y wird als die Differenz zwischen <u>wahrem Wert</u> t und der Messung definiert:

$$\varepsilon := Y - t.$$

Da nun der <u>wahre Wert</u> t nie bekannt ist und üblicherweise nie bekannt sein wird, so wird man dementsprechend auch nie den Fehler ε selbst kennen. (Daher wird auch normalerweise genau zwischen dem (unbekannten) wahren Wert und einem angenommenen Schätzwert unterschieden; ein solcher Schätzwert ist z.B. das arithmetische Mittel. Die "Fehler" zwischen Beobachtungen und Schätzwert nennt man Residuen.)

Die einzelnen Fehler ε_i einer Meßreihe Y_i sind somit in ihrer Größe nicht bestimmbar. Eine mathematische Theorie der Beobachtungsfehler ist nicht möglich, solange man sich auf die einzelnen Messungen Y_i als individuelle, fixe Größen bezieht. Die Aufgabe der Fehlertheorie kann in gewisser Weise nur darin bestehen, einen Zusammenhang zwischen diesen einzelnen Beobachtungen bzw. genauer zwischen den den Beobachtungen anhaftenden Fehlern herauszufinden und zu erforschen. Und gerade die Wahrscheinlichkeitstheorie vermochte mit dem sich in diesem Anwendungszusammenhang herausbildenden Begriff der Verteilung eine mögliche Erklärung für solch' einen Zusammenhang zu geben. Fehler wurden unter der wahrscheinlichkeitstheoretischen Perspektive nun nicht mehr als absolute, fixe Größen, sondern als Zufallsereignisse interpretiert. Ein erster Anschluß an die Methoden der Wahrscheinlichkeitstheorie war damit hergestellt. Es konnten nun für die angestrebte Begründung des "Prinzips des arithmetischen Mittels" ähnliche Forderungen aufgestellt werden, wie man sie für die relative Häufigkeit im Theorem von Bernoulli bearbeitet und bewiesen hatte. Gesucht wird die Wahrscheinlichkeit P dafür, daß das arithmetische Mittel der Beobachtungsfehler

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

innerhalb vorgegebener Grenzen liegt:

$$P(|\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i|<\eta) .$$

Ging man jedoch in dem (teilweise) vergleichbaren Fall der relativen Häufigkeit (als einem Schätzwert für die Wahrscheinlichkeit) im Grunde immer von der Kenntnis der Wahrscheinlichkeit aus, so bleiben hier "wahrer Wert" und damit die Größen der Fehler immer unbekannt.

Die Dringlichkeit praktischer Meßprobleme führte jedoch dazu, die hier quasi "fehlende Information" der Kenntnis des "wahren Parameters" (dieser wird ja gerade gesucht) durch eine andere zu ersetzen: Man machte Hypothesen über die Wahrscheinlichkeitsverteilung der unterliegenden Fehler. Die prinzipielle Unmöglichkeit der expliziten Kenntnis der einzelnen Fehlergrößen selbst zwang dazu, sich den Zusammenhang dieser Fehler genauer anzusehen.

Im Verlaufe der Entwicklung der Fehlertheorie hat man dann versucht, die notwendigen Voraussetzungen über die Verteilung der Fehler so weit wie möglich abzuschwächen, indem einerseits die Annahmen über die Fehlerverteilungen so allgemein wie möglich gehalten wurden und man zum anderen anstrebte, ein "universelles" Fehlergesetz aufzufinden.

Des weiteren möchten wir hier schon besonders hervorheben, daß das Problem der genauen Abschätzung der Sicherheit P bei einer endlichen Anzahl n von Versuchen im Zusammenhang mit praktischen Messungen eine grundsätzliche Bedeutung erfährt; hier kommt es darauf an, möglichst genau zu wissen, wieviel Beobachtungen, die ja oft sehr mühsam und zeitaufwendig sind, insgesamt angestellt werden müssen, um mit großer Sicherheit davon ausgehen zu können, daß etwa das arithmetische Mittel oder sonst ein "Schätzer" innerhalb vorgegebener Exaktheitsgrenzen liegt. Auch die anschließend durchzuführenden langen Rechnungen erfordern es, die Anzahl der Beobachtungen so klein wie möglich zu halten.

II.2.1.2 Simpson: Die Verteilung zur Bearbeitung des Zusammenhangs der Beobachtungsfehler

Doch kommen wir nun nach dieser ersten allgemeinen Skizzierung der mit der Anwendung der Wahrscheinlichkeitstheorie auf die Beobachtungsfehler verbundenen Probleme zu Simpson zurück und sehen, wie er die ersten Begründungen des "Prinzips des arithmetischen Mittels" gab, indem er einfache Verteilungen für die unterliegenden Fehler annahm.

Der wichtige erste Schritt einer weitreichenden und fruchtbaren Anwendung wahrscheinlichkeitstheoretischer Methoden in der Fehlertheorie wurde mit der Arbeit Simpsons unternommen:"A Letter to the Right Honourable George Earl of Macclesfield, President of the Royal Society, on the Advantage of taking the Mean of a Number of Observations, in Practical Astronomy" (1755). Zu Beginn seines Briefes beschreibt Simpson zunächst kurz die damals übliche Praxis bei der Behandlung von Beobachtungsfehlern, um dann sein Anliegen zu formulieren:"My Lord, it is well known to your Lordship, that the method practised by astronomers, in order to diminish the errors arising from the imperfections of instruments, and of the organs of sense, by taking the Mean of several observations, has not been so generally received, but that some persons, of considerable note, have been of opinion, and even publickly maintained, that one single observation, taken with due care, was as much to be relied on as the Mean of a great number.
As this appeared to me to be a matter of much importance, I had a strong inclination to try whether, by the application of mathematical principles, it might not recieve some new light; from whence the utility and advantage of the method in practice might appear with a greater degree of evidence. In the prosecution of this design (the result of which I have now the honour to transmit to your Lordship) I have, indeed, been obliged to make use of an hypothesis, or to assume a series of numbers, to express the respective chances for the different errors to which any single observations is subject; which series, to me, seems not ill-adapted ... .
Should not the assumption, which I have made use of, appear to your Lordship so well chosen as some others might be, it will, however, be sufficient to answer the intended purpose: and your Lordship will find, on calculation, that, whatever series is assumed for the chances of the happening of the different errors, the result will turn out greatly in favour of the method now practised, by taking a mean value. But I shall no longer detain your Lordship with general observations,

but proceed to the matter proposed ... ." (Simpson, 1755, S.82-84)

In der dann folgenden ersten Proposition setzt Simpson voraus, daß die Fehler alle einer <u>diskreten uniformen</u> Verteilung gehorchen; dies formuliert er so:"Proposition I. Supposing that the several chances for the different errors that any single observation can admit of, are expressed by the terms of the progression r^{-v} .... r^{-3}, r^{-2}, r^{-1}, r^{o}, r^{1}, r^{2}, r^{3} .... r^{v} (where the exponents denote the quantities and qualities of the particular errors, and the terms themselves the respective chances for their happening):'tis proposed to determine the probability, or odds, that the error, by taking the Mean of a given number (n) of observations, exceeds not a given quantity $(\frac{m}{n})$." (Simpson, 1755, S.84)

In der zweiten Proposition geht er von einer diskreten (gleichschenkligen) Dreiecksverteilung aller Fehler aus: "Proposition II. Supposing the respective chances, for the different errors which any single observation can admit of, to be expressed by the terms of the series $r^{-v} + 2r^{1-v} + 3r^{2-v}$ ...... $+ \overline{v+1}.r^{o}$ ..... $+ 3r^{v-2} + 2r^{v-1} + r^{v}$ (whereof the coefficients, from the middle one (v+1), decrease, both ways, according to the terms of an arithmetical progression): 'tis proposed to determine the probability, or odds, that the error, by taking the Mean of a given number (t) of observations, exceeds not a given quantity $(\frac{m}{t})$." (Simpson, 1755, S.87)

Bei aller Vorläufigkeit und Unausgereiftheit der von Simpson eingeführten "Verteilungen", bringen diese doch erstmals systematisch den Zusammenhang zwischen dem System der Zufallsereignisse bzw. genauer dem System der Zufallsgrößen oder - variablen (bei Simpson die Werte -v, -v+1, ... -1, 0, 1, ... v,) und ihren zugehörigen Wahrscheinlichkeiten zum Ausdruck.

Mit den Techniken der erzeugenden Funktionen, wie sie schon ansatzweise bei Montmort und de Moivre in Glücksspielproblemen benutzt wurden, konnte Simpson in Analogie zum Würfelspiel beide Probleme lösen; im Grunde berechnete er je aus den beiden vorgegebenen Verteilungen, die Verteilung der Summe von n (unabhängigen) Fehlern und daraus dann die Verteilung des arithmetischen Mittels von n Fehlern.

Besonders augenfällig wird dieser Zusammenhang mit dem Würfelspiel, wenn man etwa in der ersten Proposition den Spezialfall $r = 1$ annimmt, den auch Simpson in einer Anmerkung ausführlich diskutiert. Hierbei kann ein Fehler die verschiedenen Werte

$$-v, -v+1, \ldots, -1, 0, 1, \ldots, v-1, v$$

mit der (Gleich-)Wahrscheinlichkeit

$$p = \frac{1}{2v+1}$$

annehmen. Dies läßt sich auch als ein Würfel mit $(2v+1)$ Seiten und den Augenzahlen $-v, \ldots -1, 0, 1, \ldots, v$ interpretieren. Gesucht ist nun die Wahrscheinlichkeit, daß bei n Würfen (oder bei einem Wurf mit n solcher Würfel) das arithmetische Mittel der gefallenen Augenzahlen

$$\frac{1}{n} \sum_{i=1}^{n} a_i \ , \quad a_i \varepsilon \{-v, \ldots -1, 0, 1, \ldots, v\}$$

innerhalb vorgegebener Grenzen liegt. Hieran wird exemplarisch deutlich, wie man die "alten" kombinatorischen Methoden in die Behandlung von Problemen bei Beobachtungsfehlern einbezog und gleichzeitig durch eine neue Sicht, in welcher man den "Würfel" (in verallgemeinerter Form) als "Fehler" auffaßte, für die Berechnung eines unbekannten Meßparameters den Begriff der <u>Verteilung</u> in Ansätzen entwickelte.

Im Anschluß an eine Beispielsrechnung faßt Simpson abschließend seine Arbeit so zusammen: "Upon the whole of which it appears, that the taking of the Mean of a number of observations, greatly diminishes, the chances for all the smaller errors, and cuts off almost all possibility of any great ones: which last consideration, alone, seems sufficient to recommend use of the method, not only to astronomers, but to all others concerned in making of experiments of any kind (to which the above reasoning is equally applicable). And the more observations or experiments there are made, the less will the conclusion be liable to err, provided they admit of being repeated under the same circumstances." (Simpson, 1755, S.92/93)

In einer zweiten Arbeit zu dem gleichen Problem "... to show the advantage by taking the mean of a number of observations", (Simpson, 1757), berechnete er auf der Grundlage einer stetigen dreiecksförmigen Verteilung der Fehler, indem er zum Grenzwert immer "feiner" werdender Fehlerwerte überging, die Verteilung des arithmetischen Mittels von n Beobachtungsfehlern.

Insgesamt zeigt sich in den Arbeiten von Simpson schon deutlich, daß dem Funktionsbegriff in Form der Verteilung bzw. des sogenannten Fehlergesetzes nicht mehr nur wie etwa noch bei de Moivre ausschließlich der Charakter eines technischen Hilfsmittels zukommt; der Verteilungsbegriff wird in der Auseinandersetzung mit den Problemen der Beobachtungsfehler zu einem zentralen Grundbegriff der Wahrscheinlichkeitstheorie.

Diesen wichtigen Anfang bei Simpson beurteilt Eisenhart so: "It should be noted that Simpson did not prove that 'taking of the arithmetic mean' was the best thing to do, but merely that it is advantageous. However, in accomplishing this goal he did something much more important: he took the bold step of regarding errors, not as individual unrelated happenings, but as properties of the measurement process itself and the observer involved. He thus opened the way to a mathematical theory of measurement based on the mathematical theory of probability." (Eisenhart, 1964, S.28/29)

II.2.1.3. Die Vielfalt der ersten Fehlergesetze

Lagrange nahm in seiner Arbeit "Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations, ..." (1774) die Idee der Verteilung auf und behandelte ähnliche Probleme wie Simpson, jedoch viel tiefergehende; zudem benutzt er dazu systematischer die Methoden der erzeugenden Funktionen. Des weiteren finden sich bei ihm schon mehrere verschiedene stetige Verteilungen als mögliche Fehlergesetze. Dabei geht er folgendermaßen vor: Zunächst führt er die "courbe des erreurs" als eine, heute würde man sagen, stückweise lineare Funktion einer diskreten Verteilung ein (vgl. Lagrange, 1774, S.201), um später durch Verfeinerung der Abstände zwischen den einzelnen Fehlern im Grenzfall eine stetige Verteilung zu erhalten. Neben der konstanten und der dreiecksförmigen Verteilung betrachtet Lagrange u.a. auch die folgenden:"He then takes two new examples; in one he supposes that

$$\phi(x) = K\sqrt{c^2-x^2},$$

the errors lying between $-c$ and c; in the other he supposes that

$$\phi(x) = K \cos x,$$

the errors lying between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$". (Todhunter, 1865, S.313; vgl. auch Lagrange, 1774, S.228, S.232).

Auch Daniel Bernoulli, Neffe von Jakob Bernoulli, befaßte sich mit der Fehlertheorie. Interessant ist es hier festzustellen, daß er bei der Behandlung des Grenzwertsatzes von Jakob Bernoulli und de Moivre (1770-1771) die durch die e-Funktion gegebene Funktionskurve selbst nicht betrachtete: "... he pays little attention to the curve itself." (Sheynin, 1977, S. 104). Bei der Behandlung von Problemen der Beobachtungsfehler in seiner Arbeit:"The most probable choice between several discrepant observations and the formation therefrom of the most likely induction" (1777, auf englisch wiederabgedruckt in Pearson/Kendall 1970, S.157-172) stellt er jedoch u.E. besonders schön anschaulich am Beispiel des Schießens auf eine Zielscheibe die Bedeutung der Verteilungskurve heraus.

Einleitend gibt D. Bernoulli ähnlich wie Simpson eine Skizzierung der Problemsituation:"Astronomers as a class are men of the most scrupulous sagacity; it is to them therefore that I choose to propound those doubts that I have sometimes entertained about the universally accepted rule for handling several slightly discrepant observations of the same event. By this rule the observations are added together and the sum divided by the number of observations; the quotient is then accepted as the true value of the required quantity, until better and more certain information is obtained. In this way, if the several observations can be considered as having, as it were, the same weight, the centre of gravity is accepted as the true position of the objects under investigation. This rule agrees with that used in the theory of probability when all errors of observation are considered equally likely.

But is it right to hold that the several observations are of the same weight or moment, or equally prone to any and every error? Are errors of some degrees as easy to make as others of as many minutes? Is there everywhere the same probability?" (Bernoulli, 1970, S.157)

Das Modell des Schießens als eine Veranschaulichung der Beobachtungsfehler beschreibt Bernoulli anschließend:"Let us compare the observer with an archer aiming his arrows at a set mark with all the care that he can muster. Let his mark be a continuous vertical line so that only deviations in a horizontal direction are taken into account; let the line be supposed to be drawn in the middle of a vertical plane erected perpendicular to the axis of vision, and let the whole of the plane on either side be divided into narrow vertical bands of equal width. Now if the arrow be loosed several times, and for each shot the point of impact be examined and its distance from the vertical mark noted on a sheet, though the outcome cannot in the least be exactly predicted, yet there are many assumptions that can reasonably be made and which can be useful to our inquiry, provided all the errors are such as may easily be in one direction as the other, and their outcome is quite uncertain, being decided only as it were by unavoidable chance." (Bernoulli, a.a.O., S.157/58)

Zudem macht er deutlich auf den Unterschied zwischen korrigierbaren und "zufälligen" Fehlern aufmerksam:"In astronomy, likewise, anything which admits of correction a priori is not reckoned as an error. When all those corrections have been made which theory enjoins, any further correction which is necessary in order to reconcile the several slightly discrepant observations which differ slightly from each other is a matter solely for the theory of probability. What in particular happens in the course of observation, ex hypothesi we scarcely know, but this very ignorance will be the refuge to which we are forced to flee when we take our stand on what is not truest but most likely, not certain but most probable (non verissimum sed verisimillimum, non certum sed probabilissimum), as the theory of probability teaches. Whether that is always and everywhere identical with the usually accepted arithmetical mean may reasonably be doubted." (Bernoulli, a.a.O., S.158)

Anhand einer ausführlichen Diskussion seines Beispielmodells stellt Bernoulli dann fünf plausible Annahmen über die Verteilung der Wahrscheinlichkeit bei Beobachtungsfehlern auf, die im wesentlichen konstatieren, daß die Verteilung symmetrisch um den "wahrscheinlichsten Wert" ("centre of forces") ist, dort ein Maximum besitzt und nach beiden Seiten schnell abfällt.

Der mathematische Hauptteil von Bernoullis Arbeit läßt sich so zusammenfassend darstellen:"... Bernoulli proposed (1) a semi-circular law of error

$$f(x) = \frac{2}{\pi a^2} \sqrt{a^2 - x^2}, \quad -a \leqslant x \leqslant + a,$$

where $x = y - \tau$ is the error of y as an observed value of the true value τ and $\pm$ a are limits which an error will never exceed: and (2) advocated maximization of the product

$$f(x_1)\ f(x_2)\ \dots\ f(x_n) = \left(\frac{2}{\pi a^2}\right)^n \prod_{i=1}^{n} \left[a^2 - (y_i - \tau)^2\right]^{1/2}$$

with respect to τ to obtain the 'most likely value' of τ indicated by the observations $y_1, y_2, \ldots, y_n$. Today we would call this 'most likely value', $T = T(y_1, y_2, \ldots, y_n)$, the maximum likelihood estimate of τ corresponding to the law of error $f(x)$." (Eisenhart, 1964, S.29)

Darüber hinaus zeigt dieses Zitat, wie Bernoulli die oben geäußerten Zweifel, ob das arithmetische Mittel immer der "beste" Schätzwert sei, in Form eines Vorläufers der "Maximum-Likelihood" aufgenommen hat.

Abschließend wollen wir kurz auf einige der ersten Arbeiten von Laplace zur Theorie der Beobachtungsfehler eingehen. In dem Aufsatz "Determiner le milieu que l'on doit prendre entre trois observations données d'un même phémomène" (1774) wählte er als Wahrscheinlichkeitsverteilung die Funktion

$$\phi(x) = \frac{m}{2} e^{-m|x|} , \quad -\infty < x < +\infty$$

und als "besten Schätzer" schlug er die Funktion $T(Y_1, Y_2, Y_3)$ der drei Beobachtungen vor, für welche das Mittel von $|T - \tau|$ minimal wird.

"Today we would call his T <u>minimum mean absolute error estimator</u> of τ." (Eisenhart, 1964, S.21)

In einer weiteren Untersuchung "Mémoire sur les Probabilités" (1781) nimmt Laplace u.a. die Funktion

$$\phi(x) = \frac{1}{2a} \ln \frac{a}{|x|} , \quad -a \leqslant x \leqslant +a$$

als Fehlerverteilung; zudem stellt er mehrere mögliche "beste" Schätzer zur Debatte, so etwa das arithmetische Mittel, den Median und den Maximum-Likelihood-Schätzer. (Vgl. zu diesen Arbeiten von Laplace vor allem O.B. Sheynin, 1977a, S.2-10.)

Insgesamt ist an unseren Ausführungen deutlich geworden, welche Vielfalt verschiedenartiger Verteilungsgesetze für die Beobachtungsfehler schon in der allerersten Entstehungsphase des Verteilungsbegriffs benutzt worden sind. Jedoch, und diese überraschende Feststellung trifft auch Adams (1974, S.38/39), taucht bis um 1800 herum in keiner der Arbeiten zur Fehlertheorie die heute übliche Normalverteilung

$$f(x) = \frac{h}{\sqrt{\pi}} \quad e^{-h^2x^2}$$

als eine Fehlerverteilung auf. Überraschend ist dabei, daß die e^{-x^2}-Funktion seit de Moivres Arbeiten weiterhin benutzt und untersucht wurde, um möglichst gute Abschätzungen für erste Verallgemeinerungen des Grenztheorems zu bekommen; dennoch wurde ihre wichtige Bedeutung als eine sehr allgemeine Verteilungsfunktion erst im Zusammenhang mit den Arbeiten zur Fehlertheorie (von Gauß, 1809, und Adrain, 1808) erkannt. Sehr deutlich wird diese Tatsache auch aus den Arbeiten von Laplace, einem der bedeutendsten Wahrscheinlichkeitstheoretiker überhaupt. "... its study in the roles of law of error and probability distribution was preceded by the use of integrals of e^{-x^2} as an approximation tool, an application pioneered by de Moivre in his 1733 Approximatio and further developed by Laplace in a series of works published between 1781 und 1805. Although Laplace studied several functions in the role of law of error, there is no evidence that he considered the normal exponential function as such until after the publication of Gauss' Theoria Motus in 1809." (Adams, 1974, S.38/39)

II.2.2. Die Methode der kleinsten Quadrate und das Fehlergesetz von Gauß

Erstmals in der Literatur erwähnt, findet sich das Exponentialgesetz e^{-x^2} als Fehlerverteilung 1808 in einer Arbeit des Amerikaners Robert Adrain:"Research concerning the probabilities of errors which happen in making observations." Eine wichtige Annahme, die Adrain der Herleitung seiner Verteilung zugrunde legt, besteht darin, daß er die Größe eines Fehlers direkt proportional zu der gemessenen Distanz ansetzt (Adrain bezieht sich auf eine Längenmessung):"... we assume as an evident principle that the most probable distances AB, BC are proportional to the measures Ab, bc; and therefore the errors belonging to AB, BC are proportional to their lengths, ..." (Adrain, 1808, S.93). Daß diese Annahme weit davon entfernt ist, wirklich evident zu sein, kritisiert Glaisher (1872) in einer Übersicht zum Fehlergesetz. Es gelingt Adrain mit seinen Annahmen, die Verteilungsfunktion

$$f(x) = e^{a' + \frac{mx^2}{2a}} \qquad (m < 0)$$

abzuleiten (Adrain, 1808, S.93/94; eine Herleitung findet sich auch in Maistrov, 1974, S.149f, und in Glaisher, 1872, S. 76f).

Insgesamt gesehen hat Adrains Arbeit auf die Entwicklung der Fehlertheorie keinen Einfluß gehabt, da sie überhaupt nicht rezipiert und erst 1871 "wiederentdeckt" wurde (vgl. hierzu Merriman, 1877, S.163). Auch inhaltlich wurde sie nicht hoch geschätzt:"Dr. Adrain's proof, however, seems to me much inferior, both in point of rigour and conclusiveness, to any of the usual investigations, ..." (Glaisher, 1872, S.75).

In der lebhaften Diskussion in Europa um die Probleme der Fehlertheorie war die Herausbildung des heute sogenannten Gaußschen Fehlergesetzes sehr eng mit der Entstehung der "Methode der kleinsten Quadrate" verwoben. Seit Ende des 18. Jahrhunderts wurde diese Methode der kleinsten Quadrate als ein praktisch orientiertes Prinzip bei indirekten Beobachtungen benutzt. (Zur Entstehungsgeschichte vgl. Eisenhart, 1964.) In dem Bestreben dieser Methode über die bloße pragmatische Zweckmäßigkeit hinaus eine theoretische Begründung zu geben, wurde die Normalverteilung herangezogen; die Wechselbeziehung zwischen der Methode der kleinsten Quadrate und der e^{-x^2} - Verteilung war ein außerordentlich wichtiger Anstoß für die Weiterentwicklung des zentralen Grenzwertsatzes.

Der Beginn dieser Entwicklung wurde vor allem durch die Arbeiten von C.F. Gauß (1777-1855), P.S. de Laplace (1749-1827) und A. Legendre (1752-1833) bestimmt. Überschattet wurden diese bedeutsamen Forschungen jedoch durch einen Prioritätenstreit zwischen Legendre und Gauß um die Erstveröffentlichung der Methode der kleinsten Quadrate. In späteren Jahren beurteilte Laplace zutreffend diese Auseinandersetzung so:"M. Legendre eut l'idée simple de considérer la somme des carrés des erreurs des observations, et de la rendre un minimum, ce qui fournit directement autant d'équations finales, qu'il y a d'élements à corriger. Ce savant géomètre est le premier qui ait publié cette méthode; mais on doit à M. Gauss la justice d'observer qu'il avait eu, plusieurs années avant cette publication, la même idée dont il faissait un usage habituel, et qu'il avait communiquée à plusieurs astronomes." (Laplace, 1820, S.353; für eine ausführliche Darstellung dieses Prioritätenstreits vgl. R.L. Plackett, 1972).

II.2.2.1 Legendre: Die Methode der kleinsten Quadrate

In der Vielzahl der Beobachtungsexperimente kann man die Werte gewisser Konstanten

$$x_1, \ldots, x_n$$

nicht direkt durch die Beobachtung selbst bestimmen; diese Konstanten sind oft nur <u>indirekt</u> über die Messung anderer, von x_j funktional abhängiger Werte

$$y_1, \ldots, y_m$$

$$\text{mit} \quad y_i = f_i(x_1, \ldots x_n) \qquad i = 1, \ldots, m$$

zu ermitteln. Aus den y_i versucht man dann, die x_j so zu bestimmen, daß die obigen Gleichungen erfüllt sind. Man muß dazu mindestens $m \geqslant n$ Beobachtungen durchführen; für $m > n$ hat man ein "überbestimmtes" Gleichungssystem vor sich. Mit solch' einem Gleichungssystem hat man es in den allermeisten Fällen gerade bei Beobachtungen zu tun, denn es wird ja angestrebt, durch <u>mehr</u> als unbedingt notwendige Beobachtungen mögliche auftretende Fehler zu eliminieren. Welche "optimale" Lösung soll man jedoch diesem System geben?

Die Fehler ε_i kann man offensichtlich auffassen als:

$$\varepsilon_i = y_i - f_i(x_1, \ldots x_n), \; i = 1, \ldots, m.$$

Für den einfachen Fall linearer Abhängigkeiten (man kann dies jedoch auf den allgemeinen Fall übertragen) schlägt nun Legendre zur Bestimmung eines "besten" Wertes für dieses System das folgende Verfahren vor:"Of all the principles which can be proposed for that purpose, I think there is none more general, more exact, and more easy of application, than that of which we have made use in the preceding researches, and which consists of rendering the sum of the squares of the errors a minimum. By this means there is established among the errors a sort of equilibrium which, preventing the extremes from exerting an undue influence, is very well fitted to reveal that state of the system which most nearly approaches the truth." (Legendre, 1805, auf Englisch abgedruckt in Smith, 1959, S.577).

Ohne jeglichen Begründungsversuch gab Legendre diesem Prinzip praktische Bedeutung dank einfacher und exakter Rechnungen; zudem sind keine wahrscheinlichkeitstheoretischen Betrachtungen in seine Darstellung der "Methode der kleinsten Quadrate" einbezogen worden. (vgl. Eisenhart, 1964, S.25 und Merriman, 1877, S.161ff). "No probability considerations were involved. And his (Legendres') 'discovery' simply marked the culmination of the attempts by Euler, Mayer, Boscovich, Laplace, and others to develop a practicable objective method of adjustment based solely on consideration of residuals." (Eisenhart, 1964, S.27)

II.2.2.2 Gauß: Die Herleitung des "normalen" Fehlergesetzes zur ersten Begründung der Methode der kleinsten Quadrate

Eine erste theoretische Grundlegung der "Methode der kleinsten Quadrate" mit Hilfe der Wahrscheinlichkeitstheorie hat Gauß vorgenommen. "Wir tragen kein Bedenken anzunehmen, daß ihm zuerst das praktische Bedürfnis, also die Naturbeobachtung selbst, zu dieser Epoche machenden Entdeckung geführt hat." (v. Waltershausen, 1965, S.42). So charakterisiert der Freund und Biograph von Gauß dessen grundlegendes Interesse an diesem Problemzusammenhang. "Gauß's first formulation of the method of least squares dates from his student days. In the autumn of 1794 he read ... Lambert's discussion (1765) of the determination of the coefficients of a linear relationship $y = \alpha + \beta x$ from a set of n (> 2) observational points (Y_i, x_i) by the method of averages. In 1795 Gauß conceived the simpler and more objective procedure of taking for α and β the values a and b that minimize the sum of squared residuals

$$\sum_i (Y_i - a - bx_i)^2 ,$$

and worked out the computational details, ... In 1797 he attempted to justify his minimum-sum-of-squared-residuals technique by the theory of probability ..." (Eisenhart, 1968, S.76). So faßt Eisenhart in einer Gauß - Biographie kurz die Entwicklung der Untersuchungen, die diesen zu seinem Begründungsversuch führten, zusammen.

In der Ankündigung zu seiner später folgenden Arbeit: "Theoria Combinationis observationum erroribus minimis obnoxiae" (1821, 1823), die eine zweite Begründung der Methode der kleinsten Quadrate enthält, erinnert Gauß sich an seine ersten frühen Überlegungen zu diesem Problem: "Der Verfasser gegenwärtiger Abhandlung, welcher im Jahr 1797 diese Aufgabe nach den Grundsätzen der Wahrscheinlichkeitsrechnung zuerst untersuchte, fand bald, daß die Ausmittelung der wahrscheinlichsten Werthe der unbekannten Größe unmöglich sei, wenn nicht die Function, die die Wahrscheinlichkeit der Fehler darstellt, bekannt ist. Insofern sie dies aber nicht ist, bleibt nichts übrig, als hypothetisch eine solche Function anzunehmen. Es schien ihm das natürlichste, zuerst den umgekehrten Weg einzuschlagen und die Function zu suchen, die zum Grunde gelegt werden muß, wenn eine allgemein als gut anerkannte Regel für den einfachsten aller Fälle daraus hervorgehen soll, die nemlich, daß das arithmetische Mittel aus mehreren für eine und dieselbe unbekannte Größe durch Beobachtungen von gleicher Zuverlässigkeit gefundenen Werthen als der wahrscheinlichste betrachtet werden müsse. Es ergab sich daraus, daß die Wahrscheinlichkeit eines Fehlers x, einer Exponentialgröße von der Form e^{-hhxx} proportional angenommen werden müsse, und daß dann gerade diejenige Methode, auf die er schon einige Jahre zuvor durch andere Betrachtungen gekommen war, allgemein nothwendig werde. Diese Methode, welche er nachher besonders seit 1801 bei allerlei astronomischen Rechnungen fast täglich anzuwenden Gelegenheit hatte, und auf welche auch Legendre inzwischen gekommen war, ist jetzt unter dem Namen Methode der kleinsten Quadrate im allgemeinen Gebrauch, und ihre Begründung durch die Wahrscheinlichkeitsrechnung, so wie die Bestimmung der Genauigkeit der Resultate selbst, nebst andern damit zusammenhängenden Untersuchungen sind in der Theoria Motus Corporum Coelestium ausführlich entwickelt." (Gauß, Werke Bd. IV, 1873, S.98/99).

Wir wollen nun kurz die Herleitung der normalen Fehlerverteilung, wie Gauß sie gegeben hat, vorstellen. (vgl. hierzu neben Gauß: Aus der Theorie der Bewegung der Himmelskörper, welche in Kegelschnitten die Sonne umlaufen, 2. Buch 3. Abschnitt: Bestimmung der Bahn, die beliebig viele Beobachtungen möglichst genau erfüllt (deutsche Übersetzung der "Theoria Motus ...") in:Gauß, 1964; vor allem Czuber, 1897, S.154f, und 1941, S.297f; Goldstine, 1977, S.213ff; Heyde/Seneta, 1977, S.62ff; Seal, 1970, S.209ff; Sprott, 1978, S.186ff).

Gauß betrachtet μ Funktionen

$V, V', V'', \ldots$ von ν Variablen

$p, q, r, s, \ldots$:

$$V^{(i)}(p,q,r, \ldots) \quad .$$

Die Werte der Funktionen

$$V = M, \quad V' = M', \quad V'' = M'', \ldots$$

seien durch direkte Beobachtungen erhalten worden.
Ist $\mu > \nu$, so ist im allgemeinen das System zur Bestimmung von $p,q,r,s, \ldots$ überbestimmt; hiermit befaßt Gauß sich.
Warum sollte man zur Bestimmung eines "besten Wertes" für dieses System nun die Methode der kleinsten Quadrate wählen?

Zur Beantwortung dieser Frage ist es zunächst notwendig, wie es Gauß schon im obigen Zitat selbst sagte, die Fehlerverteilung zu ermitteln. Als allgemeine Merkmale einer solchen Verteilung nimmt er die folgenden an:"Wir nehmen zuerst an, es sei bei allen Beobachtungen die Sachlage derartig gewesen, daß kein Grund vorhanden ist, die eine für weniger genau als die andere zu erachten, oder daß man gleich große Fehler bei den einzelnen für gleich wahrscheinlich halten muß. Die Wahrscheinlichkeit, welche irgend einem Fehler Δ beizulegen ist, wird daher durch eine Funktion von Δ ausgedrückt, welche wir mit $\phi(\Delta)$ bezeichnen wollen. Wenn man nun auch diese Funktion nicht genau angeben kann, so kann man doch wenigstens versichern, daß ihr Werth ein Maximum für $\Delta = 0$ werden müsse, daß er im allgemeinen für gleiche und entgegengesetzte Werthe von Δ der gleiche sei, und endlich, daß er verschwinde, wenn man für Δ den größten Fehler oder einen noch größeren Werth annimmt. ...

Ferner wird die Wahrscheinlichkeit, daß der Fehler zwischen den Grenzen Δ und $\Delta + d\Delta$ liege, welche um die unendlich kleine Differenz $d\Delta$ von einander abstehen, durch $\phi(\Delta)d\Delta$ auszudrücken sein; hiernach wird allgemein die Wahrscheinlichkeit, daß der Fehler zwischen D und D' liege, durch das von $\Delta = D$ bis $\Delta = D'$ genommene Integral $\int \phi(\Delta)d\Delta$ dargestellt. Nimmt man dieses Integral von dem größten negativen Werthe bis zum größten positiven Werthe von Δ, oder allgemeiner von $\Delta = -\infty$ bis $\Delta = +\infty$, so muß es nothwendig = 1 werden." (Gauß, 1964, S.97/98).

Als "Fehler" lassen sich die Differenzen:

$$M - V,\ M' - V',\ M'' - V'',\ \ldots$$

auffassen, wenn man annimmt, in die $V^{(i)}$ sei ein (später als optimal zu bestimmendes) Wertesystem p,q,r,s, ... eingesetzt worden. Die Wahrscheinlichkeit der Fehler:

$$\varepsilon_i = V^{(i)} - M^{(i)}$$

ist entsprechend der Verteilung:

$$\phi(\varepsilon_i) = \phi(V^{(i)} - M^{(i)}) .$$

"Deshalb wird, wenn man nur alle Beobachtungen als voneinander unabhängige Ereignisse ansehen darf, das Produkt

$$\phi(M-V)\ \phi(M'-V')\ \phi(M''-V'')\ \ldots = \Omega$$

die Erwartung oder die Wahrscheinlichkeit ausdrücken, daß alle diese Werte gleichzeitig aus den Beobachtungen hervorgehen werden." (Gauß, 1964, S.98)

Im folgenden erhält Gauß dann mit Hilfe des Theorems von Bayes, ohne diesen zu erwähnen, folgendes Resultat: Unter der Annahme, daß die Werte von p,q,r,s, ... <u>a priori gleichwahrscheinlich</u> sind, läßt sich <u>nach</u> den durchgeführten Beobachtungen

$$V = M, \quad V' = M', \quad V'' = M'', \ \ldots$$

die Wahrscheinlichkeit P dafür, daß die Werte von p,q,r,s, ... zwischen p und p+dp, q und q+dq, ... liegen <u>a posteriori</u> mit Bayes Theorem bestimmen als:

$$P\Big(p\ \varepsilon\,(p,p+dp),\ q\,\varepsilon\ (q,q+dq),\ \ldots\ \big/\ M=V,\ M'=V',\ \ldots\Big)$$

$$= \frac{\Omega\ dpdq\ \ldots}{\int_{-\infty}^{+\infty} \Omega\ dpdq\ \ldots} = \frac{\phi(\varepsilon_1)\ \ldots\ \phi(\varepsilon_\nu)dpdq\ \ldots}{\int_{-\infty}^{+\infty}\ \phi(\varepsilon_1)\ldots\phi(\varepsilon_\nu)dpdq\ \ldots}\ .$$

(vgl. Gauß, 1964, S. 99/100)

Bis hierher verläuft im wesentlichen die Argumentation von Gauß "direkt"; auf der Grundlage der Fehlerverteilung ϕ wurde mit dem Bayesschen Prinzip die Wahrscheinlichkeit bestimmt, daß die unbekannten Werte p,q,r, ... nach den Beobachtungsexperimenten innerhalb vorgegebener Grenzen liegen.

Um jedoch die Fehlerverteilung ϕ explizit bestimmen zu können, <u>kehrt</u> Gauß, wie angekündigt, die Fragestellung <u>um</u>: Unter der Annahme, daß das arithmetische Mittel der wahrscheinlichste Wert ist, wird die Fehlerverteilung deduziert. Gauß sagt:

"Hieraus wird man also ... die völlig bestimmte Lösung der Aufgabe ableiten können, sobald nur die Natur der Funktion ϕ bekannt ist. Da diese aber a priori nicht definirt werden kann, so wollen wir die Sache von einer anderen Seite angreifen, und nachforschen, auf welcher stillschweigend gleichsam als Grundlage angenommenen Funktion ein landläufiges Princip eigentlich beruht, dessen Vortrefflichkeit allgemein anerkannt ist. Wie ein Axiom pflegt man nämlich die Hypothese zu behandeln, wenn irgend eine Größe durch mehrere unmittelbare, unter gleichen Umständen und mit gleicher Sorgfalt angestellte Beobachtungen bestimmt worden ist, daß alsdann das arithmetische Mittel zwischen allen beobachteten Werthen, wenn auch nicht mit absoluter Strenge, so doch wenigstens sehr nahe den wahrscheinlichsten Werth gebe, so daß es immer das sicherste ist, an diesem festzuhalten." (Gauß, 1964, S.101).

Gauß setzt nun die Funktionen V, V', V", ... gleich dem arithmetischen Mittel der beobachteten Werte M, M', M", ...:

$$V = V' = V'' = \ldots = p = \frac{1}{\mu}(M+M'+M''+\ldots).$$

Die Verteilung ϕ hat für den Wert p ein Maximum, und somit wird auch das Produkt

$$\phi(M-p)\ \phi(M'-p)\ \phi(M''-p)\ \ldots$$

und dessen Logarithmus

$$\log \phi(M-p) + \log \phi(M'-p) + \ldots$$

maximal. Dies bedeutet, daß die Ableitung Null wird:

$$\frac{\phi'(M-p)}{\phi(M-p)} + \frac{\phi'(M'-p)}{\phi(M'-p)} + \ldots = 0 .$$

Hierfür sind die Parameter M, M', M", ... und μ frei wählbar; Gauß setzt

$$M' = M'' = \ldots = M - \mu N$$

und erhält so

$$p = M + (1-\mu)N ;$$

und aus:

$$\frac{\phi'(M-p)}{\phi(M-p)} = \frac{(1-\mu)\ \phi'(M-\mu N-p)}{\phi(M-\mu N-p)}$$

ergibt sich:

$$\frac{\phi'((\mu-1)N)}{(1-\mu)\phi((\mu-1)N)} = \frac{\phi'(-N)}{\phi(-N)} = \text{konstant}$$

für jede natürliche Zahl μ. Daraus kann geschlossen werden, daß

$$\frac{\phi'(\Delta)}{\Delta\phi(\Delta)} = K$$

für alle Δ ist. (Die Übergänge von den natürlichen zu den positiven rationalen und den positiven reellen Zahlen können leicht ergänzt werden; siehe Goldstine, 1977, S.215.)

"Hieraus folgt

$$\log \phi(\Delta) = \frac{1}{2} K \Delta^2 + \text{Const.},$$

oder wenn wir die Basis der hyperbolischen Logarithmen mit e bezeichnen und Const. = $\log K$ setzen,

$$\phi(\Delta) = Ke^{\frac{1}{2}K\Delta^2}.$$

Ferner ist leicht einzusehen, daß K notwendig negativ sein muß, damit Ω wirklich ein Maximum werden könne, weshalb wir

$$\frac{1}{2}K = -h^2$$

setzen; und da nach einem zuerst von Laplace gefundenen, eleganten Lehrsatz das von $\Delta = -\infty$ bis $\Delta = +\infty$ genommene Integral

$$\int e^{-h^2\Delta^2} d\Delta = \frac{\sqrt{\pi}}{h}$$

wird ..., so wird unsere Funktion

$$\phi(\Delta) = \frac{h}{\sqrt{\pi}} e^{-h^2\Delta^2}$$

." (Gauß, 1964, S.101/02)

Soweit die mathematische Herleitung der Gaußschen Fehlerverteilung, die er erstmals 1798 durchgeführt hatte. Als eine wichtige Folgerung kann nun aus dieser Verteilung, wie zu Beginn angestrebt, die Methode der kleinsten Quadrate deduziert werden: "Damit das Produkt

$$\Omega = \left(\frac{h}{\sqrt{\pi}}\right)^{\mu} \cdot e^{-h^2(v^2 + v'^2 + \ldots)}$$

ein Maximum werde, muß offenbar die Summe $v^2+v'^2+ \ldots$ ein Minimum werden. Das wahrscheinlichste Wertsystem der Unbekannten p,q,r,s, ... wird daher dasjenige sein, bei welchem die Quadrate der Differenzen zwischen den beobachteten und berechneten Werten der Funktionen V, V', V", ... die kleinste Summe ergeben, wenn nur bei allen Beobachtungen der gleiche Grad von Genauigkeit vorausgesetzt werden darf.

Dieses Prinzip, welches bei allen Anwendungen der Mathematik auf die Naturwissenschaften sehr häufig von Nutzen ist, muß überall mit demselben Recht als Axiom gelten, mit welchem das arithmetische Mittel zwischen mehreren beobachteten Werten derselben Größe als wahrscheinlichster Wert angenommen wird." (Gauß, 1964, S.103).

Ähnlich hatte R. Adrain nach seiner Herleitung der normalen Fehlerverteilung die Methode der kleinsten Quadrate begründet (vgl. Adrain, 1808, S.95).

Bemerkenswert ist, daß Gauß im Zusammenhang mit den Meßproblemen die wichtige Bedeutung der Größe h, die wir heute in der Form $h^2 = \frac{1}{2\sigma^2}$ als (reziproke) Standardabweichung, also eine Normierung der Skala bezeichnen, gesehen hat: "Übrigens wird man die Constante h als das Maß für die Genauigkeit der Beobachtungen ansehen können." (Gauß, 1964, S.102) Für die Kombination von Beobachtungen mit unterschiedlichem Genauigkeitsmaß werden im weiteren Regeln angegeben; zudem entwickelt Gauß in den folgenden Abschnitten seine Theorie ausführlicher für den Fall, daß die Funktionen V, V', V", ... der Unbekannten p,q,r,s, ... linear sind. Zum Teil schon hier, vor allem jedoch in seiner Arbeit "Aus der Untersuchung über die elliptischen Elemente der Pallas ..." (Gauß, 1964, S.118-128) findet sich fast nebenbei in den Rechnungen das heute sogenannte Gaußsche Eliminationsverfahren; zusammen mit mehreren anderen Lösungsverfahren für Matrizengleichungen (so auch schon Iterationsverfahren, vgl. Heinrich, 1978, S.109-122) rundeten diese die hervorragende Fehlertheorie von Gauß zu einem äußerst eindrucksvollen praktischen Werkzeug ab. Die Wirksamkeit und Bedeutung seiner neuen Methoden wurden in geradezu überragender Weise der gesamten Naturwissenschaftlergemeinde Europas durch die so berühmt gewordene Berechnung der Bahn des verlorengegangenen Planetoiden Ceres deutlich:"Einer der ersten eklatanten Erfolge des rechnenden Astronomen Gauß waren seine Ergebnisse, die es den Astronomen ermöglichten, am 7. Dezember 1801 den nach nur etwa vierzigtätiger Beobachtung wieder verschwundenen, am Neujahrstag 1801 von dem Italiener Piazzi in Palermo entdeckten ersten Planetoiden Ceres wieder aufzufinden. Dahinter steckten als Leistungen einerseits die Entwicklung einer analytischen Methode, die charakteristischen Elemente einer elliptischen Planetenbahn aus nur drei Ortsbestimmungen zu berechnen, andererseits die Anwendung der Fehlerquadratmethode, die es gestattete, die mehr als drei vorhandenen Beobachtungen in systematischer, von Willkür freier Weise zu einem optimal zuverlässigen Ergebnis zu kombinieren." (Heinrich, 1978, S.113)

Gerade diese konkrete Nützlichkeit seiner Theorie in den damals so wichtigen praktischen Problemen, konnte den wissenschaftlichen Ruf von Gauß festigen. "Der große Erfolg von Gauß beeindruckte die astronomische Welt zu Recht ganz außerordentlich; fast umgehend erhielt er ein Berufungsangebot als Leiter der Sternwarte an der Petersburger Akademie der Wissenschaften. Doch Gauß lehnte ab, ..." (Wussing, 1974, S.38).

In der folgenden Zeit ist an dieser hier vorgestellten Herleitung der Fehlerverteilung und der Methode der kleinsten

Quadrate Kritik geübt worden, die sich vor allem auf den "schwachen" Punkt der Voraussetzung des "Prinzips des arithmetischen Mittels" konzentrierte. (vgl. Merriman, 1877, S.165, und Czuber, 1897, 6. Abschnitt). Später hat Gauß selbst seine erste Herleitung aufgegeben und eine zweite vorgestellt, die sich weder auf das"Prinzip des arithmetischen Mittels" noch auf die explizite Kenntnis der Fehlerverteilung stützte (vgl. Gauß, 1964, S.1-91; wir werden weiter unten ausführlicher hierauf zu sprechen kommen).

"Nevertheless, because of its elegance, and, of course, because, generally speaking, the normal distribution did hold (approximately) in astronomy and geodesy, his first approach came to be widely known and even over-popularized so that possibly down even to our time some astronomers and physicists believe that the method of least squares is inalienably connected with the normal distribution, although the real connection of this method with the distribution of the corresponding observational errors is formulated in terms of statistical properties of the least squares' estimators." (Sheynin, 1973, S.123).

Gauß' Nachweiß war der erste, der, wenn auch in recht künstlicher Weise wie oft kritisch angemerkt wurde, die in der Natur so häufig anzutreffende Normalverteilung über das bloße Konstatieren hinaus "plausibel" machte, sie ansatzweise zu rechtfertigen suchte. Dies wurde konkret im Bereich der Beobachtungsfehler auf der Grundlage der Wahrscheinlichkeitstheorie möglich; und in diesem Zusammenhang machte umgekehrt auch die Entwicklung in der Wahrscheinlichkeitstheorie selbst einen wichtigen Schritt voran:"It was in connection with Gauß' first proof of the method of least squares, that the normal distribution came to be recognized as a probability distribution in its own right and as an error distribution." (Adams, 1974, S.42)

Gauß hatte mit seinen Forschungen zur Fehlertheorie Methoden zur Lösung wichtiger praktischer Meßprobleme geliefert und dabei gleichzeitig zentrale Fragestellungen für die Weiterentwicklung grundlegender Begriffe der Wahrscheinlichkeitstheorie eröffnet. Im folgenden ging es nun darum, vor allem die "künstlichen" bzw. zu "starken" Voraussetzungen, denen sich Gauß bei der Herleitung seines Fehlergesetzes bedient hatte, "abzuschwächen" bzw. nach "plausibleren" Annahmen zu suchen. Zunächst, besonders in den folgenden Arbeiten von Laplace und Gauß, blieb die Begründung der "Methode der kleinsten Quadrate" das zentrale und treibende Moment der Forschungen; in deren Verlauf ergab sich dann eine erste Formulierung des (modernen) zentralen Grenzwertsatzes auf dem Hintergrund der Fehlertheorie und anschließend konzentrierten sich die Untersuchungen immer stärker auf die Suche nach (letztlich notwendigen und hinreichenden) Bedingungen für normales "Verhalten".

II.2.2.3 Laplace: Der Zentrale Grenzwertsatz und Verallgemeinerungen der Methode der kleinsten Quadrate

Die vielfältigen theoretischen Arbeiten von Laplace und insbesondere seine Beiträge zur Wahrscheinlichkeitstheorie sind nur auf dem Hintergrund seines großen Interesses an den Anwendungen, vor allem den Problemen der Himmelsmechanik, zu verstehen. "He realized the application of probability theory in the most diverse fields of man's thinking and acting. ... In his Mécanique céleste Laplace advanced probabilistic theories to explain astronomical facts. Like Gauß he applied the theory of errors to astronomical and geodetic operations. He made various applications of his limit theorems. Of course, he studied the usual problems of human statistics, insurances, deaths, marriages. He considered questions concerned with legal matters ... . As soon as Laplace discovered a new method, a new theorem, he investigated its applicability." (Geiringer, 1973, S.613).

Wir wollen und können auch nur aus dieser großen Vielfalt im folgenden beispielhaft kurz die Resultate der beiden zentralen Arbeiten von Laplace zur Problematik der Fehlertheorie und Normalverteilung vorstellen. Diese Arbeiten, die wohl zu den wichtigsten und zugleich auch schwierigsten Untersuchungen, welche Laplace zur Wahrscheinlichkeitstheorie gemacht hat, gehören, beinhalten zum einen eine erste "moderne" Formulierung des zentralen Grenzwertsatzes und zum anderen eine Weiterentwicklung zur Begründung der Methode der kleinsten Quadrate. Obwohl mit vielerlei Mängeln, sowohl in der Herleitung der Ergebnisse als auch der Allgemeinheit der Darstellung, behaftet, sind beide deshalb so überaus bedeutsam, weil hier zum ersten Mal im Zusammenhang mit der Herausarbeitung des Begriffs der Verteilung die "Konvergenz" einer Summe von Zufallsvariablen bzw. genauer einer Summe von Fehlern, gegen die Normalverteilung studiert wird. Neben der Bedeutsamkeit der Normalverteilung ist mit diesem Problem der Konvergenz von Summen (unabhängiger) Zufallsvariablen auch die Frage nach den Voraussetzungen für die Konvergenz gegen die Normalverteilung gestellt, d.h. die Suche nach den spezifischen Charakteristika der den Zufallsvariablen selbst zugrundeliegenden Verteilungen und damit ist die weitere Analyse der fundamentalen Merkmale des unbekannten Untersuchungsgegenstandes in Gang gesetzt.

Bei den angesprochenen Arbeiten handelt es sich einmal um den "Mémoire sur les Approximations des Formules qui sont Fonctions de très grands Nombres et sur leur Application aux Probabilités" samt einem "Supplément" (1809/1810), und zum anderen um den "Mémoire sur les Intégrales définies et leur Application aux Probabilités, et spécialement a la Recherche du Milieu qu'il faut choisir entre les Résultats des Observations" (1810/11); beide Arbeiten sind in leicht veränderter und ergänzter Form später wieder aufgenommen worden im Kapitel IV der "Théorie Analytique des Probabilités" (1812; wir beziehen uns auf Oeuvres complètes, 1820 bzw. 1878-1912, Bd. 7,2. Für eine ausführliche historische Analyse beider Arbeiten vgl.: O.B. Sheynin, 1977a, Abschnitte 4,5 und 6; siehe auch Sheynin, 1976; Adams, 1974, Kap.5; Glaisher, 1872; Todhunter, 1865, S.560ff und Czuber, 1897, S.181ff).

In den "... Approximation des Formules ..." gelangte Laplace in der Auseinandersetzung mit Problemen der Fehlertheorie zu den folgenden wichtigen Ergebnissen. Mit Hilfe der von ihm entwickelten Methoden der charakteristischen Funktionen konnte er für identisch verteilte Zufallsgrößen, die er meist als diskret verteilte Fehler interpretierte, deren Verteilung gegen eine stetige konvergiert, indem die "Abstände" zwischen diesen Fehlern immer kleiner werden, zeigen, daß die Summe dieser Zufallsgrößen gegen die Normalverteilung "konvergiert" (vgl. Sheynin, 1977a, S.10-15).

Für den ersten Fall ergibt sich unter der Annahme, daß die Fehler bzw. Zufallsgrößen:

$$\xi_1, \ldots, \xi_n$$

mit gleicher Wahrscheinlichkeit die diskreten Werte:

$$0, \frac{h}{2m}, \frac{2h}{2m}, \ldots, \frac{(2m-1)h}{2m}, h$$

annehmen, daß für großes n die Wahrscheinlichkeit P dafür, daß die Summe der Fehler

$$[\xi] = \xi_1 + \xi_2 + \ldots + \xi_n$$

bzw. das arithmetische Mittel

$$\frac{[\xi]}{n}$$

innerhalb der vorgegebenen Grenzen

$$\frac{h}{2} - \frac{rh}{2\sqrt{n}} \quad \text{und} \quad \frac{h}{2} + \frac{rh}{2\sqrt{n}}$$

liegt, gegeben wird durch:

$$P = \sqrt{\frac{3}{2\pi}} \int_{-\ell}^{\ell} + e^{-\frac{3r^2}{2}} dr - \sqrt{\frac{3}{2\pi}} \cdot \frac{3r}{20n} (1-r^2)\, e^{-\frac{3r^2}{2}}$$

mit $\pm \ell = \pm \frac{h}{2} r \sqrt{n}$. (vgl. Laplace, 1878-1912, Bd.12, S.314). Werden die Glieder der Größenordnung $\frac{1}{n}$ vernachlässigt, so bekommt man:

$$P = \sqrt{\frac{3}{2\pi}} \int_{-\ell}^{\ell} e^{-\frac{3r^2}{2}} dr.$$

Sheynin hat diese Aussage in die folgende moderne Formulierung übertragen:

$$P\left(-s < \frac{[\xi]}{n} - \frac{h}{2} \leq s\right) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \int_{0}^{s} e^{-\frac{x^2}{2\sigma^2}} dx.$$

(Sheynin, 1977a, S.13).

Hatte Laplace bisher vorausgesetzt, daß die identischen Fehlerverteilungen uniform waren, so läßt er im anschließenden Abschnitt "beliebige" identische Verteilungen zu (auch hier in der Form, daß sich die stetigen Verteilungen aus den diskreten durch einen Grenzübergang ergeben), und erzielt, wenn auch mit gravierenden mathematischen Fehlern behaftet, ein ähnliches Resultat wie oben (Sheynin, 1977, S.15). Im 4. Kapitel, §§ 18 und 19, werden diese Probleme in ähnlicher Form mit Ergänzungen behandelt; in dieser Arbeit studiert Laplace zusätzlich zur Verteilung der Summe $[\xi]$ von n (unabhängigen) Zufallsvariablen zudem die Verteilungen von $[|\xi|]$ und $[\xi^2]$, und kann auch hier zeigen, daß diese für große n "normal" verteilt sind. (vgl. Laplace,1878-1912, Bd.7,2, S.314-318; Sheynin, 1977a, S.25-30).

Insbesondere in der zweiten von uns genannten Arbeit "... sur les Intégrales définies ..." und in den §§ 20 und 21 des Kapitels IV der "Théorie analytique ..." unternimmt Laplace auf der Grundlage und unter Verwendung der bisher gewonnenen Resultate nun den Versuch der "Methode der kleinsten Quadrate" eine noch bessere Begründung zu geben, als dies bei Gauß bisher geschah. Anstelle der bloßen Summe der Fehler betrachtet Laplace hier zunächst lineare Gleichungen mit einer Unbekannten z:

$$p_i z - \phi_i = \varepsilon_i\,, \quad i = 1, \ldots, s$$

mit p_i als bekannter, Fehler-unabhängiger Größe, ϕ_i (direkten) Beobachtungsgrößen und den Fehlern ε_i; es geht hier darum, die Verteilung des arithmetischen Mittels dieser Fehler (bzw. linearen Gleichungen) zu ermitteln, (vgl. Laplace, 1878-1912, Bd. 12, S.388ff). Auch hier gelingt es Laplace u.a., bei relativ schwachen Annahmen über die Verteilungen der Fehler (identisch und gleichverteilt über ein Intervall $[-k, k]$, große Anzahl von Fehlern und von Beobachtungen) zu zeigen, daß das arithmetische Mittel (bzw. hier ein "allgemeineres" Mittel) von n unabhängigen Fehlern normalverteilt ist und daß für $n \to \infty$ der Fehler der Approximation O wird, (vgl. Adams, 1974, S.55ff; Sheynin, 1977a, S.18ff).

"From this it follows directly that the Method of Least Squares as developed by Gauß leads to 'most probable values' (under 'very general conditions') when the number of independent observations involved is large. The Method of Least Squares was, therefore, regarded as firmly established, not merely on grounds of algebraic and arithmetical convenience, but also via the calculus of probabilities - at least when the number of independent observations is large!" (Eisenhart, 1964, S.31)

II.2.2.4 Gauß: Zweite Begründung der Methode der kleinsten Quadrate

Gauß hat im Anschluß an Laplace in gewissem Sinne eine noch allgemeinere Begründung der "Methode der kleinsten Quadrate gegeben (1821/23, in Gauß, 1964, S. 1-91). Unter der allgemeinen Annahme, daß die Verteilungen der unabhängigen Beobachtungsfehler ε_i identisch und symmetrisch sind, konnte Gauß zeigen, daß für ein überbestimmtes lineares Gleichungssytem von Beobachtungsdaten

$$Y = X\beta$$

(mit β als $m \times 1$ Vektor der der abzuschätzenden indirekten Beobachtungsgrößen, Y als $m \times 1$ Vektor der direkten Beobachtungsgrößen und X als bekannter $m \times n$ Matrix; $n \geqslant m$) folgendes gilt:
Der durch die Methode der kleinsten Quadrate gegebene Schätzwert dieses linearen Systems ist identisch mit dem "natürlicherweise" besten Wert, welcher durch das Minimum des Mittels

der quadratischen Fehler gegeben wird.

Im Vergleich mit einem Glücksspiel hatte Gauß für die Beobachtungsfehler angemerkt, daß man es hierbei immer mit Verlusten zu tun habe. Wie sollte dieser Verlust (bzw. der mittlere Verlust) oder auch die Abweichung gemessen werden?
"Den Verlust dem Fehler selbst gleichzusetzen, ist offenbar nicht erlaubt; würden nämlich positive Fehler wie Verluste behandelt, so mußten negative als Gewinne gelten. Die Größe des Verlustes muß vielmehr durch eine solche Funktion des Fehlers ausgedrückt werden, die ihrer Natur nach immer positiv ist. Bei der unendlichen Mannigfaltigkeit derartiger Funktionen scheint die einfachste, welche diese Eigenschaft besitzt, vor den übrigen den Vorzug zu verdienen, und diese ist unstreitig das Quadrat. Somit ergiebt sich das oben aufgestellte Princip.
Laplace hat die Sache zwar auf eine ähnliche Weise betrachtet, er hat aber den immer positiv genommenen Fehler selbst als Maass des Verlustes gewählt. Wenn wir jedoch nicht irren, so ist diese Festsetzung sicherlich nicht weniger willkürlich, als die unsrige ... ." (Gauß, 1964, S.6)

So macht Gauß plausibel, was er unter der "besten" Schätzung des Fehlers verstehen will, die dann in seinem 2. Nachweis der Methode der kleinsten Quadrate eingeht:"Gauß's famous theorem on least squares states: Among all linear error-consistent estimates, the least squares estimate has minimum mean square error." (Sprott, 1978, S.194) Im Vergleich beider Begründungen von Gauß sagt Sprott weiter unten:"The former is more restrictive in requiring a normal distribution, but is wider in scope in allowing any functions $\xi_i = \xi_i(\theta_1, \ldots \theta_k)$ (der Abhängigkeit zwischen direkten und indirekten Beobachtungsgrößen).. The later is more general in allowing any distribution (with mean O and finite σ). But the penalty for this freedom is the limitation to (approximately) linear functions. Gauß emphasized this linearity requirement, saying that the observational errors are assumed sufficiently small that their squares and high powers can be ignored." (Sprott, 1978, S.198 ; für eine vollständige Darstellung der Gaußschen

Herleitung vgl. Goldstine, 1977, S.217-224, siehe auch Heyde/Seneta, 1977, S.64f und Whittaker/Robinson, 1924, Kap. IX).

II.2.2.5 Die große Anzahl der Beobachtungen und ihre Unabhängigkeit als Charakteristika des wahrscheinlichkeitstheoretischen Untersuchungsgegenstandes

Insgesamt ist deutlich geworden, wie sich in der Auseinandersetzung mit den drängenden Problemen der Fehlertheorie und dabei insbesondere der theoretischen Begründung für bisher einfach praktisch bewährte Schätzer der wahren Parameter (arithmetisches Mittel, Methode der kleinsten Quadrate etc.) die Wahrscheinlichkeitstheorie durch den in diesem Zusammenhang sich herausdifferenzierenden fundamentalen Begriff der (Fehler-)Verteilung weiterentwickelte. Das hierbei entstehende praktische Problem der Bestimmung bzw. Approximation der Verteilung für die Summe (bzw. das Mittel etc.) einer großen Anzahl von Fehlern für deren Verteilungen gleichzeitig die Suche nach möglichst "schwachen" bzw. allgemeinen Bedingungen auf dem Forschungsprogramm stand, enthielt im Kern eine neue und für die Entwicklung der Wahrscheinlichkeitstheorie überaus zentrale Fragestellung: Es ging um die allgemeine Analyse der Bedingungen der Konvergenz von Summen unabhängiger Zufallsvariablen auf der Grundlage der mathematisch noch zu präzisierenden Wahrscheinlichkeitsverteilung als einem speziellen Funktionsbegriff.

Gerade die im Rahmen dieser Anwendungen der Wahrscheinlichkeitstheorie auf Meßprobleme angestellten Untersuchungen über die zu treffenden Voraussetzungen bezüglich der Fehler und ihrer Verteilungen werfen ein deutliches Licht auf die dem Anwendungsgegenstand der Wahrscheinlichkeitstheorie eigenen spezifischen Charakteristika. Im Verlaufe der sich verschärfenden Anwendungsprobleme traten immer mehr die folgenden Merkmale hervor: Es handelt sich um ein System mit einer <u>großen</u> Anzahl <u>unabhängiger</u> Fehler bzw. Zufallsgrößen mit identischen (und zunächst sogar symmetrischen oder noch expliziter bestimmten) Verteilungen. Die Ersetzung der Bedingung, daß die Verteilungen identisch (und evtl. auch symmetrisch) sind durch eine schwächere, die jedoch zugleich

"Normalverhalten" des Systems garantiert, war das Anliegen der weiteren wahrscheinlichkeitstheoretischen Behandlung der Fehlertheorie unter der sog. "Hypothese der Elementarfehler", die wir im nächsten Abschnitt behandeln wollen. An dieser Stelle sei hierzu nur kurz angemerkt, daß Poisson, ein Schüler von Laplace, 1837 in der Bearbeitung und im Anschluß an die vorhin vorgestellten Arbeiten von Laplace das Theorem von Jakob Bernoulli auf "Würfel" mit verschiedenen Wahrscheinlichkeitsverteilungen zu seinem "Gesetz der großen Zahlen" verallgemeinert hat; hier wird also explizit auf die identische Verteilung der Zufallsgrößen verzichtet (vgl. Czuber, 1897, S.78ff).

Für das Problem der "Unabhängigkeit" ist es nun außerordentlich interessant festzustellen, daß es im Verlaufe der hier diskutierten Meßprobleme in einer Weise explizit berücksichtigt wurde, wie dies bis dahin überhaupt nicht der Fall gewesen war. Dies geht aus Äußerungen von Laplace und vor allem von Gauß, der eigentlich nur praktisch-orientiertes Interesse an der Wahrscheinlichkeitstheorie hatte, zur Unabhängigkeit, gerade bezogen auf konkrete Meß- oder Beobachtungssituationen, hervor.

O.B. Sheynin merkt hierzu an:"The condition of independence for the ξ_i was usually not discussed in those times; nevertheless Gauß remarked on its necessity ... several times. Even Laplace himself...recalls the independence of events as a necessary condition for the theorem on multiplication of their probabilities, and in one of his last memoirs ... he notes explicitly the independence of the errors of measurements which he investigates. Besides this, Laplace mentions at least twice the condition of independence of observations ... .
'(1) Chaque angle doit être uniquement déterminé par ses mesures, sans égard aux deux autres angles du triangle auquel il appartient; autrement, l'erreur de la somme des trois angles ne serait pas le simple résultat des observations, comme les formules de probabilité le supposent. Cette remarque me parait importante, pour démêler la vérité au milieu des légères

incertitudes que les observations présentent.
(2) Cela (a procedure for adjusting angles measured with a repeating theodolite) serait juste, en effet, si toutes ces valeurs (of the angle) étaient indépendantes. Mais leur dépendance mutuelle fait que les mêmes angles simples sont employés plusieurs fois et d'une manière différente pour chacun d'eux, ce qui doit changer ... la probabilité de la valeur moyenne. C'est un nouvel exemple des illusions auxquelles on est exposé dans ces recherches délicates.' (vgl.Laplace, 1878-1912, Bd.7, S.581-616)." (Sheynin, 1977a, S.11)

Erwähnte Gauß in seiner ersten Arbeit zur Methode der kleinsten Quadrate (1809) nur einmal fast nebenbei das Problem der Unabhängigkeit (vgl. Gauß, 1964, S.98), so geht er hierauf in seiner zweiten Arbeit (1821/23) insgesamt ausführlicher ein. Zunächst erörtert er jedoch, was für die Beschreibung des Gegenstandes der Fehlertheorie unter der wahrscheinlichkeitstheoretischen Perspektive außerordentlich wichtig ist, den Unterschied zwischen regelmäßigen und zufälligen Fehlern:"Beobachtungen, welche sich auf Größenbestimmungen aus der Sinnenwelt beziehen, werden immer, so sorgfältig man auch verfahren mag, größeren oder kleineren Fehlern unterworfen bleiben. Die Fehler der Beobachtungen sind im allgemeinen nicht einfache, sondern entspringen gleichzeitig mehreren Quellen, bei denen zwei Arten genau unterschieden werden müssen. Gewisse Fehlerursachen sind nämlich so beschaffen, daß ihr Einfluß auf jede Beobachtung von veränderlichen Umständen abhängt, die unter sich und mit der Beobachtung selbst in keinem wesentlichen Zusammenhang stehen; die so entstehenden Fehler werden unregelmäßige oder zufällige genannt; und insoweit jene Umstände der Rechnung nicht unterworfen werden können, gilt dieses auch von den Fehlern selbst. Dahin gehören die von der Unvollkommenheit unserer Sinne herrührenden Fehler und solche, die von unregelmäßigen äußeren Ursachen abhängen, z.B. von der durch das Wallen der Luft bewirkten Unsicherheit beim Sehen; auch rechnen wir hierher manche, selbst den besten Instrumenten anhaftende Unvollkommenheiten, z.B. Ungleichförmigkeiten der inneren Wandungen der Libellen, Mangel an absoluter Festigkeit usw. Dagegen haben andere Fehlerursachen bei sämtlichen Beobachtungen derselben Art ihrer Natur nach entweder einen vollkommen constanten Einfluß, oder doch einen solchen, dessen Größe in gesetzmäßig bestimmter Weise allein von Umständen abhängt, welche mit der Beobachtung wesentlich verknüpft sind. Fehler dieser Art werden constante oder regelmäßige genannt."(Gauß, 1964, S.1/2)

Bemerkenswert ist es nun, daß Gauß, bezogen auf konkrete Meßprobleme, besonders die Relativität der Unterscheidung zwischen zufälligen und regelmäßigen Größen hervorhebt; ob nun Größen zufällig oder regelmäßig sind, hängt nicht nur von diesen selbst ab, ist keine "Eigenschaft" der einzelnen Größen, sondern erst im Zusammenhang der Problemstellung, der Erkenntnisziele und der jeweiligen Fragen zu entscheiden. Größen können in gewisser Weise sowohl zufällig als auch regelmäßig sein:"Übrigens ist es klar, daß diese Unterscheidung gewissermaßen nur relativ ist und von dem weiteren oder engeren Sinne abhängt, in welchem man den Begriff von Beobachtungen derselben Art fassen will. So bringen z.B. unregelmäßige Fehler der Theilung der Instrumente bei Winkelmessungen einen constanten Fehler hervor, wenn es sich nur um eine beliebig oft zu wiederholende Beobachtung desselben Winkels handelt, und wenn dabei immer dieselben fehlerhaften Theilstriche benutzt werden; während der aus derselben Quelle stammende Fehler als ein zufälliger angesehen werden kann, wenn man irgendwie Winkel von beliebiger Größe zu messen hat, und eine Tafel, die für jeden Theilstrich den zugehörigen Fehler angiebt, nicht zu Gebote steht." (Gauß, 1964, S.2)

Die Aufgabe der Wahrscheinlichkeitstheorie besteht nun in der Untersuchung der zufälligen Größen:"Die Betrachtung der regelmäßigen Fehler soll von unseren Untersuchungen ausdrücklich ausgeschlossen bleiben. Es ist nämlich Sache des Beobachters, alle Ursachen, welche constante Fehler hervorzubringen vermögen, sorgfältig aufzusuchen und dieselben entweder abzustellen, oder wenigstens ihrer Wirkung und Größe nach auf das genaueste zu erforschen, um ihren Einfluß auf jede einzelne Beobachtung bestimmen und diese von jenem befreien zu können, so daß ein Ergebnis erzielt wird, als ob der Fehler überhaupt nicht vorhanden gewesen wäre. Ganz verschieden hiervon ist aber das Wesen der unregelmäßigen Fehler, welche ihrer Natur nach der Rechnung nicht unterworfen werden können. Diese wird man daher in den Beobachtungen zwar dulden, ihren Einfluß aber auf die aus den Beobachtungen abzuleitenden Größen durch eine geschickte Combination der ersteren möglichst abschwächen müssen. Dieser wichtigen Aufgabe ist die folgende Untersuchung gewidmet." (Gauß, 1964, S.2)

Nach dieser so ausführlichen Diskussion des Begriffs der zufälligen Größe kommt Gauß weiter unten auf die Unabhängigkeit zu sprechen und betont ihre Bedeutsamkeit für den Wahrscheinlichkeitskalkül, wie sie sich in der Multiplikation von Wahrscheinlichkeiten der Fehler zeigt:"Man kann unsere Aufgabe auch auf den Fall ausdehnen, wo die Werte der Größen V, V', V" etc. nicht unmittelbar aus den Beobachtungen gefunden, sondern irgendwie aus Combinationen der Beobachtungen abgeleitet werden, wenn nur die Bestimmungen der einzelnen voneinander unabhängig sind, d.h. auf verschiedenen Beobachtungen beruhen: sobald aber diese Bedingung nicht erfüllt ist, würde die Formel M falsch werden. Wäre z.B. eine oder die andere zur Bestimmung des Werthes von V verwendete Beobachtung auch zur Bestimmung des Werthes von V' benutzt worden, so würden die Fehler e und e' nicht mehr voneinander unabhängig, und der mittlere Werth des Produkts ee' deshalb auch nicht mehr = 0 sein." (Gauß, 1964, S.22)

Die deutlichste Beschreibung der mit der Unabhängigkeit verbundenen Schwierigkeiten gibt Gauß im Jahre 1845 in seiner Arbeit zur "Anwendung der Wahrscheinlichkeitsrechnung auf die Bestimmung der Bilanz für Witwenkassen":"Bei aller Anwendung des Kalküls sowohl auf Gegenstände der Natur als auf sociale Verhältnisse, pflegen die Erfahrungsdata selten in der reinen Gestalt, wie man sie eigentlich braucht, aufzutreten, sondern fast immer mehr oder weniger behaftet mit Störungen oder Schwankungen, die in ihrem Wechsel keiner Regel gehorchen, und man sucht dann, wie jedermann weiß, den daraus entstehenden Nachtheil wenn auch nicht aufzuheben, doch so viel thunlich zu vermindern, daß man aus vielen einzelnen Resultaten das Mittel nimmt. Man rechnet darauf, daß bei einer solchen Benutzung einer großen Zahl von Fällen die zufälligen Schwankungen einander größtentheils compensiren, und legt dann dem Mittelwerthe eine desto größere Zuverlässigkeit bei, je mehr partielle Resultate zugezogen sind. Dieses ist auch im allgemeinen vollkommen richtig, und durch consequente weitere Entwicklung und umsichtige Ausbeutung dieses Princips sind besonders in den Naturwissenschaften nicht selten die belohnendsten Früchte, selbst glänzende Resultate, gewonnen. Allein die Sicherheit des Grundprincips beruhet auf einer wesentlichen Bedingung, die, häufig genug, auch von Gelehrten vom Fach außer Acht gelassen wird, und die darin besteht, daß die an den einzelnen Beobachtungen oder Erfahrungen haftenden regellosen Störungen oder Schwankungen voneinander ganz unabhängig sein müssen. Das Urtheil, ob eine solche Unabhängigkeit vorhanden sei oder nicht, kann zuweilen sehr schwierig und ohne tiefes Eindringen in das Sachverhältniss unmöglich sein, und wenn darüber Zweifel zurückbleiben, so wird auch das den Endresultaten beizulegende Gewicht ein precäres sein." (Gauß, 1873, S.143)

II.2.3 <u>Die Hypothese der Elementarfehler und die Normalverteilung</u>

Bisher hatte ausschließlich der erste Beweis von Gauß zur Methode der kleinsten Quadrate gezeigt, daß (unter bestimmten Annahmen) die Verteilung der Fehler <u>selbst</u> normal ist, während z.B. Laplace "nur" für die "Mittel" von Fehlern die Normalverteilung nachgewiesen hat. Es gab jedoch genügend Erfahrungsmaterial, welches darauf hinwies, daß in der Tat die Beobachtungsfehler selbst normal verteilt sind und nicht erst ihre Summen. Wie konnte man nun dies theoretisch begründen?

Die Antwort auf diese Frage war die sog. "Hypothese der Elementarfehler"; man nahm an, daß sich jeder Fehler aus der Summe einer ungeheuer großen Anzahl von unabhängigen Elementarfehlern zusammensetze. Zu dieser Begründung des Fehlergesetzes merkt Czuber an:"Unter den Auffassungen dieses Gegenstandes steht ohne Zweifel wissenschaftlich am höchsten diejenige, welche von der Annahme ausgeht, jeder Beobachtungsfehler sei das Resultat der Vereinigung einer großen Anzahl sehr kleiner Fehler, welche verschiedenen, von einander unabhängigen Ursachen entspringen. Diese Annahme ist aus der Natur des Gegenstandes hergeholt; ..." (Czuber, 1897, S.163/64).Und in seiner Übersicht zu den verschiedenen Begründungen und Herleitungen des Fehlergesetzes beurteilt Glaisher die Bedeutung der "Hypothese der Elementarfehler"folgendermaßen:"As the result of the examination of all the proofs described in this paper it seems to me that the only sound philosophical basis on which the law of facility $e^{-h^2r^2}$ rests, is the supposition that an actual error is formed by the accumulation of a great number of small errors due to different independent sources, and subject to the arbitrary laws of facility $\phi_1(x)$, $\phi_2(x)$ ...; and as I think it is clear that this is the way in which an error really does arise, the law $e^{-h^2r^2}$, for an individual error, is, in my opinion, proved by Laplace." (Glaisher, 1872, S.120)

Mit dieser letzten Anmerkung bezieht Glaisher sich auf die Arbeit von Laplace über die Normalverteilung einer Summe von ("beliebig") identisch verteilten Zufallsgrößen.

Im Laufe der Ausarbeitung der Hypothese der Elementarfehler ist immer deutlicher der Zusammenhang zum Zentralen Grenzwertsatz, wie ihn Laplace als erster formulierte, sichtbar geworden: Es ist in gewisser Weise gleichwertig, aus der als "beliebig" angenommenen Verteilung (mit gewissen Einschränkungen) der Fehler das Normalverhalten von Summen bzw. der Mittel der Fehler zu deduzieren, oder schon das Normalverhalten eines Fehlers selbst aus der Annahme herzuleiten, daß dieser sich aus einer Vielzahl von ("beliebig" verteilten) Elementarfehlern zusammensetzt. In beiden Fällen handelt es sich, mathematisch gesprochen, um den Nachweis, daß die Summe (unabhängiger, identisch bzw. "beliebig" verteilter) Zufallsgrößen gegen die Normalverteilung konvergiert. Die hierzu wichtige Unterscheidung von Fehlern und Elementarfehlern wurde z.T. erst durch die Verbesserung der Meßinstrumente möglich, die zum einen ein Ausschalten "grober" Fehler erlaubte, jedoch nicht verhindern konnte, daß überhaupt keine Fehler mehr auftreten.

II.2.3.1 Die drei Entwicklungsstadien der Hypothese der Elementarfehler

Kommen wir zur Beschreibung der Entwicklung der Hypothese der Elementarfehler. Czuber teilt sie in die drei folgenden Phasen ein:"Die Entwicklung dieses Gedankens läßt deutlich drei Stadien erkennen, welche in den Voraussetzungen zum Ausdruck kommen,die man über die Wirkungsweise der Fehlerquellen aufgestellt hat. Wird im ersten Stadium an eine bestimmte Größe der Elementarfehler und ein bestimmtes Gesetz in ihrem Auftreten gedacht, so besteht der Fortschritt zum zweiten Stadium darin, daß zwar das Häufigkeitsgesetz für die einzelnen Elementarfehler als verschieden und beliebig angenommen wird, mit der Einschränkung jedoch, daß Fehlerbeträge, die nur im Vorzeichen sich unterscheiden, als gleichwahrscheinlich gelten; im dritten Stadium wird auch diese Einschränkung fallen gelassen." (Czuber, 1897, S.164)

Die Entwicklung ist also darauf konzentriert, möglichst weit die zu machenden Annahmen zu verallgemeinern, d.h. im Grunde die dem (unbekannten) Untersuchungsgegenstand anhaftenden spezifischen und wichtigen Charakteristika herauszupräparieren.

Der erste, welcher die Vorstellung zum Ausdruck bringt, ein Beobachtungsfehler setze sich aus einer Vielzahl von Elementarfehlern zusammen, ist Thomas Young (1819). Die grundlegende Idee formulierte er so:"... the combination of a multitude of independent sources of error, each liable to incessant fluctuation, has a natural tendency, derived from their multiplicity and independence, to diminish the aggregate variation of their joint effect ..." Weiter unten präzisiert er diese Überlegung an einem Beispiel:"This statement may be rendered more intelligible by the simple case of supposing an equal large number of black and white balls to be thrown into a box, and 100 of them to be drawn out either at once or in succession. It may then be demonstrated, as will appear hereafter, from the number of ways in which the respective numbers of each kind of balls may happen to be drawn, that there is 1 chance in $12\frac{1}{2}$ that exactly 50 of each kind may be drawn and an even chance, that there will not be more than 53 of either, though it still remains barely possible that even 100 black balls or 100 white may be drawn in succession." (Young, 1819, S.71/72).

Dieser Vergleich mit der Urne samt der dann vorgenommenen Berechnung der Koeffizienten des Binoms $(1+1)^n$ erinnert stark an das Vorgehen de Moivres; jedoch wie unterschiedlich sind im Vergleich die jeweiligen Sichtweisen auf das Problem: Während de Moivre an der "Messung" der Wahrscheinlichkeit eines "Elementarfehlers", z.B. eines Münzwurfes interessiert ist, so befaßt sich Young, bezogen auf die Probleme der Beobachtungsfehler im Grunde mit der Aufstellung der Wahrscheinlichkeitsverteilung der zusammengesetzten Fehler, bzw. man könnte sagen eines Ereignisses, bestehend aus einer n-fachen Ziehung. Hieran läßt sich der unterschiedliche Gebrauch des Funktionsbegriffs in wahrscheinlichkeitstheoretischen Problemen deutlich machen. Während de Moivre die e-Funktion, wie oben ausführlich dargestellt, ausschließlich in ihren kalkülmäßigen Aspekten zur präzisen Berechnung der Sicherheit P benutzt, erlangt der Funktionsbegriff in der Form der Verteilung erst seine sinnvolle Bedeutung für die Wahrscheinlichkeitstheorie als eine Abbildung zwischen (zusammengesetzten) Fehlern und entsprechenden Wahrscheinlichkeiten.

Aufgenommen wurde die Youngsche Idee von Hagen (1837), welcher dann explizit die Annahme machte, ein Beobachtungsfehler setze sich aus einer großen Anzahl unabhängiger, gleicher sehr kleiner Elementarfehler zusammen, mit der Wahrscheinlichkeit $\frac{1}{2}$ für je positive und negative Abweichungen.

"Die analytische Verfolgung dieses Gedankens, von Encke (1853) mit der wünschenswerten Strenge durchgeführt, liefert für den Gesamtfehler z das Gesetz

$$\phi(z) = \frac{1}{\sqrt{2\pi\varepsilon}\, E}\, e^{-\frac{z^2}{2\varepsilon E}} ,$$

das die Herleitung als ein nur angenähert gültiges deutlich erkennen läßt; ε bedeutet die absolute Größe des Elementar-, E jene des größtmöglichen Gesamtfehlers." (Czuber, 1897, S.164) Angemerkt sei, daß sich unter den vielen weiteren zum ersten Stadium gehörenden Ableitungen der normalen Fehlerverteilungen auch eine Arbeit von Quetelet (1848) befindet, der sich einen großen Namen als einer der Begründer der modernen Statistik gemacht hat; dies weist auf den engen Zusammenhang zwischen Fehlertheorie und Statistik in der ersten Hälfte des 19. Jahrhunderts hin (vgl. H.M. Walker, 1929, Kap. II).

Der Übergang zum zweiten Stadium vollzog sich mit einer Arbeit von Bessel (1838); hierin wurde die Annahme aufgegeben, die Verteilungen der Elementarfehler seien identisch. Bessel war ein bekannter Astronom, zudem in den theoretischen Arbeiten zur Fehlertheorie bewandert; er hatte schon 1818 die "Methode der kleinsten Quadrate" von Gauß empirisch getestet. "A close agreement was found, and this may perhaps be called a practical proof of the principle of Least Squares" (Merriman, 1877, S.170).

Der Arbeit Bessels von 1838 "Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler" spricht Czuber maßgebende Bedeutung für die Entwicklung der Fehlertheorie zu. "Den wertvolleren Teil der Besselschen Untersuchung bildet die Ableitung des Fehlergesetzes aus der Annahme, der Fehler einer Beobachtung entstehe durch das Zusammenwirken einer großen Anzahl unabhängiger Fehlerursachen, deren jede nach irgendeinem Gesetze und zwischen irgendwelchen Grenzen wirkt, jedoch so, daß gleich große positive wie negative Fehler gleich häufig vorkommen und daß die mittleren Quadrate der einzelnen Elementarfehler Größen gleicher Ordnung sind. Die letzte Voraussetzung enthält die Forderung einer gewissen Ebenmäßigkeit in der Wirkungsweise der einzelnen Fehlerquellen, durch welche ausgeschlossen werden soll, daß eine oder die andere die übrigen derart überragt, daß sie auf den Gesamtfehler einen vorwaltenden Einfluß ausübt; bei gut ausgebildeten Messungsmethoden und Meßinstrumenten wird die Voraussetzung erfüllt sein.

Ist $\phi_i(x_i)$ das Gesetz des Elementarfehlers $x_i (i = 1, 2, \ldots n)$, $K_i'' = \int x_i^2 \phi_i(x_i)dx_i$ sein mittleres Quadrat (das Integral über das Gebiet von x_i erstreckt), so findet Bessel unter den in der Hypothese ausgesprochenen Voraussetzungen für den Totalfehler $z = x_1 + x_2 + \ldots + x_n$ das näherungsweise geltende Gesetz:

$$\phi(z) = \frac{1}{\sqrt{2\pi\,[K'']}}\; e^{-\frac{z^2}{2[K'']}}$$

worin sich die Summe $[K'']$ über alle K_i'' ausdehnt."
(Czuber, 1897, S.166)

Bessel könnte somit die relativ starke Voraussetzung unabhängig identisch verteilter Elementarfehler "abschwächen" durch quasi beliebig unabhängig verteilte, jedoch mit der Einschränkung, daß die Verteilungen symmetrisch sind und daß keine der anzutreffenden Fehlerursachen dominiert. Das hier neu auftretende Merkmal der "Dominanz" "schwächt" in gewisser Weise die identische Verteilung "ab". Es stellt, so könnte man sagen, ein Kriterium der "Gleichheit" oder Ebenmäßigkeit der Verteilungen der Elementarfehler dar, so daß unter dieser Annahme die Summe einer sehr großen Anzahl solcher kleiner Fehler näherungsweise normal verteilt ist. Die mathematische Präzisierung dessen, was man unter Dominanz versteht, ist im folgenden zentraler Untersuchungsgegenstand in der Frage nach den Bedingungen für "normales Verhalten" geworden.

Bezogen auf die im Zusammenhang der verschiedensten Messungen auftretenden Probleme verdeutlicht Bessel in seiner Arbeit sehr anschaulich, welche Bedeutung den von ihm getroffenen Annahmen über die Elementarfehler und ihre Verteilungen zukommen und wie man diese auch durch praktische Erfahrungen rechtfertigen kann.

In dem folgenden längeren Zitat wird exemplarisch für die gesamte Theorie der Beobachtungsfehler ausführlich und deutlich die Vielfältigkeit verschiedenartiger unabhängiger Fehler konkret sichtbar gemacht.

"Fälle, in welchen nicht viele von einander unabhängige Ursachen zusammenwirkten, um einen Beobachtungsfehler zu erzeugen, sind wahrscheinlich sehr selten; selbst in sehr einfach erscheinenden Beobachtungsarten können oft zahlreiche Ursachen ihrer Fehler nachgewiesen werden. Um dieses durch ein Beispiel zu erläutern, werde ich annehmen, daß eine Reihe von Entfernungen eines Fixsterns von dem Scheitelpunkte oder Pole, mit einem, nach Reichenbachscher Art eingerichteten Meridiankreise beobachtet sei, und versuchen, die Ursachen der Fehler aufzuzählen, welche sich in der Zusammenstellung ihrer Resultate verrathen. Das Instrument muß zuerst auf den Stern eingestellt werden, und diese Einstellung kann aus verschiedenen Ursachen fehlerhaft werden, nämlich erstens, weil eine Grenze der Kraft des Fernrohrs vorhanden ist, innerhalb welcher seine Richtung willkürlich bleibt; zweitens weil der Punkt des Bildes

des Sterns, den man in die Absehenslinie zu bringen beabsichtigt, innerhalb gewisser Grenzen willkürlich sein kann, welche bei großen und hellen Sternen ohne Zweifel weiter auseinander liegen, als bei kleineren weniger hellen, und woraus hervorgehen kann, daß bei Nacht und bei Tage, oder bei hellerem und weniger hellem Himmel, verschiedene Punkte gewählt werden; drittens weil der Stern sich selten oder nie ruhig, sondern in zitternder, von dem Mangel des Gleichgewichts der Luft herrühender Bewegung zeigt, und also eine, zwischen den äußersten Grenzen dieser Bewegung liegende Wahl getroffen werden muß. Hierzu gesellen sich Fehlerursachen, welche von der Einstellung des Instruments ganz unabhängig sind, z.B. viertens ein Einfluß der Elasticität seines Metalls, welcher, zufälligen äußeren Umständen zufolge, bald diesen, bald jenen Werth erhalten, auch zur Folge haben kann, daß die Richtung des Fernrohrs, in dem Augenblicke des Ablesens der Beobachtung, nicht mehr dieselbe ist, welche sie bei seiner Einstellung war; fünftens eine Unsicherheit der Angabe des Kreises, welche aus kleinen Ungleichheiten der Entfernungen seiner eigenen Theilstriche und der Theilstriche der Nonien hervorgeht, und welcher sich als veränderlicher Fehler äußert, da gewöhnlich, bei jeder Wiederholung der Beobachtung, andere Theilstriche zur Coincidenz gelangen; sechstens die aus der begrenzten Schärfe des optischen Hülfsmittels, wodurch die Ablesungen erlangt werden, hervorgehende Unsicherheit; siebtens die aus dem Umstande hervorgehenden Fehler, daß die Schätzung der Angaben der Nonien nur z.B. bis auf die Hälfte des kleinsten Zwischenraumes von 2", welchen sie angeben, getrieben werden kann, wodurch alle an den vier Nonien dieser Instrumente abgelesenen Beobachtungen, sich immer mit einer vollen, einer viertel, halben oder dreiviertel Secunde, nie aber mit anderen Theilen derselben schließen. Ferner kommen dazu äußere Umstände, z.B. achtens der Einfluß der Körperwärme des Beobachters auf den Kreis oder andere Theile des Apparats; neuntens der Einfluß einer, im allgemeinen vorhandenen, Verschiedenheit der Wärme zwischen dem unteren und oberen Rande des Kreises, welcher Spannungen in seinem Metalle

und Veränderungen seiner Figur erzeugt. Auch veranlaßt zehntens die Voraussetzung, daß die Wasserwage der Alhidade bei jeder Ablesung, sich im nicht beeinträchtigten Zustande des Gleichgewichts befinde, einen zufälligen Fehler; elftens geht ein solcher aus der Annahme hervor, daß das Instrument zwischen zwei miteinander zu vergleichenden Beobachtungen in vollkommen gleichem Zustande geblieben sei, während doch die Bemerkung von Änderungen, welche es in kürzerer oder längerer Zeit erfährt, nicht selten ist. Mit dem sogenannten Beobachtungsfehler vermischt sich auch zwölftens der Einfluß, welchen die fehlerhafte Annahme hat, daß der Zustand der Atmosphäre, so wie Barometer und Thermometer ihn angeben, genau der sei, wonach die Größe der jedesmaligen Strahlenbrechung sich richtet, und dreizehntens der Einfluß kleiner Unvollkommenheiten der Reductionselemente der Beobachtungen. Ich werde vermuthlich in dieser Aufzählung von Ursachen, welche zur Erzeugung eines scheinbaren Beobachtungsfehlers zusammenwirken, mehrere übersehen haben, so wie ich der zufälligen Unachtsamkeit in der Ausführung einzelner Momente der Beobachtungen, nicht vorteilhafter oder unruhiger Beleuchtung der Fäden und der Theilstriche, der Einflüsse der Kälte auf das Instrument usw. nicht habe erwähnen wollen. Immer aber wird durch diese Aufzählung von Fehlerursachen der Zweck erreicht, bemerklich zu machen, daß selbst diese einfache Beobachtungsart einen Gesamtfehler zeigen muß, welcher aus zahlreichen Ursachen entsteht, deren jede von den übrigen unabhängig wirkt."
(Bessel, 1838, S.398-400)

Im Anschluß an die Schilderung der vielen verschiedenen Fehlerursachen beschreibt Bessel nun, in welcher Weise sich die zweite wichtige Annahme, nämlich die der "Gleichmäßigkeit" aller dieser Fehlerursachen in der Problematik der Messungen darstellt:"Es ist das Bestreben des Künstlers, welcher ein Instrument verfertigt, seine einzelnen Theile so anzuordnen, daß sie das, was sie leisten sollen, mit gleichmäßiger Genauigkeit leisten. Es würde unnütz sein, einem Kreise einen großen Halbmesser und bis auf Kleinigkeiten sichere Theilungen zu geben, wenn er nur ein kleines, wenig sichere Einstellungen

gewährendes Fernrohr tragen sollte. Wenn es dagegen in der Absicht liegt, in den ersteren Beziehungen das Äußerste zu leisten, so ist jedes Mal auch die Absicht vorhanden, ein dieser Leistung angemessenes Fernrohr anzuwenden. Auch der Wasserwaage, wodurch der Scheitelpunkt erkannt werden soll, gibt der Künstler gleichmäßige Vollendung, und sein ganzes Nachdenken wendet er an, um alle Theile des Instruments so auszuführen, daß nicht die Mangelhaftigkeit des einen, den Vortheil vernichte, welchen die Vollendung der übrigen hervorbringt. Der Beobachter, der das Instrument anwendet, bestrebt sich gleichfalls, dieser Anwendung eine der Genauigkeit des Instruments gleichmäßige Sicherheit zu geben. Er wird Beobachtungen als ungenügend erkennen, wenn die äußeren Umstände so ungünstig sind, daß sie ihm Zweifel erzeugen, welche er für ungleichmäßig mit der Genaugikeit des Apparats hält. Während er z.B. eine, bis auf einige Secunden gehende Unsicherheit, welche das Zittern der Luft verursacht, für sehr bedeutend hält, wenn er mit einem großen und genauen Instrumente beobachtet, muß sie ihm unbedeutend erscheinen, wenn er ein mit einem schwachen Fernrohre versehehes Instrument von kleinem Halbmesser, welches, auch bei der ruhigsten Luft, viel größere Unsicherheiten übrig lassen würde, anwendet. Es ist nicht die Größe der Unsicherheit, welche diesen Unterschied veranlaßt, sondern nur ihr verschiedenes Verhältnis zu anderen vorhandenen Fehlerursachen.
Aus dieser Darstellung der Beobachtungen im Allgemeinen und der wesentlichen Beschaffenheit eines guten Apparats und einer, ihm angemessenen Beobachtungsreihe, scheint mir hervorzugehen, daß man die beiden Annahmen, unter welchen das im 9^{ten} Art. erlangte Resultat näherungsweise richtig ist, nicht für so selten gerechtfertigt halten darf, als man, ohne genauere Betrachtung der Beobachtungsarten und Apparate, vielleicht geneigt sein mögte. Die erste dieser Annahmen ist, daß viele Ursachen zur Hervorbringung des Beobachtungsfehlers zusammenwirken; die zweite, daß unter den, aus den einzelnen Ursachen hervorgehenden mittleren Fehlern, keiner die übrigen beträchtlich übertreffe. Wenn diese Annahmen erlaubt sind, nähert sich immer die Wahrscheinlichkeit des Ge-

sammtfehlers n einer Beobachtung, der Form:

$$\psi\, n = \frac{1}{m\sqrt{(2\,\pi\,)}} \cdot e^{-\frac{nn}{mm}}$$

d.h. demselben Gesetze, wovon Gauß zuerst gezeigt hat, daß es das von der Vorschrift des arithmetischen Mittels geforderte ist." (Bessel, 1838, S.400-01).

An Beispielen hat Bessel zudem deutlich gemacht, daß von der Normalverteilung abweichende Fehlergesetze zu erwarten sind, wenn eine "die übrigen beträchtlich übertreffende Fehlerursache vorhanden" ist.

So gelangte er in einem speziellen Meßproblem etwa zu einer Verteilung der Form:

$$\phi\,(x) = \frac{1}{\pi} \frac{1}{\sqrt{a^2-x^2}}, \quad -a \leqslant x \leqslant +a ,$$

in welcher die Wahrscheinlichkeit des betrachteten Fehlers x mit seiner Größe zunimmt, anstatt, wie zu erwarten, für kleine Fehler am größten zu sein.

"Ein ..., dem Beispiele ... entsprechender Fall ist mir in meiner eigenen Praxis vorgekommen: ich hatte den Unterschied zweier Längen durch eine Mikormeterschraube, oft wiederholt gemessen, und bemerkte, als ich die Beobachtungen untereinander verglich, daß größere positive oder negative Abweichungen von dem mittleren Resultate der Messungen häufiger vorkamen als kleinere; was sowohl der Hypothese, auf welcher die Methode der kleinsten Quadrate beruhet, als der gewöhnlichen Erfahrung widersprach. Ich konnte nicht zweifeln, daß in diesem Falle ein ganz anderes Gesetz der Wahrscheinlichkeit der Fehler stattfinden müsse, und fand wirklich bei einer hierdurch veranlaßten näheren Prüfung des Apparates, daß die Mikrometerschraube, in dem Umfange jeder ganzen Drehung, sich nicht den Angaben ihrer Trommel proportional fortbewegte, aber in verschiedenen Drehungen wiederkehrende, dem Sinusse des von einem gewissen Anfangspunkte an gewählten Drehungswinkels proportionale Ungleichheiten zeigte. Dieser Fall entsprach also dem Beispiele. Als ich die erkannte Fehlerursache durch Rechnung beseitigte, hatte ich das Vergnügen, meine Messungen in sehr befriedigender Übereinstimmung zu finden; woraus also hervorging, daß der Apparat nur geringe sonstige Fehlerursachen besaß." (Bessel, 1838, S.375).

Die an dieser Stelle ausführlich wiedergegebenen Überlegungen Bessels zur Fehlertheorie machen sehr schön deutlich, welch bewußte Reflexion sich gegenüber dem "Gegenstand" der Fehlertheorie und seinen Bedingungen, infolge der sich verschärfenden praktischen Probleme und wachsenden Anforderungen nach Präzision herausgebildet hat; der Untersuchungsgegenstand ist nicht mehr einfach vorgegeben, es geht darum, genauer seine unter dem jeweiligen Erkenntnisinteresse spezifischen Charakteristika herauszuarbeiten. Und das am Beispiel der Fehlertheorie gewonnene Gegenstandsverständnis, daß es die Wahrscheinlichkeitstheorie mit Systemen einer großen Anzahl von unabhängigen, "beliebig" verteilten, "gleichmäßigen" Elementarfehlern und der Berechnung der Verteilung des Totalfehlers zu tun hat, konnte in der sich hierbei zeigenden "reinen" Form schließlich auch auf andere Anwendungsgebiete, wie z.B. in der Physik oder auch in der sich entwickelnden Statistik übertragen werden.

Erstmals wird 1869 in einer Arbeit von Crofton die bisher immer noch gemachte Annahme, die Verteilungen der Elementarfehler seien symmetrisch, fallen gelassen. Hiermit wurde der Übergang in das von Czuber gekennzeichnete dritte Stadium der "Hypothese der Elementarfehler" vollzogen. Der "Verzicht" auf die Symmetrie hatte indes wiederum eine Präzisierung der Auffassung von "Gleichmäßigkeit" zur Folge.

"My object in this paper is to give the mathematical proof, in its most general form, of the law of single errors of observation, on the hypothesis, that an error in practice arises from the joint operation of a large number of independent sources of error, each of which, did it exist alone, would produce errors of extremely small amount as compared generally with those arising from all the other sources combined." (Crofton, 1869, S.175), so beginnt Crofton seine Untersuchung. Zunächst gibt er auch, ähnlich wie es Bessel getan hat, eine detaillierte Vorstellung von der Vielfalt möglicherweise auftretender Fehler. Interessant ist hier sein Hinweis zur Geschichte der Astronomie, wonach man anfangs nur drei oder vier Quellen grober Fehler in Betracht gezogen hatte: etwa die Lichtbrechung, ein ungenaues Meßinstrument, das bloße Auge bei Fixierung der Beobachtung usw. Die dann immer mehr möglich gewordene fast vollständige Ausschaltung dieser groben Fehler brachte schlagartig eine neue Stufe der Präzision. "But when astronomers, not content with the degree of accuracy they had reached, prosecuted their researches into the remaining sources of error, they found that, not three or four, but a great number of minor sources of error, of nearly coordinate importance, began to reveal themselves, having been till then masked and overshadowed by the graver errors, which had been now approximately removed. It was as if a small number of forest trees had been cut cown, leaving an innumerable growth of shrubs' and brushwood at their feet, remaining to be cleared." (Crofton, 1869, S.177/78) Anschließend zählt Crofton beispielhaft mögliche Fehlerursachen auf, ähnlich denen, die schon Bessel nannte. Erst die Verbesserung der geforderten Meßqualität ließ die Vielfalt einwirkender Elementarfehler sichtbar werden, und rechtfertigte in gewisser Weise den Gebrauch wahrscheinlichkeitstheoretischer Methoden. Neben der großen Anzahl betrachteter Zufallsgrößen bleibt jedoch, so Crofton, weiteres zu beachten:"... thus it is not enough for our purpose to show, could we do so conclusively, that each error in practice is compounded of a large number of smaller errors; we must also show that they are independent, at least for the most part. Thus we may conceive one of the minute errors affecting astronomical magnitude, to be an error in the refraction proceeding from a rise in the general temperature, and another affecting the same observation to be an error of time arising from the expansion of the pendulum through the same cause; now these two minute errors are not independent, and would have to be mathematically combined in quite a different way from two that were independent; and indeed, such a change of temperature would influence the actual error of the observation in other ways also. However, we may at least savely conclude that the hypothesis in question is not a mere arbitrary assumption, but a reasonable and probable account of what does in fact take place in the case of careful and refined observations." (Crofton, 1869, S.179)

Auf dieser Grundlage kommt Crofton zu dem von Czuber so zusammengefaßten Ergebnis:"Das Resultat ist folgendes: Bestimmt $\phi_i(x_i)$ das Gesetz des Elementarfehlers $x_i (i=1,2,\ldots,n)$, sind

$$K_i' = \int x_i \; \phi_i(x_i)dx_i$$

und

$$K_i'' = \int x_i^2 \; \phi_i(x_i)dx_i$$

die mittleren Werte seiner ersten und zweiten Potenz (die Integrale über das Gebiet von x_i ausgedehnt), so befolgt der Totalfehler

$$z = x_1 + x_2 \; \ldots . + x_n$$

näherungsweise das Gesetz

$$\phi(z) = \frac{1}{\sqrt{2\pi([K''] - [K'^2])}} \; e^{-\frac{(z-[K'])^2}{2([K''] - [K'^2])}} \quad ;$$

darin erstrecken sich die durch die eckigen Klammern angedeuteten Summen über alle Elementarfehler." (Czuber, 1897, S.167)

Für die grundlegenden Fragen der Fehlertheorie unter der Hypothese der Elementarfehler war somit im dritten Stadium in gewisser Weise ein Abschluß erreicht worden; es war nun möglich, aus sehr allgemeinen Annahmen über den Gegenstand, die zudem für die Meßproblematik als äußerst plausibel angesehen wurden, die "universelle" Normalverteilung als eine Fehlergesetzmäßigkeit abzuleiten. Aus den verschiedensten praktischen Fragestellungen hatte man darüber hinaus gelernt, wie sich unter verändernden Bedingungen andere Fehlerverteilungen als die vom Typ e^{-x^2} einstellten; vor allem spielte hierbei neben der "Unabhängigkeit" die Annahme der "Gleichmäßigkeit" bzw. "Dominanz" eine wichtige Rolle.

Insgesamt haben die Probleme der Beobachtungsfehler und die Anforderungen nach größerer Meßgenauigkeit zur Herausbildung des Fehlergesetzes und der Entwicklung einer universellen normalen Fehlerverteilung geführt, deren ganz allgemeine Bedingungen man im Verlaufe der Entwicklung der Hypothese der Elementarfehler immer deutlicher erkannt hat. Wie drängend die theoretischen Probleme der Beobachtungsfehler im 19. Jahrhundert waren, mag die Übersicht, welche Merriman zu den Arbeiten über die Fehlertheorie 1877 aufgestellt hat, verdeutlichen; sie enthält mehr als 400 Titel, darunter etwa 70 Bücher und 20 Teile von Büchern. Über die Lösung konkreter Probleme der Fehlertheorie hinaus führten diese Arbeiten dann zu einer recht allgemeinen Erklärung für die zahlreichen Naturerscheinungen, welche einer Normalverteilung unterworfen sind. Glaisher formuliert diese Einsicht so:"We thus have as a consequence of what appears to be a true conception of an error the form $e^{-h^2x^2}$ for the law of facility, and the great accuracy with which the errors in a set of observations made apparently under similar circumstances agree with this law, strongly confirms the hypothesis. It is for this reason, no doubt, that the number in so many statistical tables follow this law; the deviations from the average being due to the combined action of a very great number of causes." (Glaisher, 1872, S.105).

II.2.3.2 Die Wahrscheinlichkeitsverteilung als Grundlage der beginnenden Differenzierung in die "reine" und "angewandte Wahrscheinlichkeitstheorie

Die Anwendung der Wahrscheinlichkeitstheorie auf Probleme der Beobachtungsfehler führte in der Bearbeitung dieser Probleme zu wichtigen Weiterentwicklungen und grundlegenden Begriffen. Vor allem der Begriff des "Fehlergesetzes", der es zunächst erlaubte, sich mit wahrscheinlichkeitstheoretischen Methoden den Beobachtungsfehlern zuzuwenden, erlangte für die Wahrscheinlichkeitstheorie selbst außerordentliche Bedeutung, indem er sich während der Herausarbeitung der allgemeinen Bedingungen des Untersuchungsgegenstandes zum (abstrakten) Begriff der Wahrscheinlichkeitsverteilung herauskristallisierte.

Auf der Grundlage der wichtigen Herausarbeitung eines präzisierten Objektverständnisses, und zwar dem des Systemcharakters des Untersuchungsgegenstandes (unabhängig, "gleichmäßig" verteilte Zufallsvariablen), erlaubte es nun der Begriff der Wahrscheinlichkeitsverteilung, die "alte" kombinatorisch-kalkülhafte Bernoullische Meßvorschrift:

$$P(|p-h_n|<\varepsilon) > 1 - \eta$$

in einer qualitativ neuartigen Weise "einheitlich" zu interpretieren.

Faßt man die relative Häufigkeit h_n als "durchschnittliche" Summe entsprechender Zufallsvariablen x_i auf mit

$$x_i = \begin{cases} 1, & \text{wenn im i-ten Experiment das Ereignis A eintrifft} \\ 0, & \text{sonst} \end{cases} ,$$

$$n \cdot h_n = S_n = x_1 + \ldots + x_n ,$$

und sei der Erwartungswert von S_n

$$E(S_n) = n \cdot p , \quad \text{und die Streuung}$$

$$\sigma(S_n) = \sqrt{n \cdot pq} ,$$

so läßt sich Bernoullis Theorem "übersetzen" in:

$$\lim_{n\to\infty} P(-\varepsilon < \frac{S_n - E(S_n)}{\sigma(S_n)} < +\varepsilon)$$

$$= \lim_{n\to\infty} P(|\frac{S_n - np}{\sqrt{n \cdot pq}}| < \varepsilon)$$

$$= \lim_{n\to\infty} P\left(\left|\frac{1}{n} S_n - p\right| < \varepsilon\sqrt{\frac{pq}{n}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{+\varepsilon} e^{-\frac{x^2}{2}} dx$$

Die Sicherheit P und die Exaktheit ε lassen sich anhand der Wahrscheinlichkeitsverteilung gleichzeitig einheitlich berücksichtigen; d.h. die Sicherheit P ist (approximativ) gegeben durch das Integral über die Wahrscheinlichkeitsverteilung von der unteren bis zur oberen Exaktheitsgrenze ε und zwar um den (transformierten) Erwartungswert herum. Zudem läßt sich nun aus der Gestalt der Wahrscheinlichkeitsverteilung selbst ablesen, ob man zugleich eine große Sicherheit P bei kleinen Exaktheitsgrenzen ε erwarten kann. Kurz, die Wahrscheinlichkeitsverteilung ist eine einheitliche Beschreibung des betrachteten Systems, welche es erlaubt, einzelne, das System kennzeichnende Paramter ("wahrscheinlichster Wert", etc.) in angemessener Form darzustellen, indem etwa solche Parameter zum Maximum der Verteilung werden.

Diese Bedeutung und wichtige Rolle des Begriffs der Wahrscheinlichkeitsverteilung als einer mathematisch-wahrscheinlichkeitstheoretischen Beschreibung des betrachteten Zufallssystems, einer Darstellung, die alle wichtigen Strukturmerkmale einheitlich aufeinander bezieht und zusammenfaßt, erlaubte eine sich im folgenden immer stärker durchsetzende "Arbeitsteilung" innerhalb der Wahrscheinlichkeitstheorie, in einerseits die anwendungsorientierte Statistik und andererseits die "reine" Wahrscheinlichkeitstheorie der Petersburger Schule. Der mathematische Funktionsbegriff in Form der Wahrscheinlichkeitsverteilung war gleichzeitig Forschungsgegenstand der reinen Wahrscheinlichkeitstheorie und Analyseinstrument der modernen Statistik.

So bearbeiteten die Mathematiker der Petersburger wahrscheinlichkeitstheoretischen Schule vor allem im Anschluß an die Arbeiten von Laplace und zur Fehlertheorie die mit dem Zentralen Grenzwertsatz verbundenen Probleme der "Konvergenzgeschwindigkeit" und der notwendigen und hinreichenden Bedingungen für "normales Verhalten" auf einer mathematisch exakten Grundlage. Die Untersuchungen Tschebyscheffs, Markovs und Lyapunovs (vgl. A.N. Kolmogoroff, 1947; Adams, 1974, Kap. 7 + 8; Maistrov, 1974, S.188-224) erbrachten nach mehreren Verbesserungen der Laplaceschen Arbeiten erste, allgemeine Beweise des Zentralen Grenzwertsatzes und führten schließlich 1935 in Form der sog. Lindeberg-Bedingung zu der lange gesuchten notwendigen und hinreichenden Bedingung für normales Verhalten" (vgl. W. Feller, 1935 und 1945, S.818; H. Cramer, 1976, S.522ff; Maistrov, 1974, S.223/24) Diese Bedingung drückt in allgemeinster Form die Forderung nach der "Gleichmäßigkeit" der Vielzahl der betrachteten Zufallsvariablen aus.

Die wichtige zweite Grundannahme, nämlich die der Unabhängigkeit, konnte zudem von Markov 1907 in seiner Arbeit "Extension de la loi des grands nombres aux événements dépendants les uns les autres" erstmals verallgemeinert werden; ein weiterer wichtiger Schritt zu allgemeineren wahrscheinlichkeitstheoretischen Grenzwertsätzen war getan.

Die erste stürmische Entwicklung in der frühen Statistik um 1830 fand im Anschluß an den in der Fehlertheorie entwickelten Begriff der Wahrscheinlichkeitsverteilung statt (vgl. Westergaard, 1932, Kap. XIII und Walker, 1929, Kap. II). Dieser Begriff ermöglichte über die bisher fast ausschließlich vorgenommene bloße Sammlung von Daten hinaus eine angemessene Charakterisierung und Strukturierung solcher Datenmassen mit Hilfe von speziellen Verteilungen und den solche Verteilung auszeichnenden einzelnen Parametern wie beispielsweise der Varianz und der Standardabweichung (vgl. Walker, 1929, Kap. IIIff).

Auch die ersten spezifisch statistischen Analyseverfahren wie das von Galton und Pearson entwickelte Korrelationsverfahren, benutzten den Verteilungsbegriff und hatten ihren Ursprung z.T. in den Methoden und Begriffen der Fehlertheorie.

Kurz, mit dem sich aus dem Fehlergesetz herausbildenden Begriff der Wahrscheinlichkeitsverteilung erhält die Wahrscheinlichkeitstheorie einen fundamentalen Begriff, welcher in einheitlicher

Weise den Grundzusammenhang von Sicherheit und Exaktheit ausdrückt bzw. allgemeiner gesagt, das statistische Verhalten eines Zufallssystems charakterisiert, und somit die Trennung in theoretische und anwendungsorientierte Bereiche für die Wahrscheinlichkeitstheorie langsam in Gang setzen konnte.