Wir wollen richtige Fehler !

Ein Statistik - Feuilleton zur Lektüre am Feierabend

J. Stiewe
Kirchhoff - Institut für Physik, Universität Heidelberg
E-Mail: stiewe@kip.uni-heidelberg.de

I Eine etwas längliche, aber notwendige Einleitung

Fehlerrechnung hat nur einen begrenzten Sex - Appeal. Dies gilt nicht nur für die Physik, sondern auch für alle Adepten der sogenannten exakten Naturwissenschaften. Auf das "exakt" kommen wir noch zurück.
Aber das ist ungerecht. Wer einmal an einem wissenschaftlichen Projekt mitgearbeitet hat, der weiß, daß sich die Gelehrten meistens heftiger um die Größe ihres Meßfehlers streiten als um den Wert der weltbewegenden Zahl, die sie veröffentlichen wollen.
Warum ist das so? Zunächst: Wer eine bestimmte Größe messen will, etwa - um ein simples Beispiel zu wählen - die Entfernung zwischen zwei Orten, der wird den "wahren" Wert dieser Entfernung nie ergründen. Denn schon bei der zweiten Messung wird er feststellen, daß der Wert von der ersten Messung abweicht - vorausgesetzt, sein Meßgerät ist empfindlich genug. Wer Entfernungen schlicht in "Tagesreisen" mißt, hat diese Probleme natürlich nicht.
Die gleiche Erfahrung wird man machen, wenn ein zweiter Experimentator die Messung durchführt: Auch er wird eine - im günstigen Fall geringe - Abweichung von der Messung des Kollegen feststellen, und zwar auch dann, wenn er dessen Meßgerät benutzt. (Wir wollen für den Augenblick ausschließen, daß eines der Geräte schlicht fehlerhaft arbeitet.)
Was also kann man nach mehreren Messungen guten Gewissens behaupten? Man kann eine Vorschrift konstruieren, die auf die "beste Schätzung" des wahren Wertes führt. Wir ahnen, daß dies im einfachsten Fall der Mittelwert (das arithmetische Mittel der Einzelmessungen) sein wird. Wir werden sehen, daß man das begründen kann.

I.1 Vertrauen hat Grenzen

Man kann aber auch von einem anderen Standpunkt an die Sache herangehen: Wir fragen nicht nach der besten Schätzung, also dem Wert, der dem unbekannten "wahren" Wert unserer Meinung nach am nächsten kommt, sondern nach dem **Intervall**, in dem der wahre Wert mit einer bestimmten Wahrscheinlichkeit, z.B. 95 % oder 99 %, liegt. Wir brauchen also zwei Zahlen a und b, so daß nach unseren Forschungen der wahre Wert mit einer Wahrscheinlichkeit von (z.B.) 95 % zwischen diesen beiden Zahlen liegt:

$$a < \text{wahrer Wert} < b \quad (95\ \%\ CL).$$

Das "CL" steht für "Confidence Level" ("Vertrauensniveau") und gibt eben die Wahrscheinlichkeit an, mit der der wahre Wert in diesem Intervall zu finden wäre, wenn man ihn denn finden könnte. Natürlich möchte jeder Experimentator sein Ergebnis mit einem möglichst schmalen Vetrauensbereich publizieren. Ein solcher Vertrauensbereich wird im Volksmund "Fehler" genannt. Das ist kein sehr glücklicher Terminus, da das Wort "Fehler" suggeriert, daß man etwas falsch gemacht hat. Aber der "Fehler" ist auch dann endlich, wenn man garantiert alles richtig macht. Man sagt deshalb auch oft "Unsicherheit" und vermeidet den "Fehler". (Übrigens: Diesen Fehler nennen die Angelsachsen "error" (oder "uncertainty"), aber nicht "mistake".)
Darüber, wie die Fehler einer Messung zu bestimmen sind, wurden viele Bücher und Artikel verfaßt. Wir werden am Schluß einige von ihnen nennen. Man braucht aber ein Minimum an Kenntnissen von der Kunst der "Statistik", also der Wissenschaft, die untersucht, an welche Regeln sich der Zufall hält. Dort geht es um Wahrscheinlichkeitsdichten, ihre Darstellung, und die Parameter, mit denen man sie beschreibt. Wir werden die wichtigsten kennen (und lieben!) lernen.

I.2 Richtige Fehler sind wichtig, z.B. für den Nobelpreis!

Zum Schluß dieser Einleitung noch ein anschauliches Beispiel, das zeigen soll, wie wichtig die Bestimmung des "richtigen" Fehlers ist: Nehmen wir an, drei Wissenschaftler wollen unabhängig voneinander eine wichtige Naturkonstante messen, für deren Veröffentlichung es internationalen Ruhm zu gewinnen gibt. Der erste zieht seine Messung ordentlich und nach bestem Wissen und Gewissen

durch und veröffentlicht sein Resultat. Der zweite macht es ebenso; natürlich hat er ein etwas anderes Resultat bekommen, aber sein "Vertrauensbereich" und der des ersten überlappen sich. Der dritte hat ein Resultat, das von denen der beiden Kollegen abweicht. Er hat aber auch - Hochmut kommt vor dem Fall! - seinen Fehler gnadenlos unterschätzt und gibt ein so schmales Intervall an, daß es mit keinem der beiden anderen überlappt. Folge: Diese Messung nimmt niemand ernst. Wohlgemerkt: Nicht die wichtige Naturkonstante wurde "falsch" gemessen, sondern ihre Unsicherheit wurde falsch bestimmt! Denn durch die Angabe eines winzigen Fehlers behauptet man implizit, daß die Messungen anderer schlicht inkompatibel sind: Ein Fehlerbalken ist eben eine harte Aussage darüber, daß der Experimentator den "Wahren Wert" mit z.B. 68 % ("Gaußscher Fehler", darauf kommen wir noch) Wahrscheinlichkeit innerhalb seiner Fehlergrenzen wähnt.
Ach ja: Da waren noch die "exakten" Naturwissenschaften. Wie paßt das "exakt" zu den allgegenwärtigen, unvermeidlichen Fehlern? Es ist eben **kein** Widerspruch; denn gerade die korrekte (nicht immer einfache!) Bestimmung von Meßunsicherheiten, die nicht ohne Grund "Confidence Levels" heißen, gibt dem Wissenschaftler in Süd - Alaska die Möglichkeit, sein Ergebnis mit dem des Kollegen aus Ost - Japan zu vergleichen: Sind unsere Ergebnisse "innerhalb der Fehler" kompatibel? Warum hat der Kollege die geringere Unsicherheit? Was muß ich tun, um meinen Fehler zu verkleinern? Etc etc.. Kurzum: Die exakten Naturwissenschaften heißen exakt, weil sie genau angeben, wie ungenau sie sind.

II Meßreihen und ihre Parameter

Die meisten von Ihnen werden die Begriffe "Mittelwert", "Varianz" und "Standardabweichung" schon einmal gehört haben, auch wenn Ihnen deren Definitionen gerade nicht gegenwärtig sind. Für die echten Fans: Es gibt noch weitere interessante Parameter wie "Schiefe" ("skewness") und "Wölbung" ("kurtosis"), auf die wir aber nicht eingehen werden. Alle diese Parameter dienen zur Charakterisierung der Verteilung von "Wahrscheinlichkeitsdichten". Dies sind eben jene statistischen Verteilungen, denen man entnehmen kann, wie wahrscheinlich es (z.B.) ist, älter als 101 Jahre zu werden oder mehr als eine Million Euro im Jahr zu verdienen. Die berühmteste Verteilung einer Wahrscheinlichkeitsdichte ("probability density") ist natürlich die Gaußverteilung, die auch "Normalverteilung" genannt wird.
Wir wollen uns diesen schrecklichen Dingen aber von einer ganz einfachen Seite nähern: Wir wollen die Zeit messen, die ein Läufer für eine bestimmte Strecke, z.B. 1000 m, braucht. Dazu brauchen wir (von den Startpistolen etc abgesehen) natürlich (mindestens) eine Uhr. Wir wollen annehmen, daß dies eine "analoge" Uhr ist, also eine mit Zeigern, und daß die Ablesegenauigkeit begrenzt bis mäßig ist. Und da wir hier ein "Gedankenexperiment" (englisch: "gedanken experiment") durchführen, soll es möglich sein, den Läufer immer wieder in genau der gleichen "wahren" Zeit die Strecke durchlaufen zu lassen.

II.1 Was ist der wahre Wert?

Wenn wir das tun, werden wir, wie schon oben angedeutet, feststellen, daß die Meßwerte, die wir nacheinander registrieren, immer ein wenig voneinander abweichen. Wir fragen uns also, was wir tun können, um dem "wahren" Wert der zu messenden Zeit möglichst nahe zu kommen. Dazu betrachten wir das Protokoll unserer Messungen, denn wir haben natürlich jede Einzelzeit notiert. Aber bevor wir weitermachen, erlauben wir uns wieder einen kleinen intellektuellen Schlenker: Anstatt daß wir den armen Läufer "immer wieder" starten lassen, lassen wir ihn nur einmal laufen; dafür lassen wir die Zeit jetzt von 20 Leuten gleichzeitig bestimmen: Jeder von diesen hat eine Stoppuhr in der Hand, die den anderen nach Bauweise und Ablesegenauigkeit gleicht. Auch soll keine defekt sein oder vor- oder nachgehen.

Der Läufer läuft also, und die 20 Linienrichter geben ihre gemessenen Zeiten zu Protokoll. Um eine bessere Übersicht zu haben, trägt der Ober - Linienrichter sie in ein simples Diagramm ein, er markiert nämlich auf einer geraden Linie, die die "Zeitachse" darstellt, jede gemessene Einzelzeit mit einem kurzen senkrechten Strich an der jerweiligen Position.

Wir übersehen jetzt mit einem Blick, wie sich die Meßwerte verteilen: Irgendwo "häufelt" es, und an den Rändern wird es dünn. Natürlich haben wir den Verdacht, daß die Meßwerte da, wo sie dichter liegen, in der Nähe des wahren Wertes liegen, während die "dünnen" Werte weiter von diesem entfernt sind. Und jetzt gehen wir ganz tapfer los und behaupten: "Die beste Schätzung für den (unbekannten!) wahren Wert ist das arithmetische Mittel aller Einzelmessun". Das leuchtet unmittelbar ein, und wir werden gleich sehen, daß der Mittelwert tatsächlich vor allen anderen Werten ausgezeichnet ist. Wenn wir unsere Einzelmessungen mit t_i bezeichnen, wobei der Index i von 1 bis 20 läuft, dann erhalten wir den Mittelwert, indem wir alle 20 Messungen aufsummieren und die Summe durch 20 dividieren:

$$\langle t \rangle \;=\; \frac{\sum_{i=1}^{20} t_i}{20}.$$

Für die, denen diese Formel seltsam vorkommt: Das große griechische Sigma mit den beiden Indizes 1 und 20 ist nichts als eine Abkürzung für die Vorschrift, die zwanzig einzelnen Messungen t_i aufzusummieren. Wenn man nicht von vornherein weiß, um wie viele Messungen es geht, nennt man deren Anzahl ganz allgemein N und summiert entsprechend:

$$\text{Summe} = \sum_{i=1}^{N} s_i,$$

wobei die s_i die N Summanden sind.
Natürlich kann man die Summe, je nach Problemstellung, auch bei Null oder -3 anfangen und bei $N-2$ oder $N+5$ etc enden lassen.
Aber zurück zu unserem 1000 m - Läufer. Wir haben aus den 20 Einzelmessungen den Mittelwert ("mean value") $\langle t \rangle$ herausgekocht, den wir für die beste Schätzung des wahren Wertes halten. Bevor wir aber den Charme des Mittelwertes (MW) ganz erkennen, wollen wir unser Gedankenexperiment erweitern: Wir wiederholen die Zeitmessung, allerdings mit neuen Stoppuhren, die deutlich weniger genau sind als die vorigen. Wenn wir dann wieder die Einzelmessungen auf unserer Geraden einzeichnen, dann werden wir sehen, daß die Werte jetzt wesentlich weiter auseinanderliegen als bei der ersten Messung. Wie können wir diese beiden unterschiedlichen Meßreihen quantifizieren?

II.2 Varianz und Standardabweichung

Wir brauchen offenbar ein Maß für die Streuung, und das erhalten wir folgendermaßen: Wir rechnen uns zunächst den MW aus und nennen ihn $\langle t \rangle$. Wie man das macht, wissen wir bereits. Dann bilden wir für jede Einzelmessung die Differenz zum MW und **quadrieren** diese: $(\langle t \rangle \;-\; t_i)^2$. Diese quadrierten Differenzen addieren wir alle und dividieren dann durch (20 - 1) = 19. Das Ergebnis nennen wir "Varianz" und kürzen es mit σ^2 ("sigma - Quadrat") ab:

$$\sigma^2 \;=\; \frac{\sum_{i=1}^{20} (\langle t \rangle \;-\; t_i)^2}{19},$$

oder allgemein, bei N Messungen:

$$\sigma^2 \;=\; \frac{\sum_{i=1}^{N} (\langle t \rangle \;-\; t_i)^2}{N-1}.$$

Durch die Konstruktion der Varianz als Summe aus (quadrierten) Differenzen zwischen Mittelwert und Einzelmessungen wird unmittelbar klar, daß sie für "schmale" Verteilungen klein und für "breite" Verteilungen groß wird. - Die Varianz wird auch "mittlere quadratische Abweichung" (engl. mean square deviation) genannt.
Natürlich fragt sich hier der scharfsinnige Leser / die scharfsinnige Leserin sofort, warum man bei der Berechnung der Varianz durch "$N-1$" und nicht durch N dividiert, wie bei der MW - Bildung. Ein Grund dafür liegt darin, daß man eben den wahren Wert nicht kennt und ihn durch dem MW ersetzen muß, den man aber auch aus den 20 bzw. N Einzelmessungen ableiten muss. Dies bedeutet eine Informationseinbuße, die man mit einer größeren Varianz "bezahlen" muß. Oft wird auch gefragt, warum man denn die Abweichungen vom MW quadrieren muß - tut es nicht auch der Absolutwert der Differenz? Er tut es nicht; erstens aus formalen Gründen: Die Varianz ist das "zweite Moment" der Verteilung, und die ist "quadratisch" definiert (s. auch Abschnitt 3.2), und zweitens auch anschaulich: Nur eine parabolische Abhängigkeit führt auf ein eindeutiges Minimum. Darauf kommen wir gleich zurück. Der Mittelwert ist das "erste Moment".
Mit der Varianz (englisch: variance) haben wir jetzt ein Maß für die "Breite" einer Verteilung gewonnen. Wenn zwei Meßreihen derselben Größe zwei verschiedene Varianzen zeitigen, so werden wir die mit der kleineren Varianz als die genauere Messung bezeichnen.
Wenn man aus der Varianz σ^2 die Quadratwurzel zieht, so erhält man "die Standardabweichung" σ. (Englisch: standard deviation.) Da die Quadratwurzel eine monotone Funktion ist, ist auch σ ein Maß für die Breite einer Verteilung. Man nennt die Standardabweichung deshalb auch den "Fehler der Einzelmessung". Wir schreiben diese Größe noch einmal hin und **merken uns ihre Definition fürs Leben**:

$$\sigma \;=\; \sqrt{\frac{\sum_{i=1}^{N} (\langle t \rangle \;-\; t_i)^2}{N-1}}.$$

Daß σ auch "Fehler der Einzelmessung" genannt wird, liegt daran, daß man bei Kenntnis des Mittelwertes und der Standardabweichung im allgemeinen (und im besonderen bei der Gaußverteilung) die Wahrscheinlichkeit angeben kann,

daß ein Meßwert in einem Intervall (MW - σ, MW + σ) um den Mittelwert liegt. Gerade die Größe dieses Intervalls kennzeichnet ja die "Genauigkeit" einer Messung. - Die Standardabweichung wird in der angelsächsischen Literatur oft als "r.m.s." ("root mean square") bezeichnet.

II.3 Der diskrete Charme des Mittelwerts

Wir wollen aber noch einmal auf die besondere Eigenschaft des Mittelwertes zurückkommen. Dazu denken wir z B. an die verschiedenen Meßwerte, die wir gewonnen haben, als wir die Zeit des 1000 m - Läufers gestoppt haben. Wir schreiben uns jetzt die Varianz dieser Verteilung auf (wobei wir uns nicht mehr auf 20 Messungen festlegen), ersetzen aber den MW durch eine zunächst unbekannte Größe x:

$$\sigma^2 \;=\; \frac{\sum_{i=1}^{N}\;(x\;-\;t_i)^2}{N-1}.$$

Wir haben jetzt keine wohldefinierte Zahl mehr vor uns, sondern, da x eine zunächst beliebige Zahl ist, eine **Funktion** von x; deshalb schreiben wir korrekt:

$$f(x) \;=\; \frac{\sum_{i=1}^{N}\;(x\;-\;t_i)^2}{N-1}.$$

Jetzt wollen wir wissen, für welchen Wert von x die Funktion $f(x)$ ein Minimum annimmt. Wie macht man das? Schulwissen /Hausaufgabe! Man muß die Funktion $y = f(x)$ nach x differenzieren (= ableiten) und das Ergebnis = 0 setzen. (Die t_i und N sind jetzt bekannte und konstante Zahlen.) Dann erhalten wir eine Gleichung, die wir nach x auflösen können. Wir tun das und bekommen für den Wert, der die Funktion minimalisiert,

$$x_{min} \;=\; \frac{\sum_{i=1}^{N}\;t_i}{N} \;=\; \langle t\rangle.$$

Dies ist aber - oh Wunder! - nichts anderes als der Mittelwert. Wir haben also gelernt: Der Mittelwert einer Verteilung macht ihre Varianz zum Minimum. Diese Eigenschaft zeichnet den MW vor allen anderen Werten (auf unserer Meßgeraden) aus.

Wir haben jetzt den Mittelwert und die Varianz bzw. Standardabweichung einer Verteilung kennengelernt. Wir erinnern uns aber daran, daß unsere Verteilung (nämlich die der Einzelzeiten des Läufers) aus einzelnen "diskreten" Meßpunkten bestand. Wir konnten uns bei der MW - Bildung also der einfachen Summation bedienen.

De facto ist aber die Verteilung der Zeiten kontinuierlich, denn der Läufer muß keine "Quantisierungsvorschrift" seiner Laufzeiten beachten. Im Falle einer kontinuierlichen Verteilung muß die Summation durch ein Integral ersetzt werden. Wir kommen darauf zurück.

Wir wollen zum Schluß dieses Abschnitts noch einmal auf die "beste Schätzung" ("best estimate") des wahren Wertes und ihre Unsicherheit zurückkommen. Denn natürlich hat auch der Mittelwert selbst einen "Fehler" oder besser eine Unsicherheit. Die Theorie sagt nun (vgl. z.B. [1, 2]), daß dieser Fehler durch

$$\sigma(MW) = \frac{\sigma}{\sqrt{N}}$$

gegeben ist. Dabei bedeutet $\sigma(MW)$, wie gesagt, den statistischen Fehler des Mittelwertes, während das σ auf der rechten Seite die Standardabweichung der Verteilung der Meßwerte ist.

Diese Beziehung ist wichtig und weitreichend. Sie besagt einerseits, daß man die (statistische) Unsicherheit einer Größe unter jeden vorgegebenen Wert drücken kann, wenn man nur oft genug mißt Andererseit wird der Preis (und dies, bei teuren Expertimenten, im Wortsinn) immer höher, denn um den Fehler um den Faktor 10 zu reduzieren, muß man 100 mal so oft messen. Dies führt bei jedem praktischen Experiment zu Begrenzungen.

II.4 Systematische Fehler

Außerdem ist da noch eine andere Schwierigkeit im Spiel: Wir hatten in unserem Gedankenexperiment angenommen, daß alle Stoppuhren von gleicher Qualität sind und daß keine vor- oder nachgeht. Dies ist natürlich eine Idealisierung, im wirklichen Leben wird jedes Meßgerät einen, vielleicht nur winzigen, Fehler aufweisen. Die Verfälschung eines Meßergebnisses durch fehlerhafte Geräte kann man natürlich **nicht** durch noch so viele Messungen kompensieren. Man muß also auch seinen **systematischen** Fehler zu bestimmen wissen, bevor man sich Gedanken über den Umfang einer Meßreihe macht: Sobald der systematische Fehler in die Größenordnung des statistischen rückt, lohnt es sich nicht mehr, die Anzahl der Messungen in die Höhe zu treiben.

III Von Meßreihen zu Histogrammen und Wahrscheinlichkeitsdichten: Die Gaußverteilung

Wir wollen uns jetzt um die Darstellung und Beschreibung von Verteilungen gemessener Größen kümmern. Unser Ziel ist, zum Begriff der "Wahrscheinlichkeitsdichte" und schließlich zur Gaußverteilung, die die wichtigste kontinuierliche Verteilung ist, zu kommen. Wir werden auch noch echte "diskrete" Verteilungen kennenlernen, nämlich die wichtige Binomial- und die ebenso wichtige Poisson - Verteilung.

Wir kehren zu unserer Zeitmessung zurück. Wir wollen jetzt annehmen, daß die Zahl der Zeitmessungen sehr groß wird, z.B. in die Tausende geht. Etwas unrealistisch, aber schließlich ist es ein Gedankenexperiment. Bei einer sehr großen Zahl von Messungen ist es unökonomisch, jede einzelne mit einem kleinen senkrechten Strich auf unserer Geraden zu markieren. Wir werden stattdessen den Meßbereich - etwa von 172 Sekunden bis 188 Sekunden - in Intervalle einteilen, deren Anzahl und damit Dichte von der Zahl der Messungen abhängt. Ein solches Intervall wird auch "Bin" (von englisch bin = Kasten, Behälter) genannt.

Der nächste Schritt führt uns zum "Histogramm": Wir zählen die Meßwerte in jedem Intervall und tragen die Summe als Ordinate nach oben auf. Damit haben wir wieder einen anschaulichen Eindruck von unserer Verteilung (Abb. 1):

Dieses Histogramm stellt den Fall dar, daß der Läufer "im Mittel" der Messungen 180 s für die 1000 m braucht, und daß die Standardabweichung 2 s beträgt.

An den Rändern links und rechts gibt es nur wenige Einträge ("entries"), während wir etwa in der Mitte der Verteilung ein Maximum ("peak") sehen. Außerdem ist die Verteilung so gut wie symmetrisch. An der Ordinate können wir ablesen, wie viele Einträge etwa jedes Bin zählt. Allerdings haben wir gegenüber unserer Geraden mit den Strichelchen einiges an Information verloren: Wir wissen nur noch, wie viele Meßpunkte z.B. zwischen den Marken "174 s" und "176 s" liegen; wie sie sich aber innerhalb des Intervalls verteilen, wissen wir nicht mehr - wir haben "darüber hinwegintegriert".

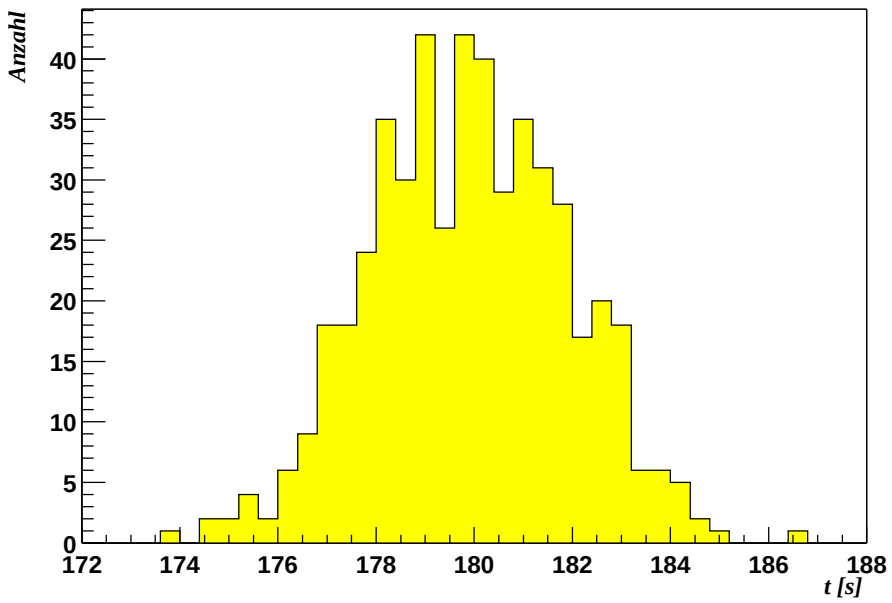

Abbildung 1: "Histogramm" der gemessenen Zeiten des Läufers mit 500 Einträgen

III.1 Vom Histogramm zur Wahrscheinlichkeitsdichte

Eine solche Darstellung nennt man ein "Histogramm" (englisch: histogram). In diesem Wort stecken die griechischen Wurzeln "histos" = Gewebe und "gramma" = Buchstabe, Schrift. Wir wollen nun unser Histogramm etwas aufbohren und neu interpretieren:
Wir lassen jetzt die Zeit unseres bedauernswerten 1000 m - Läufers von mehr und mehr und mehr Leuten messen. Wenn wir die Anzahl der stoppuhrbewerten Meßdiener "gegen unendlich" gehen lassen, können wir die Intervallbreite ("bin size") immer kleiner und kleiner wählen - wir haben ja genug "Statistik". Damit wird das Histogramm auch immer glatter, denn der Polygonzug, den die Ecken der Intervall - Inhalte bilden, geht in allmählich in eine glatte Kurve über. Wir sind eigentlich zufrieden. Aber da kommt ein Kollege und sagt: "Unendlich

reicht mir nicht, ich möchte zwei mal unendlich."

Um ihm diesen Gefallen zu tun, müßten wir die Zahlen auf der Ordinate auch mit zwei multiplizieren. Aber wir ahnen, daß wir keine Information gewinnen würden, unser Histogramm sähe noch genau so aus. Das bringt uns auf den Gedanken, die Verteilung zu "normieren" ("to normalize"). Was bedeutet das? Wir merken uns die Zahl **aller** Einträge des Histogramms und dividieren dann den Inhalt eines jeden Bins durch diese Zahl. Ergebnis: Wenn wir jetzt die Summe aller Bin - Inhalte ("bin contents") bilden, erhalten wir immer Eins - das Histogramm ist "auf Eins normiert".

Und jetzt, in diesem Grenzfall unendlich vieler Einträge, können wir das Histogramm neu interpretieren: Es stellt sicherlich ein gutes "Persönlichkeitsprofil" unseres Läufers dar, und wir deuten es jetzt als "Wahrscheinlichkeitsdichte". Was heißt das? Die gute alte Massendichte bezeichnete "Masse pro Volumen". Die Wahrscheinlichkeitsdichte in diesem Fall bedeutet "Wahrscheinlichkeit pro Zeit". Anschaulich: Wir schraffieren die Fläche zwischen (z.B.) 176 s und 184 s; da wir die Gesamtfläche auf Eins normiert haben, gibt uns die schraffierte Fläche die Wahrscheinlichkeit, mit der der Läufer ein Ergebnis zwischen 176 s und 184 s erzielt. Dabei haben wir vorausgesetzt, daß er nie weniger als 172 s und nie mehr als 188 s braucht. Mathematisch haben wir die Verteilung zwischen den beiden Intervallen "integriert". Trivialerweise (da nach Konstruktion) ist das Integral über die gesamte Verteilung gleich Eins. Das bedeutet, der Läufer wird - wegen unserer einschränkenden Voraussetzung - **auf jeden Fall** (also mit der Wahrscheinlichkeit Eins) ein Ergebnis zwischen 172 s und 188 s erzielen. Noch einmal: Die Verteilung selbst gibt uns die Wahrscheinlichkeitsdichte. Eine Wahrscheinlichkeit erhält man erst durch Integration über ein vorgegebenes Intervall. Hieraus folgt auch, daß die Wahrscheinlichkeit, **genau** einen bestimmten Wert zu messen, gleich Null ist. Warum ist das so? Denken Sie über diese Merkwürdigkeit nach.

Damit haben wir die Wahrscheinlichkeitsdichte verstanden. In unserem Beispiel hatten wir die Ergebnisse des Läufers auf das Intervall von 172 s bis 188 s eingeschränkt. Eine solche Einschränkung gilt natürlich im allgemeinen nicht, vielmehr sind im allgemeinsten Fall die Grenzen "minus unendlich" und "plus unendlich" zu wählen.

Wenn wir uns die Verteilungsdichte unseres Läufers genauer ansehen, dann hat sie die Form einer Glocke. Dies ist die Form der Gauß- oder Normalverteilung; in der Tat läßt sich eine Theorie der Meßfehler ("Laplacesches Fehlermodell") aufstellen, die just auf die Gaußverteilung führt. Wir wollen das als gegeben hinnehmen (siehe z.B. [1]) und uns die Gaußverteilung etwas näher ansehen.

III.2 Annäherung an Herrn Gauß

Die Verteilung ist offensichtlich positiv und symmetrisch. Wenn sie um den Punkt x_0 (wir wählen jetzt allgemeine Variablen für die zu messenden Größen, es könnte ja auch eine Temperatur sein) symmetrisch ist, so empfiehlt sich eine Form (das Zeichen $\propto$ bedeutet "proportional zu")

$$\propto \exp\left(-(x-x_0)^2\right).$$

(Mit "$\exp(x)$" meinen wir natürlich die Exponentialfunktion "e hoch x": $\exp(x) = e^x$. Wir wollen sie hier nicht im Detail diskutieren. Wer ihre Eigenschaften vergessen hat, grabe bitte in seinen Mathematikbüchern nach.)
Damit haben wir die Glockenform. Um die Glocke "breit" oder "schmal" zu machen, führen wir im Exponenten den Faktor $1/(2\ \sigma^2)$ ein. Wird σ groß, so ist die Glocke breit, und umgekehrt. Jetzt müssen wir noch für die Normierung sorgen, denn das Integral über die gesamte Verteilung - von "minus unendlich" bis "plus unendlich" soll ja Eins sein. Dies besorgt der Faktor $1/\sqrt{2\ \pi\ \sigma^2}$. Alles in der Literatur nachzulesen [1, 2, 3, 4]. Insgesamt sieht also eine Gaußverteilung (wohlgemerkt, wieder eine Wahrscheinlichkeitsdichte!) so aus:

$$g(x) = \frac{1}{\sqrt{2\ \pi\ \sigma^2}}\ \exp\left(-\frac{(x-x_0)^2}{2\ \sigma^2}\right)$$

Diese etwas kompliziert aussehende Funktion malen wir uns noch einmal in aller Schönheit auf (Abb. 2):
Wir haben für dieses Bild mit Absicht eine standardisierte Gaußverteilung gewählt, nämlich eine mit dem MW Null und der Breite $\sigma = 1$. Die Verteilung in der Abbildung ist übrigens - zur besseren Darstellung - **nicht** korrekt auf Eins normiert. Aus dieser Verteilung kann man leicht jede andere herstellen, indem man den MW verschiebt und die Breite durch Multiplikation mit dem gewünschten Faktor verkürzt oder streckt.
Und jetzt wollen wir uns ihre Eigenschaften ansehen. Zunächst: Die Kurve kommt von $-\infty$ und geht nach $+\infty$. Ihr Integral ist Eins (dafür sorgt der Normierungsfaktor):

$$\int_{-\infty}^{+\infty} g(x)\ dx\ =\ 1.$$

Wenn wir die Wahrscheinlichkeit wissen wollen, ein Ergebnis im Intervall mit den Grenzen a und b zu erzielen, dann ist diese gegeben durch

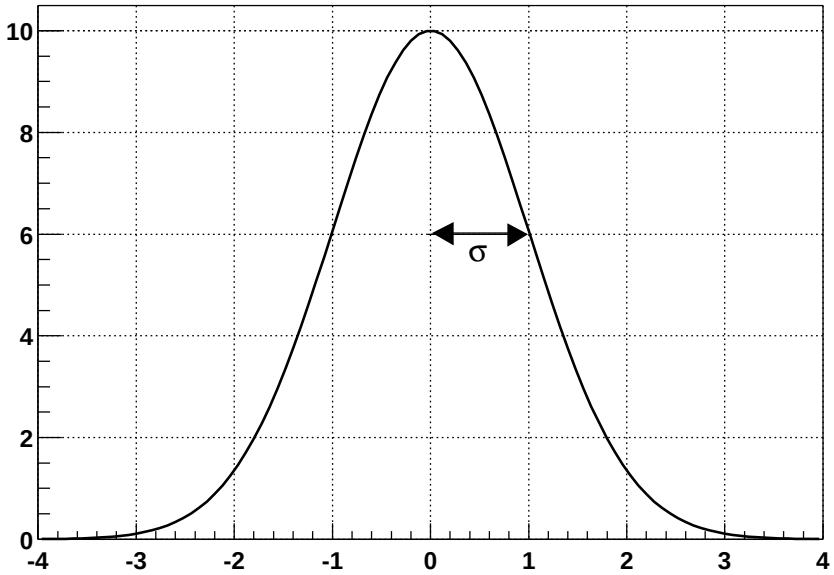

Abbildung 2: Gaußverteilung mit dem Mittelwert Null und der Breite (σ) Eins

$$G(a,b) \;=\; \int_a^b g(x)\; dx.$$

Die Gaußverteilung hängt von genau zwei Parametern ab, nämlich dem Mittelwert x_0 und der Breite σ. Tatsächlich gilt

$$\int_{-\infty}^{+\infty} x\; g(x)\; dx \;=\; x_0.$$

Auf der linken Seite steht die Definition des Mittelwertes ("erstes Moment") bei einer kontinuierlichen Verteilung; sie entspricht der Summation bei diskreten Meßwerten. Außerdem gilt:

$$\int_{-\infty}^{+\infty} x^2\; g(x)\; dx \;=\; \sigma^2.$$

Das ist das Analogon zur Bestimmung der Varianz ("zweites Moment") bei diskreten Meßwerten. Übrigens: Man findet die Position "$\pm$ 1 σ" an den Stellen, an denen die Gaußkurve auf den Wert "Maximum/$\sqrt{e}$" abgesunken ist. Dies sind zugleich die Positionen der Wendepunkte $\rightarrow$ nachrechnen!
Wir halten fest: Eine Gaußverteilung wird durch ihren Mittelwert (hier x_0) und ihre Breite ("Gaussian width") σ beschrieben. σ ist die Standardabweichung der Verteilung. (Übrigens: Sprechen Sie im Englischen "Gaussian" niemals "Gooschn" aus, das ist igitt! Der gebildete Angelsachse tut das nicht.)
Jetzt haben wir fast alles Wichtige über die Gaußverteilung beisammen. Und deshalb wollen wir noch einmal zum Anfang unserer Überlegungen, nämlich zu den "Vetrauensintervallen" zurückkommen. Wie sieht es beim Gauß aus?

III.3 Vertrauen bei Herrn Gauß

Die Mathematik sagt uns, wie groß das Integral über die Gaußverteilung "von $-\sigma$ bis $+\sigma$" ist, nämlich 0.683. Das bedeutet anschaulich: Wenn man eine Größe mißt, die "gaußverteilt" ist (und diese Annahme macht man meistens, wenn auch oft nur als Näherung), dann ist die Wahrscheinlichkeit 68.3 %, daß das Ergebnis in eben jenes Intervall fällt. Man nennt dies die "1 σ - Umgebung. Wer eine gemessene Größe mit einem Fehler veröffentlicht, der meint mit der Größe seines Fehler immer implizit dieses "Gaußsche σ". Das heißt, die Angabe eines Fehlers ist auch immer eine klare quantitative Botschaft: Ich behaupte, daß der "wahre Wert" der Größe, die ich messen will, mit einer Wahrscheinlichkeit von 68.3 % in dem angegebenen Intervall liegt.
Natürlich gibt es auch die "2 σ"- und "3 σ"- Intervalle. Wieder sagt uns die Mathematik, daß die ensprechenden Integrale über die Gaußverteilung 0.955 und 0.997 sind. Mit anderen Worten: Mit einer Wahrscheinlichkeit von 95.5 % fällt eine Messung in das 2 σ-, und mit 99.7 % in das 3 σ- Intervall. Die Wahrscheinlichkeit, daß die Messung außerhalb dieses Intervalls landet, ist also nur 0.3 %. Wenn etwa zwei Messungen derselben Größe - bei Berücksichtigung der jeweiligen Fehler - um (mindestens) "3 σ" differieren, dann nennt man sie "signifikant verschieden", und man hält eine von beiden für falsch. Hier lauert die Tücke: Wenn man zwar seine Messung ordentlich durchgeführt, seine Fehler aber versehentlich oder aus Dummheit zu klein bestimmt hat, dann schließt man sich selbst aus der Reihe der seriösen Wissenschaftler aus.

Wenn man die Gaußsche Glockenkurve zeichnet, dann kann man die "1σ- Umgebung" leicht ablesen, denn 1σ ist gerade der Abstand vom Mittelwert (dem Maximum) zu den Wendepunkten. Manchmal möchte man auch wissen, wie breit die Verteilung, mit der man es zu tun hat, "auf halber Höhe" ist. Diese "volle Breite auf halber Höhe" ("full width at half maximun", FWHM) beträgt, wie man leicht nachrechnet, etwa 2.3 σ.

Damit haben wir die Gaußverteilung im Wesentlichen verstanden. Wir fassen zusammen: Man unterstellt Messungen einer kontinuierlichen Größe, die - im weitesten Sinne - mit einem "Maßstab" durchgeführt werden, daß sie "gaußisch" sind, d.h. daß die Meßwerte, wenn man sie histographiert (s. o.), sich durch die berühmte Glockenkurve interpolieren lassen. Das ist in Wirklichkeit so gut wie nie der Fall; aber man rechnet dennoch sehr oft "gaußisch" weiter, z.B. wenn es um Hypothesen - Tests mit Hilfe der χ^2 - Methode ("chi - Quadrat") geht, auf die wir hier nicht eingehen wollen. Auch die Abschätzungen der "Vertauensintervalle" mit den oben genannten Zahlen gelten nur für "echte" Gaußverteilungen. Man muß deshalb im täglichen leben die "Gaußizität" seiner Meßergebnisse kritisch überprüfen.

Dazu gibt es einen einfachen Trick: Wenn man eine Gaußfunktion logarithmiert (z.B. mit Hilfe des beliebten Logarithmenpapiers), erhält man eine nach unten offene Parabel (eben gerade den Exponenten). Das menschliche Auge erkennt aber Abweichungen von der Parabelform und damit von der "Gaußizität" sofort!

Und nun wenden wir uns neuen Ufern zu.

IV Wir sind ganz diskret: Binomial- und Poisson - Verteilung

Wir haben es im vorigen Abschnitt mit Messungen "kontinuierlicher" Größen wie Zeiten zu tun gehabt. Ähnliches gilt für Längen, Temperaturen, Stromstärken, etc.

Es gibt aber eine ganz andere Art von Messungen, bei denen "etwas gezählt" wird. Klassische Beispiele sind die Zahl der Zerfallsakte eines radioaktiven Präparates pro Zeiteinheit oder die Zahl von Bakterien in einer Nährlösung. Hier wird gezählt: "Eins, zwei, drei..", und die natürlichen Zahlen sind eine diskontinuierliche, "diskrete" Variable.

IV.1 Alea iacta sit!

Ein einfaches Beispiel, das uns auch die Möglichkeit gibt, den Begriff der Wahrscheinlichkeit etwas zu beleuchten, ist ein symmetrischer Würfel mit den "Augen" 1 bis 6. Wegen der Symmetrie sind alle Augenzahlen gleichberechtigt. Da die Summe aller Einzelwahrscheinlichkeiten Eins sein muß, ist die Wahrscheinlichkeit, eine der Zahlen 1 bis 6 zu würfeln, gerade 1/6.
Das wundert uns auch nicht. Wir wollen aber dieses einfache Beispiel benutzen, um zwischen "und" und "oder" bei Wahrscheinlichkeiten zu unterscheiden: Die Wahrscheinlichkeit ("probability"), z.B. eine 2 **oder** eine 5 zu würfeln, ist $1/6 + 1/6 = 1/3$. Wenn ich aber mit zwei Würfeln spiele, dann ist die Wahrscheinlichkeit, eine 2 **und** eine 5 zu würfeln, gleich $1/6 \cdot 1/6 = 1/36$. Es gibt nämlich $6 \cdot 6 = 36$ Möglichkeiten von Zahlenpaaren. Diese Multiplikationsregel gilt aber nur dann, wenn die beiden Wahrscheinlichkeiten unabhängig voneinander sind. Dies können wir aber bei den Würfeln voraussetzen, solange sie nicht miteinander verklebt sind.
Es lohnt sich übrigens, einmal den MW dieser "flachen" (alle Zahlen sind gleich wahrscheinlich) Verteilung auszurechnen: Es ist $1/6 \cdot (1 + 2 + 3 + 4 + 5 + 6) = 21/6$; der Mittelwert der "Augenzahlen" eines Würfels ist also eine gebrochene Zahl, die selbst nicht das Ergebnis eines Wurfes sein kann.
Wir wollen uns in diesem Abschnitt der äußerst wichtigen Poisson - Verteilung nähern. Ihre Kenntnis ist von herausragender Bedeutung. Und da trotzdem kaum jemand über sie bescheidweiß, bedeutet es so etwas wie "Herrschaftswissen", wenn man sich mit ihr auskennt.

IV.2 Tossing a Coin: Head or Tail?

Die Poissonverteilung läßt sich aus der "Binomialverteilung" herleiten - wie übrigens auch die Gaußverteilung. Die Binomialverteilung ist also sozusagen die "Mutter aller Verteilungen". Mit ihr kann man die Ergebnisse eines "Bernoulli - Experimentes" beschreiben, das ist eines mit genau zwei möglichen Ergebnissen. Ein einfaches Bernoulli - Experiment ist der Wurf einer Münze. Mit Hilfe der Binomialverteilung (BiV) kann ich dann z.B. die Frage beantworten: "Wie groß ist die Wahrscheinlichkeit, einmal "Zahl" und 9 mal "Adler" zu bekommen, wenn ich eine (symmetrische) Münze 10 mal werfe?" Man kann sich das Ergebnis eigentlich "zu Fuß" ausrechnen, wenn man alle Möglichkeiten abzählt. Wir wollen das hier tun, bevor wir uns die BiV im Detail ansehen. Also:

Wir wollen das Erscheinen von "Zahl" nach dem Wurf den "Erfolg" nennen. Dann bedeutet "Adler" Mißerfolg. Die Wahrscheinlichkeit für einen Erfolg nennen wir p, die für Mißerfolg q. Da ich nur entweder Erfolg oder Mißerfolg haben kann, gilt $p+q = 1$. Andererseits ist die Münze symmetrisch, daher $p = q = 1/2$. Die Anzahl der "Erfolge" sei r, also hier $r = 1$. Die Anzahl der Würfe ist $n = 10$. Dann ist die Wahrscheinlichkeit, bei 10 voneinander unabhängigen Würfen einmal "Erfolg" zu haben, $p^r = 0.5^1$. Entsprechend ist die Wahrscheinlichkeit, 9 mal "Mißerfolg" zu haben, gleich $q^9 = (1-p)^9 = 0.5^9$. Damit sind wir schon fast fertig, wir müssen aber noch bedenken, daß wir das Ergebnis auf 10 verschiedene Art und Weisen erreichen können, nämlich (A für Adler, Z für Zahl): ZAAAAAAAAA, AZAAAAAAAA, AAZAAAAAAA, etc. Damit haben wir unser Ergebnis beisammen: Die gesuchte Wahrscheinlichkeit ist

$$B(1;10,0.5) \;=\; 10 \;\cdot\; 0.5 \;\cdot\; 0.5^9 \;=\; 10/1024\;,$$

also etwa 1 %.
Für den allgemeinen Fall (p und q verschieden, aber natürlich immer $p+q = 1$; wir können hier an eine manipulierte Münze denken, die auf einer Seite schwerer ist) lautet die Formel

$$B(r;n,p) = \frac{n!}{r!\;(n-r)!}\;p^r\;q^{n-r}.$$

Das sieht ein bißchen angsteinflößend aus, aber nach unserer vorherigen Abzählübung verstehen wir es: $B(r;n,p)$ gibt die Wahrscheinlichkeit an, bei n Würfen r Erfolge zu erzielen, wenn die Wahrscheinlich für den Erfolg gleich p (und damit die Wahrscheinlichkeit für den Mißerfolg gleich $q = 1-p$) ist. Unnötig zu sagen, daß p nur eine Zahl zwischen Null und Eins sein kann.
Der Bruch auf der rechten Seite ist ein sogenannter "Binomialkoeffizient"; er gibt an, auf wie viele verschiedene Weisen man r Erfolge bei n Würfen realisieren kann. (Raten Sie, was der Ausdruck 49! / (6! (49−6)!) bedeutet - Woche für Woche!) In unserem Beispiel war er gleich 10! / (1! (10−1)!) = 10. Ach ja: Der Ausdruck 10! (sprich "zehn - Fakultät") bedeutet $10! = 1 \cdot 2 \cdot \cdots \cdot 9 \cdot 10$. Merke fürs Leben: $1! = 0! = 1$. Außerdem: (Jede Zahl)$^0 = 1$, auch $0^0 = 1$. Unser Beispiel hat uns also gelehrt, daß ich bei 100 Serien von jeweils 10 Würfen im Mittel einmal das Ergebnis "einmal Zahl und neunmal Adler" erzielen werde. Alles verstanden? Gut. Rechnen Sie als Hausaufgabe das Ergebnis für $r = 4$ aus!
Der Vollständigkeit halber sei noch erwähnt, daß der MW der BiV gleich $n \cdot p$ ist, und ihre Varianz gleich $n \cdot p \cdot q$.

IV.3 Von "Herrn Binom" zu Herrn Poisson

Wenn man in der Binomialverteilung die Wahrscheinlichkeit für den "Erfolg" immer kleiner, die Anzahl der "Würfe" immer größer werden läßt, dabei aber das Produkt $n \cdot p$ (also den Mittelwert) konstant hält, dann erhält man die Poisson - Verteilung. Die Poissonverteilung greift immer dann, wenn man Ereignisse in einem bestimmten "Intervall x" zählt, typischerweise von Raum oder Zeit, die voneinander und von x unabhängig sind (d.h.einander nicht kausal beeinflussen, und deren Häufigkeit auch nicht von der Wahl des Intervalls abhängen soll). Ein Paradebeispiel ist der Zerfall eines (langlebigen) Radio - Isotops: Wir wollen annehmen, daß die mittlere Lebensdauer sehr groß gegen die Beobachtungszeit sei (so daß die mittlere Zerfallsrate sich nicht ändert).
Dann könnte unser Experiment folgendermaßen aussehen: Wir bewaffnen uns mit einem Zählrohr und messen sehr oft, wie viele Zerfälle in jeweils einer Minute stattfinden. Wir wissen natürlich, daß die mittlere Zerfallsrate pro Minute durch Halbwertszeit und Menge des radioaktiven Materials (und durch andere Parameter) gegeben ist. Wir interessieren uns aber vor allem um die Fluktuationen.
Nun ist aber der radioaktive Zerfall ein durch und durch statistisches ("stochastisches") Phänomen. In unserer Probe befindet sich eine immense Zahl von Kernen unseres Radionuklids; aber keiner dieser Kerne "weiß" etwas von den anderen - es gibt keine kausale Beeinflussung der Kerne untereinander. Jeder Kern "entscheidet" sich allein, wann er "zerfallen will", und gehorcht dabei nur übergeordneten physikalischen Gesetzen, die z.B. die Halbwertszeit für just dieses Nuklid festlegen.
Es kann also sein (vorausgesetzt, daß ich den Umfang meiner Probe entsprechend gewählt habe), daß ich in einer Minute einmal neun, einmal zwei oder sieben Zerfälle, oder auch einmal gar keinen Zerfall registriere. Nur wenn ich nach jeweils (z.B.) 100 Meßreihen den Mittelwert der Zahlen meiner Zerfälle bilde, dann bekomme ich - mit kleinen Fluktuationen - immer die gleiche Zahl. Wir stellen nun das Ergebnis unserer Meßreihen graphisch dar, und zwar in der bewährten Form des Histogramms. Wir teilen unsere Abszisse in Bins ein, die wir mit 0, 1, 2, usw. bezeichenen, im Prinzip bis "unendlich". Wenn wir keinen Zerfall registriert haben, dann machen wir ein Kreuzchen in das "nullte" Bin, bei einem Zerfall in das erste, und so weiter. Wenn wir das lange genug gemacht haben, dann bekommen wir eine veritable Poisson - Verteilung. Die Form dieser Verteilung hängt nur (das ist wichtig!) von **einem** Parameter ab, nämlich ihrem Mittelwert, den wir μ nennen wollen. Tapfer, wie wir sind, stellen wir uns gleich der mathematischen Formel für die Poisson - Verteilung:

$$P(n;\mu) \;=\; \frac{\mu^n\; e^{-\mu}}{n!}$$

In dieser Formel bedeutet μ (der griechische Buchstabe "my"), wie schon gesagt, den MW der Verteilung. n ist schlicht ein Laufindex, der von Null bis unendlich geht: $n = 0, 1, 2, ..\infty$. Und die Formel wird folgendermaßen interpretiert (wir bleiben bei unserem Beispiel): Wenn der Mittelwert der Zerfälle meines Nuklids in einem gegebenen Zeitintervall (z.B. einer Minute) gleich μ ist, dann wird die Wahrscheinlichkeit, genau n Zerfälle in diesem Intervall zu messen, durch $P(n;\mu)$ gegeben. Alles klar? Übrigens: Die Poisson - Verteilung ist zwar eine diskrete Verteilung - n ist eine ganze Zahl - , der Mittelwert μ jedoch kann eine beliebige reelle Zahl zwischen Null und Unendlich sein!
In den folgenden Abbildungen 3 und 4 sind zwei Poisson - Verteilungen dargestellt, und zwar einmal für einen kleinen MW (1.2), und einmal für einen großen (15). Wir sehen, daß die Verteilung für kleine MWe stark asymmetrisch ist und offenbar für größer werdende MWe immer symmetrischer wird: In der Tat geht die Poisson - Verteilung für große MWe in eine Gaußverteilung über. In der Abbildung 4 sehen wir, daß die Poisson - Verteilung für einen MW von 15 schon deutlich weniger asymmetrisch ist und eine gewisse "Gauß - Ähnlichkeit" zeigt. Aber wir müssen offensichtlich noch deutlich größere MWe wählen, bis wir den Unterschied nicht mehr erkennen können.
Hier machen wir eine Pause: Glauben wir das? Wir haben gelernt, daß die Gaußverteilung eine kontinuierliche Verteilung ist und daß sie durch **zwei** Parameter definiert wird, nämlich durch MW und Varianz σ^2. Die Poisson - Verteilung wird aber durch nur einen Parameter definiert, durch den Mittelwert. Außerdem geht die Gaußverteilung von $-\infty$ bis $+\infty$, die Poisson - Verteilung jedoch fängt erst bei Null an.
Bevor wir weiter argumentieren, lüften wir noch ein Geheimnis der Poisson - Verteilung. Ihre Varianz ist gleich ihrem Mittelwert, nämlich

$$\sigma^2_{\text{Poisson}} \;=\; \mu_{\text{Poisson}}.$$

Auf dieser Eigenschaft beruht das "Wurzel - N - Gesetz" bei der Fehlerbestimmung von gezählten Größen. Aber darauf kommen wir noch.
Nach dem, was wir bis jetzt zusammengestellt haben, kann die Poissonverteilung niemals in eine "echte" Gaußverteilung übergehen. Tut sie auch nicht. Aber tatsächlich läßt sich die Form der Poisson - Verteilung für große MWe mathematisch durch die Gaußverteilung approximieren. Man muß dann aber in der Formel für den Gauß das "σ^2" durch "μ" ersetzen - auch dieser "Pseudo

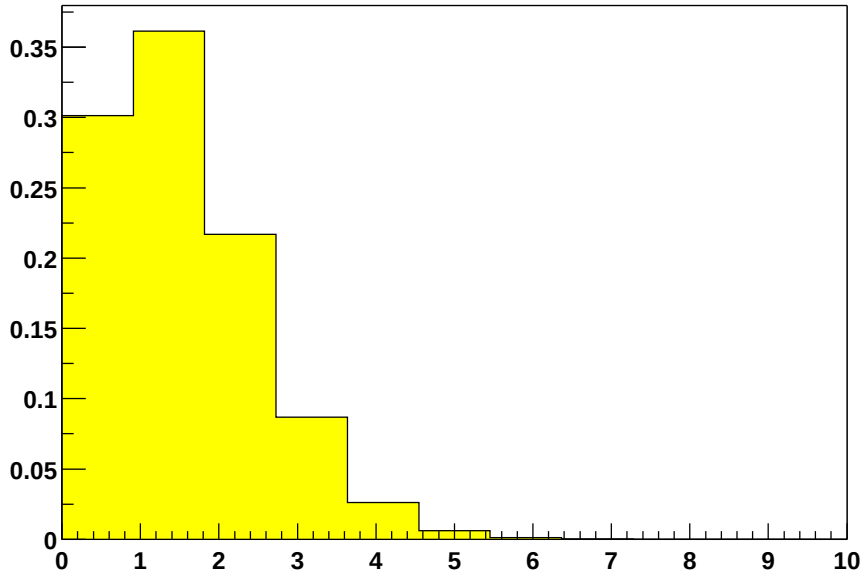

Abbildung 3: Poisson - Verteilung mit $\mu \;=\; 1.2$

- Gauß" hängt nur von einem Parameter ab! Einen Vorteil bietet die Vergaußung der Poisson - Verteilung allerdings: Für große MWe wird die Fakultät ($n!$ im Nenner) sehr unhandlich - dies wird durch die Gauß - Form vermieden.

Jetzt haben wir also auch die Poisson - Verteilung verstanden. Sie wird immer dann wichtig, wenn irgendetwas in irgendeinem Intervall gezählt wird, dessen Vorkommen völlig statistisch erfolgt. Dies können, wie in unserem Beispiel, Zerfälle eines Radionuklids pro Zeiteinheit sein, aber auch Sterne pro Raumwinkel - Segment am Nachthimmel. Regentropfen pro Quadrat - Dezimeter im Garten tun es auch, oder - beliebt bei Physikern - Bläschen entlang einer Teilchenspur in einer Blasenkammer. Mediziner dagegen haben eine Vorliebe für Leukozyten im Blut oder das Auftreten seltener Krankheiten.

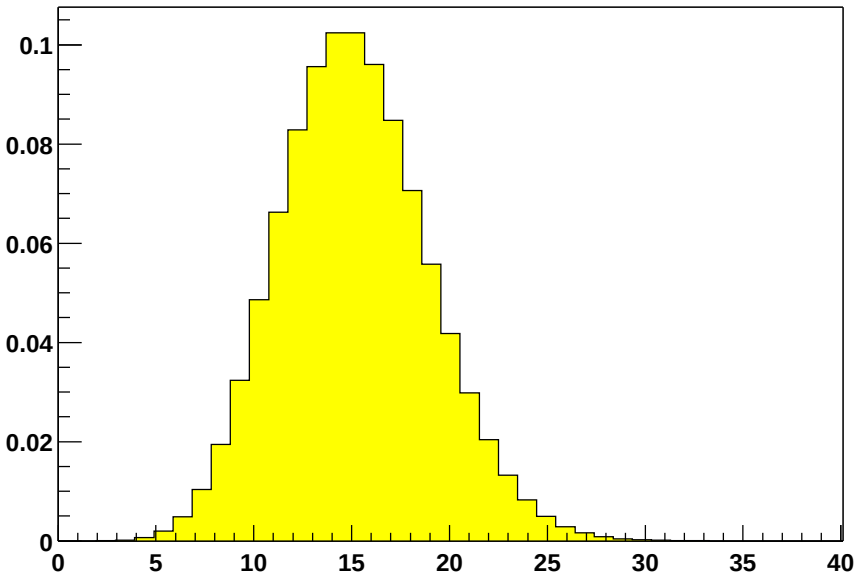

Abbildung 4: Poisson - Verteilung mit $\mu = 15$

IV.4 Das "Wurzel - N - Gesetz"

Schließlich noch die versprochene Bemerkung zum "Wurzel - N - Gesetz". Wenn jemand mit einem Geigerzähler eine Zählrate mißt, und er hat während der Meßzeit - sagen wir - 4711 "Counts" gezählt, dann wird er sagen, der Fehler (oder die statistische Unsicherheit) auf diese Zahl sei $\sqrt{4711}$. Warum?
Wie gesagt, unser Experimentator soll nur einmal gemessen haben. Was er eigentlich hätte tun müssen, aber aus verständlichen Gründen nicht tun kann, wäre "unendlich oft" zu messen. Dann könnte er nach jeder Messung sein Egebnis in ein Histogramm (s. oben) eintragen und hätte am jüngsten Tag eine perfekte Poisson - Verteilung (vorausgesetzt, sein Präparat lebt noch länger). Diese Poisson - Verteilung wird dann einen Mittelwert haben, der gleich ihrer Varianz ist. Und hier einnern wir uns daran, daß die Standardabweichung die Wurzel aus der Varianz ist.

Unser Experimentator hat nun einen doppelten intellektuellen Schlenker gemacht: Da er nur einmal messen kann, interpretiert er sein Ergebnis als Schätzung für den Mittelwert; und wenn er den erst hat, zieht er die Wurzel daraus und verkauft sie als Standardabweichung - die klassische "Unsicherheit auf die Einzelmessung". Dies sollte man im Hinterkopf haben, wenn man das "Wurzel - N - Gesetz" anwendet.
(Natürlich kann man auch hier subtiler vorgehen: Man kann nach dem Intervall fragen, in dem der wahre MW mit einer gewissen Wahrscheinlichkeit liegt, wenn ich eben bei nur einer Messung meine z.B. 4711 Zerfälle zähle. Aber das geht über den Rahmen dieser Betrachtung hinaus.)
Im Übrigen: $\sqrt{N}$ ist der **absolute** Fehler auf die gemessene Zahl N. Der **relative** Fehler ist dann

$$\frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}}.$$

Wenn ich diesen relativen Fehler mit 100 multipliziere, bekomme ich den "prozentualen" Fehler. Wir sehen, daß der relative Fehler mit "höherer Statistik" schrumpft. Man kann ihn also im Prinzip unter jede vorgegebene Grenze drücken - aber auch hier gilt wieder das, was schon im zweiten Abschnitt gesagt wurde.

IV.5 Normierung der Poisson - Verteilung

Zum Schluß noch etwas für den echten Fan: Wir wissen, daß jede Verteilung einer Wahrscheinlichkeitsdichte "auf Eins normiert" sein muß, denn irgendein Ergebnis kommt "mit Sicherheit", also der Wahrscheinlichkeit Eins, heraus. Wie steht es mit der Poisson - Verteilung?
Um die Normierung zu überprüfen, muß die Summe über alle Einzelwahrscheinlichkeiten Eins werden, also:

$$\sum_{n=0}^{\infty} \frac{\mu^n \; e^{-\mu}}{n!} = 1.$$

Stimmt das? Wir schreiben dieselbe Formel etwas anders, indem wir die Exponentialfunktion vor die Summe ziehen:

$$e^{-\mu} \sum_{n=0}^{\infty} \frac{\mu^n}{n!} = 1.$$

Aber nun sieht unser scharfes Auge sofort, was unter dem Summenzeichen steht: Es ist die Taylor - Reihenentwicklung der Exponentialfunktion, die (mit x statt mit μ) ausgeschrieben so aussieht:

$$1 \; + \; x \; + \; \frac{x^2}{2!} \; + \; \frac{x^3}{3!} \; + \; \cdot \; \cdot \; = \; e^x.$$

Damit wird unsere Summe schlicht zu

$$e^{-\mu} \; e^{\mu} \; = \; e^{(\mu - \mu)} \; = e^0 \; = \; 1,$$

was zu beweisen war.

IV.6 Die "Duplizität der Ereignisse"

Man beobachtet im täglichen Leben des öfteren, daß zwei außergewöhnliche Ereignisse rasch nacheinander stattfinden. Die öffentliche Aufmerksamkeit nimmt dies besonders bei Katastrophen wie z.B. Flugzeugabstürzen wahr. Können wir das mit dem, was wir bisher gelernt haben, verstehen? Wir versuchen eine Annäherung:
Wir bleiben zunächst bei unseren radioaktiven Nukliden; wir schreiben aber den Mittelwert μ (für die Anzahl der Zerfälle pro Zeitintervall) etwas um. Wir führen die "mittlere Häufigkeit pro Zeiteinheit" ein und nennen sie λ ("lambda"). Dann können wir uns den MW für jedes beliebige Zeitintervall t konstruieren, er ist schlicht $\mu \; = \; \lambda \; t$. Unser Poisson sieht jetzt so aus:

$$P(n; \mu = \lambda \; t) \; = \; \frac{(\lambda \; t)^n \; e^{-\lambda \; t}}{n!}.$$

Und jetzt fragen wir uns: Wie groß ist die Wahrscheinlichkeit, daß im Intervall t **kein** Zerfall stattfindet? Offenbar:

$$P(0; \lambda \; t) \; = \; \frac{(\lambda \; t)^0 \; e^{-\lambda \; t}}{0!} \; = \; e^{-\lambda \; t}.$$

(Wir erinnern uns: $0! = 1$.) Und diese einfache Formel (die nichts anderes ist als das Gesetz des radioaktiven Zerfalls) sagt uns, daß **kurze** Intervalle zwischen zwei Zerfällen wahrscheinlicher sind als lange, denn $e^{-\lambda \; t}$ wird um so kleiner, je größer t ist. Es ist also wahrscheinlicher, daß zwei Zerfälle dicht aufeinander folgen, als daß sie durch lange Intervalle getrennt sind.
Was hat das nun mit Katastrophen zu tun? Natürlich nur bedingt etwas. Wesentlich ist, daß bei dieser Betrachtung die einzelnen "Ereignisse" einander eben **nicht** kausal beeinflussen, wie das bei radioaktiven Zerfällen der Fall ist. Wenn wir also annehmen, daß Katastrophen wie Flugzeugabstürze "statistisch verteilt" vorkommen (und nicht, weil in einem bestimmten Flugzeugtyp etwa zwei Kabel vertauscht wurden), dann würde man tatsächlich häufiger kurze Abstände zwischen zwei solchen bedauerlichen Vorkommnissen erwarten als lange. Der Volksmund nennt das die "Duplizität der Ereignisse". Und damit endgültig genug von Herrn Poisson.

V Ein Wort zur Fehlerfortpflanzung

Die Größen, für die man sich interessiert, sind oft nicht die gemessenen selbst, sondern daraus abgeleitete. Beispiel: Das Volumen einer Kiste ist das Produkt der Kantenlängen. Wie hängt die Unsicherheit (der "Fehler") des Volumens von den Meßfehlern der Kantenlängen ab?
Diese und andere Fragen beantworten die Rechenregeln der Fehlerfortpflanzung ("error propagation"). Diese basieren aber auf einfachen Regeln der Differentialrechnung und sollen hier nicht diskutiert werden. Außerdem findet man sie in jedem Lehrbuch und auch in der Anleitung zum Physikalischen Praktikum. Es soll nur kurz der "asymmetrische Fehler" erwähnt werden; das ist ein Fehler, dessen Balken verschieden lang sind. Wie kann es dazu kommen?
Wenn eine Größe "linear" von einer anderen abhängt, wenn also beide nur durch einen konstanten Faktor verbunden sind wie z.B. beim Ohmschen Gesetz ($U = R \cdot I$) Strom und Spannung, dann skaliert natürlich auch der Fehler mit dieser Konstanten. Graphisch kann man das durch "Spiegelung an einer Geraden" darstellen. Sobald aber der Zusammenhang nicht mehr linear ist, wie z.B. bei der kinetischen Energie, die quadratisch von der Geschwindigkeit abhängt, dann spiegelt man nicht mehr an einer Geraden, sondern an einer Parabel. Und dann haben die Spiegelbilder des - symmetrischen - Fehlers z.B. der Geschwindigkeit nicht mehr gleiche Abstände vom Spiegelbild des Zentralwertes - der Fehler wird "asymmetrisch". Übrigens, in logarithmischer Darstellung erscheint natürlich auch ein symmetrischer Fehler asymmetrisch.

VI Schlußbemerkung

So, ganz langsam geht uns die Puste aus, es ist Zeit, Schluß zu machen. Was haben wir gelernt? Wir haben uns mit zwei wichtigen Methoden der Messung und ihren Unsicherheiten beschäftigt: Mit der Messung einer kontinuierlichen Größe

mit Hilfe eines (wie auch immer gearteten) “Maßstabes”, und einer Messung, die auf der Zählung von diskreten “Ereignissen” wie z.B. Zerfallsprozessen beruht. Und wir haben das dazugehörige Handwerkszeug kennengelernt, nämlich die Gauß- und die Poisson - Verteilung.
Damit haben wir in der Tat ein ansehnliches Wissen erworben, denn auf Gauß und Poisson basieren viele Gebäude der Statistik, insbesondere beim “Test von Hypothesen” und der Parameterschätzung. Hier sollte kurz das Wort “Anpassung” (“Fit”) fallen. Einfaches Beispiel: Jemand hat an einem langen Draht eine Spannung angelegt und mißt nun den Strom als Funktion der Spannung. Offenbar liegen die Meßpunkte “mehr oder weniger” auf einer Geraden. Welches ist die “beste” Gerade, die ich durch die Meßpunkte legen kann, und was sind meine Kriterien? Wenn ich ein “Fit - Verfahren” für diese Prozedur kenne, kann ich das Result (offenbar den reziproken Ohmschen Widerstand des Drahtes) und auch noch dessen Unsicherheit angeben.
Wir sehen, es ist ein weites und spannendes Feld. Jede Messung hat es mehr oder weniger mit dem Zufall zu tun. Aber der Zufall hat einmal (natürlich rein zufällig) Herrn Gauß und Monsieur Poisson getroffen - Herrn Binom sei Dank.
Last but not least: Der Verfasser dankt seinem Nachbarn und Mitstreiter T. Berndt für tätige Hilfe bei der Erzeugung der Diagramme mit Hilfe des geheimisvollen Programmpaketes “Root”.