

*Version Datum: 29/11/07*

---

Stochastische Beschreibung physikalischer Systeme

Prof. Reiner Kree

---

Göttingen WS 2007/2008

## Inhaltsverzeichnis

<b>1</b>	<b>Wahrscheinlichkeit</b>	<b>4</b>
1.1	Die Sache mit dem Zufall . . . . .	4
1.2	Der klassische Wahrscheinlichkeitsbegriff . . . . .	5
1.3	Zufallsvariablen . . . . .	8
1.4	Gemeinsame W., bedingte W., Korrelation . . . . .	11
1.5	Gesetz der großen Zahlen und zentraler Grenzwertsatz . . . . .	14
1.6	Einige gebräuchliche ZV . . . . .	19
1.7	Datenanalyse . . . . .	21
1.8	Erzeugung von Zufallszahlen auf dem Computer . . . . .	30
<b>2</b>	<b>Zustände und Ensembles</b>	<b>34</b>
2.1	Zeitentwicklung reiner Zustände . . . . .	34
2.2	Observable . . . . .	35
2.3	Ensembles und gemischte Zustände . . . . .	36
2.4	Zeitentwicklung gemischer Zustände . . . . .	37
2.5	Verkürzung von Zuständen . . . . .	40
<b>3</b>	<b>Langevin Gleichung des Brownschen Teilchens</b>	<b>42</b>
3.1	Mittleres Auslenkungsquadrat. Diffusion . . . . .	45
3.2	Überdämpfte Brownsche Bewegung mit äußerer Kraft . . . . .	47
<b>4</b>	<b>Markov-Prozesse und Mastergleichung</b>	<b>48</b>
4.1	Detaillierte Balance und das Streben ins Gleichgewicht . . . . .	56
4.2	Anwendungsbeispiel: Die chemische Mastergleichung . . . . .	58
<b>5</b>	<b>Fokker-Planck-Gleichung und Langevin Prozesse</b>	<b>61</b>
5.1	Die Kramers-Moyal Entwicklung . . . . .	61
5.2	Inkremete von Langevin Prozessen . . . . .	64
5.3	Transformation von Langevin Prozessen. Die Ito Formel . . . . .	67
<b>6</b>	<b>Die Tücken des Modellierens mit Langevingleichungen</b>	<b>68</b>
<b>7</b>	<b>Monte Carlo Simulation von Mastergleichungen</b>	<b>74</b>
7.1	Gillespie's Algorithmus . . . . .	75
7.2	Algorithmus mit festem Zeitschritt . . . . .	76

## Hinweis

Dieses Skript soll den 1. Teil meiner Vorlesung "Statistische Physik und Thermodynamik" unterstützen. Es ist sozusagen die "Version 0.1.0". Es hat alle Vor- und Nachteile eines Skripts. Vorteil: Es ist sehr nah am tatsächlich gelesenen Stoff<sup>1</sup>. Nachteil: Es ist kein Lehrbuch, d.h. möglicherweise noch tippfehlverseucht, und nicht mit jeder Menge netter Bildern und geschliffener, didaktischer Darstellung versehen. Also, — nutzen Sie es weise! Für Hinweise auf Fehler jeder Art und Vorschläge zur Verbesserung bin ich immer dankbar. Vielleicht wird so mal eine "Version 1.0.0" daraus.

Reiner Kree

---

<sup>1</sup>Ein paar Dinge wurden hinzugefügt, insbesondere dann, wenn von Seiten der Studierenden entsprechende Fragen gestellt wurden. Einige Dinge, die in Übungsaufgaben ausführlich bearbeitet wurden, sind hingegen weggelassen

# 1 Wahrscheinlichkeit

## 1.1 Die Sache mit dem Zufall

N Wiederholungen desselben physikalischen Experiments ergeben i.d.R. verschiedene (“zufällige”) Ergebnisse. *Statistische Physik* beschäftigt sich mit solchen Systemen, bei denen die zufälligen Ergebnisse aufgrund physikalischer Gesetze zustande kommen (und nicht, z.B., durch Schlamperei des Experimentators). Dabei gibt es unterschiedliche Gründe für den Zufall:

- 1) Es gelingt aus prinzipiellen Gründen nicht, den Anfangszustand, in dem das Experiment ausgeführt wird, hinreichend genau festzulegen.
- 2) Während der Messungen unterliegt das System äußeren Störungen, die sich prinzipiell nicht “abschirmen” lassen.
- 3) Das Experiment führt Messungen an einem Quantenzustand aus

Obwohl die statistische Physik sich mit “Zufällen” aus allen 3 Gründen befasst, muss deutlich zwischen dem “klassischen Zufall” (1) und (2) und dem “Quantenzufall” (3) unterschieden werden. Insbesondere können nicht alle Aspekte des “Quantenzufalls” mit der in diesem Abschnitt vorgestellten klassischen Wahrscheinlichkeitstheorie beschrieben werden. Andernfalls könnte man ja eine klassische, stochastische Beschreibung formulieren, die äquivalent zu Quantenmechanik wäre und das geht aus sehr grundsätzlichen Erwägungen schief (s. “Bellsche Ungleichung”). Als Daumenregel merke man sich: Eigenschaften, die durch *mehrere, nicht-kommutative Observablen* ausgedrückt werden, folgen nicht (!) der klassischen W.theorie. Genaueres dazu später.

Damit (1) und/oder (2) auch wirklich zu zufälligen Ergebnissen führen, muss das System sehr empfindlich auf kleine Zustandsänderungen reagieren, es muss, etwas technischer gesagt, eine *chaotische Dynamik* besitzen, bei der kleine Änderungen seines Zustands  $x$ ,  $\delta x_0$  durch die interne Dynamik stets stark (exponentiell) anwachsen  $\delta x_t = \delta x_0 \exp(\lambda t)$  ( $\lambda$  wird in der nichtlinearen Dynamik als *Lyapunov-Exponent* bezeichnet.) Die erzeugten Abweichungen müssen außerdem nach kurzer Zeit (Messzeit) so groß sein, dass sie zu unterschiedlichen Messergebnissen führen. Es ist eines der Grundprobleme der Physik zu verstehen, warum sich eigentlich “fast alle Systeme” (was immer das heißt) so verhalten. Beachten Sie, dass die typischen Lehrbuchbeispiele (harmonischer Oszillator) dieses Verhalten nicht zeigen! Eine Störung der Bahn eines harmonischen Oszillators wächst nicht in der Zeit. Diese Gegenbeispiele müssen in einem bestimmten Sinn “atypische Ausnahmen” sein.

## 1.2 Der klassische Wahrscheinlichkeitsbegriff

Aus Experimenten erhält man eine *Messwertstatistik*, aus dieser lassen sich "Frequenzen" (wie der Statistiker sagt) bzw. "relative Häufigkeiten" gewinnen. Man beobachtet "Ereignisse" (z.B. die Zahl der Augen beim Würfeln oder die Anzahl der Klicks in einem Teilchenzähler,...). Dann ist die relative Häufigkeit  $r(E)$  eines Ereignisses "E" gerade

$$r(E) = \frac{\text{Anzahl der Beobachtungen von E}}{\text{Gesamtzahl der Messungen}}$$

Eigentlich ist  $r(E)$  selber eine zufällige Größe, denn sie hängt ja von den zufälligen Messergebnissen ab. Erhöht man aber die Zahl der Messungen, so stellt man fest, dass  $r(E)$  immer weniger "schwankt", so dass man zu der Auffassung kommt, im Limes unendlich vieler Messungen erhält man einen nicht-zufälligen Wert

$$\lim_{N \rightarrow \infty} r(E) = P(E)$$

Auf diese Weise ermittelt die Physik (und alle anderen, quantitativ messenden Wissenschaften) Wahrscheinlichkeiten  $P(E)$  für das Auftreten von Ereignissen. Man könnte nun vielleicht denken, dass eine *mathematische Wahrscheinlichkeitstheorie* diese "Frequenzdefinition" irgendwie präzisiert. Dem ist nicht so (obwohl frühe Ansätze das versucht haben). Die **Ziele der mathematischen W.Theorie** sind vergleichsweise bescheiden:

- a) klärt, wie man einer ganzen Menge von Ereignissen *widerspruchsfrei* Wahrscheinlichkeiten zuordnen kann
- b) untersucht, wie man W. von einer Menge von Ereignissen vermittelt geeigneter Abbildungen auf andere Mengen von Ereignissen "transportiert".

Insbesondere versucht sie Eines nicht: *sie zaubert keine Regeln zur Berechnung von W. ab initio!* Das ist ihre Stärke und ihre Schwäche.

Die statistische Physik bedient sich der Ergebnisse der W.theorie, um aus "mikroskopischen" physikalischen Theorien (klassische Mechanik, Quantenmechanik, Elektrodynamik, ...) Messwertstatistiken (insbesondere für Experimente an Systemen mit sehr vielen Freiheitsgraden) vorherzusagen.

Also lohnt sich schon ein kurzer Seitenblick auf die Mathematik. Was ist denn nun im Sinne der mathematischen Theorie überhaupt eine Wahrscheinlichkeit? Um dies zu beantworten, brauchen wir ein paar Begriffe. Da ist zunächst

die Formalisierung des *Zufallsexperiments*. Es ist gekennzeichnet durch einen Raum  $S$ , der alle möglichen Ergebnisse des Experiments enthält.  $S$  kann Zahlen als Elemente haben, aber auch Farben, Orte, Töne, .... Elemente aus  $S$  bezeichnen wir hier mal mit  $\omega \in S$ .

Haben wir  $S$ , können wir den Begriff "Ereignis" genauer fassen. Ereignisse  $E$  sind (bestimmte, s.u.) Teilmengen von  $S$ , also  $E \subseteq S$ . Liegt das Ergebnis  $\omega$  einer Realisierung des Zufallsexperiments in  $E$ , so sagen wir: "E ist eingetroffen". Bevor wir der Frage nachgehen, welche "bestimmten Teilmengen" denn nun Ereignisse sein können (und welche nicht), wollen wir die 3 definierenden Eigenschaften der  $W$ . hinschreiben. Genau diese 3 Eigenschaften wünschen wir uns für ordentliche  $W$ .:

$$(i) P(E) \geq 0$$

$$(ii) P(S) = 1$$

(iii) Seien  $E_n$  höchstens abzählbar viele Ereignisse.

$$P\left(\bigcup_n E_n\right) = \sum_n P(E_n)$$

für paarweise disjunkte Ereignisse, d.h.  $E_n \cap E_m = \emptyset$  für alle  $n \neq m$

Alle drei Eigenschaften sind wichtig, aber die ersten beiden sind ja leicht zu durchschauen mit unserem intuitiven Begriff der  $W$ :  $W$ . sind Zahlen zwischen 0 und 1 und die  $W$ . irgendein Ergebnis bei einem Experiment zu erzielen, ist genau 1.  $S$  selbst ist also auch ein Ereignis, und zwar das "sichere Ereignis" (trifft immer ein). Die für alle Schlussfolgerungen wichtigste Eigenschaft ist (iii), sie ist das Herzstück der axiomatischen Definition, also mache man sich wirklich ganz (!) klar, dass auch diese Eigenschaft völlig in Übereinstimmung mit dem "intuitiven  $W$ . begriff" ist (wenn man mal von der Tatsache absieht, dass hier unendlich viele Ereignisse zugelassen werden). Die  $W$ . das eines von 2 ganz unabhängigen (d.h. disjunkten) Ereignissen eintritt ist gerade die Summe der  $W$ . für das Eintreffen der beiden Einzelereignisse.

Nun wenden wir uns wieder der Frage zu: Was ist ein System von Teilmengen von  $S$ , denen man widerspruchsfrei  $W$ . zuordnen kann? Ein solches, logisch konsistentes System von Ereignissen bezeichnet man auch (vor allem in der Mathematik) als  $\sigma$ -Algebra. Der Schlüssel, solche Systeme zu konstruieren ist für Physiker ganz einfach zu verstehen;

Ereignisse muss man messen können. Hat man aber einmal Messgeräte für zwei Ereignisse  $E_1$  und  $E_2$  konstruiert, so kann man natürlich daraus durch logische Verknüpfungen Messgeräte für weitere Ereignisse basteln, insbesondere:

- $E_1$  UND  $E_2$  fragt ab, ob  $\omega \in E_1 \cap E_2$
- $E_1$  ODER  $E_2$  fragt ab, ob  $\omega \in E_1 \cup E_2$
- NOT  $E_1$  fragt ab, ob  $\omega \in \bar{E}_1$  (Komplement)

womit also klar ist, dass mit  $E_1, E_2$  auch  $E_1 \cap E_2, E_1 \cup E_2$  und  $\bar{E}_1, \bar{E}_2$  Ereignisse sein müssen. Genau das ist die definierende Struktur einer  $\sigma$ -Algebra. So ist garantiert, dass alles "postprocessing" mit Hilfe Boolescher Algebra immer zu konsistenten Aussagen über wirklich existierende Ereignisse führt.

Schließlich kommen wir zur Antwort auf die erste, oben genannte Grundfrage der W. theorie. Wie ordnet man nun einer  $\sigma$ -Algebra widerspruchsfrei W. zu? Hier gibt es konzeptionell wichtige Unterschiede zwischen  $S$  mit höchstens abzählbar vielen Elementen und  $S$ , die Kontinua sind.

Ist  $S$  höchstens abzählbar ("Würfel"), so bildet die Potenzmenge (Menge aller Teilmengen) die größte  $\sigma$ -Algebra, und deren Elementen kann man konsistent W. zuordnen, indem man einfach jedem Element  $\omega_i \in S, i = 1, 2, \dots$  eine W.  $P(\omega_i) = P_i$  zuordnet, so dass  $0 \leq P_i \leq 1$  und  $\sum_i P_i = 1$  ist. Da jede Teilmenge sich aus disjunkten Einzelelementen zusammenbasteln lässt, liegen die Wahrscheinlichkeiten aller Ereignisse (unter Benutzung von definierender Eigenschaft (iii)) damit fest, und sie sind alle miteinander konsistent. Manchmal benutzt man aus praktischen Erwägungen auch kleinere  $\sigma$ -Algebren. Wenn man einen Würfel benutzen will, um eine Ja oder Nein Entscheidung zu treffen, kann man z.B. alle geraden Zahlen mit "Ja" und alle ungeraden mit "Nein" identifizieren. Diese beiden Teilmengen bilden zusammen mit ganz  $S$  und der leeren Menge wieder eine  $\sigma$ -Algebra.

Ist  $S$  hingegen ein Kontinuum, ist die Sache komplizierter. Betrachten wir als Musterbeispiel das Kontinuum der reellen Zahlen. Eine "natürliche"  $\sigma$ -Algebra können wir aus den abgeschlossenen Intervallen  $[a, b]$  gewinnen durch Anwendung aller möglichen Vereinigungen, Durchschnitte und Komplementbildungen.<sup>2</sup> Wie kann man nun diesen unübersichtlich vielen Mengen konsistent W. zuordnen? z.B. so, wie es für Physiker am wichtigsten ist: Man betrachtet Intervalle  $I_x = (-\infty, x)$ . Die W.  $P(I_x)$  ist eine bezüglich  $x$  von 0 (für

<sup>2</sup>Für diejenigen, die in Mathematikvorlesungen das Lebesguesche Integral kennengelernt haben: Ja, hier handelt es sich um die Borel Mengen und überhaupt hat die Zuordnung von W. im Kontinuum eine extrem enge Verwandtschaft zur Integrations- sprich Maßtheorie. In der W. theorie kommt nur die Bedingung: "Das Maß des ganzen Raums ist immer gleich 1" hinzu.

$I_{-\infty}$ ) auf 1 (für  $I_{+\infty} = \mathbb{R}$ ) monoton wachsende Funktion (Machen Sie sich klar, warum das so sein muss. Es folgt aus der axiomatischen Definition der  $W$ .) Umgekehrt kann man mit Hilfe einer monoton von 0 auf 1 wachsenden Funktion auf der reellen Achse allen Mengen der "natürlichen"  $\sigma$ -Algebra konsistent  $W$  zuordnen (Wenn Sie von einer Polizeistreife angehalten werden und von Ihnen ein Beweis dieser Tatsache verlangt wird, sollten Sie jetzt notfalls in der Lage sein, den zu liefern).

Beachten Sie, dass auch die einzelnen reellen Zahlen Elemente der  $\sigma$ -Algebra sind, allerdings wird ihnen für glatte Funktionen  $P(I_x)$  stets die  $W$ . 0 zugeordnet, d.h. jedes mögliche Ergebnis  $\omega$  des Zufallsexperiments tritt mit  $W$ . Null auf! Erst kleine Intervalle können mit endlicher  $W$ . realisiert werden. Weniger einsichtig mag Ihnen dagegen erscheinen, dass es im Kontinuum auch Teilmengen gibt, die definitiv keine Ereignisse sein können (ganz anders als im Fall eines abzählbaren  $S$  !)

Hier ein in der Mathematik sehr übliches Beispiel: Wir betrachten Zufallsexperimente mit Ergebnissen in  $S = [0, 1]$  und die  $W$ . eines Intervalls  $P([a,b])=b-a$ . Für Ergebnisse  $x, y$  in  $S$  definieren wir eine Äquivalenzrelation:  $x$  und  $y$  seien äquivalent ( $x \sim y$ ), wenn  $x-y \in \mathbb{Q}$ , also eine rationale Zahl ist. Da es abzählbar viele rationale Zahlen gibt, enthält jede Äquivalenzklasse also auch abzählbar viele Elemente. Da alle Äquivalenzklassen zusammen wiederum ganz  $S$  ausfüllen, muss es überabzählbar viele solche Klassen geben. Jetzt bilden wir unsere "verrückten Teilmengen", indem wir aus jeder Äquivalenzklasse einen Repräsentanten auswählen. So eine Menge enthält also überabzählbar viele Elemente. Andererseits kann man ganz  $S$  in abzählbar viele solche Mengen zerlegen (entsprechend den abzählbar vielen Elementen, die in jeder Äquivalenzklasse als Repräsentanten zur Verfügung stehen.) Nun versuchen wir den "verrückten Teilmengen"  $W$ . zuzuordnen. Wäre die zugeordnete  $W$ . ungleich Null ( $p_0$ ) (wenn für eine Teilmenge, dann für jede), so wäre die Gesamtw. von  $S = \infty \cdot p_0$ , also sicher nicht 1. Wählen wir hingegen 0 als  $W$ . der "verrückten Teilmenge" so wäre die Gesamtw. von  $S$  gerade 0! Leben Sie also mit der Tatsache, dass nicht jede Teilmenge auch ein Ereignis sein kann!

### 1.3 Zufallsvariablen

Als messende Physikerinnen und Physiker haben wir es ständig mit folgender Situation zu tun: Wir untersuchen ein System, in dem sich zufällige Ereignisse abspielen (hier kein Quantenzufall!). Diese Ereignisse messen wir, d.h. genauer, wir bilden sie mittels eines Messgerätes auf eine "Messskala" ab. Meist besteht diese Skala aus reellen Zahlen (muss aber nicht, sie könnte auch

aus Farben einer Pseudocolorierung, Tönen oder sonst was bestehen). Aus der Messwertstatistik (abgelesen auf der Messskala) möchten wir dann Aussagen über die Statistik der zufälligen Ereignisse im System machen. Damit das auch wirklich funktioniert, muss die Abbildung, die unser Messgerät von den Ereignissen auf unsere Messskala vermittelt, eine wichtige Eigenschaft haben:

Jedem auf der Messskala beobachteten Ereignis muss auch ein Ereignis im System entsprechen

Ist also  $E_M$  ein Ereignis der Messskala (die möglichen Messwerte liegen in  $S_M$ ), so müssen wir ganz sicher sein können, dass immer, wenn auf dem Messgerät " $E_M$  ist eingetroffen" blinkt, im System die beobachtete Größe <sup>3</sup>  $\omega$  auch tatsächlich in genau einem Ereignis  $E_S$  liegt. Mathematisch ausgedrückt:

Die Urbilder von Ereignissen müssen Ereignisse sein!

Abbildungen  $X : S \rightarrow S_M$ , die diese Eigenschaft haben, können die W.  $P(E_S)$  konsistent, vollständig und widerspruchsfrei vom System  $S$  auf den Bildraum  $S_M$  (sprich die Messskala) transportieren durch

$$E_M = X(E_S) \rightarrow P(E_M) = P(X(E_S))$$

Solche Abbildungen heißen *messbar*. Für den wichtigen Spezialfall, dass  $S_M \subseteq \mathbb{R}$ , spricht man von einer *reellen Zufallsvariablen* (ZV)  $X$ . Eine ZV lässt sich vollständig charakterisieren, indem man (nach dem im letzten Abschnitt angegebenen Rezept) den Intervallen  $I_x$  in  $S_M \subseteq \mathbb{R}$  die W. der Urbilder der  $I_x$  in  $S$  zuordnet., also

$$F_X(\lambda) := P(\{\omega \in S | X(\omega) < \lambda\})$$

Die Funktion  $F_X$  wird als *Verteilungsfunktion* der ZV  $X$  bezeichnet. Die Ableitung der Verteilungsfunktion

$$\rho_X(\lambda) = \frac{dF_X(\lambda)}{d\lambda}$$

bezeichnet man als die *Wahrscheinlichkeitsdichte* (WD) der ZV  $X$ .

Wir weisen hier noch auf einen nicht selten auftretenden Spezialfall hin:  $S$  muss kein Kontinuum sein, sondern kann auch höchstens abzählbar viele Elemente beinhalten. Dann hat  $F_X(\lambda)$  die Gestalt einer aufsteigenden Treppe, deren Stufenhöhen gerade die W.  $P(\omega_i)$  sind. In diesem Fall besteht die WD aus einer Summe von  $\delta$ -Funktionen

$$\rho_X(\lambda) = \sum_i P(\omega_i) \delta(\lambda - X(\omega_i)).$$

<sup>3</sup>I.a. keine Zahl, sondern Element der "physikalischen Wirklichkeit" also ein Ort, eine Energie, eine Farbe, eine Molekülkonformation,...

Sprechweise: Man spricht häufig davon, dass eine ZV "soundso verteilt" ist, oder sogar "nach  $\rho_X$  verteilt". In der physikalischen Praxis sind WD viel gebräuchlicher als Verteilungsfunktionen.

In der Praxis kennt man meist nicht die gesamte Verteilungsfunktion oder WD (es sei denn aus einer Theorie), sondern nur gewisse Charakteristika. Die wichtigsten sind:

- a) **Erwartungswerte**  $\langle g(X) \rangle := \int dx \rho_X(x)g(x)$  insbesondere
- b) **Mittelwert** der ZV  $\langle X \rangle = \int dx \rho_X(x)x$  und
- c) **Momente** der ZV  $\langle X^n \rangle, n \in \mathbb{N}$ . Man beachte, dass Mittelwert und Momente nicht unbedingt existieren müssen! So hat z.B. die in der Physik übliche *Lorenz Verteilung* ("Linienform")

$$\rho_X(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + x^2}$$

gar keine endlichen Momente.

- d) **Varianz**  $var(X) = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2$  gibt ein Mass für Fluktuationen um den Mittelwert, das tun auch die
- e) **zentrale Momente**  $\langle (X - \langle X \rangle)^n \rangle$
- f) **Quantile**  $x_\alpha$  sind die Stellen, für die  $P(X < x_\alpha)$  gerade gleich  $\alpha$  ist. Also Werte links von  $x_\alpha$  treten mit W.  $\alpha$  auf, Werte rechts von  $x_\alpha$  mit  $1 - \alpha$ . Das spezielle Quantil  $x_{1/2}$  heißt Median. Manchmal führt man auch Intervalle  $[x_0 - \Delta x_\alpha, x_0 + \Delta x_\alpha]$  (z.B. um den Maximumswert, oder um den Mittelwert) ein, in denen  $x$  mit W.  $\alpha$  realisiert wird.

Die Momente einer WD kann man elegant aus **Momente erzeugenden Funktionen** erhalten. Die bekannteste ist die Fouriertransformierte der WD, die **charakteristische Funktion der ZV** heisst:

$$\phi_X(k) := \langle e^{ikx} \rangle = \int \rho_X(x)e^{ikx}$$

Die Momente erzeugt man aus  $\phi_X(k)$  durch Ableitungen bei  $k = 0$

$$\langle X^n \rangle = (-i)^n \left. \frac{d^n}{dk^n} \phi_X(k) \right|_{k=0}$$

$\phi_X(k)$  ist für viele Rechnungen sehr nützlich, wir werden ihr noch wiederholt begegnen. Leitet man statt  $\phi_X$  den Logarithmus  $\ln \phi_X$  ab, so erzeugt man Kombinationen aus Momenten, die als **Kumulanten** bekannt sind.

$$\kappa_n := (-i)^n \left. \frac{d^n}{dk^n} \ln \phi_X(k) \right|_{k=0}$$

$$\kappa_1 = \langle X \rangle, \quad \kappa_2 = var(X), \quad \kappa_3 = \langle X^3 \rangle - 3 \langle X \rangle \langle X^2 \rangle + 2 \langle X \rangle^3$$

## 1.4 Gemeinsame W., bedingte W., Korrelation

Statt einer reellen Zufallsvariable können wir auch ein ganzes n-Tupel von Zufallsvariablen (Zufallsvektor) betrachten

$$\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

Dadurch entstehen einige neue Begriffe und Konzepte, die wir jetzt kurz einführen.

### Gemeinsame W.

Für 2 Ereignisse  $E_1, E_2$  ist  $P(E_1 \cap E_2)$  die W., dass ein Ergebnis  $\omega$  sowohl in  $E_1$  als auch in  $E_2$  liegt, also, dass beide Ereignisse eingetroffen sind. Entsprechend können wir für 2 Zufallsvariablen  $X_1$  und  $X_2$  die *gemeinsame Verteilungsfunktion* definieren als

$$F_{X_1 X_2}(\lambda_1, \lambda_2) = P(\{\omega \mid X_1(\omega) < \lambda_1\} \cap \{\omega \mid X_2(\omega) < \lambda_2\}),$$

also als W., dass ein Zufallsergebnis  $\omega$  Werte für  $X_1$  ergibt, die kleiner sind als  $\lambda_1$  und gleichzeitig Werte für  $X_2$ , die kleiner sind als  $\lambda_2$ . Entsprechend kann man die *gemeinsame WD* für  $X_1$  und  $X_2$  aus

$$\rho_{X_1 X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1 X_2}(x_1, x_2)}{\partial x_1 \partial x_2}$$

erhalten. Eine wichtige, wenn auch triviale Aussage über gemeinsame W. folgt aus der Tatsache, dass eines der Ereignisse  $E_1, E_2$ , die man zusammen betrachtet auch das sichere Ereignis  $S$  sein kann. Insbesondere ist  $P(E_1 \cap S) = P(E_1)$ . Stellt man nun  $S$  als die Vereinigung einer Menge von paarweise disjunkten Ereignissen dar,  $S = \cup_j \hat{E}_j$ , so kann man mithilfe der definierenden Eigenschaft (iii) der W. (s.o.) die interessanten Relationen

$$P(E) = P\left(E \cap \left(\bigcup_j \hat{E}_j\right)\right) = \sum_j P(E, \hat{E}_j)$$

erhalten. Übertragen wir das auf die WD von ZV so ist

$$\rho_{X_1}(x_1) = \int dx_2 \rho_{X_1 X_2}(x_1, x_2)$$

Die Verallgemeinerung dieser Begriffe von 2 auf n Variablen ist eine einfache Notationsübung, die wir hier weglassen. Statt dessen merken wir an, dass Physiker die hier gebrauchte Notation auf die Dauer lästig finden. Für sie ist es

natürlich, eine Abbildung und ihre Werte mit demselben Buchstaben zu bezeichnen (siehe  $y = y(t)$ , seit Newton beliebt). Daher neigen sie dazu, gemeinsame WD so zu schreiben

$$\rho(x_1, x_2, \dots, x_n)$$

und meinen damit natürlich  $\rho_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ . Allgemein lassen sich also aus gemeinsamen WD für  $k$  ZV die WD für  $k - 1$  oder  $k - 2$  oder  $k - l$  ZV gewinnen, indem man über die anderen integriert. Man nennt diese "teilweise ausintegrierten" WD auch *Randdichten* zur ursprünglichen WD.

### Bedingte W.

Betrachten wir mal wieder zunächst 2 Ereignisse  $E_1$  und  $E_2$  sowie  $P(E_1 \cap E_2)$  wie oben. Nun stellen wir uns vor, wir wüssten bereits mit Sicherheit, dass  $E_2$  eintritt und fragen, mit welcher W. trifft  $E_1$  unter dieser Bedingung ein. Da die W. für das Eintreffen von  $E_2$  gerade  $P(E_2)$  ist, können wir diese *bedingte W* wie folgt mit der gemeinsamen W. verknüpfen:

$$P(E_1 \cap E_2) = P(E_1|E_2) P(E_2)$$

Lies: Die gemeinsame W. dass  $E_1$  und  $E_2$  eintreffen, ist gleich der W., dass  $E_2$  eintritt, multipliziert mit der bedingten W., dass  $E_1$  eintritt, wenn  $E_2$  sicher ist. In der Tat ist dies die Definition der bedingten W.  $P(E_1|E_2)$ . Man beachte die Notation, die wir später extensiv verwenden: Variablen und Bedingungen sind durch senkrechten Strich voneinander getrennt. Von der bedingten W. gelangen wir zur *bedingten WD* von ZV. So ist beispielsweise

$$\rho(x_1, x_2) = \rho(x_1|x_2)\rho(x_2)$$

die definierende Relation für die WD, dass wir  $X_1$  bei  $x_1$  finden, falls mit Sicherheit  $X_2 = x_2$  ist<sup>4</sup>. Haben wir mehr als 2 ZV, können wir auch weitere, verschiedene Typen von Bedingungen stellen, z.B. ist

$$\rho(x_1|x_2, x_3)$$

---

<sup>4</sup>Nun ja, für alle die W.theorie gehört haben (alle Anderen weghören): Es gibt da natürlich noch ein kleines Problem, das darin besteht, dass die Bedingung, nämlich  $X_2 = x_2$  ja die W. Null hat. Händewedelnd könnte man sich zunächst mal die Bedingung durch  $X_2 \in [x_2 - \epsilon, x_2 + \epsilon]$  ersetzt denken, und dann an jeder geeigneten Stelle den  $\epsilon \rightarrow 0$  Limes betrachten. Tatsächlich gibt es in der W.theorie eine elegante Verallgemeinerung der elementaren bedingten W., die sogenannte **bedingte Erwartung**, die einem diese Klimmzüge erspart. Siehe jeden Standardtext zur W.theorie

die WD für  $X_1$  unter den Bedingungen  $X_2 = x_2$  und  $X_3 = x_3$ , also ist

$$\rho(x_1, x_2, x_3) = \rho(x_1|x_2, x_3)\rho(x_2, x_3).$$

Schematisch gilt ganz allgemein:

$$\rho(x_1, \text{Bedingung}) = \rho(x_1 | \text{Bedingung})\rho(\text{Bedingung}).$$

Es kann natürlich vorkommen, dass  $\rho(x_1 | \text{Bedingung})$  ganz unabhängig von der Bedingung ist, d.h.  $\rho(x_1 | \text{Bedingung}) = \rho(x_1)$ , dann sagt man,  $X_1$  **ist statistisch unabhängig von** *Bedingung*. Insbesondere heißt  $X_1$  statistisch unabhängig von  $X_2$  falls

$$\rho(x_1|x_2) = \rho(x_1).$$

Für statistisch unabhängige ZV gilt offenbar

$$\langle g(X_1)h(X_2) \rangle = \langle g(X_1) \rangle \langle h(X_2) \rangle$$

Andernfalls sind die ZV statistisch abhängig. Dann braucht man quantitative Maße für die statistische Abhängigkeit. Die für praktische Zwecke wichtigsten Abhängigkeitsmaße beruhen auf der **Kovarianz**

$$\text{cov}(X_1, X_2) := \langle (X_1 - \langle X_1 \rangle)(X_2 - \langle X_2 \rangle) \rangle = \langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle$$

Für statistisch unabhängige ZV ist  $\text{cov}(X_1, X_2) = 0$ . In der Literatur der Statistik heißt

$$\text{cor}(X_1, X_2) := \frac{\text{cov}(X_1, X_2)}{[\text{var}(X_1)\text{var}(X_2)]^{1/2}}$$

die **Korrelation**. Offenbar ist

$$|\text{cor}(X_1, X_2)| \leq 1$$

und die Korrelation ist betragsmäßig gleich 1, wenn  $X_1 = \alpha X_2$  (sozusagen maximale statistische Abhängigkeit).

Sprechweisen: Physiker nennen Kovarianz auch "Korrelationsfunktion" (sorry!)

Für  $n$  ZV bilden die Kovarianzen

$$C_{ij} = \text{cov}(X_i, X_j)$$

eine symmetrische (pos. def.)  $n \times n$  Matrix (auch **Korrelationsmatrix** oder auch Kovarianzmatrix genannt).

### 1.5 Gesetz der großen Zahlen und zentraler Grenzwertsatz

In der W.theorie gibt es ein "Standardmodell" für Messungen, nämlich ein Zufallsexperiment mit einer zugeordneten ZV ( $X$ ) (sozusagen der "Zeiger" eines Messinstruments). Die "Messdaten" sind  $N$  Werte von  $X$  ( $x_i = X(\omega(i))$ ,  $i = 1, \dots, N$ ), die bei  $N$  Wiederholungen des Zufallsexperiments entstehen. Statt ein  $x_i$  nach dem anderen zu erzeugen, kann man sich äquivalent auch  $N$  identische Kopien des Zufallsexperiments vorstellen (mit  $N$ , statistisch unabhängigen ZV  $X_i$ ,  $i = 1 \dots, N$ ), die gewissermaßen parallel durchgeführt werden, so dass alle  $x_i$  auf einen Schlag entstehen. Um die  $x_i$  und die  $X_i$  deutlich zu unterscheiden, wollen wir die  $x_i$  die *Messreihe* nennen, die  $X_i$  hingegen *Stichprobe*. Eine Messreihe besteht bei uns also aus Zahlen, eine Stichprobe hingegen aus ZV. Wir warnen allerdings die Leser: diese Notation ist keineswegs die überall gebräuchliche <sup>5</sup>. Beachten Sie, dass die  $X_i$  sowohl *statistisch unabhängig* als auch *identisch verteilt* sind, dh.

$$\rho(x_1, \dots, x_N) = \prod_{i=1}^N \rho(x_i)$$

mit ein und demselben  $\rho$  für alle  $X_i$ . In der Physik sind die  $N$  identischen Kopien als *Ensemble* bekannt. Von Ensembles wird im Verlaufe dieser Vorlesung noch viel die Rede sein.

Manchmal ist es nützlich, sich daran zu erinnern, dass man auch eine ganze Messreihe wiederholen (sagen wir  $M$  mal) und so Messreihen von Messreihen einführen kann. Im Ensemblebild würde man sagen, die  $N$  ZV  $X_i$  bilden ein  $N$ -tupel  $\mathbf{X}$  und wir betrachten  $M$  Realisierungen von  $\mathbf{X}$ .

Hat man nun Messdaten gesammelt, so möchte man anhand dieser (endlichen) Datenmenge Aussagen über die ZV  $X$  (die Messgröße) und ihre Verteilung machen. Eine für alle messenden Wissenschaften besonders wichtige Aussage betrifft den **empirischen Mittelwert**:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i.$$

So definiert ist  $\bar{X}$  eine ZV (!), die sich als (lineare) Funktion aus den ZV  $X_i$  ergibt. Als messende Wissenschaftlerin glaubt man fest daran, dass diese ZV für große  $N$  den wirklichen Mittelwert  $\langle X \rangle$  der Messgröße "gut approximiert". Wieso? Nun, für jedes  $N$  gilt schon mal

$$\langle \bar{X} \rangle = \frac{1}{N} \sum_i \langle X_i \rangle = \langle X \rangle .$$

<sup>5</sup>Eine überall gebräuchliche Standnotation gibt es in diesem Punkt leider nicht, sonst hätten wir sie verwendet.

Man sagt dazu:  $\bar{X}$  schätzt  $\langle X \rangle$  erwartungstreu. Als Maß für die Abweichungen  $\bar{X} - \langle X \rangle$  betrachten wir die Varianz

$$\sigma_N^2 := \text{var} \bar{X} = \langle (\bar{X} - \langle \bar{X} \rangle)^2 \rangle$$

Setzen wir einfach die Definition von  $\bar{X}$  ein, so erhalten wir

$$\sigma_N^2 = \frac{1}{N^2} \sum_{i,j} (\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle).$$

Da nun aber die  $X_i$  statistisch unabhängig sind, ist für  $i \neq j$  gerade  $\langle X_i X_j \rangle = \langle X_i \rangle \langle X_j \rangle$  und die entsprechenden Terme heben sich in der obigen Doppelsumme folglich weg. Es bleibt

$$\sigma_N^2 = \frac{1}{N^2} \sum_i (\langle X_i^2 \rangle - \langle X_i \rangle^2) = \frac{1}{N^2} \sum_i \text{var}(X_i)$$

Für identisch verteilte  $X_i$  sind alle Varianzen gleich und wir erhalten

$$\sigma_N^2 = \frac{1}{N} \text{var}(X) =: \frac{\sigma^2}{N}$$

Beachten Sie, dass  $\sigma^2 = \text{var}(X)$  (wenn es denn endlich ist) nicht (!) mit dem Stichprobenumfang  $N$  wächst, so dass folglich

$$\sigma_N^2 = O(N^{-1}) \quad \lim_{N \rightarrow \infty} \sigma_N^2 = 0$$

ist. Schon dies ist als *Gesetz der großen Zahlen* bekannt(s.u.). Beziehen wir die Fluktuationen auf den Mittelwert, so können wir dieses Gesetz auch so schreiben

$$\frac{\sigma}{\langle X \rangle} = O\left(\frac{1}{\sqrt{N}}\right)$$

(falls denn  $\langle X \rangle \neq 0$ ). In Worten: Die relativen Schwankungen des empirischen Mittels unabhängiger ZV verschwinden für große  $N$  wie  $1/\sqrt{N}$ .<sup>6</sup>

Muss nun aber das Verschwinden der Varianz notwendigerweise bedeuten, dass  $\bar{X} \rightarrow \langle X \rangle$  und falls ja, in welchem Sinn soll man diesen Grenzwert eigentlich verstehen? Antwort darauf gibt die sog. **Tschebyscheffsche Ungleichung**, die die W. für Fluktuationen einer beliebigen ZV mit deren Varianz verknüpft. Bezeichnen wir die W. für Ereignisse der ZV  $X$  mit  $P_X(\text{Ereignis})$ , so lautet sie:

$$P_X(|x - \langle X \rangle| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2},$$

<sup>6</sup>Wie Sie gmerkt haben werden, braucht man nicht einmal identisch verteilte ZV, so wie bei unserem Modell der Messung, es reicht, wenn die einzelnen  $\text{var}(X_i)$  existieren.

d.h. sie liefert eine obere Schranke für die W. von "Ausflügen um den Mittelwert". Der Beweis dieser Ungleichung ist erstaunlich einfach. Wir müssen nur bemerken, dass

$$P_X(|x - \langle X \rangle| \geq \epsilon) = \int dx \Theta(|x - \langle X \rangle| - \epsilon) \rho(x) = \langle \Theta(|x - \langle X \rangle| - \epsilon) \rangle$$

denn die Theta-funktion schränkt das Integrationsgebiet gerade auf die gewünschte Menge ein. Da sie nur dann nicht verschwindet, wenn  $|x - \langle X \rangle| \leq \epsilon$ , kann man folgende, sehr einfache Ungleichungskette aufschreiben:

$$\Theta(|x - \langle X \rangle| - \epsilon) \leq \left( \frac{|x - \langle X \rangle|}{\epsilon} \right)^2 \Theta(|x - \langle X \rangle| - \epsilon) \leq \left( \frac{|x - \langle X \rangle|}{\epsilon} \right)^2$$

und daraus folgt sofort die Behauptung, wenn wir den Erwartungswert bilden.

Wenden wir diese Ungleichung auf unser obiges Problem an, so sehen wir, dass

$$P_{\bar{X}}(|\bar{x} - \langle X \rangle| \geq \epsilon) \leq \frac{\sigma^2/N}{\epsilon^2}$$

d.h.

$$\lim_{N \rightarrow \infty} P_{\bar{X}}(|\bar{x} - \langle X \rangle| \geq \epsilon) = 0 \quad \forall \epsilon > 0$$

Diese Aussage heißt üblicherweise Gesetz der großen Zahlen. Sie klärt zweifelsfrei, dass das Verschwinden der Varianz impliziert, dass noch so kleine Ausflüge um den Mittelwert für  $N \rightarrow \infty$  die W. 0 haben.

### Zentraler Grenzwertsatz

Nun wollen wir eine etwas anspruchsvollere Frage stellen:

*Wie ist  $\bar{X}$  für große  $N$  verteilt?*

Antwort auf diese Frage gibt der *zentrale Grenzwertsatz* (ZGS). Zu seiner Ableitung benutzen wir eine Technik, die ein sehr breites Anwendungsspektrum in der W.theorie hat, und die uns auch in der Vorlesung wiederholt begegnen wird. Daher gießen wir sie in ein

#### Lemma (Verteilung transformierter ZV)

$\mathbf{X} = (X_1, \dots, X_n)$  seien ZV.  $f(\mathbf{X}) : \mathbb{R}^n \rightarrow \mathbb{R}$  sei glatte Funktion.

Dann ist  $\rho_f(y) = \langle \delta(f(\mathbf{X}) - y) \rangle$

Beweisidee

Hier ist die charakteristische Funktion der ZV  $f$  sehr hilfreich,

$$\phi_f(k) = \langle e^{ikf} \rangle = \int d^n x \rho_{\mathbf{X}}(\mathbf{x}) e^{ikf(\mathbf{x})}$$

Aus ihr gewinnt man die WD  $\rho_f$  durch Fourier(rück)transformation

$$\rho_f(y) = \int \frac{dk}{2\pi} e^{-iky} \int d^n x \rho_{\mathbf{X}}(\mathbf{x}) e^{ikf(\mathbf{X})}$$

Vertauschen wir hier die Integrationsreihenfolge so sind wir schon fertig, denn

$$\delta(y - f(\mathbf{X})) = \int \frac{dk}{2\pi} e^{-ik(y-f(\mathbf{X}))}$$

Dieses Lemma wenden wir nun an auf die Funktion  $\Delta X = \bar{X} - \langle X \rangle$ . Die WD dieser ZV ist nach obigem Lemma

$$\rho_{\Delta X}(z) = \langle \delta(z - \bar{X} + \langle X \rangle) \rangle = \int d^N x \prod_{i=1}^N \rho(x_i) \delta\left(z - \frac{1}{N} \sum_j x_j + \langle X \rangle\right)$$

Hier haben wir bereits ausgenutzt, dass die  $X_i$  statistisch unabhängig und identisch verteilt sind. Setzen wir die Fourierdarstellung der  $\delta$ -Funktion ein, so erkennen wir, dass wir über die einzelnen  $x_j$  *unabhängig voneinander* integrieren können, wobei jedes Integral gerade die charakteristische Funktion von  $\rho$  ergibt. Insgesamt:

$$\rho_{\Delta X}(z) = \int \frac{dk}{2\pi} e^{-ik(z + \langle X \rangle)} \prod_i \phi\left(\frac{k}{N}\right)$$

Beachten Sie, dass  $\phi$  für alle  $X_i$  die gleiche Funktion ist, daher brauchen wir auch keinen Index  $X_i$  zur Unterscheidung der charakteristischen Funktionen. Nun fragen wir uns nach dem Verhalten von  $\phi(k/N)^N$  für große  $N$ . Wir wollen hier voraussetzen, dass  $\phi(k)$  für kleine Argumente mindestens bis zur 2. Ordnung Taylor-entwickelt werden kann (das ist gleichbedeutend mit der Aussage: die ersten 2 Momente der ZV existieren!) Also erhalten wir mit

$$\phi(k/N)^N = e^{N \ln \phi(k/N)}$$

und

$$\phi(k/N) = 1 + i \frac{k}{N} \langle X \rangle - \frac{k^2}{2N^2} \langle X^2 \rangle + O(N^{-3})$$

sowie

$$\ln(1 + C) = C - C^2/2 + O(C^3)$$

schließlich

$$\phi(k/N)^N = \exp\left(ik \langle X \rangle - \frac{\text{var}(X)}{2N} k^2 + O(N^{-2})\right)$$

Für  $\text{var}(X)$  führen wir wieder die Bezeichnung  $\sigma^2 := \text{var}(X)$  ein, und erhalten so den folgenden Ausdruck für  $\rho_{\Delta X}$ ;

$$\rho_{\Delta X}(z) = \int \frac{dk}{2\pi} e^{-ikz - \sigma^2 k^2 / (2N) + O(N^{-2})}$$

Wenn wir an dieser Stelle  $N \rightarrow \infty$  gehen lassen, so verschwinden im Exponenten des Integranden alle Terme bis auf  $ikz$  und die Grenzverteilung wird zu  $\delta(z)$ . Das ist auch ganz in Übereinstimmung mit dem Gesetz der großen Zahlen, es sagt uns, dass nicht nur die Varianz verschwindet sondern dass auch die WD gegen  $\delta(\bar{X} - \langle X \rangle)$  geht! Wenn wir für große  $N$  die Korrekturterme  $O(N^{-2})$  einmal vernachlässigen (gleich werden wir das noch genauer formulieren), so wird die Verteilung eine Gaussische Verteilung, wie man mit Hilfe des "quadratischen Ergänzungstricks" leicht sieht:

$$\frac{\sigma^2}{2N} \left( k^2 + i \frac{2N}{\sigma^2} ikz \right) = \frac{\sigma^2}{2N} \left( k + \frac{iNz}{\sigma^2} \right)^2 + \frac{N}{2\sigma^2} z^2$$

Der erste Term auf der rechten Seite ergibt (nach Verschieben der Integrationsvariablen) ein einfaches Gauss Integral (s. Bronstein), während der 2. Term aus dem Integral herausgezogen werden kann und offenbar dazu führt, dass

$$\rho_{\Delta X}(z) \propto e^{-Nz^2/(2\sigma^2)} \approx \sqrt{\frac{N}{2\pi\sigma^2}} e^{-Nz^2/(2\sigma^2)}$$

wobei der Vorfaktor für die richtige Normierung sorgt. Dies entspricht einer mit wachsendem  $N$  immer schmaler werdenden Gauss-Glockenkurve.

Man sieht aber nun auch sehr einfach, dass es tatsächlich eine "richtig skalierte" Abweichung des  $\bar{X}$  vom  $\langle X \rangle$  gibt, die im Limes  $N \rightarrow \infty$  sogar eine Gaussische Grenzverteilung besitzt, nämlich

$$\hat{z} = \sqrt{N}(\bar{X} - \langle X \rangle).$$

Ihre Verteilung erhalten wir aus der obigen Gauss-Verteilung durch die Substitution  $\hat{z} = \sqrt{N}z$  und wenn man auf die vernachlässigten Terme schaut, so findet man, dass sie tatsächlich im Limes  $N \rightarrow \infty$  verschwinden. Also ist

$$\lim_{N \rightarrow \infty} \rho_{(\sqrt{N}\Delta X)}(\hat{z}) = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\hat{z}^2/(2\sigma^2)}$$

Dies ist eigentlich der *zentrale Grenzwertsatz* (ZGS), obwohl das obige Resultat für  $z$  in der Praxis ebenso nützlich ist.

Abschließend wollen wir noch einmal den Hauptnutzen des ZGS hervorheben. Er macht eine quantitative Aussage über die WD von (normierten bzw skalierten) Summen unabhängiger, identisch verteilter ZV, wobei die Verteilung der einzelnen ZV *ganz unbekannt* ist. In die Verteilung der Summen gehen nur die ersten beiden Momente der unbekanntem Verteilung ein. Das heißt einerseits, dass viele WDs zur gleichen Grenzverteilung für die Summenvariable führen und andererseits, dass WDs, deren erste 2 Momente nicht existieren, keine Gaussische Grenzverteilung der Summenvariablen besitzen.

## 1.6 Einige gebräuchliche ZV

### Gleichverteilung

$X(\omega) \in [a, b]$  mit

$$\rho(x) = \frac{1}{b-a}$$

Mittelwert:  $\langle X \rangle = (b+a)/2$

Varianz:  $\sigma^2 = \frac{1}{6}(b-a)^2$

Für diskrete  $X(\omega) \in S = \{x_i\}$  mit  $|S|$  Elementen ist

$$P(x_i) = \frac{1}{|S|}$$

### Gaussverteilung

Für  $N$  ZV  $\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)) \in \mathbb{R}^N$  ist die gemeinsame WD

$$\rho(\mathbf{x}) = \frac{(2\pi)^{-N/2}}{\sqrt{\det(C)}} \exp \left\{ \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t C^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

Mit symmetrischer, pos. def. Matrix  $C$ . Da man eine solche Matrix stets durch orthogonale Transformation diagonalisieren kann (und die transformierten Variablen dann statistisch unabhängig Gaussisch verteilt sind!) ist die Bestimmung des Normierungsfaktors in dieser transformierten Form besonders einfach. Man erhält ein  $1/\sqrt{2\pi\lambda_i}$  für jede transformierte Variable, wobei  $\lambda_i$  die Eigenwerte der Matrix  $C$  sind. Das Produkt der Eigenwerte ist gerade die Determinante und die ist invariant unter der Transformation!

Mittelwerte:  $\langle \mathbf{x} \rangle = \langle \boldsymbol{\mu} \rangle$

Varianz:  $\text{cov}(x_i, x_j) = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle = C_{ij}$

Eine Besonderheit der Gaussverteilung, die sie von allen anderen WD auszeichnet:

Die Gaussverteilung besitzt nur 2 nichtverschwindende Kumulanten!

### Binomial (Bernoulli) Verteilung

Wir gehen aus von einer **binären** ZV  $X$ , die also irgend 2 Werte annimmt, z.B.:  $\{0, 1\}$ ,  $\{Ja, Nein\}$ , etc. Nehmen wir mal die erstere Wahl. Die Verteilung ist durch einen einzigen Parameter bestimmt:

$P(X = 1) = p \rightarrow P(X = 0) = 1 - p$ . Nun fragen wir danach, wie häufig bei  $N$  Realisierungen der ZV  $X$  die "1" vorkommt, d.h. die ZV, die dies gerade misst, ist

$$N_+ = \sum_{i=1}^N \delta_{X(\omega_i), 1}$$

( $\delta$  ist hier Kronecker's Delta). Welche W. charakterisieren die  $N_+ \in \mathbb{N}$ ? Offenbar ist

$$P_{N_+}(M) \propto p^M (1-p)^{N-M},$$

d.i. gerade die W.,  $M$  der (statistisch unabhängigen)  $N$  Realisierungen mit  $X = 1$  zu finden und  $N - M$  mit  $X = 0$ . Die Normierung  $\sum_{M=0}^N P_{N_+}(M) = 1$  legt die Verteilung fest:

$$P_{N_+}(M) = \binom{M}{N} p^M (1-p)^{N-M}$$

Erinnern Sie sich dazu an die *Binomialformel*

$$(a+b)^N = \sum_{M=0}^N \binom{M}{N} a^M b^{N-M}$$

wobei die *Binomialkoeffizienten* gegeben sind durch

$$\binom{M}{N} = \frac{N!}{M!(N-M)!}.$$

Die Berechnung von Momenten dieser Verteilung wird sehr vereinfacht, wenn man eine momenten-erzeugende Funktion einführt, die (wg. des diskreten Charakters der ZV) eine etwas andere Form als die charakteristische Funktion (s.o.) hat

$$g(\lambda) := \sum_{M=0}^N \lambda^M p^M (1-p)^{N-M} = (\lambda p + (1-p))^N = \{1 + p(\lambda - 1)\}^N$$

Offenbar ist  $g(1) = 1$  und Differenzieren nach  $\lambda$  bei  $\lambda = 1$  erzeugt die Momente, z.B.

$$\left. \frac{dg}{d\lambda} \right|_{\lambda=1} = \langle N_+ \rangle$$

$$\left. \frac{d^2g}{d\lambda^2} \right|_{\lambda=1} = \langle N_+(N_+ - 1) \rangle = \langle N_+^2 \rangle - \langle N_+ \rangle^2$$

etc.

## 1.7 Datenanalyse

Wir wenden uns nun noch einmal der statistischen Analyse von Daten zu, die man aus Experimenten oder Simulationen gewinnt.  $N$  Wiederholungen des Experiments ergeben eine Messreihe (bzw. Stichprobe, s.o.) einer Observablen (=ZV)  $X(\omega)$ . Die Verteilung dieser ZV (hier als WD  $\rho_X(x) = \rho(x)$  beschrieben) ist das "unbekannte Naturgesetz", dem man mittels der Experimente auf die Spur kommen will. Daher lautet die Grundfrage der Datenanalyse

### (Statistische Inferenz)

Welche Informationen über  $\rho$  kann man aus der Messreihe gewinnen?

Dafür gibt es 3 grundsätzliche Methoden, die nicht unabhängig sind, sondern zumeist zusammen verwendet werden:

- a) Schätzen
- b) Konfidenzbereiche bestimmen
- c) Testen

Beginnen wir mit dem Schätzen. Ein bekanntes Beispiel ist das Schätzen des Mittelwerts aus einer Messreihe. Man benutzt eine **Schätzfunktion**, in diesem Fall

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

die den wahren Mittelwert  $\langle X \rangle$  annähern soll. In welchem Sinn? Dazu ist es nützlich, eine entsprechende ZV  $\bar{X}_N$  zu untersuchen, die man über die Summe der Stichprobe ( $X_i()$  lauter ZV) definiert:

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i.$$

Über diese ZV lassen sich folgende Aussagen treffen:

- (i)  $\langle \bar{X}_N \rangle = \langle X \rangle$ , und zwar für Stichproben beliebigen Umfangs  $N$ . Man sagt daher,  $\bar{X}_N$  sei eine **erwartungstreue** Schätzfunktion
- (ii)  $\lim_{N \rightarrow \infty} \bar{X}_N = \langle X \rangle$  (Gesetz der großen Zahlen). Man sagt daher,  $\bar{X}_N$  sei eine **konsistente** Schätzfunktion

(iii)  $\rho_{\bar{X}_N}$  wird Gaussisch für große  $N$  (zentraler Grenzwertsatz). Diese Aussage ist für die weiteren Methoden der statistischen Inferenz sehr wichtig.

Um diese Aussage gleich zu illustrieren, wenden wir uns kurz der 2. Methode zu, der Bestimmung von Konfidenzbereichen. Gehen wir vorläufig davon aus, dass die Varianz der ZV  $X$  bekannt sei (sagen wir  $\sigma^2$ ) (ist natürlich nicht der Fall, s.u.). Dann ist  $\sigma_{\bar{X}_N}^2 = \sigma^2/N$  (Gesetz der großen Zahlen). Wenn wir nun auch noch  $\langle X \rangle = \langle \bar{X}_N \rangle$  kennen *würden* (tun wir natürlich nicht), so könnten wir Quantile für die ZV  $\bar{X}_N$  angeben, z.B. ist <sup>7</sup>

$$P_{\bar{X}_N}(\bar{x} \in [\langle X \rangle - \sigma_{\bar{X}_N}, \langle X \rangle + \sigma_{\bar{X}_N}]) = 0.68 \dots,$$

Die interessantere Frage ist aber:

Mit welcher W.  $\alpha$  liegt der (unbekannte) wahre Mittelwert  $\langle X \rangle$  in einem bestimmten Intervall um die Realisierung  $\bar{x}$  der Messreihe?

Oder noch etwas anders ausgedrückt:

In welchem stochastischen Intervall um  $\bar{x}$  liegt  $\langle X \rangle$  mit einer vorgegeben W.  $\alpha$ ? Dies nennt man **Konfidenzintervall zum Konfidenzniveau  $\alpha$**

Zum Glück ist diese Frage einfach zu beantworten, wenn  $\sigma$  bekannt ist, denn die Bedingung

$$\langle X \rangle - \sigma_{\bar{X}_N} \leq \bar{x} \leq \langle X \rangle + \sigma_{\bar{X}_N}$$

ist äquivalent zu

$$\bar{x} - \sigma_{\bar{X}_N} \leq \langle X \rangle \leq \bar{x} + \sigma_{\bar{X}_N}$$

Daher liegt  $\langle X \rangle$  mit W.  $0.68 \dots$  im Intervall der Breite  $2\sigma/\sqrt{N}$  um  $\bar{x}$  und die Konfidenzintervalle zu anderen Niveaus sind analog leicht zu bestimmen.

Nun kommen wir auf die Frage zurück, was zu tun ist, wenn wir  $\sigma$  nicht kennen (also den üblichen Fall). Zunächst einmal suchen wir uns wieder eine Schätzfunktion für die Varianz. In den Übungsaufgaben haben Sie gezeigt, dass

$$s_N^2 := \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X}_N)^2$$

eine erwartungstreue Schätzfunktion ist. Die nächste Frage ist nun: Wie die ZV  $s_N^2$  verteilt? Da wir die Verteilung der  $X_i$  nicht kennen, ist diese Frage nicht

<sup>7</sup>was man aus dem Integral

$$\int_{\langle x \rangle - \sigma}^{\langle x \rangle + \sigma} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x - \langle x \rangle)^2 / (2\sigma^2)}$$

und numerischer Berechnung der resultierenden "Error Funktion" gewinnt.

ganz einfach zu beantworten, denn den ZGS können wir nicht unmittelbar heranziehen.

Für Mutige wollen wir hier andeuten, wie man bei unbekanntem  $\langle X \rangle$  und unbekanntem  $\sigma$  vorgehen kann. Die Beweisidee des ZGS können wir nämlich ohne Weiteres auf diesen komplizierteren Fall übertragen: Wir suchen die *gemeinsame WD*

$$\rho(\bar{x}, s_N^2) = \left\langle \delta \left( \bar{x} - \frac{1}{N} \sum_i x_i \right) \delta \left( s_N^2 - \frac{1}{N-1} \sum_i x_i^2 - \frac{N}{N-1} \bar{x}^2 \right) \right\rangle$$

und übertragen die Schritte der Beweisidee des ZSG. So kommen wir zu einer gemeinsamen WD, die wir hier aber nicht weiter untersuchen wollen. Wir verweisen auf die einschlägige Literatur zur statistischen Datenanalyse. Wichtiger ist uns, dass wir mit Hilfe von  $\rho(\bar{x}, s_N^2)$  die Idee der Konfidenzintervalle auch auf *zwei-dimensionale Bereiche* erweitern können, also wir fragen:

In welchem 2-dim Bereich der  $(x - z)$  Ebene um  $(\bar{x}, s_N^2)$  liegt der Punkt  $(x = \langle X \rangle, z = \sigma^2)$  mit vorgegebener W.  $\alpha$ ? (**Konfidenzbereich zum Konfidenzniveau  $\alpha$** ).

In der Praxis benutzt man oft auch bei unbekanntem Mittelwert und unbekannter Varianz ein-dimensionale Konfidenz-Intervalle. Die Idee ist, dass die reskalierte Variable

$$\frac{(\bar{X} - \langle X \rangle)}{\sigma/\sqrt{N}}$$

bei bekannter Varianz für große  $N$  eine Gaussverteilung mit Mittelwert Null und Varianz 1 ergibt. Ersetzt man hier  $\sigma/\sqrt{N}$  durch den erwartungstreuen Schätzer  $s_N$ , so erhält man eine neue ZV,

$$t = \frac{(\bar{X} - \langle X \rangle)}{s_N/\sqrt{N}}$$

die ein Quotient aus 2 ZV ist. Deren Verteilung lässt sich (für große  $N$ ) berechnen, sie wird auch *t-Verteilung* genannt (s. statistische Tabellen) und kann nun herangezogen werden, um Konfidenzintervalle festzulegen.

Für viele Anwendungen in der Physik ist das **Schätzen von Modellparametern** wichtig, dem wir uns nun zuwenden. Physikalische Theorien liefern (idalerweise) ein **Modell** der Observablen (ZV)  $X$ , das wir mal mit

$$\rho_{theor}(x; \lambda)$$

bezeichnen. Dabei ist/sind  $\lambda$  Modellparameter, die die Theorie leider nicht bestimmen konnte. Die Fragen sind nun:

- 1) Wie schätzen wir den besten Wert für  $\lambda$  aus der Messreihe  $\mathbf{x} = (x_1, \dots, x_N)$ ?
- 2) Folgt denn die Messreihe dem theoretischen Modell?

Die erste Frage ist eine klassische Frage aus dem Bereich der Schätztheorie, die man häufig mit der Methode der **maximal likelihood** (ML) behandelt. Für eine vorgegebene Messreihe  $\mathbf{x} = (x_1, \dots, x_N)$  ist die likelihood definiert als

$$L(\lambda) := \prod_{i=1}^N \rho_{theor}(x_i, \lambda)$$

Es ist also gerade die Vorhersage des Modells für die W. der gefundenen Messreihe. Man geht nun davon aus, dass die Messreihe eine typische Realisierung einer "typischen Verteilung" ist und sucht das *Maximum von  $L$  bezüglich  $\lambda$* , d.h. diejenigen Werte der Modellparameter, die die typische Messreihe mit möglichst großer W. auftreten lassen. Um die Qualität einer ML Schätzung zu überprüfen, kann man (wenn man kann!) weitere Messreihen heranziehen. Jede Messreihe liefert einen Schätzwert für  $\lambda$ , die Gesamtheit der Schätzwerte nennt man dann die Schätzung. Eine gute Schätzung hat einen (und nur einen!) deutlichen Häufungspunkt für die geschätzten Parameter.

Sehr häufige Verwendung in der Praxis findet die ML Schätzung mit einem *Gaussischen Modell*, die auch als **least square Schätzungen** bekannt sind. Das Gaussische Modell (für eine Variable) enthält 2 Parameter, den Mittelwert  $\mu$  und die Varianz  $\sigma$ . Die einfachste Gaussische likelihood Funktion

$$L(\mu, \sigma) = \prod_{i=1}^N \frac{1}{2\pi\sigma^2} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

hat ihr Maximum bei

$$\mu = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

ganz unabhängig von  $\sigma$ , d.h. die ML Schätzung des Mittelwerts reproduziert den uns schon bekannten, konsistenten und erwartungstreuen Schätzer. Nun brauchen wir nur noch die Funktion  $L(\bar{x}, \sigma)$  zu maximieren, und diese Prozedur liefert uns eine Schätzung für die Varianz. Das Ergebnis ist jedoch nicht — wie Sie vielleicht gehofft haben — unsere erwartungstreue Schätzfunktion  $s_N^2$ , sondern

$$s^2 = \frac{N-1}{N} s_N^2$$

wie Sie ganz leicht nachrechnen können.

Eine etwas anspruchsvollere Anwendung der least square Schätzer ist die **Methode der kleinsten Fehlerquadrate** (Gauss!). Wir betrachten ein deterministisches Modell (als Grenzfall natürlich in WDs enthalten), dessen Vorhersage

eine Kurve  $x(t, \lambda)$  (z.B. eine Trajektorie eines Planeten) ist, wobei aber  $\lambda$  eben noch nicht bekannt ist (die Elemente der Planetenbahn). Auf der experimentellen Seite haben wir Beobachtungen  $x_i$  zu Zeiten  $t_i$ , die aber mit Messfehlern behaftet sind. Hier betrachten wir nur den Fall, dass Messfehler in den  $y_i$  auftreten, während die Zeit  $t$  als beliebig genau messbar angenommen wird. Der Vergleich zwischen Modellvorhersage und Experiment muss also die Messfehler  $\epsilon_i$  in Betracht ziehen, d.h.

$$x_i = x(t_i, \lambda) + \epsilon_i$$

Gute Experimentatoren kontrollieren ihre Messfehler durch Messreihen und liefern eine Schätzung der Messfehlervarianz  $\sigma_i$  des  $i$ -ten Datenpunkts mit. Da nicht mehr als die Varianz des Fehlers bekannt ist (und natürlich sein Mittelwert: der ist per def. Null, denn sonst hätte man *systematische Fehler* gemacht), kann man auch guten Gewissens nur eine Gaussische Verteilung der Messfehler in Betracht ziehen (für alles Weitere fehlt die Information, die in höheren Kumulanten der Messfehler steckt). Also geht eine ML Schätzung aus von einer Messreihe von Datenpunkten  $(x_i, \sigma_i)$ ,  $i = 1 \dots M$  und betrachtet eine Gaussische Likelihood Funktion

$$L(\lambda) = C \prod_{i=1}^M e^{-(x_i - x(t_i, \lambda))^2 / (2\sigma_i^2)} = C e^{-E(\lambda)}$$

Das Maximum dieser Funktion liegt beim Minimum der *gewichteten Fehlerquadrate*!

$$\min_{\lambda} E = \min_{\lambda} \sum_{i=1}^M \left( \frac{x_i - x(t_i, \lambda)}{2\sigma} \right)^2$$

Aber warum erzähle ich Ihnen das? Nach den zahlreichen Praktikumsversuchen während Ihres bisherigen Studiums sind sie natürlich längst in der Datenanalyse perfekt und wissen auch, wie man die Methode der kleinsten Fehlerquadrate effektiv benutzt. Daher schenke ich mir hier alles Weitere zu diesem Thema.

Nun wenden wir uns der dritten Methode statistischer Inferenz zu, dem **Testen**. Haben wir eine Messreihe gewonnen und haben wir ein theoretisches Modell, so werden wir (zumindest der Theoretiker wird) die Hypothese aufstellen, dass die Messwerte mit dem Modell verträglich sind. Wie überprüfen wir diese Hypothese? Natürlich können wir einzelne Momente (Mittelwert, Varianz) vergleichen, aber damit reduzieren wir unsere experimentelle Information erheblich, die ja schließlich in einer **empirischen Verteilungsfunktion**

oder **empirischen WD** besteht, die zumeist in Form eines **Histogramms** vorliegt, d.h. die relativen Häufigkeiten, aufgetragen gegen die Messwerte. Der systematische Weg zum Überprüfen solcher Hypothesen ist die Konstruktion eines statistischen Tests.

Ein **statistischer Test** ist eine Regel, eine Hypothese aufgrund einer Messreihe zu akzeptieren oder abzulehnen. Formal kann man sagen: Ist  $\omega = \mathbf{x}$  eine Realisierung (Messreihe), so ist ein Test eine *binäre ZV*  $\tau(\omega) \in \{0(= \text{abgelehnt}), 1(= \text{akzeptiert})\}$

Die Hypothese selbst ist eine Aussage, die entweder wahr oder falsch ist (sicher!). Da das Modell jedoch noch Parameter  $\lambda$  enthält, so wird die Hypothese i.a. für eine gewisse Menge  $\Lambda_w$  von Modellparametern wahr<sup>8</sup> und für den Rest  $\bar{\Lambda}_w = \Lambda_f$  falsch sein.

Allerdings können wir aufgrund des endlichen Umfangs der Messreihe keine sichere Zuordnung des Wahrheitswertes vornehmen. Selbst wenn beispielsweise eine Messreihe  $x_i, i = 1, \dots, N$  tatsächlich aus Realisierungen einer vorgegebenen ZV mit WD  $\rho_{theor}$  besteht, kann sie ja *zufällig* sehr untypisch sein und ein Kriterium, das wir benutzen (z.B. Vergleich des empirischen Varianz mit der Modellvarianz) würde uns suggerieren, die Hypothese abzulehnen. Dann haben wir einen Fehler gemacht, und solche Irrtümer sind bei statistischer Inferenz *unvermeidlich*. Es gilt lediglich, sie genau zu quantifizieren. Wenn der Umfang der Messreihe wächst, wissen wir schon aus dem Gesetz der großen Zahl, dass Fehler immer seltener passieren müssen. Daher ist es eine eigene Untersuchung wert, den *Umfang einer Stichprobe* geeignet festzulegen. Darauf gehen wir hier nicht ein. Wir beschäftigen uns nur mit der Konstruktion von Tests.

Zunächst einmal gibt es wohl zwei Typen von Irrtümern, die wir begehen können:

- (i) die Hypothese ablehnen, obwohl sie wahr ist (Fehler 1. Art)
- (ii) die Hypothese akzeptieren, obwohl sie falsch ist (Fehler 2. Art)

Man beachte, dass man den einen oder den anderen Fehler stets auf triviale Art vermeiden kann. Fehler 1. Art treten nicht auf, wenn wir die Hypothese für alle Messreihen akzeptieren, Fehler 2. Art treten nicht auf, wenn wir alles ablehnen. Nützen würden uns solche "trivialen Tests" natürlich nix, denn sie beinhalten keinerlei Erkenntniszuwachs. Wir müssen aber schon etwas deutli-

<sup>8</sup>Wenn man Pech hat, ist  $\Lambda_w = \emptyset$ , als Theoretiker hat man dann seinen Beruf verfehlt

cher sagen, was wir nun eigentlich wollen.<sup>9</sup> Meistens einigt man sich auf eine positive Sicht der Dinge und sagt: Wir suchen nach einem *mächtigen* Test, das ist einer, bei dem die Wahrscheinlichkeit  $P(\tau = 1)$  dann sehr groß ist, wenn die Hypothese stimmt (d.h. für das Modell:  $\lambda \in \Lambda_w$ ). Mächtige Tests machen also mit geringer W. Fehler 1. Art. Quantitativ gesagt: Wir nennen einen (mächtigen) **Test zum Signifikanzniveau  $\alpha$**  einen, bei dem

$$P_\lambda(\tau = 0) \leq \alpha \quad \forall \lambda \in \Lambda_w$$

also, die W. die Hypothese abzulehnen, wenn sie richtig ist, ist kleiner gleich  $\alpha$ . Unter diesen Tests suchen wir nun solche, für die

$$P_\lambda(\tau = 1) \quad \text{möglichst groß für } \lambda \in \Lambda_w$$

Wenn man kann, kann man die sog. *Gütefunktion*

$$\beta(\lambda) = P_\lambda(\tau = 1)$$

für alle  $\lambda$  bestimmen, sie umfasst die interessanten Eigenschaften des Tests (natürlich muss auch  $\Lambda_w$  bekannt sein).

Wie konstruiert man nun einen "guten" Test? Antwort: Heuristisch. Eine typische Form, die Entscheidung "akzeptieren oder ablehnen" herbeizuführen ist eine *Schwellenwertentscheidung*. Man konstruiert eine "aussagekräftige" ZV  $T$  (für Test), von der man die WD berechnen kann, *wenn die Hypothese wahr ist*. "Aussagekräftig" ist eine ZV die man unmittelbar oder mittelbar als eine Schätzung von Parametern aus  $\Lambda_w$  auffassen kann. Zumindest muss klar sein, für welche Wertebereiche von  $T$  die Hypothese immer sicherer, bzw. immer sicherer falsch wird. Oft ergibt sich die Wahl von  $T$  unmittelbar aus der Hypothese (s.u.) Da man die WD von  $T$  berechnet hat, weiß man, in welchen Intervallen man  $T$  mit welchen W. findet, falls die Hypothese stimmt, und aus der Kenntnis von  $T$  kennt man zumindest die Topologie der *Annahme-* bzw. *Ablehnungsbereiche*.  $T$  kann z.B. alle Werte auf der reellen Achse annehmen,

<sup>9</sup>Statistiker haben mit dem Begriff der sog. *Null-Hypothese* all die obigen, eigentlich sonnenklaren Dinge wieder verdunkelt. In ihrer Welt gibt es Hypothesen, die interessant sind und die Verneinung einer solchen Hypothese ist dann weniger interessant. (z.B. Hypothese: Da ist eine tolle Anomalie in den Daten. Verneinung: Die Daten sind total gewöhnlich (gäh!), so wie immer). Also stellt man eine (interessante) Hypothese  $H_1$  auf, und deren Alternative ist dann die ("es tut sich nichts") Null-Hypothese  $H_0$ . Nun fangen sie an, über die Wahrheit und Falschheit von  $H_0$  und  $H_1$  so lange zu reden, bis nichts mehr so recht klar ist. Für messende Wissenschaftlerinnen, die auch noch modellbildende und vorhersagende Kolleginnen zur Seite haben, stellt sich das nämlich oft ganz anders dar: Spannend ist es, wenn ein Modell mit dem Experiment übereinstimmt (für den Statistiker eigentlich die Null-Hypothese), frustrierend, wenn es nicht übereinstimmt. Wir wollen daher hier lieber keine Null-Hypothesen einführen!

der Annahmebereich liegt jedoch um einen bestimmten Punkt (z.B. das Maximum) herum, d.h.  $T$  sollte weder zu groß noch zu klein sein und die Ablehnungsbereiche erstrecken sich von  $\pm\infty$  bis irgendwo in die Nähe des speziellen Punktes. Dann spricht man von einem *zweiseitigen Test*. Ist hingegen der Ablehnungsbereich bei kleinen, der Annahmebereich bei großen Werten von  $T$  (oder umgekehrt) heißt der Test *einseitig*.

Man quantifiziert nun die Größe der Bereiche durch das *Signifikanzniveau*, das **Quantile von T** festlegt, die zu Grenzen zwischen Akzeptanz und Ablehnung werden. Dann schaut man auf die vorliegende Messreihe, berechnet aus ihr  $T(\omega)$  und sieht nach, ob das in einen Annahme- oder ein Ablehnungsbereich fällt.

Betrachten wir zur Illustration zwei gängige Tests der Hypothese:

Die Messreihe besteht aus Realisierungen von  $X$ , wobei  $X$  nach einer vorgegebenen WD  $\rho_{theor}$  verteilt ist.

### Der $\chi^2$ Test:

Wir fassen die  $x_1 \dots x_M$  der Messreihe zu  $k$  Klassen zusammen ("bins" wie bei einem Histogramm), so dass in jeder Klasse  $N_l \gg 1$  Messwerte liegen. Aus dem Modell verschaffen wir uns die  $W_l$ , eine Realisierung in einer Klasse zu finden, d.h. die  $P_l$ ,  $l = 1 \dots k$ . Die  $W_l$ , von den  $M$  Werten der Messreihe gerade  $N_l$  in der Klasse  $l$  (und  $M - N_l$  nicht in der Klasse  $l$ ) zu finden, ist natürlich eine Binomialverteilung (s.o.)

$$P(N_l) = \binom{M}{N_l} (P_l)^{N_l} (1 - P_l)^{M - N_l}$$

Für  $N_l \gg 1$  geht sie (und zwar rasch) gegen eine Gaussverteilung, wie das nach dem zentralen Grenzwertsatz auch sein muss (denn schließlich ist ja  $N_l$  die Summe von statistisch unabhängigen, binären ZV:  $s_i = 1$  falls Messwert in lter Klasse,  $s_i = 0$  andernfalls.) Mittelwert und Varianz von  $N_l$  ist  $\langle N_l \rangle = MP_l$   $\langle (N_l - \langle N_l \rangle)^2 \rangle = MP_l(1 - P_l)$ . Durch Reskalieren zur ZV

$$z_l := \frac{N_l - MP_l}{\sqrt{MP_l(1 - P_l)}}$$

erhält man (für  $N_l \gg 1$ ) Gaussisch verteilte ZV  $z_l$ , die alle Mittelwert 0 und Varianz 1 haben. d.h. identisch verteilt sind. Diese  $z_l$  sind ein Maß für die Abweichung der *empirischen Verteilung*  $N_l$  von der vorhergesagten Verteilung  $MP_l$

(vorausgesetzt natürlich: die  $N_l$  folgen auch der vorhergesagten Verteilung). Aus ihnen bauen wir nun unsere Testvariable  $T$  in der Form

$$T := \chi^2 = \sum_{l=1}^k z_l^2$$

Die  $z_l$  sind auch *fast* statistisch unabhängig, — bis auf eine Nebenbedingung: Da

$$\sum_{l=1}^k N_l = M$$

ist auch

$$\sum_{l=1}^k z_l \sqrt{M P_l ((1 - P_l))} = 0$$

In einer Übungsaufgabe haben Sie die WD der ZV  $T$  bestimmt und gefunden, dass es eine  $\chi^2$  Verteilung  $P_{\chi^2}(z, k - 1)$  zu  $k - 1$  statistisch unabhängigen Quadraten ist. Der  $\chi^2$  Test ist offenbar einseitig: großes  $T$  bedeutet Ablehnung! Also bestimmen wir die entsprechenden Quantile

$$[0, \chi_{1-\alpha}^2]$$

als Annahmehereich. Diese Quantile sind tabelliert. Fällt eine Realisierung in den Ablehnungsbereich, so macht man durch das Ablehnen, wie oben gesagt, mit W.  $\alpha$  einen Fehler, denn das ist gerade die W.,  $T$  bei Gültigkeit der Hypothese im Ablehnungsbereich zu finden.

Der Knackpunkt bei der Testkonstruktion ist, wie Sie hieran sehen, eine "ausagekräftige" Testvariable zu finden, deren WD man auch berechnen kann. Tests für eine Hypothese kann es beliebig viele geben. Eine anderer Test der gleichen obigen Hypothese ist der

### Kolmogorov-Smirnov Test

Hierbei vergleicht man zur Konstruktion der Testvariablen die **empirische Verteilungsfunktion**

$$F(x; M) = \frac{1}{M} (\text{Anzahl der } x_i \text{ mit } x_i < x)$$

mit der Verteilungsfunktion des Modells

$$F_{theor}(x) = \int_{-\infty}^x dy \rho_{theor}(y)$$

und benutzt als Fehlermaße (= Testvariablen) die maximalen Abweichungen zwischen den beiden Funktionen nach oben und nach unten, d.h.

$$T_1 = \max_x (F(x; N) - F_{theor}(x))$$

$$T_2 = \max_x (F_{theor}(x) - F(x; N))$$

Auch hier ist der entscheidende Punkt: Die WD dieser Testvariablen lassen sich berechnen. (Machen wir hier aber nicht).

### 1.8 Erzeugung von Zufallszahlen auf dem Computer

Ein Computer ist (abgesehen von Betriebsstörungen) ein deterministischer Automat. Mit ihm kann man deterministische Zahlenfolgen erzeugen, aber keine, die im strikten Sinn zufällig sind. Daher ist es eine Kunst, Zahlenfolgen  $z_1, z_2 \dots$  zu generieren, die wenigstens bezüglich einiger Eigenschaften wie Zufallszahlen aussehen. Diese nennt man **Pseudozufallszahlen**.

Die Grundaufgabe besteht darin, gleichverteilte, statistisch unabhängige  $x_n \in [0, 1]$  zu erzeugen, also Realisierungen der auf  $[0, 1]$  gleichverteilten ZV (s.o.). Ein Grundalgorithmus ist die sogenannte *lineare Kongruenzmethode*, d.i. eine Folge mit dem Bildungsgesetz

$$z_{n+1} = (az_n + b \bmod m)$$

mit "geeignet gewählten"  $a, b, m$ . Die erzeugten, natürlichen Zahlen liegen im Intervall  $[0, m - 1]$ . Die Folge ist völlig deterministisch, d.h. startet man beim gleichen  $z_0$ , so erhält man stets dieselben  $z_n$ . Die Zahlen

$$\xi_n = z_n / (m - 1) \in [0, 1]$$

liegen zwar im geforderten Intervall, ob sie aber gleichverteilten Zufallszahlen ähneln, hängt stark von den gewählten Konstanten  $a, b, m$  ab. Geeignete Zahlen muss man mit Hilfe zahlentheoretischer Überlegungen und Kenntnissen der jeweiligen Computer-Hardware bestimmen. Ein ganz ordentlicher Startpunkt ist die Wahl

$$(a = 714025, b = 1366, m = 150889)$$

die dem Generator RAN2 in den *Numerical Recipes* zugrunde liegt. Wie testet man nun eine Pseudozufallszahlenfolge? Mittels statistischer Datenanalyse! Man wird zunächst mal ein Histogramm der  $z_n$  aufstellen, um nach der Gleichverteilung zu schauen (quantitativ mittels  $\chi^2$  oder *KS Test* s.o.). Dann kann man nach der statistischen Unabhängigkeit fragen, indem man z.B.  $z_n$

gegen  $z_{n+k}$  plottet oder Korrelationsfunktionen schätzt. Keine Pseudozufallszahlenfolge kann alle denkbaren Tests passieren, denn dann wäre sie ja wirklich zufällig. Für eine bestimmte Anwendung muss man stets die wichtigsten Aspekte testen, insbesondere, wenn man viele Zufallszahlen braucht; eine lineare Kongruenzfolge kann ja nur  $m$  verschiedene Zahlen produzieren und häufig braucht man bedeutend mehr Zufallszahlen als (s.o.) 150889. Eine Rechnung, die stark auf Zufallszahlen basiert (s. Monte Carlo) sollte man mit verschiedenen Generatoren wiederholen, um zu schauen, ob systematische Einflüsse der jeweiligen Pseudozufallszahlenfolge in Ergebnissen sichtbar werden.

Hat man einmal einen Generator für akzeptable  $\xi_n \in [0,1]$  unabhängig gleichverteilt gefunden, gibt es einige Methoden, aus diesen (Pseudo) Zufallszahlen mit anderen Verteilungen zu erzeugen. Zu den wichtigsten Methoden gehören:

#### a) Transformationen

Sei  $X$  eine ZV mit WD  $\rho_X(x)$ . Der Genauigkeit halber wollen wir noch vermerken, dass die WD in der Menge  $I$  liege (kann auch ganz  $\mathbb{R}$ ) sein. Es sei  $Y$  eine *ein-eindeutige* Abbildung (Umkehrabbildung  $X(y)$ ). Dann ist auch  $Y(X)$  eine ZV. Aus

$$1 = \int_{x \in I} dx \rho_X(x) = \int_{y \in Y(I)} dy \left| \frac{dX}{dy} \right| \rho_X(X(y)) = \int_{y \in Y(I)} dy \rho_Y(y)$$

liest man ab

$$\rho_Y(y) = \left| \frac{dX}{dy} \right| \rho_X(X(y))$$

Nun gehen wir aus von der Gleichverteilung auf  $[0,1]$ , d.h.  $\rho_X(x) = 1$  und  $I = [0,1]$  und suchen geeignete Transformationen, so dass eine gewünschte WD  $\rho_Y$  entsteht. Offenbar ist

$$\rho_Y(y) = \left| \frac{dX}{dy} \right|$$

die entstehende Dichte auf  $I' = Y([0,1])$ . Wir suchen  $Y(X)$  bei vorgegebenem  $\rho_Y$ . Die Umkehrabbildung ist

$$X(y) = \int^y dy' \rho_Y(y') = F_Y(y),$$

also gerade die Verteilungsfunktion der neuen ZV  $Y$ , so dass

$$Y(X) = F_Y^{-1}(X)$$

ist. Die Methode ist also immer dann praktikabel, wenn man die Umkehrfunktion der Verteilungsfunktion von  $Y$  leicht bestimmen kann.

Beispiel: (Exponentiell verteilte ZV):

gewünschte WD:  $\rho_Y(y) = \lambda \exp(-\lambda y)$ ,  $I' = [0, \infty]$

$$X(y) = \int_0^y dy' \lambda \exp(-\lambda y') = 1 - \exp(-\lambda y)$$

$$\rightarrow Y(x) = -\frac{1}{\lambda} \ln(1 - x)$$

mit  $I = X([0, \infty]) = [0, 1]$ . Also kann man mittels eines Logarithmus aus gleichverteilten Zufallszahlen aus  $[0, 1]$  exponentiell verteilte Zufallszahlen erzeugen.

Die Methode ist auch auf mehrere ZV übertragbar. Man benutzt entsprechend

$$\rho_{\mathbf{Y}}(\mathbf{y}) = \left| \frac{\partial \mathbf{X}}{\partial \mathbf{y}} \right| \rho_{\mathbf{X}}(\mathbf{X}(\mathbf{y}))$$

Beispiel: Gaussverteilte ZV mit dem Box-Müller Algorithmus

Es seien  $X_1$  und  $X_2$  in  $[0, 1]$  gleichverteilt. Wir betrachten die Abbildung

$$Y_1(x_1, x_2) = \sqrt{-2 \ln x_1} \cos(2\pi x_2)$$

$$Y_2(x_1, x_2) = \sqrt{-2 \ln x_1} \sin(2\pi x_2)$$

(so dass  $(Y_1, Y_2) : [0, 1] \times [0, 1] \rightarrow \mathbb{R}^2$ ) und die zugehörige Umkehrabbildung

$$X_1(y_1, y_2) = \exp\left(-\frac{1}{2}(y_1^2 + y_2^2)\right)$$

$$X_2(y_1, y_2) = \frac{1}{2\pi} \arctan\left(\frac{y_2}{y_1}\right)$$

und berechnen den Jacobian

$$\left| \frac{\partial(X_1, X_2)}{\partial(y_1, y_2)} \right| = \frac{1}{2\pi} e^{-(y_1^2 + y_2^2)/2}$$

Also ist  $\rho_{Y_1, Y_2}(y_1, y_2) = \frac{1}{2\pi} e^{-(y_1^2 + y_2^2)/2}$  und wir können über die Abbildung  $(Y_1, Y_2)$  aus 2 Zufallszahlen gleichverteilt aus  $[0, 1]$  zwei Gaussverteilte Zufallszahlen gewinnen.

### b) Reduktion

Aus mehreren, insbesondere statistisch unabhängigen und identisch verteilten ZV lassen sich durch Reduktionen gelegentlich neue, nützliche Verteilungen konstruieren. In der Statistik spielt z.B. die sog.  $\chi^2$ -Verteilung eine besondere Rolle. Man gewinnt sie als WD einer Summe von *Quadraten* Gaussischer, unabhängiger ZV  $X_i$ , d.h.

$$\chi^2 := \sum_{i=1}^k x_i^2$$

Wir betrachten die  $X_i$  als so reskaliert, dass ihre Gaussverteilung Mittelwert 0 und Varianz 1 hat. Nun unserem Lemma über die WD transformierter ZV berechnen wir die WD von  $\chi^2$  über

$$\begin{aligned}\rho_{\chi^2}(z; k) &= \left\langle \delta \left( z - \sum_1^k x_i^2 \right) \right\rangle \\ &= \left( \frac{1}{\sqrt{2\pi}} \right)^k \int d^k x \delta \left( z - \sum_1^k x_i^2 \right) e^{-\sum_i x_i^2/2} \\ &= \left( \frac{1}{\sqrt{2\pi}} \right)^k e^{-z/2} S_k(1)\end{aligned}$$

wobei  $S_k(1)$  gerade die Oberfläche einer k-dimensionalen Einheitskugel bezeichnet (vgl. Übungsaufgabe). So erhalten wir schließlich

$$\rho_{\chi^2}(z; k) = \frac{z^{(k/2)-1} e^{-z/2}}{2^{k/2} \Gamma(k/2)}$$

Die Gammafunktion ist definiert als

$$\Gamma(x) = \int_0^\infty dt t^{x-1}$$

und ist folglich verwandt mit der Fakultätsfunktion

$$n! = \Gamma(n + 1) \quad \text{für } n \in \mathbb{N}_0$$

### c) von Neumann's Rejektionsalgorithmus

Was tut man nun, wenn die obigen Methoden nicht funktionieren? Man beachte, dass die Transformationsmethode viele spezielle Voraussetzungen hat (ein-eindeutige Transformation, Jacobian leicht berechenbar,...) die in vielen praktischen Fällen nicht erfüllt sind. In solchen Fällen kann man Gebrauch von einer **algorithmischen Lösung** machen, die leicht auf einem Computer implementiert werden kann. Dazu betrachten wir die Figur(1), in der eine WD  $\rho(x)$  zu sehen ist, für die wir Realisierungen erzeugen wollen. Die WD sei auf ein Intervall  $[a, b]$  beschränkt und wir haben eine horizontale Gerade  $w = C \geq \max_x(\rho(x))$  eingezeichnet, die oberhalb des Maximums der WD verläuft. Nun erzeugt man Punkte  $(x_i, w_i), i = 1, 2 \dots M$ , die gleichverteilt in  $[a, b] \times [0, C]$  liegen. Ein solcher Punkt wird *akzeptiert*, falls

$$w_i \leq \rho(x_i),$$

andernfalls wird es verworfen. Man beachte, dass die akzeptierten Punkte nach und nach die Fläche unter  $\rho$  homogen ausfüllen. Die *relative Häufigkeit* von Punkten mit  $x \leq x_i \leq x + dx$  geht für  $M \rightarrow \infty$  gegen

$$\frac{\text{number of points with } x_i \in [x, x + dx]}{M} \rightarrow$$

$$\frac{\text{area of } [x, x + dx] \times [0, \rho(x)]}{\text{area beneath graph of } \rho} = \rho(x)dx$$

daher sind die  $x_i$  der akzeptierten Punkte Realisierungen von  $\rho(x)$ .

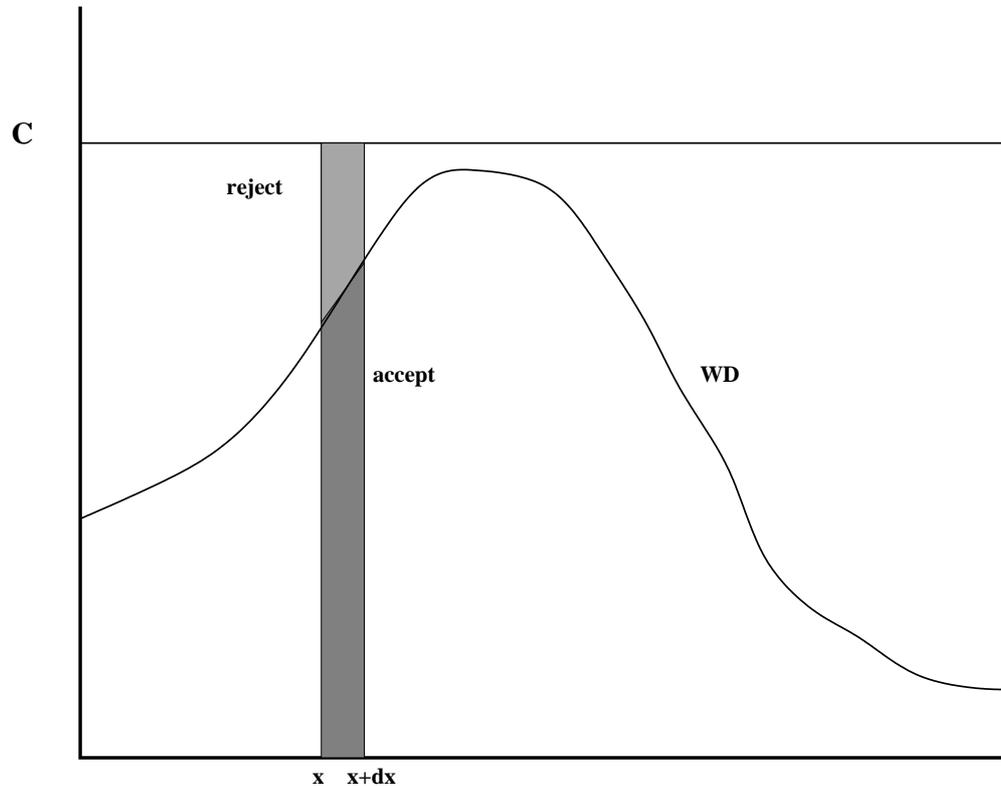


Abbildung 1: Der Rejektionsalgorithmus.

## 2 Zustände und Ensembles

### 2.1 Zeitentwicklung reiner Zustände

Die **Mikrozustände** in der klassischen Mechanik (klM) eines System aus  $N$  Punktteilchen (in 3 Dimensionen) sind Punkte des  $6N$ -dimensionalen Phasenraums ( $\Gamma$ )

$$\text{klassisch : } (p_1 \dots p_{3N}, q_1 \dots, q_{3N}) = (\mathbf{p}, \mathbf{q})$$

von Impulsen und Koordinaten.

In der Quantenmechanik (QM) sind die Mikrozustände Vektoren in einem Hilbertraum  $|\psi\rangle$  oder alternativ und gleichberechtigt die Projektoren auf diese Zustände  $\hat{P}_\psi = |\psi\rangle\langle\psi|$ .<sup>10</sup>

<sup>10</sup>Wir verwenden hier "bra-ket" Notation. Alle Aussagen lassen sich problemlos in andere

Die Mikrozustände der klassischen und der Quantenmechanik werden auch als **reine Zustände** bezeichnet.

Die Zeitentwicklung reiner Zustände ist in der kLM durch die **kanonischen Bewegungsgleichungen** gegeben

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}$$

die durch eine **Hamiltonfunktion**  $H(\mathbf{p}, \mathbf{q})$  bestimmt werden. Die kanonischen DGL definieren an jedem Punkt des Phasenraums eine "Phasenraumgeschwindigkeit"

$$\mathbf{v} = (\dot{\mathbf{p}}, \dot{\mathbf{q}}) = \left( \frac{\partial H}{\partial p_1}, \dots, \frac{\partial H}{\partial p_{3N}}, \frac{\partial H}{\partial q_1}, \dots, \frac{\partial H}{\partial q_{3N}} \right)$$

Die **Divergenz dieses Geschwindigkeitsfeldes** verschwindet, denn

$$\operatorname{div} \mathbf{v} = \sum_i^{3N} \frac{\partial^2 H}{\partial q_i \partial p_i} - \frac{\partial^2 H}{\partial p_i \partial q_i} = 0$$

Die Strömungslinien eines solchen Geschwindigkeitsfeldes sind — analog zu den Feldlinien des magnetischen Feldes— frei von Quellen oder Senken (wie auch bei einer inkompressiblen Flüssigkeit).

In der QM ist die Dynamik der reinen Zustände durch die **Schrödingergleichung** bestimmt

$$i\hbar \partial_t |\psi\rangle = \hat{H} |\psi\rangle,$$

wobei  $\hat{H}$  der Hamiltonoperator des Systems die Observable "Energie" charakterisiert. Die Transformation  $|\psi\rangle_{t=0} \rightarrow |\psi\rangle_t$ , die durch diese Dynamik erzeugt wird ist eine **unitäre** Transformation. Man kann den unitären Zeitentwicklungsoperator durch Integration der Schrödingergleichung in folgender Form schreiben:

$$|\psi\rangle_t = \hat{U}(t, t_0) |\psi\rangle_{t_0} = e^{-\frac{i}{\hbar} \hat{H}(t-t_0)} |\psi\rangle_{t_0}$$

Durch hermitesche Konjugation erhält man hieraus

$$\langle \psi|_t = \langle \psi|_{t_0} e^{\frac{i}{\hbar} \hat{H}(t-t_0)}$$

## 2.2 Observable

Observable in der kLM sind Funktionen  $A(\mathbf{p}, \mathbf{q})$  auf dem Phasenraum. Ist der aktuelle Wert des Mikrozustands  $\mathbf{p}(t), \mathbf{q}(t)$ , so ist der aktuelle Wert der Observablen  $A(\mathbf{p}(t), \mathbf{q}(t))$ . Etwas aufwändiger, aber äquivalent kann man das auch so schreiben

$$A(t) = \int d\Gamma \delta^{(3N)}(\mathbf{p} - \mathbf{p}(t)) \delta^{(3N)}(\mathbf{q} - \mathbf{q}(t)) A(\mathbf{p}, \mathbf{q})$$

---

Notationen überführen.

wobei wir

$$d\Gamma = d^{3N} p_i d^{3N} q_i$$

zur Abkürzung eingeführt haben. In einem reinen Zustand hat jede Observable einen scharfen Wert. Fluktuationen oder Zufälligkeiten gibt es keine.

In der QM wird eine Observable durch einen **hermiteschen Operator**  $\hat{A}$  charakterisiert. In einem reinen Zustand gibt es i.a. eine nichttriviale Messwertstatistik von  $\hat{A}$ . Der Erwartungswert ist

$$\langle \hat{A} \rangle = \langle \psi | \hat{A} | \psi \rangle = Sp(P_\psi \hat{A})$$

wobei die letzte Form ( $P_\psi = |\psi\rangle\langle\psi|$ ) sich gleich noch als verallgemeinerungsfähig erweisen wird. Falls  $|\psi\rangle$  nicht Eigenzustand von  $\hat{A}$  ist, verschwinden auch Varianz und höhere Kumulanten nicht. Man kann die Momente der Messwertstatistik immer noch (wie im klassischen) Fall aus der charakteristischen Funktion erzeugen

$$\phi_{\hat{A}}(k) = \langle e^{ik\hat{A}} \rangle .$$

### 2.3 Ensembles und gemischte Zustände

Es kommt praktisch ständig vor, dass die Auswahl eines reinen Zustands als Anfangszustand eines Experiments selbst ein Zufallsexperiment ist (Wer kann schon  $10^{23}$  Orte und Impulse *beliebig genau* festlegen. Wir zeigen nun, dass sich der Zustandsbegriff in QM und kIM so erweitern lässt, dass er zufällige Anfangszustände mit einschließt.

Betrachten wir zunächst den qm Fall und gehen davon aus, dass aus einer Menge  $\{|\psi_a\rangle\}$  von reinen Zuständen der a-te gerade mit W  $P_a$  ausgewählt wird (also abzählbar viele, mögliche reine Zustände). Man beachte, dass die  $\psi_a$  hier *keine Basis* und auch nicht ein *orthogonales System* bilden sollen. Die Messwertstatistik einer Observablen  $\hat{A}$  wird nun durch 2 Arten von Zufall bestimmt: dem qm Zufall und dem klassischen Zufall der  $P_a$ . Der Erwartungswert von  $\hat{A}$  nimmt folgende Form an:

$$\begin{aligned} \langle \hat{A} \rangle &= \sum_a P_a \langle \psi_a | \hat{A} | \psi_a \rangle = \sum_a Sp(P_a |\psi_a\rangle\langle\psi_a| \hat{A}) \\ &= Sp(\hat{\rho} \hat{A}) \quad \text{mit} \quad \hat{\rho} = \sum_a P_a |\psi_a\rangle\langle\psi_a| \end{aligned}$$

Der Operator  $\rho$  kann benutzt werden, um jeden beliebigen Erwartungswert zu bilden, d.h. beim Übergang von einem reinen Zustand zu diesem *Ensemble von Zuständen* muss man lediglich die "Erwartungswertklammer"  $\langle \dots \rangle$  statt mit  $Sp(P_\psi \dots)$  mit  $Sp(\hat{\rho} \dots)$  bilden. Also beschreibt  $\hat{\rho}$  den Zustand mit

qm und klassischem Zufall vollständig. Dieser Operator heißt: **Zustandsoperator, Dichtematrix, statistischer Operator** (je nach Autor). Er beschreibt ein sogenanntes *Gemisch* oder einen **gemischten Zustand**, wenn er nicht zu einem Projektor (=reinem Zustand) degeneriert ist. Die folgenden 3 definierenden Eigenschaften liest man leicht aus der obigen Darstellung des Zustandsoperators ab

- (i)  $\hat{\rho} = \hat{\rho}^+$ , der Zustandsoperator ist hermitesch
- (ii)  $\hat{\rho} \geq 0$ , d.h.  $\langle \phi | \hat{\rho} | \phi \rangle \geq 0 \forall |\phi \rangle$  des Hilbertraums
- (iii)  $Sp(\hat{\rho}) = 1$

Man beachte, dass  $\hat{\rho}$  zwar gemäß seines physikalischen Ursprungs eine Darstellung über die Zustände des Ensembles wie oben angegeben hat, dass man ihn jedoch auch in anderen Darstellungen betrachten kann. Insbesondere kann man seine *Diagonaldarstellung* betrachten, denn er besitzt ja als hermitescher Operator ein VONS<sup>11</sup> Im System seiner Eigenzustände ist

$$\hat{\rho} = \sum_{\alpha} w_{\alpha} |\alpha \rangle \langle \alpha|$$

und die Eigenwerte  $w_{\alpha}$  sind nicht-negativ und

$$\sum_{\alpha} w_{\alpha} = Sp(\hat{\rho}) = 1.$$

Wir können also jedes gegebene Ensemble  $\{|\psi_{\alpha} \rangle\}$  durch ein äquivalentes Ensemble von  $|\alpha \rangle$  ersetzen, d.h. ein Ensemble bestimmt zwar den Zustandsoperator eindeutig, aber nicht umgekehrt.

In der kIM ist ein gemischter Zustand eines Ensembles durch eine WD auf dem Phasenraum definiert, (also eine ZV mit  $6N$  Komponenten) d.h. durch

$$\rho(\mathbf{p}, \mathbf{q}) \geq 0,$$

mit der man wie üblich Erwartungswerte bildet

$$\langle A \rangle = \int d\Gamma A(\mathbf{p}, \mathbf{q}) \rho(\mathbf{p}, \mathbf{q})$$

## 2.4 Zeitentwicklung gemischer Zustände

In der QM ist man es gewohnt, Zustände als zeitabhängig, Observable als zeitunabhängig zu betrachten (**Schrödingerbild**), obwohl es auch eine äquivalente Darstellung gibt, in der die Zustände zeitunabhängig und die Observablen

<sup>11</sup>vollständiges orthonormales System von Eigenzuständen

zeitabhängig sind (**Heisenbergbild**). Die Zeitabhängigkeit von  $\hat{\rho}$  erhält man ganz einfach aus der Definition und der Zeitabhängigkeit der reinen Zustände:

$$\begin{aligned}\hat{\rho}(t) &= \sum_a P_a |\psi_a(t)\rangle \langle \psi_a(t)| = \sum_a P_a e^{\frac{-i}{\hbar} \hat{H}t} |\psi_a\rangle \langle \psi_a| e^{\frac{i}{\hbar} \hat{H}t} \\ &= \hat{U}(t, 0) \hat{\rho} \hat{U}^\dagger(t, 0)\end{aligned}$$

Leitet man diese Form nach der Zeit ab, so erhält man die **Bewegungsgleichung für den Zustandsoperator**

$$\frac{d\hat{\rho}}{dt} = \frac{i}{\hbar} [\hat{\rho}, \hat{A}]_-$$

oder auch (**Liouville**)-**vonNeumann Gl.** Da es sich hier um die Bewegungsgleichung eines (hermiteschen) Operators handelt, kommt man leicht auf die Idee, diese Gleichung sei einfach die Bewegungsgleichung einer Observablen im Heisenbergbild. Das ist aber falsch! Sehen wir uns den Übergang zum Heisenbergbild für gemischte Zustände einmal etwas genauer an. Die das Heisenbergbild definierende Beziehung lautet:

$$\langle \hat{A} \rangle = Sp(\hat{\rho}(t) \hat{A}) = Sp(\hat{\rho} \hat{A}_H(t))$$

Hier setzen wir die zeitentwickelte Form von  $\hat{\rho}$  ein und stellen die Reihenfolge der Operatoren unter der Spur etwas um (zyklische Invarianz der Spur)

$$Sp(\hat{U}(t, 0) \hat{\rho} \hat{U}^\dagger(t, 0) \hat{A}) = Sp(\hat{\rho} \hat{U}^\dagger(t, 0) \hat{A} \hat{U}(t, 0))$$

woraus wir ablesen

$$\hat{A}_H(t) = \hat{U}^\dagger(t, 0) \hat{A} \hat{U}(t, 0)$$

Daraus ergibt sich die Zeitableitung von Operatoren im Heisenbergbild

$$\frac{d\hat{A}_H}{dt} = \frac{-i}{\hbar} [\hat{A}_H, \hat{H}]_-$$

Beachten Sie das Vorzeichen. Vorwärtszeitentwicklung bei Zuständen entspricht Rückwärtszeitentwicklung bei Observablen (und umgekehrt).

In der kLM ist man eher gewohnt, die Observablen als zeitabhängig zu betrachten und die Zustände als zeitunabhängig (s.o.) Mit Hilfe der sog. *Poisson-Klammer* lassen sich die Bewegungsgleichungen für Zustände und Observablen sehr analog zu QM darstellen. Die **Poissonklammer** ist definiert als eine Art Produkt, denn sie verknüpft 2 Phasenraumfunktionen  $f, g$  wie folgt:

$$\{f, g\} := \sum_i \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} - \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i}$$

Ganz offensichtlich ist die Poissonklammer anti-symmetrisch, so wie der Kommutator.

Betrachten wir nun die Zeitableitung einer Observablen  $A(\mathbf{p}(t), \mathbf{q}(t))$ :

$$\begin{aligned} \frac{dA}{dt} &= \sum_i \dot{q}_i \frac{\partial A}{\partial q_i} + \dot{p}_i \frac{\partial A}{\partial p_i} \\ &= \sum_i \frac{\partial H}{\partial p_i} \frac{\partial A}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial A}{\partial p_i} \\ &= \{H, A\} = -\{A, H\} \end{aligned}$$

Das sieht ganz analog zur Bewegungsgleichung für Observablen im Heisenbergbild aus. Um zu einer Bewegungsgleichung für  $\rho(\mathbf{p}, \mathbf{q}, t)$  zu kommen, müssten wir wieder die "Bilder wechseln", diesmal vom Heisenberg- ins Schrödingerbild (der kLM). Es geht aber auch einfacher. Da

$$\int d\Gamma \rho = 1$$

ist  $\rho$  die Dichte einer erhaltenen Größe und muss folglich eine Kontinuitätsgleichung erfüllen

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{j}$$

Die Stromdichte  $\mathbf{j}$  entsteht dadurch, dass jeder Ensemblepunkt (z.B.  $\mathbf{p}, \mathbf{q}$ ) sich im Phasenraum mit der entsprechenden Geschwindigkeit  $\mathbf{v} = (\dot{\mathbf{p}}, \dot{\mathbf{q}})$  bewegt, d.h.

$$\mathbf{j} = \rho \mathbf{v}$$

Setzen wir das zusammen, so ergibt sich

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= -\sum_i \frac{\partial}{\partial p_i} (\rho \dot{p}_i) + \frac{\partial}{\partial q_i} (\rho \dot{q}_i) \\ &= -\sum_i \frac{\partial}{\partial p_i} (\rho \dot{p}_i) + \frac{\partial}{\partial q_i} (\rho \dot{q}_i) \end{aligned}$$

Hier setzen wir die kanonischen Bewegungsgleichungen ein und benutzen schließlich noch die Divergenzfreiheit des Geschwindigkeitsfeldes (s.o.). Dann erhalten wir

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \sum_i \frac{\partial H}{\partial q_i} \frac{\partial \rho}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial \rho}{\partial q_i} \\ &= \{\rho, H\} \end{aligned}$$

wiederum in völliger Analogie zu QM! Diese Bewegungsgleichung ist als **Liouville Gl.** bekannt. Wem die Ableitung der Bewegungsgleichung noch nicht genau und ausführlich genug ist, den verweisen wir auf den Abschnitt zur Fokker-Planck Gl., denn die Liouville Gleichung wird dort als Spezialfall wieder erscheinen.

## 2.5 Verkürzung von Zuständen

Sehr oft interessiert man sich nicht für alle Details eines riesengroßen Systems (z.B: des gesamten Labors), sondern nur für die Eigenschaften eines *Untersystems*. Klassisch würde man sagen: Ein sehr großes System mag sich in einem reinen oder gemischten Zustand befinden, aber es ist aus 2 Untersystemen (1 und 2) zusammengesetzt, so dass

$$\rho_{1,2}(\mathbf{p}_1, \mathbf{q}_1, \mathbf{p}_2, \mathbf{q}_2)$$

Interessiert mich nur die Statistik der Phasenraumpunkte des Systems 1, so kann ich, den Regeln der klassischen W.theorie folgend, einfach die entsprechende Randdichte (s.o.) heranziehen, die über

$$\rho_1(\mathbf{p}_1, \mathbf{q}_1) = \int d\Gamma_2 \rho_{1,2}(\mathbf{p}_1, \mathbf{q}_1, \mathbf{p}_2, \mathbf{q}_2)$$

definiert ist. Mit dieser WD lassen sich alle Erwartungswerte von Funktionen bilden, die nur von Phasenraumpunkten des Systems 1 abhängen.

$$\langle A \rangle_{\rho_{1,2}} = \langle A \rangle_{\rho_1} \quad \text{falls} \quad A(\mathbf{p}_1, \mathbf{q}_1)$$

Man bezeichnet in der Physik  $\rho_1$  als **reduzierten Zustand**. Es ist unschwer zu erkennen, dass in der klm die Reduktion eines reinen Zustands zu einem reinen reduzierten Zustand führt, denn

$$\begin{aligned} \rho_1(\mathbf{p}_1, \mathbf{q}_1) &= \int d\Gamma_2 \delta(\mathbf{p}_1) \delta(\mathbf{q}_1) \delta(\mathbf{p}_2) \delta(\mathbf{q}_2) \\ &= \delta(\mathbf{p}_1) \delta(\mathbf{q}_1) \end{aligned}$$

In der QM ist die Reduktion von Zuständen geringfügig komplizierter, da man dort nicht von einer WD, sondern von einem Zustandsoperator  $\hat{\rho}_{12}$  ausgehen muss. Aus Basen  $\{|\psi_i\rangle\}$  des Hilbertraums von System 1,  $\mathcal{H}_1$ , und  $\{|\phi_n\rangle\}$  von  $\mathcal{H}_2$  kann man eine Basis des Gesamtsystems ( $\mathcal{H}_{12}$ ) als Produkt bilden, d.h.

$$|i, n\rangle := |\psi_i\rangle |\phi_n\rangle = |\psi_i\rangle \otimes |\phi_n\rangle$$

Observablen, die sich auf das Untersystem 1 beziehen, haben die Gestalt

$$\hat{A} = \hat{A}_1 \otimes \mathbb{I}$$

Ein (auf das Untersystem 1) reduzierter Zustandsoperator ist eine **partielle Spur** über  $\hat{\rho}_{12}$ , d.h.

$$\hat{\rho}_1 = Sp_2(\hat{\rho}_{12}) := \sum_n \sum_{i,j} |i\rangle \langle i, n| \hat{\rho}_{12} |j, n\rangle \langle j|,$$

was man auch kürzer (aber weniger klar) so schreibt:

$$\hat{\rho}_1 = Sp_2(\hat{\rho}_{12}) := \sum_n \langle n | \hat{\rho}_{12} | n \rangle$$

Die partielle Spur ist das Analogon zum klassischen "Ausintegrieren" des Phasenraums von System 2 in der WD. Offenbar ist die partielle Spur eines beliebigen Operator auf  $\mathcal{H}_{12}$  ein Operator auf  $\mathcal{H}_1$  (insbesondere hängt die Definition hängt nicht von der speziellen, verwendeten Basis ab). Man prüft auch leicht nach, dass die partielle Spur die Eigenschaften "hermitesch" und "positiv" nicht stört und das  $Sp_1(\hat{\rho}_1) = 1$ . Mit anderen Worten:  $\hat{\rho}_1$  ist ein Zustandsoperator auf  $\mathcal{H}_1$ . Will man die partielle Spur eines Operators als Operator auf ganz  $\mathcal{H}_{12}$  auffassen, so erweitert man ihn um den Einheitsoperator auf  $\mathcal{H}_2$ , d.h.

$$Sp_2(\hat{A}_{12}) \rightarrow Sp_2(\hat{A}_{12}) \otimes \mathbb{I}_2.$$

In diesem Sinne gilt dann offenbar, dass

$$Sp(Sp_2(\hat{A}_{12})) = Sp(\hat{A}_{12}).$$

Fasst man die partielle Spur als Operator auf dem Teilraum auf, so gilt stets

$$Sp(\hat{A}_{12}) = Sp_1(Sp_2(\hat{A}_{12})).$$

Mit dem reduzierten Zustandsoperator lassen sich alle Erwartungswerte von Operatoren bilden, die nur auf  $\mathcal{H}_1$  leben, denn

$$Sp_2\{(\hat{A}_1 \otimes \mathbb{I}_2) \hat{\rho}_{12}\} = \hat{A}_1 Sp_2(\hat{\rho}_{12})$$

also

$$Sp\{(\hat{A}_1 \otimes \mathbb{I}_2) \hat{\rho}_{12}\} = Sp\{Sp_2[(\hat{A}_1 \otimes \mathbb{I}_2) \hat{\rho}_{12}]\} = Sp_1(\hat{A}_1 \hat{\rho}_1)$$

In einem sehr wichtigen Punkt unterscheiden sich reduzierte Zustände in der kLM und in der QM: *Im Gegensatz zur kLM führt die Zustandsreduktion in der QM von einem reinen Zustand des Gesamtsystems i.a. auf einen gemischten Zustand des Teilsystems* Es gilt sogar noch mehr: *Jeder (!) Zustand eines Systems kann als Reduktion eines reinen Zustands von einem größeren System aufgefasst werden.* Um dies einzusehen, betrachte man den Zustandsoperator eines beliebigen Zustands in seiner Diagonaldarstellung

$$\hat{\rho} = \sum_n P_n |n\rangle\langle n|$$

Die  $|n\rangle$  bilden eine orthonormale Basis im Hilbertraum  $\mathcal{H}$  des Systems. Nun vergrößern wir unser System um eine "Kopie", d.h. der Hilbertraum des größeren Systems ist  $\mathcal{H} \otimes \mathcal{H}$ . In diesem betrachten wir den reinen Zustand

$$|\psi\rangle = \sum_n P_n^{1/2} |n\rangle \otimes |n\rangle$$

Der Projektor auf diesen Zustand hat die Gestalt

$$|\psi\rangle\langle\psi| = \sum_{n,m} P_n^{1/2} P_m^{1/2} (|n\rangle\langle m|)_1 \otimes (|n\rangle\langle m|)_2$$

Bilden wir nun die partielle Spur über 2, so ergibt sich

$$Sp_2 |\psi\rangle\langle\psi| = \sum_n P_n (|n\rangle\langle n|)_1$$

also gerade der beliebig vorgegebene Zustand. Dieser Mechanismus erschließt **Zustandsreduktion als Quelle klassischen Zufalls aus der QM!**

### 3 Langevin Gleichung des Brownschen Teilchens

Häufig hat man es mit folgender Problemstellung zu tun: In einem komplizierten N-Teilchensystem mit vielen Freiheitsgraden und Wechselwirkungen interessiert man sich für die Dynamik weniger, makroskopisch beobachtbarer Freiheitsgrade. Der Prototyp eines solchen Problems ist die *Brownsche Bewegung* eines "Pollenkorns" (=makroskopisch beobachtbares Teilchen ohne innere Freiheitsgrade) in Wasser ( $10^{23}$  Wassermoleküle, die untereinander und mit dem Pollenkorn wechselwirken).

Schon die prinzipielle Betrachtung der Problemstellung zeigt, dass es vielleicht nicht aussichtslos ist, eine Bewegungsgleichung aufzustellen, in der nur die Pollenkorneigenschaften (und eine Anfangsbedingung der Wassermoleküle) vorkommen. Schließlich haben wir es ja (klassisch betrachtet) "nur" mit sehr vielen Newtonschen Bewegungsgleichungen zu tun. Stellen wir uns vor, wir könnten die Bewegungsgleichungen der Wassermoleküle tatsächlich lösen, und zwar für alle Anfangsbedingungen und für alle Phasenraumtrajektorien  $x(t), v(t)$  des Pollenkorns <sup>12</sup> Diese Lösungen fassen wir mal symbolisch zusammen zu  $\mathbf{z}(t; \{x(\cdot), v(\cdot)\}, \mathbf{z}_0)$ , d.h.  $\mathbf{z}$  enthält alle Orte und Geschwindigkeiten der Wassermoleküle bei vorgegebener Trajektorie des Brownschen Teilchens und vorgegebenen Anfangsbedingungen  $\mathbf{z}_0$ . Nun können wir diese Lösungen

<sup>12</sup>Um Indizes, die für das Problem unerheblich sind, erst mal wegzulassen, hier eine 1-dimensionale Bewegung. Sie dürfen gern überall 3-d Indizes einfügen.

in die Newtonsche Bewegungsgleichung des Pollenkorns einsetzen, die ja so aussieht

$$M\ddot{x} = F(x, \mathbf{z}) + F^{ext}(x, v)$$

und —schwupps— schon hängt die rechte Seite nur noch von  $x, v$  und  $\mathbf{z}_0$  ab. Es gibt tatsächlich systematische Methoden, diesen Vorschlag in die Tat umzusetzen, die allerdings den Rahmen dieser Vorlesung sprengen und Gegenstand einer "Statistischen Physik II" Vorlesung sein können. Wir geben uns hier mit weniger zufrieden, wir wollen das Ergebnis solcher Untersuchungen nur plausibel machen. Zunächst einmal haben wir eine praktische Quelle des Zufalls, nämlich die sicher nicht kontrollierbaren, mikroskopischen Anfangsbedingungen. Daher führen wir gleich  $W$ . ein, um die Dynamik des Pollenkorns zu beschreiben. Als nächstes ist auch klar, dass es makroskopische, mittlere Kraftwirkungen der Wassermoleküle auf das Pollenkorn gibt, die wir als *Reibungskräfte* beobachten. Für nicht zu große  $v$  können wir das Stokesche Reibungsgesetz für diese Kräfte verwenden, also

$$\langle F \rangle = -\gamma v$$

(Mittelung über eine (noch unbekannt) Verteilung von Anfangsbedingungen, die Wassermoleküle im Gleichgewicht beschreibt). Die Newtonsche Bewegungsgleichung nimmt so die Form an

$$M\dot{v} = -\gamma v + f(t),$$

wobei  $f(t)$  alle Kraftwirkungen beschreibt, die im Mittel verschwinden, d.h.

$$\langle f(t) \rangle = 0$$

Die Details von  $f$  kennen wir nicht, der geniale Trick (und das Ergebnis der oben genannten Untersuchungen) besteht darin,  $f(t)$  durch *Zufallskräfte* zu ersetzen, deren statistische Eigenschaften konsistent mit unserem beschränkten Wissen über  $f$  sind. Mit "beschränktem Wissen" ist insbesondere gemeint:

- a)  $f$  ist zu jeder (makroskopischen) Zeit die Summe sehr vieler, molekularer Beiträge ("Stöße"). Die Verteilung der Kräfte bei Einzelstößen kennen wir zwar nicht, aber hier hilft der ZGS weiter und sagt uns, dass  $f$  einer Gaußschen Verteilung folgen sollte
- b)  $f(t)$  und  $f(t + \tau)$  sind zwei ZV, die aber (auf der makroskopischen Zeitskala betrachtet) schon nach sehr kurzer Zeit statistisch unabhängig bzw unkorreliert werden sollten. Tatsächlich kommen ja die mit dem Pollenkorn zusammenstoßenden Moleküle aus einem riesigen Reservoir und

können dorthin auch wieder verschwinden. Sie wechselwirken mit dem Pollen über eine Zeit, die in etwa gleich

$$\tau_c \sim \frac{\text{Radius des Pollenkorns}}{\text{Geschwindigkeit der Moleküle}}$$

ist und nur Ereignisse auf dieser Zeitskala können merklich korreliert sein. Wir erwarten also, dass  $\langle f(t)f(t+\tau) \rangle = g(\tau/\tau_c)$  ist, wobei  $g$  eine symmetrische, aber sehr schnell abfallende Funktion ist.

Eine grobe Abschätzung von  $\tau_c$  ist möglich, wenn man die zur Wassertemperatur gehörige, mittlere Geschwindigkeit der Moleküle kennt. Die werden wir uns später tatsächlich berechnen (s. Maxwell'sche Geschwindigkeitsverteilung), dann können Sie nachprüfen, dass  $\tau_c \sim 10^{-11} \text{ sec}$ . Schauen wir nicht so häufig hin, so können wir eine sehr nützliche Idealisierung vornehmen, und die unbekannte Funktion  $g$  durch eine  $\delta$ -Funktion ersetzen,

$$\langle f(t)f(t+\tau) \rangle = \lambda\delta(\tau) \quad (\text{weißes Rauschen}).$$

Damit haben wir alle 1. und 2. Momente (Korrelationsmatrix) aller ZV  $f(t)$  beschrieben und parametrisiert. Da die  $f$  Gaussisch sind, liegen ihre statistischen Eigenschaften somit fest. Nun können wir versuchen, mit Hilfe der Newton'schen Bewegungsgleichung mit Gaussisch verteiltem weißem Rauschen etwas Beobachtbares vorherzusagen.

### **Bewegungsgleichungen mit weißem Rauschen nennt man Langevin Gl.**

Langevingleichungen finden sich in vielen Gebieten der Physik und anderer Naturwissenschaften, sowie in den Finanzwissenschaften, Sozialwissenschaften, ... als wichtige Werkzeug der theoretischen Modellierung.

#### **Einstein Relation:**

Die Langevingleichung

$$M\dot{v} = -\gamma v + f(t)$$

ist, als gewöhnliche DGL betrachtet, sehr einfach zu lösen:

$$v(t) = v_0 e^{-\zeta t} + \frac{1}{M} e^{-\zeta t} \int_0^t d\tau e^{\zeta \tau} f(\tau) \quad (1)$$

wobei  $\zeta := \gamma/M$ . Der Mittelwert der Lösung ergibt keine neuen Einsichten, denn

$$\langle v(t) \rangle = v_0 e^{-\zeta t}$$

ist genau die Lösung, die zur mittleren Kraftwirkung (Stokessche Reibung) gehört. Interessanter ist da schon  $\langle v^2(t) \rangle$ , denn  $M \langle v^2 \rangle / 2$  ist ja gerade

die *mittlere kinetische Energie* des Brownschen Teilchens. Wir wollen an dieser Stelle ein Resultat der statistischen Physik des Gleichgewichts vorwegnehmen (s. Maxwell'sche Geschwindigkeitsverteilung), das aussagt: Die mittlere kinetische Energie eines Teilchens *im Gleichgewicht* ist  $kT/2$  (k: Boltzmann Konstante, T: Temperatur), im Gleichgewicht erwarten wir also

$$\langle v^2 \rangle = \frac{kT}{M}$$

Nun führen wir die Mittelung mit Hilfe der Lösung der Langevingl. aus

$$\langle v^2(t) \rangle = v_0^2 e^{-2\zeta t} + \frac{1}{M^2} e^{-2\zeta t} \int_0^t d\tau_1 \int_0^t d\tau_2 e^{\zeta(\tau_1 + \tau_2)} \langle f(\tau_1) f(\tau_2) \rangle$$

Beachten Sie, dass alle in  $f$  linearen Terme durch die Mittelung verschwinden. Setzen wir nun für die Korrelationsfunktion der  $f$  die  $\delta$ -Funktion ein (weißes Rauschen), so finden wir

$$\langle v^2(t) \rangle = v_0^2 e^{-2\zeta t} + \frac{\lambda}{2M\gamma} (1 - e^{-2\zeta t}).$$

Wir erkennen hier transiente Terme, für  $\zeta t \gg 1$  bleibt nur

$$\langle v^2(t) \rangle \rightarrow \frac{\lambda}{2M\gamma}$$

Sind alle Transienten abgeklungen, befindet sich das Pollenkorn im Gleichgewicht mit dem Wasserbad und durch Vergleich der beiden Ausdrücke für  $\langle v^2 \rangle$  schließen wir

$$\lambda = 2\gamma kT \quad (\text{Einstein Relation}).$$

Diese Relation ist bemerkenswert, denn sie verknüpft die *Stärke der Fluktuationen der stochastischen Kräfte* (sprich:  $\lambda$ ) mit der *Stärke der Dissipation* (sprich:  $\gamma$ ) über die Forderung, dass die Dynamik für lange Zeiten konsistent mit den Resultaten im Gleichgewicht sein muss. So wird sichergestellt, dass das System stets ins "richtige Gleichgewicht" strebt! Relationen dieser Art heißen auch *Fluktuations-Dissipations-Theoreme* (FDT).

### 3.1 Mittleres Auslenkungsquadrat. Diffusion

Eine der wichtigsten, experimentell überprüfbareren Aussagen über die Bewegung des Brownschen Teilchens betrifft das *mittlere Auslenkungsquadrat*. Starten wir unser Teilchen zur Zeit  $t = 0$  bei  $x = 0$ , so ist dieses Auslenkungsquadrat gegeben durch

$$\langle x(t)^2 \rangle = \int_0^t dt_1 \int_0^t dt_2 \langle v(t_1) v(t_2) \rangle.$$

Zur Berechnung muss man folglich die sog. *Geschwindigkeits-Korrelationsfunktion* bestimmen. Dazu benutzen wir die Lösung (1) und erhalten zunächst nach Mittelung:

$$\begin{aligned} \langle v(t_1)v(t_2) \rangle &= v_0^2 e^{-\zeta(t_1+t_2)} + \\ &+ \frac{1}{M^2} e^{-\zeta(t_1+t_2)} \int_0^{t_1} d\tau_1 \int_0^{t_2} d\tau_2 e^{\zeta(\tau_1+\tau_2)} \lambda \delta(\tau_1 - \tau_2) \end{aligned}$$

Bei den Integrationen gilt es zu beachten, dass für  $t_2 < t_1$  zuerst über  $\tau_2$  und für  $t_1 < t_2$  zuerst über  $\tau_1$  integriert werden muss, damit die Nullstelle des Arguments der  $\delta$ -Funktion im Integrationsbereich liegt. Der 2. Term auf der rechten Seite nimmt somit die Form an

$$\dots + \frac{1}{M^2} e^{-\zeta(t_1+t_2)} \int_0^{\min(t_1, t_2)} d\tau e^{2\zeta\tau}$$

Führen wir die  $\tau$  Integration aus, so erhalten wir

$$\dots + \frac{1}{M^2} e^{-\zeta(t_1+t_2)} \left( e^{2\zeta \min(t_1, t_2)} - 1 \right)$$

Dieser Ausdruck verschönert sich erheblich, wenn wir verwenden, dass

$$t_1 + t_2 - 2 \min(t_1, t_2) = |t_1 - t_2|$$

und so erhalten wir schließlich

$$\langle v(t_1)v(t_2) \rangle = \frac{\lambda}{2M\gamma} e^{-\zeta|t_1-t_2|} + \left( v_0^2 - \frac{\lambda}{2M\gamma} \right) e^{-\zeta(t_1+t_2)}$$

Der erste Term auf der rechten Seite hängt von der Zeitdifferenz ab, der zweite von der Summe der Zeiten. Für  $t_1, t_2 \gg \zeta$  können wir den 2. Term vergessen (Sie dürfen ihn natürlich auch gern noch ein bisschen mitschleppen und gucken, welche Zusatzterme er erzeugt.) und die asymptotische Form für die weitere Berechnung des mittleren Auslenkungsquadrats verwenden:

$$\langle x(t)^2 \rangle \rightarrow \int_0^t dt_1 \int_0^t dt_2 \frac{\lambda}{2M\gamma} e^{-\zeta|t_1-t_2|}$$

Das verbleibende Doppelintegral vom Typ  $I = \int_0^t dt_1 \int_0^t dt_2 f(t_1 - t_2)$  schlachten wir, indem wir zunächst über  $t_1$  integrieren (Stammfunktion von  $f$  sei  $F$ ). Wir erhalten

$$I = \int_0^t dt_2 [F(t - t_2) - F(-t_2)] = \int_0^t du [F(u) - F(-u)]$$

wobei wir im ersten Term  $u = t - t_2$  substituiert haben. Jetzt integrieren wir einmal partiell,

$$I = t[F(t) - F(-t)] - \int_0^t du u [f(u) + f(-u)],$$

und schreiben dieses Ergebnis schließlich so:

$$I = \int_0^t dt_1 \int_0^{t_1} dt_2 f(t_1 - t_2) = \int_0^t du (t - u)[f(u) + f(-u)],$$

Angewandt auf unser Auslenkungsquadrat ergibt sich also:

$$\langle x(t)^2 \rangle = \frac{\lambda}{2M\gamma} 2 \int_0^t du (t - u) e^{-\zeta u}$$

Wiederum für  $t \gg \zeta$  erhalten wir das Resultat

$$\langle x(t)^2 \rangle \rightarrow \frac{\lambda}{\gamma^2} t = \frac{kT}{\gamma} t =: 2Dt$$

Diese Abhängigkeit (Quadrat der Auslenkung  $\propto$  Zeit) ist das Markenzeichen der Brownschen Bewegung! Die Konstante  $D$  bezeichnet man auch als Diffusionskonstante, denn tatsächlich folgt die zeitliche Änderung der Aufenthaltswahrscheinlichkeitsdichte des Brownschen Teilchens  $\rho(x, t)$  einer Diffusionsgleichung, wie wir später sehen werden (vgl. Abschnitt über Fokker-Planck Gleichung).

### 3.2 Überdämpfte Brownsche Bewegung mit äußerer Kraft

Wirkt auf das Brownsche Teilchen neben der Reibungskraft noch eine weitere, äußere Kraft  $K_{ext}(x)$ , so muss man 2 DGLn 1. Ordnung betrachten

$$M\dot{v} = -\gamma v + K_{ext}(x) + f(t)$$

$$\dot{x} = v$$

was die analytische Behandlung erschwert. Zum Glück ist in vielen Fällen die Bedingung

$$\gamma v \gg M\dot{v}$$

erfüllt, d.h. die Reibungskräfte sind weiter größer als die Trägheit. Dann verläuft die Bewegung überdämpft (wie Sie aus der Mechanik wissen). Wir können einfach die linke Seite der ersten der beiden obigen Gleichungen auf 0 setzen und erhalten eine einzige Gleichung für  $x$ :

$$\gamma\dot{x} = K_{ext}(x) + f(t)$$

## 4 Markov-Prozesse und Mastergleichung

Die Langevingleichung ist ein sehr unmittelbarer Weg, eine stochastische Dynamik einzuführen, sie schöpft allerdings das Potential der Methode, die Dynamik eines komplexen Systems durch einen stochastischen Prozess zu beschreiben, bei Weitem nicht aus. Wir wollen hier etwas systematischer Resultate für die wohl wichtigste Klasse von stochastischen, dynamischen Modellen zusammenstellen, die uns noch gute Dienste leisten werden und die jeder Physiker "im Werkzeugkasten" haben sollte: die Markov Prozesse.

Zunächst klären wir noch etwas genauer, was eigentlich ein stochastischer Prozess im Sinne der W.theorie ist. Im Grunde ist die Definition sehr einleuchtend:

Ein **stochastischer Prozess** ist eine Familie von ZV  $\{X(t)\}_{t \in I}$ , wobei die Indexmenge  $I$  abzählbar, aber auch kontinuierlich sein kann.

Greifen wir uns also irgendeine (endliche) Menge von Zeitpunkten  $t_i$ ,  $i = 1 \dots k$  und zugehörigen ZV heraus, so können wir deren Eigenschaften vollständig durch ihre gemeinsame WD

$$\rho(x_k, t_k; \dots; x_1, t_1)$$

charakterisieren.

Hier haben wir eine Reihe von Notationen eingeführt, die im Zusammenhang mit WD stochastischer Prozesse sehr nützlich sind:

- (i) Statt  $\rho_{X(t_k) \dots X(t_1)}(x_k, \dots, x_1)$  schreiben wir  $\rho(x_k, t_k; \dots; x_1, t_1)$ . Das entspricht der oben schon eingeführten Physikernotation für WD von Zufallsvariablen, notiert aber die Zeitpunkte (die ja die ZV eindeutig festlegen) explizit mit. Diese Notation verliert keine Informationen, ist aber deutlich entschlackt.
- (ii) Wir wollen von nun an eine Ordnung in den gewählten Zeitpunkten einführen und zwar so dass wir vereinbaren:

$$t_k \geq t_{k-1} \dots \geq t_1 \geq t_0$$

- (iii) Außerdem sei (aus Gründen, die später das Leben vereinfachen) vereinbart, dass die Argumente der gemeinsamen WD bei stochastischen Prozessen stets so angeordnet werden, dass die ZV mit zunehmender Zeit von *rechts nach links* angeordnet werden, also die ZV zur spätesten Zeit am weitesten links steht.

Wollen wir uns nur mit den Eigenschaften einer Untermenge dieser ZV beschäftigen, so können wir die WD über die nicht benötigten ZV integrieren und zu den entsprechenden Randdichten (siehe oben) übergehen, also z.B.

$$\rho(x_{k-1}, t_{k-1}; \dots; x_1, t_1) = \int dx_k \rho(x_k, t_k; \dots; x_1, t_1).$$

Natürlich hätten wir genausogut über irgendein anderes (oder mehrere)  $x_i$  integrieren können, und so entsprechende Randdichten gewinnen. — So weit so gut. Nach einem berühmten Satz von Kolmogorov kann man aber auch den Spieß umdrehen und einen stochastischen Prozess *vollständig* durch WD von ZV zu *abzählbar vielen* Zeitpunkten charakterisieren. Mit anderen Worten: Eine Menge von WD

$$\begin{aligned} &\rho_1(x_1, t_1) \\ &\rho_2(x_2, t_2; x_1, t_1) \\ &\dots\dots\dots \\ &\rho_k(x_k, t_k; \dots, x_1, t_1) \\ &\dots\dots\dots \end{aligned}$$

definiert einen stochastischen Prozess (wohlgemerkt, hier ist  $\rho_1$  als Funktion von  $x_1$  und  $t_1$  zu verstehen, die weiteren WD entsprechend!). Nur eine Bedingung muss erfüllt sein: Die  $\rho_j$  müssen sich immer als Randdichten der WD mit mehr Variablen ( $\rho_k$ ,  $k > j$ ) ausdrücken lassen (**Konsistenzbedingung**). Diese Art der Beschreibung erscheint natürlich auf den ersten Blick recht mühselig, denn im Ernstfall muss man bis zu abzählbar unendlich vielen Funktionen mit bis zu abzählbar unendlich vielen Argumenten festlegen. Es gibt aber zumindest zwei Klassen von stochastischen Prozessen, die sich sehr effektiv über ihre WD charakterisieren lassen: die **vollständig unkorrelierten Prozesse (weißes Rauschen)** und die **Markov Prozesse**. Weißes Rauschen haben wir bereits im Zusammenhang mit den Zufallskräften in Langevingleichungen eingeführt. Die definierende Eigenschaft lässt sich durch WD ganz einfach ausdrücken: Alle ZV  $X(t)$  sind statistisch unabhängig, folglich faktorisieren alle gemeinsamen WD

$$\rho(x_k, t_k; \dots; x_1, t_1) = \prod_{i=1}^k \rho^{(i)}(x_i, t_i)$$

Sind außerdem alle  $X(t)$  identisch verteilt, so wird der ganze Prozess durch eine einzige WD einer ZV vollständig beschrieben, d.h. alle  $\rho^{(i)} = \rho$ .

### Die Markov Eigenschaft

Vollständig unkorrelierte Prozesse haben sozusagen überhaupt kein Erinnerungsvermögen; die Werte zu einer späteren Zeit hängen in keiner Weise von den Werten zu früheren Zeiten ab. Als vorhersagende Modelle sind sie daher nicht sonderlich nützlich. Allgemein kann man das Erinnerungsvermögen eines stochastischen Prozesses mit Hilfe von bedingten WD der Form

$$\rho(x_k, t_k | x_{k-1}, t_{k-1}; \dots x_1, t_1)$$

charakterisieren. Man fragt nach der WD einer ZV  $X(t_k)$  (können auch mal mehrere ZV sein) *unter der Bedingung*, dass der Prozess zu früheren Zeiten bestimmte Werte mit Sicherheit angenommen hat ( $X(t_{k-1}) = x_{k-1}$  etc). Als Physikerinnen und Physiker sollte uns sofort eine Klasse als besonders ausgezeichnet erscheinen, nämlich diejenige, bei der die Zukunft des Prozesses nur vom *jüngsten, mit Sicherheit angenommen Zustand abhängt und nicht von der weiteren Vorgeschichte*. Genau so sind vorhersagende Modelle physikalischer Theorien nämlich gestrickt. Diese Bedingung sieht, formal ausgedrückt folgendermaßen aus

$$\rho(x_k, t_k | x_{k-1}, t_{k-1}; \dots x_1, t_1) = \rho(x_k, t_k | x_{k-1}, t_{k-1})$$

und ist die definierende **Markov Eigenschaft**.

Die auf der rechten Seite auftretende, bedingte WD nennt man auch die **Übergangswahrscheinlichkeitsdichte (ÜWD)** des Markov Prozesses. Sie bestimmt den gesamten Prozess (beinahe) vollständig, denn alle gemeinsamen WD lassen sich rekursiv auf ÜWD und die Anfangsbedingung, d.h. die WD  $\rho(x_0, t_0)$  zur frühesten, überhaupt vorkommenden Zeit zurückführen. Das geht so: Zunächst machen wir von der Definition bedingter WD Gebrauch (siehe oben):

$$\rho(x_k t_k; \dots; x_1, t_1) = \rho(x_k, t_k | x_{k-1}, t_{k-1}; \dots x_1, t_1) \rho(x_{k-1}, t_{k-1}; \dots x_1, t_1)$$

Dann benutzen wir die Markov- Eigenschaft für den ersten Faktor, der damit zu einer ÜWD wird. Der zweite Faktor ist wieder eine gemeinsame WD, allerdings mit einem Argument weniger, und wir wiederholen daher unsere Prozedur rekursiv. Auf diese Art erhalten wir

$$\rho(x_k t_k; \dots; x_1, t_1) = \rho(x_k, t_k | x_{k-1}, t_{k-1}) \cdots \rho(x_1, t_1 | x_0, t_0) \rho(x_0, t_0)$$

und wir schließen daraus:

Um einen Markov Prozess festzulegen muss man nur die ÜWD und die AnfangsWD angeben.

Also: Statt —wie im allg. stochastischen Prozess— abzählbar unendlich viele gemeinsame WD angeben zu müssen, brauchen wir für Markov Prozesse nur 2 Funktionen mit überschaubar vielen Variablen. Das ist der erste Teil der großen “Markov Vereinfachung”.

Es kommt aber noch besser. Tatsächlich braucht man gar nicht die ganze ÜWD sondern nur die ÜWD für “ganz kleine Zeitschritte”, um den Prozess festzulegen. Um dies einzusehen, starten wir einmal von der gemeinsamen WD zweier ZV als Randdichte

$$\rho(x_3, t_3; x_1, t_1) = \int dx_2 \rho(x_3, t_3; x_2, t_2; x_1, t_1)$$

einer WD mit 3 ZV, wobei die zusätzliche ZV zu einer “Zwischenzeit” ( $t_3 > t_2 > t_1$ ) gehört. Nun benutzen wir auf der linken wie auch auf der rechten Seite unseren obigen “Reduktionstrick”:

$$\rho(x_3, t_3|x_1, t_1)\rho(x_1, t_1) = \int dx_2 \rho(x_3, t_3|x_2, t_2)\rho(x_2, t_2|x_1, t_1)\rho(x_1, t_1)$$

Abschließend dividieren wir noch durch  $\rho(x_1, t_1)$  und erhalten die sog. **Chapman-Kolmogorov Gleichung**:

$$\rho(x_3, t_3|x_1, t_1) = \int dx_2 \rho(x_3, t_3|x_2, t_2)\rho(x_2, t_2|x_1, t_1)$$

Damit haben wir unser Ziel erreicht, denn nun können wir diese Gleichung rekursiv verwenden und z.B. das  $\rho(x_2, t_2|x_1, t_1)$  wiederum durch ein Integral über die Werte einer ZV zu einer Zeit  $t_a$  mit  $t_2 > t_a > t_1$  ausdrücken usf. Mit anderen Worten:

Die ÜWD eines Markov Prozesses läßt sich aus der ÜWD für beliebig kleine Zeitschritte durch wiederholte Anwendung der Chapman-Kolmogorov Gleichung gewinnen

Das klingt schon ganz nett, aber irgendwie noch unhandlich! Man kann sich jedoch des Gefühls nicht erwehren, dass sich daraus eine *Bewegungsgleichung für die ÜWD* gewinnen lassen sollte, dann was ist eine Bewegungsgleichung schon anderes als eine Berechnung der Funktion über ihre Änderungen in beliebig kleinen Zeitschritten. Bevor wir das in Angriff nehmen, lohnt sich jedoch eine etwas weitergehende Fallunterscheidung von Markovprozessen.

### spezielle Klassen von Markovprozessen

- (i) Der “Zustandsraum”, d.h. der Bildraum der ZV  $X(t)$  kann (!) die reelle Achse (oder ein Stück davon) sein, kann aber auch eine *diskrete, d.h. höchstens abzählbare Menge* sein (z.B.  $\mathbb{N}$  für Teilchenzahlen). Im letzteren Fall

sind die WD natürlich entbehrlich (außerdem bestehen sie aus Summen eklektisch singulärer  $\delta$ -Funktionen) und wir können direkt mit den W. rechnen. Damit wir solche diskreten Zustandsräume gleich am Variablenamen erkennen, schreiben wir  $P(n_1, t_1)$  etc. Wir ersetzen also in allen obigen Diskussionen WD durch W (was problemlos möglich ist, wie Sie sich überzeugen können) und Integrale über  $x_i$  durch Summen über  $n_i$ . Die Chapman-Kolmogorov Gl. wird beispielsweise zu

$$P(n_3, t_3 | n_1, t_1) = \sum_{n_2} P(n_3, t_3 | n_2, t_2) P(n_2, t_2 | n_1, t_1)$$

- (ii) In der Definition stochastischer Prozesse gab es ja auch die Möglichkeit einer *diskreten Indexmenge für die Zeiten*. Gehen Sie die Diskussionen dieses Abschnitts durch, so stellen Sie wieder fest, dass sich für diesen Fall eigentlich fast nichts ändert, — außer eben der Tatsache, dass es jetzt überhaupt nur noch abzählbar viele Zeiten gibt. In der auf die Chapman Kolmogorov Gleichung folgenden Diskussion kann man dann natürlich nicht mehr von *beliebig kleinen* Zeitschritten sprechen, denn kleiner als zwischen 2 aufeinanderfolgenden, diskreten Zeiten kann der Zeitschritt nicht werden. Eine wichtige Unterklasse dieser Unterklasse sind Markov Prozesse in diskreter Zeit mit **festem Zeitschritt**, d.h.  $t_{i+1} - t_i = \Delta$  für zwei aufeinanderfolgende Zeiten ist konstant.
- (iii) Hat man einen Markov Prozess mit Eigenschaft (i) und Eigenschaft (ii), so nennt man das eine **Markov Kette**.
- (iv) Hängt die ÜWD bzw die ÜW nur von der Differenz der Zeitargumente ab, also  $\rho(x_2, t_2 | x_1, t_1) = \rho(x_2 | x_1; t_2 - t_1)$  bzw  $P(n_2, t_2 | n_1, t_1) = P(n_2 | n_1; t_2 - t_1)$ , so heißt der Markov-Prozess **stationär** oder **zeitlich homogen**.

### Mastergleichung

So, nun sind wir vorbereitet, um Bewegungsgleichungen für ÜW oder ÜWD abzuleiten. Fangen wir mit dem einfachsten Fall an, den Markovketten mit konstantem Zeitschritt. Die Bewegungsgleichung ist nichts anderes als die Chapman Kolmogorov Gleichung, für  $t_3 = t_2 + \Delta = t + \Delta$ , also

$$P(n_3, t + \Delta | n_1, t_1) = \sum_{n_2} P(n_3 | n_2; \Delta) P(n_2, t | n_1, t_1)$$

Sie beschreibt, wie aus der ÜW zur (diskreten) Zeit  $t$  die ÜW einen Zeitschritt später entsteht, nämlich mit Hilfe der **1-Schritt ÜW**  $P(n | m; \Delta)$ , **die folglich den ganzen Prozess (bis auf die Anfangsw.) festlegen**. Zweckmäßigerweise legt man den willkürlich wählbaren Zeitpunkt  $t_1$  auf die Anfangszeit  $t_0$ .

Man kann diese Gleichung auch in etwas anderer, aber äquivalenter Form angeben, indem man von der Normierung der ÜW Gebrauch macht. Da  $P(n|m; \Delta)$  ja eine W für die Variable  $n$  ist, ist sie so normiert:

$$\sum_n P(n|m; \Delta) = 1$$

Nun betrachten wir die *Änderung* der ÜW in einem Zeitschritt.

$$P(n, t + \Delta | n_0, t_0) - P(n, t | n_0, t_0) = \sum_m P(n|m; \Delta) P(m, t | n_0, t_0) - P(n, t | n_0, t_0)$$

und wegen der Normierung können wir das auch so schreiben:

$$P(n, t + \Delta | n_0, t_0) - P(n, t | n_0, t_0) = \sum_m [P(n|m; \Delta) P(m, t | n_0, t_0) - P(m|n; \Delta) P(n, t | n_0, t_0)].$$

Hier haben wir lediglich einen Faktor 1 in komplizierter Form eingefügt, und man fragt sich, wozu das nützen soll.

Die obige Form ist dann sehr nützlich, wenn wir einen *Limes zu beliebig kleinen Zeitschritten  $\Delta$ , sprich, zu kontinuierlichen Zeiten ausführen wollen*. Dazu dividieren wir die Gleichung durch  $\Delta$  und stellen zunächst mal fest, dass ein Limes nicht so ohne Weiteres existiert. Beachten Sie, dass die ÜW *per definitionem* die Eigenschaft haben

$$P(n|m; \Delta \rightarrow 0) = \delta_{nm}$$

so dass es scheint, als würden Terme der Art  $\delta_{nm}/\Delta$  herumgeistern. Nun sieht man aber sehr leicht, dass sich diese Terme auf der rechten Seite der obigen Gleichung wegheben. Daher reicht es zu fordern

$$P(n|m; \Delta) = \delta_{nm} + r(n|m)\Delta + o(\Delta)$$

wobei  $o(\Delta)$  Terme bezeichnet, die schneller als  $\Delta$  verschwinden. So erhält man

$$\begin{aligned} \frac{dP(n, t | n_0, t_0)}{dt} &= \lim_{\Delta \rightarrow 0} \frac{P(n, t + \Delta | n_0, t_0) - P(n, t | n_0, t_0)}{\Delta} \\ &= \sum_m [r(n|m)P(m, t | n_0, t_0) - r(m|n)P(n, t | n_0, t_0)]. \end{aligned}$$

Dies ist die sogenannte **Mastergleichung**.

Die  $r(n|m)$  für  $n \neq m$  nennt man **Übergangsraten**, denn es sind ÜW *pro Zeit*. Beachten Sie, dass die W., in einem kleinen Zeitschritt  $dt$ <sup>13</sup> von  $m$  nach  $n(\neq m)$  zu springen, gerade (bis auf  $o(dt)$  Terme)

$$P(n, t + dt | m, t) = \delta_{nm} + r(n|m)dt = r(n|m)dt$$

<sup>13</sup>Falls  $t$  kontinuierlich ist, bezeichnen wir kleine Zeitschritte stets als  $dt$ , egal ob "infinitesimal" oder noch endlich.

ist, während die W. *nicht zu springen* gerade

$$P(n, t + dt | n, t) = 1 + r(n|n)dt$$

ist. Die Normierung von  $P(n, t + dt | m, t)$  verlangt aber

$$\sum_n P(n, t + dt | m, t) = \left( \sum_{n(\neq m)} r(n|m)dt \right) + 1 + r(n|n)dt = 1$$

Daher ist  $\sum_m r(m|n) = 0$  und wir können  $r(n|n)$  durch die Übergangsraten ausdrücken, d.h., die W. *nicht zu springen* ist

$$P(n, t + dt | n, t) = 1 - \sum_{n(\neq m)} r(n|m)dt.$$

Tatsächlich hängt die rechte Seite der Mastergleichung gar nicht von  $r(n|n)$  ab, denn dieser Term hebt sich heraus.

Man prüft auch leicht nach, dass die Mastergleichung die Normierung der ÜW stets erhält, d.h

$$\frac{d}{dt} \sum_n P(n, t | n_0 t_0) = 0.$$

Die beiden Terme mit unterschiedlichem Vorzeichen auf der rechten Seite der Mastergleichung haben eine intuitive Interpretation: sie bilanzieren die Änderungen der ÜW. Der positive Term vergrößert die ÜW bei  $n$  durch Terme, die Sprünge von  $m$  nach  $n$  enthalten (sog. *scattering in*), während der negative Term alle Sprünge von  $n$  weg zu irgendeinem anderen  $m$  beschreibt (sog. *scattering out*). Die einzelnen Terme kann man graphisch darstellen, indem man die  $n$  zu Knoten eines Graphen macht und für jedes nicht-verschwindende  $r(n|m)$  eine gerichtete Kante von  $m$  nach  $n$  in den Graphen einfügt (der man das "Gewicht"  $r(n|m)$  gibt.) Solche *transition diagrams* sind nützlich, wenn man rasch überblicken will, ob von jedem Ausgangspunkt  $n_0$  aus jedes andere  $n$  mit nicht-verschwindender W. erreicht werden kann. In diesem Fall nennt man den Markov Prozess **ergodisch**. Damit er ergodisch ist, muss man offenbar zwischen je zwei beliebig wählbaren Knoten immer mindestens einen gerichteten Weg finden.

Man kann die Mastergleichung in einer noch kompakteren Form angeben, indem man eine **Vektor-Matrix Notation** einführt. Dazu betrachtet man  $P(n, t | n_0, t_0)$  als die  $n$ -te Komponente eines Vektors  $\mathbf{P}(t)$ <sup>14</sup>, also  $[\mathbf{P}(t)]_n = P(n, t | n_0, t_0)$ . Die 1-Schritt ÜW für Markov-Ketten fasst man entsprechend als eine Matrix  $\mathbb{T}$  auf, d.h.

$$\mathbb{T}_{nm} = P(n|m; \Delta)$$

<sup>14</sup>Die Erinnerung an den Anfangszustand lässt man beim Vektorsymbol weg

Für kleine Zeitschritte soll der Limes zu kontinuierlichen Zeiten möglich sein, d.h. (s.o.) in Matrixnotation

$$\mathbb{T} = \mathbb{I} + \mathcal{R}\Delta + o(\Delta)$$

wobei  $\mathbb{I}$  die Einheitsmatrix bezeichnet und die Matrix  $\mathcal{R}$  gerade die Übergangsraten enthält, dh.

$$\mathcal{R}_{nm} = r(n|m) \quad \text{für } n \neq m$$

$$\mathcal{R}_{nn} = - \sum_{m(\neq n)} r(m|n).$$

Jetzt können wir die Chapman-Komogorov Gl. für einen Schritt so kurz schreiben

$$\mathbf{P}(t + \Delta) = \mathbb{T}\mathbf{P}(t),$$

denn die rechte Seite entspricht genau die Anwendung der Matrix  $\mathbb{T}$  auf den Vektor  $\mathbf{P}$  (s.o.). Die Aufstellung der Mastergl. ist ein Klacks:

$$\mathbf{P}(t + \Delta) = (\mathbb{I} + \mathcal{R}\Delta)\mathbf{P}(t) = \mathbf{P}(t) + \mathcal{R}\Delta\mathbf{P}(t),$$

$$\frac{d\mathbf{P}(t)}{dt} = \mathcal{R}\mathbf{P}(t)$$

Man bezeichnet die Matrix  $\mathbb{T}$  als *Übergangsmatrix (transition matrix)* oder *stochastische Matrix*, während die Matrix  $\mathcal{R}$  *Generator* des Markovprozesses heißt. Man kann sich nun fragen, ob man die Mastergleichung auch für einen *kontinuierlichen Zustandsraum* benutzen kann, also statt diskreter  $n$  kontinuierliche, reelle  $x$ . Dabei stößt man auf eine weitere Klasseneinteilung von Markovprozessen. Geht man nämlich von der Chapman-Kolmogorov Gl. für einen zeitlich diskreten Prozess aus, also

$$\rho(x, t + \Delta | x_0, t_0) = \int dy \rho(x|y; \Delta) \rho(y, t | x_0, t_0),$$

so sieht Alles recht ähnlich zum Fall der diskreten Zustände aus. Man muss nur ÜW durch ÜWD ersetzen und Summen durch Integrale. Die ÜWD müssen

$$\rho(x|y; \Delta \rightarrow 0) = \delta(y - x)$$

erfüllen, so dass man auch wieder Übergangsraten einführen kann, indem man für kleiner werdende Zeitschritte verlangt

$$\rho(x|y; \Delta) = \delta(x - y) + r(x|y)\Delta + o(\Delta).$$

Falls nun aber die  $r(x|y) \neq 0$  für  $|x - y|$  endlich, so bedeutet dies, dass der Prozess **unstetige Realisierungen** haben muss, denn er springt ja (in verschwindend kleinen Zeitintervallen) von  $y$  nach  $x$ . Will man jedoch Markov Prozesse

mit **stetigen Realisierungen**, so muss man die Bedingungen an  $\rho(x|y; \Delta)$  etwas anders stellen. Das werden wir im nächsten Abschnitt tun.

Für unstetige Realisierungen hingegen sieht die Mastergleichung für ÜWD einfach so aus, wie man es naiv erwartet, wenn man von diskreten  $n$  zu kontinuierlichen  $x$  übergeht:

$$\frac{\partial \rho(x, t|x_0, t_0)}{\partial t} = \int dy [r(x|y)\rho(y, t|x_0, t_0) - r(y|x)\rho(x, t|x_0, t_0)]$$

#### 4.1 Detaillierte Balance und das Streben ins Gleichgewicht

Betrachten wir einmal stationäre Lösungen der Mastergleichung, d.h. Lösungen  $\mathbf{P}^0$ , die nicht von der Zeit abhängen. Sie erfüllen die Gleichung

$$\mathcal{R}\mathbf{P}^0 = 0,$$

mit anderen Worten, der Vektor  $\mathbf{P}^0$  ist Lösung eines homogenen, linearen Gleichungssystems. Es hat die Form

$$\sum_m r(n|m)P^0(m) - r(m|n)P^0(n) = 0$$

Man sagt: Der Markov-Prozess erfüllt **detaillierte Balance** (DB), falls die obige Summe termweise verschwindet, also falls

$$r(n|m)P^0(m) = r(m|n)P^0(n)$$

für alle  $n \neq m$  ist. In einem transition diagram findet man also stets Paare von Mikrozuständen, die durch Paare von gerichteten Kanten ("hin" und "her") verknüpft sind.

Obwohl in einem stationären Zustand die W. zeitunabhängig ist, finden natürlich weiterhin Übergänge zwischen den Mikrozuständen  $n$  statt. Wenn wir bei dieser *mikroskopischen Dynamik* nun den Zeitpfeil umdrehen, so laufen alle Übergängen  $n \rightarrow m$  mit unveränderten Raten in umgekehrter Richtung  $m \rightarrow n$  ab, falls diese mikroskopische Dynamik *invariant unter Zeitumkehr* ist. Von der klassischen Mechanik und der Quantenmechanik wissen wir das, daher liegt es nahe, die Konsequenzen dieser sogenannten **Mikroreversibilität** für konsistente stochastische Beschreibungen zu fordern. Die Übergänge zwischen je zwei Mikrozuständen sind prinzipiell beobachtbar, und nur im Falle detaillierter Balance findet man keine Änderungen bei einer Umkehr der Richtung aller gerichteten Kanten im transition diagram. Also

**Mikroreversibilität  $\rightarrow$  Detaillierte Balance**

Markov Prozesse mit detaillierter Balance haben eine grundlegend wichtige, dynamische Eigenschaft, die wir hier vorstellen und beweisen wollen:

### Streben ins Gleichgewicht

Falls DB, so strebt jede W.  $P(n, t = 0)$  für  $t \rightarrow \infty$  gegen das eindeutige  $P^0$  (stationäre Lösung)

### Die dynamische Entropie

$$S := - \sum_n P(n) \ln \left( \frac{P(n)}{P^0(n)} \right)$$

ist sogar eine *Lyapunov Funktion*.

Letzteres ist auch gleich der Schlüssel zum Beweis, denn Lyapunov Funktion zu sein bedeutet:

- (i)  $S[P^0(n)] = 0$
- (ii)  $S$  ist maximal für  $P(n) = P^0(n)$
- (iii)  $\dot{S} \geq 0$  und  $\dot{S} = 0$  für  $P(n) = P^0(n)$

Wenn alle diese Eigenschaften gezeigt sind, ist es offensichtlich, dass jede Anfangsverteilung ins Gleichgewicht streben muss, denn die Dynamik läßt  $S$  wachsen bis es seine Maximumsstelle (sprich  $P^0$ ) erreicht hat.

Wie beweisen wir also (i), (ii) und (iii)?

(i) ist trivial!

(ii) folgt aus der Ungleichung

$$\ln x \leq 1 - x$$

die wir anwenden auf

$$\begin{aligned} S &= \sum_n P(n) \ln \left( \frac{P^0(n)}{P(n)} \right) \\ &\leq \sum_n P(n) \left( 1 - \frac{P^0(n)}{P(n)} \right) = \sum_n (P(n) - P^0(n)) = 0 \end{aligned}$$

Und nun noch (iii):

Wir starten von

$$\dot{S} = - \sum_n \dot{P}(n) \ln \left( \frac{P(n)}{P^0(n)} \right) - \sum_n \dot{P}(n)$$

Der 2. Term auf der rechten Seite ist ganz klar Null, denn  $\sum_n P(n) = 1$ . Für die Zeitableitung der  $W$ . setzen wir nun die Mastergleichung ein

$$\dot{S} = - \sum_n \sum_m (r(n|m)P(m) - r(m|n)P(n)) \ln \left( \frac{P(n)}{P^0(n)} \right)$$

Jetzt führen wir

$$x(n) := \left( \frac{P(n)}{P^0(n)} \right)$$

ein, so dass

$$\dot{S} = - \sum_n \sum_m (r(n|m)P^0(m)x(m) - r(m|n)P^0(n)x(n)) \ln x(n)$$

und definieren noch

$$w(n|m) := r(n|m)P^0(m) = w(m|n) \text{ (wegen DB)}$$

Damit wird

$$\begin{aligned} \dot{S} &= - \sum_{n,m} w(n|m) [x(m) - x(n)] \ln x(n) \\ &= \sum_{m,n} w(n|m) [x(n) - x(m)] \ln x(m), \end{aligned}$$

Letzteres durch Vertauschen der Summationsindizes  $m, n$ . Nun setzen wir  $\dot{S}$  aus jedem der beiden obigen Formen je zur Hälfte zusammen

$$\dot{S} = -\frac{1}{2} \sum_{n,m} w(n|m) [x(m) - x(n)] \ln \left( \frac{x(n)}{x(m)} \right)$$

In dieser Gestalt können wir nun eine Standard-Ungleichung benutzen, nämlich

$$[x - y] \ln \frac{x}{y} \geq 0,$$

um schließen zu können

$$\dot{S} \geq 0,$$

was behauptet wurde.

## 4.2 Anwendungsbeispiel: Die chemische Mastergleichung

Eine wichtige Klasse von Anwendungen der Mastergleichung ist die stochastische Dynamik chemischer Reaktionen. Wir betrachten  $n$  Molekülsorten  $A_i$ ,  $i = 1 \cdots n$ . Von der Sorte  $A_i$  gebe es  $N_i$  Moleküle, die in einem Reaktionsvolumen  $\Omega$  homogen verteilt sind (ihre Dichte ist also  $n_i = N_i/\Omega$ ). Wir fassen die Teilchenzahlen in ein  $n$ -Tupel zusammen, das den Mikrozustand des Systems festlegt.

$$\mathbf{N} = (N_1, \cdots, N_n).$$

Die  $N_i$  können sich durch *chemische Reaktionen* verändern. Wir betrachten  $r$  verschiedene Reaktionen, wobei sich in der Reaktion Nr.  $\alpha$  ( $\alpha = 1 \cdots r$ ) die Anzahl  $N_i$  gerade um  $\Delta_{\alpha i}$  ändert, d.h. jede Reaktion  $\alpha$  ändert



$\Delta_{\alpha}$  bezeichnet man auch als "stöchiometrischen Vektor". Chemische Reaktionen sind stochastische Ereignisse, von denen bekannt ist, dass die W. für die Reaktion  $\alpha$  im Zeitintervall  $[t, t + dt]$  nur vom Zustand  $\mathbf{N}$  zur Zeit  $t$  abhängt (und nicht von der weiteren Vorgeschichte), mit anderen Worten: chemische Reaktionen bilden einen Markovprozess. Die Übergangsraten (in der chem. Literatur "propensities") notiert man wie folgt

$$r(\mathbf{N} + \Delta_{\alpha} | \mathbf{N}) dt = r_{\alpha}(\mathbf{N}) dt$$

und nennt die  $r_{\alpha}$  *Reaktionsraten*. Sie legen den Prozess fest. Die chemische Mastergleichung ist in dieser Notation

$$\frac{dP(\mathbf{N})}{dt} = \sum_{\alpha} r_{\alpha}(\mathbf{N} - \Delta_{\alpha}) P(\mathbf{N} - \Delta_{\alpha}) - r_{\alpha}(\mathbf{N}) P(\mathbf{N})$$

Die Abhängigkeit der Reaktionsraten von den Teilchenzahlen läßt sich unter sog. "well-stirred conditions" einfach angeben.

Für *monomolekulare* Reaktionen

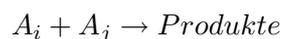


wird angenommen, dass alle Umwandlungen (statistisch) unabhängig voneinander in jedem  $A_i$  Molekül mit gleicher Rate stattfinden, dann ist

$$r(N_i) = k_1 N_i$$

mit einer Konstanten  $k_1$ , die man auch **spezifische Reaktionsrate** nennt, und die aus mikroskopischen Theorien oder aus Experimenten ermittelt werden muss.

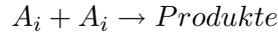
Für *bimolekulare* Reaktionen



müssen sich 2 Moleküle begegnen, da chemische Reaktionen nur stattfinden, wenn die Reaktionspartner sich auf atomare Abstände angenähert haben. "Well-stirred" bedeutet nun, dass *die W. sich zu begegnen, für alle möglichen Paare gleich groß ist*. Dann ist die Reaktionsrate

$$r(N_i, N_j) = k_2 N_i N_j$$

denn  $N_i N_j$  ist gerade die Anzahl der möglichen Paare. Falls sich ein Paar gleicher Molekülsorte treffen müssen, also für Reaktionen



ist entsprechend

$$r(N_i) = k_2 N_i (N_i - 1).$$

Da sich die Moleküle wie Brownsche Teilchen bewegen, muss offenbar zwischen 2 Reaktionen so viel Zeit vergangen sein, dass alle möglichen Paare sich einmal via Brownscher Bewegung begegnen können. Daraus ergibt sich eine wichtige Einschränkung an die Größe eines Reaktionsvolumens, in dem Reaktionen noch "well-stirred" ablaufen können. Nehmen wir z.B. eine kubische Box als Reaktionsvolumen, so ist jede Kantenlänge  $L = \Omega^{1/3}$ . Die Zeit zwischen zwei Reaktionen ist ungefähr  $(\text{Reaktionsrate } r_\alpha)^{-1}$ . In dieser Zeit muss ein Molekül per Brownscher Bewegung in etwa die Systemlänge durchmessen. Im Mittel kommt ein Brownsches Teilchen überhaupt nicht voran, aber die "Ausflüge" können wir über das mittlere Auslenkungsquadrat beschreiben, das, wie wir oben gesehen haben linear in der Zeit wächst

$$\langle x^2 \rangle \sim Dt$$

wobei  $D$ , bis auf eine dimensionsabhängige numerische Konstante ( $2 \dots 6$ ) gerade die Diffusionskonstante der betrachteten Molekülsorte ist (s.u.). (Beachten Sie auch, dass aufgrund des Charakters der Brownschen Bewegung das Molekül "überall im System" herumgeirrt ist, wenn  $\langle x^2 \rangle \gg L^2$ ). Also muss gelten, dass

$$\frac{D}{r_\alpha} \gg L^2 = \Omega^{2/3}$$

sein muss, damit die Reaktionen "well-stirred" ablaufen können. Als Beispiel setzen wir einmal typische Werte für biochemische Reaktionen in einer biologischen Zelle ein. Für Proteine in der Zelle ist  $D \approx 10 \dots 50 \mu\text{m}^2/\text{sec}$  und typische Reaktionsraten liegen bei  $0.1 \dots 100 \text{sec}^{-1}$ . Damit reicht  $D/r$  von  $0.1 \dots 500 \mu\text{m}^2$ , so dass "well-stirred" Volumen Längenskalen

$$L \ll 22.4 \dots 0.32 \mu\text{m}$$

haben müssen. Eine typische Zelle (kein Bakterium, sondern so eine, wie Sie im Körper haben) hat eine Längenskala von  $\approx 10 \mu\text{m}$ . Hieraus können Sie bereits schließen, dass biologische Zellen i. a. keine "well-stirred" Volumina sein können und dass man einige tausend (!) "well-stirred" Reaktionsvolumina braucht, um eine Zelle zu beschreiben.<sup>15</sup>

<sup>15</sup>Tatsächlich braucht man noch viel mehr, denn die Zelle ist in viele Kompartments unterteilt, die Molekülsorten auf noch kleineren Längenskalen lokalisieren können.

## 5 Fokker-Planck-Gleichung und Langevin Prozesse

### 5.1 Die Kramers-Moyal Entwicklung

Wir wollen nun eine Bewegungsgleichung für die ÜWD von Markovprozessen konstruieren, die **stetige Trajektorien** haben (die sollten aus naheliegenden Gründen in der Physik ziemlich wichtig sein). Insbesondere werden wir dadurch Bewegungsgleichungen für ÜWD finden, die genau unsere früher schon eingeführten Langevin-Prozesse beschreiben. Die resultierende Form der Mastergleichung heißt **Fokker-Planck Gleichung** und der Weg, sie zu konstruieren heißt **Kramers-Moyal Entwicklung**.

Beginnen wir mit einer Erleichterung der Notation. Damit wir nicht immer die Erinnerung an den Anfangszustand mit uns herumschleppen müssen, schreiben wir in diesem Abschnitt *ausnahmsweise* mal

$$\rho(x, t) := \rho(x, t | x_0, t_0)$$

Aber Achtung: Nur der Anfangszustand wird nicht explizit notiert, alle anderen Zustände in ÜWD schon. Wir gehen nun aus von

$$\frac{\partial \rho(x, t)}{\partial t} = \lim_{dt \rightarrow 0} \frac{1}{dt} [\rho(x, t + dt) - \rho(x, t)]$$

und drücken  $\rho(x, t + dt)$  mit Hilfe der Chapman Kolmogorov Gleichung aus

$$\rho(x, t + dt) = \int dz \rho(x, t + dt | z, t) \rho(z, t)$$

wodurch wir einen Zwischenpunkt namens  $z$  eingeführt haben, der natürlich für  $dt \rightarrow 0$  gegen  $x$  rutschen muss. Für die Differenz  $x - z$  gönnen wir uns einen extra Bezeichner

$$x =: \Delta x + z$$

$\Delta x$  nennen wir einen "kleinen Schritt" oder "Inkrement". Wir können jederzeit statt  $x$  auch  $\Delta x$  als Zufallsvariable benutzen (unterscheiden sich ja nur um die nicht-zufällige Konstante  $z$ ). Falls wir diese Auffassung ganz explizit machen wollen, schreiben wir

$$\rho(x, t + dt | z, t) = \rho(z + \Delta x, t + dt | z, t) =: \rho(\Delta x, t + dt | z, t)$$

In  $\Delta x$  soll nun entwickelt werden. Bevor wir diese Entwicklung starten, führen wir noch eine "Glättungsprozedur" durch: wir betrachten statt der Zeitableitung das Integral

$$\int dx g(x) \partial_t \rho(x, t),$$

wobei die Funktion beliebig, aber glatt sein soll und auf dem Rand des betrachteten Gebietes von  $x$  (also evtl auch im Unendlichen) glatt verschwinden soll. Man kann dies auch so ausdrücken: statt der Zeitableitung der ÜWD betrachten wir die Zeitableitung beliebiger Erwartungswerte von Observablen  $g(x)$

$$\partial_t \langle g \rangle = \int dx g(x) \partial_t \rho(x, t) \quad (*)$$

Daraus gewinnen wir hinterher die Zeitableitung von  $\rho$  aufgrund der Beliebbarkeit von  $g$ . Der Grenzwert nimmt nun folgende Gestalt an:

$$\begin{aligned} & \lim_{dt \rightarrow 0} \frac{1}{dt} \left[ \int dx \int dz g(x) \rho(x, t + dt | z, t) \rho(z, t) - \int dx g(x) \rho(x, t) \right] \\ & \lim_{dt \rightarrow 0} \frac{1}{dt} \left[ \int d\Delta x \int dz g(z + \Delta x) \rho(\Delta x, t + dt | z, t) \rho(z, t) - \int dx g(x) \rho(x, t) \right] \end{aligned}$$

Nun entwickeln wir

$$g(z + \Delta x) = g(z) + g'(z) \Delta x + \frac{g''(z)}{2} \Delta x^2 + O(\Delta x^3),$$

setzen diese Entwicklung ein und vertauschen schließlich noch die Integrationsreihenfolge von  $\Delta x$  und  $z$ . Die  $\Delta x$  Integrationen produzieren Momente der Variablen  $\Delta x$ , die wir so bezeichnen:

$$\langle (\Delta x)^n \rangle := \int d\Delta x (\Delta x)^n \rho(\Delta x, t + dt | z, t)$$

Beachten Sie, dass der erste Term der Entwicklung sich genau heraushebt! Somit bleibt

$$\lim_{dt \rightarrow 0} \frac{1}{dt} \left[ \int dz \left( g'(z) \langle \Delta x \rangle + \frac{g''(z)}{2} \langle (\Delta x)^2 \rangle + O(\langle (\Delta x)^3 \rangle) \right) \rho(z, t) \right]$$

Die Striche am  $g(z)$  bezeichnen Ableitungen nach  $z$ . Diese wälzen wir nun durch partielle Integrationen auf den Rest des jeweiligen Integranden ab. Die Randterme sind wegen des glatten Verschwindens von  $g$  alle Null. Das Ergebnis kann man schon als Kramers-Moyal Entwicklung (KME) bezeichnen

$$\begin{aligned} & \lim_{dt \rightarrow 0} \int dz g(z) \left[ -\frac{\partial}{\partial z} \left( \frac{\langle \Delta x \rangle}{dt} \rho(z, t) \right) + \frac{1}{2} \frac{\partial^2}{\partial z^2} \left( \frac{\langle (\Delta x)^2 \rangle}{dt} \rho(z, t) \right) + \right. \\ & \left. + O\left( \frac{\langle (\Delta x)^3 \rangle}{dt} \right) \right] \quad (**) \end{aligned}$$

Entscheidend für die Existenz des Limes und für die nichtverschwindenden Beiträge der Entwicklung ist das Verhalten der Momente der kleinen Schritte für kleine  $dt$ .

Der Limes existiert falls  $\langle (\Delta x)^k \rangle = A_k(z, t)dt + o(dt)$  und nur Terme mit  $A_k \neq 0$  tragen bei

Einen wichtigen und uns allen lange bekannten Spezialfall finden wir sofort wieder: genügt  $x$  einer gewöhnlichen Differentialgleichung (keine Zufallskräfte) der Form

$$\frac{dx}{dt} = v(x)$$

so ist  $\Delta x = v(x(t))dt + O(dt^2)$ , was wiederum bedeutet, dass nur der erste Term der Entwicklung beiträgt und die Bewegungsgleichung für die ÜWD die Form

$$\partial_t \rho(x, t) = -\frac{\partial}{\partial x} (v(x)\rho(x, t))$$

annimmt. Hier haben wir bereits die Beliebigkeit von  $g$  benutzt und die Integranden in (\*) und in (\*\*) identifiziert. Es mag Sie vielleicht überraschen, aber **Differentialgleichungen von 1. Ordnung in der Zeit sind nach allen unseren Kriterien ebenfalls Markov Prozesse!** Natürlich ist die Dynamik völlig deterministisch, aber der Anfangszustand kann ja eine WD sein, und dann entstehen durch die Dynamik ÜWD, die genau der obigen Gleichung genügen. Tatsächlich haben wir in dem Abschnitt über die Zeitentwicklung "gemischter Zustände" genau so argumentiert. Angewandt auf die kanonischen Bewegungsgleichungen ist die obige Gleichung also gerade die **Liouville Gleichung in der klassischen Mechanik**. Wie Sie sehen, hat die Gleichung die Struktur einer *Kontinuitätsgleichung* und das ist kein Zufall. Sie schreibt ja die Zeitentwicklung der Dichte einer erhaltenen Größe, nämlich der Wahrscheinlichkeit des sicheren Ereignisses

$$\int dx \rho(x, t) = 1.$$

Im allgemeinen Fall der KME könnten nun aber vielleicht Terme bis zu irgendeiner  $k$ -ten Ordnung beitragen. Es gibt einen sehr hübschen Satz, der folgendes besagt:

$A_k = o(dt)$  entweder für  $k > 1$  oder für  $k > 2$  oder nie!

Den Beweis wollen wir hier auslassen, denn unser Hauptanliegen ist es, die Bewegungsgleichung für ÜWD zu finden, die Langevin-Prozesse beschreiben. Es ist aber für den Seelenfrieden nützlich, wenn man weiß, dass wir am Ende dieses Abschnitts *wirklich alle möglichen Typen der Bewegungsgleichungen für ÜWD von Markovprozessen* erschlagen haben.

Beschließen wir diesen Unterabschnitt, indem wir die Bewegungsgleichung für die ÜWD, wie sie aus der KME entsteht noch einmal vollständig aufschrei-

ben

$$\partial_t \rho(x, t) = \sum_{k=1}^{\infty} \frac{1}{k!} \left( -\frac{\partial}{\partial x} \right)^k [A_k(x) \rho(x, t)]$$

## 5.2 Inkremente von Langevin Prozessen

Nun wollen wir untersuchen, wie sich die  $\langle (\Delta x)^k \rangle$  bzw die  $A_k$  für Langevin Prozesse verhalten, die einer Gleichung

$$\dot{x} = f(x) + \xi(t)$$

genügen, wobei  $\xi(t)$  weißes Rauschen mit Varianz  $\sigma^2$  ist, dh  $\langle \xi \rangle = 0$  und  $\langle \xi(t)\xi(t') \rangle = \sigma^2 \delta(t - t')$ . Den zufälligen, kleinen Schritt erhalten wir durch Integration

$$\Delta x = x(t + dt) - x(t) = \int_t^{t+dt} d\tau f(x(\tau)) + \int_t^{t+dt} d\tau \xi(\tau)$$

Wenn wir die Notation vollständig dem vorigen Unterabschnitt angleichen wollen, müssen wir  $x(t) = z$  setzen, worauf wir aber hier verzichten.

Entscheidend ist nun, dass die beiden Terme auf der rechten Seite ein ganz unterschiedliches Verhalten für kleine  $dt$  zeigen. Von einer gewöhnlichen Differentialgleichung würden wir erwarten (s.o.), dass die Integrale dem Mittelwertsatz der Integralrechnung folgen und folglich proportional zu  $dt$  sind. Dann ist nur  $A_1 \neq 0$ . Schauen wir uns aber den 2. Term auf der rechten Seite etwas genauer an. Er spielt eine so große Rolle, dass wir ihm einen eigenen Bezeichner

$$dW_t^{t+dt} = \int_t^{t+dt} d\tau \xi(\tau)$$

geben und dem zugehörigen Prozess einen eigenen Namen.

$$W(t; t_0) = \int_{t_0}^t d\tau \xi(\tau)$$

heißt **Wiener Prozess**. In physikalischer Sprechweise würden wir sagen: überdämpfte, kräftefreie Brownsche Bewegung, denn die zugehörige Langevin-Gleichung ist

$$\dot{x} = \xi.$$

Wenn keine Verwechslungsgefahren bestehen, werden wir die Anfangs- und Endpunkte bei den Inkrementen  $dW_t^{t+dt}$  auch einfach weglassen,  $dW$  bezeichnet stets das Inkrement zwischen  $t$  und  $t + dt$ .

Warum sollte denn nun mit dem Integral über  $\xi$  etwas nicht in Ordnung sein, dass so gravierend ist, dass nicht mal mehr der gute, alte Mittelwertsatz der Integralrechnung anwendbar ist? Nun, — werfen Sie einen schauernden Blick

auf die Realisierungen des weißen Rauschens. Zu jedem Zeitpunkt  $t$  wird (ganz unabhängig was noch kürzeste Zeiten vorher passierte) ein neues  $\xi$  gemäß einer Gaussverteilung gewürfelt. Diese Trajektorien sind nicht nur unstetig, — nein viel schlimmer, sie springen überall !! Werfen wir also einen Blick auf die Inkremente des Wiener Prozesses, d.h. auf die durch Integration geglätteten Trajektorien des weißen Rauschens. Folgende 3 Eigenschaften sind elementar:

a)  $\langle dW \rangle = 0$

b)  $\langle (dW)^2 \rangle = \int_t^{t+dt} d\tau_1 \int_t^{t+dt} d\tau_2 \langle \xi(\tau_1)\xi(\tau_2) \rangle = \sigma^2 dt$

c)  $\langle dW_{t_1}^{t_1+dt} dW_{t_2}^{t_2+dt} \rangle = 0$ , falls die Intervalle  $[t_1, t_1 + dt]$  und  $[t_2, t_2 + dt]$  nicht überlappen

Ebenfalls erben die Wiener Inkremente von den  $\xi$  die Gaussische Statistik, es sind also kleine, unabhängige Schritte mit einer Gaussischen Schrittweiten-WD<sup>16</sup>

$$\rho(dW) = \frac{1}{\sqrt{2\pi\sigma^2 dt}} \exp\left(-\frac{dW^2}{2\sigma^2 dt}\right)$$

Da ist zunächst gar nichts Erschreckendes, aber die Eigenschaft b) bringt einen ins Grübeln. Sollte daraus folgen, dass  $dW = O(\sqrt{dt})$ ? Tatsächlich, man kann die  $dW$  folgendermaßen schreiben

$$dW = \sigma\eta(t)\sqrt{dt}$$

wobei nun  $\eta(t)$  eine Gaussische Zufallsvariable mit Mittelwert Null und Varianz 1 ist ( $\langle \eta(t) \rangle = 0$ ,  $\langle \eta(t)^2 \rangle = 1$ ). Diese einfache Transformation macht den Knackpunkt offensichtlich. Das Wiener Inkrement verhält sich für kleine  $dt$  wie  $\sqrt{dt}$  und wird daher viel größer als ein regulärer  $O(dt)$  Term in einer gewöhnlichen Differentialgleichung.

Zu den merkwürdigen Folgen dieser “Wiener Anomalie” gehört auch, dass man eine Langevin Gleichung eigentlich gar nicht als Differentialgleichung auffassen kann, denn wenn

$$\Delta x \propto \sqrt{dt}$$

so ist

$$\dot{x} = \lim_{dt \rightarrow 0} \frac{\Delta x}{dt}$$

leider nicht existent! Anders ausgedrückt: Die Realisierungen eines Langevinprozesses sind zwar stetig (immerhin!) aber nirgends differenzierbar.

<sup>16</sup>Nebenbei bemerkt: So lassen sich Trajektorien des Wiener Prozesses auch sehr gut auf dem Computer erzeugen

Langevingleichungen sollte man also vielmehr als Gleichungen für kleine Schritte (Inkrementen)

$$\Delta x = f(x(t))dt + dW + o(dt)$$

auffassen und als solche sind sie auch wohldefiniert. Beachten Sie, dass wir beim ersten Term anstandslos den Mittelwertsatz der Integralrechnung anwenden können, denn  $x(t)$  ist ja sogar stetig.

So, — nun können wir ohne Mühe die  $\langle (\Delta x)^k \rangle$  ausrechnen:

$$\begin{aligned} \langle \Delta x \rangle &= f(x(t))dt + o(dt) \\ \langle (\Delta x)^2 \rangle &= \langle dW^2 \rangle + o(dt) = \sigma^2 dt + o(dt) \\ \langle (\Delta x)^k \rangle &= o(dt) \text{ für } k > 2 \end{aligned}$$

Diese Ergebnisse setzen wir in die KME des vorigen Unterabschnitts ein und erhalten die gesuchte Bewegungsgleichung für die ÜWD der Langevinprozesse

$$\frac{\partial \rho(x, t)}{\partial t} = -\frac{\partial}{\partial x} [f(x)\rho(x, t)] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} \rho(x, t),$$

genannt **Fokker-Planck Gleichung**. Für  $f = 0$  wird sie zur wohlbekannteren Diffusionsgleichung, deren Lösung

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

die Anfangsbedingung

$$\rho(x, t = 0) = \delta(x - x_0)$$

erfüllt, mit anderen Worten, sie ist die ÜWD  $\rho(x, t|x_0, 0)$  des Wiener Prozesses. Die Gaussverteilung für kleine Schritte dieses Prozesses (s.o.) lesen wir übrigens auch an der Lösung für endliche Zeiten  $t$  ab!

Für das oben betrachtete Brownsche Teilchen mit Bewegungsgleichung

$$\dot{v} = -\zeta v + \frac{f(t)}{M}$$

mit  $\langle f \rangle = 0$  und  $\langle f(t)f(t') \rangle = 2\gamma kT$ , wobei  $\zeta = \gamma/M$  ergibt sich

$$\begin{aligned} \langle \Delta v \rangle &= -\zeta v dt + o(dt) \\ \langle (\Delta v)^2 \rangle &= \frac{2\gamma kT}{M^2} dt + o(dt) \end{aligned}$$

und somit die Fokker-Planck Gl

$$\partial_t \rho(v, t) = \zeta \partial_v [v\rho(v, t)] + \frac{\zeta kT}{M} \partial_v^2 \rho(v, t)$$

Für die überdämpfte Brownsche Bewegung ist

$$\dot{x} = \frac{1}{\gamma}(F_{ext}(x) + f(t))$$

so dass

$$\begin{aligned} \langle \Delta x \rangle &= \frac{F_{ext}(x)}{\gamma} dt + o(dt) \\ \langle (\Delta x)^2 \rangle &= \frac{2kT}{\gamma} dt + o(dt) := 2Ddt + o(dt). \end{aligned}$$

Die zugehörige Fokker-Planck Gleichung lautet

$$\partial_t \rho(x, t) = -\frac{1}{\gamma} \frac{\partial}{\partial x} [F_{ext}(x) \rho(x, t)] + D \frac{\partial^2}{\partial x^2} \rho(x, t)$$

### 5.3 Transformation von Langevin Prozessen. Die Ito Formel

Es kommt häufig vor, dass man aus einem Langevinprozess per Transformation einen neuen Prozess generiert und sich dann für die Eigenschaften dieses Prozesses interessiert. Ist also z.B. die Variable  $x$  die Geschwindigkeit eines Brownschen Teilchens, so ist  $E = x^2$  proportional zur kinetischen Energie. Die ÜWD für solche transformierten Prozesse kann man sich per Variablentransformation verschaffen, d.h. für (glatte, ein-eindeutige)  $z(x)$  ist

$$\rho(z, t) = \left| \frac{dx}{dz} \right| \rho(x(z), t).$$

Die Fokker-Planck Gleichung für  $\rho(z, t)$  erhält man einfach durch Anwenden der Kettenregel

$$\frac{\partial}{\partial x} = \frac{\partial z}{\partial x} \frac{\partial}{\partial z}$$

Nun erfüllt ja  $x$  eine Langevingleichung und man könnte auch auf die Idee kommen, diese Gleichung, bzw. die Inkremente zu transformieren. Dabei erlebt man allerdings eine Überraschung, weil wieder mal eine "Wiener Anomalie" auftritt. Betrachten wir des  $z$ -Inkrement

$$\Delta z = z(x(t) + \Delta x) - z(x(t))$$

wobei natürlich  $\Delta x = f(x(t))dt + dW$  sein soll. Taylorentwicklung zeigt uns

$$\Delta z = z'(x(t))\Delta x + \frac{z''(x(t))}{2}(\Delta x)^2 + \dots$$

Nun enthält  $\Delta x$  ja mit  $dW$  einen Term der Ordnung  $\sqrt{dt}$ , folglich muss man (entgegen üblichen Weisheiten aus Analysis I) hier auch Terme proportional

zu  $z''$  mitnehmen. Man kann  $\Delta x^2$  natürlich durch den einzigen  $O(dt)$  Term, nämlich  $dW^2$  ersetzen, so dass

$$\Delta z = z'(x(t))\Delta x + \frac{z''(x(t))}{2}dW^2 + o(dt).$$

Da der Prozess aber durch die ersten beiden Momente seiner Inkremente vollständig bestimmt ist, kann man auch statt  $dW^2$   $\sigma^2 dt$  schreiben, denn das reproduziert  $\langle \Delta z \rangle$  und  $\langle (\Delta z)^2 \rangle$ .

$$\Delta z = z'(x(t))\Delta x + \frac{z''(x(t))}{2}\sigma^2 dt + o(dt).$$

Diese Formel für den transformierten kleinen Schritt heisst auch **Ito's Lemma** oder **Ito Formel**. Die Terme der Fokker-Planck Gleichung lassen sich nun aus den transformierten Inkrementen ablesen

$$A_1(z) = z'(x(z))f(x(z)) + \frac{\sigma^2}{2}z''(x(z))$$

$$A_2 = z'(x(z))^2\sigma^2$$

( $z' = dz/dx$ ).

## 6 Die Tücken des Modellierens mit Langevingleichungen

Es folgen noch einige Notizen für Interessierte, die sich mit dem Problem des Modellierens mit Langevinprozessen beschäftigen. Sie sind in Englisch und tragen den Titel:

### Stochastic Calculus: Extremely Poor Man's Version

#### Scenario 1

$$dS(t) = rS(t)dt$$

You get an asset with a guaranteed interest rate  $r$ . No fluctuations. You buy  $S_0$  at time  $t = 0$ . At a later time you get

$$S(t) = S_0 \exp(rt)$$

No fluctuations, no problems.

### Scenario 2

You are offered an ensemble of assets  $S_i(t)$ ,  $i = 1 \dots N$ , with fluctuating interest rates. You observe the market for some time and try to estimate the expected, ensemble averaged interest rate by plotting  $\ln(S(t))$  vs.  $t$ , averaging over your ensemble, i.e.

$$\langle \ln(S(t)) \rangle \approx \frac{1}{N} \sum_{i=1}^N \ln(S_i(t))$$

and finally fitting this to a straight line. For the best fit slope, you find a number  $r$ . Furthermore you study the fluctuations of  $\ln(S(t))$  and find that for  $y = \ln(S(t)) - \langle \ln(S(t)) \rangle$  the increments for small time intervals

$$y(t+dt) - y(t) = dW(t)$$

are nicely Gaussian and uncorrelated for disjoint small time intervals. Thus you set up the following Langevin equ. for  $l(t) = \ln(S(t))$

$$dl(t) = rdt + dW(t)$$

to model the situation. The solution of this equation is

$$l(t) = l(0) + rt + \int_0^t dW(t)$$

Now, if you want to calculate your expected pay-off at time  $t$ , you have to study

$$\langle S(t) \rangle = \langle \exp(l(t)) \rangle = S_0 \exp(rt) \langle \exp\left(\int_0^t dW(t)\right) \rangle .$$

This can be done directly for a finite time interval, if you decompose the interval into an (increasing) number of small subintervals (length  $dt$ ) and use the fact that the  $dW(t)$  are Gaussian (and uncorrelated for different subintervals). Thus

$$I = \langle \exp\left(\int_0^t dW(t)\right) \rangle = \lim \langle \exp\left(\sum_i dW(t_i)\right) \rangle = \lim \prod_i \langle \exp(dW(t_i)) \rangle$$

The averages over

$$\rho(dW) = \frac{1}{\sqrt{2\pi\sigma^2 dt}} \exp\left(-\frac{dW^2}{2\sigma^2 dt}\right)$$

give the result

$$\langle \exp(dW(t_i)) \rangle = \exp(\sigma^2 dt/2)$$

and thus

$$\langle S(t) \rangle = S_0 \exp[(r + \sigma^2/2)t]$$

Oops! Is it mysterious that the interest rate derived from the expected pay-off differs from the estimate  $r$ ? Not at all, as we all know

$$\langle \exp(l) \rangle \neq \exp(\langle l \rangle)$$

Alternatively, you may calculate small increments of  $S(t)$  and derive a Langevin equ. for the value. Using Ito's Lemma for the map  $S = \exp(l)$  this leads to

$$dS(t) = \left(r + \frac{\sigma^2}{2}\right)S(t)dt + S(t)dW(t)$$

What is the solution of this stochastic differential equation? It is (as you again may verify by Ito's Lemma)

$$S(t) = S_0 \exp\left(rt + \int_0^t dW(t)\right)$$

i.e.  $S(t) = \exp(l(t))$  as it should be!

### Scenario 3

Now you use a different (for some people "more natural") procedure to estimate your model parameter "risk free interest rate"  $r$ . You plot  $\ln \langle S(t) \rangle$  vs.  $t$  and fit to a straight line. Let us call this interest rate  $\tilde{r}$ . Of course, from scenario 2 we know by now that

$$\tilde{r} = r + \frac{\sigma^2}{2}$$

Note that the Langevin equation for  $S(t)$  now looks as follows

$$dS(t) = \tilde{r}S(t)dt + S(t)dW(t),$$

which nicely reflects the role of  $\tilde{r}$ . On the other hand it is the solution which now looks a bit strange:

$$S(t) = S_0 \exp\left[(\tilde{r} - \sigma^2/2)t + \int_0^t dW\right]$$

Note that scenarios 2 and 3 describe the same situation! The differences arise from your procedure of introducing model parameters. For physicists this is a very simple example of what they call *renormalization*.

### Scenario 4

Now, instead of an asset, consider a different situation, which nevertheless leads to very similar model equations. Consider a growing bacterial colony. Let  $S$  be the density of the bacteria. It grows exponentially, and again

$$dS(t) = rS(t)dt$$

if no fluctuations are present. Due to environmental fluctuations, the growth rate fluctuates, thus we write the model equation as

$$dS(t) = (r + \zeta(t))S(t)dt$$

Here it is natural to assume, that  $\zeta$  is Gaussian, but correlated

$$\langle \zeta(t + \tau)\zeta(t) \rangle = \phi(\tau)$$

with some (sharply peaked) correlation function  $\phi$ . Note that for a stationary noise  $\phi(\tau) = \phi(-\tau)$ . Let  $\phi(\tau)$  have a finite maximum at  $\tau = 0$  and be smooth in the vicinity of  $\tau = 0$ . Small increments of  $\zeta$

$$d\zeta(t) = \zeta(t + dt) - \zeta(t)$$

then are “harmless”. They are Gaussian and

$$\langle (\zeta(t + dt) - \zeta(t))^2 \rangle = 2\phi(0) - 2\phi(dt) = -\frac{1}{2}\phi''(0)dt^2$$

Thus, there are no “Wiener” anomalies. ( $d\zeta = O(dt)$ ). We may safely integrate the differential equation to obtain

$$S(t) = S_0 \exp\left(rt + \int_0^t \zeta(\tau)d\tau\right)$$

What happens, if we let  $\phi$  tend to a  $\delta$ -function? Where do the Wiener anomalies show up?

To demonstrate what happens, write the differential equation in integral form

$$dS_t := S(t + dt) - S(t) = \int_t^{t+dt} d\tau_1 S(\tau_1)(r + \zeta(\tau_1))$$

and (little trick), reinsert this integral form for  $S(\tau_1)$  by writing

$$dS(\tau_1) := S(\tau_1) - S(t) = \int_t^{\tau_1} d\tau_2 S(\tau_2)(r + \zeta(\tau_2))$$

so that

$$dS_t = \int_t^{t+dt} d\tau_1 (r + \zeta(\tau_1)) (S(t) + dS(\tau_1))$$

which gives the following 4 terms for  $dS_t$ :

$$dS_t = rS(t)dt + S(t)dW_{t+dt} + r \int_t^{t+dt} d\tau_1 dS(\tau_1) + \int_t^{t+dt} d\tau_1 \zeta(\tau_1) dS(\tau_1)$$

For simplicity, we have already named  $dW_{t+dt} = \int_t^{t+dt} \zeta$ , which will become the Wiener increment if  $\zeta$  tends to white noise. We have indicated the upper

bound of the small time interval, implying that the lower bound is always  $t$ . Now we look for terms which are up to  $O(dt)$  in the "white noise limit"  $\phi() \rightarrow \sigma^2 \delta()$ , where  $\sigma^2 = \int_{-\infty}^{\infty} d\tau \phi(\tau)$  is kept fixed. Note that the leading terms in  $dS(\tau_1)$  are obtained if we put  $S(\tau_2) = S(t) + dS(\tau_2) \approx S(t)$ . Sceptics may check explicitly, which terms arise from  $dS(\tau_2)$ . This gives us to leading order

$$dS(\tau_1) = rS(t)(\tau_1 - t) + S(t)dW_{\tau_1}$$

Thus we see that the 3rd term in  $dS_t$  does not contain contributions of  $O(dt)$  (only  $O(dt^2)$  and  $O(dt^{3/2})$ ).

Now consider the 4th term. It contains something, which looks as if it may contribute to  $O(dt)$ , and which we therefore keep, namely

$$S(t) \int_t^{t+dt} d\tau_1 \zeta(\tau_1) \int_t^{\tau_1} d\tau_2 \zeta(\tau_2).$$

So our increment has taken on the form

$$dS_t = rS(t)dt + S(t)dW_{t+dt} + S(t) \int_t^{t+dt} d\tau_1 \zeta(\tau_1) \int_t^{\tau_1} d\tau_2 \zeta(\tau_2) + o(dt).$$

The average of this increment takes on the form:

$$\langle dS(t) \rangle = rS(t)dt + S(t) \int_t^{t+dt} d\tau_1 \int_t^{\tau_1} d\tau_2 \phi(|\tau_1 - \tau_2|)$$

Note that the integral would be ill defined if we just replace  $\phi$  by a  $\delta$  function (argument of  $\delta$  function at upper limit of integration). However, the integral exists for all  $\phi$  in the sequence towards the  $\delta$  function and tends to

$$\int_t^{\tau_1} d\tau_2 \phi(|\tau_1 - \tau_2|) \rightarrow \frac{\sigma^2}{2}$$

because in the limit we exactly integrate over half of the area below  $\phi$ . Thus we obtain in the white noise limit

$$\langle dS_t \rangle = \left(r + \frac{\sigma^2}{2}\right) S(t)dt$$

The second moment  $\langle dS^2 \rangle$  is less spectacular. Here we simply get

$$\langle dS^2 \rangle = S^2(t)\sigma^2 dt$$

without any reinsertion tricks. The white noise limit causes no problems here. But now we have got something really remarkable! We have started from a well-defined differential equation

$$\dot{S} = rS(t) + \zeta(t)S(t)$$

with correlated noise. It can be integrated according to usual rules of calculus. Then we have performed the “white noise limit” and found the above average increment. Note that this result differs from what we would have obtained from a corresponding interest rate model

$$\langle dS_t \rangle = rS(t)dt + S(t) \langle dW(t) \rangle = rS(t)dt$$

The difference between the model buildings (interest rate vs bacterial growth rate), which leads to these different results now begins to transpire: In the interest rate models, the  $S(t)$  in the term  $S(t)dW_t$  is statistically independent from  $dW_t$  (by financial causality": it is always the fixed amount of money at the beginning of a period which is incremented. The interest rate and the money at the beginning of the period are strictly independent!) In the bacterial growth model, the noise started out with a finite correlation time, which also implies that  $S(t)$  is not statistically independent from  $\zeta(t)$  because it was generated by some  $\zeta(t - \tau)$ , which are not independent from  $\zeta(t)$ ! This correlation effect survives the white noise limit.

Mathematically, the interest rate model building is called *Ito stochastic calculus* and the bacterial growth model building is called *Stratonovich stochastic calculus*. We are not going to develop them any further. FAPP (for all practical purposes) you are done if you can calculate the  $\langle dS \rangle$  and  $\langle dS^2 \rangle$ , because then you can set up the Fokker-Planck equation, which does not need any stochastic calculus magics.

Note, however, that once you have unambiguously set up a Langevin equation—be it Ito or Stratonovich—you may easily set up the equivalent equation in terms of the other calculus and may use whatever stochastic calculus you want for your computations.

### Summary:

Here is a short summary: Ito will be denoted by (*I*), Stratonovich by (*S*)

#### 1.) Same process — different Langevin equations

$$(I, S) \quad dl = rdt + dW_t$$

Equivalent Langevin equs. for  $S = e^l$

$$(I) \quad dS = \left( r + \frac{\sigma^2}{2} S(t) \right) dt + S(t) dW_t$$

$$(S) \quad dS = rS(t)dt + S(t)dW_t$$

Solution of (*I*)

$$(I, S) \quad S(t) = S_0 \exp \left( rt + \int_0^t dW_t \right)$$

Expectation

$$(I, S) \quad \langle S(t) \rangle = S_0 e^{(r+\sigma^2/2)t}$$

Fokker-Planck equ

$$\partial_t \rho(S) = -\partial_S [(r + \sigma^2/2)\rho(S)] + \frac{\sigma^2}{2} \partial_S^2 (S^2 \rho(S))$$

## 2.) Same Langevin equation — Different processes

$$(I, S) \quad dS(t) = rS(t)dt + S(t)dW_t$$

Solutions

$$(I) \quad S(t) = S_0 \exp \left( (r - \sigma^2/2)t + \int_0^t dW_t \right)$$

$$(S) \quad S(t) = S_0 \exp \left( rt + \int_0^t dW_t \right)$$

Expectations

$$(I) \quad \langle S(t) \rangle = S_0 e^{rt}$$

$$(S) \quad \langle S(t) \rangle = S_0 e^{(r+\sigma^2/2)t}$$

Fokker-Planck equation:

$$(I) \quad \partial_t \rho(S) = -\partial_S [r\rho(S)] + \frac{\sigma^2}{2} \partial_S^2 (S^2 \rho(S))$$

$$(S) \quad \partial_t \rho(S) = -\partial_S [(r + \sigma^2/2)\rho(S)] + \frac{\sigma^2}{2} \partial_S^2 (S^2 \rho(S))$$

Now you know more about stochastic calculus than many people do. Start using it. Good luck! R.K.

## 7 Monte Carlo Simulation von Mastergleichungen

Das Lösen von Mastergleichungen ist i.A. nicht einfach bis unmöglich, obwohl Mastergleichungen *per definitionem* lineare Gleichungen sind (allerdings häufig sehr viele bis unendlich viele).

Ganz allgemein sind **Monte Carlo Methoden numerische Methoden, die Näherungen durch statistische Abschätzungen gewinnen.**

Mastergleichungen sind ein Paradebeispiel für den Erfolg dieser Methoden.

Die Idee ist dabei folgende: Vorgegeben sei eine Mastergleichung, z.B.

$$\frac{dP(n, t)}{dt} = \sum_m r(n|m)P(m, t) - r(m|n)P(n, t)$$

Statt  $P(n, t)$  zu berechnen, erzeuge man sich eine Stichprobe aus den  $P(n, t)$ , d.h.  $M$  Realisierungen  $\{n_a(t)\}_{a=1, \dots, M}$ , die zu jeder Zeit  $t$  gemäß  $P(n, t)$  verteilt sind. Mit dieser Stichprobe schätze man die Größen, die man eigentlich berechnen will, also z.B. die W.  $P(n, t)$  über die relative Häufigkeit oder den Mittelwert  $\langle n(t) \rangle$  über das arithmetische Stichprobenmittel  $(1/M) \sum_a n_a(t)$ .

Auf den ersten Blick scheint etwas an diesem Ansatz komisch: Wie soll man Stichproben erzeugen, wenn man  $P(n, t)$  gar nicht kennt? Das genau ist der Schlüssel jeder guten Monte Carlo Methode. Wir präsentieren hier 2 Varianten eines Algorithmus, der diese Aufgabe löst.

## 7.1 Gillespie's Algorithmus

Wir betrachten eine Realisierung. Zur Zeit  $t$  liege der Zustand  $n$  vor. Wie bestimmen wir die weitere Zeitentwicklung der Realisierung? Wir zerlegen sie in 2 Schritte, die wir der Reihe nach behandeln:

- 1.) Wann erfolgt der nächste Sprung in ein  $m \neq n$ ?
- 2.) Zu welchem  $m$  springt dann die Realisierung?

Zur Bestimmung der nächsten Sprungzeit gehen wir aus von der Gesamtrate, aus  $n$  herauszuspringen, d.i.

$$\Gamma(n) = \sum_{m(\neq n)} r(m|n)$$

Die W., in einem, beliebig kleinen Zeitschritt bei  $n$  zu bleiben ist daher

$$1 - \Gamma(n)dt$$

Also ist die W. in  $k$  aufeinanderfolgenden Zeitschritten bei  $n$  zu bleiben gerade

$$[1 - \Gamma(n)dt]^k$$

Schließlich ist die W., im  $(k + 1)$ -ten Schritt zu springen, nachdem man  $k$  Schritte bei  $n$  verweilt

$$[1 - \Gamma(n)dt]^k \Gamma(n)dt$$

Denken wir uns nun diese Zeitschritte als eine Unterteilung des endlichen Zeitintervalls der Länge  $\tau$ , so dass

$$dt = \frac{\tau}{k + 1}$$

und verfeinern wir diese Zerlegung immer weiter, so erhalten wir eine WD für die ZV  $T$ , "Wartezeit bis zum nächsten Sprung"

$$\rho_T(\tau)d\tau = \lim_{k \rightarrow \infty} [1 - \Gamma(n) \frac{\tau}{k + 1}]^k \Gamma(n) d\tau$$

$$= e^{-\Gamma(n)\tau} \Gamma(n) d\tau$$

Die Wartezeiten zwischen Sprüngen sind also *exponentiell verteilte* Zufallszahlen (s.o.). Daher sieht der 1. Schritt von Gillespie's Algorithmus so aus

**Schritt 1:**

Ziehe ein  $\xi$  gleichverteilt au  $[0, 1]$ . Berechne  $\tau = -(\ln \xi)/\Gamma(n)$ . Setze  $t \rightarrow t + \tau$

Nun wenden wir uns der Frage zu, wohin die Realisierung denn nun zur Zeit  $t + \tau$  springt. Hier brauchen wir offenbar eine bedingte W, nämlich

$$P(n \rightarrow m) := P(m, t + \tau + dt | (n, t + \tau) \text{ UND: Ein Sprung findet z.Zt. } t + \tau \text{ statt}^{17})$$

Hier wird neben der Bedingung für den Ausgangspunkt  $n$  des Sprungs eine weitere Bedingung gestellt, nämlich dass auch tatsächlich mit Sicherheit gesprungen wird. Wir kennen zunächst nur  $P(m, t + \tau + dt | n, t + \tau) = r(m|n)dt$ , aber wir kennen natürlich auch die W. für das Auftreten der Bedingung

$$P(\text{Ein Sprung findet z. Zt. } t + \tau \text{ statt}) = \Gamma(n)dt$$

so dass nach den Regeln der W. theorie (s.bedingte W.)

$$P(n \rightarrow m) = \frac{r(m|n)}{\Gamma(n)} := z_m$$

Die  $z_m$  liegen alle im Intervall  $[0,1]$  und

$$\sum_{m(\neq n)} z_m = 1$$

Fassen wir die  $z_m$  als Längen von nicht überlappenden Intervallen  $I_m \subset [0, 1]$  auf, so füllen diese Intervalle genau  $[0, 1]$  (die Reihenfolge der Intervalle spielt hier keine Rolle). Ziehen wir nun ein  $\eta$  gleichverteilt aus  $[0, 1]$ , so ist W. es in  $I_m$  zu finden gerade gleich  $z_m$ . Daraus ergibt sich nun unmittelbar der 2. Schritt von Gillespie's Algorithmus:

**Schritt 2:**

Ziehe ein  $\eta$  gleichverteilt aus  $[0, 1]$ . Für  $\eta \in I_m$ , springe nach  $m$

## 7.2 Algorithmus mit festem Zeitschritt

Wir legen einen Zeitschritt  $dt$  fest und fangen wieder auf der Realisierung zur Zeit  $t$  bei  $n$  an. Die Sprungwahrscheinlichkeit im nächsten Schritt ist  $\Gamma(n)dt$ . Man beachte, dass  $dt$  so klein gewählt werden muss, dass stets  $\Gamma(n)dt < 1$ . Wiederum zerlegen wir die W. für einen Sprung mit Ziel  $m$  wie oben, nämlich

$$r(m|n)dt = \left( \frac{r(m|n)}{\Gamma(n)} \right) \Gamma(n)dt$$

und ziehen im 1. Schritt ein  $\xi \in [0, 1]$  gleichverteilt. Falls

$$\xi < \Gamma(n)dt$$

gehen wir zu Schritt 2 (Sprungzielwahl), andernfalls wiederholen wir Schritt 1. Schritt 2 führen wie aus wie bei Gillespie's Algorithmus. Der *Vorteil dieses Algorithmus*: Man vermeidet die (numerisch kostspielige) Berechnung eines Logarithmus in jedem Schritt. Der *Nachteil*: Falls einige oder alle Raten sehr klein, d.h.  $\Gamma(n)dt \ll 1$  für einige  $n$ , so wird Schritt 1 sehr häufig ausgeführt, ohne dass die Realisierung von der Stelle kommt.