ELEMENTARE STOCHASTIK

Vorlesung gehalten von
Martin Aigner

Wintersemester 2008/9

Inhaltsverzeichnis

1 Grundbegriffe

1.1 Zufall

Stochastik ist die Lehre von den Gesetzmäßigkeiten des Zufalls. Dies scheint zunächst ein Widerspruch in sich zu sein. Trotzdem wird jeder annehmen, dass bei wiederholtem Münzwurf in etwa der Hälfte der Fälle Kopf bzw. Zahl erscheint. Dies ist genau so eine Gesetzmäßigkeit, die wir studieren wollen.

Die Stochastik besteht aus zwei Teilen. Die *Wahrscheinlichkeitstheorie* beschreibt zufällige Vorgänge mit Modellen, die *Statistik* zieht Schlußfolgerungen aus den Beobachtungen auf den realen Vorgang und das Modell.

Beispiel. Die W-Theorie modelliert das Werfen einer Münze mit einem sogenannten Bernoulli-Modell. Die Statistik schließt aus den Beobachtungen, ob die Münze fair ist.

Grundlegend für die Behandlung zufälliger Vorgänge sind folgende Bedingungen:

A. Das Zufallsexperiment ist beliebig oft wiederholbar.

B. Es gibt keine äußeren Informationen über den Ausgang.

C. Die Erfahrung über bisherige Ausgänge nützt nichts.

1.2 Wahrscheinlichkeitsräume

Die Stochastik, wie wir sie heute betreiben, beginnt mit den Axiomen, die Kolmogorov 1933 postuliert hat.

Zunächst ist eine Menge Ω von *Elementarereignissen* gegeben. Zum Beispiel ist $\Omega = \{K(\text{opf}), Z(\text{ahl})\}$ beim Münzwurf oder $\Omega = \{1, 2, \dots, 6\}$ beim Würfeln. Jede Teilmenge $A \subseteq \Omega$ heißt ein *Ereignis*. Zum Beispiel $A = \{\text{mindestens } k \text{ Mal Kopf bei } n \text{ Münzwürfen}\}$ oder $A = \{\text{gerade Zahl}\}$ beim Würfeln.

Definition. $\mathcal{E} \subseteq 2^\Omega$ ist eine *σ-Algebra über* Ω, falls gilt:

1. $\varnothing \in \mathcal{E}$,

2. $A \in \mathcal{E} \Rightarrow A^c = \Omega \smallsetminus A \in \mathcal{E}$,

3. $A_1, A_2, A_3, \ldots \in \mathcal{E} \Rightarrow \bigcup\limits_{i\geq 1} A_i \in \mathcal{E}$.

Beispiel. $\mathcal{E} = \{\varnothing, \Omega\}$, $\mathcal{E} = 2^\Omega$. Sind $\mathcal{F}_i$ σ-Algebren über Ω, dann auch $\bigcap \mathcal{F}_i$. Ist also $\mathcal{E}_0 \subseteq \Omega$. beliebig, so gibt es eine kleinste σ-Algebra $[\mathcal{E}_0]$, die $\mathcal{E}_0$ enthält, nämlich $[\mathcal{E}_0] = \bigcap \mathcal{F}$, $\mathcal{E}_0 \subseteq \mathcal{F}$, $\mathcal{F}$ σ-Algebra. $[\mathcal{E}_0]$ heißt die von $\mathcal{E}_0$ *erzeugte* σ-Algebra.

Beispiel. Die *Borelmengen* in $\mathbb{R}$ (oder $\mathbb{R}^n$) sind die von den offenen Intervallen erzeugte σ-Algebra.

Wir kennen die Gesetze von de Morgan:

$$(\bigcup A_i)^c = \bigcap A_i^c, \quad (\bigcap A_i)^c = \bigcup A_i^c .$$

Aus der Definition können wir sofort ein paar Folgerungen ziehen.

1. $\mathcal{E}$ enthält alle endlichen Vereinigungen und Durchschnitte, denn mit $A_1, \ldots, A_n \in \mathcal{E}$ ist

$$A_i \cup \ldots \cup A_n \cup \varnothing \cup \varnothing \cup \ldots = A_1 \cup \ldots \cup A_n \in \mathcal{E} .$$

Ferner ist mit $A_i \in \mathcal{E}$ auch $A_i^c \in \mathcal{E}$, also

$$(A_1 \cap \ldots \cap A_n)^c = A_1^c \cup \ldots \cup A_n^c \in \mathcal{E} ,$$

und somit $A_1 \cap \ldots \cap A_n \in \mathcal{E}$.

2. Mit $A, B \in \mathcal{E}$ ist auch $A \setminus B \in \mathcal{E}$. Es gilt

$$A \setminus B = A \cap B^c \in \mathcal{E} .$$

3. Alle abzählbaren Vereinigungen und Durchschnitte von $\mathcal{E}$-Mengen A_i sind in $\mathcal{E}$.

Für die Vereinigung ist dies Axiom 3), und für die Durchschnitte verwenden wir de Morgans Gesetz.

Definition. Ein *Wahrscheinlichkeitsraum* ist ein Tripel $(\Omega, \mathcal{E}, P)$, wobei

1. $\mathcal{E} \subseteq 2^\Omega$ ist σ-Algebra,
2. $P : \mathcal{E} \to [0, 1]$ heißt das *Wahrscheinlichkeitsmaß* und erfüllt

i) $P(\Omega) = 1$,

ii) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, falls $A_1, A_2, \ldots$ paarweise disjunkte Mengen in $\mathcal{E}$ sind.

Bemerkung. Falls Ω endlich oder abzählbar ist, so nehmen wir immer die σ-Algebra $\mathcal{E} = 2^{\Omega}$. Für überabzählbare Mengen Ω (wie z.B. die Menge der Ausgänge beim unendlich oft wiederholten Münzwurf) funktioniert 2^{Ω} nicht. Das heißt, es gibt kein geeignetes W-Maß. (Beweis siehe Georgii).

Wir ziehen wiederum ein paar unmittelbare Folgerungen.

1. $P(\varnothing) = 0$.

Es ist $P(\varnothing) = P(\varnothing \cup \varnothing \cup \varnothing \cup \ldots) = \sum P(\varnothing)$, und dies ist nur für $P(\varnothing) = 0$ erfüllt.

2. $P(A) + P(A^c) = 1$.

Es ist $\Omega = A \cup A^c \cup \varnothing \cup \varnothing \cup \ldots$, also $1 = P(\Omega) = P(A) + P(A^c)$.

3. $A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$.

Das folgt aus $A \cup B = A \cup B \cup \varnothing \cup \varnothing \ldots$

4. $A \subseteq B \Rightarrow P(B) = P(A) + P(B \setminus A) \geq P(A)$.

Das W-Maß ist also monoton steigend.

5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Wir haben $(A \cup B) \setminus A = B \setminus (A \cap B)$, also nach 4)

$$P(A \cup B) - P(A) = P(B) - P(A \cap B)\,.$$

6. $P(\bigcup_{n\geq 1} A_n) \leq \sum_{n\geq 1} P(A_n)$.

Wir stellen $\bigcup_{n\geq 1} A_n$ als Vereinigung disjunkter Mengen dar:

$$\bigcup_{n\geq 1} A_n = \bigcup_{n\geq 1} (A_n \setminus \bigcup_{j<n} A_j)\,,$$

und erhalten mit 4)

$$P(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} P(A_n \setminus \bigcup_{j<n} A_j) \leq \sum_{n\geq 1} P(A_n)\,.$$

Beispiel. 1. Werfen einer Münze. $\Omega = \{0,1\}$, $1 =$“Kopf”, $0 =$“Zahl”, $P(1) = p$, $P(0) = 1-p$, $0 \leq p \leq 1$.

2. Würfeln. $\Omega = \{1,2,\ldots,6\}$, $P(k) = \frac{1}{6}$.

3. $\Omega = [0,1] \subseteq \mathbb{R}$, $P([a,b]) = b-a$.

Wir werden öfters Produkte von W-Räumen betrachten. Seien $\mathcal{E}_i$ σ-Algebren auf Ω_i, $i = 1,\ldots,n$, dann ist die Produktalgebra $\mathcal{E}^{\otimes}$ auf $\Omega_1 \times \ldots \times \Omega_n$ die von $\mathcal{E}_1 \times \ldots \times \mathcal{E}_n$ erzeugte σ-Algebra.

Die ersten interessanten Ergebnisse sind die folgenden.

Lemma 1.1. a) *Sei $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ eine aufsteigende Kette von Ereignissen, dann gilt*

$$P(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n)\,.$$

b) *Ist $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$ eine absteigende Kette, so gilt*

$$P(\bigcap_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n)\,.$$

Beweis. Wir haben

$$A_n = A_1 \cup (A_2 \setminus A_1) \cup \ldots \cup (A_n \setminus A_{n-1})\,,$$

also

$$P(A_n) = P(A_1) + P(A_2 \setminus A_1) + \ldots + P(A_n \setminus A_{n-1})\,,$$

und es folgt mit $A_0 = \varnothing$

$$\begin{aligned} P(\bigcup_{n\geq 1} A_n) &= P(\bigcup_{n\geq 1}(A_n \setminus A_{n-1})) = \sum_{n\geq 1} P(A_n \setminus A_{n-1}) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i \setminus A_{i-1}) = \lim_{n\to\infty} P(A_n)\,. \end{aligned}$$

Aussage b) folgt mit de Morgan. $\square$

Folgerung 1.2. a. *Ist* $A_1 \subseteq A_2 \subseteq \ldots$ *mit* $P(A_n) = 0$ *für alle* n*, so gilt* $P(\bigcup A_n) = 0$.

b. *Ist* $A_1 \supseteq A_2 \supseteq \ldots$ *mit* $P(A_n) = 1$ *für alle* n*, so gilt* $P(\bigcap_{n\geq 1} A_n) = 1$.

Wir kennen $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Die Verallgemeinerung auf n Mengen ist das *Prinzip der Inklusion - Exklusion*.

Lemma 1.3. *Es gilt*

$$P(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Beweis. Für $n = 1, 2$ ist das richtig, nun verwenden wir Induktion:

$$\begin{aligned} P(A_1 \cup \ldots \cup A_n) &= P((A_1 \cup \ldots \cup A_{n-1}) \cup A_n) \\ &= P(A_1 \cup \ldots \cup A_{n-1}) + P(A_n) - P((A_1 \cup \ldots \cup A_{n-1}) \cap A_n) \\ &= P(A_1 \cup \ldots \cup A_{n-1}) + P(A_n) - P(\bigcup_{i=1}^{n-1} (A_i \cap A_n)). \end{aligned}$$

Der erste Summand ergibt alle Summen mit $1 \leq i_1 < \ldots < i_k \leq n-1$ nach Induktion, der zweite die Summe mit $k = 1$, $i_1 = n$, und für den dritten gilt wieder nach Induktion

$$P(\bigcup_{i=1}^{n-1} (A_i \cap A_n)) = \sum_{k=2}^{n-1} (-1)^{k-2} \sum_{1 \leq i_1 < \cdots < i_{k-1} \leq n-1} P(A_{i_1} \cap \ldots \cap A_{i_{k-1}} \cap A_n),$$

und mit dem Minuszeichen folgt die Formel. $\square$

Schreiben wir Ω für den leeren Durchschnitt, so können wir die Formel auch auf folgende Form bringen:

$$P(\Omega \setminus (A_1 \cup \ldots \cup A_n) = \sum_{k=0}^{n} (-1)^{k} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \ldots \cap A_{i_k}).$$

1.3 Diskrete Wahrscheinlichkeitsräume

Wenn Ω endlich oder abzählbar ist, so heißt $(\Omega, 2^{\Omega}, P)$ ein *diskreter* Wahrscheinlichkeitsraum. Da $\sum_{\omega \in A} P(\omega) \leq 1$ ist für jedes $A \subseteq \Omega$, so konvergiert

$\sum_{\omega\in A} P(\omega) = P(A)$. Ein disjunkter W-Raum ist also durch Angabe von $P(\omega)$, $\omega \in \Omega$, vollkommen determiniert, wobei $\sum_{\omega\in\Omega} P(\omega) = 1$ ist.

Beispiele.
1. **Gleichverteilung.** Ω endlich, $P(\omega) = \frac{1}{|\Omega|}$, also $P(A) = \frac{|A|}{|\Omega|}$. Man spricht von einem *Laplace Modell*, und interpretiert $\frac{|A|}{|\Omega|}$ als "günstige" durch "mögliche" Fälle. Als Beispiele haben wir eine faire Münze, Würfel oder Lotto.

2. **Binomialverteilung.** Wir führen ein Experiment mit den Ausgängen 1 und 0 n Mal aus, wobei die W-keit von 1 gleich p ist, von 0 gleich $1-p$. Wir interpretieren 1 als "Erfolg" und 0 als "Mißerfolg" und sprechen von einem *Bernoulli Modell*. Was uns interessiert, ist die Anzahl k der Erfolge bei n Versuchen. Es ist also $\Omega = \{0, 1, 2, \ldots, n\}$, $P(k) = \binom{n}{k} p^k (1-p)^k$, und wir setzen

$$b(k, n; p) = \binom{n}{k} p^k (1-p)^{n-k} .$$

3. **Poissonverteilung.** Hier ist $\Omega = \{0, 1, 2, 3, \ldots\}$,

$$P(n) = p(n; \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}, \;\; \lambda > 0 .$$

Wir sprechen von einem *Poisson Modell*.

4. **Geometrische Verteilung.** Wir führen wie beim Bernoulli Modell ein Experiment beliebig oft aus, mit $p = W$-keit für 1, $q = 1 - p = W$-keit für 0. Was uns interessiert, ist, in welchem Versuch Erfolg zum ersten Mal auftritt. Also $\Omega = \mathbb{N}$,

$$P(n) = q^{n-1}(1-q), \quad 0 \le q < 1 .$$

Als Beispiel sei A_0 das Ereignis, dass der erste Erfolg in einem Versuch mit gerader Nummer auftritt. Dann ist

$$\begin{aligned} P(A_0) &= \; q(1-q) + q^3(1-q) + \ldots = q(1-q) \sum_{i\ge 0} q^{2i} \\ &= \; q(1-q) \frac{1}{1-q^2} = \frac{q}{1+q} . \end{aligned}$$

Für das komplementäre Ereignis $A_1 = \mathbb{N} \smallsetminus A_0$ haben wir daher

$$P(A_1) = 1 - \frac{q}{1+q} = \frac{1}{1+q} > \frac{q}{1+q} = P(A_0) .$$

5. **Urnenmodelle.** Dies sind die klassischen Modelle. Gegeben seien N Kugeln in einer Urne, von denen n gezogen werden. Wir numerieren die Kugeln $\{1, 2, \ldots, N\}$. Eine Ziehung ist also ein Wort $a_1 a_2 \ldots a_n$, $a_i \in \{1, 2, \ldots, N\}$, und Ω sei die Menge der Ziehungen. Wir haben vier Klassen:

(A) Mit Zurücklegen:

- (A1) geordnet, das heißt die Reihenfolge spielt eine Rolle. Dann ist $|\Omega| = N^n$.
- (A2) ungeordnet, die Reihenfolge, in der die Kugeln gezogen werden, ist belanglos. Dann ist $|\Omega| = \binom{N+n-1}{n} = \frac{N(N+1)\ldots(N+n-1)}{n!}$.

(B) Ohne Zurücklegen:

- (B1) geordnet, $|\Omega| = N(N-1)\cdots(N-n+1)$,
- (B2) ungeordnet, $|\Omega| = \binom{N}{n} = \frac{N(N-1)\cdots(N-n+1)}{n!}$.

Beweis von (A2). Jede Ziehung entspricht einem Wort $1 \leq a_1 \leq a_2 \leq \ldots \leq a_n \leq N$. Zu $a_1 a_2 \ldots a_n$ assoziieren wir das Wort $1 \leq a_1 < a_2 + 1 < a_3 + 2 < \ldots < a_n + n - 1 \leq N + n - 1$. Dies ergibt eine Bijektion auf die Menge aller n-Untermengen von $\{1, 2, \ldots, N + n - 1\}$, also ist $|\Omega| = \binom{N+n-1}{n}$.

Beispiel. Geburtstagsparadox. Bei einer Party sind n Personen. Was ist die W-keit, dass keine zwei Personen am selben Tag Geburtstag haben? Wir schließen Schaltjahre aus, also ist $N = 365$. Die Geburtstage entsprechen dann einer Ziehung nach Modell (B1). Unter der Voraussetzung der Gleichverteilung aller Tage haben wir also für die W-keit "keine zwei Personen haben am selben Tag Geburtstag"

$$\frac{365 \cdot 364 \cdot \cdots (365 - n + 1)}{365^n} = 1(1 - \frac{1}{365})(1 - \frac{2}{365}) \cdots (1 - \frac{n-1}{365})$$

und das ist $< \frac{1}{2}$ für $n \geq 23$.

Von 1936 - 1978 gab es 24 Fieldsmedaillenpreisträger, und tatsächlich haben Ahlfors (1936) und Fefferman (1978) beide am 18. April Geburtstag.

6. **Hypergeometrische Verteilung.** Gegeben sind n Kugeln, davon r rote und $n - r$ weiße. Wir ziehen m Kugeln ohne Zurücklegen, und interessieren uns für die W-keit, dass genau k rote unter den m gezogenen Kugeln sind.

Wir haben Modell (B2), die Anzahl der möglichen Ziehungen ist $\binom{n}{m}$ und die Anzahl der günstigen $\binom{r}{k}\binom{n-r}{m-k}$. Also ist die W-keit

$$h(k,m;r,n) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} .$$

Wir sprechen von einem *hypergeometrischen* Modell.

Beispiel. Eine übliche Lottoausspielung 6 aus 49 entspricht dieser Situation mit $n = 49$, $r = 6$ (richtige), $m = 6$ (angekreuzte Zahlen). Also ist die W-keit für k richtige

$$h(k,6;6,49) = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}} .$$

Die W-keit für 6 richtige ist $\frac{1}{\binom{49}{6}} \sim 7 \cdot 10^{-8} \sim \frac{1}{14 \text{ Mill}}$. Die W-keit für ≥ 3 richtige ist $\sim 0,018$, und die W-keit für 0 richtige ist $\frac{\binom{43}{6}}{\binom{49}{6}} \sim 0,436$, also immer noch fast 50%.

1.4 Wahrscheinlichkeitsräume mit Dichtefunktion

Die andere wichtige Klasse von W-Räumen betrifft kontinuierliche Prozesse. Ω ist $\mathbb{R}$ oder eine "einfache" Teilmenge wie $[a,b]$, oder allgemein $\mathbb{R}^n$ bzw. einfache Teilmengen davon. Die σ-Algebra enthält alle Borelmengen in Ω.

Wir erklären das W–Maß P durch eine sogenannte *Dichtefunktion* $f(x)$. Dabei wird vorausgesetzt:

1. $f(x) \geq 0$,
2. $f(x)$ ist stückweise stetig,
3. $\int\limits_{\Omega} f(x)dx = 1$.

Die entsprechenden Forderungen gelten für $\mathbb{R}^n$.

Wir erklären nun für $A \in \mathcal{E}$

$$P(A) = \int\limits_A f(x)dx .$$

Zum Beispiel ist $P([a,b]) = \int\limits_a^b f(x)dx$.

Nach den Regeln der Integrationsrechung erfüllt P die Forderungen an ein W-Maß. Wir können $f(x)$ auf ganz $\mathbb{R}$ (oder $\mathbb{R}^n$) erklären, indem wir $f(x) = 0$ für $x \notin \Omega$ setzen. Dies wird in den uns interessierenden Fällen keine Schwierigkeiten bereiten.

Die *Verteilungsfunktion* ist

$$F(x) := \int_{-\infty}^{x} f(t)dt\,.$$

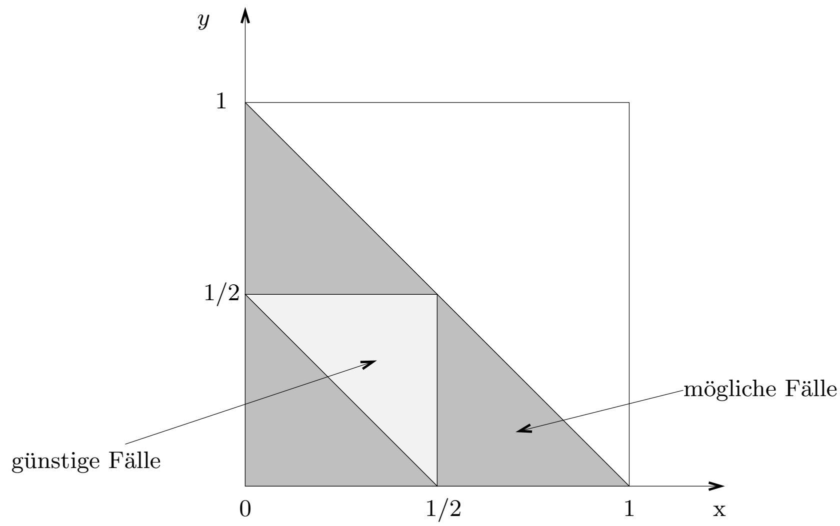
$F(x)$ ist eine monoton wachsende differenzierbare Funktion, und es gilt

$$F'(x) = f(x)\,.$$

Beispiele.

1. Gleichverteilung. Es sei $\Omega = [a,b]$, $f(x) = \frac{1}{b-a}$ für alle $x \in \Omega$. Offenbar ist $\int\limits_a^b f(x)dx = 1$. Die Verteilungsfunktion ist $F(x) = \int\limits_a^x \frac{1}{b-a}dt = \frac{x-a}{b-a}$. Als Beispiel haben wir $P([\alpha,\beta]) = \frac{\beta-\alpha}{b-a}$ für $[\alpha,\beta] \subseteq [a,b]$. Für allgemeines Ω setzen wir $I = \int\limits_\Omega 1dx \in (0,\infty)$ und erklären die Gleichverteilung durch $f(x) = \frac{1}{I}$ für alle x.

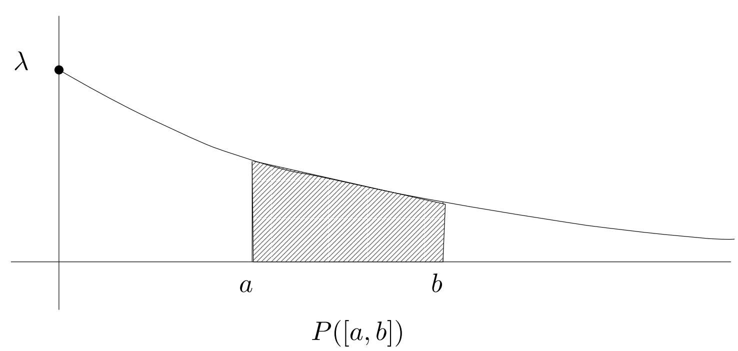
Beispiel. Der Weihnachtsmann bricht seinen Stab in drei Teile. Was ist die W-keit p, dass die drei Teile ein Dreieck bilden? Wir nehmen an, der Stab hat Länge 1, und die Teile können auch Länge 0 haben (die W-keit dafür ist natürlich 0). Seien x, y zwei Längen, dann ist $\Omega = \{(x,y) : x \geq 0, y \geq 0, x+y \leq 1\}$. Aufgrund der Dreiecksungleichung können wir genau dann ein Dreieck bilden, wenn $x+y \geq \frac{1}{2}$, $x \leq \frac{1}{2}$, $y \leq \frac{1}{2}$ gilt. Damit haben wir folgende Situation



Unter der Annahme der Gleichverteilung ist somit $p = \frac{1}{4}$.

2. Exponentialverteilung. Hier ist $\Omega = [0, \infty)$, $\lambda > 0$, und $f(x) = \lambda e^{-\lambda x}$. Wir haben $\int\limits_0^\infty \lambda e^{-\lambda x} dx = -e^{\lambda x}\Big|_0^\infty = 1$, also ist $f(x)$ Dichtefunktion. Die Verteilungsfunktion ist

$$F(x) = \int\limits_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x}\,.$$



Die Exponentialverteilung ist wichtig für Wartezeitenmodelle, die wir später behandeln werden.

3. Normalverteilung. Hier ist $\Omega = \mathbb{R}$, $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. Es ist nicht unmittelbar einsichtig, dass $\int\limits_{-\infty}^{\infty} f(x)dx = 1$ ist. Zum Beweis verwenden wir den folgenden Trick.

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy = \int\limits_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \left(\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) dy = \left(\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right)^2 .$$

Transformation auf Polarkoordinaten $x = r\cos\vartheta$, $y = r\sin\vartheta$, $dxdy = rdrd\vartheta$ ergibt

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy = \int\limits_{0}^{2\pi} \left(\int\limits_{0}^{\infty} re^{-\frac{r^2}{2}} dr \right) d\theta = \int\limits_{0}^{2\pi} \left(-e^{-\frac{r^2}{2}} \Big|_0^{\infty} \right) d\vartheta$$

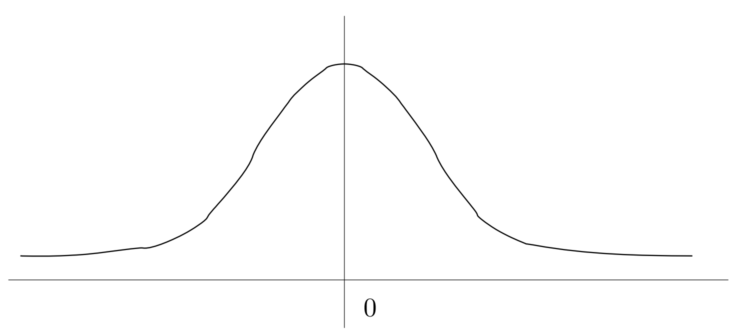
$$= \int\limits_{0}^{2\pi} d\vartheta = 2\pi \,,$$

und es folgt $\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}\,.$

Wir sprechen von der *Standard Normalverteilung* $N(0,1)$, und bezeichnen die Verteilungsfunktion mit

$$\phi(x) = \int\limits_{\infty}^{x} f(t)dt \,.$$

Die Dichte ist die berühmte *Gaußsche Glockenkurve.*



Sie wird uns in den fundamentalen Grenzwertsätzen wieder begegnen.

Die Normalverteilung $N(\mu, \sigma^2)$ ist erklärt durch die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

4. Gammaverteilung. Die Funktion

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx \quad (r \geq 0)$$

heißt die Γ-*Funktion*. Sie ist eine der fundamentalen Funktionen der Analysis. Wir stellen einige Eigenschaften zusammen:

1. $\Gamma(1) = 1$. Dies folgt durch Integration $\int_0^\infty e^{-x} dx = -e^{-x}\Big|_0^\infty = 1$.

2. $\Gamma(r+1) = r\Gamma(r)$. Wir erhalten durch partielle Integration

$$\begin{aligned}\Gamma(r+1) &= \int_0^\infty x^r e^{-x} dx = \underbrace{x^r(-e^{-x})\Big|_0^\infty}_{=0} + r \int_0^\infty x^{r-1} e^{-x} dx \\ &= r\Gamma(r) .\end{aligned}$$

Aus $\Gamma(1) = 1$ folgt insbesondere $\Gamma(n+1) = n!$ für $n \in \mathbb{N}$.

Die $\gamma_{\alpha,r}$-Verteilung ist erklärt durch die Dichtefunktion

$$f(x) = \frac{\alpha^r x^{r-1} e^{-\alpha x}}{\Gamma(r)} \text{ auf } [0, \infty), \;\; \alpha, r > 0 .$$

Mit der Transformation $y = \alpha x$ haben wir

$$\int_0^\infty \alpha^r x^{r-1} e^{-\alpha x} dx = \alpha \int_0^\infty y^{r-1} e^{-y} \frac{1}{\alpha} dy = \Gamma(r) ,$$

also ist $f(x)$ Dichtefunktion.

Der Spezialfall $r = 1$ ergibt $f(x) = \alpha e^{-\alpha x}$, also die Exponentialverteilung.

Beispiel. Was ist $\Gamma(\frac{1}{2})$? Wir haben für $r = \alpha = \frac{1}{2}$

$$\Gamma(\frac{1}{2}) = \frac{1}{\sqrt{2}} \int\limits_0^\infty x^{-\frac{1}{2}} e^{-\frac{x}{2}} dx \,.$$

Mit der Transformation $x = y^2$ erhalten wir mit $dx = 2ydy$

$$\int\limits_0^\infty x^{-\frac{1}{2}} e^{-\frac{x}{2}} dx = \frac{1}{2} \int\limits_{-\infty}^\infty \frac{1}{y} e^{-\frac{y^2}{2}} 2ydy = \sqrt{2\pi}\,,$$

und somit

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}\,.$$

Die Γ-Verteilung wird in der Statistik eine große Rolle spielen.

1.5 Bedingte Wahrscheinlichkeit

Es sei ein W-Raum $(\Omega, \mathcal{E}, P)$ gegeben, $A, B \in \mathcal{E}$.

Definition. Die *bedingte W-keit von A unter der Annahme* von B ist

$$P(A|B) := \frac{P(A \cap B)}{P(B)}\,,$$

wobei $P(B) > 0$ vorausgesetzt wird.

Definition. Die Ereignisse A und B heißen *unabhängig*, falls

$$P(A \cap B) = P(A)P(B)$$

gilt oder äquivalent

$$P(A|B) = P(A), \quad P(B|A) = P(B)\,.$$

Die Gleichungen bedeuten, dass wir aus der Kenntnis von B keine zusätzliche Information über A erhalten, und umgekehrt.

Beispiel. 1. Würfeln. Sei $A = \{\text{gerade Zahl}\}$, $B = \{x \geq 3\}$. Dann ist $P(A \cap B) = \frac{1}{3}$, $P(A) = \frac{1}{2}$, $P(B) = \frac{2}{3}$, also $P(A|B) = \frac{1}{2}$, $P(B|A) = \frac{2}{3}$. Die Ereignisse A, B sind unabhängig. Wählt man $B = \{x \geq 4\}$, so ist $P(A \cap B) = \frac{1}{3}$, $P(B) = \frac{1}{2}$, also $P(A|B) = P(B|A) = \frac{2}{3}$. Die Ereignisse sind nicht unabhängig.

2. Geometrische Verteilung. Es sei $A = \{1, 2\}$, $B = \{\text{gerade}\}$. Wir haben $P(A) = (1-q) + q(1-q) = 1 - q^2$, $P(B) = \frac{q}{1+q}$, $P(A \cap B) = q(1-q)$, somit $P(A|B) = 1 - q^2 = P(A)$. A und B sind unabhängig.

Satz 1.4. *Es sei $\Omega = \bigcup\limits_k B_k$ disjunkte Vereinigung, $P(B_k) > 0$ für alle k.*

a. $P(A) = \sum\limits_k P(B_k)P(A|B_k)$.

b. *Formel von Bayes: Sei $P(A) > 0$, dann gilt*

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum\limits_k P(B_k)P(A|B_k)}\,.$$

Beweis. a. Wir haben

$$\begin{aligned} P(A) &= P(A \cap \bigcup_k B_k) = P(\bigcup_k (A \cap B_k)) = \sum_k P(A \cap B_k) \\ &= \sum_k \frac{P(A \cap B_k)P(B_k)}{P(B_k)} = \sum_k P(B_k)P(A|B_k)\,. \end{aligned}$$

b. Hier ist

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A \cap B_i)}{P(A)} \cdot \frac{P(B_i)}{P(B_i)} = \frac{P(B_i)P(A|B_i)}{P(A)}\,. \qquad \square$$

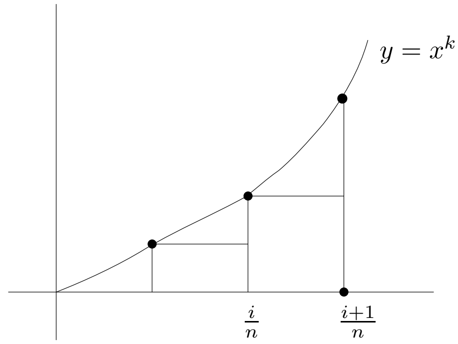
Beispiele. 1. Es sind n Urnen mit je n Kugeln gegeben. In Urne i befinden sich i rote und $n - i$ weiße.

Problem I. Eine Urne wird uniform mit W-keit $\frac{1}{n}$ gewählt. Was ist die W-keit p_k, dass bei k-maligem Ziehen mit Zurücklegen aus der gewählten Urne k Mal eine rote Kugel gezogen wird?

Problem II. Eine Urne wird mit W-keit $\frac{1}{n}$ gewählt. Es wird k Mal gezogen mit Zurücklegen und jedes Mal erscheint rot. Was ist die W-keit q_k, dass auch beim $k+1$-ten Mal rot gezogen wird?

Lösung von Problem I. Sei $A = \{k \text{ Mal rot}\}$, $B_i = \{\text{Urne } i\}$. Wir haben $P(B_i) = \frac{1}{n}$, $P(A|B_i) = (\frac{i}{n})^k$, also $P(A) = \sum\limits_{i=1}^{n} (\frac{i}{n})^k \frac{1}{n}$. Schauen wir auf die

Kurve $y = x^k$



so sieht man $p_k = P(A) \sim \int_0^1 x^k dx = \frac{1}{k+1}$ für große n.

Lösung von Problem II. Sei $B = \{$die ersten k Kugeln rot$\}$, $A = \{$alle $k+1$ Kugeln rot$\}$. Wir haben

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}$$

wegen $A \subseteq B$, also $q_k = P(A|B) \sim \frac{\frac{1}{k+2}}{\frac{1}{k+1}} = \frac{k+1}{k+2}$.

2. Es wird ein medizinischer Test zu einer Krankheit durchgeführt. Sei $B = \{$Krankheit$\}$, $A = \{$Test positiv$\}$. Der Patient ist interessiert an der W-keit, dass er bei einem positiven Test tatsächlich krank ist, also $P(B|A)$. Angenommen, wir haben die folgenden Daten:

$$P(A|B) = 0{,}95,\ P(A|B^c) = 0{,}10,\ P(B) = 0{,}06\,.$$

Dann ist nach Bayes

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{0{,}95 \cdot 0{,}06}{0{,}95 \cdot 0{,}06 + 0{,}10 \cdot 0{,}94} = 0{,}378\,.$$

3. Ziegenproblem. In einer Fernsehshow sind drei Türen, hinter einer befindet sich ein Auto (das der Kandidat gewinnen möchte), hinter den beiden anderen Ziegen. Der Kandidat zeigt auf Tür 1. Der Moderator (der weiß, wo das Auto ist) öffnet Tür 3, hinter der sich eine Ziege befindet. Der Kandidat hat ein zweite Chance. Soll er wechseln?

Sei $B_i = \{\text{Auto hinter Tür } i\}$, $P(B_i) = \frac{1}{3}$. $A = \{\text{Quizmaster zeigt, dass hinter Tür 3 eine Ziege ist}\}$. Der Kandidat ist also an $P(B_1|A)$ und $P(B_2|A)$ interessiert. Es ist $P(A|B_1) = \frac{1}{2}$, $P(A|B_2) = 1$, $P(A|B_3) = 0$, also $P(A) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$. Mit Bayes erhalten wir

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{1}{3}\,,$$

$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)} = \frac{2}{3}\,.$$

Der Kandidat sollte also wechseln.

Seien nun Ereignisse $A_1, A_2, \ldots, A_n$ gegeben. Mit Induktion beweist man sofort

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \ldots \cap A_{n-1})\,.$$

Definition. Die Ereignisse $A_1, \ldots, A_n$ heißen *unabhängig*, falls $P(A_{i_1} \cap \ldots \cap A_{i_r}) = \prod_{j=1}^{r} P(A_{i_j})$ für alle $I = \{i_1, \ldots, i_r\} \subseteq \{1, \ldots, n\}$. Falls $A_1, \ldots, A_n$ unabhängig sind, so sind sie auch paarweise unabhängig, aber nicht umgekehrt.

Beispiele. 1. Sei $\Omega = \{1, 2, 3, 4\}$ mit Gleichverteilung, $A_1 = \{1, 2\}$, $A_2 = \{2, 3\}$, $A_3 = \{1, 3\}$. Dann gilt $P(A_i \cap A_j) = P(A_i)P(A_j)$ für alle $i \neq j$, aber $P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3)$.
2. Wir werfen eine Münze drei Mal, $\Omega = \{K, Z\}^3$. $A = \{\text{mindestens 2 Mal Kopf}\}$, $B = \{\text{1. Wurf ist Kopf}\}$, $C = \{\text{der 2. und 3. Wurf zeigen dasselbe}\}$. Wir haben $|\Omega| = 8$, $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(A \cap B) = \frac{3}{8}$ (KKK, KKZ, KZK), $P(A \cap C) = \frac{1}{4}$ (KKK, ZKK), $P(B \cap C) = \frac{1}{4}$ (KKK, KZZ), $P(A \cap B \cap C) = \frac{1}{8}$ (KKK). A, B, C sind nicht unabhängig.

Lemma 1.5. *Sind $A_1, \ldots, A_n$ unabhängig, so auch die komplementären Ereignisse $A_1^c, \ldots, A_n^c$.*

Beweis. Sei $A_{i_1}, \ldots, A_{i_r}$ gegeben, wobei wir uns auf $A_1, \ldots, A_r$ beschränken können. Nach Inklusion-Exklusion haben wir

$$P(A_1^c \cap \ldots \cap A_r^c) = P((A_1 \cup \ldots \cup A_r)^c) = \sum_{k=0}^{n} (-1)^k \sum_{1 \le i_1 < \cdots < i_k \le r} P(A_{i_1} \cap \ldots \cap A_{i_k})\,.$$

Andererseits ist

$$P(A_1^c)\cdots P(A_r^c) = \prod_{i=1}^{r}(1-P(A_i)) = \sum_{k=0}^{r}(-1)^k \sum_{1\le i_1<\dots<i_k\le r} P(A_{i_1})\cdots P(A_{i_k})$$

$$= \sum_{k=0}^{r}(-1)^k \sum_{1\le i_1<\dots<i_k\le r} P(A_{i_1}\cap\ldots\cap A_{i_k})$$

wegen der Unabhängigkeit der A_i. $\square$