Wahrscheinlichkeit
Mathematische Bissen zu Kursvorlesung Theoretische Physik

---

1 Zufall und Notwendigkeit

Auf dem Papier verspricht die klassische Physik, also die Newton'sche Mechanik und die Maxwellsche Elektrodynamik, dass alles notwendig ist, dass nichts dem Zufall überlassen bleibt. Im Labor wird dieses Ideal nie erreicht. Messungen weisen immer gewisse Schwankungen auf, und Messresultate können nie mit absoluter Sicherheit, sondern lediglich mit gewissen **Wahrscheinlichkeiten** (im folgenden abgekürzt W'keiten) vorausgesagt werden. Pragmatisch führt man Zufallselemente ein, die darin ihre Rechtfertigung finden, dass man schlicht nicht über die notwendigen Mittel verfügt, um alle Freiheitsgrade im Labor kontrollieren zu können. Solche Zufallselemente konstituieren einen Zufallsbegriff, der in der Wissenschaftstheorie **epistemischer Zufall** genannt wird. Zufall aus Unkenntnis, bei prinzipieller Möglichkeit, die Unkenntnis durch feinere Methoden zu eliminieren.

Aber selbst wenn man über alle Mittel und die feinsten Methoden verfügte würde die Quantenmechanik ein Strich durch die Rechnung machen. In der Quantenmechanik ist der Zufall *absolut*, dort genannt **ontologischer Zufall**, das heißt jegliches Bemühen den Zufall aus quantenmechanischen Experimenten zu eliminieren, ist fruchtlose Zeitverschwendung. Zeit also, den W'keitsbegriff kurz zu erläutern.

2 W'keitsraum bzw. Ensemble

In der W'keitstheorie wird jedem **Experiment** (mit einem Würfel, einem Atom oder einer Katze) eine Menge X zugeordnet, genannt **Ergebnismenge**, oder *Stichprobenraum* (engl. sample space). Im einfachsten Fall ist X eine endliche Menge, beim einfachen Wurf eines Würfel etwa $X = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$ – die Menge der möglichen Muster auf den oben liegenden Seiten nach dem Wurf.[1] Etwas schwieriger gelagert der Fall, in dem X ein Kontinuum, wie etwa beim Glücksrad, wo die Ergebnismenge ein reelles Winkelintervall $X = [0, 2\pi)$. Die möglichen Ergebnisse sind hier reelle Zahlen, und die lassen sich nicht abzählen ...

Wir betrachten vorläufig nur Experimente mit endlich abzählbarer Ergebnismenge. In diesem Fall kann das Experiment (ein Wurf) durch die Angabe eine Wahrscheinlichkeit p_x für jedes Ergebnis $x \in X$ charakterisiert werden. Die Funktion $p : X \to \mathbb{R}$ kodiert die **Präparation** des Systems bzw. den statistischen **Zustand**, in dem das System sich befindet bevor das Experiment durchgeführt ist.[2] Ein fairer Würfel, beispielsweise, ist charakterisiert durch $p_x = 1/6$ für alle $x \in X$, aber es gibt bekanntlich auch gezinkte Würfel.[3] Die

[1] Auch wenn Sie versucht sind, die Muster mit Zahlen zu identifizieren – tun Sie es nicht: Schliesslich kann man Muster nicht addieren.

[2] Man beachte, dass x u.U. lediglich ein Symbol, nicht notwenig eine Zahl, dass aber p_x eine Zahl.

[3] Experimente mit fairen Würfeln nennt man auch *Laplace-Experimente*.

einzige Einschränkung an die p_x sind (i) $p_x \geq 0$, und (ii) $\sum_{x \in X} p_x = 1$.

Eine *Messgröße*, auch genannt **Observable** wird für unsere Zwecke beschrieben durch eine reellwertige(!) Funktion $f : X \to \mathbb{R}$. "Augenzahl", beispielsweise, ist eine Observable, also $f(\text{⚀}) = 1, f(\text{⚁}) = 2$ usw. In der Sprache der W'keitstheorie heißen die f **Zufallsvariable**, zuweilen *stochastische Variable*. Der Wert $y = f(x)$ wird als derjenige Messwert interpretiert, den man an einem f-Messgerät abläse, ginge das Experiment mit Ergebnis x aus.

In einer typischen Versuchsreihe von N statistisch unabhängigen Einzelexperimenten darf im Limes $N \to \infty$ erwartet werden, dass der Messwert $f(x)$ mit Häufigkeit Np_x abgelesen wird (vgl. allerdings die Ausführungen zu relativen Häufigkeiten im nächsten Abschnitt). Die gewichtete Summe $\sum_x Np_x f(x)$ definiert – nach Division durch N – den statistischen **Mittelwert**

$$\langle f \rangle_p := \sum_{x \in X} p_x f(x)\,, \tag{1}$$

auch genannt der **Erwartungswert** der Observablen f im Zustand p. Verallgemeinert: Das n**-te Moment** der Observabe f ist definiert als der Erwartungswert den n-ten Potenz von f,

$$\langle f^n \rangle_p = \sum_{x \in X} p_x f(x)^n\,, \tag{2}$$

und die (quadratische) **Varianz**, auch genannt mittlere quadratische Schwankung von f

$$\begin{aligned} \sigma_p(f)^2 &:= \langle (f - \langle f \rangle_p)^2 \rangle_p & (3)\\ &= \langle f^2 \rangle_p - 2\langle f \rangle_p \langle f \rangle_p + \langle f \rangle_p^2 & (4)\\ &= \langle f^2 \rangle_p - \langle f \rangle_p^2 & (5) \end{aligned}$$

Die Wurzel, also die Varianz $\sigma_p(f)$, ist ein Maß für die Breite der statistischen(!) Verteilung der f-Werte. Die vollständige Verteilung der f-Werte erhält man nach der Vorschrift

$$p_f(y) = \langle \delta_y(f) \rangle_p = \sum_x \delta_y(f(x)) p_x\,. \tag{6}$$

wo $\delta_y(f(x)) = 1$ falls $f(x) = y$, und 0 sonst.[4]

Ganz wichtig sind Zufallsvariable, hier bezeichnet χ, die nur die Werte 1 oder 0 annehmen können. Solche *Indikatorvariable* sind eindeutig charakterisiert durch die Angabe eine Teilmenge $A \subset X$ auf der sie den Wert 1 annehmen,

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{sonst}\,. \end{cases} \tag{7}$$

Ihre Bezeichnung verdanken die Indikatorvariable dem Umstand, dass ein χ_A-Messgerät lediglich das Fallen (oder Nicht-Fallen) des Ergebnisses x in die Menge A anzeigt, beispielsweise durch das Aufleuchten einer grünen Birne im Falle $x \in A$, und einer roten Birne

[4] Da in diesem Abschnitt X eine diskrete Menge, ist auch der Wertebereich von f eine diskrete Menge, und δ ist eine Art Kronecker-Delta.

falls x nicht in A, ohne genaueren Aufschluss über x zu geben.[5]

Eine Teilmenge $A \subset X$ nennt man ein **Ereignis** (*nicht* Ergebnis bzw. *Elementar*ereignis). Die Teilmenge $A = \{$⚁,⚃,⚅$\}$, beispielsweise, bezeichnet das Ereignis "die Zahl der gewürfelten Augen ist gerade". Würfelt man eine ⚃, ist dieses Ereignis eingetreten, der Wert der Indikatorfunktion ist $\chi_A($⚃$) = 1$. Bei einer ⚂ ist es nicht eingetreten, der Wert der Indikatorfunktion ist $\chi_A($⚂$) = 0$. Die mit Abstand wichtigsten Ereignisse sind das **sichere Ereignis** X – interpretiert "irgendein *Ergebnis* (egal welches)" – und das **unmögliche Ereignis** $\emptyset$ – beim Würfeln z.B. die Augenzahl 7.

Der Erwartungswert einer Indikatorvariable

$$P(A) := \langle \chi_A \rangle = \sum_x p_x \chi_A(x) = \sum_{x \in A} p_x \,. \tag{8}$$

gibt die W'keit für das Auftreten des Ereignisses A. Durch scharfes Hinschauen bestätigt man nämlich: für jede Teilmenge $A \in X$ ist P ist nicht-negativ, $P(A) \geq 0$, (2) für das unmögliche Ereignis $\emptyset$ ist der Wert gleich Null, $P(\emptyset) = 0$, sowie (3) für disjunkte Teilmengen $A_i \subset X$ ist P additiv, $P(\cup_i A_i) = \sum_i P(A_i)$. Kurz: die Mengenfunktion P ist ein *Maß*. Und da obendrein P normiert, $P(X) = 1$, ist P ein **Wahrscheinlichkeitsmaß** in X.

Mit ein bisschen Mengenlehre ist nun schnell bewiesen

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \,. \tag{9}$$

Insbesondere $P(A \cup B) = P(A) + P(B)$ für zwei sich gegenseitig ausschließende Ereignisse $A \cap B = \emptyset$, was Ihnen möglicherweise als "entweder-oder-Regel" aus der Schulzeit in Erinnerung ist.

Sei nun B fest mit $P(B) > 0$, dann definiert

$$A \mapsto P(A|B) := \frac{P(A \cap B)}{P(B)} \tag{10}$$

ein W'keitsmaß in X (Beweis!), genannt **bedingte W'keit**, interpretiert "W'keit, dass A vorliegt gegeben B ist realisiert". Sei beispielsweise, $A = \{$⚁,⚃,⚅$\}$, entsprechend der Eigenschaft gerade, und $B = \{$⚁,⚂,⚄$\}$, entsprechend der Eigenschaft prim, dann ist für faire Würfel $P(B) = \frac{1}{2}$, $P(A \cap B) = P(\{$⚁$\}) = \frac{1}{6}$ und also $P(\mathsf{gerade}|\mathsf{prim}) = \frac{1}{3}$. Wenn man irgendwie informiert wird, dass die soeben gewürfelte Augenzahl eine Primzahl ist, dann ist mit W'keit $\frac{1}{3}$ die Augenzahl eine gerade Zahl (also 2).

Die W'keit $P(A \cap B)$ heißt auch **Verbundwahrscheinlichkeit**. Gilt $P(A|B) = P(A)$ (was bei unserem Beispiel nicht der Fall ist) nennt man die Ereignisse A und B **statistisch unabhängig**, auch *unkorreliert*. In diesem Fall offenbar $P(A \cap B) = P(A)P(B)$, was Ihnen möglicherweise als "sowohl-als-auch-Regel" aus der Schulzeit in Erinnerung ist.

Das Komplement einer Teilmenge $A \subseteq X$, für $A = \mathsf{gerade}$ gegeben $A^c = \{$⚀,⚂,⚄$\}$, steht für das Ereignis "Nicht-A" und betrachtet man zwei verschiedene Ereignisse, neben dem

[5]Indikatorvariable werden von Quantenmechanikern der Ludwig-Schule auch *Entscheidungseffekt* genannt. Zufallsvariable $0 \leq f(x) \leq 1$ für alle $x \in X$, mit $f(x)$ interpretiert "Ansprech-W'keit bei Ergebnis x" (und *nicht* Messwertanzeige $f(x)$!), heißen bei Ludwigerianern *Effekt*. Ein schönes Beispiel vermittelt $f(x) = q\chi_A(x)$, wo $0 \leq q \leq 1$ ein Maß für die Effizienz eines A-Detektors. Mehr dazu bei: Günther Ludwig *Einführung in die Grundlagen der Theoretischen Physik – Band 3: Quantentheorie*, Vieweg 1976 [ISBN 3 528 09183 1].

bereits erwähnten Ereignis A das Ereignis $B =$ prim, kann man sich für Vereinigungsmenge $A \cup B$ interessieren, in unserem Beispiel $A \cup B = \{$⚁,⚂,⚃,⚄,⚅$\}$, interpretiert "die Augenzahl ist entweder gerade oder prim", bzw. die Durchschittsmenge $A \cap B = \{$⚁$\}$, interpretiert "die Augenzahl ist sowohl gerade als auch prim". Teilmengen von X, so die Moral, stehen für Ereignisse (bzw. Eigenschaften) und die mengentheoretischen Verknüpfungen finden ihrer natürliche Interpretation in der Aussagenlogik der Ereignisse.

Für jeden Ergebnismenge X, ob nun abzählbar endlich (wie beim Würfel), abzählbar unendlich (wie beim Lotto mit rationalen Zahlen), oder gar überabzählbar (wie beim Dartspiel), ist die Menge der Ereignisse gegeben durch ein System von Teilmengen von X, die leere Menge und die Ergebnismenge mit eingeschlossen, das unter Komplementbildung und abzählbaren Vereinigungen stabil ist. Solcherart Mengensysteme laufen in der Mathematik unter dem Begriff **σ-Algebra**, genauer: Ein System $\mathcal{A}$ von Teilmengen einer Menge X heißt eine *σ-Algebra über X*, falls (1) $X \in \mathcal{A}$, (2) mit A in $\mathcal{A}$ ist auch das Komplement in $\mathcal{A}$, (3) für jede abzählbare Folge (A_i) von Mengen aus $\mathcal{A}$ liegt auch die Vereinigung $\cup_i A_i$ in $\mathcal{A}$. Das Tupel $(X, \mathcal{A})$, worin X irgendeine Menge und $\mathcal{A}$ eine σ-Algebra über X fungiert in der Mathematik unter dem Begriff **Messraum**, auch *messbarer Raum*. Ist auf $(X, \mathcal{A})$ obendrein ein W'keitsmaß verabredet, fungiert das Tripel $(X, \mathcal{A}, P)$ in der Mathematik unter dem Begriff **W'keitsraum**. In der (klassischen) Physik versteht man unter dem Begriff **Ensemble** eine gedachte Gesamtheit unendlich vieler, gleichartig präparierter physikalischer Systeme, wobei jedes einzelne System (und die mit ihm gegebenen Observablen) durch den Maßraum $(X, \mathcal{A})$ und die Präparation durch die Wahl eines W'keitsmaßes P spezifiziert sind.

Leicht bestätigt man, dass die Menge aller Teilmengen von X, genannt die Potenzmenge $\mathcal{P}(X)$, eine σ-Algebra darstellt. Aber auch das System $\{\emptyset, X\}$ bilden eine σ-Algebra über X – wenn auch keine sonderlich interessante. Kurz: die Ereignismenge $\mathcal{A}$ ist nicht automatisch schon mit der Ergebnismenge gegeben. Nur im Falle einer endlichen Ergebnismenge, $|X| < \infty$, wie er etwa beim Würfel vorliegt, ist die Potenzmenge $\mathcal{P}(X)$ ein natürlicher Kandidat für die σ-Algebra der Ereignisse. In allen anderen Fällen muss jeweils eine geeignete Untermenge von $\mathcal{P}(X)$ bestimmt werden, die zumindest alle interessierenden Ereignisse umfasst, dabei in jedem Fall aber den Anforderungen an eine σ-Algebra genügt. Welche Ereignisse interessieren, wird dabei durch den physikalischen Messaufbau bestimmt. Kann beispielsweise beim Würfeln nur gemessen werden, ob die Augenzahl gerade ist (Ereignis A) oder nicht (Ereignis A^c), definiert bereits $\mathcal{A} := \{\emptyset, X, A, A^c\}$ eine ausreichende σ-Algebra. Für ein Messgerät, das alle Ergebnisse eines Wurfes wohl unterscheiden kann, ist allerdings die volle Potenzmenge $\mathcal{P}(X)$ die einzig adäquate σ-Algebra.

Beim Glücksrad ist die Sache schon schwieriger. Der Ausgang eines Experiments (das Ergebnis) ist hier vollständig durch den Winkel $x \in [0, 2\pi)$ charakterisiert, bei dem das Glücksrad stehen bleibt. Die Ergebnismenge $X := [0, 2\pi)$ ist überabzählbar. Sieht man von pathologisch präparierten Glücksrädern ab, ist die W'keit, irgendeinen bestimmten Winkel (Ergebnis), beispielsweise $x = 1$, genau zu treffen, Null. Allerdings – und das ist der Grund für die Unterteilung mit Nägeln – ist die W'keit, dass das Glücksrad in irgendeinem endlichen Winkelintervall $[a, b) \subset [0, 2\pi)$ stehen bleibt, im Allgemeinen nicht Null. Nun erzeugen solche Intervalle durch Komplementbildung und abzählbare Vereinigung eine bestimmte σ-Algebra in $[0, 2\pi)$, genannt die σ-Algebra der **Borel'schen Mengen** des $[0, 2\pi)$, bezeichnet $\mathcal{B}([0, 2\pi))$, und es ist genau diese Algebra die beim Glücksrad (nun wieder ohne Teilungsnägel) die Ereignismenge darstellt: Glücksrad-Ergebnisse sind Winkel, Glücksrad-Ereignisse sind die Borelschen Mengen des Intervalls $[0, 2\pi)$. Ein Maß auf $\mathcal{B}([0, 2\pi))$ ist das **Lebesgue-Borel Maß** λ, das man erhält wenn das Elementarmaß $\lambda([a,b)) := b-a$ auf die Vereinigung abzählbar vieler disjunkter Intervalle übertragen wird (jede Borel'sche Menge kann als solch eine Vereinigung dargestellt werden). Im rellen Zahlenraum $\mathbb{R}$ hat das Lebesgue-Borel Maß übrigens schöne Eigenschaften, unter anderem die Translationsinvarianz, aber es ist bei weitem nicht das einzige Maß auf $\mathcal{B}(\mathbb{R})$ ... vgl. auch die Ergänzung.

Da die Glücksrad-Ereignisse aus Intervallen erzeugt werden, reicht es für die Angabe eines W'keitsmaßes aus, Wahrscheinlichkeiten für eine monoton wachsende Schar von Intervallen anzugeben, bezeichnet

$$m(x) := P\left([0,x)\right)\,, \qquad 0 \leq x \leq 2\pi \tag{11}$$

mit $m(0) = 0$ (entsprechend $P(\emptyset) = 0$), $m(2\pi) = 1$ (entsprechend $P([0,2\pi)) = 1$), und m monoton nicht abnehmend (entsprechend $P(A \cup B) = P(A) + P(B)$ für nicht überlappende Winkelintervalle A, B.) Eine Funktion m mit diesen Eigenschaften heißt auch **Verteilungsfunktion**. Offensichtlich für ein beliebiges Intervall $[a,b) \subset [0,2\pi)$

$$P\left([a,b)\right) = m(b) - m(a)\,. \tag{12}$$

Ein W'keitsraum, bei dem das W'keitsmaß über eine Verteilungsfunktion definiert ist, ist vom Typ **Lebesgue-Stieltjes Maßraum**. In so einem Maßraum können einzelne Punkte ein Maß ungleich Null haben – genau dort wo m einen Sprung macht – aber ganze Intervalle können das Maß Null haben – wo nämlich m konstant bleibt. Wählt man $m := (2\pi)^{-1}\mathrm{id}$, wo id die Identität $\mathrm{id}(x) = x$, schaut man offensichtlich auf eine Variante des bereits bekannten Lebesgue-Borel Maßes – die **Gleichverteilung** auf $[0,2\pi)$.

Führen wir an dieser Stelle die abkürzende Notation ein, $P(dx) := P\left([x, x+dx)\right)$ (lies: die (differentielle) W'keit in einem bei x lokalisierten Intervall der (differentiellen) Größe dx zu landen), werden die Erwartungswerte von stochastischen Variablen $f : [0,2\pi) \to \mathbb{R}$ notiert[6]

$$\langle f \rangle = \int_0^{2\pi} f(x) P(dx) \tag{13}$$

und ist die Verteilungsfunktion m gar stetig differenzierbar, das heißt existiert $\rho(x) := \frac{d}{dx}m(x)$ für alle $x \in [0,2\pi)$ (das entsprechende Maß heißt dann *absolut stetig*),

$$\langle f \rangle = \int_0^{2\pi} f(x)\rho(x) dx\,. \tag{14}$$

Die Funktion ρ nennt man **Wahrscheinlichkeitsdichte**, und $\rho(x)dx$ wird ebenso wie oben interpretiert "die (differentielle) W'keit in einem bei x lokalisierten Intervall der (differentiellen) Größe dx zu landen".[7]

Ein wunderbares Beispiel vermittelt das faire Glücksrad ohne Teilungsnägel, bei dem ρ definitionsgemäß konstant, $\rho(x) = (2\pi)^{-1}$, bzw. $m(x) = (2\pi)^{-1}x$ (Gleichverteilung). Der entsprechende W'keitsraum ist vom Typ *Lebesgue Maßraum*. Im Lebesgue Maßraum (und

[6] Die Notation mit dem $P(dx)$ ist zugegebenermaßen gewöhnungsbedürtig – und nicht etwa zu lesen "Funktion P der Variablen dx, wie etwa $P(dx) = dx^{17}\sin(dx)/\ln(dx)$". Nein, Nein. $P(dx)$ ist lediglich die (differentielle) W'keit im Intervall $[x, x+dx)$ zu landen. In der Eingewöhnungsphase dürfen Sie mit Blick auf (14) ruhig $P(dx) = \rho(x)dx$ lesen – und sich mit dem dx an der gewohnten Stelle wie zu Hause fühlen.

[7] W'keitsdichten sind das A und O der Quantenmechanik. Das Absolutquadrat $|\Psi(x)|^2$, beispielsweise, ist die W'keitsdichte für die Verteilung der Ortsmesswerte eines Punktteilchens, das im quantenmechanischen Zustand (Wellenfunktion) $\Psi(x)$ präpariert wurde.

immer dann wenn eine glatte W'keitsdichte gegeben ist) haben Singletons $\{x_0\}$, worin x_0 Punkt in $[0, 2\pi)$, und alle abzählbaren Vereinigungen von Singletons, notwendig das Maß Null. Interpretiert: die W'keit, bei einem fairen Glücksrad bei irgendeinem *bestimmten* Winkel zu landen, beispielsweise $x_0 = 1$, ist Null. Die W'keit in einem kleinen, um x_0 herum zentrierten Intervall dx zu landen, ist natürlich nicht Null, sonderen $\rho(x_0)dx$.

Das Gegenstück zum fairen Glücksrad ist das maximal gezinkte Glücksrad. Es bleibt immer bei einem bestimmten Winkel x_0 stehen. Eine derartig unfaire Präparation des Glücksrads wird durch ein sog. *Punktmaß* δ_{x_0} beschrieben, für beliebigen Messraum $(X, \mathcal{A})$ und festen Punkt $x_0 \in X$ definiert $\delta_{x_0}(A) := \chi_A(x_0)$ bzw.

$$\delta_{x_0}(A) := \begin{cases} 1 & \text{falls } A \ni x_0 \\ 0 & \text{sonst} \end{cases} \tag{15}$$

für alle Mengen $A \in \mathcal{A}$. Hat man dann eine abzählbare Teilmenge von X, bezeichnet X', ist mit

$$P := \sum_{x \in X'} p_x \delta_x \tag{16}$$

ein *diskretes W'keitsmaß* gegeben, sofern nur $p_x \geq 0$ und $\sum_{x \in X'} p_x = 1$.[8]

Die Definition des Punktmaßes wird in der Physik selten via (15) vorgenommen. Das maximal gezinkte Glücksrad, beispielsweise, wird dort vielmehr durch eine Verteilungsfunktion $m(x) = \theta(x - x_0)$ beschrieben, worin die θ die Sprungfunktion, entsprechen einer W'keitsdichte $\rho(x) = \delta(x - x_0)$, worin δ die **Dirac'sche Deltafunktion** (PhysikerInnen-Jargon). Angesichts $\int_A \delta(x - x_0)dx = 1$ falls $x_0 \in A$, und $= 0$ sonst, darf identifiziert werden $\int_A \delta(x - x_0)dx = \delta_{\varphi_0}(A)$. Kurz: das *Diracmaß* der Physik und das Punktmaß der Mathematik beschreiben ein-und-dasselbe (singuläre) Maß.

3 W'keit und Häufigkeit

Wahrscheinlichkeiten erfahren eine natürliche Interpretation durch sog. **relative Häufigkeiten**, die in langen Messreihen ermittelt werden. Würfelt man in einer Versuchsreihe N mal, und wertet das Ergebnis x eines Wurfes als "A-Treffer" falls $x \in A$, so ist die relative Häufigkeit die in der t-ten Versuchreihe aus den Messdaten (gewürfelten Augenzahlen) ermittelt wird

$$r(A; N, t) = \frac{[\text{Zahl der } A\text{-Treffer}]_t}{N} \,. \tag{17}$$

Das große Versprechen der Natur ist nun, daß im Grenzfall $N \to \infty$ die $r(A; N, t)$ mit Sicherheit einen von t unabhängigen (also objektiven) Wert $P(A)$ annehmen,

$$\lim_{N \to \infty} r(A; N, t) = P(A) \,. \tag{18}$$

Wir glauben der Natur, müssen aber darauf hinweisen daß diese Konstruktion philosophisch durchaus umstritten ist. "Mit Sicherheit" heisst genauer "mit an Sicherheit grenzender W'keit", und damit bewegt sich der Versuch, den Wahrscheinlichkeits-Begriff über relative

[8] Mit $\mu := \sum_{x \in X'} \delta_x$ ist ein Maß in X verabredet, genannt **Zählmaß**: für beliebige Teilmenge $A \in X$ zählt $\mu(A)$ wieviele Punkte x aus X' in A angetroffen werden.

Häufigkeiten zu definieren, im Kreis. "Relative Häufigkeit" – das lehrt der nächste Abschnitt – ist vielmehr selbst eine Zufallsvariable.

Auch sind W'keitsaussagen grundsätzlich weder verifizierbar noch falsifizierbar. Das Versprechen "Dieser Würfel ist nicht gezinkt" behaupet unter anderem, dass alle Augenzahlen mit der gleichen W'keit auftreten. Haben Sie Pech (oder Glück?), und würfeln permanent eine ⚄, könnte man meinen, Sie hätten die behauptete Fairness des Würfels widerlegt. Haben Sie aber nicht. Schließlich haben Sie sicherlich nicht unendlich oft gewürfelt. Finden Sie andererseits bei einer langen Messreihe eine relative Häufigkeit $r \approx 1/6$ ist damit auch nichts bewiesen: schließlich könnte der Würfel trotzdem gezinkt sein (ein kleiner Motor im Inneren sorgt dafür, dass nach der scheinbar zufälligen Sequenz immer nur die sechs oben liegt).

Die Mathematik braucht sich um derartige Einwände nicht zu kümmern. "W'keitstheorie" ist aus ihrer Sicht angewandte Maßtheorie, die durch eine Liste spezieller Axiome, den sog. Kolmogoroffschen Axiomen (des W'keitsraums – s.o.), genauestens präzisiert wird. Für den Physiker bleibt bei der probabilistischen Quantenmechanik allerdings ein Rest Unbehagens, muss er sich doch aus dem Paradies unverrückbarer Gewissheiten verabschieden. Trotzdem – auch der mit Unbehagen geplagte Physiker tut gut daran, bei einer beoachteten relativen Häufigkeit $r \approx 1/6$ auf einen fairen Würfel zu setzen. Eine gegenteilige Wette könnte seine bürgerliche Existenz schnell ruinieren. Mit Sicherheit.

4 Binomialverteilung

Ein Teilchen hüpft mit W'keit p nach rechts (Elementarereignis $\texttt{R}$) und mit W'keit $q := 1-p$ nach links (Elementarereignis $\texttt{L}$). Stichprobenraum ist hier $X = \{\texttt{R}, \texttt{L}\}$, der Ereignisraum $\mathcal{P}(X) = \{\emptyset, \{\texttt{R}\}, \{\texttt{L}\}, \{\texttt{R}, \texttt{L}\}\}$, und das W'keitsmaß auf $\mathcal{P}(X)$ ist durch die Angabe von p bereits vollständig charakterisiert,

$$P(\emptyset) = 0\,, \quad P(\{\texttt{R}\}) = p\,, \quad P(\{\texttt{L}\}) = 1-p\,, \quad P(\{\texttt{R},\texttt{L}\}) = 1\,. \tag{19}$$

Sind dem Teilchen zwei Schritte erlaubt – ist das Experiment also auch ein anderes – gibt es offensichtlich 4 verschiedene Resultate. Beispielsweise kann das Teilchen beide Schritte nach rechts machen mit dem Resultat $\texttt{RR}$, oder den ersten Schritt nach Rechts, den zweiten aber nach links mit dem Resultat $\texttt{RL}$ usw. Der Stichprobenraum ist in diesem Fall das kartesische Produkt $X^2 := X \times X$ – die Menge aller geordneten Sequenzen $s_1 s_2$ worin s_i einer der Schritte $\texttt{L}$ oder $\texttt{R}$.

Für ein Teilchen das N Schritte machen kann ist der Stichprobenraum das N-fache kartesiche Produkt $X^N := X \times \cdots \times X$. Jedes Elementarereignis $x \in X^N$ kann mit einem bestimmten Pfad identifiziert werden, beispielsweise

$$\underbrace{\texttt{LLRL}\cdots\texttt{RL}}_{\text{N Schritte}} \tag{20}$$

Da die einzelnen Schritte hier zufällig genommen werden, heißt so ein Pfad auch *Zufallspfad*. Unter den Annahme, dass jeder Schritt unabhängig von der vorangegangenen Schrittfolge "ausgewürfelt" wird, und dass die $\texttt{R}$-W'keit immer die Gleiche für jeden einzelnen Schritt, ist die W'keit, einen bestimmten Pfad x zu realisieren, gegeben $p^n(1-p)^{N-n}$, worin n die Gesamtzahl der Schritte `R`, und $N-n$ entsprechend die Gesamtzahl der Schritte `L`.

Meistens interessiert man sich aber nicht für einen bestimmten Pfad, sondern für eine Klasse von Pfade. Eine solche Klasse wäre zum Beispiel die Klasse S_n^N aller Pfade mit insgesamt n Schritten nach rechts und dementsprechend $N-n$ Schritten nach links, wobei Reihenfolge der Schritte beliebig ist. Jeder einzelne Pfad in dieser Klasse tritt mit der gleichen W'keit $p^n(1-p)^{N-n}$ auf, und da die Klasse insgesamt $\binom{N}{n}$ Mitglieder umfasst ist die W'keit, dass ein Pfad in die Klasse S_n^N fällt, gegeben

$$P_N(n) = \binom{N}{n} p^n(1-p)^{N-n} \,. \tag{21}$$

Die hier eingeführte Verteilung $P_N(n)$ heißt *Binomialverteilung* und spielt in der W'keitstheorie eine wichtige Rolle. Der binomische Lehrsatz

$$\sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} = (p+q)^N \tag{22}$$

in Verbindung mit $q = 1-p$ beweist, dass die Binomialverteilung korrekt normiert ist, $\sum_{n=0}^{N} P_N(n) = 1$.[9]

Die Binomialverteilung ist die W'keitsverteilung für die Zufallsgröße "Gesamtzahl der Schritte nach rechts", hier bezeichnet $\hat{n}$ (der Hut hat nichts mit Quantenmechanik zu tun). Wir schreiben $\hat{n}(\texttt{LLRL}\cdots\texttt{RL}) = n$ falls in der Sequenz $\texttt{LLRL}\cdots\texttt{RL}$ das Symbol `R` genau n mal vorkommt. Momente der Binomialverteilung

$$\langle \hat{n}^\nu \rangle := \sum_{n=0}^{N} n^\nu \binom{N}{n} p^n(1-p)^{N-n} \tag{23}$$

lassen sich mit einem "Ableitungstrick" leicht berechnen

$$\langle \hat{n}^\nu \rangle = \sum_{n=0}^{N} n^\nu \binom{N}{n} p^n q^{N-n} \Bigg|_{q=1-p} = \left[\left(p\frac{\partial}{\partial p} \right)^\nu \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} \right]_{q=1-p} \tag{24}$$

$$= \left[\left(p\frac{\partial}{\partial p} \right)^\nu (p+q)^N \right]_{q=1-p} \tag{25}$$

Die ersten beiden Momente berechnen sich zu

$$\langle \hat{n} \rangle = Np \,, \qquad \langle \hat{n}^2 \rangle = N^2p^2 + Np(1-p) \,, \tag{26}$$

und also die *mittlere quadratische Schwankung*

$$\langle (\hat{n} - \langle \hat{n} \rangle)^2 \rangle \equiv \langle \hat{n}^2 \rangle - \langle \hat{n} \rangle^2 = Np(1-p) \,. \tag{27}$$

[9] Im Anschluss an den vorhergehenden Abschnitt wäre die Binomialverteilung zu notieren $P(S_n^N)$. Der Nachweis der korrekten Normierung wäre dann mengentheoretisch zu führen: Für $n \neq m$ ist $S_n^N \cap S_m^N = \emptyset$, und da $\cup_{n=0}^{N} S_n^N = \Omega^N$ ist $\sum_{n=0}^{N} P(S_n^N) = 1$.

Prominenten Grenzfälle der Binomialverteilung im Limes $N \to \infty$ sind die **Gaussverteilung** (falls mit $N \to \infty$ auch $Np \to \infty$) und die Poissonverteilung (falls Np konstant im Limes $N \to \infty$).

Ist Np groß können $N!$, $n!$ und $(N-n)!$ durch die **Stirlingsche Formel**,

$$x! \approx \sqrt{2\pi x}\left(\frac{x}{e}\right)^x \tag{28}$$

gültig für große x, genähert werden, und man schaut auf

$$P_N(n) \approx \sqrt{\frac{N}{2\pi n(N-n)}}N^N\left(\frac{p}{n}\right)^n\left(\frac{1-p}{N-n}\right)^{N-n} . \tag{29}$$

Als Funktion von n (bei festem N und p) ist das Maximum via

$$\frac{\partial \ln P_N(n)}{\partial n} = 0 \tag{30}$$

schnell bestimmt,

$$n_{\max} = Np = \langle \hat{n} \rangle . \tag{31}$$

Taylorentwicklung um den Wert $n_{\max}$ liest sich

$$\ln P_N(n) = \ln P_N(Np) - \frac{1}{2}\frac{(n-Np)^2}{Np(1-p)} + \ldots \tag{32}$$

wobei die Korrekturterme $\ldots = \frac{1}{3!}\frac{1-2p}{(Np(1-p))^2}(n-Np)^3 + \ldots$ von Ordnung $O(1/\sqrt{N})$ im Limes $N \to$ vernachlässigt werden können.[10] Im Ergebnis

$$P_N(n) = \frac{1}{\sqrt{2\pi Np(1-p)}}\exp\left\{-\frac{(n-Np)^2}{2Np(1-p)}\right\} , \tag{33}$$

also um Np zentrierte **Gaussglocke** der Breite $\sqrt{Np(1-p)}$. Im Limes $N \to \infty$ zieht sich die Glocke immer mehr zusammen – die Verteilung wird **scharf** um den Mittelwert zentriert.

Für N groß, aber $Np = \langle n \rangle$ fest (somit $p \sim N^{-1}$ winzig)) kann $n!$ nicht durch die Stirling-Formel genähert werden, wohl aber $N!$ und $(N-n)!$,

$$\frac{N!}{(N-n)!} \approx \frac{N^N e^{-n}}{N^{N-n}(1-n/N)^{N-n+\frac{1}{2}}} . \tag{34}$$

Da nun $n/N \ll 1$ für relevante n, darf weiter approximiert werden $(1-n/N)^{N-n+\frac{1}{2}} \approx (1-n/N)^N \approx e^{-n}$. Im Ergebnis

$$P_N(n) = \frac{(Np)^n}{n!}e^{-Np} , \tag{35}$$

also **Poissonverteilung** zum Mittelwert $\langle \hat{n} \rangle = Np$. Angesichts ihrer quadratischen Varianz $\langle \hat{n}^2 \rangle - \langle \hat{n} \rangle^2 = \langle \hat{n} \rangle = Np$ ist die Poissonverteilung durch ihren Mittelwert bereits vollständig

[10] Beachte $(n-Np)^2/(Np(1-p)) = (n-\langle n \rangle)^2/\sigma^2 \sim O(1)$.

festgelegt. Poisson-verteilt sind beispielsweise die Photonenzahlen im Strahl eines idealen single-mode Lasers.

Die relative Häufigkeit das Symbol R in einer gegebenen Sequenz $x \in X^N$ anzutreffen ist eine Zufallsvariable, definiert $\hat{r} := \frac{\hat{n}}{N}$. Offensichtlich $\langle \hat{r} \rangle = p$: die relative Häufigkeit für einen Schritt nach rechts ist *im Mittel* genau die W'keit p, einen Schritt nach rechts zu machen! Für die Varianz ergibt sich $\sigma(\hat{r}) = \sqrt{\frac{p(1-p)}{N}}$. Für lange Sequenzen $N \to \infty$ geht die Varianz nach Null $\sim 1/\sqrt{N}$, formuliert als **Gesetz der großen Zahlen**: bei langen Sequenzen gleicht die relative Häufigkeit der R-Schritte mit hoher W'keit ihrem Mittelwert p, im Limes $N \to \infty$ gar mit W'keit Eins.

Die Verallgemeinerung findet sich wieder im **zentralen Grenzwertsatz**: Hat man eine Kollektion N unabhängigiger Zufallsvariabler X_i mit endlichen Momenten $\langle X_i^n \rangle$, so ist die Zufallsvariable $Y := \frac{1}{N} \sum X_i$ im Grenzfall $N \to \infty$ Gauß-verteilt, auch genannt **normal verteilt**, mit Mittelwert

$$\langle Y \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle X_i \rangle \,, \tag{36}$$

und mittlerer quadratischer Schwankung

$$\sigma(Y)^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sigma(X_i)^2 \,. \tag{37}$$

Kurz: die Verteilung $p_Y(y) = \langle \delta(y - Y) \rangle$ ist im Limes $N \to \infty$ eine Gauss-Funktion, und das unabhängig davon, wie die X_i verteilt sind.

5 Ergänzung: Borel, Lebesgue und so

Wohlvertraut das *Intervall* $[a, b) \subset \mathbb{R}$, das hier aus technischen Gründen halboffen gewählt wird, dessen *Elementarinhalt* durch die reelle Zahl $\lambda([a, b)) := b - a$ definiert ist. Das System aller Intervalle kann zu einer Menge zusammengefasst werden, bezeichnet $\mathcal{I}$. Jede endliche Vereinigung von Intervallen aus $\mathcal{I}$ ist eine Teilmenge des $\mathbb{R}$, genannt eine *Figur*. Für das System der Figuren, bezeichnet $\mathcal{F}$, gilt nun: Die leere Menge ist Element von $\mathcal{F}$, und mit F und G zwei Figuren aus $\mathcal{F}$ sind auch $F \setminus G$ sowie $F \cup G$ Figuren aus $\mathcal{F}$. Kurz: die Menge der Figuren in $\mathbb{R}$ bilden einen *Ring* über $\mathbb{R}$.

Jede Figur $F \in \mathcal{F}$ gestattet eine eindeutige Darstellung als endliche Vereinigung von paarweise elementfremden Intervallen $I_i \in \mathcal{I}$, also $F = \cup_{i=1}^n I_i$, und da Intervalle über ihren Elementarinhalt messbar, ist auch jede Figur messbar, für paarweise elementfremde Mengen $\lambda(F) = \sum_{i=1}^n \lambda(I_i)$, genannt *Lebesguesches Prämaß*. Das Adelsprädikat “Maß” wird hier verwehrt, da der Definitionsbereich von λ lediglich ein Ring, und keine σ-Algebra (beispielsweise ist $\mathbb{R}$ selber nach beiden Seiten offen, und also kein Element von $\mathcal{F}$).

Nun existiert zu jedem System $\mathcal{S}$ von Teilmengen von Ω eine kleinste, $\mathcal{S}$ enthaltende *sigma*-Algebra $\mathcal{A}(\mathcal{S})$, für das Mengensystem der Figuren genannt die **Borelsche σ-Algebra**, bezeichnet $\mathcal{B}$, und es sind genau die Elemente von $\mathcal{B}$, genannt *Borelsche Mengen*, die eine Fortsetzung des Lebesgueschen Prämaßes zu einem Maß gestatten, genannt das *Lebesgue-Borel Maß in* $\mathbb{R}$. Fortsetzung bedeutet dabei, dass λ für alle Figuren $F \in \mathcal{F} \subset \mathcal{B}$ mit $\bar{\lambda}$ identisch.

Das Lebesgue-Borel Maß ist soetwas wie der Archimedische Punkt der Maßtheorie. Mit seiner Hilfe lassen sich unzählige andere Maße erklären die in allen möglichen Kontexten von Interesse sind. Außerdem kommt es mit einigen netten Eigenschaften, wovon vielleicht die Translationsinvarianz die wichtigste.

`[Hier noch Beispiel von Vitali?]`

DIe Borel'schen Mengen des n-dimensionalen Zahlenraums $\mathbb{R}^n$ $(n = 1, 2, \ldots)$: Für je zwei Punkte $a = (\alpha^1, \ldots, \alpha^n)$ und $b = (\beta^1 \ldots, \beta^n)$ des $\mathbb{R}^n$ sind die Relationen $a \leq b$ bzw. $a < b$ über die Koordinaten erklärt $\alpha^i \leq \beta^i$ bzw. $\alpha^i < \beta^i$ für alle $i = 1, \ldots, n$. In Verallgemeinerung des Intervalls $[a, b) \in \mathbb{R}$ Für jede Menge der Gestalt

$$[a, b) := \{x \in \mathbb{R}^n | a \leq x < b\} \qquad \text{wobei } a \leq b \tag{38}$$

geometrisch ein achsparalleler, "nach rechts hin offener" Quader, unter Gebildeten genannt *Parallelotop*, ist mit

$$\lambda^n \left([a, b)\right) := (\beta^1 - \alpha^1) \cdot (\beta^2 - \alpha^2) \cdots (\beta^n - \alpha^n) \tag{39}$$

eine reelle Zahl gegeben, genannt der *Elementarinhalt* von $[a, b)$. Der Elementarinhalt ist offensichtl genau dann gleich Null, wenn $[a, b)0\emptyset$, wenn also zwar $a \leq b$, nicht aber $a < b$. Der Elementarinhalt wird nun herangezogen, um auch abzählbaren Vereinigungen solcher Quadermengen einen Inhalt zuzuweisen. Fasst man alle Quader zu einer Menge $\mathcal{I}$ zusammen, wird für paarweise elementfremde Quader $[a_i, b_i) \in \mathcal{I}$ $(i = 1, \ldots, k)$ verabredet

$$\lambda^n \left(\cup_{i=1}^k [a_i, b_i)\right) = \sum_{i=1}^k \lambda^n \left([a_i, b_i)\right) \tag{40}$$

UndSoWeiterUndSoFort ...