

ANLAGE VI  
STATISTIK

GPI

Statistische Schwankungen

Physikalische Vorgänge verlaufen aufgrund der quantelten Natur der Systeme stochastisch. Deterministisch werden Vorgänge in Strenge für den Grenzfall unendlich vieler oder praktisch bei einer großen Anzahl elementarer Beiträge; so dass die zu erwartenden Schwankungen gegen Null gehen. Dies ist die Situation der *makroskopischen* Physik, deren Aussagen sich z.B. auf "unendlich" viele Teilchen eines Gasvolumens oder auf "unendlich" viele schwingenden Ladungen einer strahlenden Antenne beziehen.

Für die Wertausprägungen von Systemen, die durch endlich viele oder eine vergleichsweise kleine Anzahl atomarer Systeme bestimmt wird, gib es keine streng deterministische Vorhersage; und die Ergebnisse zeigen (statistische) Schwankungen mit einer Verteilung der Werte.

Ist analog nicht die Zahl der mitwirkenden mikroskopischen Systeme, sondern die Zahl der Beobachtungen unendlich, so gibt es wieder eine strenge Vorhersage, jedoch nur für die *Parameter* der Verteilung, wie z.B. den Mittelwert oder die Standardabweichung (Schwankungsbreite).

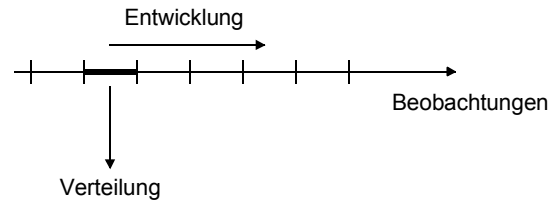
Zusätzlich treten die Schwankungen experimentell nur in Erscheinung, wenn sie aus der Empfindlichkeit der Meßapparatur herausragen. Eine reale Meßreihe liefert dann eine Näherung der erwarteten Verteilungsfunktion.

Klassische Beispiele auch aus dem Bereich des Grundpraktikums sind die Zerfälle angeregter atomarer oder Kernzustände, der radioaktive Zerfall oder die stochastische Wechselwirkung eines Teilchen- oder Strahlungsstroms mit Materie.

Typisch ist dabei die Betrachtung einer *begrenzten Beobachtung*, d.h. einer begrenzten Beobachtungszeit oder einer begrenzten Wechselwirkungsstrecke, für den dann der stochastische Prozess durch zeit- oder längenbezogene Wahrscheinlichkeiten beschrieben werden kann (Übergangswahrscheinlichkeit, Stoßwahrscheinlichkeit).

Mathematisch-statistisch stellt dies ein *Bernoulli-Experiment* mit zweiwertigem Ausgang dar: *Zerfall/Stoß* oder *nicht, JA* oder *NEIN, 0* oder *1, SCHWARZ* oder *WEISS, ADLER* oder *ZAHL*. In diesem Skript sollen Verteilungsfunktionen entwickelt werden, die sich im Zusammenhang mit derartigen *Bernoulli-Experimenten* ergeben.

Dabei gibt es zwei Fragestellungen. Einmal die Frage nach der Verteilung der Ergebnisse innerhalb einer solchen Beobachtung. Zum anderen die Entwicklung eines endlichen Ensembles aufgrund solcher Prozesse.

Exponentialentwicklung

Zunächst soll eine zeitliche Entwicklung am Beispiel des radioaktiven Zerfalls betrachtet werden (Versuch *RADIOAKTIVER ZERFALL*). Wenn die zeitbezogene Zerfallswahrscheinlichkeit (Übergangswahrscheinlichkeit) eines Kerns  $\lambda$  ist, so ergibt sich als mittlere Zahl der Zerfälle eines Ensembles von  $n$  radioaktiven Kernen in einem Beobachtungszeitraum  $\Delta t$ :

$$(1) \quad \bar{n} = p n = \lambda \Delta t n$$

Hinsichtlich einer endlichen Menge von Kernen stellt  $\bar{n}$  eine Abnahme dar. Näherungsweise soll die Menge als groß betrachtet, und  $n$  als kontinuierliche Variable aufgefasst werden. Dann ergibt sich aus (1) eine Differentialgleichung für die zeitliche Entwicklung des Ensembles:

$$(2) \quad dn = -\lambda n dt$$

mit der Lösung

$$(3) \quad n = n_0 e^{-\lambda t}$$

Dies ist das Zeitgesetz des radioaktiven Zerfalls, wobei man ein genau entsprechendes Ergebnis für den Fall eines Absorptionsgesetzes von Strahlung durch Materie erhält, wenn man modellmäßig stochastisch verteilte

Strahlungsquanten und Stoßpartner und "singuläre" Stoßprozesse annimmt, bei denen ein Strahlungsquant endgültig aus dem Strahlungsstrom entfernt wird.

Binomialverteilung

Während im vorgehenden Fall die Anzahl der Ereignisse bei einer Beobachtung "integral" durch den Mittelwert beschrieben wurde, soll jetzt "differentiell" nach der Verteilung dieser Anzahl gefragt werden. Zur Verallgemeinerung wird die Zerfallswahrscheinlichkeit für einen Beobachtungszeitraum  $\lambda \Delta t$  als  $p$  geschrieben, so dass die Wahrscheinlichkeit des Komplementäreignisses  $(1-p)$  ist. Bei  $N$  "Durchführungen" des Einzelexperiments (bei  $N$  Kernen) folgt aus dem Multiplikationsgesetz für statistisch unabhängige Ereignisse als Wahrscheinlichkeit, daß in  $n$  bestimmten Fällen das Ereignis (der Zerfall) eintritt, und in den übrigen  $(N-n)$  Fällen nicht:

$$(4) \quad p^n (1-p)^{N-n}$$

Jedoch interessiert im allgemeinen nicht, in welchen bestimmten Fällen das Ergebnis eintritt, sondern nur deren Gesamtzahl. Der Wert (4) ist dann mit der Zahl der Möglichkeiten zu multiplizieren, aus den  $N$  Experimenten  $n$  auszuwählen, in denen das Ereignis eintritt. Diese Zahl ist durch einen Binomialkoeffizient gegeben:

$$(5) \quad \binom{N}{n} = \frac{N!}{n! (N-n)!}$$

Als gesuchte Verteilung erhält man die *Binomialverteilung*:

$$(6) \quad B(n) = \frac{N!}{n! (N-n)!} p^n (1-p)^{N-n}$$

Die diskrete *Zufallsvariable*  $n$  beschreibt die tatsächlich möglichen Ergebnisse, und die Verteilung (6) direkt deren Wahrscheinlichkeitswerte (z.B. für die Anzahlen von Kernzerfällen in einem Beobachtungsintervall).

Die Binomialverteilung entspricht dem *Normierungssaxiom* für Wahrscheinlichkeiten. Die Summation über alle möglichen Ereignisse, d.h. über alle  $n$ , stellt gerade die binomische Potenz mit  $p$  und  $(1-p)$  als Nomen dar:

(7)

$$\sum_{n=0}^N B(n) = \sum_{n=0}^N \binom{N}{n} p^n (1-p)^{N-n} = [p + (1-p)]^N = 1^N = 1$$

Verteilungsparameter

Die Werte eines Verteilungsmodells, oder die Daten einer beobachteten oder erhobenen Verteilung, stellen die umfassende, aber umfangreiche und damit unbequeme Information einer Verteilung dar. Es sind daher *statistische Maßzahlen* für Verteilungen definiert, mit denen die wichtigen Eigenschaften quantitativ repräsentiert werden können.

(8)	Diskrete Verteilung F(x <sub>i</sub> )	Verteilungsfunktion ρ(x)
Lage- maß	Mittelwert $\bar{x} = \sum x_i F(x_i)$	Erwartungswert $\mu = \int x\rho(x)dx$
Streu- maß	Mittlere quadratische Abweichung $s^2 = \sum (x_i - \bar{x})^2 F(x_i)$	Varianz $\sigma^2 = \int (x - \mu)^2 \rho(x)dx$

Die Bildung der quadratischen Abweichungen als Streumaß berücksichtigt die unterschiedlichen Vorzeichen der Abweichungen (die mittlere Abweichung selbst würde bei einer symmetrischen Verteilung verschwinden). Die *Standardabweichung* ist die Wurzel aus der mittleren quadratischen Abweichung bzw. der Varianz, die damit in die Ursprungsdimension der Verteilungsvariablen zurückkehrt.

Die Berechnung von Mittelwert und Varianz der Binomialverteilung ist auf direktem Weg algebraisch aufwendig, aber einfach unter Anwendung der (plausiblen) Summenregeln für die Maßzahlen:

(9)

$$\overline{(x_i + y_i)} = \bar{x}_i + \bar{y}_i \quad \text{und} \quad s^2(x_i + y_i) = s^2(x_i) + s^2(y_i)$$

Die zweiwertige *Bernoulli-Verteilung* als Grenzfall der Binomialverteilung mit n = 1 hat die Werte n = 0 mit der Wahrscheinlichkeit (1-p) und n = 1 mit der Wahrschein-

lichkeit p. Als Mittelwert und Varianz folgen damit nach den Definitionen (8):

$$\bar{n} = 0(1-p) + 1p = p \quad \text{und}$$

$$s^2 = (0-p)^2(1-p) + (1-p)^2 p = (1-p)p$$

Und daraus mit (9) sofort für die Binomialverteilung mit N Ereignissen:

$$(10) \quad \bar{n} = pN \quad \text{und} \quad s^2 = (1-p)pN = (1-p)\bar{n}$$

Poisson-Verteilung

Die *Bernoulli-Verteilung* ist praktisch unhandlich wegen der Binomialkoeffizienten, die schnell sehr große Werte annehmen. Für "seltene" Ereignisse mit kleinen Mittelwerten und kleiner Wahrscheinlichkeit, die häufig vorliegen, und für Werte in der Umgebung des Mittelwerts lässt sich eine bequemere Näherungsverteilung entwickeln. Die Annahmen bedeuten:

$$(11) \quad \bar{n}, n \ll N \quad \text{und} \quad p \ll 1$$

Aus dem Binomialkoeffizienten wird (N-n)! herausgekürzt und die verbleibenden Faktoren im Zähler von (N-n+1) bis N näherungsweise gleich N gesetzt:

$$(12) \quad \frac{N!}{(N-n)!} \approx N^n \quad \text{und} \quad N^n p^n = \bar{n}^n$$

Die Potenz (1-p)<sup>N-n</sup> wird als Exponentialterm geschrieben, und der benötigte Logarithmus in der Umgebung von 1 linear genähert:

$$(13) \quad (1-p)^{N-n} = e^{\ln(1-p)^{N-n}} = e^{(N-n)\ln(1-p)} \approx e^{-pN} = e^{-\bar{n}}$$

wegen

$$(14) \quad \ln(1-p) \approx -p.$$

Die Bernoulli-Verteilung geht damit in die *Poisson-Verteilung* über, die nur noch durch den einen Parameter  $\bar{n}$  bestimmt wird:

$$(15) \quad P(n) = \frac{\bar{n}^n}{n!} e^{-\bar{n}}$$

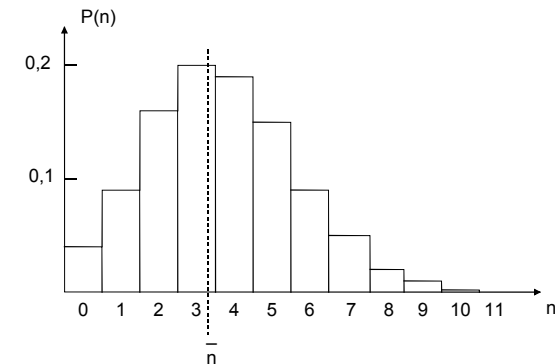
Auch die *Poisson-Verteilung* ist normiert:

$$(16) \quad \sum_{n=0}^{\infty} P(n) = 1 \quad \text{wegen} \quad \sum_{n=0}^{\infty} \frac{\bar{n}^n}{n!} = e^{\bar{n}}$$

Dabei wird der Gültigkeitsbereich von (15) abweichend von der Ausgangseinschränkung für n → ∞ angenommen, was wegen des starken Abfalls der Funktion P(n) zu großen Werten hin berechtigt ist. Für den Mittelwert berechnet sich analog zu (9):

$$(17) \quad \sum_{n=0}^{\infty} n P(n) = \bar{n}$$

Die folgende Abbildung zeigt eine Poisson-Verteilung für  $\bar{n} = 3,8$ .



Die *Poisson-Verteilung* ist unsymmetrisch. Mit wachsendem  $\bar{n}$  geht sie aber in eine Verteilung über, die in der Umgebung von  $\bar{n}$  symmetrisch zu diesem Wert wird.

Gauß-Verteilung

Die *Bernoulli-* und die *Poisson-Verteilung* sind diskrete Verteilungen, wobei es in vielen Fällen praktisch vorteilhaft ist, mit einer kontinuierlichen Verteilungsfunktion arbeiten zu können.

Zur Entwicklung einer Funktion aus der *Poisson-Verteilung* soll die Forderung (11) dadurch ergänzt werden, dass der Mittelwert zwar klein gegen N, aber dennoch groß gegen 1 sein soll:

$$(18) \quad \bar{n} \gg 1$$

nd die Überlegungen auf eine Umgebung von  $\bar{n}$  beschränkt werden, so dass:

$$(19) \quad |n - \bar{n}| \ll \bar{n}$$

Der Übergang zu einer kontinuierlichen Verteilungsfunktion geschieht durch Aufstellung eines Differenzenquotienten für die diskrete Poisson-Verteilung, die mit den Näherungsannahmen als Differentialgleichung aufgefasst werden kann:

$$(20) \quad \frac{P(n) - P(n-1)}{\Delta n = 1} = P(n) - P(n-1) = -\frac{n - \bar{n}}{\bar{n}} P(n)$$

Der rechte Ausdruck in (20) stellt wegen der Voraussetzung (18) eine mit  $n$  nur langsam variierende Größe dar, so dass  $n$  dann als kontinuierliche Variable aufgefasst (und deshalb jetzt mit  $x$  bezeichnet wird), und (20) als Differentialgleichung einer Verteilungsfunktion  $G(x)$  interpretiert werden kann:

$$(21) \quad \frac{dG(x)}{dx} = -\frac{x - \bar{x}}{\bar{x}} G(x)$$

Sie hat die Lösung:

$$(22) \quad G(x) = C e^{-\frac{(x - \bar{x})^2}{2\bar{x}}}$$

Das ist die *Gauß-* oder *Normalverteilung*. Die Integrationskonstante wird so bestimmt, daß die Funktion wieder normiert ist. Mit dem Integral:

$$(23) \quad \int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi} \quad \text{folgt} \quad C = \frac{1}{\sqrt{2\pi \bar{x}}} \quad \text{bzw.}$$

$$(24) \quad G(x) = \frac{1}{\sqrt{2\pi \bar{x}}} e^{-\frac{1}{2} \left( \frac{x - \bar{x}}{\sqrt{\bar{x}}} \right)^2}$$

Durch den Übergang von einer diskreten zu einer kontinuierlichen Zufallsvariablen werden die einzelnen ( $x$ -) Werte infinitesimal, und deren Wahrscheinlichkeiten gehen gegen Null. Tatsächliche Ereignisse werden jetzt durch Intervalle der Zufallsvariablen gebildet (mit einem durch die Breite bestimmten "Volumen"), deren Wahrscheinlichkeit dann durch die Fläche der Verteilung über dem Intervall beschrieben wird. Die Funktion (24)

stellt die *Wahrscheinlichkeitsdichte* der Zufallsvariablen dar.

#### Allgemeine Normalverteilung

Die Form (24) ist eine spezielle Normalverteilung mit nur einem Parameter, wobei  $\bar{X}$  Erwartungswert (in trivialer Weise wegen der Symmetrie der Verteilung) und Varianz gleichzeitig ist. Die allgemeine Form der Normalverteilung mit unabhängigen Werten von Erwartungswert  $\mu$  und Varianz  $\sigma^2$  lautet:

$$(25) \quad N(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2}$$

In der folgenden Abbildung sind zwei Normalverteilungen zum gleichen Erwartungswert ( $\mu = 5,0$ ), aber unterschiedlichen Varianzen bzw. Standardabweichungen dargestellt ( $\sigma = 0,5$  und  $S = 1,5$ ).

