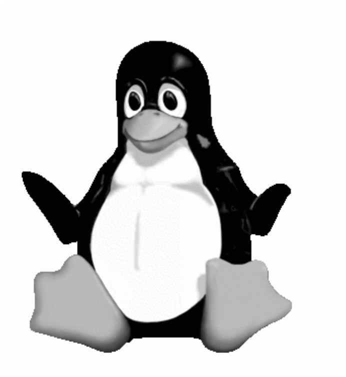
Stochastik I

Dr. Grübel

SS 1999

Stand: 15. Oktober 2001



Gewidmet allen freizeitlosen Mathematikern.

Dieses Dokument ist meine Mitschrift der Vorlesung "Stochastik I" von Dr. Grübel im Sommersemester 1999.
Die Nummerierung der Lemmata, Sätze, etc. habe ich teilweise etwas meinen Normen angepaßt.
Für eventuelle Fehler aller Art übernehme ich keine Verantwortung !
Vielen Dank allen Entwicklern von TEX, X-Fig, GnuPlot und Linux !
Tobias Müller

Inhaltsverzeichnis

1 Grundbegriffe, Kolmogorov-Axiome

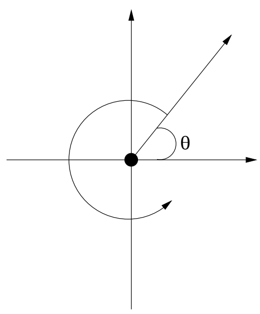
1.1 Einführung

Stochastik (Wahrscheinlichkeitstheorie und mathematische Statistik) beschäftigt sich mit Zufallsexperimenten, bei denen das Ergebnis nicht durch die "Randbedingungen" des Experiments festgelegt ist.
Der Ergebnisraum Ω ist eine Menge, die die möglichen Ergebnisse(Resultate) des Experiments enthält.
Ereignisse werden durch Teilmengen von Ω beschrieben.

Beispiel 1.1.1

1. *Würfelwurf* $\Omega = \{1,2,3,4,5,6\}$, *das Ergebnis "eine gerade Zahl erscheint" ist* $A = \{2,4,6\}$.

2. *Eine Münze wird* n*-mal geworfen. Schreibt man* 0 *für Kopf und* 1 *für Zahl, so ist*
   $\Omega = \{(i_1, \cdots, i_n) : i_j \in \{0,1\} \forall j\} (= \{0,1\}^n)$
   ein geeigneter Ergebnisraum. Hierbei bedeutet $\omega = (i_1, \cdots, i_n)$, *daß im* j*-ten Wurf* i_j *erscheint,* $1 \leq j \leq n$.
   Das Ergebnis "in allen n *Würfen erscheint dieselbe Seite" wird beschrieben durch* $\{(0, \cdots, 0), (1, \cdots, 1)\}$.



$\Theta = 2\pi x$, $0 \leq x < 1$ *Winkel bei Stillstand*
$\Omega := [0,1)$
"Zeiger stoppt im rechten oberen Quadranten" $= (0, \frac{1}{4})$

Ein Ereignis A mit exakt einem Element, also $A = \{\omega\}$, heißt Elementarereignis.
Kombinationen von Ereignissen können durch mengentheoretische Operationen beschrieben werden:

$A \cap B$:	A und B treten beide ein
$A \cup B$:	A tritt ein, oder B tritt ein, oder A und B treten ein
A^c:	A tritt nicht ein

Beispiel 1.1.2 *Kombination von mehr als zwei Ereignissen*

1. *"Genau eines der Ereignisse A,B,C tritt ein"*
   $A \cap B^c \cap C^c + A^c \cap B \cap C^c + A^c \cap B^c \cap C$
   ($A + B$ steht für $A \cup B$ im Falle $A \cap B = \emptyset$)

2. *Es sei $A_1, A_2, \cdots$ eine Folge von Ereignissen.*
   Dann wird das Ereignis "unendlich viele der A_n´s treten ein" durch den Limes superior der Mengenfolge beschrieben:
   $$\boxed{\lim\sup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m}$$
   $\omega \in \lim\sup_{n\to\infty} A_n$
   $\Leftrightarrow \forall n \in \mathbb{N}: \omega \in \bigcup_{m=n}^{\infty} A_m$
   $\Leftrightarrow \forall n \in \mathbb{N} \exists_{m \geq n} : \omega \in A_m$
   $\Leftrightarrow \#\{n \in \mathbb{N}: \omega \in A_n\} = \infty$

Die Menge der Ereignisse *(eine Menge von Mengen !)* bildet ein Mengensystem $\mathfrak{A}(\subset \mathfrak{P}(\Omega)$, Potenzmenge von Ω).
Bei endlichen oder abzählbar unendlichen Ω kann man mit $\mathfrak{A} = \mathfrak{P}(\Omega)$ arbeiten, bei überabzählbaren Ω stößt dies auf Schwierigkeiten (später mehr.)

Definition 1.1.1
$\mathfrak{A} \subset \mathfrak{P}(\Omega)$ *heißt σ-Algebra über Ω, wenn gilt:*

1. $\Omega \in \mathfrak{A}$

2. $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$

3. $A_1, A_2, \cdots \in \mathfrak{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$

(Das System $\mathfrak{A}$ enthält die Ergebnismenge und ist abgeschlossen gegenüber Komplementbildung und abzählbarer Vereinigung.)
Was ist nun "Wahrscheinlichkeit" ? Dies ist streng genommen keine mathematische Frage.
(vergleiche: "Was ist eine Gerade ?" "Was ist eine Menge ?")
Als "mathematischer Gegenstand" ist Wahrscheinlichkeit eine Funktion, die Ereignissen Zahlen zwischen 0 und 1 zuordnet und dabei gewissen "Axiomen" genügt.
Die Axiome werden durch den umgangssprachlichen Wahrscheinlichkeitsbegriff motiviert.
Betrachte zur Erläuterung die Aussage *"das Ereignis A hat Wahrscheinlichkeit p"* (z.B. *"beim Wurf eines Würfels erscheint mit Wahrscheinlichkeit $\frac{1}{2}$ eine gerade Zahl".*)
Es gibt zwei hauptsächliche Interpretationen:

1. Die Häufigkeitsauffassung($\rightsquigarrow$ Frequentisten)
   Es sei $N_n(A)$ die Häufigkeit des Auftretens von A bei n Wiederholungen des Zufallsexperiments, $\frac{1}{n} * N_n(A)$ ist die relative Häufigkeit von A. Bei großem n sollte die relative Häufigkeit von A in der Nähe von p liegen.

2. Die Glaubens- oder Plausibilitätsauffassung($\rightsquigarrow$ Subjektivitisten)
   Der Wert p gibt auf einer Skala von 0 bis 1 die *"Stärke meines Glaubens"* an das Eintreten von A wieder. Diese Auffassung kann über Wetten formalisiert werden.

(2) ist im Gegensatz zu (1) auch bei nicht-wiederholbaren Experimenten anwendbar, aber eben subjektiv.
Diese Auffassungen sind nicht disjunkt.

Für relative Häufigkeiten gelten folgende Regeln:

1. $\frac{1}{n} * N_n(\Omega) = 1$
2. $\frac{1}{n} * N_n(A) \geq 0 \ \forall A \in \mathfrak{A}$
3. $\frac{1}{n} * N_n(A_1 + \cdots + A_k) = \frac{1}{n} * N_n(A_1) + \cdots + \frac{1}{n} * N_n(A_k)$ für alle paarweise disjunkten $A_1, \cdots, A_k \in \mathfrak{A}$

1.2 Das mathematische Modell für Zufallsexperimente

Definition 1.2.1 *Die Kolmogorov-Axiome*
Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathfrak{A}, P)$, bestehend aus einer nichtleeren Menge Ω (dem Ergebnisraum), einer σ-Algebra $\mathfrak{A}$ über Ω (dem Ereignissystem) und einer Abbildung $P : \mathfrak{A} \rightarrow \mathbb{R}$ mit den Eigenschaften:

1. $P(\Omega) = 1$
2. $P(A) \geq 0 \ \forall A \in \mathfrak{A}$
3. $\boxed{P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)}$
   für alle paarweise disjunkten $A_1, A_2, \cdots \in \mathfrak{A}$

Eine Abbildung mit diesen Eigenschaften heißt Wahrscheinlichkeitsmaß .
Eigenschaft (3) nennt man auch σ-Additivität.

$\Omega \neq \emptyset$ Ergebnismenge, $(\Omega, \mathfrak{A}, P)$ Wahrscheinlichkeitsraum
$\mathfrak{A}$, σ-Algebra über Ω, Ereignisalgebra
$P : \mathfrak{A} \rightarrow \mathbb{R}$, Wahrscheinlichkeitsmaß $P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

Beispiel 1.2.1

1. *Ist Ω endlich, so wird durch*
   $P(A) = \frac{\#A}{\#\Omega} \forall A \subset \Omega \qquad P(\{\omega\}) = \frac{1}{N}$*, wenn $N = \#\Omega$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{P}(\Omega))$ definiert.*
   (Hier, bei endlichem oder abzählbar unendlichen Ω, arbeiten wir mit $\mathfrak{A} = \mathfrak{P}(\Omega)$)
   Diese werden häufig durch Symmetrieüberlegungen nahe gelegt, wie im Beispiel auf Seite 6.
   Beim Wurf eines fairen(symmetrischen) Würfels ergibt sich als Wahrscheinlichkeit

dafür, daß eine gerade Zahl geworfen wird

$$P(A) = P(\{2,4,6\}) = \frac{\#\{2,4,6\}}{\#\{1,2,3,4,5,6\}} = \frac{3}{6} = \frac{1}{2}$$

(Anzahl der "günstigen" Fälle, dividiert durch die Anzahl der möglichen Fälle)

2. *Ein Experiment, das deterministisch ist, in dem Sinne, daß nur ein Ereignis ω_0 möglich ist, kann als Zufallsexperiment $(\Omega, \mathfrak{A}, \delta_{\omega_0})$ betrachtet werden:*
   Ω ist hierbei irgendeine Menge, die ω_0 enthält, $\mathfrak{A}$ irgendeine σ-Algebra über Ω, δ_{ω_0} ist das Dirac-Maß in ω_0 (auch: Einpunktmaß):
   $\delta_{\omega_0}(A) = \begin{cases} 1 & ; \quad \omega_0 \in A \\ 0 & ; \quad \omega_0 \notin A \end{cases}$ *(δ_{ω_0} ist ein Wahrscheinlichkeitsmaß)*

   $\delta_{\omega_0}(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \delta_{\omega_0}(A_n)$
   $\omega_0 \in \sum_{n=1}^{\infty} A_n$ *? trivial ja !*

Folgerungen aus den Axiomen

Satz 1.2.1 *Rechenregeln*
Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum.
Dann gilt:

1. $P(\emptyset) = 0$
   $P(A) \leq 1$ *für alle* $A \in \mathfrak{A}$

2. $P(A^c) = 1 - P(A)$ *für alle* $A \in \mathfrak{A}$

3. *(endliche Additivität)*
   $P(A_1 \cup \cdots \cup A_k) = P(A_1) + \cdots + P(A_k)$ *für alle paarweise disjunkten* $A_1, \cdots, A_k \in \mathfrak{A}$

4. *(Monotonie)*
   $A \subset B \Rightarrow P(A) \leq P(B) \ \forall A, B \in \mathfrak{A}$

5. *(Boole'sche Ungleichung)*
   $P(A_1 \cup \cdots \cup A_k) \leq P(A_1) + \cdots + P(A_k) \ \forall A_1, \cdots, A_k \in \mathfrak{A}$
   (die A_i's brauchen hier nicht disjunkt zu sein)
   $\left(P(A_1 \cup A_2) \begin{cases} = & P(A_1) + P(A_2) \quad \textit{wenn } A_1 \cap A_2 = \emptyset \\ \leq & P(A_1) + P(A_2) \quad \textit{sonst} \end{cases} \right)$

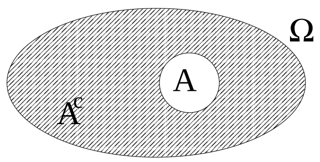
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

7. *(Formel von Poincaré, auch Einschluß-Ausschlußformel, Siebformel,$\cdots$)*
   $\boxed{P(A_1 \cup \cdots \cup A_k) = \sum_{H \subset \{1,\cdots,k\}, H \neq \emptyset} (-1)^{\# H - 1} * P(\bigcap_{i \in H} A_i)}$

Beweis:

1. $\emptyset \in \mathfrak{A}$, da $\Omega \in \mathfrak{A}$, und $\mathfrak{A}$ abgeschlossen gegenüber Komplementbildung ist.
   Verwendet man die σ-Additivität mit $A_1 = A_2 = \cdots = \emptyset$,
   so folgt $P(\emptyset) = P(\emptyset) + P(\emptyset) + \cdots$, also $P(\emptyset) = 0$.
   $P(A) \leq 1$ folgt aus $P(\Omega) = 1$ (Bestandteil der Axiome)

2. $A \cup A^c = \Omega$, $A \cap A^c = \emptyset$, verwende nun die endliche Additivität.



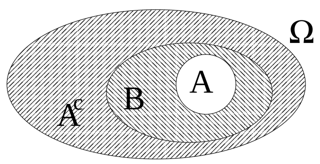
$$(\underbrace{P(A \cup A^c)}_{=P(\Omega)=1} = P(A) + P(A^c))$$

3. Setze $A_{k+1} = A_{k+2} = \cdots = \emptyset$, verwende σ-Additivität und $P(\emptyset) = 0$

4. $B = A + B \cap A^c$, also $P(B) = P(A) + \underbrace{P(B \cap A^c)}_{\geq 0 (Axiom)} \geq P(A)$

5. Im Falle $k = 2$ folgt dies aus (6) und $P(A \cap B) \geq 0$. Ansonsten Induktion.

6. $A = A \cap B + A \cap B^c$, also $\underline{P(A) = P(A \cap B) + P(A \cap B^c)}$



   $A \cup B = B + A \cap B^c)$, also: $\underline{P(A \cup B) = P(B) + P(A \cap B^c)}$
   Insgesamt: $P(A \cup B) = P(B) + P(A) - P(A \cap B)$

7. Im Fall $k = 2$ erhält man (6). Induktionsschritt: Übungsaufgabe.

$\square$

Beispiel 1.2.2 *Einfache Anwendung*
Mit welcher Wahrscheinlichkeit erscheinen beim n-fachen Wurf einer (fairen) Münze Kopf und Zahl beide mindestens einmal ?
Bezeichnet A das Ereignis, so gilt:

$A^c = \{(0, \cdots, 0), (1, \cdots, 1)\}$
man erhält also
$P(A) = 1 - P(A^c) = 1 - \frac{\#A^c}{\#\Omega} = 1 - \frac{2}{2^n} = 1 - \frac{1}{2^{n-1}}$

Warum σ-Additivität ?
Würde endliche Additivität nicht reichen ?
σ-Additivität kann als Stetigkeitseigenschaft aufgefaßt werden:

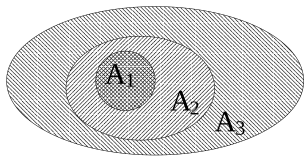
Satz 1.2.2
Sei $\mathfrak{A}$ eine σ-Algebra über $\Omega(\neq \emptyset)$ und $P : \mathfrak{A} \to \mathbb{R}$ eine Abbildung mit:

1. $P(\Omega) = 1$
2. $P(A) \geq 0 \ \forall A \in \mathfrak{A}$
3. $P(A_1 + \cdots + A_k) = P(A_1) + \cdots + P(A_k)$ *für* $A_1, \cdots, A_k \in \mathfrak{A}$ *paarweise disjunkt*

Dann sind äquivalent:

1. *P ist Wahrscheinlichkeitsmaß (also sogar $\underline{\underline{\sigma}}$-additiv)*
2. *P ist stetig von unten, d.h. für jede Folge $A_1, A_2, \cdots$ von Ereignissen, die isoton ist im Sinne von $A_n \subset A_{n+1}$ $\forall n \in \mathbb{N}$, gilt:*
   $lim_{n\to\infty} P(A_n) = P(\bigcup_{n=1}^{\infty} A_n)$
3. *P ist stetig von oben, d.h. für jede Folge $A_1, A_2, \cdots$ von Ereignissen, die antiton ist im Sinne von $A_n \supset A_{n+1}$, $\forall n \in \mathbb{N}$, gilt:*
   $lim_{n\to\infty} P(A_n) = P(\bigcap_{n=1}^{\infty} A_n)$
4. *P ist stetig in $\emptyset$, d.h. $lim_{n\to\infty} P(A_n) = 0$ für jede Folge $(A_n) \subset \mathfrak{A}$ mit $n \in \mathbb{N}$ und $A_n \downarrow \emptyset$*
   (d.h. $A_n \supset A_{n+1}$ $\forall n, \bigcap_{n=1}^{\infty} A_n = \emptyset$)

Beweis:
(1)⇒(2)



Es sei $B_1 = A_1$, $B_n = A_n \cap A_{n-1}^c$ $\forall n > 1$
Klar: $B_n \in \mathfrak{A}$ für alle $n \in \mathbb{N}$, $A_n = B_1 + \cdots + B_n$
Die σ-Additivität liefert:
$P(\bigcup_{n=1}^{\infty} A_n)$
$= P(\bigcup_{n=1}^{\infty} B_n)$
$= P(\sum_{n=1}^{\infty} B_n)$
$= \sum_{n=1}^{\infty} P(B_n)$
$= lim_{n\to\infty} \sum_{m=1}^{n} P(B_n)$
$= lim_{n\to\infty} P(\sum_{m=1}^{n} B_m)$ (endliche Additivität)
$= lim_{n\to\infty} P(A_n)$ (2)→(1):
Wir verwenden die umgekehrte Konstruktion:

Ist (A_n) mit $n \in \mathbb{N}$ eine disjunkte Mengenfolge in $\mathfrak{A}$ und setzt man $B_n := \sum_{m=1}^{n} A_m$, so gilt $B_n \uparrow \bigcup_{n=1}^{\infty} A_n$, also

$\underline{P(\sum_{n=1}^{\infty} A_n)} = P(\bigcup_{n=1}^{\infty} B_n)$

$\overset{(2)}{=} lim_{n\to\infty} P(B_n)$

$= lim_{n\to\infty} P(\sum_{m=1}^{n} A_m)$

$\overset{endliche Additivität}{=} lim_{n\to\infty} \sum_{m=1}^{n} P(A_m) = \underline{\sum_{n=1}^{\infty} P(A_n)}$

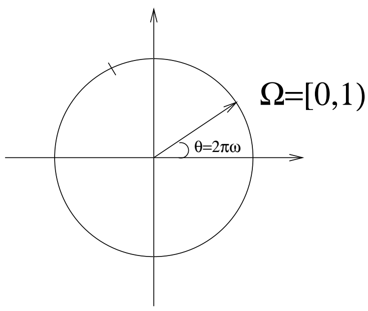
Rest: Übungsaufgabe ! $\square$

Also:

endliche Additivität ist eine natürliche Minimalanforderung.
σ-Additivität ist eine (zusätzliche) Stetigkeitseigenschaft.

Was ist "Hyperadditivität" ?

$P(\bigcup_{i\in I} A_i) = \sum_{i\in I} P(A_i)$ auch für überabzählbare I.



"Glücksrad" $P(\{\omega\}) = 0$ $\emptyset \neq \{\emptyset\}$

$P(\Omega) = P(\bigcup_{\omega\in\Omega}\{\omega\}) \overset{Hyperadditivität}{=} \sum_{\omega\in\Omega} \underbrace{P(\{\omega\})}_{=0} = 0$

Der Zeiger hält also nirgends an !

2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

2.1 Einführung

Es seien A, B Ereignisse in einem Zufallsexperiment, daß durch einen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ beschrieben wird.
Was ist die Wahrscheinlichkeit von B unter der Bedingung, daß A eintritt ?
Bei n Wiederholungen tritt A $N_n(A)$-mal ein, unter diesen ist $N_n(A \cap B)$ die (absolute) Häufigkeit von B.
Für die relative Häufigkeit von B unter den Experimenten, die A liefern, gilt:

$$\frac{N_n(A\cap B)}{N_n(A)} = \frac{\frac{1}{n} * N_n(A\cap B)}{\frac{1}{n} * N_n(A)} \qquad \begin{matrix}(\rightsquigarrow P(A\cap B))\\(\rightsquigarrow P(A))\end{matrix}$$

Dies motiviert

Definition 2.1.1
Es sei A ein Ereignis mit $P(A) > 0$.
Die bedingte Wahrscheinlichkeit eines Ereignisses B unter A wird definiert durch

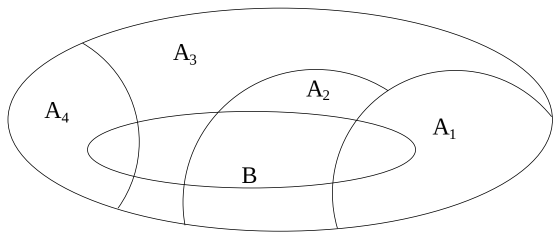
$$\boxed{P(B|A) := \frac{P(A\cap B)}{P(A)}}$$

Man sieht leicht, daß $B \rightarrow P(B|A)$ ein Wahrscheinlichkeitsmaß ist ($A \in \mathfrak{A}$ fest).
$(\Omega, \mathfrak{A}, P(*|A))$ repräsentiert das gegenüber $(\Omega, \mathfrak{A}, P)$ dahingehend veränderte Experiment, daß das Eintreten von A bekannt ist.

Satz 2.1.1

1. **Die Multiplikationsregel**
   Es seien $A_1, \cdots, A_n$ Ereignisse
   mit $P(A_1 \cdots A_n)$ Ereignisse mit $P(A_1 \cap \cdots \cap A_n) > 0$.
   Dann gilt:
   $$P(A_1 \cap \cdots \cap A_k) = P(A_1) * P(A_2|A_1) * P(A_3|A_1 \cap A_2) * \cdots * P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

2. **Das Gesetz von der totalen Wahrscheinlichkeit**
   Es sei $A_1, \cdots, A_n$ eine Ereignispartition von Ω,
   d.h. $A_1, \cdots, A_n \in \mathfrak{A}$, $\bigcup_{i=1}^n A_i = \Omega$, $A_i \cap A_j = \emptyset$ für $i \neq j$.



   Dann gilt für alle $B \in \mathfrak{A}$:
   $$P(B) = \sum_{i=1}^n P(B|A_i) * P(A_i)$$
   (Wir lassen $P(A_i) = 0$ zu und setzen dann $P(B|A_i) * P(A_i) = 0$)

3. **Die Formel von Bayes**
Es seien $A_1, \cdots, A_n, B$ wie in (2) und es gelte $P(B) > 0$.
Dann folgt: $P(B|A_i)$

$$\boxed{P(A_i|B) = \frac{P(B|A_i)*P(A_i)}{\sum_{k=1}^n P(B|A_k)*P(A_k)}}$$

<u>Beweis:</u>

1. Einsetzen der Definition
2. Einsetzen liefert:
$$\sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B \cap A_i) = P(\bigcup_{i=1}^n (B \cap A_i)) = P(B \cap \underbrace{\bigcup_{i=1}^n A_i}_{\Omega}) = P(B)$$
3. $\surd$

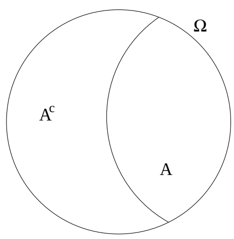
$\square$

Beispiel 2.1.1 *Ein bestimmter medizinischer Test ist zu 95% effektiv im Erkennen einer bestimmten Krankheit, liefert allerdings bei 1% der gesunden Patienten einen "falschen Alarm".*
Angenommen 0,5% der Bevölkerung leiden unter dieser Krankheit.
Mit welcher Wahrscheinlichkeit hat jemand die Krankheit, wenn der Test dies behauptet ?
Sei:

A: die getestete Person hat die Krankheit
B: der Test zeigt das Vorliegen der Krankheit an

Wir setzen die obigen Annahmen ein:

$P(A) = 0,005$
$P(B|A) = 0,95$
$P(B|A^c) = 0,01$



Bayes liefert nun:
$$P(A|B) = \frac{P(B|A)*P(A)}{P(B|A)*P(A)+P(B|A^c)*P(A^c)} = \frac{0{,}95*0{,}005}{0{,}95*0{,}005+0{,}01*(1-0{,}005)} \approx 0,323$$

<u>**Beachte:**</u> Die Übersetzung von Prozentzahlen in Wahrscheinlichkeiten liegt bestimmten Annahmen über die Auswahl der Testpersonen, etc. zugrunde.
Obiges Beispiel zeigt auch, daß es nicht immer nötig bzw. sinnvoll ist, $(\Omega, \mathfrak{A}, P)$ explizit anzugeben.

Einer der zentralen Begriffe der Stochastik ist der der (stochastischen) Unabhängigkeit.
Die mathematische Definition sollte das "intuitive Konzept" wiedergeben:
B wird von A nicht beeinflußt, wenn sich die Wahrscheinlichkeit von B nicht durch die Information ändert, daß A eingegeben ist.
Dies führt auf die Forderung $P(B|A) = P(B)$ $\qquad (\Leftrightarrow \frac{P(A\cap B)}{P(A)} = P(B))$

Definition 2.1.2 *Zwei Ereignisse A und B heißen <u>stochastisch Unabhängig</u>, wenn gilt:*

$$\boxed{P(A \cap B) = P(A) * P(B)}$$

Vorsicht bei mehr als 2 Ereignissen:

Definition 2.1.3 *Eine Familie $\{A_i : i \in I\}$ von Ereignissen heißt <u>paarweise Unabhängig</u>, wenn*

$$P(A_i \cap A_j) = P(A_i) * P(A_j) \ \forall i, j \in I, i \neq j$$

Sie heißt <u>unabhängig</u>, wenn für jede endliche Teilmenge H von I gilt:

$$\boxed{P(\bigcap_{i\in H} A_i) = \prod_{i\in H} P(A_i)}$$

unabhängig $\overset{\Rightarrow}{\not\Leftarrow}$ *paarweise unabhängig*

Beispiel 2.1.2 *Eine (faire) Münze wird zweimal geworfen (siehe auch Beispiel 1.1.1.2 auf Seite 6).*
Wir haben ein Laplace-Experiment über $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$
Es sei

$A_1 := \{(0,0), (0,1)\}$ *("Kopf im 1. Wurf")*
$A_2 := \{(0,0), (1,0)\}$ *("Kopf im 2. Wurf")*
$A_3 := \{(0,1), (1,0)\}$ *("Resultate verschieden")*

Dann:

$$P(A_1 \cap A_2) = \frac{\#A_1 \cap A_2}{\#\Omega} = \frac{1}{4} = P(A_1) * P(A_2)$$

analog:

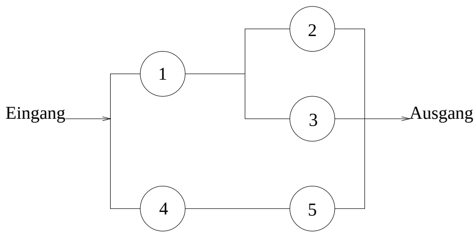
$$P(A_1 \cap A_3) = P(A_1) * P(A_3)$$
$$P(A_2 \cap A_3) = P(A_2) * P(A_3)$$

Die Familie $\{A_1, A_2, A_3\}$ ist also <u>paarweise</u> unabhängig.
Aber:

$$P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0 \neq \frac{1}{8} = P(A_1) * P(A_2) * P(A_3)$$

Die Ereignisse sind also nicht unabhängig !

Beispiel 2.1.3 *Funktion von Netzwerken*
Typische Fragestellung der angewandten Wahrscheinlichkeitsrechnung:
Funktion von Netzwerken
Wir betrachten einen einfachen Fall: Ein System bestehend aus 5 Elementen



Wir nehmen an, daß die Komponenten $\underline{\underline{\textit{unabhängig}}}$ *von einander mit Wahrscheinlichkeit p funktionieren.*
Das Gesamtnetzwerk funktioniert, wenn es einen Pfad funktionierender Komponenten vom Eingang zum Ausgang gibt.
Mit welcher Wahrscheinlichkeit funktioniert das Gesamtsystem ?
Sei A_i *das Ereignis, daß Komponente i funktioniert, sei B das interessierende Ereignis.*
Dann gilt $B = B_1 \cup B_2$ *mit:*

$B_1 := A_4 \cap A_5$ *(unterer Pfad passierbar)*

$B_2 := A_1 \cap (A_2 \cup A_3)$ *(oberer Pfad passierbar)*

Dann ist

$$P(B_1) = P(A_4 \cap A_5) = P(A_4) * P(A_5) = p^2$$

$$P(B_2) = P((A_1 \cap A_2) \cup (A_1 \cap A_3)) = P(A_1 \cap A_2) + P(A_1 \cap A_3) - P(\underline{\underline{\cancel{A_1}}} \cap A_2 \cap \underline{\underline{A_1}} \cap A_3)$$
$$= p^2 + p^2 - p^3 \text{ (unabhängig)}$$

$$P(B_1 \cap B_2) = P(A_4 \cap A_5 \cap A_1 \cap A_2) + P(A_4 \cap A_5 \cap A_1 \cap A_3) - P(A_4 \cap A_5 \cap A_1 \cap A_2 \cap A_3)$$
$$= 2 * p^4 - p^5$$

Insgesamt:

$$P(B) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = p^2 + 2 * p^2 - p^3 - 2 * p^4 + p^5$$
$$= p^2 * (3 - p - 2 * p^2 + p^3)$$

3 Laplace-Experimente

3.1 Allgemeines

$(\Omega, \mathfrak{A}, P) : \Omega \neq \emptyset$, endlich, $\mathfrak{A} = \mathfrak{P}(\Omega), P(A) = \frac{\#A}{\#\Omega}$
Laplace-Experiment; wird durch Ω festgelegt.
Taucht auf bei:

- Werfen eines symmetrischen Gegenstandes (faire Münze, Würfel, alle Seiten haben dieselbe Wahrscheinlichkeit oben zu landen)
- Mischen von Karten oder allgemeiner: Herstellen einer zufälligen Reihenfolge (alle Permutationen gleiche Wahrscheinlichkeit)
- Entnahme einer Stichprobe festen Umfangs aus einer Grundgesamtheit (alle Teilmengen dieses Umfangs haben dieselbe Wahrscheinlichkeit)

<u>Wir betrachten nun “Kopplungen” von solchen Experimenten:</u>
Sind $(\Omega_i, \mathfrak{A}_i, P_i)$, $1 \leq i \leq n$, Laplace-Experimente, so kann die Ausführung aller dieser Experimente als neues Zufallsexperiment betrachtet werden.
Ein geeigneter Ergebnisraum ist

$$\Omega = \times_{i=1}^{n} \Omega_i = \Omega_1 \times \cdots \times \Omega_n$$
$$= \{\omega = (\omega_1, \cdots, \omega_n) : \omega_i \in \Omega_i \text{ für } i = 1, \cdots, n\}$$

Dies ist eine endliche Menge, wähle also $\mathfrak{A} = \mathfrak{P}(\Omega)$.
Wie sieht ein geeignetes Wahrscheinlichkeitsmaß P hierauf aus ?
Wenn die Experimente unabhängig voneinander ausgeführt werden, dann sollten für alle $\omega = (\omega_1, \cdots, \omega_n) \in \Omega$ die Ereignisse

A_i : im i-ten Teilexperiment erscheint ω_i, $1 \leq i \leq n$,

im Sinne von §2 stochastisch unabhängig sein. Wegen

$A_i = \Omega_1 \times \cdots \times \Omega_{i-1} \times \{\omega_i\} \times \Omega_{i+1} \times \cdots \times \Omega_n$ führt dies auf

$$\underline{\underline{P(\{\omega\})}} = P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i) = \prod_{i=1}^{n} P_i(\{\omega_i\}) = \prod_{i=1}^{n} \frac{1}{\#\Omega_i} = \underline{\underline{\frac{1}{\#\Omega}}}$$

$$\{\omega_1\} \times \{\omega_2\} \times \cdots \times \{\omega_n\} = \{w\}$$

<u>Also:</u>
Alle Elementarereignisse haben dieselbe Wahrscheinlichkeit.
Die <u>unabhänige Kopplung</u> (auch: das <u>Produkt</u>) von Laplace-Experimenten ist wieder ein Laplace-Experiment.
Bei dieser Konstruktion sind die Ereignisse, die sich auf verschiedene Teilexperimente beziehen, stochastisch unabhängig im Sinne von Definition 2.1.2 (siehe Seite 15).

Beispiel 3.1.1
Zwei Münzen werden gleichzeitig geworfen. Als Model hierfür würde das Laplace-Experiment über $\Omega = \{\omega_0, \omega_1, \omega_2\}$ *mit:*

ω_0*: beide Kopf*
ω_1*: beide Zahl*
ω_2*: 1*Kopf, 1*Zahl*

auf $P(\textit{Ergebnisse verschieden}) = \frac{1}{3}$ *führen.*

nicht gleichzeitig:
$$\Omega = \{(0,0),\underline{(0,1)},\underline{(1,0)},(1,1)\}$$
$$\#\Omega = 4,\ P(\cdots) = \tfrac{2}{4} = \tfrac{1}{2}$$

Dies widerspricht der Erfahrung.
Das "korrekte" Modell führt auf das Laplace-Experiment über $\Omega = \{0,1\}^2$ *und die Wahrscheinlichkeit* $\frac{1}{2}$.

Wichtig:
Die Bestimmung des korrekten Modells ist kein rein mathematisches Problem, es muß sich aus der "Außenwelt" ergeben. Wir gehen davon aus, daß die Münzen unterscheidbar sind. Die Elementarteilchenphysik liefert Beispiele mit "ununterscheidbaren Objekten"

3.2 Etwas Kombinatorik, oder: die Kunst des Zählens

$\#A = ?$
$\#\Omega = ?$

Zwei elementare Regeln:

1. $\exists \phi : A \to B$ bijektiv $\Rightarrow \#A = \#B$
2. $A \cap B = \emptyset \Rightarrow \#(A \cup B) = \#A + \#B$

Die Verallgemeinerung auf mehr als zwei Mengen führt auf die folgende Aussage:

Ist für jedes $x \in A$ B_x eine Menge mit n Elementen, so gilt:
$$\#\{(x,y) : x \in A, y \in B_x\} = n * \#A$$
Spezialfall: $\#(A \times B) = (\#A) * (\#B)$ $\qquad (A \times B = \bigcup_{x \in A} \underbrace{\{x\} \times B}_{\#\cdots = \#B})$

Wir schreiben abkürzend $M_n = \{1, \cdots, n\}$. Regel (2) liefert:
$$\underline{\underline{\#M_n^k}} = \#\{(i_1, \cdots, i_k) : 1 \le i_j \le n \text{ für } j = 1, \cdots, k\} = \underline{\underline{n^k}}$$
Die Elemente von M_n^k werden auch <u>k-Permutationen von M_n mit Wiederholung</u> genannt.

Interpretationen:

1. Einer Menge von n Elementen kann man n^k Stichproben vom Umfang k <u>mit Zurücklegen bei Berücksichtigung der Reihenfolge</u> entnehmen.
   Das Element $(i_1, \cdots, i_k)$ steht hierbei für die Stichprobe, bei der im l-ten Zug das Element mit Nummer i_l erscheint.

2. Es gibt n^k Möglichkeiten, k Objekte auf n Plätze bei möglicher Mehrfachsetzung und Berücksichtigung der Reihenfolge zu verteilen. Hierbei steht $(i_1, \ldots, i_k)$ für die Verteilung, bei der das Objekt mit Nummer l auf dem Platz mit der Nummer i_l gelegt wurde.

Satz 3.2.1 *Für $1 \leq k \leq n$ gilt:*
$$\#\{(i_1, \cdots, i_k) \in M_n^k : i_j \neq i_l \text{ für } j \neq l\} = \frac{n!}{(n-k)!}$$

Beweis:

Es gibt n Möglichkeiten für i_1, bei gegebenem i_1 bleiben $(n-1)$ Möglichkeiten für i_2, bei gegebenem (i_1, i_2) bleiben $(n-2)$ Möglichkeiten für i_3 etc., d.h. man hat nach Regel (2)
$$n * (n-1) * (n-2) * \cdots * (n-k+1) = \frac{n!}{(n-k)!}$$
Möglichkeiten. □

Als wichtiger Spezialfall ergibt sich mit $k = n$:

Es gibt genau $n!$ Permutationen einer Menge mit n Elementen

Die Elemente der Menge aus Satz 3.2.1 (auf Seite 19) heißen auch
k-Permutationen von M_n ohne Wiederholung.

Interpretationen:

1. Einer Menge von n Elementen kann man $\frac{n!}{(n-k)!}$ verschiedene Stichproben vom Umfang k ohne Zurücklegen bei Berücksichtigung der Reihenfolge entnehmen.
2. Es gibt $\frac{n!}{(n-k)!}$ verschiedene Möglichkeiten, k Objekte auf n Plätze ohne Mehrfachbesetzung bei Berücksichtigung der Reihenfolge zu verteilen.

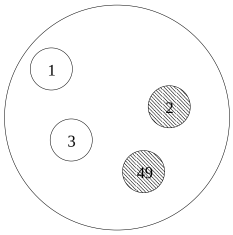
Satz 3.2.2 *Für $1 \leq k \leq n$ gilt:*
$$\#\{(i_1, \cdots, i_k) \in M_n^k : i_1 < i_2 < \cdots < i_k\} = \binom{n}{k}$$

Beweis:

Zu jedem Element dieser Menge gehören genau $k!$ Elemente der Menge aus Satz 3.2.1 (auf Seite 19), nämlich alle die k-Tupel, die durch Permutation der Koordinaten aus dem geordneten Tupel hervorgehen:
$$(\#\{\cdots\}) * k! = \frac{n!}{(n-k)!}$$ □

Die Elemente der Menge aus Satz 3.2.2 (auf Seite 19) heißen
k-Kombinationen von M_n ohne Wiederholung.

Wichtige Anwendung:
Eine Menge mit n Elementen hat $\binom{n}{k}$ Teilmengen mit genau k Elementen.



$$[\sum_{k=0}^{n}\binom{n}{k} = 2^n \qquad (a+b)^n = \sum_{k=0}^{n}\binom{n}{k} a^k * b^{n-k}]$$

Interpretation:

1. Es gibt $\binom{n}{k}$ Möglichkeiten aus n verschiedenen Objekten k herauszugreifen (ohne Berücksichtigung der Reihenfolge, ohne Zurücklegen)
2. Es gibt $\binom{n}{k}$ Möglichkeiten k Objekte auf n Plätze ohne Mehrfachbesetzung zu verteilen, wenn die Verteilungsreihenfolge nicht berücksichtigt wird.

Kombinatorische Grundformeln

	Permutationen (Reihenfolge wird berücksichtigt)	**Kombinationen** (Reihenfolge wird nicht berücksichtigt
ohne Wiederholung	$\#\{(i_1,\cdots,i_k)\in M_n^k : i_j \neq i_l \text{ für } j \neq l\}$ $= \frac{n!}{(n-k)!}$	$\#\{(i_1,\cdots,i_k)\in M_n^k : i_1 < \cdots < i_k\}$ $= \binom{n}{k}$ *(1,3,7)*
mit Wiederholung	$\# M_n^k$ $= n^k$	$\#\{(i_1,\cdots,i_k)\in M_n^k : i_1 \leq i_2 \leq \cdots \leq i_k\}$ $= \binom{n+k-1}{k}$ *(1,1,7)*

Satz 3.2.3 $\#\{(i_1,\cdots,i_k)\in M_n^k : i_1 \leq i_2 \leq \cdots \leq i_k\} = \binom{n+k-1}{k}$

Beweis:

Wir definieren eine bijektive Abbildung
$\phi : \{(i_1,\cdots,i_k)\in M_n^k : i_1 \leq \cdots \leq i_k\} \to \{(i_1,\cdots,i_k)\in M_{n+k-1}^k : i_1 < \cdots < i_k\}$
durch

$\phi((i_1,\cdots,i_k)) = (i_1, i_2 + 1.i_3 + 2, \cdots, i_k + k - 1)$

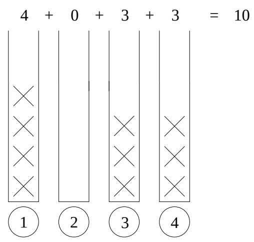
Verwende nun Regel 1 und Satz 3.3.2 (Seite 19). □

Die Elemente der obigen Menge heißen auch k-Kobinationen von M_n mit Wiederholung.

Interpretationen:

1. Einer Menge von n Elementen kann man $\binom{n+k-1}{k}$ verschiedene Stichproben vom Umfang k entnehmen, wenn zurückgelegt wird und die Ziehungsreihenfolge nicht berücksichtigt wird.
2. Es gibt $\binom{n+k-1}{k}$ Möglichkeiten, k Objekte mit möglicher Mehrfachbesetzung auf n Plätze zu verteilen, wenn die Verteilungsreihenfolge nicht berücksichtigt wird.

Bild für $n = 10$, $k = 4$:



Anwendung:
Es gibt $\binom{n+k-1}{k}$ Möglichkeiten, die natürliche Zahl k als Summe von n nicht-negativen Zahlen zu schreiben:

$$\#\{(i_1, \cdots, i_n) \in \mathbb{N}_0^n : i_1 + i_2 + \cdots + i_n = k\} = \binom{n+k-1}{k}$$

Hierbei ist i_l die Anzahl der Objekte auf Platz l, $1 \leq l \leq n$; ein leeres Fach beispielsweise entspricht einem Summanden 0.

3.3 Einige typische Probleme

3.3.1 Geburtstagsproblem

In einem Raum befinden sich n Personen, $n \leq 365$.
Mit welcher Wahrscheinlichkeit haben mindestens zwei am selben Tag Geburtstag ?
Vereinfachende Annahmen:

- kein 29. Februar
- Zwillinge werden vernachlässigt *gewagte These, anm. des Autors*
- saisonale Schwankungen der Geburtenrate werden vernachlässigt

Dann ist ein Laplace-Experiment über

$$\Omega = \{(i_1, \cdots, i_n) : 1 \leq i_1, \cdots, i_n \leq 365\} \qquad (= \{1, \cdots, 365\}^n)$$

plausibel, wobei $i_j = k$ bedeutet, daß Person j am Tag k Geburtstag hat.
Es geht um

$$A = \{(i_1, \cdots, i_n) \in \Omega : \exists\, l \neq j : i_l = i_j\}$$

Man hat

$$A^c = \{(i_1, \cdots, i_n) \in \Omega : i_l \neq i_j \text{ für } l \neq j\}$$

Also gemäß des vorigen Paragraphen

$$\boxed{P(A) =} \ 1 - P(A^c) = 1 - \frac{\# A^c}{\Omega} \ \boxed{= 1 - \frac{365!}{(365-n)! * 365^n}}$$

$$= 1 - \frac{365}{365} * \frac{364}{365} * \cdots * \frac{365-n+1}{365}$$

Dies ist eine in n steigende Folge, ab $n = 23$ ist dies $\geq 0,5$, bei $n = 50$ hat man bereits

$$P(A) \approx 0,97$$

3.3.2 Bridge

Beim Kartenspiel "Bridge" werden 52 Karten (die üblichen) an 4 Spieler (Nord, Ost, Süd und West) gleichmässig ausgeteilt.
Wir wollen die Wahrscheinlichkeit der Ereignisse

A: einer der Spieler erhält alle 4 Asse
B: jeder Spieler erhält ein As

bestimmen.
Das Mischen der Karten liefert eine zufällige Reihenfolge, also ein Laplace-Experiment über

$$\Omega' = \{(\omega_1, \cdots, \omega_{52}) \in \{1, \cdots, 52\}^{52} : \omega_i \neq \omega_j \text{ für } i = j\}$$

Ω' ist die Menge der Permutationen von $\{1, \cdots, 52\}$.
A und B hängen nicht von der Reihenfolge ab, in der die Karten bei den Spielern ankommen, man könnte also auch mit

$$\Omega := \{(D_1, D_2, D_3, D_4) : D_i \subset \{1, \cdots, 52\}, \# D_i = 13, \text{ paarweise disjunkt}\}$$

rechnen.
Hierbei ist D_i die Menge der Karten von Spieler i, $i = 1, \cdots, 4$.
Die Austeilreihenfolge definiert eine Abbildung von Ω' nach Ω, die jeweils $(13!)^4$ Elemente auf ein Element von Ω abbilden, d.h. wir haben wieder ein Laplace-Experiment über

Ω. (Bei Beispiel 3.1.1 auf Seite 17 war dies anders.)

Dies liefert auch
$$\#\Omega = \frac{\#\Omega'}{(13!)^4} = \frac{52!}{(13!)^4}$$
Alternativ:
$$\#\Omega = (\text{Möglichkeiten für } D_1) * \cdots * (\text{Möglichkeiten für } D_4)$$
also
$$\#\Omega = \binom{52}{13} * \binom{39}{13} * \binom{26}{13} * \binom{13}{13} \qquad \binom{13}{13} = 1$$
Es sei A_i das Ereignis, daß Spieler i alle 4 Asse erhält (oBdA: diese haben die Nummern 1 bis 4).
Dann:
$$A_1 = \{(D_1, \cdots, D_4) \in \Omega : \{1,2,3,4\} \subset D_1\}$$
Für $D_1' := D_1 \cap \{1,2,3,4\}^c$ bleiben $\binom{48}{9}$ Möglichkeiten (9 Karten aus der Menge der 48 Nicht-Asse)
Die Anzahl bei D_2, D_3, D_4 bleibt unverändert, also
$$P(A_1) = \frac{\#A_1}{\#\Omega} = \frac{\binom{48}{9}*\binom{39}{13}*\binom{26}{13}}{\binom{52}{13}*\binom{39}{13}*\binom{26}{13}} = \frac{\binom{48}{9}}{\binom{52}{13}} = \frac{13*12*11*10}{52*51*50*49}$$
Dieselben Argumente funktionieren bei A_2, A_3, A_4 und liefern dasselbe Ergebnis (auch aus Symmetriegründen klar), also:
$$P(A) = P(A_1) + P(A_2) + P(A_3) + P(A_4) \qquad \textit{(da die } A_i\textit{'s paarweise disjunkt sind)}$$
$$= 4 * \frac{13*12*11*10}{52*51*50*49} \approx 0,01056$$
(in einem von 100 Spielen hat ein Spieler alle 4 Asse)

Bei der Behandlung von B kann man ganz analog verfahren, oder das Ganze etwas abkürzen. (B: jeder Spieler hat (genau) ein As)
Es gibt 4! Möglichkeiten die 4 Asse so an die 4 Spieler zu verteilen, daß jeder Spieler genau ein As bekommt.
Sind die Asse verteilt, so bleiben
$$\binom{48}{12} * \binom{36}{12} * \binom{24}{12} * \binom{12}{12} = \frac{48!}{(12!)^4} \qquad \left(\binom{12}{12} = 1\right)$$
Möglichkeiten für die übrigen Karten. Also
$$P(B) = \frac{\#N}{\#\Omega} = \frac{4! * \frac{48!}{(12!)^4}}{\frac{52!}{(13!)^4}} = \frac{4! * (13)^4}{52*51*50*49} \approx 0,1055$$
(in ungefähr einem von 10 Spielen sind die Asse gleichmäßig verteilt.)

3.3.3 Postbote

Ein Postbote verteilt n Briefe zufällig auf n Briefkästen (einen pro Kasten); zu jedem Kasten gehört genau ein Brief.
Wieviele Personen erhalten den für sie bestimmten Brief ?
Wir nummerieren die Briefe und Briefkästen so, daß Brief i in Kasten i gehört, $1 \leq i \leq n$.
Die möglichen Austeilungen entsprechen den Permutationen von 1 bis n. "Zufällig" soll heißen, daß ein Laplace-Experiment über
$$\Omega_n := \{(\omega_1, \cdots, \omega_n) : \omega_i \in \{1, \cdots, n\} \forall i, \omega_i \neq \omega_j \text{ für } i \neq j\}$$
Sei $A_n := \{\omega \in \Omega_n : \omega_i \neq i \text{ für } i = 1, \cdots, n\}$ die Menge der fixpunktfreien Permutationen, $B_{ni} := \{\omega \in \Omega_n : \omega_i = i\}$. Dann $A_n^c = \bigcup_{i=1}^n B_{ni}$, also

folgt mit Satz 1.2.1(7) von Seite 9:

$$P(A_n) = 1 - P(A_n^c) = 1 - \sum_{\substack{J\subset\{1,\cdots,n\}\\ J\neq\emptyset}} (-1)^{\#J-1} P(\bigcap_{i\in J} B_{ni})$$

Klar: $\bigcap_{i\in J} B_{ni} = \{\omega \in \Omega_n : \omega_i = i \text{ für alle } i \in J\}$

Für ein ω aus dem Durchschnitt sind $\#J$ Positionen festgelegt, die übrigen $n - \#J$ können beliebig permutiert werden, also $\#\bigcap_{i\in J} B_{ni} = (n - \#J)!$

Schließlich ist die Anzahl aller $J \subset \{1, \cdots, n\}$ mit k Elementen gleich $\binom{n}{k}$

also

$$\begin{aligned} P(A_n) &= 1 - \sum_{J\subset\{1,\cdots,n\},\#J\neq 0} (-1)^{\#J-1} * \frac{(n-\#J)!}{n!} \\ &= 1 - \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} * \frac{(n-k)!}{n!} \\ &= 1 + \sum_{k=1}^{n} \frac{(-1)^k}{k!} \\ &= \underline{\sum_{k=0}^{n} \frac{(-1)^k}{k!}} \end{aligned}$$

Was passiert mit $n \to \infty$?

Aus der Analysis bekannt:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

also: die Wahrscheinlichkeit definiert, daß das Ereignis, daß keiner der n Briefe beim korrekten Empfänger landet, mit $n \to \infty$ gegen

$$\boxed{\frac{1}{e} \approx 0,3679}$$

geht.

Da eine alternierende Reihe vorliegt, hat man sogar eine Fehlerabschätzung:

$$\boxed{|P_n(A_n) - e^{-1}| \leq \frac{1}{(n+1)!}}$$

Gleichzeitig haben wir eine Aussage bewiesen, die nicht auf Wahrscheinlichkeit bezug nimmt:

Die Anzahl der fixpunktfreien Permutationen einer Menge von n Elementen ist

$$\boxed{n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}}$$

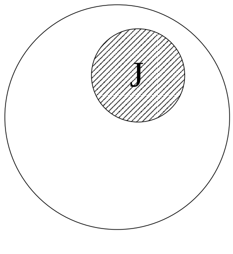
Sei nun

$$X_n : \Omega_n \to \mathbb{N}_0,$$
$$X_n(\omega) := \#\{1 \leq i \leq n : \omega_i = i\}$$

die Anzahl der Fixpunkte *(richtig zugestellte Briefe)*

Es gilt

$$X_n(\omega) = k \Leftrightarrow \exists J \subset \{1, \cdots, n\}, \#J = k : n_i \begin{cases} = i & , \quad i \in J \\ \neq i & , \quad i \notin J \end{cases}$$



Die Anzahl aller $\omega \in \Omega_n$ mit $X_n(\omega) = k$ ist das Produkt der Anzahl der Möglichkeiten für J und der jeweiligen Anzahl der fixpunktfreien Permutationen der Menge $\{1, \cdots, n\} - J$ (Regel (2) aus §3.2 von Seite 18):

$$\begin{aligned} P_n(X_n = k) &= \tfrac{\#\{\omega\in\Omega_n : X_n(\omega)=k\}}{\#\Omega_n} \\ &= \underbrace{\frac{1}{n!}}_{\Omega_n} * \underbrace{\binom{n}{k}}_{J} * (n-k)! * \textstyle\sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \\ &= \tfrac{1}{k!} * \textstyle\sum_{j=0}^{n-k} \frac{(-1)^j}{j!},\ 0 \le k \le n \end{aligned}$$

Im Limes erhält man

$$\boxed{lim_{n\to\infty} P_n(X_n = k) = \tfrac{1}{k!} * e^{-1}\ \forall k \in \mathbb{N}_0}$$

4 Diskrete Wahrscheinlichkeitsräume und Zufallsgrößen

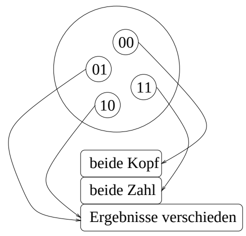
4.1 Allgemeines

Wir nennen $(\Omega, \mathfrak{A}, P)$ einen diskreten Wahrscheinlichkeitsraum, wenn Ω endlich oder abzählbar unendlich ist und $\mathfrak{A} = \mathfrak{P}(\Omega)$ gilt. Aufgrund der σ-Additivität ist P dann durch die zugehörige Wahrscheinlichkeitsmassenfunktion p, $p : \Omega \to \mathbb{R}$, $p(\omega) = P(\{\omega\})$ eindeutig festgelegt:

$$P(A) = \sum_{\omega \in A} p(\omega) \text{ für alle } A \in \mathfrak{A} \ (A \subset \Omega)$$

$(A = \sum_{\omega \in \Omega} \{\omega\})$
$(P(A) = P(\sum) = \sum(P))$

Bei Laplace-Experimenten: $p(\omega) \equiv \frac{1}{\#\Omega}$
Auch hier lassen sich Produkte bilden: das Produkt von endlich vielen diskreten Wahrscheinlichkeitsräumen ist wieder ein diskreter Wahrscheinlichkeitsraum. *Münze zweimal werfen*



Kein Laplace mehr !
Wahrscheinlichkeitsmassenfunktion nicht konstant !

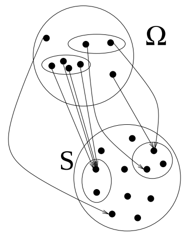
Definition 4.1.1
Es sei $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitsraum, S eine nichtleere Menge. Dann heißt eine Abbildung $X : \Omega \to S$ eine (S-wertige) diskrete Zufallsgröße,
im Falle $S = \mathbb{R}$ Zufallsvariable (ZV)
Bei $S = \mathbb{R}^d$: Zufallsvektor:
Mit ω ist i.a. auch $X(\omega)$ zufällig.

Es wird bei der Behandlung von Zufallsgrößen also nicht darum gehen können, welchen Wert X annimmt, sondern darum, mit welcher Wahrscheinlichkeit X in einer Teilmenge von S liegt.

Satz 4.1.1 *(und Definition)*
Es sei $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to S$ eine diskrete Zufallsgröße.
Dann wird durch

$$\boxed{\begin{array}{lll} P^X := \mathfrak{P}(S) & \to & \mathbb{R} \\ \quad P^X(A) & := & P(X^{-1}(A)) \text{ für alle } A \subset S \end{array}}$$

mit $X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}$ ein Wahrscheinlichkeitsmaß definiert.
P^X heißt die Verteilung von X, alternative Schreibweise: $\mathcal{L}(X)$ ("Law" of X).
Wir schreiben auch $P(X \in A)$ für $P(X^{-1}(A))$.



Beweis:

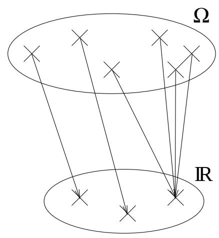
1. $P^X(S) = P(\{\omega \in \Omega : X(\omega) \in S\}) = P(\Omega) = 1$ $\surd$

2. Sind $A_1, A_2, \cdots \subset S$ paarweise disjunkt, so sind auch die Mengen $B_1, B_2, \cdots \subset \Omega$ mit $B_i := \{\omega \in \Omega : X(\omega) \in A_i\}$ $(= X^{-1}(A_i))$ paarweise disjunkt, und es folgt:
$$\underline{P^X(\sum_{i=1}^{\infty} A_i)}$$
$$= P(X^{-1}(\sum_{i=1}^{\infty} A_i))$$
$$= P(\sum_{i=1}^{\infty} \underbrace{X^{-1}(A_i)}_{B_i})$$
$$\overset{\sigma-Additivit\ddot{a}t\ von\ P}{=} \sum_{i=1}^{\infty} P(X^{-1}(A_i))$$
$$= \underline{\sum_{i=1}^{\infty} P^X(A_i)}$$
Also ist P^X σ-additiv $\square$

Beispiel 4.1.1 *Münzwurf*
Wie oft erscheint "Zahl" beim 5 maligen Wurf einer fairen Münze ?
Das Ausgangsexperiment ist ein Laplace-Experiment über $\Omega = \{0,1\}^5$ *(0: Kopf, 1:Zahl).*
Die Anzahl der Zahlwürfe ist
$$X(\omega) = \omega_1 + \omega_2 + \cdots + \omega_5 \ \forall \omega = (\omega_1, \cdots, \omega_5) \in \Omega$$
Als Bildbereich kommt beispielsweise
$$S = \{0, 1, \cdots, 5\}$$
in Frage.
Als Wahrscheinlichkeitsmaß auf eine endliche Menge wird $\mathcal{L}(X)$ *wieder durch die zugehörige Massenfunktion beschrieben, wir benötigen also die Werte*
$$P(X = k) = P(\{\omega \in \Omega : X(\omega) = k\}) = P(X^{-1}(\{k\}) \textit{ für } k = 0, 1, \cdots, 5$$
$$P(\{\omega \in \Omega : X(\omega) = k\}) = \frac{\#\{\omega \in \Omega : X(\omega) = k\}}{\#\Omega}$$
$$= \frac{\#\{(\omega_1, \cdots, \omega_5) \in \{0,1\}^5 : \sum_{i=1}^{5} \omega_i = k\}}{2^5}$$
$$= \binom{5}{k} * \frac{1}{32}, \textit{ für } k = 0, 1, \cdots, 5$$
(Es gibt $\binom{5}{k}$ *Möglichkeiten,*
die k Zahlwürfe auf
die 5 Möglichkeiten zu verteilen)
z.B. $k = 3$*:*
0 1 0 1 1

Agenommen, Y ist die Anzahl der Kopfwürfe.
$P(Y = k) = \binom{5}{k} * \frac{1}{32}$, $k = 0, \cdots, 5$.
X und Y haben dieselbe Verteilung (!!!), aber X und Y sind <u>nicht</u> gleich.

Zufallsgröße=Abbildung



$X : \Omega \to \mathbb{R}$
$\mathcal{L}(X)$, *Verteilung von X*

4.2 Einige wichtige Verteilungen

4.2.1 Binomialverteilung

X heißt <u>binomial verteilt</u> mit Parameter $n(\in\mathbb{N})$ und $p(\in [0, 1])$, wenn

$$\boxed{P(X = k) = \binom{n}{k} * p^k * (1-p)^{n-k}, k = 0 \cdots, n}$$

Die Zufallsvariable X aus Beispiel 4.1.1 (s. Seite 27) ist $Bin(5, \frac{1}{2})$-verteilt (kurz: $\mathcal{L}(X) = Bin(5, \frac{1}{2}), X \sim Bin(5, \frac{1}{2})$).
Diese Verteilungen tauchen auf, wenn man bei n unabhängigen Versuchswiederholungen zählt, wie oft ein bestimmtes Ereignis, das in jedem Einzelexperiment die Wahrscheinlichkeit p hat, auftritt.
<u>Begründung:</u>
Jedes Ereignis

$$\underbrace{A \times A \times A^c \times \cdots \times A \times A^c \times \cdots \times A}_{insgesamt\ n\ Faktoren,\ davon\ k\ mal\ A}$$

$$(p * p * (1-p) \cdots P(1-p) \cdots p) = p^k (1-p)^{n-k}$$

im Produktexperiment hat die Wahrscheinlichkeit $p^k(1-p)^{n-k}$, und es gibt $\binom{n}{k}$ Möglichkeiten, die k "Erfolge" auf die n Positionen zu verteilen.

<u>Spezialfall: $n = 1$</u>

$$P(X = 1) = p,\ P(X = 0) = 1 - p$$

heißt auch <u>Bernoulli-Verteilung</u> (mit Parameter p).

4.2.2 Poissonverteilung

X heißt <u>Poisson-verteilt</u> mit Parameter $\lambda(> 0)$ wenn

$$\boxed{P(X = k) = e^{-\lambda} * \frac{\lambda^k}{k!} \text{ für alle } k \in\mathbb{N}_0}$$

und spielt eine wichtige Rolle als Grenzverteilung (s. z.B. 3.3.1 auf Seite 22, Anzahl der Fixpunkte einer zufälligen Permutation, bei $n \to \infty$ erhählt man die Poisson-Verteilung mit Parameter 1).

Poisson-Verteilungen treten auch als Grenzwerte von $Bin(n,p)$ bei großen n und kleinem p auf:

Satz 4.2.1 *(Gesetz der seltenen Ereignisse)*
Ist $(P_n) \subset [0,1]$ *mit* $n \in \mathbb{N}$ *eine Nullfolge mit*
$lim_{n\to\infty} n * p_n = \lambda (\in (0,\infty))$,
so gilt für alle $k \in \mathbb{N}_0$
$lim_{n\to\infty} \binom{n}{k} * p_n^k * (1-p_n)^{n-k} = e^{-\lambda} * \frac{\lambda^k}{k!}$

<u>Beweis:</u>

$$\binom{n}{k} * p_n^k * (1-p_n)^{n-k} = \underbrace{\frac{n*(n-1)*\cdots*(n-k+1)}{n^k}}_{\to 1} * \underbrace{\frac{(n*p_n)^k}{k!}}_{\frac{\lambda^k}{k!}} * (1-\frac{\overbrace{n*p_n}^{\to\lambda}}{n})^n \overbrace{-k}^{egal}$$

$$\approx (1-\frac{\lambda}{n})^n \to e^{-\lambda}$$

<u>In Worten:</u>
Bei einer großen Anzahl von Wiederholungen (n) und kleiner Erfolgswahrscheinlichkeit (p) ist die Zahl X der Erfolge näherungsweise Poison-verteilt (mit Parameter $\lambda = n*p$).
Zahlreiche Anwendungen:

- Anzahl Druckfehler pro Seite
- emittierte Partikel

4.2.3 geometrische Verteilung

Angenommen, wir werfen einen (fairen) Würfel solange, bis eine '6' erscheint, sei X die hierfür nötige Wurfzahl (<u>einschließlich</u> der 6).
Offensichtlich gilt $X = n$ genau dann, wenn die ersten $n-1$ Würfe keine '6' liefert. Aufgrund der Unabhängigkeit der Würfe hat dieses Ereignis die Wahrscheinlichkeit
$$\underbrace{(1-\frac{1}{6})*\cdots*(1-\frac{1}{6})}_{n-1-mal} * \frac{1}{6}$$
Allgemeiner, wenn X nur Werte aus $\mathbb{N}$ annimmt und
$P(X = n) = (1-p)^{n-1} * p$ für alle $n \in \mathbb{N}$ gilt, dann heißt X <u>geometrisch verteilt</u> mit Parameter $p(\in (0,1])$: Anzahl der Versuche bis (einschließlich) zum ersten Erfolg bei Erfolgswahrscheinlichkeit p.
Wartet man bis zum r-ten Erfolg ($r = 2,3,\cdots$), so erhält man eine Zufallsvariable X, die nur Werte $\geq r$ annimmt, mit
$P(X = k) = \binom{k-1}{r-1} * (1-p)^{k-r} * p^r$
Dies ist die <u>negative Binomialverteilung</u> mit Parameter $r(\in\mathbb{N})$ und $p(\in (0,1))$ (der Faktor $\binom{k-1}{r-1}$ ist die Anzahl der Möglichkeiten, die ersten $r-1$ Erfolge auf die $k-1$ Möglichkeiten zu verteilen.) (Mathematische Subtilität: Was ist Ω ?

Eine Möglichkeit:
Betrachte unendlich viele Wiederholungen, also: $\Omega_0^{\mathbb{N}}$, wobei Ω_0 das "Basisexperiment" repräsentiert.
Wichtig: dies führt auf überabzählbare Ergebnisräume.
Alternative: $\Omega = \{(0, \cdots, 0, 1) \in 0, 1^k, k \in \mathbb{N}\})$

4.2.4 Hypergeometrische Verteilung

Eine Urne enthält N Kugeln, M weiße und $N - M$ schwarze. Es werden n Kugeln ohne zurücklegen entnommen ($n \leq N.M \leq N$). Sei X die Anzahl der weißen Kugeln in dieser "Stichprobe".
Dann gilt:

$$P(X = k) = \frac{\binom{M}{k} * \binom{N-M}{n-k}}{\binom{N}{n}}$$

für $k = 0, \cdots, min\{M, n\}$,
mit:

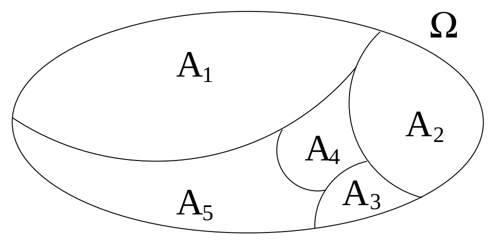
- $\binom{M}{k}$: Möglichkeiten für die weißen Kugeln in der Stichprobe
- $\binom{N-M}{n-k}$: Möglichkeiten für die schwarzen Kugeln
- $\binom{N}{n}$: Anzahl der möglichen Stichproben

Dies ist die hypergeometrische Verteilung mit Parametern n, N und M.
Diese Verteilung ist in §3.3.2 auf Seite 22 aufgetaucht. (Die Anzahl der Asse von Spieler Nord ist hypergeometrisch verteilt mit Parametern $n = 13$, $N = 52$ und $M = 4$).

Ein weiteres typisches Beispiel:
Die Wahrscheinlichkeit für k Richtige beim Zahlenlotto '6 aus 49' ist $\frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}$. Dies ist eine hypergeometrische Verteilung mit Parametern $n = 6$, $N = 49$, $M = 6$.

4.2.5 Multinomialverteilung

Es sei $(\Omega, \mathfrak{A}, \mathfrak{P})$ ein Zufallsexperiment und $A_1, \cdots, A_r$ eine Ereignispartition von Ω, $p_i = \mathfrak{P}(A_i)$. $\qquad p_1 + \cdots + p_r = 1$
Dieses Experiment wird n-mal (unabhängig) wiederholt, $X = (X_1, \cdots, X_r)$ sei der Zufallsvektor, dessen l-te Komponente zählt, wie oft das Ereignis A_l eingetreten ist.



Dann gilt, in Verallgemeinerung von §4.2.1 auf Seite 28 ($r = 2$)

$$P(X = (k_1, \cdots, k_r)) = \frac{n!}{k_1! * \cdots * k_r!} * p_1^{k_1} * p_2^{k_2} * \cdots * p_r^{k_r}$$

für all $k_1, \cdots, k_r \in \mathbb{N}_0$ mit $k_1 + \cdots + k_r = n$. $\binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots$

Man nennt diese Verteilung die Multinomial-Verteilung mit Parametern $p = (p_1, \cdots, p_r)$ und $n \in \mathbb{N}$; es muß $p_i \geq 0$, $p_1 + \cdots + p_r = 1$ gelten.

Zählt man beispielsweise beim n-fachen Wurf eines fairen Würfels, wie oft die Augenzahlen $1, \cdots, 6$ eingetreten sind, so erhählt man die Multinomialverteilung mit Parametern n und $p = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$.

4.3 Erwartungswerte

Vorgeplänkel

$(\{0,1\}^5, \mathfrak{P}(\{0.1\}^5)$ Gleichverteilung)

Wahrscheinlichkeitsraum für den 5-maligen Wurf einer fairen Münze

X

Y $\quad X(\omega) = \omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5$

g $\quad g(x) = (x-2)^2$

Massenfunktion zu Y

<u>Mit:</u>

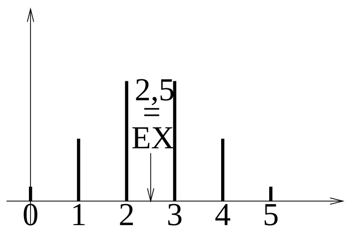
X: Anzahl der Kopfwürfe

Verteilung von X: $Bin(5, \frac{1}{2})$

Massenfunktion zu X:

X	0	1	2	3	4	5
$p(X)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Diagramm:



(Stab mit Massen)

$$EX = 0 * \frac{1}{32} + 1 * \frac{5}{32} + 2 * \frac{10}{32} + 3 * \frac{10}{32} + 4 * \frac{5}{32} + 5 * \frac{1}{32} = \frac{80}{32} = 2,5$$

(EX ≙ Schwerpunkt)

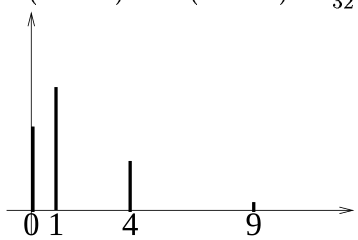
Massenfunktion zu Y:

$P(Y=0) = P(X=2) = \frac{10}{32}$

$P(Y=1) = P(X \in \{1,3\}) = \frac{15}{32}$

$P(Y=4) = P(X \in \{0,4\}) = \frac{6}{32}$

$P(Y=9) = P(X=5) = \frac{1}{32}$



Definition 4.3.1 *Erwartungswert*
Es sei X eine diskrete Zufallsvariable auf einem diskreten Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit Massenfunktion p.
Der Erwartungswert von X wird definiert durch

$$EX = \sum_{\omega\in\Omega} X(\omega) * P(\{\omega\})$$

vorausgesetzt die Summe konvergiert absolut, d.h. $\sum_{\omega\in\Omega} |X(\omega)| * P(\{\omega\}) < \infty$ (sonst: Erwartungswert existiert nicht).

Satz 4.3.1
$EX = \sum_{x\in\mathbb{R}} x * p_X(x)$, wenn p_X die Wahrscheinlichkeitsmassenfunktion zu X bezeichnet.
Ist $f : \mathbb{R} \to \mathbb{R}$ eine Funktion, so ist auch $y := f(X)$ eine Zufallsvariable und es gilt

$$\boxed{EY = \sum_{y\in\mathbb{R}} y * p_Y(y) = \sum_{x\in\mathbb{R}} f(x) * p_X(x)}$$

vorausgesetzt, die beteiligten Summen konvergieren absolut.

Beweis:

Die Mengen $A_x := \{\omega \in \Omega : X(\omega) = x\}, x \in Bild(X)$, bilden einen Partition von Ω, also

$$\sum_{\omega\in\Omega} X(\omega) * P(\{\omega\}) = \sum_{x\in Bild(X)} \underbrace{\sum_{\omega\in A_x} \overbrace{X(\omega)}^{=x} * P(\{\omega\})}_{P(A_x)=P(X=x)=p_X(x)}$$

Umordnungen wegen absoluter Konvergenz erlaubt

$$\begin{aligned}
&= \sum_{x\in Bild(X)} x * p_X(x) = \sum_{x\in\mathbb{R}} x * p_X(x)\\
EY &= \sum_{\omega\in\Omega} Y(\omega) * P(\{\omega\})\\
&= \sum_{\omega\in\Omega} f(X(\omega)) * P(\{\omega\})\\
&= \sum_{x\in Bild(X)} \sum_{\omega\in A_x} \underbrace{f(X(\omega))}_{f(x)} * P(\{\omega\})\\
&= \sum_{x\in Bild(X)} f(x) * p_X(x)\\
&= \sum_{x\in\mathbb{R}} f(x) p_X(x)
\end{aligned}$$

□

Wichtige Konsequenz:

EX hängt nur von der Verteilung von X ab.

Beispiel 4.3.1 *Erwartungswert der Binomialverteilung*
Im Falle $X \sim Bin(n,p)$ erhalten wir

$$\begin{aligned}
EX &= \sum_{k=0}^{n} k * P(X = k)\\
&= \sum_{k=1}^{n} k * \binom{n}{k} * p^k * (1-p)^{n-k}\\
&= n * p * \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)! * ((n-1)-(k-1))!} * p^{k-1} * (1-p)^{(n-1)-(k-1)}\\
&= n * p * \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} * p^k * (1-p)^{(n-1)-k}}_{=(p+(1-p))^{n-1}} && \textit{Binomische Formel !}\\
&\underline{\underline{= n * p}}
\end{aligned}$$

Definiert man Y durch $Y := X(X-1)$, so ergibt sich

$$\begin{aligned} EY &= \sum_{k=0}^{n} k*(k-1)*\binom{n}{k}*p^k*(1-p)^{n-k} \\ &= \sum_{k=2}^{n} k*(k-1)*\binom{n}{k}*p^k*(1-p)^{n-k} \\ &= n*(n-1)*p^2 \end{aligned}$$

Satz 4.3.2 *Rechnen mit Erwartungswerten*

Es seien X und Y diskrete Zufallsvariablen mit existierendem Erwartungswert und $c \in \mathbb{R}$.

1. *Dann existieren auch $E(X+Y), E(c*X)$ und es gilt*
   $E(X+Y) = EX + EY$
   $E(c*X) = c*EX$ — *"der Erwartungsert ist linear"*
2. $X \leq Y \Rightarrow EX \leq EY$ — *"der Erwartungswert ist monoton"*
   $X \leq Y$ *heißt:* $X(\omega) \leq Y(\omega) \forall \omega \in \Omega$

Beweis:

Für den Nachweis der Existenz von $E(X+Y)$ ist die absolute Konvergenz der Reihe $\sum_{\omega\in\Omega}(X+Y)(\omega)*P(\{\omega\})$ zu zeigen:

$$\sum_{\omega\in\Omega}|(X+Y)(\omega)|*P(\{\omega\}) \leq \sum_{\omega\in\Omega}(|X(\omega)|+|Y(\omega)|)*P(\{\omega\}) < \infty$$

$$((X(\omega)+Y(\omega))$$

Außerdem:

$$\begin{aligned} E(X+Y) &= \sum_{\omega\in\Omega}(X(\omega)+Y(\omega))*P(\{\omega\}) \\ &= \sum_{\omega\in\Omega}X(\omega)*P(\{\omega\}) + \sum_{\omega\in\Omega}Y(\omega)*P(\{\omega\}) = (EX)+(EY) \qquad \square \end{aligned}$$

Wichtiger Spezialfall:

$$\left.\begin{array}{rclcl} X \leq |X| & \Rightarrow & EX \leq E|X| & & \\ -X \leq |X| & \Rightarrow & E(-X) \leq E|X| & \Rightarrow & -EX \leq E|X| \end{array}\right\} \Rightarrow \boxed{|EX| \leq E|X|}$$

Definition 4.3.2 *k-tes Moment*

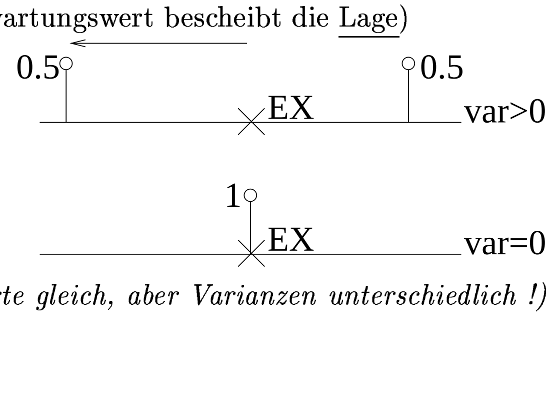
Das k-te Moment einer Zufallsvariablen X ist EX^k (vorausgesetzt, er existiert). Existiert das zweite Moment, so nennen wir

$$var(X) := E(X-EX)^2,\ \sigma(X) := (var(X))^{\frac{1}{2}}$$

die Varianz bzw. Standardabweichung von X.

(Varianz $\hat{=}$ Streuung)

(Beide messen die Variabilität in der Verteilung von X, bei der Standardabweichung hat man dieselben Einheiten wie bei X; der Erwartungswert bescheibt die Lage)



(Erwartungswerte gleich, aber Varianzen unterschiedlich !)

Lemma 4.3.1 *Varianz*
$var(X) = E(X^2) - (EX)^2$

<u>Beweis:</u>

$$\begin{aligned} var(X) &= E(X - EX)^2 \\ &= E(X^2 + (EX)^2 - 2 * (EX) * X) \\ &= (E(X^2)) + E(EX)^2 + E(-2 * (EX) * X) \\ &= E(X^2) + (EX)^2 - 2 * (EX) * (EX) \\ &= E(X^2) - (EX)^2 \end{aligned}$$ □

(wesentlich: nimmt X nur den Wert $c \in \mathbb{R}$ an, so gilt $EX = c$)

Beispiel 4.3.2

1. *Im Fall* $X \sim Bin(n,p)$ *gilt* $EX = n * p$
   $E(X(X-1)) = n * (n-1) * p^2$ (→*Beispiel 4.3.1*),
   also $EX^2 = EX(X-1) + EX = n * (n-1) * p^2 + n * p$
   und damit
   $\underline{var(X)} = EX^2 - (EX)^2 = n * (n-1) * p^2 + n * p - (n * p)^2 = -n * p^2 + n * p = \underline{n * p * (1-p)}$

2. *Ist* X *Poisson-verteilt mit Paramter* λ, *so gilt*
   $$\begin{aligned} \underline{\underline{EX}} &= \textstyle\sum_{k=0}^{\infty} k * e^{-\lambda} * \frac{\lambda^k}{k!} \\ &= \textstyle\sum_{k=1}^{\infty} k * e^{-\lambda} * \frac{\lambda^k}{k!} \\ &= \textstyle\sum_{k=1}^{\infty} e^{-\lambda} * \frac{\lambda^k}{(k-1)!} \\ &= e^{-\lambda} * \lambda * \textstyle\sum_{k=1}^{\infty} \frac{\lambda^{(k-1)}}{(k-1)!} \\ &= \underline{\underline{\lambda}} \end{aligned}$$
   $EX(X-1) = \sum_{k=0}^{\infty} k * (k-1) * e^{-\lambda} * \frac{\lambda^k}{k!} = \lambda^2$, *also*
   $$\begin{aligned} \underline{\underline{varX}} &= EX(X-1) + EX - (EX)^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \underline{\underline{\lambda}} \end{aligned}$$
   Bei der Poisson-Verteilung stimmen also Erwartungswert und Varianz überein.

Bemerkung 4.3.1 *Indikatorfunktion*
Ist M *eine beliebiege Menge und* $A \subset M$, *so heißt*

$$1_A : M \to \mathbb{R},\ 1_A(x) := \begin{cases} 1 & , \ x \in A \\ 0 & , \ sonst \end{cases}$$

die <u>*Indikatorfunktion*</u> *zu* A.

Ist $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitsraum und $A \subset \Omega$ ein Ereignis, so ist $X := 1_A$ eine Zufallsvariable; diese zeigt an, ob ein Ereignis A eingetreten ist (Wert 1) oder nicht (Wert 0).

<u>**Klar:**</u>

$\mathcal{L}(X) = Bin(1,p)$ mit $p = P(A)$

Mit dieser Konstruktion sieht man auch, daß Erwartungswerte Wahrscheinlichkeiten verallgemeinern:

$$E1_A = 0 * P(1_A = 0) + 1 * P(1_A = 1) = P(A)$$

$P : \mathfrak{A} \to \mathbb{R}$
$\mathfrak{A} \rightsquigarrow$ Funktionen raum
$A \to 1_A$

4.4 Bedingte Verteilung und Unabhängigkeit

Sind $X : \Omega \to S_1$, $Y : \Omega \to S_2$ Zufallsgrößen auf einem diskreten Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$, so ist

$$Z : \Omega \to S_1 \times S_2,\ Z(\Omega) := (X(\omega), Y(\omega))$$

mit Werten in $S_1 \times S_2$.
Die Verteilung P^Z von Z nennt man die gemeinsame Verteilung von X und Y.

Beispiel 4.4.1
In der Situation von Absatz 3.3.2 auf Seite 22 (Bridge) sei X die Anzahl der Asse von "Nord", Y die Anzahl der Asse von "Süd".
Dann ist $Z := (X, Y)$ eine Zufallsgröße mit Werten in $\{0,1,2,3,4\} \times \{0,1,2,3,4\}$ und es gilt

$$P(Z = (k,l)) = \frac{\binom{4}{k} * \binom{48}{13-k} * \binom{4-k}{l} * \binom{35+k}{13-l} * \binom{26}{13}}{\frac{(52!)}{(13!)^4}}$$

Tabelle der mit 20825 multiplizierten Werte:

$Y \backslash X$	0	1	2	3	4	*Zeilen-*$\sum$
0	1150	2600	1950	572	55	6327
1	2600	4225	2028	286	0	9139
2	1950	2028	468	0	0	4446
3	572	286	0	0	0	858
4	55	0	0	0	0	55
Spalten-$\sum$	6327	9139	4446	858	55	20825

Die Nullen unten rechts ergeben sich, da es nie mehr als 4 Asse geben kann !

Aus der gemeinsamen Verteilung kann man die Einzelverteilungen zurückerhalten:

$$P(Y = 0) = P(Y = 0, X = 0) + P(Y = 0, X = 1) + \cdots + P(Y = 0, X = 4)$$

Umgekehrt geht es nicht (zumindest nicht ohne Zusatzvoraussetzungen):
aus den Werten

$$P(X = i), P(Y = j), i, j = 0, \cdots, 4$$

lassen sich die Werte

$$P(X = i, Y = j)$$

nicht ermitteln.

Die gemeinsame Verteilung enthält "mehr Informationen" als die Marginalverteilungen. Man kann die Wahrscheinlichkeit von Ereignissen bestimmen, die von beiden Zufallsvariablen abhängen, z.B.:

$$P(X = Y) = P(X = 0, Y = 0) + P(X = 1, Y = 1) + \cdots + P(X = 4, Y = 4)$$
$$= \frac{1150+4225+468+0+0}{20825} \approx 0,2805$$

Man kann mit der gemeinsamen Verteilung von X und Y auch den Erwartungswert einer Zufallsvariablen $Z = f(X, Y)$ bestimmen, die von X und Y abhängt

$$\begin{array}{ccccc} & & S_1 & & \\ & \overset{X}{\nearrow} & & & \\ \Omega & & \times & \xrightarrow{f} & \mathbb{R} \\ & \underset{Y}{\searrow} & & & \\ & & S_2 & & \end{array}$$

$$\boxed{Ef(X,Y) = \sum_{x,y} f(x,y) * P(X = x, Y = y)}$$

$P(X = x, Y = y)$ *gemeinsame Wahrscheinlichkeitsmassenfunktion zu* X *und* Y

$Z = X \cup Y$

Zur Motivation:

Wenn X eine Zufallsvariable mit endlichem zweiten Moment ist, dann ist für jedes $a \in \mathbb{R}$ die mittlere quadratische Abweichung

$$\phi(a) := E(X - a)2$$

wohldefiniert.

Bei welchem a ergibt sich die minimale mittlere quadratische Abweichung ?

$$\phi(a) = \underbrace{EX^2}_{const} - 2 * a * \underbrace{EX}_{const} + a^2$$

ϕ nimmt sein globales Minimum in $a = EX$ an.

Satz 4.4.1 *(bedingte Verteilung, bedingter Erwartungswert)*

$(\Omega, \mathfrak{A}, P), S_1, S_2, X, Y$ *wie oben.*

Dann gilt für alle $x \in S_1$ *mit* $P(X = x) > 0$:

Durch

$$A \to P(Y \in A | X = x) \qquad \left(= \frac{P(\{\omega \in \Omega : Y(\omega) \in A, X(\omega) = x\})}{P(\{\omega \in \Omega : X(\omega) = x\})}\right)$$

wird ein Wahrscheinlichkeitsmaß auf $(S_2, \mathfrak{P}(S_2))$ *definiert, die bedingte Verteilung von* Y *unter* $X = x$, $P^{Y|X=x}$

Ist $S_2 = \mathbb{R}$ *und* $\sum |y| * P^{Y|X=x}(\{y\}) < \infty$, *so nennen wir*

$$E[Y|X = x] = \sum_{y \in \mathbb{R}} y * P^{Y|X=x}(\{y\}) \qquad \left(= \frac{1}{P(X=x)} \sum y * P(X = x, Y = y)\right)$$

den bedingten Erwartungswert von Y *unter* $X = x$.

Für die Verknüpfung der Abbildung $X : \Omega \to S_1$ und $x \mapsto P^{Y|X=x}$ bzw. $x \mapsto E[Y|X = x]$ schreiben wir kurz:

$$P^{Y|X} \text{ bzw. } E[Y|X]$$

Später:

Die Funktion ϕ, $\phi(x) := E[Y|X = x]$, minimiert die mittlere quadratische Abweichung $E(Y - \psi(X))^2$ unter allen Funktionen $\psi :\mathbb{R}\to\mathbb{R}$.

Beweis:

Klar ! $\square$

In der Situation von Beispiel 4.4.1 von Seite 36 ergibt sich beispielsweise als bedingte Erwartung der Anzahl der Asse des Partners, wenn man selbst zwei Asse hat,

$$E[Y|X = 2]$$
$$= 0 * \overbrace{P(Y = 0|X = 2)}^{=\frac{P(Y=0,X=2)}{P(X=2)}=\frac{\frac{1950}{20825}}{\frac{4446}{20825}}} + 1 * P(Y = 1|X = 2) + \cdots + 4 * P(Y = 4|X = 2)$$
$$= 0 * \tfrac{1950}{4446} + 1 * \tfrac{2028}{4446} + 2 * \tfrac{468}{4446} + 3 * \tfrac{0}{4446} + 4 * \tfrac{0}{4446}$$
$$= \tfrac{2964}{4446}$$
$$= \tfrac{2}{3}$$

Beachte:

$EY = 0 * P(Y = 0) + \cdots + 4 * P(Y = 4) = 1$

$E(Y_{Nord} + Y_{Ost} + Y_{Süd} + Y_{West}) = 4$

Symmetrieüberlegungen:

$EY_{Nord} = EY_{Süd} = \cdots = EY_{West}$

Linearität:

$E(\cdots) = EY_{Nord} + EY_{Süd} + \cdots + EY_{West}$

Insgesamt:

$\boxed{EY = 1}$

Beispiel 4.4.2 *Es sei $(\Omega', \mathfrak{A}', P')$ das Modell für ein Zufallsexperiment, in dem ein bestimmtes Ereignis A mit Wahrscheinlichkeit $p > 0$ eintritt. Unser Modell für das n-malige unabhängige Wiederholen dieses Experiments ist*

$(\Omega, \mathfrak{A}, P)$ *mit* $\Omega = \Omega'^n$ *(*$= \Omega' \times \cdots \times \Omega'$*, n Faktoren),* $\mathfrak{A} = \mathfrak{P}(\Omega)$

und

$P(\{(\omega_1, \cdots, \omega_n)\}) = P'(\{\omega_1\}) * \cdots * P'(\{\omega_n\})$

Es sei $X(\omega) := \#\{1 \le i \le n : \omega_i \in A\}$ die Anzahl der Versuche, bei denen A eingetreten ist, und $Y : \Omega \to \mathfrak{P}(\{1, \cdots, n\}), Y(\omega) := \{1 \le i \le n; \omega_i \in A\}$ die Menge der Versuchsnummern, in denen A eingetreten ist.

Die gemeinsame Verteilung von X und Y ist offensichtlich auf

$\{(k, B) : k \in \{0, \cdots, n\}, B \subset \{1, \cdots, n\}$ *mit* $\#B = k\}$

konzentriert und für jedes Element dieser Menge gilt:

$P(X = k, Y = B) = \prod_{j\in B} p * \prod_{j\notin B}(1 - p) = p^k(1 - p)^{n-k}$

Aus Absatz 4.2.1 von Seite 28 ist bereits

$P(X = k) = \binom{n}{k} * p^k * (1 - p)^{n-k}$

bekannt, also folgt

$P^{Y|X=k}(\{B\}) = \frac{p^k*(1-p)^{n-k}}{\binom{n}{k}*p^k*(1-p)^{n-k}} = \frac{1}{\binom{n}{k}}$ *(wenn $\#B = k$)*

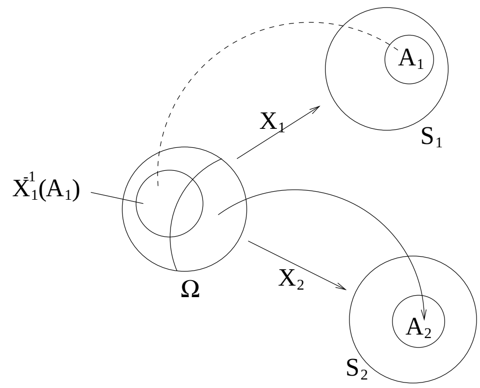
Die bedingte Verteilung von Y unter X hängt nicht mehr von p ab; man erhält die Laplace-Verteilung auf der Menge aller Teilmengen vom Umfang k von $\{1, \cdots, n\}$:

Alle möglichen Anordnungen der k "Erfolge" auf die n möglichen Positionen sind gleich wahrscheinlich.

Wir dehnen den Unabhängigkeitsbegriff auf Zufallsgrößen aus.

Definition 4.4.1 *(stochastisch unabhängig)*

Für jede $i \in I$ sei $X_i : \Omega \to S_i$ eine diskrete Zufallsgröße. Die Familie $\{X_i : i \in I\}$ heißt stochastisch unabhängig, wenn für jede Wahl von $A_i \in S_i, i \in I$, die Ereignisfamilie $\{X_i^{-1}(A_i) : i \in I\}$ stochastisch unabhängig ist im Sinne von Definition 2.1.3 von Seite 15



Satz 4.4.2 *(stochastische Unabhängigkeit)*

Eine Familie $\{X_i : i \in I\}$ von Zufallsgrößen auf einem diskreten Wahrscheinlichkeitsraum ist genau dann unabhängig, wenn für alle $\{i_1, \cdots, i_n\} \subset I$ gilt:

$$P(X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}, \cdots, X_{i_n} = x_{i_n}) = P(X_{i_1} = x_{i_1}) * \cdots * P(X_{i_n} = x_{i_n})$$

für alle $x_{i_1} \in S_{i_1}, \cdots, x_{i_n} \in S_{i_n}$.

Beweis:

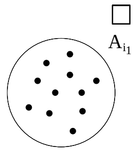
Wäre $A_i = \{x_i\}$, um zu sehen, daß die Bedingung notwendig ist.

$(X_i^{-1}(A_i) = \{\omega : X_i(\omega) = x_i\})$

Für beliebige $A_i \subset S_i$ und $\{i_1, \cdots, i_n\} \subset I$ gilt:

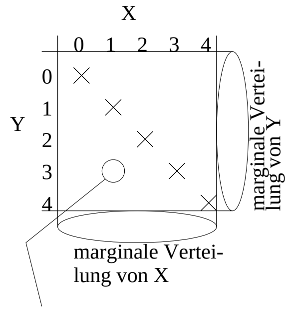
$$\begin{aligned} & P(\textstyle\bigcap_{j=1}^n X_{i_j}^{-1}(A_{i_j})) \\ &= \textstyle\sum_{x_{i_1} \in A_{i_1}, \ldots, x_{i_n} \in A_{i_n}} P(X_{i_1} = x_{i_1}, \cdots, X_{i_n} = x_{i_n}) \\ &= \textstyle\sum_{x_{i_1} \in A_{i_1}, \ldots, x_{i_n} \in A_{i_n}} P(X_{i_1} = x_{i_1}) * \cdots * P(X_{i_n} = x_{i_n}) \\ &= \textstyle\sum_{x_{i_1} \in A_{i_1}} \sum_{x_{i_2} \in A_{i_2}} \cdots \sum_{x_{i_n} \in A_{i_n}} P(X_{i_1} = x_{i_1}) * \cdots * P(X_{i_n} = x_{i_n}) \\ &= (\textstyle\sum_{x_{i_1} \in A_{i_1}} P(X_{i_1} = x_{i_1})) * \cdots * (\sum_{x_{i_n} \in A_{i_n}} P(X_{i_n} = x_{i_n})) \\ &= P(X_{i_1} \in A_{i_1}) * \cdots * P(X_{i_n} \in A_{i_n}) \end{aligned}$$

$\square$



Bei einer endlichen Familie $X_1, \cdots, X_n$ hat man also Unabhängigkeit genau dann, wenn die gemeinsame Massenfunktion p, $p(x_1, \cdots, x_n) = P(X_1 = x_1, \cdots, X_n = x_n)$, das Produkt der marginalen Massenfunktionen $p_1, \cdots, p_n$, $p_i(x_i) = P(X_i = x_i)$ ist:

$\boxed{p(x_1, \cdots, x_n) = p_1(x_1) * \cdots * p_n(x_n)}$ für alle $x_i \in S_i$



Bei unabhängigen X, Y steht in Position i, j das Produkt der Randverteilungen

Bei unabhängigen Zufallsgrößen ergibt sich die gemeinsame Verteilung aus den Randverteilungen.

4.5 Reellwertige diskrete Zufallsgrößen

Satz 4.5.1 *(Multiplikationsregel für Erwartungswerte)*
Sind X und Y unabhängige Zufallsvariablen mit existierendem Erwartungswert, so existiert auch der Erwartungswert zu $X * Y$, und es gilt:

$$\boxed{E(X * Y) = (EX) * (EY)}$$

$$A \rightsquigarrow X = 1_A$$

$$E1_A = 0 * P(1_A = 0) + 1 * \underbrace{P(1_A = 1)}_{P(A)} = P(A)$$

$$E1_A 1_B = E1_{A \cap B} = P(A \cap B)$$

$$\|$$

$$E1_A * E1_B = P(A) * P(B)$$

<u>Beweis:</u>

Die Menge $A_{xy} := \{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}$,
$x \in Bild(X), y \in Bild(Y)$, bilden eine Partition von Ω, also

$\sum_{\omega \in \Omega} |X * Y(\omega)| * P(\{\omega\})$
$= \sum_{\substack{x \in Bild(X) \\ y \in Bild(Y)}} |x| * |y| * \underbrace{P(X = x, Y = y)}_{P(X=x)*P(Y=y)}$
$= (\sum_{x \in Bild(X)} |x| * P(X = x)) * (\sum_{y \in Bild(Y)} |y| * P(Y = y))$
$= E|X| * E|Y| < \infty$,

d.h. der Erwartungswert zu $X * Y$ existiert.

Dasselbe Argument, nur ohne Betragsstriche, beweist die Gleichheit:

$$
\begin{aligned}
&EXY \\
&= \sum_{\omega\in\Omega} X * Y(\omega) * P(\{\omega\}) \\
&= \sum_{\substack{x\in Bild(X)\\ y\in Bild(Y)}} x * y * P(X=x, Y=y) \\
&= \overbrace{\Big(\sum_{x\in Bild(X)} x * P(X=x)\Big)}^{EX} * \overbrace{\Big(\sum_{y\in Bild(Y)} y * P(Y=y)\Big)}^{EY} \qquad \Box
\end{aligned}
$$

Satz 4.5.2 *(Cauchy-Schwarz-Ungleichung)*
Existiert zu den Zufallsvariablen X, Y das zweite Moment, so existiert auch EXY und es gilt:

$$\boxed{(EXY)^2 \leq (EX^2) * (EY^2)}$$

Beweis:

$|X * Y(\omega)| = |X(\omega)| * |Y(\omega)| \leq X(\omega)^2 + Y(\omega)^2 \qquad \forall \omega \in \Omega$

Es gilt also:

$$\sum_{\omega\in\Omega} |X * Y(\omega)| * P(\{\omega\}) \leq \sum_{\omega\in\Omega} X(\omega)^2 * P(\{\omega\}) + \sum_{\omega\in\Omega} Y(\omega)^2 * P(\{\omega\}) = EX^2 + EY^2 < \infty,$$

also existiert EXY.

Für ein beliebiges $t \in \mathbb{R}$ existiert dann auch das zweite Moment zu $X + t * Y$ und ist ≥ 0:

$$0 \leq E(X + t * Y)^2 = EX^2 + t^2 * EY^2 + 2 * t * EXY \qquad \forall t \in \mathbb{R}.$$

Fall 1: $EY^2 = 0$

Dann steht rechts, als Funktion von t, eine Gerade mit Funktionswert ≥ 0 $\forall t$, also muß die Steigung 0 sein, und das heißt $EXY = 0$.

Fall 2: $EY^2 > 0$

Berechne den kleinsten Wert der Parabel rechts (Schulmathematik), man erhält

$$\frac{1}{EY^2} \underbrace{(EX^2 * EY^2 - (E(XY))^2)}_{\geq 0 (muß\ gelten)}. \qquad \Box$$

Bemerkung 4.5.1
Ist $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitsraum mit der Eigenschaft $P(\{\omega\}) > 0$ $\forall \omega \in \Omega$, so ist

$H := \{X : \Omega \to \mathbb{R}: EX^2 < \infty\}$ *Vektorraum*

mit *mit Skalarprodukt*

$< X, Y > := EXY$ *bilinear*

ein Hilbert-Raum. *"Im Krieg, in der Liebe und im Hilbert-Raum ist alles erlaubt"*
(ein zentraler Gegenstand der Funktionalanalysis 3)

Ist Z irgendeine Zufallsgröße mit Werten in irgendeiner Menge S, so wird durch

$H = (Z) := \{X \in H : \exists \phi : S \to \mathbb{R} \text{ mit } X = \phi(Z)\}$

ein Unterraum definiert.

$$\begin{array}{ccc} & \Omega & \xrightarrow{Z} & S \\ X & \downarrow & \swarrow^{\phi} & \\ & \mathbb{R} & & \end{array}$$

Die Abbildung

$H \to H(Z)$

$X \mapsto E[X|Z]$

ist die Orthogonalprojektion von H auf $H(Z)$ ($\rightsquigarrow$ Übungen)

Dies ist einer der Punkte, in dem die Verbindung Stochastik$\leftrightsquigarrow$ Funktionalanalysis deutlich wird.

Definition 4.5.1 *(Kovarianz)*
Es seien X, Y Zufallsvariablen mit endlichem zweiten Moment und Standardabweichungen σ_x, σ_y.
Dann heißt

$$\boxed{cov(X,Y) := E(X - EX) * (Y - EY)}$$

die Kovarianz von X und Y.

(das meint:
$cov(X,Y) = E((X - EX) * (Y - EY))$*)*

Im Falle $cov(X,Y) = 0$ nennt man X und Y unkorreliert.
Ist $\sigma_x * \sigma_y > 0$, so nennt man

$$\boxed{\rho(X,Y) := \frac{cov(X,Y)}{\sigma_x * \sigma_y}}$$

den Korrelationskoeffizienten zu X und Y.

Satz 4.5.3 *(Äquivalenzen)*
Es seien X und Y Zufallsvariablen mit existierendem zweiten Moment. Dann gilt:

1. $cov(X,Y) = EXY - (EX) * (EY)$
2. X, Y *unabhängig* $\Rightarrow X, Y$ *unkorreliert*
3. $-1 \le \rho(X,Y) \le +1$ *(wenn definiert)*

Beweis:

1. Mit der Linearität des Erwartungswertoperators folgt
$$\begin{aligned} cov(X,Y) &= E(X * Y - (EX) * Y - X * (EY) + (EX) * (EY)) \\ &= E(XY) - (EX) * (EY) - (EX) * (EY) + (EX) * (EY) \\ &= E(XY) - (EX) * (EY) \end{aligned}$$

2. Verwende (1) und Satz 4.5.1 von Seite 40

3. $var(X) * var(Y) * \rho(X,Y)^2$
$$= (\overbrace{E\,\underbrace{(X - EX)}_{X'} * \underbrace{(Y - EY)}_{Y'}}^{cov(X,Y)})^2 \qquad \sigma_x = \sqrt{var(X)}$$

$$\overset{Cauchy-Schwarz}{\leq} (E(X-EX)^2) * (E(Y-EY)^2)$$
$$\underline{= var(X) * var(Y)}$$
$$\Rightarrow \underline{\underline{\rho(X,Y)^2 \leq 1}}$$

Multiplikationsregel: X, Y unabhängig $\Leftrightarrow EXY = EX * EY$
$cov(X,Y) = EXY - EX * EY$

Insbesondere: X, Y unabhängig $\underset{\not\Leftarrow}{\Rightarrow}$ X, Y unkorreliert.

$cov(X,X) = EX^2 - (EX^2) = var(X)$

Satz 4.5.4 *(Gleichheit von Bienaymé)*
Es seien $X_1, \cdots, X_n$ Zufallsvariablen mit existierendem zweiten Moment.
Dann gilt:

$$var(\sum_{i=1}^n X_i) = \sum_{i=1}^n var(X_i) + \sum_{i \neq j} cov(X_i, X_j)$$

Sind speziell $X_1, \cdots, X_n$ unabhängig (unkorreliert reicht), so gilt

$$\boxed{var(\sum_{i=1}^n X_i) = \sum_{i=1}^n var(X_i)}$$

Gleichheit von Bienaymé

Beweis:

$$\begin{aligned} var(\sum_{i=1}^n X_i) &= E(\sum_{i=1}^n X_i)^2 - (E(\sum_{i=1}^n X_i))^2 \\ &= E(\sum_{i,j=1}^n (X_i * X_j) - \sum_{i,j=1}^n EX_i * EX_j \\ &= \sum_{i,j=1}^n (EX_iX_j - EX_i * EX_j) \\ &= \sum_{i=1}^n (\underbrace{EX_i^2 - (EX_i)^2}_{var(X_i)}) + \sum_{i \neq j} (\underbrace{EX_iX_j - EX_i * EX_j}_{cov(X_i,X_j)}) \end{aligned}$$

□

Beispiel 4.5.1
In einem Zufallsexperiment sei A ein Ereignis mit Wahrscheinlichkeit p. Das Experiment werde n-mal unabhängig wiederholt; X_i zeige an, ob A im i-ten Durchgang eingetreten ist ($X_i = 1$), oder nicht ($X_i = 0$).
Dann sind $X_1, \cdots, X_n$ unabhängig mit

$$EX_i = 0 * P(X_i = 0) + 1 * P(X_i = 1) = p, \; EX_i^2 = EX_i = p,$$

da X_i nur die Werte 0 und 1 annimmt gilt ferner

$$var(X_i) = EX_i^2 - (EX_i)^2 = p - p^2 = p * (1-p)$$

Also gilt für $S_n := \sum_{i=1}^n X_i$

$$ES_n = n * p, \; var(S_n) = n * p * (1-p)$$ *(Bienaymé, Seite 43)*

Bereits bekannt:

$S_n \sim Bin(n,p)$ (Anzahl der Erfolge bei n Wiederholungen)

also hat man einen neuen Beweis für die Formel aus Beispiel 4.3.2(1) von Seite 35

$p : \mathbb{R} \to [0,1], p(x) = P(X = x)$
$\mathbb{Z} \qquad (p_k), k \in \mathbb{Z}$

Weitere Spezialisierung: Zufallsvariable mit Werten in $\mathbb{Z}$,
schreibe p_k für $P(X = k), k \in \mathbb{Z}$

Satz 4.5.5 *(und Definition) Faltung*

1. *Es seien P, Q Wahrscheinlichkeitsmaße auf $\mathbb{Z}$ mit Massenfunktionen p, q.*
   Dann ist auch $r : \mathbb{Z} \to \mathbb{R}$,
   $r_n = \sum_{k\in\mathbb{Z}} p_k * q_{n-k} \;\; \forall n \in \mathbb{Z}$ *(Cauchy-Produkt)*
   eine Wahrscheinlichkeitsmassenfuntion.
   Das zugehörige Wahrscheinlichkeitsmaß R nennen wir die Faltung von P und Q, $R = P \star Q$

2. *Sind X und Y unabhängige Zufallsvariablen mit Verteilung P bzw. Q, so ist $P\star Q$ die Verteilung von $X + Y$.*

Beweis:

1. Offensichtlich hat man $r_n \geq 0$ für alle $n \in \mathbb{Z}$, sowie

$$\begin{aligned}\sum_{n\in\mathbb{Z}} r_n &= \sum_{n\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} p_k * q_{n-k} \\ &= \sum_{k\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} p_k * q_{n-k} \\ &= \sum_{k\in\mathbb{Z}} p_k * (\sum_{m\in\mathbb{Z}} q_m) \\ &= \sum_{k\in\mathbb{Z}} p_k * 1 \\ &= 1\end{aligned}$$

Rechtfertigung: großer Umordnungssatz ($\rightsquigarrow$ Analysis)

2. Wir zerlegen nach dem Wert von X:

$$\begin{aligned}P(X+Y=n) &= \sum_{k\in\mathbb{Z}} P(X+Y=n.X=k) \\ &= \sum_{k\in\mathbb{Z}} P(Y=n-k, X=k) \\ &= \sum_{k\in\mathbb{Z}} P(Y=n-k) * P(X=k) \\ &= \sum_{k\in\mathbb{Z},k=0} p_k * q_{n-k} \\ &= r_n\end{aligned}$$

□

(general abstract nonsense: "$\star$" ist eine kommutative Verknüpfung von Wahrscheinlichkeitsmaßen auf $\mathbb{Z}$.)

($P\star? = P \quad ? = \delta_0$, das im Punkt 0 konzentrierte Wahrscheinlichkeitsmaß)

Beispiel 4.5.2
Seien X, Y unabhängig und poisson-verteilt mit Parameter λ bzw. μ

$$p_k = e^{-\lambda} * \frac{\lambda^k}{k!}, k \geq 0$$
$$q_k = e^{-\mu} * \frac{\mu^k}{k!}, k \geq 0$$

$$\begin{aligned}P(X+Y=n) &= \sum_{k\in\mathbb{Z}} e^{-\lambda} * \frac{\lambda^k}{k!} * e^{-\mu} * \frac{\mu^{(n-k)}}{(n-k)!} \\ &= \sum_{k=0}^{n} e^{-\lambda} * \frac{\lambda^k}{k!} * e^{-\mu} * \frac{\mu^{(n-k)}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} * \frac{1}{n!} * \underbrace{\sum_{k=0}^{n} \binom{n}{k} * \lambda^k * \mu^{n-k}}_{\textit{Binomische Formel}}\end{aligned}$$

$$= \boxed{e^{-(\lambda+\mu)} * \frac{(\lambda+\mu)^n}{n!}}$$

Also ist $X+Y$ *__wieder__ poisson-verteilt, und zwar mit Parameter* $\lambda+\mu$.
(g.ä.h.n.: die Poisson-Verteilungen bilden eine Faltungshalbgruppe)

4.6 Wahrscheinlichkeitserzeugende Funktion

Ist $(a_n), n \in \mathbb{N}_0$ eine Folge reeller Zahlen, so nennt man

$$\boxed{\hat{a}(z) := \sum_{n=0}^{\infty} a_n * z^n} \qquad \hat{a}(0) = a_0, \hat{a}'(0) = a_1, \cdots$$

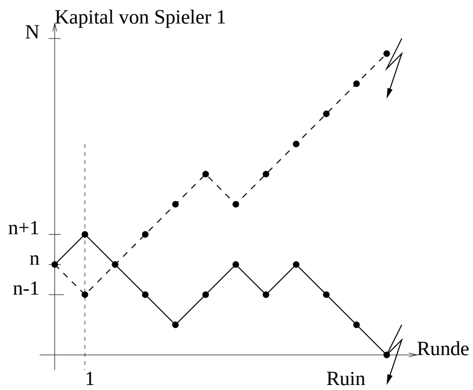
die erzeugende Funktion zu $(a_n), n \in \mathbb{N}_0$.

Gilt $sup_{n \in N_0} |a_n| \leq 1 (< \infty)$, so kann $\hat{a}$ in $\{z \in \mathbb{C}\colon |z| < 1\}$ beliebig oft "unter dem Summenzeichen" differenziert werden; insbesondere:

$$\boxed{a_n = \frac{1}{n!} * \frac{d^u}{dz^u} \hat{a}(z) \Big|_{z=0}}$$

Manche Probleme, insbesondere die Behandlung von Differentialgleichungen, können durch den Übergang zu erzeugenden Funktionen vereinfacht werden. ("Transformationsmethode")

Beispiel 4.6.1 *(Ein Ruin-Problem)*
Spieler I besitzt nDM, Spieler II $(N - n)DM$. In jeder Runde gewinnt I von II $1DM$ mit Wahrscheinlichkeit p und verliert sonst $1DM$ an II.



Das Spiel wird fortgesetzt, bis einer der Spieler sein gesamtes Geld verloren hat. Mit welcher Wahrscheinlichkeit gewinnt I das Spiel ?
Sei $N \in \mathbb{N}$ fest,

A_n: I gewinnt, bei Anfangskapital n

B: I gewinnt die erste Runde

Sei $p_n := P(A_n)$
Das Gesetz von der totalen Wahrscheinlichkeit (Satz 2.1.1(2), Seite 13) liefert

$$P(A_n) = P(A_n|B) * P(B) + P(A_n|B^c) * P(B^c)$$

also:

$$p_n = p_{n+1} * p + p_{n-1} * (1 - p)$$

für $1 \leq n \leq N - 1$
Randbedingung: $p_0 = 0$, $p_N = 1$

Nochmal:

Differentialgleichung:	$p_n = p * p_{n+1} + (1 - p) * p_{n-1}$
Randbedingung:	$p_0 = 0, p_N = 1$

Umformung:
$r := \frac{1-p}{p}$, *wir setzen (zunächst)* $r \neq 1$ *voraus (also* $p \neq \frac{1}{2}$*). Löst man nach* p_{n+1} *auf, so erhält man*

$$\boxed{p_{n+1} = (1+r) * p_n - r * p_{n-1}}$$

und damit für $\hat{p}(z) := \sum_{n=0}^{\infty} p_n * z^n : \hat{p}(z) - p_1 * z = (1+r) * z * \hat{p}(z) - r * z^2 * \hat{p}(z)$

Multiplizieren mit z^{n+1}*, über* $n \in \mathbb{N}$
summieren, $p_0 = 0$ *anwenden*

Auflösen nach $\hat{p}$*:*

$$\hat{p}(z) = \frac{p_1 * z}{1-(1+r)*z+r*z^2} = \frac{p_1}{r-1} * (\overbrace{\frac{1}{1-r*z}}^{geom.Reihe} - \overbrace{\frac{1}{1-z}}^{geom.Reihe})$$

Partialbruchzerlegung

Also:
$p_n = \frac{p_1}{r-1} * (r^n - 1)$

Die übrige Randbedingung $p_N = 1$ *liefert* $p_1 = \frac{r-1}{r^N-1}$*, also insgesamt:*

$$\boxed{p_n = \frac{r^n-1}{r^N-1}}$$

Ähnlich (oder auch direkter) erhält man im Falle $r = 1$ *(also* $p = \frac{1}{2}$*):*
$p_n = \frac{n}{N}$ *für* $n = 0, \cdots, N$.

Anwendung:
Ich betrete ein Kasino mit 100DM Kapital und setze beim Roulette in jeder Runde 1DM auf Rot.
Rot erscheint mit einer Wahrscheinlichkeit von $\frac{18}{37}$ und bringt 2DM.
Ich höre auf, wenn ich 100DM gewonnen, oder aber alles verloren habe.
Dieses Experiment paßt in obige Situation mit $p = \frac{18}{37}, N = 200, n = 100$. Die zugehörige Erfolgswahrscheinlichkeit ist $\frac{\left(\frac{19}{18}\right)^{100}-1}{\left(\frac{19}{18}\right)^{200}-1} \approx 0,0047$

Moral:
Es ist offensichtlich geschickter, alles auf einen Schlag auf Rot zu setzen. (Erfolgswahrscheinlichkeit: $\frac{18}{37} \approx 0,4865$)

Definition 4.6.1
Ist X *eine* $\mathbb{N}_0$*-wertige Zufallsvariable, so heißt*
$\hat{p}_X(z) := \sum_{k=0}^{\infty} P(X = k) z^k$ *(=*Ez^X*)*
die Wahrscheinlichkeitserzeugende Funktion zu(r Verteilung von) X.

Warum lohnt sich der Übergang in die "ˆ-Welt" ?

Satz 4.6.1

1. *Ist* X *eine* $\mathbb{N}_0$*-wertige Zufallsvariable mit Wahrscheinlichkeitserzeugender Funktion* $\hat{p}$*, so gilt für alle* $k \in \mathbb{N}$*:*
   das k-te faktorielle Moment
   $EX(X-1) * \cdots * (X-k+1)$

existiert genau dann, wenn $lim_{Z\uparrow 1}\hat{p}^{(k)}(Z)$
existiert, dann gilt: $\quad$ *(*$^{(k)}$ *meint die k-te Ableitung)*

$$EX(X-1)*\cdots*(X-k+1) = lim_{z\uparrow 1}\hat{p}^{(k)}(z)$$

2. *Sind X und Y unabhängige* $\mathbb{N}_0$*-wertige Zufallsvariablen mit Wahrscheinlichkeitserzeugenden Funktionen* $\hat{p}_X(z), \hat{p}_Y(z)$*, so gilt für die Wahrscheinlichkeitserzeugende Funktion*

$$\boxed{\hat{p}_{X+Y}(z) = \hat{p}_X(z) * \hat{p}_Y(z)} \text{ für alle } z \text{ mit } |z| \leq 1.$$

Beweis:

1. Innerhalb des Konvergenzradius darf man Summe und Differentiation vertauschen, d.h.

$$\hat{p}^{(k)}(z) = \sum_{n=k}^{\infty} \underbrace{\overbrace{n*(n-1)*\cdots*(n-k+1)*P(X=n)}^{a_n}*z^{n-k}}_{\geq 0}$$

$$\hat{p}'_X(z) := \sum_{k=1}^{\infty} k*P(X=k)*z^{k-1}$$

Nach dem Satz von Abel ($\rightsquigarrow$ Analysis) gilt für Potenzreihen $\sum_{n=0}^{\infty} a_n z^n$ mit nichtnegativen Koeffizienten

$$lim_{z\uparrow 1}\sum_{n=0}^{\infty} a_n * z^n = \sum_{n=0}^{\infty} a_n,$$

wobei sogar bestimmte Divergenz zugelassen ist.
Damit:

$$lim_{z\uparrow 1}\hat{p}^{(k)}(z) = \sum_{n=0}^{\infty} n*(n-1)*\cdots*(n-k+1)*P(X=n)$$

(oder ab n=k, egal)

$$= EX(X-1)*\cdots*(X-k+1)$$

(Satz 4.6, Seite 33)

2. **Prüfungsrelevanter Ansatz**
$\hat{p}_{X+Y}(z) = Ez^{X+Y} = Ez^X * z^Y = (Ez^X)*(Ez^Y) = \hat{p}_X(z)*\hat{p}_Y(z)$
Da z^Y und z^Y wieder unabhängig sind.
(nach Stundenübung 14: X, Y unabhängig $\Rightarrow f(X), g(Y)$ unabhängig) $\square$

Beispiel 4.6.2
Ist X Poisson-verteilt mit Parameter λ*, so erhält man*

$$\underline{\underline{\hat{p}_X(z)}} = \sum_{k=0}^{\infty} e^{-\lambda} * \frac{\lambda^k}{k!} * z^k = e^{-\lambda} * \overbrace{\sum_{k=0}^{\infty}\frac{(\lambda*z)^k}{k!}}^{e^{\lambda*z}} = \underline{\underline{exp(\lambda*(z-1))}}$$

Hieraus folgt:

$$\hat{p}'(z) = \lambda * \hat{p}_X(z),\ \hat{p}''_X(z) = \lambda^2 * \hat{p}_X(z),$$

also

$$EX = lim_{1\uparrow 1}\hat{p}'_X(z) = \lambda * \hat{p}_X(1) = \lambda$$

und

$$EX(X-1) = \lambda^2 \text{ (vergleiche Beispiel 4.3.2(2), Seite 35)}$$

Ist Y eine weitere, von X unabhängige Zufallsvariable, Poisson-verteilt mit Parameter μ*, so erhält man*

$$\hat{p}_{X+Y}(z) = \hat{p}_X(z) * \hat{p}_Y(z) = e^{\lambda*(z-1)} * e^{\mu*(z-1)} = exp((\lambda + \mu) * (z - 1))$$

Da p durch $\hat{p}$ festgelegt ist, folgt hieraus, daß $X + Y$ Poisson-verteilt ist mit Parameter $\lambda + \mu$ (in Übereinstimmung mit Beispiel 4.5.2, Seite 44).

4.7 Das schwache Gesetz der großen Zahlen

Seien $X_1, \cdots, X_n$ unabhängig und identisch verteilt mit Erwartungswert μ und Varianz σ^2.

Sei $\overline{X}_n := \frac{1}{n} * \sum_{i=1}^n X_i$ der Mittelwert dieser Zufallsvariablen.

$\underline{\underline{E\overline{X}_n}} = \frac{1}{n} * \sum_{i=1}^n X_i = \underline{\underline{\mu}}$,

$$\begin{aligned} var(\overline{X}_n) &= \frac{1}{n^2} * var(\textstyle\sum_{i=1}^n X_i) \\ &= \frac{1}{n^2} * \textstyle\sum_{i=1}^n var(X_i) \text{ (Bienaymé, Seite 43)} \\ &= \frac{1}{n} * \sigma^2 \end{aligned}$$

Für große n ist die Verteilung des Mittelwerts mit kleiner Variabilität um ihren Erwartungswert herum konzentriert.

Satz 4.7.1

1. **(Die Markovsche Ungleichung)**
   Es sei $p > 0$ und $E|X|^p < \infty$. Dann gilt für alle $\alpha > 0$
   $$\boxed{P(|X| \geq \alpha) \leq \frac{1}{\alpha^p} * E|X|^p}$$

2. **(Ungleichung von Chebychev)**
   Es sei $EX^2 < \infty$. Dann gilt für alle $\alpha > 0$:
   $$\boxed{P(|X - EX| \geq \alpha) \leq \frac{1}{\alpha^2} * var(X)}$$

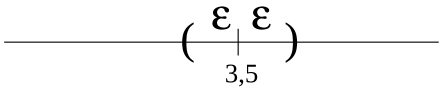
<u>Beweis:</u>

1. **Prüfungsrelevanter Ansatz**
   Sei $\alpha > 0, Y(\omega) := \begin{cases} \alpha & , X(\omega) \geq \alpha \\ 0 & , sonst \end{cases}$
   Offensichtlich gilt
   $|Y(\omega)|^p \leq |X(\omega)|^p$ für alle $\omega \in \Omega$
   Die "Monotonieeigenschaft des Erwartungswertoperators" (Satz 4.3.2, Seite 34) liefert
   $E|Y|^p \leq E|X|^p$
   Da Y nur die beiden Werte 0 und α annimmt, folgt mit Satz 4.3.1 (Seite 33)
   $E|Y|^p = 0^p * P(Y = 0) + \alpha^p * P(Y = \alpha) = \alpha^p * P(|X| \geq \alpha)$
   <u>Zusammen:</u>
   $\alpha^p * P(|X| \geq \alpha) \leq E|X|^p$

2. Sei $Y := X - EX$, verwende Teil (1) mit $p = 2$:
   $$\begin{aligned} P(|X - EX| \geq \alpha) &= P(|Y| \geq \alpha) \\ &\leq \frac{1}{\alpha^2} * E|Y|^2 \\ &= \frac{1}{\alpha^2} * EY^2 \\ &= \frac{1}{\alpha^2} * E(X - EX)^2 \\ &= \frac{1}{\alpha^2} * var(X) \end{aligned}$$
   □

Satz 4.7.2 *(Das schwache Gesetz der großen Zahlen - genauer: eine einfache Version)*
Es sei $X_1, X_2, \cdots$ eine Folge von paarweise unkorrelierten Zufallsvariablen mit Erwartungswert μ und Varianz σ^2; $\overline{X}_n := \frac{1}{n} * \sum_{i=1}^{n} X_i$.
Dann gilt

$$\boxed{lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \epsilon) = 0 \ \forall \epsilon > 0}$$



Beweis:

Bienaymé liefert $var(\overline{X}_n) = \frac{1}{n} * \sigma^2$,
also folgt mit Chebychev
$P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{1}{\epsilon^2} * var(\overline{X}_n) \to 0.$ □

Wichtiger Spezialfall:
Ein Zufallsexperiment, in dem ein bestimmtes Ereignis A mit Wahrscheinlichkeit p auftritt $(0 < p < 1)$, wird unabhängig wiederholt.

Es sei $X_i = \begin{cases} 1 & ,A \text{ tritt bei } i\text{-ter Wiederholung auf} \\ 0 & ,\text{sonst} \end{cases}$

In dieser Situation ist voriger Satz anwendbar:

$EX_i = 0 * P(X_i = 0) + 1 * P(X_i = 1) = p$,

also:

$$lim_{n\to\infty} P(|\overbrace{(\frac{1}{n} * \sum_{i=1}^{n} X_i)}^{\overline{X}_n} - p| \geq \epsilon) = 0 \ \forall \epsilon > 0.$$

Interpretation von $\overline{X}_n$:

relative Häufigkeit von A in den ersten n Versuchen.
p: Wahrscheinlichkeit von A

relative Häufigkeiten konvergieren gegen Wahrscheinlichkeiten

Beispiel 4.7.1 *(Eine Anwendung dieser Ideen in der Analysis)*
Der Approximationssatz von Weierstraß besagt, daß eine reellwertige Funktion auf einem kompakten Intervall $[a, b] \subset \mathbb{R}$ gleichmäßig durch Polynome approximiert werden kann.
Wir wollen diesen Satz mit Hilfe der Stochastik beweisen - sogar konstruktiv (!). Wir können $[a, b] = [0, 1]$ annehmen:
Sei

$p_n : [0, 1] \to \mathbb{R}$, $p_n(x) := \sum_{k=0}^{n} f(\frac{k}{n}) * \binom{n}{k} * x^k * (1 - x)^{n-k}$

das n-te Bernstein-Polynom zu f.
Wir behaupten:

$$\boxed{\forall \epsilon > 0 \ \exists n_0 \in \mathbb{N} \ \forall n \geq n_0 \ \forall x \in [0, 1] : |f(x) - p_n(x)| \leq \epsilon} \qquad \heartsuit$$

Sei $\epsilon > 0$. Stetige Funktionen sind auf kompakten Intervallen sogar gleichmäßig stetig, d.h. es existiert ein $\delta = \delta(\epsilon) > 0$ mit

$\forall x, y \in [0, 1] : |x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\epsilon}{2}$

Außerdem sind stetige Funktionen auf kompakten Intervallen beschränkt, d.h. es gibt ein k *mit*

$$|f(x)| \leq k \ \forall x \in [0,1]$$

Verbindung zur Stochastik:
Wir betrachten den n*-fachen unabhängig wiederholten Wurf einer Münze, die mit Wahrscheinlichkeit* x *das Ergebnis "Kopf" liefert.*

Sei $X_i := \begin{cases} 1 & \text{,"Kopf" im } i\text{-ten Wurf} \\ 0 & \text{,sonst} \end{cases}$

$\overline{X}_n := \frac{1}{n} * \sum_{i=1}^n X_i$

Bekannt: $n * \overline{X}_n \sim Bin(n,x)$*, also*

$$E \underbrace{f(\overline{X}_n)}_{\underbrace{f(\frac{1}{n}*\sum_{i=1}^{n} X_i)}_{?}} = \sum_{k=0}^n f(\tfrac{k}{n}) * \underbrace{P(n * \overline{X}_n = k)}_{\binom{n}{k}*x^k*(1-x)^{n-k}} = p_n(x)$$

Wie im Beweis zu Satz 4.7.2(Seite 50) erhalten wir

$$P(|\overline{X}_n - x| \geq \delta) \leq \frac{x*(1-x)}{n*\delta^2} \leq \frac{1}{4*n*\delta^2}$$

nach Chebychev. $\qquad x*(1-x) \leq \frac{1}{4}$

Wähle nun $n_0 \in \mathbb{N}$ *so groß , daß für alle* $n \geq n_0$ *gilt:*

$$\frac{2*k}{4*n*\delta^2} < \frac{\epsilon}{2}$$

Für alle solche n *gilt dann:*

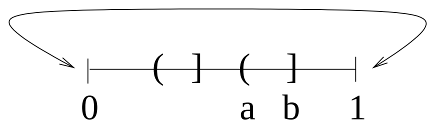
$$\begin{aligned}
|f(x) - p_n(x)| &= |Ef(\overline{X}_n) - f(x)| \qquad (\leq E|f(\overline{X}_n) - f(x)|) \\
&\leq E|\underbrace{f(\overline{X}_n) - f(x)}_{<\frac{\epsilon}{2}}| * 1_{\{|\overline{X}_n - x| < \delta\}} + E|\underbrace{f(\overline{X}_n) - f(x)}_{\leq 2*k}| * 1_{\{|\overline{X}_n - x| \geq \delta\}} \\
&\leq \frac{\epsilon}{2} * \underbrace{P(|\overline{X}_n - x| \leq \delta)}_{\leq 1} + \underbrace{2*k*\underbrace{P(|\overline{X}_n - x| \geq \delta)}_{\leq \frac{1}{4*n*\delta^2}}}_{\leq \frac{\epsilon}{2}} \\
&\leq \epsilon
\end{aligned}$$

Da diese Oberschranke nicht von x *abhängt, ist* $(\heartsuit)$ *damit bewiesen.*

5 Allgemeine Wahrscheinlichkeitsräume

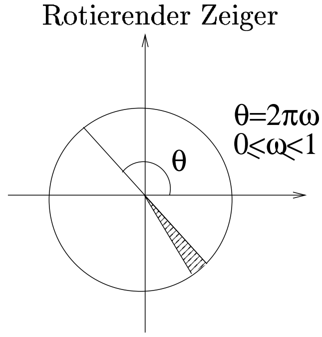
5.1 Vorbemerkungen

Warum reichen diskrete Wahrscheinlichkeitsräume nicht ?
Für viele Zufallsgrößen ist der natürliche Wertebereich überabzählbar (Lebensdauer, Meßfehler,$\cdots$)
"Innermathematisch" gibt es Parallelen zum Übergang von $\mathbb{Q}$ auf $\mathbb{R}$.
Was erhält man als limes der Gleichverteilung auf $\{\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}, 1\}$ mit $n \to \infty$?
Außerdem: die Idealisierung eines unendlich oft unabhängig wiederholten Zufallexperiments spielt in der gesamten Theorie eine große Rolle. Selbst im einfachsten Fall (Münzwurf, nur zwei Ergebnisse, 0 und 1) wird man auf überabzählbare Ereignismengen geführt (die Menge $\{0,1\}^N$ aller 0-1-Folgen ist überabzählbar). Will man Satz 4.7.2 von Seite 50 ohne implizit anwesende, überabzählbare Ergebnisräume formulieren, so müßte man eine Folge $((\Omega_n, \mathfrak{A}_n, P_n))$ mit $n \in \mathbb{N}$ von diskreten Wahrscheinlichkeitsräumen verwenden (unästhetisch).



In Beispiel 1.1.1(2) auf Seite 6 haben wir ein Experiment genannt, das eine "Gleichverteilung auf $\Omega = [0,1)$" erfordert:

Rotierender Zeiger



Forderung 5.1:

$P(x+A) = P(A)$ für alle $x \in [0,1)$, $A \in \mathfrak{A}$,
wobei "+" modulo 1 zu verstehen ist.

$$(x + A = \{x + y : y \in A\})$$

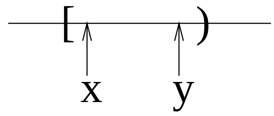
Satz 5.1.1
Es gibt **kein** *Wahrscheinlichkeitsmß auf* $\mathfrak{P}([0,1))$ *mit der Eigenschaft (5.1)*

Beweis:(unter Verwendung des Auswahlaxioms)
Auf $[0,1)$ wird durch

$$\boxed{x \sim y :\Leftrightarrow x - y \in \mathbb{Q}}$$

eine Äquivalenzrelation definiert.
Das Auswahlaxiom erlaubt es, aus jeder Äquivalenzklasse ein Element auszuwählen, sei A die so erhaltene Menge.
Da die Äquivalenzklassen disjunkt sind, enthält A von jeder Äquivalenzklasse genau ein Element.
Wir behaupten:

1. $(A+x) \cap (A+y) = \emptyset$ für alle $x, y \in \mathbb{Q} \cap [0,1), x \neq y$
2. $\bigcup_{x \in Q \cap [0,1)} (A+x) = [0,1)$



<u>zu 1.</u> Angenommen, man hat $a + x = b + y$ mit $x, y \in \mathbb{Q} \cap [0,1), x \neq y, a, b \in A$.
Wegen $x - y \notin \mathbb{Z}$ bedeutet dies $a \neq b$, wegen $a - b \in \mathbb{Q}$ würde A also im Widerspruch zur Konstruktion <u>zwei</u> Elemente einer Äquivalenzklasse enthalten.

<u>zu 2.</u> "$\subset$" ist klar, da die Addition modulo 1 geschieht.
Ist andererseits $z \in [0,1)$, dann existiert ein $a \in A$ mit $a \sim z$, d.h.
$\boxed{x := a - z} \in \mathbb{Q}$. (mit dem "üblichen" -)
Ersetzt man gegebenenfalls x durch $x + 1$, so enthält man die gewünschte Darstellung von z, also (2).
Ist nun P ein Wahrscheinlichkeitsmaß auf $\mathfrak{P}([0,1))$ mit Eigenschaft (5.1) von Seite 52, so muß P auch der Menge A einen Wert zuordnen.
Mit (5.1),(2),(1) und der σ-Additivität würde nun folgen:

$$1 = P([0,1)) \overset{(2)}{=} P(\bigcup_{\underbrace{x \in Q \cap [0,1)}_{abz\ddot{a}hlbar}} (A+x)) \qquad (A+x) \text{ disjunkt, } (2)$$

$$\overset{\sigma-Add.}{=} \sum_{x \in Q \cap [0,1)} P(A+x)$$

$$\overset{(5.1)}{=} \sum_{x \in Q \cap [0,1)} P(A)$$

Dies ist unmöglich.
($P(A) = 0$ würde $1 = 0$, $P(A) > 0$ würde $1 = \infty$ liefern) $\square$

Die Potenzmenge ist also <u>zu groß</u> , wir werden uns mit einer kleineren σ-Algebra zufrieden geben müssen.

5.2 Mengensysteme

Man wird zumindestens gewisse Mengen, beispielsweise den Intervallen im Falle $\Omega = \mathbb{R}$, Wahrscheinlichkeiten zuordnen wollen.

Definition 5.2.1

Es sei $\Omega \neq \emptyset$ und $\mathfrak{E} \subset \mathfrak{P}(\Omega)$. Dann heißt

$$\sigma(\mathfrak{E}) := \boxed{\bigcap_{\substack{\mathfrak{A}\ \sigma-Algebra \\ \mathfrak{A} \supset \mathfrak{E}}} \mathfrak{A}}$$

die von $\mathfrak{E}$ erzeugte $\sigma-$Algebra; $\mathfrak{E}$ nennt man ein Erzeugendensystem (zu $\sigma(\mathfrak{E})$).

Implizit wird hierbei verwendet, daß der Durchschnitt beliebig vieler (!) σ-Algebren wieder eine σ-Algebra ist.

($A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$ etc.)
(einfach zu überprüfen)

Außerdem taucht rechts auf jeden Fall die Potenzmenge von Ω auf, d.h. der Durchschnitt ist nicht leer.

Interpretation:
$\sigma(\mathfrak{E})$ ist die kleinste σ-Algebra, die alle Mengen aus $\mathfrak{E}$ enthält.
Besonders wichtig: $\Omega = \mathbb{R}$.

Definition 5.2.2

Die von den LORA-Intervallen $(a, b], -\infty < a < b < \infty$ erzeugte σ-Algebra heißt die σ-Algebra der Borel-Mengen von $\mathbb{R}$.
Schreibweisen:

$$\mathfrak{B}, \mathfrak{B}(\mathbb{R}), \mathfrak{B}_{\mathbb{R}}$$
$$(= \sigma(\{(a, b] : -\infty < a < b < \infty\}))$$

Eine σ-Algebra (wie $\mathfrak{B}_{\mathbb{R}}$) kann durchaus verschiedene Erzeugendensysteme haben.
(trivialerweise gilt: $\sigma(\mathfrak{A}) = \mathfrak{A}$ für jede σ-Algebra $\mathfrak{A}$)

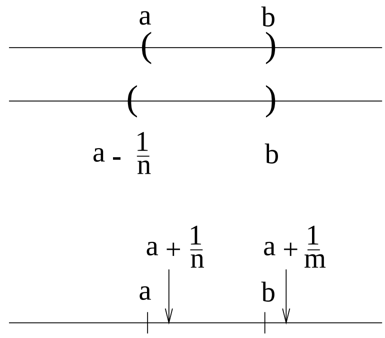
Satz 5.2.1 *(Erzeugendensysteme für Borelmengen)*
$\mathfrak{B}(\mathbb{R})$ *wird erzeugt von den Mengensystemen*

$\mathfrak{E}_1 := \{[a, b) : -\infty < a < b < \infty\}$ *("ROLA-Intervalle")*
$\mathfrak{E}_2 := \{(-\infty, a] : a \in \mathbb{R}\}$
$\mathfrak{E}_3 := \{U \subset \mathbb{R} : U \ offen\}$

Beweis:
Es sei $\mathfrak{E} := \{(a, b] : -\infty < a < b < \infty\}$.
Zum Nachweis von
$(\mathfrak{B}(\mathbb{R}) =)\sigma(\mathfrak{E}) = \sigma(\mathfrak{E}_i)$
reicht es jeweils
$\mathfrak{E}_i \subset \mathfrak{B}$ (das impliziert $\sigma(\mathfrak{E}_i) \subset \mathfrak{B}$)
und
$\mathfrak{E} \subset \sigma(\mathfrak{E}_i)$ (dies impliziert $(\mathfrak{B} =)\sigma(\mathfrak{E}) \subset \sigma(\mathfrak{E}_i)$)
zu zeigen.

Hierbei können wir die "mengenalgebraischen Abgeschlossenheitseigenschaften" von σ-Algebren (gegenüber Komplementbildung, endlichen und abzählbar unendlichen Verinigungen und Durchschnitten) verwenden.
In diesem Sinne ergibt sich $\sigma(\mathfrak{E}_1) = \mathfrak{B}$ aus

$$[a,b) = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m=1}^{\infty} (a - \frac{1}{n}, b - \frac{1}{m}]}_{=(a - \frac{1}{n}, b)}$$

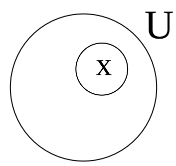


$$(a,b] = \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} [a + \tfrac{1}{n}, b + \tfrac{1}{m})$$

Analog folgt $\sigma(\mathfrak{E}_2) = \mathfrak{B}(\mathbb{R})$

$$(-\infty, a] = \bigcup_{n=1}^{\infty} (a - n, a]$$
$$(a,b] = (-\infty, b] \cap (-\infty, a]^c$$

Bei $\mathfrak{E}_3$ verwenden wir, daß es zu jedem x aus einer offenen Menge U ein x enthaltendes Intervall $(a,b] \subset U$ gibt - wir können sogar $a, b \in \mathbb{Q}$ annehmen.



Damit:

$$U = \bigcup_{\{(a,b) \in \mathbb{Q} \times \mathbb{Q} : (a,b] \subset U\}} (a,b]$$

(a,b) *Paar, keine Intervalle*

Jede offene Teilmenge von $\mathbb{R}$ kann als abzählbare Vereinigung von LORA-Intervallen geschrieben werden und ist somit eine Borel-Menge.
Dies liefert $\sigma(\mathfrak{E}_3) \subset \mathfrak{B}$, die Gegenrichtung folgt mit

$$(a,b] = \bigcap_{n=1}^{\infty} (a, b + \tfrac{1}{n}).$$

$(a, b + \frac{1}{n})$ *offene Mengen*

□

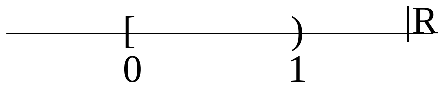
Dieser Satz impliziert, daß die Intervalle $[a,b), (-\infty, a]$ Borel-Mengen sind, ebenso offene Mengen.
Wegen

$$\{a\} = \bigcap_{n=1}^{\infty} (a - \tfrac{1}{n}, a]$$

sind auch alle Einpunktmengen, und damit alle endlichen und abzählbar unendlichen Mengen wie $\mathbb{Q}$ Borelmengen, auch alle kompakten Intervalle, die irrationalen Zahlen, etc $\cdots$.

<u>**Grob:**</u>
$\mathfrak{B}(\mathbb{R})$ ist für alle praktischen Zwecke reichhaltig genug.

$\mathfrak{B}$:σ-Algebra der Borelschen Teilmengen von $\mathbb{R}$.



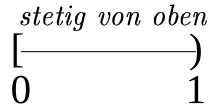
Ist A eine nicht-leere Teilmenge von $\mathbb{R}$, so wird durch $\mathfrak{B}_A := \{B \cap A : B \in \mathfrak{B}\}$ eine σ-Algebra über A definiert (Stundenübung 21), die <u>Spur</u> von $\mathfrak{B}$ auf A. Die Elemente von $\mathfrak{B}_A$ heißen <u>Borel-Mengen</u> von A. Der folgende Satz wird in der Maßtheorie bewiesen:

Satz 5.2.2 *Gleichverteilung auf dem Einheitsintervall*
Es gibt ein Wahrscheinlichkeitsmaß P auf $([0,1), \mathfrak{B}_{[0,1)})$ mit (5.2, Seite 54)
$$P([a,b)) = b - a \forall a, b \in [0,1) \text{ mit } a \leq b$$
("Gleichverteilung auf dem Einheitsintervall")

Bemerkung 5.2.1

1. *Man kann zeigen, daß (5.1, Seite 52) auf (5.2,Seite 54) führt (Übungsaufgabe). Gegenrichtung: später: Satz 5.2.2 zeigt also, daß durch die Verkleinerung des Definitionsbereichs (die für praktische Anwendung akzeptabel ist) das Problem aus §5.1, Seite 52 gelöst werden kann.*

2. *Man kann P auf $(\mathbb{R}, \mathfrak{B})$ fortsetzen durch*
$$P_{\mathbb{R}}(B) := P(B \cap [0,1)) \forall B \in \mathfrak{B}$$
(ist wieder ein Wahrscheinlichkeitsmaß). Umgekehrt erhält man aus einem Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B})$ ein Wahrscheinlichkeitsmaß $P_{[0,1)}$ auf $[0,1)$ durch $P_{[0,1)}(B) := P(B)$ für alle Borelschen Teilmengen von $[0,1)$, wenn nur $P([0,1)) = 1$ gilt.
Dies läßt sich verallgemeinern von $[0,1)$ auf beliebiges $A \in \mathfrak{B}$.
Man kann das P aus Satz 5.2.2 also auch als Wahrscheinlichkeitsmaß auf $\mathbb{R}$ oder $(0,1]$ oder $[0,1]$ oder $(0,1)$ auffassen:
$$P(\{x\}) = lim_{n\to\infty} \underbrace{P([x, x + \frac{1}{n}))}_{x+\frac{1}{n}-x=\frac{1}{n}} = 0$$
Wahrscheinlichkeitsmaße sind stetig von oben



<u>*Also:*</u> *einzelne Punkte spielen keine Rolle.*

3. *In der Maßtheorie nennt man ein Paar $(\Omega, \mathfrak{A})$ mit $\Omega \neq \emptyset$ und $\mathfrak{A}$ σ-Algebra über Ω einen <u>meßbaren Raum</u>, und eine Abbildung*
$$\mu : \mathfrak{A} \to [0, \infty]$$
ein Maß , wenn
$$\mu(\emptyset) = 0$$

$\mu(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$
für alle paarweise disjunkten $A_1, A_2, \dots \in \mathfrak{A}$ gilt.
In diesem Sinne sind Wahrscheinlichkeiten einfach normierte Maße.
Die geometrische Variante des Problems aus §5.1 (Seite 52) lautet:
Läßt sich allen Teilmengen von $\mathbb{R}$ ($\mathbb{R}^d$) ein Länge (Volumen) zuordnen ?
Es ist wieder eine Einschränkung des Definitionsbereichs nötig, man erhält:
Es gibt ein Maß l (das Lebesgue-Maß auf $(\mathbb{R}, \mathfrak{B})$ mit $l((a,b]) = b - a \forall a \leq b$.
Man kann $unif(0,1)$ als "Spur" von l auf dem Einheitsintervall auffassen.

Was ist Eindeutigkeit ?
Hilfsmittel:

Definition 5.2.3 *(Dynkin-System)*
$\Omega \neq \emptyset$. *$\mathfrak{D} \subset P(\Omega)$ heißt Dynkin-System, wenn gilt:*

1. $\Omega \in \mathfrak{D}$
2. $A \in \mathfrak{D} \Rightarrow A^c \in \mathfrak{D}$
3. $A_1, A_2, \dots \in \mathfrak{D}, A_i \cap A_j = \emptyset \forall i \neq j \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathfrak{D}$

$\exists P : \mathfrak{B}_{(0,1)} \to [0,1)$ *mit* $\cdots$ $P((a,b]) = b - a$
$\mathbb{Q} \in \mathfrak{B}$
$unif(0,1)(\mathbb{Q}) = 0$

Gegenüber σ-Algebren wird also die Forderung "abgeschlossen gegenüber abzählbaren Verinigungen" auf disjunkte Verinigungen abgeschwächt.
Der Durchschnitt von (beliebig vielen) Dynkin-Systemen über einer festen Menge ist wieder ein Dynkin-System, wir können also von

$$\delta(\mathfrak{E}) := \bigcap_{\mathfrak{D} \supset \mathfrak{E}, \mathfrak{D}\ Dynkin-System} \mathfrak{D}$$

als von dem von $\mathfrak{E}$ erzeugten Dynkin-System sprechen.
Wir nennen ein Mengensystem $\mathfrak{E}$ durchschnittsstabil (kurz: $\cap$-stabil), wenn gilt:

$$A, B \in \mathfrak{E} \Rightarrow A \cap B \in \mathfrak{E}$$

Satz 5.2.3

1. *Ein $\cap$-stabiles Dynkin-System ist eine σ-Algebra*
2. *Ist $\mathfrak{E}$ $\cap$-stabil, so gilt*
$$\delta(\mathfrak{E}) = \sigma(\mathfrak{E})$$

Beweis:

1. Es seien $A_1, A_2, \dots \in \mathfrak{D}$ (nicht notwendiger Weise disjunkt), z.Z. $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{D}$
   Setze $\boxed{\begin{array}{lll} B_1 & := & A_1 \\ B_n & := & A_n \cap A_1^c \cap \dots \cap A_{n-1}^c \end{array}}$
   $\cap$-Stabilität und Eigenschaft (2) liefern $B_n \in \mathfrak{D} \forall n \in \mathbb{N}$.
   Auch klar: die B_n's sind paarweise disjunkt.
   Eigenschaft (3) liefert daher $\bigcup_{n=1}^{\infty} B_n \in \mathfrak{D}$
   Beachte man noch: $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$
   (siehe auch den Beweis von Satz 1.2.2 (Seite 11)).

2. Da jede σ-Algebra ein Dynkin-System ist, folgt

   $\delta(\mathfrak{E}) \subset \sigma(\mathfrak{E})$

   unmittelbar aus den Definitionen.

   Es sei nun, für jedes $A \in \delta(\mathfrak{E})$

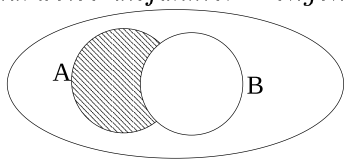
   $\boxed{\mathfrak{D}_A := \{B \subset \Omega : B \cap A \in \delta(\mathfrak{E})\}}$

   Dann ist $\mathfrak{D}_A$ ein Dynkin-System:

   1) ist trivial

   2) folgt mit $B^c \cap A = (A^c + B \cap A + \Omega^c + \Omega^c + \cdots)^c$ $\qquad \Omega^c = \emptyset$

   abzählbare Vereinigung paarweise disjunkter Mengen



   also $\mathfrak{E} \subset \mathfrak{D}_E$ und damit $\delta(\mathfrak{E}) \subset \mathfrak{D}_E \forall E \in \mathfrak{E}$

   denn $\mathfrak{D}_E$ ist ja ein Dynkin-System.

   Dies heißt

   $D \in \delta(\mathfrak{E}), E \in \mathfrak{E} \Rightarrow D \cap E \in \delta(\mathfrak{E})$

   also $E \in \mathfrak{D}_D$ für alle $E \in \mathfrak{E}$, $D \in \delta(\mathfrak{E})$

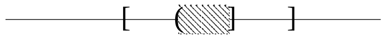
   Dies wiederum liefert $\mathfrak{E} \subset \mathfrak{D}_D$, also

   $\delta(\mathfrak{E}) \subset \mathfrak{D}_D \forall D \in \delta(\mathfrak{E})$

   und damit

   $A \in \delta(\mathfrak{E}), D \in \delta(\mathfrak{E}) \Rightarrow A \cap D \in \delta(\mathfrak{E})$

   Damit ist $\delta(\mathfrak{E})$ $\cap$-stabil und (1) kann verwendet werden.



Anwendung dieser Konstruktion:

Die Mengen $[a, b), 0 \le a \le b < 1$, bilden ein $\cap$-stabiles Erzeugendensystem von $\mathfrak{B}_{[0,1)}$ (Aufgabe 21(ii) in Verbindung mit Satz 5.2.1 (Seite 54)).

Satz 5.2.4

Es sei $\mathfrak{A}$ eine σ-Algebra mit $\cap$-stabilem Erzeuger $\mathfrak{E}$. Sind dann P und Q Wahrscheinlichkeitsmaße auf $\mathfrak{A}$ mit der Eigenschaft $P(E) = Q(E) \forall E \in \mathfrak{E}$, so gilt $P = Q$ (d.h. $P(A) = Q(A) \forall A \in \mathfrak{A}$).

(Stimmen zwei Wahrschinlichkeitsmaße auf einem $\cap$-stabilen Erzeuger überein, so sind sie gleich.)

Beweis:

Es sei $\mathfrak{D} := \{A \in \mathfrak{A} : P(A) = Q(A)\}$

Dann ist $\mathfrak{D}$ ein Dynkin-System.

$\Omega \in \mathfrak{D}$, da $P(\Omega) = 1 = Q(\Omega)$,

$A_1, A_2, \cdots$ aus $\mathfrak{D}$, paarweise disjunkt

$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} Q(A_i) = Q(\bigcup_{i=1}^{\infty} A_i)$

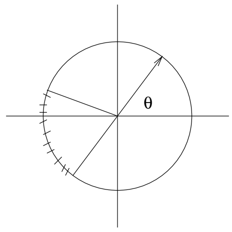
also: $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{D}$

klar: $\mathfrak{E} \subset \mathfrak{D}$

Satz 5.2.3(2) (Seite 57) liefert $\mathfrak{D} \supset \delta(\mathfrak{E}) = \sigma(\mathfrak{E}) = \mathfrak{A}$. $\square$

Insbesondere gibt es genau ein Wahrscheinlichkeitsmaß auf $\mathfrak{B}_{[0,1)}$ mit der Eigenschaft (5.2) (Seite 54)

5.3 Zufallsgrößen und Verteilungen



$\theta = 2 * \pi * \omega,\ \omega \in [0,1]$

$\cancel{\mathfrak{P}([0,1))}$

$\mathfrak{B}_{[0,1)}$

Wie im diskreten ist oft nicht $\omega \in \Omega$ selbst, sondern nur der Wert $X(\omega)$ einer Funktion $X : \Omega \to \Omega'$ interessant.

Definition 5.3.1

Es seien $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathfrak{A}')$ ein meßbarer Raum. Eine Abbildung $X : \Omega \to \Omega'$ heißt Zufallsgröße (auf $(\Omega, \mathfrak{A}, P)$ mit Werten in $(\Omega', \mathfrak{A}')$), wenn X $(\mathfrak{A}, \mathfrak{A}')$-meßbar ist, d.h. wenn gilt

$$X^{-1}(A') \ (:= \{\omega \in \Omega : X(\omega) \in A'\}) \in \mathfrak{A} \forall A' \in \mathfrak{A}'$$

(Urbild von A' unter X)

(Grund: damit ist $P(X \in A')$ für $A' \in \mathfrak{A}'$ definiert.)
(Wenn X beispielsweise eine Lebensdauer bezeichnet, so soll $P(X > 100)$ etc. natürlich definiert sein.
Hier hätte man $A' = (100, \infty)$.)

Analogie zur mengentheoretischen Topologie:

Man hat auf einer Menge M ein System $\mathfrak{U}$ offener Mengen und nennt $f : M \to M'$ stetig, wenn Urbilder offener Mengen offen sind, d.h.

$$f^{-1}(U') \in \mathfrak{U} \forall U' \in \mathfrak{U}'$$

Klar:

Im Falle $\mathfrak{A} = \mathfrak{P}(\Omega)$ ist $X^{-1}(A') \in \mathfrak{A}$ automatisch erfüllt. (Dies ist der Grund dafür, daß Meßbarkeit in der diskreten Situation nicht explizit auftaucht.)

Satz 5.3.1 *(und Definition: Verteilung von X)*

Ist X ein $(\Omega', \mathfrak{A}')$-wertige Zufalssgröße auf $(\Omega, \mathfrak{A}, P)$, so wird durch

$$\mathfrak{A}' \ni A' \mapsto P(X \in A') \ (= P(\underbrace{\{\omega \in \Omega : X(\omega) \in A'\}}_{\in \mathfrak{A}}))$$

ein Wahrscheinlichkeitsmaß auf $(\Omega', \mathfrak{A}')$ definiert, die Verteilung von X.

$(\Omega, \mathfrak{A}, P)$
$\downarrow X$

$$\begin{array}{c}(\Omega', \mathfrak{A}') \\ \uparrow \\ P^X\end{array}$$

Schreibweise: P^X *oder* $\mathcal{L}(X)$ *"law"*

Beweis:

identisch zum diskreten Fall (Satz 4.1.1, Seite 26)

$$\begin{array}{rl} & \{0,1\}^{10} \text{ (Laplace)} \\ X(\omega) = & \sum_{i=1}^{10} \downarrow \\ & \{0,1,\cdots,10\} \\ P^X = & Bin(10, \tfrac{1}{2}) \end{array}$$

Beim Nachweis der Meßbarkeit kann man sich auf Erzeugendensysteme beschränken:

Satz 5.3.2 *(Erzeugendensysteme und Meßbarkeit)*
Es seien $(\Omega, \mathfrak{A})$ *und* $(\Omega', \mathfrak{A}')$ *meßbare Räume,* $X : \Omega \to \Omega'$ *eine Abbildung. Ist* $\mathfrak{E}'$ *ein Erzeugendensystem von* $\mathfrak{A}'$ *und gilt*

$$\boxed{X^{-1}(E') \in \mathfrak{A} \forall E' \in \mathfrak{E}'},$$

so ist X $(\mathfrak{A}, \mathfrak{A}')$*-meßbar.*

Beweis:

Es sei $\mathfrak{A}_0 := \{A' \subset \Omega' : X^{-1}(A') \in \mathfrak{A}\}$

$$\left(\begin{array}{ll}\text{Ziel:} & \mathfrak{A}_0 \supset \mathfrak{A}' \\ \text{bekannt:} & \mathfrak{A}_0 \supset \mathfrak{E}'\end{array}\right)$$

Dann ist $\mathfrak{A}_0$ eine σ-Algebra:

1. $X^{-1}(\Omega') = \Omega \in \mathfrak{A}$, also $\Omega' \in \mathfrak{A}_0$
2. Es sei $A_0 \in \mathfrak{A}_0$, zu zeigen: $A_0^c \in \mathfrak{A}_0$

$$\begin{aligned} X^{-1}(A_0^c) &= \{\omega \in \Omega : X(\omega) \notin A_0\} \\ &= \{\omega \in \Omega : X(\omega) \in A_0\}^c \\ &= (\underbrace{X^{-1}(A_0)}_{\in \mathfrak{A}})^c \in \mathfrak{A}, \end{aligned}$$

also hat man auch $A_0^c \in \mathfrak{A}_0$

"Sorry, so trivial is nu auch wieder nich"

3. analog.

$$\left.\begin{array}{l}\mathfrak{A}_0 \ \sigma\text{-Algebra} \\ \mathfrak{A}_0 \supset \mathfrak{E}' \\ \sigma(\mathfrak{E}') = \mathfrak{A}'\end{array}\right\} \Rightarrow \mathfrak{A}_0 \supset \mathfrak{A}'$$

□

Beispiel 5.3.1 *Es sei* $(\Omega, \mathfrak{A}, P) = ([0,1), \mathfrak{B}_{[0,1)}, unif(0,1))$.
Für jedes $x \in \Omega$ *werde* $T_x : \Omega \to \Omega$ *definiert durch*

$$T_x(y) = \begin{cases} y - x & , \text{ wenn } y \geq x \\ y - x + 1 & , \text{ sonst} \end{cases}$$

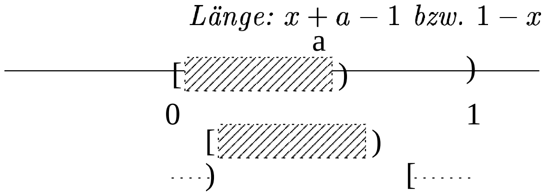
Für alle $A \in \mathfrak{A}$ *gilt*

$$T_x^{-1}(A) = \{y \in \Omega : y - x \in A \text{ oder } y - x + 1 \in A\} = x + A \bmod 1$$

Insbesondere

$$T_x^{-1}([0,a)) = \begin{cases} y[x,x+a) & \text{, wenn } x+1 \leq 1 \\ [0,x+a-1) \cup [x,1) & \text{, sonst} \end{cases} \in \mathfrak{A}$$

Länge: $x+a-1$ *bzw.* $1-x$



also ist T_x $(\mathfrak{A},\mathfrak{A})$*-meßbar (die Mengen* $[0,a), 0 \leq a \leq 1$*, erzeugen* $\mathfrak{B}_{[0,1)}$*).*
Außerdem:

$$P(T_x^{-1}([0,a))) = a = P([0,a))$$

Da P^{T_x} *und* P *somit auf einem* $\cap$*-stabilen Erzeugendensystem übereinstimmen, gilt nach Satz 5.2.4 (Seite 58)* $\boxed{P^{T_x} = P}$,
und somit

$P(x+A) = P(A)$ *für alle* $A \in \mathfrak{A}$*, d.h.* $unif(0,1)$ *ist translationsinvariant (modulo 1, d.h. hat die Eigenschaft 5.1 (Seite 52)).*

Satz 5.3.3 *("Verknüpfungen meßbarer Abbildungen sind meßbar")*
Es seien $(\Omega,\mathfrak{A}), (\Omega',\mathfrak{A}'), (\Omega'',\mathfrak{A}'')$ *meßbare Räume,*

$X : \Omega \to \Omega'$ $(\mathfrak{A},\mathfrak{A}')$*-meßbar,*

$Y : \Omega' \to \Omega''$ $(\mathfrak{A}',\mathfrak{A}'')$*-meßbar*

dann ist

$Z := Y \circ X$ *(als Abbildung von* Ω *in* Ω''*)*

$(\mathfrak{A},\mathfrak{A}'')$*-meßbar.*

<u>Beweis:</u>

$$\Omega \xrightarrow{X} \Omega' \xrightarrow{Y} \Omega''$$

$$\xrightarrow[Z]{\qquad\qquad}$$

Für alle $A'' \in \mathfrak{A}''$ gilt

$$\begin{aligned} Z^{-1}(A'') &= \{\omega \in \Omega : Z(\omega) \in A''\} \\ &= \{\omega \in \Omega : Y(X(\omega)) \in A''\} \\ &= X^{-1}(\{\omega' \in \Omega' : Y(\omega') \in A''\}) \\ &= X^{-1}(\underbrace{Y^{-1}(A'')}_{\in \mathfrak{A}'}) \in \mathfrak{A} \qquad \text{da } Y \ (\mathfrak{A}',\mathfrak{A}'')\text{-meßbar} \end{aligned}$$

$\square$

$$Z = Y \circ X$$
$$Z^{-1} = X^{-1} \circ Y^{-1}$$

5.4 Reellwertige Zufallsgrößen

(Zufalls<u>variable</u>, kurz: ZV)
Sei $(\Omega,\mathfrak{A},P)$ ein Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}$. Auf $\mathbb{R}$ verwenden wir (wenn nicht explizit anders vereinbart) die σ-Algebra $\mathfrak{B}$ der Borel-Mengen.

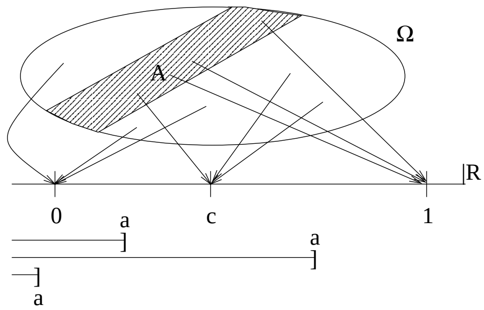
Ist X $(\mathfrak{A}, \mathfrak{B})$-meßbar, so nennen wir X eine Zufallsvariable.
Hierfür reicht

$$X^{-1}((-\infty, a]) \ (= \{\omega \in \Omega : X(\omega) \leq a\}) \in \mathfrak{A} \text{ für alle } a \in \mathbb{R}$$

Allereinfachster Fall:

$X(\omega) \equiv c$ (konstante Abbildung)

$$X^{-1}((-\infty, a]) = \begin{cases} \Omega & , c \leq a \\ \emptyset & , c > a \end{cases}$$



Also: konstante Abbildungen sind grundsätzlich meßbar.

Was passiert mit Indikatorfunktionen ?

Sei A eine beliebige Teilmenge von Ω, $X := 1_A$.

$$X^{-1}((-\infty, a]) = \begin{cases} \Omega & , 1 \leq a \\ A^c & , 0 \leq a < 1 \\ \emptyset & , a < 0 \end{cases}$$

Also: 1_A ist genau dann meßbar, wenn $A \in \mathfrak{A}$.

(Einbettung: $A \to 1_A$)

Ist mit X beispielsweise auch X^2 eine Zufallsvariable ?

Satz 5.4.1 *(Meßbarkeit von Funktionen)*
Ist $g : \mathbb{R} \to \mathbb{R}$ stetig, oder (schwach) monoton steigend, oder (schwach) monoton fallend, so ist g $(\mathfrak{B}, \mathfrak{B})$-meßbar (kurz: Borel-meßbar).

Beweis:

Ist g stetig, so ist $g^{-1}(U)$ offen für alle offenen Mengen, also in $\mathfrak{B}$.
Hieraus folgt die Behauptung mit Satz 5.2.1 und Satz 5.3.2 (Seite 54 und 60).
Monotone g's: Übungsaufgabe. □

Also: Mit X ist auch X^2 (als Verknüpfung der $(\mathfrak{A}, \mathfrak{B})$-meßbaren Abbildungen $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$) eine Zufallsvariable.

Satz 5.4.2 *(Verknüpfungen von Zufallsvariablen)*

1. *Sind X und Y Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$, so liegen die Mengen*
   $\{X < Y\}, \{X \leq Y\}, \{X = Y\}$ *und* $\{X \neq Y\}$ *in* $\mathfrak{A}$.
   (Hierbei steht $\{X < Y\}$ für $\{\omega \in \Omega : X(\omega) < Y(\omega)\}$ etc.)

2. *Sind X und Y Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$, mit $\alpha, \beta \in \mathbb{R}$, so sind auch*
   $\alpha * X + \beta, X + Y, X * Y, X \wedge Y, X \vee Y$
   wieder meßbar (also Zufallsvariablen).

3. *Ist $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$, so sind auch*
   $sup_{n \in \mathbb{N}} X_n$, $inf_{n \in \mathbb{N}} X_n$, $limsup_{n \to \infty} X_n$, $liminf_{n \to \infty} X_n$
   Zufallsvariablen (vorausgesetzt, sie sind $\mathbb{R}$-wertig).
   Gilt $X_n(\omega) \to X(\omega)$ für alle $\omega \in \Omega$, so ist auch X eine Zufallsvariable.

Beweis:

1. **Prüfungsrelevant**
   $\{X < Y\} = \bigcup_{q \in \mathbb{Q}} \{X < q\} \cap \{Y > q\} \in \mathfrak{A}$

   $\supset \subset \quad Y^{-1}(\underbrace{(q, \infty)}_{\in \mathfrak{B}}) \in \mathfrak{A}$

   $\{X \leq Y\} = \{Y < X\}^c$
   $\{X = Y\} = \{X \leq Y\} \cap \{X < Y\}^c$
   $\{X \neq Y\} = \{X = Y\}^c$
   also liegen alle diese Mengen auch in $\mathfrak{A}$.

2.
   $$(\Omega, \mathfrak{A}) \xrightarrow{X} (\mathbb{R}, \mathfrak{B}) \xrightarrow[(1)]{\alpha * + \beta} (\mathbb{R}, \mathfrak{B})$$
   $$\xrightarrow{\alpha * X + \beta}$$

   [1]*: stetig, also $(\mathfrak{B}, \mathfrak{B})$-meßbar*

   Hintereinanderschaltungen von meßbaren Abbildungen sind meßbar (Satz 5.3.3, Seite 61).
   $\{X + Y \leq a\} = \{X \leq a - Y\} \in \mathfrak{A} \forall a \in \mathbb{R}$ nach (1).
   $a - Y$: affine Transformation von Y, also meßbar
   Also: Urbilder eines Erzeugendensystems in $\mathfrak{A}$, somit meßbar.
   Prüfungsrelevant:
   $X * Y = \frac{1}{4} * ((X + Y)^2 - (X - Y)^2)$ (Polarisationstrick)
   Die rechte Seite zeigt, daß $X * Y$ durch Operationen erhalten werden kann, die die Meßbarkeit nicht zerstören, also ist auch $X * Y$ meßbar (Zufallsvariable).
   $\{X \vee Y \leq a\} = \{\underbrace{X \leq a}_{\in \mathfrak{A}}\} \cap \{\underbrace{Y \leq a}_{\in \mathfrak{A}}\} \in \mathfrak{A}$
   $\{X \wedge Y \leq a\} = \{\underbrace{X \leq A}_{\in \mathfrak{A}}\} \cup \{\underbrace{Y \leq a}_{\in \mathfrak{A}}\} \in \mathfrak{A}$
   Auch hier wurde verwendet, daß es reicht zu zeigen, daß die Urbilder eines Erzeugendensystems in $\mathfrak{A}$ liegen.

3. $\{sup_{n \in \mathbb{N}} X_n \leq a\} = \bigcap_{n=1}^{\infty} \{X_n \leq a\} \in \mathfrak{A}$
   (Beachte: hier wurde verwendet, daß die Familie, über die das Supremum gebildet wird, abzählbar ist)
   Damit: $\qquad X_n \; ZG$

$inf_{n\in\mathbb{N}} X_n = -sup_{n\in\mathbb{N}}(-X_n)$ $\Rightarrow -X_n$ ZG
$limsup_{n\to\infty} X_n = inf_{n\in\mathbb{N}} sup_{m\geq n} X_m$ $\Rightarrow sup(-X_n)$ ZG
$liminf_{n\to\infty} X_n = sup_{n\in\mathbb{N}} inf_{m\geq n} X_m$ $\Rightarrow -sup(-X_n)$ ZG
$\Rightarrow inf X_n$ ZG

Konvergiert X_n punktweise gegen X mit $n \to \infty$, so gilt $X = limsup_{n\to\infty} X_n$, also ist X meßbar. □

Bei Teil (3) existieren Varianten, bei denen die Werte $\pm\infty$ erlaubt sind.
Auf $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ erhält man eine σ-Algebra durch
$\mathfrak{B}(\overline{\mathbb{R}}) = \sigma(\mathfrak{B}(\mathbb{R}) \cup \{\{\infty\}, \{-\infty\}\})$
$(0, \infty] = (0, \infty) \cup \{\infty\} \in \mathfrak{B}(\overline{\mathbb{R}})$.

$(\Omega, \mathfrak{A})$ $(\Omega', \mathfrak{A}')$ meßbare Räume.
$X : \Omega \to \Omega'$ heißt $(\mathfrak{A}, \mathfrak{A}')$-meßbar, wenn

$$X^{-1}(A') \in \mathfrak{A} \forall A' \in \mathfrak{A}' \begin{pmatrix} \text{meßbar:} & \text{Maßtheorie} \\ \text{Zufalsgröße:} & \text{Stochastik} \end{pmatrix}$$

Formale Analogie zur mengentheoretischen Topologie:
σ-Algebra $\rightsquigarrow$ System der offenen Mengen
meßbar $\rightsquigarrow$ stetig

Motivation: Ist P ein Wahrcheinlichkeitsmaß auf $(\Omega, \mathfrak{A})$, so soll $P(X \in A')$ für $A' \in \mathfrak{A}$ definiert sein.

$$\{X \in A'\} = X^{-1}(A')$$

5.5 Verteilungsfunktionen

Die Verteilung einer reellen Zufallsgröße (Zufallsvariablen) ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$, also durch seine Werte auf dem $\cap$-stabilen Erzeugendensystem
$\mathfrak{E} = \{(-\infty, x] : x \in \mathbb{R}\}$
festgelegt.

Definition 5.5.1 *(Verteilungsfunktion)*
Die Verteilungsfunktion F zu einem Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B})$ wird definiert durch $F : \mathbb{R} \to \mathbb{R}$, $F(x) := P((-\infty, x])$ $\forall x \in \mathbb{R}$.
Ist P die Verteilung einer Zufallsvariablen X, so nennen wir F die Verteilungsfunktion von X.

Satz 5.5.1 *(Eigenschaften von Verteilungsfunktionen)*
Ist F Verteilungsfunktion (zu einem Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B})$), so gilt:

1. $lim_{x\to-\infty} F(x) = 0$, $lim_{x\to+\infty} F(x) = 1$
2. *F ist (schwach) monoton steigend*
3. *F ist stetig von rechts*

Beweis:

1. Sei $(x_n)_{n\in\mathbb{N}}$ eine gegen $-\infty$ konvergierende Folge ("bestimmte Divergenz")
   (klar: $\forall c<0 \exists n_0 \in\mathbb{N} \forall n \geq n_0 : x_n \leq c$)
   Setze $y_n := sup_{m\geq n} x_m$, klar: $y_n \downarrow -\infty$ $\quad (y_{n+1} \leq y_n)$
   also $(-\infty, y_n] \downarrow \emptyset$, also:
   $$0 \overset{1}{\leq} F(x_n) \overset{2}{\leq} F(y_n) \overset{3}{=} P((-\infty, y_n]) \xrightarrow[n\to\infty]{4} P(\emptyset) = 0$$
   [1]: *Wahrscheinlichkeiten sind* ≥ 0
   [2]: *da* $x_n \leq y_n$
   [3]: *Stetigkeit von oben von P*
   [4]: *Satz 1.2.2, Seite 11, Aufgabe 4(a)*

   $F(x_n) \to 1$ mit $x_n \to \infty$: analog

2. folgt unmittelbar aus der Monotonie von Wahrscheinlichkeitsmaßen (Satz 1.2.1(4), Seite 9):
   $$\begin{array}{rcccc} x<y & \Rightarrow & (-\infty, x] & \subset & (-\infty, y] \\ & \Rightarrow & P((-\infty, x]) & \leq & P((-\infty, y]) \\ & & \| & & \| \\ & \Rightarrow & F(x) & \leq & F(y) \end{array}$$

3. Ist $(x_n)_{n\in\mathbb{N}} \subset\mathbb{R}$ mit $x_n \geq x \ \forall n \in\mathbb{N}$, $x_n \xrightarrow{n\to\infty} x$, so gilt für $y_n := sup_{m\geq n} x_m$
   $y_n \downarrow x$, also
   $$\begin{aligned} \underline{\underline{F(x)}} &= P((-\infty, x]) \\ &\leq P((-\infty, x_n]) \\ &= F(x_n) \\ &\leq F(y_n) \\ &= P((-\infty, y_n]) \\ &\downarrow P((-\infty, x]) \qquad \textit{da Wahrscheinlichkeitsmaße stetig von oben sind.} \\ &= \underline{\underline{F(x)}} \end{aligned}$$

□

Entscheidend ist nun, daß die Eigenschaftsliste vollständig ist.

Hierzu: (**Prüfungsrelevant**)

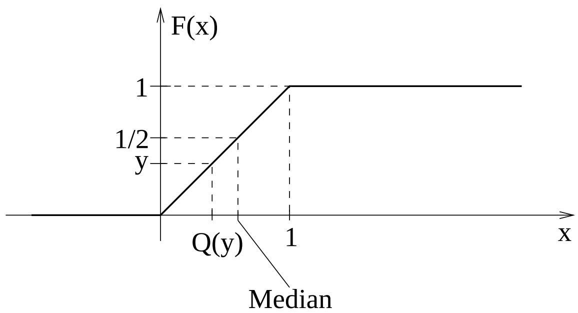
Definition 5.5.2 *Es sei F eine Funktion mit den Eigenschaften (1)-(3) aus Satz 5.5.1 (Seite 64).*
Dann definieren wir die Quantilfunktion Q zu F durch

$$Q : (0,1) \to \mathbb{R},\ Q(y) := inf\{x \in \mathbb{R} : F(x) \geq y\}$$

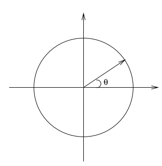
Alternative (aber etwas missverständlichere) Schreibweise:

F^{-1} *statt* Q

Ist X eine Zufallsvariable mit Verteilungsfunktion F, so nennt man $F^{-1}(\alpha)$ $(0 < \alpha < 1)$ das α-Quantil zu(r Verteilung) von X.
$q_\alpha = Q(\alpha) = F^{-1}(\alpha)$ ist der kleinste Wert mit der Eigenschaft, daß X mindestens mit Wahrscheinlichkeit α nicht größer ist.
Im Falle $X \sim unif(0,1)$:

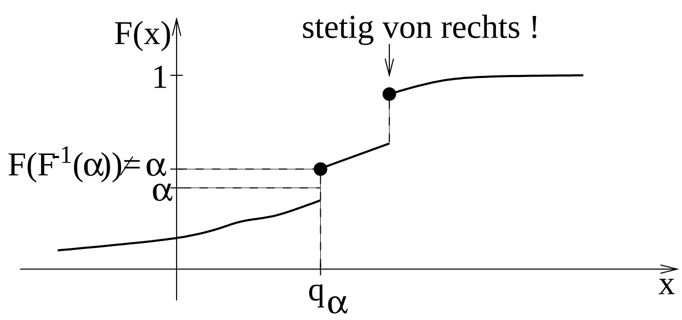


$$F(x) = P(X \leq x) = unif(0,1) = x - 0$$
$$((0,x)) \subset (0,1) \text{ für } 0 < x < 1$$
$$Q(y) = y, 0 < y < 1$$

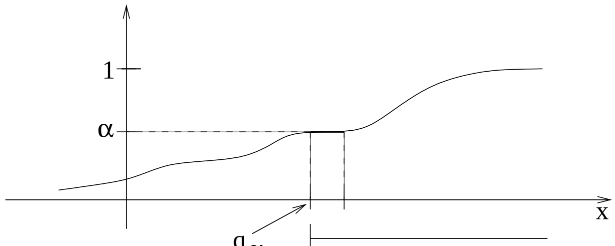


$$\theta = 2\pi x$$
$$0 \leq x < 1$$

F kann "springen"



oder "stückweise konstant" sein.



Nur wenn F streng monoton wachsend und stetig ist, ist F^{-1} die Umkehrfunktion zu F "im üblichen Sinne"

gewöhnlich:
$F(F^{-1}(y)) = y$
hier im allgemeinen $\neq$

Lemma 5.5.1
$y \leq F(x) \Leftrightarrow F^{-1}(y) \leq x$

Beweis:
"$\Rightarrow$" folgt unmittelbar aus der Definition von F^{-1}
Da außerdem

$$\underline{F(x) < y} \Rightarrow F(x + \tfrac{1}{n}) < y \text{ für ein } n \in \mathbb{N} \quad \text{(denn } F \text{ ist stetig von rechts)}$$
$$\Rightarrow F^{-1}(y) \geq x + \tfrac{1}{n} \quad (F \text{ schwach monoton steigend})$$
$$\Rightarrow \underline{F^{-1}(y) > x}$$

gilt, hat man auch die Gegenrichtung.

$A \Rightarrow B$
$\Leftrightarrow$
$\neg B \Rightarrow \neg A$
$\square$

Satz 5.5.2 *(Existenz einer Verteilungsfunktion)*
Es sei $F : \mathbb{R} \to \mathbb{R}$ *eine Funktion mit den Eigenschaften (1)-(3) aus Satz 5.5.1 (Seite 64). Dann existiert ein Wahrscheinlichkeitsmaß* P *auf* $(\mathbb{R}, \mathfrak{B})$ *mit Verteilungsfunktion* F.

Beweis:
Es sei $\Omega = (0,1)$, $\mathfrak{A} = \mathfrak{B}_{(0,1)}$ *und* $P_0 = unif(0,1)$
Definiere $X : \Omega \to \mathbb{R}$ *durch*
$$X(\omega) := F^{-1}(\omega) \ \forall \omega \in \Omega$$
Da F^{-1} *monoton ist, ist* X *eine Zufallsvariable, deren Verteilung somit ein Wahrscheinlichkeitsmaß* P *auf* $(\mathbb{R}, \mathfrak{B})$ *ist.*

Dann gilt für alle $x \in \mathbb{R}$:

$$\begin{aligned} P((-\infty,x]) &= P_0(X \leq x) \\ &= P_0(\{\omega \in \Omega : X(\omega) \leq x\}) \\ &= P_0(\{\omega \in \Omega : F^{-1}(\omega) \leq x\}) \\ &= P_0(\{\omega \in \Omega : \omega \leq F(x)\}) \\ &= P_0((0, F(x)]) && \leftarrow \textit{dies ist ein Intervall} \\ &= F(x) - 0 \\ &= F(x) \end{aligned}$$

□

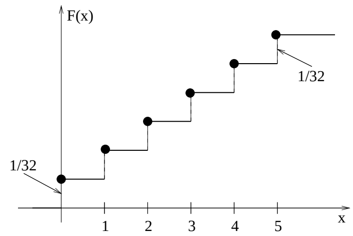
Ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B})$ *wird vollständig beschrieben durch seine Verteilungsfunktion*

$F: \mathbb{R} \to \mathbb{R}$, $F(x) := P((-\infty, x])$.

Was passiert bei <u>diskreten</u> Zufallsvariablen ?

Klar: Die Verteilung $\mathcal{L}(X)$ einer diskreten Zufallsvariablen X ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$.

Typisch: 5*-facher Münzwurf:*

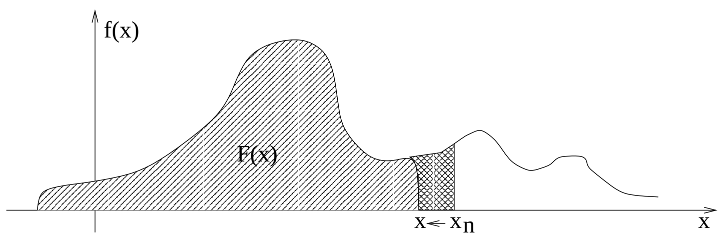


Solche Verteilungsfunktionen sind "von reinem Sprungtyp", d.h. konstant zwischen den x-Werten, die X mit Wahrscheinlichkeit > 0 annimmt.

Ist $f: \mathbb{R} \to \mathbb{R}_+$ eine Funktion mit $\int_{-\infty}^{\infty} f(x)dx = 1$ (Riemann-Integral), so wird durch

$F(x) := \int_{-\infty}^{\infty} f(y)dy$

eine Verteilungsfunktion (und damit ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B})$*) definiert.*



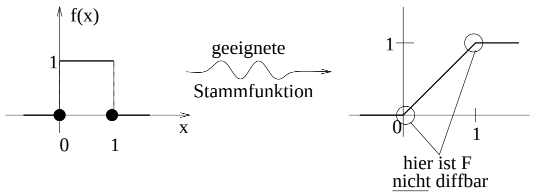
f heißt <u>Wahrscheinlichkeitsdichte</u> oder <u>Dichtefunktion</u> zu P.

<u>Schlechte</u> Sprachweise: eine Zufallsvariable X, deren Verteilung eine solche (Riemann-)Dichte aufweist, nennt man stetig. (ist Abkürzung für: X ist absolutig verteilt) (im Gegensatz zu <u>diskret</u>)

Ist f in x stetig, so gilt $\boxed{f(x) = F'(x)}$

Beispiel 5.5.1 *Im Falle* $P = unif(0,1)$ *hat man*

$$P((-\infty,x]) = \int_{-\infty}^{x} f(y)dy \ \ mit\ f(x) = \begin{cases} 1 & ,0 < x < 1 \\ 0 & ,sonst \end{cases}$$



Beachte:

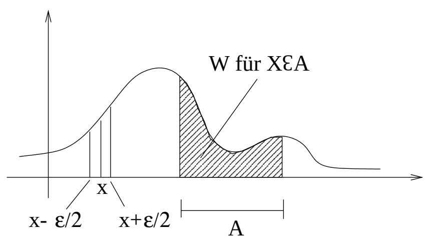
$f(x) = \begin{cases} 1 & ,0 \leq x \leq 1 \\ 0 & ,sonst \end{cases}$ *führt auf dieselbe Verteilungsfunktion* F*, also auf dasselbe Wahrscheinlichkeitsmaß .*

Konsequenz:

Dichtefunktionen sind im Allgemeinen nicht eindeutig.

Dichten können als infinitesimales Gegenstück zu Wahrscheinlichkeitsmassenfunktionen betrachtet werden, aber der Vergleich hinkt: so kann f *beispielsweise durchaus Werte* > 1 *annehmen.*

Allgemeiner: $P(X \in A) = \int_A f(y)dy$ — *entspricht der Definition mit* $A = (-\infty, x]$



W *=Fläche unter der Dichte*

$$P(x - \tfrac{\epsilon}{2} < X < x + \tfrac{\epsilon}{2}) = \int_{x-\frac{\epsilon}{2}}^{x+\frac{\epsilon}{2}} f(y)dy \approx \epsilon * f(x) \Leftrightarrow \tfrac{1}{\epsilon} * \int_{x-\frac{\epsilon}{2}}^{x+\frac{\epsilon}{2}} f(y)dy \approx f(x)$$

wenn f *stetig in* x.

im diskreten Fall:
$p(x) = P(X = x) \leq 1$
$p(x)$ *Massenfunktion*

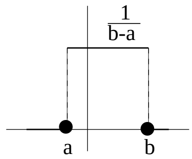
5.6 Einige wichtige Verteilungen mit Riemann-Dichten

5.6.1 Rechteckverteilung, Gleichverteilung

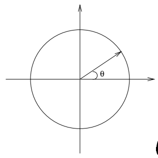
Generelle Voraussetzung: $-\infty < a < b < \infty$

$$f_{a,b} : \mathbb{R} \to \mathbb{R}_+,\ f_{a,b}(x) := \begin{cases} \frac{1}{b-a} & ,a < x < b \\ 0 & ,sonst \end{cases}$$

hat die Eigenschaften $f_{a,b} \geq 0$, $\int_{-\infty}^{\infty} f_{a,b}(y)dy = 1$, *ist also die Dichte eines Wahrscheinlichkeitsmaßes auf* $(\mathbb{R}, \mathfrak{B})$.



Dieses wird Rechteckverteilung auf (a, b) *oder auch Gleichverteilung auf* (a, b) *genannt.*
Schreibweise: $unif(a, b)$



Dies verallgemeinert $unif(0, 1)$

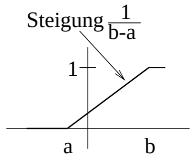
$\theta = 2 * \pi * x, \quad 0 \leq x < 1$ *für den Winkel selbst erhält man* $unif(0, 2 * \pi)$

Alle diese Verteilungen gehen durch affine Transformationen auseinander hervor:

$X \sim unif(0, 1)$

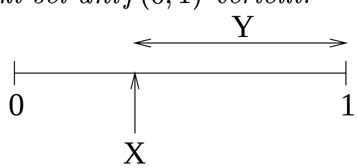
$Y := a + \underbrace{(b - a)}_{>0} * X$

$$P(Y \leq y) = P(a + (b - a) * X \leq y) = P(X \leq \tfrac{y-a}{b-a}) = \begin{cases} \frac{y-a}{b-a} & , \text{ wenn } 0 < \frac{y-a}{b-a} < 1 \\ 0 & , \text{ wenn } \frac{y-a}{b-a} \leq 0 \\ 1 & , \text{ wenn } \frac{y-a}{b-a} \geq 1 \end{cases}$$

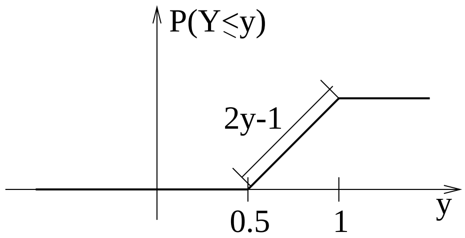


also: $Y \sim unif(a, b)$

Beispiel 5.6.1 *Ein Stab der Länge* 1 *zerbricht an einer zufälligen Stelle.*
Annahme: alle Bruchpositionen sind gleich wahrscheinlich,
genauer: Abstand X *des Bruchpunkts vom unteren Endpunkt sei* $unif(0, 1)$*-verteilt.*



Sei Y *die Länge des längeren Bruchstücks.*
Klar: wieder eine Zufallsvariable.
Welche Verteilung hat y *?*
Auch klar: $P(Y \leq y) = 0$ *für* $y < \frac{1}{2}$, $= 1$ *für* $y \geq 1$

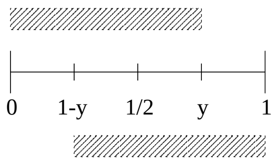


Sei $\frac{1}{2} < y < 1$.
Dann gilt:

$$\begin{aligned} P(Y = y) &= P(max\{X, 1 - X\} \leq y) \\ &= P(X \leq y \ \underline{und} \ \underbrace{1 - X \leq y}_{X \geq 1-y}) \\ &= P(1 - y \leq X \leq y) \\ &= P(X \in [1 - y, y]) \\ &= P(X \in \underbrace{(1 - y, y]}_{*}) \ (denn \ P(X = x) = 0 \ \forall x \in \mathbb{R}) \\ &= 2 * y - 1 \end{aligned}$$

$*$)*längs* $2 * y - 1$, *außerdem:* $(1 - y, y] \subset [0, 1]$

<u>*Also:*</u> $Y \sim unif(\frac{1}{2}, 1)$



5.6.2 Gammaverteilung

Die <u>*Gammaverteilung*</u> *mit Parametern* α *und* λ *(beide* > 0*), ist die Verteilung mit Dichte*

$$f_{\alpha,\lambda}(x) := \begin{cases} \frac{1}{\Gamma(\alpha)} * x^{\alpha-1} * \lambda^{\alpha} * e^{-\lambda * x} & , x > 0 \\ 0 & , sonst \end{cases}$$

mit $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} * e^{-x} dx$ *(Gamma-Funktion)* $(\Gamma(n) = (n-1)!)$
Schreibweise: $X \sim \Gamma(\alpha, \lambda)$
mit α *: "Formparameter" und* λ *: "Skalierungsparameter"*

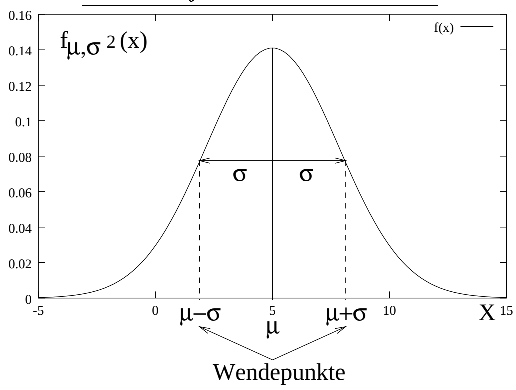
<u>*Besonders wichtig:*</u> $\alpha = 1$, <u>*Exponentialverteilungen*</u> *(siehe Aufgabe 26)*

5.6.3 Normalverteilung

Die Normalverteilung mit Parametern μ und σ^2, kurz: $N(\mu,\sigma^2)$, mit $\mu\in\mathbb{R}$, $\sigma^2>0$, hat die Dichte

$$\phi_{\mu,\sigma^2}(x) := \frac{1}{\sqrt{2*\pi*\sigma^2}} * exp(-\frac{1}{2*\sigma^2} * (x-\mu)^2), x\in\mathbb{R}$$

Die Gaußsche Glockenkurve



μ und σ beschreiben Lage und Breite von ϕ.
Im Falle $\mu=0, \sigma^2=1$ spricht man von der Standardnormalverteilung
Klar: $\phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma} * \phi_{0,1}(\frac{x-\mu}{\sigma})$ für alle $x\in\mathbb{R}$
Dichte zu $N(0,1)$: $\phi_{0,1}(x) = \frac{1}{\sqrt{2*\pi}} * e^{-\frac{x^2}{2}}$
Die zugehörige Verteilungsfunktion (zu $N(0,1)$) ist

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2*\pi}} * e^{-\frac{y^2}{2}} dy$$

Φ ist vertafelt und in den üblichen Softwarepaketen enthalten. (z.B. im Fuchs)
Besonders wichtig sind die α-Quantile μ_α für bestimmte Standardwerte von α.
($\Phi(\mu_\alpha)=\alpha$)

α	0.9	0.95	0.975	0.99	0.995
μ_α	1.2816	1.6449	1.9600	2.3263	2.5758

(±0.00005)

Lemma 5.6.1

1. $\int_{-\infty}^{\infty} \phi_{\mu,\sigma^2}(x)dx = 1 \ \forall \mu\in\mathbb{R}, \sigma^2>0$
2. $\Phi(x) = 1-\Phi(-x)$
3. $X \sim N(\mu,\sigma^2), a\neq 0, b\in\mathbb{R} \Rightarrow Y := a*X+b \sim N(a*\mu+b, a^2*\sigma^2)$

Beweis:

1. *Die Substitution $y := \frac{1}{\sigma}*(x-\mu)$ zeigt, daß es reicht $\mu=0, \sigma^2=1$ zu behandeln. Hierzu:*

$$
\begin{aligned}
(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} exp(-\tfrac{1}{2}*(x^2+y^2))dxdy \\
&= \int_0^{2*\pi} \underbrace{\int_0^{\infty} r*e^{-\frac{r^2}{2}}dr}_{-\frac{d}{dr}e^{-\frac{r^2}{2}}} d\phi
\end{aligned}
$$

(Transformation in Polarkoordinaten
$dxdy = rdrd\phi$)

$$
\begin{aligned}
&= \int_0^{2*\pi}(-e^{\frac{r^2}{2}})|_0^{\infty}d\phi \\
&= 2*\pi
\end{aligned}
$$

2. *folgt mit $\phi_{0,1}(-x) = \phi_{0,1}(x)$*

3. *Im Falle $a > 0$ hat man*
$P(Y \leq y) = P(X \leq \frac{y-b}{a})$
$= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2*\pi*\sigma^2}} * exp(-\frac{1}{2}*(x-\mu)^2)dx$ *Substitution: $a*x+b \leadsto x'$*
$= \int_{-\infty}^{y} \underbrace{\frac{1}{\sqrt{2*\pi*\sigma^2*a^2}} * exp(-\frac{1}{\sqrt{2*\pi*\sigma^2*a^2}} * (x'-(a*\mu+b))^2)}_{\text{Dichte zu } N(a*\mu+b,\sigma^2*a^2)} dx'$
Die Verteilungsfunktion zu Y ist Stammfunktion zur Dichte von $N(a*\mu+b, a^2*\sigma^2)$, d.h. Y hat, wie behauptet, die Verteilung $N(a*\mu+b, a^2*\sigma^2)$.

□

Kantormenge
$X \not\in \mathbb{N}\mathbb{R}$ $X \in [0,1]$

5.7 Erwartungswerte

Erinnerung: im diskreten Fall, also bei einer diskreten Zufallsvariablen X mit Wahrscheinlichkeitsmassenfunktion $p(x) = P(X = x)$ ist

$\boxed{EX = \sum_x x * p(x)}$

(vorausgesetzt, es gilt $\sum |x| * p(x) < \infty$) und in Verallgemeinerung hiervon

$Eg(X) = \sum_x g(x) * p(x)$

Wie sieht dies bei "stetigen" Zufallsvariablen mit Wahrscheinlichkeitsdichtefunktion f aus ?
Die Präzise Definition erfordert das Lebesgue-Integral ($\leadsto$ Stochastik II)

Heuristik:
Es sei

$\underline{X}_n := 2^{-n}\lfloor 2^n * X\rfloor$, $\overline{X}_n := 2^{-n} * \lceil 2^n * X\rceil$.

Diese Zufallsvariablen sind offensichtlich diskret und es gilt

$\underline{X}_n \leq X \leq \overline{X}_n$.

Die Monotonie führt auf die Forderung

$E\underline{X}_n \leq \underbrace{EX}_{?} \leq E\overline{X}_n$

$X = \cdot\underline{1100010}100$
$\underline{X_n} = \cdot 1100010^n 000$

$$\underline{\underline{E\underline{X}_n}} = \sum_{k\in\mathbb{Z}} \frac{k}{2^n} * P(\underline{X}_n = \frac{k}{2^n})$$
$$= \sum_{k\in\mathbb{Z}} \frac{k}{2^n} * P(\frac{k}{2^n} \leq X \leq \frac{k+1}{2^n})$$
$$= \sum_{k\in\mathbb{Z}} \frac{k}{2^n} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f(x)dx$$
$$= \sum_{k\in\mathbb{Z}} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} \frac{k}{2^n} f(x)dx \leq \sum_{k\in\mathbb{Z}} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} x * f(x)dx$$
$$= \underline{\underline{\int_{-\infty}^{\infty} x * f(x)dx}} \leq \sum_{k\in\mathbb{Z}} \frac{k+1}{2^n} * \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f(x)dx$$
$$= \sum_{k\in\mathbb{Z}} \frac{k+1}{2^n} * P\underbrace{(\frac{k}{2^n} < X \leq \frac{k+1}{2^n})}_{\Leftrightarrow \overline{X}_n = \frac{k+1}{2^n}}$$
$$= \underline{\underline{E\overline{X}_n}}$$

Klar: $\underline{X}_n \uparrow X$, $\overline{X}_n \downarrow X$, *also führt dies auf*
$\boxed{EX = \int x * f(x)dx}$ *Analogie:* $\boxed{Eg(X) = \int g(x) * f(x)dx}$
vorausgesetzt, $\int |x| * f(x)dx < \infty$, g *meßbar,* $\int |g(x)| * f(x)dx < \infty$

Beispiel 5.7.1
Im Falle $X \sim N(\mu, \sigma^2)$ *erhählt man*

$$EX = \int x * \frac{1}{\sqrt{2*\pi*\sigma^2}} * exp(-\frac{1}{2*\sigma^2} * (x-\mu)^2)dx$$
$$= \int (x-\mu) * \frac{1}{\sqrt{2*\pi*\sigma^2}} * exp(-\frac{1}{2*\sigma^2} * (x-\mu)^2)dx\ \} = 0, \text{ da antisymmetrisch um } \mu$$
$$+ \mu * \int \underbrace{\frac{1}{\sqrt{2*\pi*\sigma^2}} * exp(-\frac{1}{2*\sigma^2} * (x-\mu)^2)\,dx}_{Wahrscheinlichkeitsdichte\ zu\ N(\mu,\sigma^2)}$$
$$= \mu,$$

der erste Parameter bei $N(\mu, \sigma^2)$ *ist also gerade der Erwartungswert.*
Analog: $\sigma^2 = var(X)$, *wenn* $X \sim N(\mu, \sigma^2)$

5.8 Unabhängigkeit

σ-Algebren (bisher nur "notwendiges Übel") spielen auch bei stochastischer Unabhängigkeit und als Repräsentanten von Teilinformationen eine Rolle.

Satz 5.8.1 *(und Definition)*
Es sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit Werten in $(\Omega', \mathfrak{A}')$.Dann ist $\{X^{-1}(A) : A \in \mathfrak{A}'\}$ eine σ-Algebra, die von X <u>erzeugte</u> σ-Algebra.
Schreibweise: $\sigma(X)$

$$(\Omega, \mathfrak{A}, P)$$
$$\downarrow X$$
$$(\Omega', \mathfrak{A}')$$

Ist $\mathfrak{E}'$ ein $\cap$-stabiles Erzeugendensystem zu $\mathfrak{A}'$, so ist $\{X^{-1}(E') : E' \in \mathfrak{E}'\}$ ein $\cap$-stabiles Erzeugendensystem von $\sigma(X)$.

Beweis: *Übungsaufgabe.* □

$$\Omega = \left\{ \begin{array}{c} (1,1),(1,2),\cdots,(1,6) \\ (2,1),(2,2),\cdots,(2,6) \\ \vdots \\ (6,1),(6,2),\cdots,(6,6) \end{array} \right\} \xrightarrow{X} \{2,3,\cdots,12\} = \Omega'$$

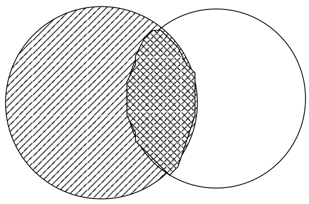
$\mathfrak{A}' = \mathfrak{P}(\Omega')$

$\{(1,1)\} \in \sigma(X)$, $\{(1,2),(2,1)\}(= X^{-1}(\{3\})) \in \sigma(X)$

$\{(1,2)\} \notin \sigma(X)$

$\sigma(x)$ repräsentiert die "in X enthaltene Information"

Bekannt: mit A und B sind auch A^c und B^c, A^c und B, A und B^c,$\cdots$ unabhängig.



$$\begin{aligned} \underline{P(A \cap B^c)} &= P(A) - P(A \cap B) \\ &= P(A) - P(A) * P(B) \\ &= P(A) * (1 - P(B)) \\ &= \underline{P(A) * P(B^c)} \end{aligned}$$

$$P(A \cap \Omega) = P(A) = P(A * 1) = P(A) * P(\Omega)$$

Mit B sind automatisch alle Elemente von $\{\emptyset, B, B^c, \Omega\} = \underline{\sigma(\{B\})}$ von A unabhängig.

Definition 5.8.1 *(Der "offizielle" Unabhängigkeitsbegriff)*
Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum.

1. *Eine Familie $\{\mathfrak{A}_i : i \in I\}$ von Unter-σ-Algebren von $\mathfrak{A}$ heißt stochastisch unabhängig, wenn für jede endlich Teilmenge $J = \{j_1, \cdots, j_n\}$ von I und alle $A_{j_1} \in \mathfrak{A}_{j_1}, \cdots, A_{j_n} \in \mathfrak{A}_{j_n}$ gilt:*
   $$\boxed{P(\bigcap_{i \in J} A_j) = \prod_{j \in J} P(A_j)} \qquad \heartsuit$$
2. *ist für jedes $i \in I$ X_i eine Zufallsgröße auf $(\Omega, \mathfrak{A}, P)$ mit Werten in $(\Omega_i, \mathfrak{A}_i)$, so heißt die Familie $\{X_i : i \in I\}$ stochastisch unabhängig (kurz: die $X_i, i \in I$ sind unabhängig), wenn die Familie $\{\sigma(X_i) : i \in I\}$ im Sinne von (1) unabhängig ist.*

Satz 5.8.2 *Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum, $I \neq \emptyset$, und für jedes $i \in I$ sei $\mathfrak{A}_i$ eine Unter-σ-Algebra von $\mathfrak{A}$ mit $\cap$-stabilen Erzeugendensystem $\mathfrak{E}_i$.*
Gilt dann

$$P(\bigcap_{j \in J} E_j) = \prod P(E_j)$$

für alle endlichen $J = \{j_1, \cdots, j_n\} \subset I$ *und*
alle $E_{j_1} \in \mathfrak{E}_{j_1}, \cdots, E_{j_n} \in \mathfrak{E}_{j_n}$*, so sind* $\mathfrak{A}_i, i \in I$*, unabhängig.*

Beweis:

Sei $J = \{j_1, \cdots, j_n\} \subset I$*. Sei* $\mathcal{D}_{j_1}$ *die Menge alle* $A \in \mathfrak{A}_{j_1}$
mit $P(A)^c \cap E_{j_2} \cap \cdots \cap E_{j_n}) = P(A^c) * P(E_{j_2}) * \cdots * P(E_{j_n})$
Klar: $\mathfrak{E}_{j_1} \subset \mathcal{D}_{j_1}$
Klar: $\mathcal{D}_{j_1}$ *Dynkin-System*
Nach Satz 5.2.3(2) (Seite 57) gilt also $\mathcal{D}_{j_1} = \mathfrak{A}_{j_1}$
("die erste E-Menge darf durch A-Mengen ersetzt werden").
Im zweiten Schritt sei $\mathcal{D}_{j_2}$ *die Menge aller* $A \in \mathfrak{A}_{j_2}$ *mit*

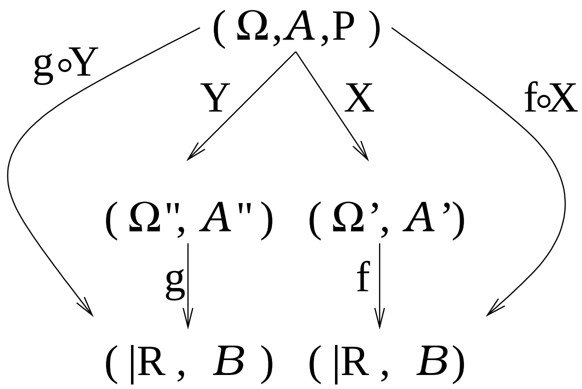
$$P(A_{j_1} \cap A \cap E_{j_3} \cap \cdots \cap E_{j_n}) = P(A_{j_1}) * P(A) * P(E_{j_3}) * \cdots * P(E_{j_n})$$

Wie oben: $\mathcal{D}_{j_2} = \mathfrak{A}_{j_2}$*, nach insgesamt n Schritten dieser Art hat man*

$$P(A_{j_1} \cap \cdots \cap A_{j_n}) = P(A_{j_1}) * \cdots * P(A_{j_n})$$

für alle $A_{j_1} \in \mathfrak{A}_{j_1}, \cdots, A_{j_n} \in \mathfrak{A}_{j_n}$ □

Bei einer diskreten Zufallsgröße X bilden die Mengen $X^{-1}(\{x\}), x \in Bild(X)$*, ein* $\cap$*-stabiles Erzeugendensystem von* $\sigma(X)$ *(man muß* $\emptyset$ *zusätzlich aufnehmen), Satz 4.4.2 (Seite 39) zeigt also, daß Teil (2) von Definition 5.8.1 (Seite 75) zu Definition 4.4.1 (Seite 39) "abwärtskompatibel" ist. Der Zugang über* σ*-Algebren bietet Vorteile, beispielsweise bei Beweisen:*



Satz 5.8.3 *(Funktionen unabhängiger Größen sind unabhängig)*
Für jedes $i \in I$ *seien* X_i *eine Zufallsgröße mit Werten in* $(\Omega_i, \mathfrak{A}_i)$*,* $(\Omega_i', \mathfrak{A}_i')$ *ein weiterer meßbarer Raum und* $g_i : \Omega_i \to \Omega_i'$ *eine* $(\mathfrak{A}_i, \mathfrak{A}_i')$*-meßbare Abbildung.*
Ist dann $\{X_i : i \in I\}$ *eine unabhängige Familie von Zufallsgrößen* (X_i *mit Werten in* Ω_i*), so ist auch* $\{Y_i : i \in I\}$ *mit* $Y_i := g_i(X_i)$ *unabhängig.*
"In Worten": Funktionen unabhängiger Größen sind unabhängig.

Beweis: $\sigma(Y_i) \subset \sigma(X_i)$ □

$(\Omega, \mathfrak{A}, P)$ $\{X_i : i \in I\}$ *Zufallsgröße.* $X_i : (\Omega, \mathfrak{A}, P) \to (\Omega_i, \mathfrak{A}_i)$
Unabhängigkeit über $\sigma(X_i) = \{X_i^{-1}(A_i) : A_i \in \mathfrak{A}_i\}$

$$P(X_{i_j} \in A_{i_j} \text{ für } j = 1, \cdots, n) = \prod_{j=1}^n P(X_{i_j} \in A_{i_j})$$

Beispiel 5.8.1

Es sei $(\Omega, \mathfrak{A}, P) = ([0,1), \mathcal{B}([0,1)), unif(0,1))$

Für jedes $n \in \mathbb{N}$ *werde* $X_n : \Omega \to \{0,1\}$ *definiert durch*

$$X_n(\omega) := \lfloor 2^n \ \omega \rfloor - 2 * \lfloor 2^{n-1} \ \omega \rfloor$$

$$011001000\overbrace{\underbrace{1}_{X_n}}^{n=9}|10\cdots$$

Dann gilt

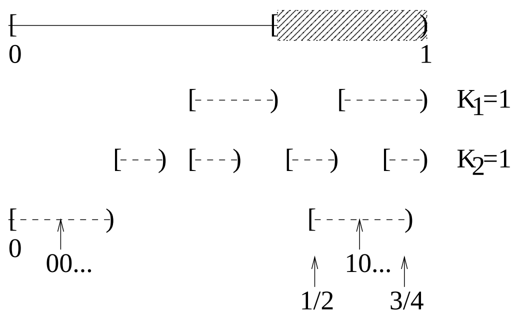
$$\omega = \sum_{n=1}^{\infty} 2^{-n} X_n(\omega),$$

d.h. die Folge $0.X_1(\omega)X_2(\omega)\cdots$ *ist (im wesentlichen) die Binärdarstellung von* ω.

$(0.1999\cdots = 0.2)$

Für alle $k_1, \cdots, k_n \in \{0,1\}$ *gilt*

$$P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n)$$
$$= P(\textstyle\sum_{l=1}^{n} 2^{-l} k_l \leq \omega < \sum_{l=1}^{n} 2^{-l} k_l + 2^{-n})$$
$$= 2^{-n}$$



Für beliebige $i_1 < i_2 < \cdots < i_n$ *erhält man somit*

$$P(\underbrace{X_{i_1} = 1, \cdots, X_{i_n} = 1}_{\substack{Ein\ Ereignis, das\ von \\ den\ Positionen\ 1\cdots i_n\ abhängt}})$$
$$= \sum_{\substack{(k_1,\cdots,k_{i_n}) \in \{0,1\}^{i_n} \\ k_{i_j}=1\ für\ j=1,\cdots,n}} P(X_1 = k_1, \cdots, X_{i_n} = k_{i_n})$$
$$= 2^{-i_n} * \#\{(k_1, \cdots, k_n) \in \{0,1\}^{i_n} : k_{i_j} = 1 \ für \ j = 1, \cdots, n\}$$
$$= 2^{-i_n} * 2^{i_n - n}$$
$$= \underline{\underline{2^{-n}}}$$

"So ungern ich mich selbst kritisiere, aber hier muß ich es wirklich mal tun"

$$P(A \cap B) = P(A) * P(B)$$
$$P(\textstyle\bigcap_{j \in J} A_{i_j}) = \prod_{j \in J} P(A_{i_j})$$
$$\{X_{i_j} = 1\}$$
$$i_1 = 5, i_2 = 7, i_3 = 13$$

1	0
2	0
⋮	⋮
5	1
6	0

7	1
	$\vdots$
13	1

Insbesondere: $P(X_{i_j} = 1) = \frac{1}{2}$ *und damit insgesamt:*

$$P(X_{i_1} = 1, X_{i_2} = 1, \cdots, X_{i_n} = 1) = P(X_{i_1} = 1) * P(X_{i_2} = 1) * \cdots * P(X_{i_n} = 1)$$

Da $\{X_i^{-1}(\{1\})\}$ *ein* $\cap$*-stabiles Erzeugendensystem von* $\sigma(X_i)$ *ist, folgt hieraus, daß die Zufallsvariablen* $X_1, X_2, \cdots$ *unabhängig sind.*
Außerdem gilt: $\mathcal{L}(X_i) = Bin(1, \frac{1}{2})$, *die gesamte Konstruktion kann also als Modell für den* ∞ *oft wiederholten Wurf einer fairen Münze dienen.*
Umgekehrt ließe sich aus einer unendlichen Folge von Münzwürfen $k_1, k_2, \cdots$ *eine auf* $[0, 1)$ *gleichverteilte Zahl* x *durch* $x := \sum_{i=1}^{\infty} k_i * 2^{-i}$ *konstruieren (!).*

"Jetzt habe ich gesagt, es wäre leicht zu durchschauen und durschaue es selbst nicht"

Sind X, Y *unabhängige Zufallsvariablen mit Verteilungsfunktionen* F_X, F_Y, *so gilt für die gemeinsame Verteilungsfunktion* $F_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$

$$\boxed{\begin{aligned} F_{X,Y}(x,y) :=\; & P(X \leq x, Y \leq y) \\ =\; & P(X \leq x) * P(Y \leq y) \\ =\; & F_X(x) * F_Y(y) \end{aligned}} \qquad \heartsuit$$

für alle $x, y \in \mathbb{R}$:

die gemeinsame Verteilungsfunktion ist das Produkt der einzelnen - **BEI UNABHÄNGIGKEIT !**

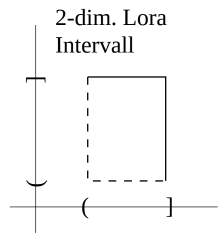
Umgekehrt folgt aus ($\heartsuit$*) die Unabhängigkeit von* X *und* Y, *da* $\{(-\infty, a] : a \in \mathbb{R}\}$ *ein* $\cap$*-stabiles Erzeugendensystem von* $\mathcal{B}(\mathbb{R})$ *ist (Satz 5.2.1 (Seite 54), Satz 5.8.2 (Seite 75)).*
Sind X *und* Y *stetige Zufallsvariablen mit Dichten* f_X, f_Y, *also*

$$F_X(x) = \int_{-\infty}^{x} f_X(z)dz, \; F_Y(y) = \int_{-\infty}^{y} f_Y(z)dz,$$

so erhält man bei Unabhängigkeit

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_Y(u) * f_X(v) * du * dv \qquad = F_X * F_Y$$

In naheliegender Verallgemeinerung des eindimensionalen Falls nennt man $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ *eine gemeinsame Dichte von* X *und* Y, *wenn* $P((X,Y) \in A) = \iint_A f_{X,Y}(x,y) * dy * dx$ *für "hinreichend viele"* $A \subset \mathbb{R}^2$ *gilt. (genaueres in Stochastik II)*



Insbesondere hat man bei Unabhängigkeit $f_{X,Y}(x,y) = f_X(x) * f_Y(y)$, *wenn* X, Y *stetig.*

Mit gemeinsamen Dichtefunktionen lassen sich Erwartungswerte ausrechnen, die sich auf mehrere Zufallsvariablen beziehen.
Erinnerung an den diskreten Fall:

Sind X, Y *diskrete Zufallsvariablen mit gemeinsamer Massenfunktion*

$p_{X,Y}(x,y) = P(X = x, Y = y)$, *so gilt*

$$\boxed{Eg(X,Y) = \sum_{x \in Bild(X)} \sum_{y \in Bild(Y)} g(x,y) * p_{X,Y}(x,y)}$$

(vorausgesetzt, die Summe konvergiert absolut).
Ganz analog gilt in dieser Situation

$$\boxed{Eg(X,Y) = \iint g(x,y) * f_{X,Y}(x,y) * dx * dy}\ !$$

(man braucht Meßbarkeit bei g etc.,$\cdots$ Stochastik II)
Hiermit erhählt man unter anderem eine Variante der Multiplikationsregel (Satz 4.5.1 (Seite 40)) für unabhängige <u>*stetige*</u> *Zufallsvariablen:*

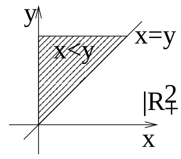
$$\begin{aligned}\boxed{EXY} &= \iint x * y * f_X(x) * f_Y(y) * dy * dx \qquad (g(x,y) = x * y)\\ &= \int x * (\underbrace{\int y * f_Y(y) * dy}_{EY}) * f_X(x) * dx\\ &= (EY) * \int x * f_X(x) dx\\ &= \boxed{(EX) * (EY)}\end{aligned}$$

$EXY = EX * EY$ gilt also bei Unabhängigkeit

1. *für diskrete Zufallsvariablen*
2. *für stetige Zufallsvariablen*

Das Lebesgue-Integral erlaubt es, (1) und (2) als Spezialfälle einer allgemeineren Konstruktion zu betrachten und erlaubt auch die Behandlung von Zufallsvariablen, die weder diskret noch stetig sind.
Weitere Analogiebetrachtungen führen auf die Faltung von Wahrscheinlichkeitsdichten, siehe beispielsweise Stundenübung 29.

Beispiel 5.8.2
Die Lebensdauer X einer Glühbirne Marke A sei exponential verteilt mit Parameter λ_A, Y sei die Lebensdauer einer Typ-B-Birne, exponential verteilt mit Parameter λ_B.
Wir setzen voraus, daß X und Y unabhängig sind.
Mit welcher Wahrscheinlichkeit brennt eine Typ-B-Birne länger als eine A-Birne ?



$$\begin{aligned}P(X < Y) &= P((X,Y) \in \{(x,y) \in \mathbb{R}^2_+ : x < y\})\\ &= \iint_{\{(x,y) \in \mathbb{R}^2_+ : x < y\}} \underbrace{f_{X,Y}(x,y) * dy * dx}_{f_X(x) * f_Y(y) = \lambda_A * \lambda_B * e^{-\lambda_A * x} * e^{-\lambda_B * y}}\\ &= \underbrace{\int_0^\infty \int_x^\infty \lambda_B * e^{-\lambda_B * y} * dy * \lambda_A * e^{-\lambda_A * x} * dx}_{Stammfunktion: -e^{-\lambda_B * y}}\\ &= \int_0^\infty e^{-\lambda_B * x} * \lambda_A\ e^{-\lambda_A * x} * dx\\ &= \lambda_A * \int_0^\infty \underbrace{e^{-(\lambda_A + \lambda_B) * x}}_{-\frac{1}{\lambda_A + \lambda_B} * e^{-(\lambda_A + \lambda_B) * x}} * dx \qquad \text{“Bischen trivial, nech”}\end{aligned}$$

$$= \frac{\lambda_A}{\lambda_A + \lambda_B}$$

(Bei $\lambda_A = \lambda_B$: $P(X < Y) = P(Y < X),\ P(X = Y) = 0$)

6 Verteilungskonvergenz und Normalapproximation

6.1 Einführung

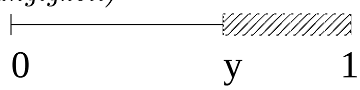
Sei X eine Zufallsvariable auf $(\Omega, \mathfrak{A}, P)$, $\mathcal{L}(X)$: Verteilung von X.
(z.B. Augensumme beim 100 fachen Würfelwurf)
Es geht um Approximation zu $\mathcal{L}(X)$, also: <u>Verteilungskonvergenz</u>

- *Gesetz der seltenen Ereignisse, führt auf Poissonverteilungen ($Bin(n,p) \approx Poisson(n * p)$) wenn n groß , p klein*
- *Extremwertverteilungen*
- *<u>Zentraler Grenzwertsatz</u>*

6.2 Verteilungskonvergenz

Beispiel 6.2.1 *Die Zufallsvariablen $X_1, \cdots, X_n$ seien unabhängig und $unif(0,1)$-verteilt;*
$Y_n := min\{X_1, \cdots, X_n\}$
Die zugehörige Verteilungsfunktion ist $(0 \leq y \leq 1)$

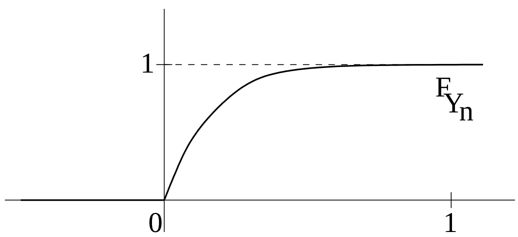
$$\begin{aligned} F_{Y_n}(y) &= P(Y_n \leq y) \\ &= 1 - P(X_1 > y, \cdots, X_n > y) \\ &= 1 - P(X_1 > y) * \cdots * P(X_n > y) \quad \textit{(Unabhängigkeit)} \\ &= 1 - (1-y)^n \end{aligned}$$



<u>Klar:</u> $F_{Y_n}(y) = 0$ für $y < 0$, 1 für $y > 1$

$$\textit{Mit } n \to \infty: \ F_{Y_n}(y) = \begin{cases} 1 & , y > 1 \\ 1-(1-y)^n & , 0 \leq y \leq 1 \\ 0 & , y < 0 \end{cases} \xrightarrow{n \to \infty} \begin{cases} 1 & , y > 1 \\ 1 & , 0 < y \leq 1 \\ 0 & , y \leq 0 \end{cases}$$

Auf diesem Level erhält man nur eine uninteressante Grenzfunktion.



Auch intuitiv klar:

$$Y_n \xrightarrow{n \to \infty} 0 \ \textit{“in einem geeigneten Sinn”}$$

z.B. $P(|Y_n - 0| > \epsilon) \xrightarrow{n \to \infty} 0 \ \forall \epsilon > 0$ (vergleiche Satz 4.7.2 auf Seite 50)
Interessanter wird es, wenn man Y_n mit einem wachsenden Faktor multipliziert, z.B. $n * Y_n$:

$$
\begin{aligned}
F_{n*Y_n}(y) &= P(n * Y_n \leq y) \\
&= P(Y_n \leq \tfrac{y}{n}) \\
&= F_{Y_n}(\tfrac{y}{n}) \\
&= 1 - (1 - \tfrac{y}{n})^n \xrightarrow{n\to\infty} 1 - e^{-y} \;, y > 0
\end{aligned}
$$

Als Grenzwert ergibt sich also die Verteilungsfunktion zur Exponentialverteilung mit Parameter 1.
Es gibt eine Reihe von wichtigen Konvergenzbegriffen in der Stochastik.
Hier zentral: (**Prüfungsrelevant**)

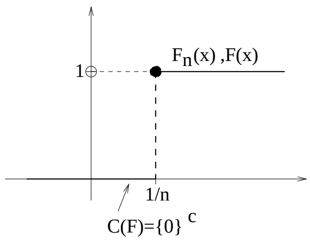
Definition 6.2.1
P, P_n ($n \in \mathbb{N}$) seien Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$ mit den Verteilungsfunktionen F, F_n ($n \in \mathbb{N}$),

$$C(F) := \{x \in \mathbb{R} : F \text{ stetig in } x\} \qquad (\textbf{c}\textit{ontinuity points})$$

ist die Menge der Stetigkeitspunkte von F.
Gilt dann $\lim_{n\to\infty} F_n(x) = F(x)$ $\forall x \in C(F)$, so sagt man, daß P_n <u>schwach</u> gegen P konvergiert. ($P_n \xrightarrow{w} P$, "weak").
Sind $X, X_n (n \in \mathbb{N})$ Zufallsvariablen mit Verteilungen $P, P_n (n \in \mathbb{N})$, so sagt man, daß X_n <u>in Verteilung</u> gegen X konvergiert. ($X_n \xrightarrow{\mathcal{D}} X$, "in <u>distribution</u>"))
(**Prüfungsrelevant**)

Beispiel 6.2.1 von Seite 81 zeigt, daß $n * \min\{X_1, \cdots, X_n\}$ in Verteilung konvergiert, wobei der Grenzwert die Exponentialverteilung mit Parameter 1 hat.

Warum nicht punktweise Konvergenz <u>überall</u> ?
Betrachte $X_n \equiv \frac{1}{n}$.
"Vernünftiger Grenzwert" $X \equiv 0$



$P(X_n \leq \frac{1}{n}) = 1$

$$F(x) = \begin{cases} 1 & , x \geq 0 \\ 0 & , < 0 \end{cases}$$

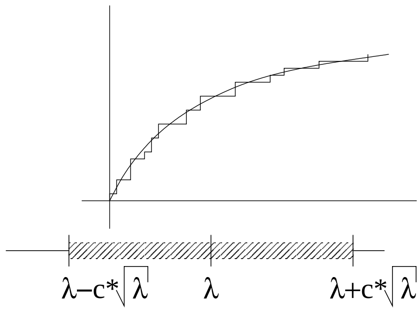
$F_n(0) = 0 \ \forall n \in \mathbb{N}$
$F(0) = 1$
$F_n(x) \to F(x)$ gilt hier tatsächlich nur <u>innerhalb</u> von $C(F)$.

6.3 Normalapproximation bei Poisson-Verteilungen

<u>Grob:</u> *Poisson(λ) sieht bei großen λ in etwa wie eine Normalverteilung mit denselben Erwartungswert und derselben Varianz aus.*

Aus Beispiel 4.3.2 (Seite 35) bekannt: dieser Erwartungswert ist λ *und die zugehörige Varianz ist auch* λ.
In welchem Sinne gilt (kann gelten) $Poisson(\lambda) \approx N(\lambda, \lambda)$ *?*
Dies ist nicht trivial, da z.B. $Poisson(\lambda)(\mathbb{N}_0) = 1,\ N(\lambda, \lambda)(\mathbb{N}_0) = 0$
Sei $I(c, \lambda) := [\lambda - c * \sqrt{\lambda}, \lambda + c * \sqrt{\lambda}]$. *Chebychev (Satz 4.7.1(2) (Seite 49) liefert für eine* $Poisson(\lambda)$*-verteilte Zufallsgröße* X*:*

$$
\begin{aligned}
P(X \in I(c, \lambda)) &= P(X - \overbrace{\lambda}^{EX} | \leq c * \sqrt{\lambda}) \\
&= 1 - P(|X - \lambda| > c * \sqrt{\lambda}) \\
&\geq 1 - \frac{1}{(c*\sqrt{\lambda})^2} * var(X) \\
&= 1 - \frac{1}{c^2}
\end{aligned}
$$



"Wenn λ *gegen* ∞ *geht,*
geht dieses Intervall Richtung
Herrenhäuser Gärten !"

Bei großem c liegt die Hauptmasse von X *in* $I(c, \lambda)$.
Wir behaupten, daß $I(c, \lambda)$ *die Wahrscheinlichkeitsmassenfunktion zu* $Poisson(\lambda)$ *durch die Wahrscheinlichkeitsdichte zu* $N(\lambda, \lambda)$ *gleichmäßig gut approximiert wird:*
Sei $\phi(x|\mu, \sigma^2) = \frac{1}{\sqrt{2*\pi*\sigma^2}} * exp(-\frac{1}{2*\sigma^2} * (x - \mu)^2)$ *die Dichte zu* $N(\mu, \sigma^2)$

Satz 6.3.1 *(Normalapproximation für Poisson-Verteilungen, lokale Form)*
Für alle $c > 0$ *gilt:*

$$sup_{k \in I(c,\lambda)} \left| \frac{e^{-\lambda} * \frac{\lambda^k}{k!}}{\frac{1}{\sqrt{2*\pi*\lambda}} * exp(-\frac{1}{2*\lambda} * (k-\lambda)^2)} - 1 \right| \xrightarrow{mit\ \lambda \to \infty} 0$$

Beweis:

Sei $c > 0$ *fest. Setze*

$$g(k|\lambda) := log P_\lambda(X = k) = -\lambda + k * log\lambda - log(k!)$$
$$Z(k, \lambda) := \frac{1}{\sqrt{\lambda}} * (k - \lambda)$$

Dann gilt:

$$g(k|\lambda) - g(k-1|\lambda) = log\lambda - logk,$$
$$|Z(k, \lambda)| \leq c \text{ für alle } k \in I(c, \lambda)$$

Mit $\boxed{log(1 + x) = x + O(x^2)}$ *(bei* $x \to 0$*) folgt:*

$$
\begin{aligned}
\underline{\underline{logk}} &= log(\lambda + \sqrt{\lambda} * Z(k, \lambda)) \\
&= log\lambda + \log(1 + \frac{1}{\sqrt{\lambda}} * Z(k, \lambda)) \\
&= \underline{\underline{log\lambda + \frac{1}{\sqrt{\lambda}} * Z(k, \lambda) + O(\frac{1}{\lambda})}},\ mit\ \lambda \to \infty,
\end{aligned}
$$

gleichmäßig auf $I(c,\lambda)$

[Einschub: $lim_{x\to 0}\frac{log(1+x)-x}{x^2} = lim_{x\to 0}\frac{\frac{1}{1+x}-1}{2*x}$
$= lim_{x\to 0}\frac{-\frac{1}{(1+x)^2}}{2}$
$= -\frac{1}{2}$*]*

<u>*Also:*</u>

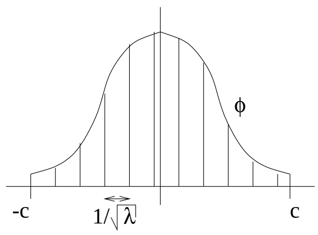
$$
\begin{aligned}
g(k|\lambda) - g(\lfloor\lambda\rfloor|\lambda) &= \textstyle\sum_{j=\lfloor\lambda\rfloor+1}(g(j|\lambda) - g(j-1|\lambda))\\
&= -\tfrac{1}{\sqrt{\lambda}} * \textstyle\sum_{j=\lfloor\lambda\rfloor+1}^{k} Z(j|\lambda) + O(\tfrac{1}{\sqrt{\lambda}})\\
&\qquad\qquad (\sqrt{\lambda} + O(\tfrac{1}{\sqrt{\lambda}}) = O(\tfrac{1}{\sqrt{\lambda}}))\\
&= -\tfrac{1}{\lambda} * \textstyle\sum_{j=\lfloor\lambda\rfloor+1}^{k}(j-\lambda) + O(\tfrac{1}{\sqrt{\lambda}})\\
&\approx -\tfrac{1}{\lambda} * \textstyle\sum_{j=1}^{k-\lfloor\lambda\rfloor} j + O(\tfrac{1}{\sqrt{\lambda}})\\
&= -\tfrac{1}{\lambda} * \tfrac{1}{2} * (k-\lfloor\lambda\rfloor) * (k+1-\lfloor\lambda\rfloor) + O(\tfrac{1}{\sqrt{\lambda}})\\
&\approx -\tfrac{1}{\lambda} * \tfrac{1}{2} * (k-\lambda)^2 + O(\tfrac{1}{\sqrt{\lambda}})\\
&= -\tfrac{1}{2*\lambda} * (k-\lambda)^2 + O(\tfrac{1}{\sqrt{\lambda}})
\end{aligned}
$$

"Ich hab' meine gelbe Kreide noch nie so vermißt wie heute !"
"Wenn wir uns die Aussage des Satze, der dummerweise weggewischt ist, nochmal anschauen..."

(Die Ungenauigkeiten landen alle im O-Term)
Setzt man $h(\lambda) := exp(g(\lfloor\lambda\rfloor|\lambda)$*, so haben wir:*
$P_\lambda(X=k) = h(\lambda) * exp(-\frac{1}{2*\lambda} * (k-\lambda)^2) * (1 + O(\frac{1}{\sqrt{\lambda}}))$
Es bleibt zu zeigen:
$lim_{\lambda\to\infty}\sqrt{2*\pi*\lambda} * h(\lambda) = 1$ $\quad(\heartsuit)$
Summiere die bereits erhaltene Approximation über $I(c,\lambda)$*:*

$$
\begin{aligned}
1 \geq P_\lambda(X \in I(c,\lambda)) &= h(\lambda) * (\textstyle\sum_{k\in I(c,\lambda)} exp(-\frac{1}{2*\lambda} * (k-\lambda)^2)) * (1 + O(\frac{1}{\sqrt{\lambda}}))\\
&= \sqrt{2*\pi*\lambda} * h(\lambda) * (1 + O(\tfrac{1}{\sqrt{\lambda}})) * \underbrace{\sum_{|k-\lambda|\leq c*\sqrt{\lambda}} \frac{1}{\sqrt{\lambda}} * \phi(\frac{k-\lambda}{\sqrt{\lambda}})}_{Riemann-Summe}
\end{aligned}
$$

mit $\phi(x) = \frac{1}{\sqrt{2*\pi}} * e^{-\frac{x^2}{2}}$



konvergiert gegen $\int_{-c}^{c}\phi(y) * dy$

"Das ist das, was da oben gegen die Decke läuft"

Insgesamt bisher:

$$1 \geq (limsup_{\lambda\to\infty}(\sqrt{2*\pi*\lambda}*h(\lambda))*\underbrace{\int_{-c}^{c}\phi(y)*dy}_{\to 1 \ mit \ c\to\infty} \ \forall c>0$$

Also:

$$\boxed{limsup\sqrt{2*\pi*\lambda}*h(\lambda)\leq 1}$$

Analog: liminf ≥ 1, es folgt ($\heartsuit$). □

Korollar 6.3.1 *(Stirling-Formel)*

$$lim_{n\to\infty}\frac{n!}{\sqrt{2*\pi*n*e^{-n}*n^n}}=1$$

Beweis: *Setze $\lambda = k = n$ in vorhergehendem Satz.*

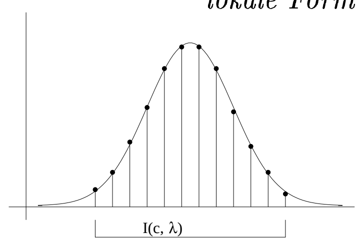
Satz 6.3.2 *(Normalapproximation für Poisson-Verteilungen, kumulative Form)*
Für alle $\lambda > 0$ sei X_λ eine Zufallsvariable, die Poissonverteilt ist mit Parameter λ. Dann gilt:

$$\boxed{\mathcal{L}\left(\frac{X_\lambda-\lambda}{\sqrt{\lambda}}\right)\xrightarrow{w}N(0,1)}$$

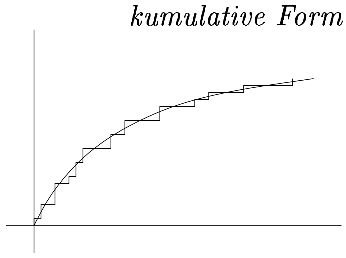
(Der Übergang von X_λ zu $\frac{(X_\lambda-\lambda)}{\sqrt{\lambda}}$ kann als Standardisierung betrachtet werden: nach dieser affinen Transformation hat man Erwartungswert 0 und Varianz 1)

$$\frac{\overbrace{P_\lambda(X=k)}^{\frac{1}{\sqrt{\lambda}}*\phi(\frac{k-\lambda}{\sqrt{\lambda}}|0,1)}}{\phi(k|\lambda,\lambda)}-1$$

lokale Form



kumulative Form



kumulativ heißt: bezieht sich auf die Verteilungsfunktion

<u>*Beweis:*</u>

Für beliebige $a, b \in \mathbb{R}$ *mit* $a < b$ *gibt es ein* C *mit*

$[a, b] \subset [-C, C]$,

also folgt wie im Beweis zu Satz 6.3.1 (Seite 83)

$$\boxed{lim_{\lambda\to\infty} P(a \le \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le b)}$$
$$= lim_{\lambda\to\infty} \sum_{\lambda + a*\sqrt{k} \le k \le \lambda + b*\sqrt{\lambda}} P(X_\lambda = k)$$
$$= lim_{\lambda\to\infty} \sum_{a \le \frac{k-\lambda}{\sqrt{\lambda}} \le b} \frac{1}{\sqrt{\lambda}} * \phi(\frac{k-\lambda}{\sqrt{\lambda}})$$
$$= \int_a^b \phi(x) * dx$$
$$= \boxed{\Phi(b) - \Phi(a)}$$

Φ *ist die Verteilungsfunktion zu* $N(0, 1)$.

z.Z.: $\boxed{P(\frac{X_\lambda}{\sqrt{\lambda}} \le b) \to \Phi(b) \quad \forall b \in \mathbb{R}}$

(Definition von $\xrightarrow{w}$ $C(\Phi) = \mathbb{R}$*)*

Erste Idee: lasse a *gegen* $-\infty$ *laufen*

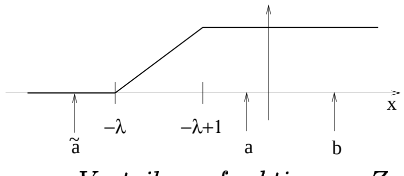
$$lim_{a\to-\infty} P(a \le \frac{X-\lambda}{\sqrt{\lambda}} \le b) = P(\frac{X-\lambda}{\sqrt{\lambda}} \le b)$$
$$lim_{a\to-\infty} (\Phi(b) - \Phi(a)) = \Phi(b)$$

Was also gebraucht wird, ist

$$lim_{a\to-\infty} lim_{\lambda\to\infty} P(a \le \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le b) = lim_{\lambda\to\infty} lim_{a\to-\infty} P(a \le \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le b)$$

Diese Limesvertauschung benötigt zusätzliche Argumente:

$Z_\lambda = Z - \lambda, \qquad Z \sim unif(0, 1)$



Verteilungsfunktion zu Z_λ

$$lim_{\lambda\to\infty} \underbrace{P(a \le Z_\lambda \le b)}_{P(Z_\lambda \le b) - P(Z_\lambda < a)} = 0 \ \forall a, b$$
$$lim_{a\to-\infty} lim_{a\to\infty} P(a \le Z_\lambda \le b) = 0$$
$$lim_{\lambda\to\infty} \underbrace{lim_{a\to-\infty} P(a \le Z_\lambda \le b)}_{= P(Z_\lambda \le b) - P_\lambda(Z_\lambda < a)} = 1$$
$$P(Z_\lambda \le b) \to 1 \ mit \ \lambda \to \infty$$
$$P_\lambda(Z_\lambda < a) \to 0 \ mit \ a \to -\infty$$

Sei $\epsilon > 0$*. Dann existiert ein* $a > -\infty$ *mit* $\Phi(a) < \frac{\epsilon}{2}$
und $P(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} < a) < \frac{\epsilon}{2}$ *für* $\lambda > 0$.
<u>*Begründung:*</u>

$$\begin{aligned} P(\tfrac{X_\lambda-\lambda}{\sqrt{\lambda}} < a) &= P(X_\lambda - \lambda \leq a * \sqrt{\lambda}) \\ &= P(-\tfrac{X_\lambda-\lambda}{\sqrt{\lambda}} > -a) \\ &\leq P(\tfrac{|X_\lambda-\lambda|}{\sqrt{\lambda}} > |a|) \text{ für } a < 0 \end{aligned}$$

"sieht etwas gepfuscht aus, oder ?"

$$\begin{aligned} &\leq P(|X_\lambda - \lambda| \geq |a| * \sqrt{\lambda}) \\ &\leq \tfrac{1}{a^2 * \lambda} * varX_\lambda \text{ nach Chebychev} \\ &= \tfrac{1}{a^2} \end{aligned}$$

Dreiecksungleichung:

$$|P(a \leq \tfrac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq b) + P(\tfrac{X_\lambda-\lambda}{\sqrt{\lambda}} < a) - (\Phi(b) - \Phi(a) + \phi(a))|$$
$$= P(\tfrac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq b) - \Phi(b)|$$
$$\leq |P(a \leq \tfrac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq b) - (\Phi(b) - \Phi(a))| + P(\tfrac{X_\lambda-\lambda}{\sqrt{\lambda}} < a) + \Phi(a)$$

Also:

$$limsup_{\lambda\to\infty}|P(\tfrac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq b) - \Phi(b)| \leq 0 + \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2}$$

Da $\epsilon > 0$ *beliebig war, folgt*

$$limsup|\cdots| = 0,$$

also die behauptete Konvergenz. □

6.4 Normalapproximation bei der Binomialverteilung

Satz 6.4.1 *("lokale Form")*
Es sei $p \in [0,1)$ *fest,* $c > 0$, $I(c,n) := [n*p - c*\sqrt{n}, n*p + c*\sqrt{n}] \cap \{0, \cdots, n\}$
Dann gilt

$$lim_{n\to\infty} sup_{k\in I(c,n)} \left| \frac{\binom{n}{k} p^k * (1-p)^{n-k}}{\phi(k|n*p, n*p(1-p))} - 1 \right| = 0$$

mit $n*p$ *Erwartungswert und* $n*p(1-p)$ *Varianz von* $Bin(n,p)$

<u>*Beweis:*</u>
Wir führen die Aussage durch einen Trick auf den Poisson-Fall zurück.
Setze $\lambda = n*p, y = n*(1-p)$*:*

$$\binom{n}{k} * p^k * (1-p)^{n-k} = \frac{e^{-\lambda} * \frac{\lambda^k}{k!} * e^{-\mu} * \frac{y^{n-k}}{(n-k)!}}{e^{-(\lambda+\mu)} * \frac{(\lambda+\mu)^n}{n!}}$$

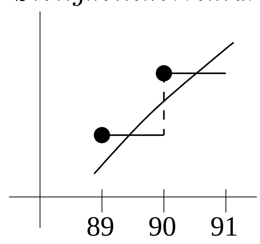
Wende nun Satz 6.3.1 (Seite 83) dreimal an.
(Rest technisch) □

Satz 6.4.2 *(de Moivre-Laplace)* **(Prüfungsrelevant)**
Es sei $X_1, X_2, \cdots$ eine Folge von unabhängigen Zufallsvariablen mit
$P(X_i = 1) = 1 - P(X_i = 0) = p$
für alle $i \in \mathbb{N}$, $p \in (0,1)$ fest.
Sei $S_n := \sum_{i=1}^{n} X_i$. Dann gilt: $S_n \sim Bin(n,p)$
$$\mathcal{L}\left(\frac{S_n - n*p}{\sqrt{n*p*(1-p)}}\right) \xrightarrow{w} N(0,1) \text{ mit } n \to \infty$$

Beweis:
Wegen $\mathcal{L}(S_n) = Bin(n,p)$ folgt dieser Satz aus Satz 6.4.1 (Seite 87) mit denselben Argumenten, wie Satz 6.3.2 (Seite 85) aus Satz 6.3.1 (Seite 83) folgt. $\square$

Beispiel 6.4.1 **(Prüfungsrelevant)**
Mit welcher Wahrscheinlichkeit erscheint beim 600 maligen Wurf eines fairen Würfels mindestens 90 mal und höchstens 105 mal eine sechs ?
Als tatsächlicher Wert ergibt sich (diese Anzahl ist $Bin(600, \frac{1}{6})$-verteilt):
$\sum_{k=90}^{105} \binom{600}{k} * (\frac{1}{6})^k * (1 - \frac{1}{6})^{600-k} = 0.60501...$
Satz 6.4.1 (Seite 87) liefert den Näherungswert
$\sum_{k=90}^{105} \phi(k|100, \frac{500}{6}) = 0,601769...,$
also eine recht gute Näherung (relativer Fehler $< 0.54\%$)
Typische Anwendung des Satzes von de Moivre-Laplace:
$$P(90 \leq S_{600} \leq 105) = P(89 < S_{600} < 106)$$
$$= P(89,5 < S_{600} \leq 105,5)$$

Stetigkeitskorrektur



$$= P(S_{600} \leq 105,5) - P(S_{600} \leq 89,5)$$
$$= P\left(\frac{S_{600}-100}{\sqrt{\frac{500}{6}}} \leq \frac{105,5-100}{\sqrt{\frac{500}{6}}}\right) - P\left(\frac{S_{600}-100}{\sqrt{\frac{500}{6}}} \leq \frac{89,5-100}{\sqrt{\frac{500}{6}}}\right)$$
(de Moivre-Laplace besagt, daß bei großen n die Verteilungsfunktion von $\frac{(S_n - n*p)}{\sqrt{n*p*(1-p)}}$ dicht bei Φ liegt):
$$\approx \Phi\left(\frac{5,5}{\sqrt{\frac{500}{6}}}\right) - \Phi\left(-\frac{10,5}{\sqrt{\frac{500}{6}}}\right)$$
$$= 0,6015504 \text{ (Tabellenwert)} \qquad \Phi(-x) = 1 - \Phi(x)$$
(relativer Fehler marginal größer als bei der lokalen Form: $0,57\%$)

Bemerkung 6.4.1
Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen mit endlichem zweiten Moment.
Man nennt $(S_n)_{n\in\mathbb{N}}$ mit $S_n := \sum_{i=1}^{n} X_i$ die Folge der *<u>Partialsummen</u>* *und $S_n^\star := \frac{S_n - ES_n}{\sqrt{var\ S_n}}$ die* *<u>n-te standardisierte</u>* *Partialsumme.*

(Klar: $ES_n^\star = 0, var(S_n^\star) = 1$*). Man sagt, daß* $(X_n)_{n\in\mathbb{N}}$ *dem zentralen Grenzwertsatz genügt, wenn*

$$\boxed{\mathcal{L}(S_n^\star) \xrightarrow{w} N(0,1)}$$

gilt.

Satz 6.4.2 (Seite 88) zeigt, daß $(X_n)_{n\in\mathbb{N}}$ *dem zentralen Grenzwertsatz genügt, wenn die* X_i *unabhängig und* $Bin(1,p)$*-verteilt sind. Auch die Poisson-Situation paßt in diesen Rahmen.*
Allgemeine Form $\rightarrow$ *Stochastik II.*

7 Goodies, alle Angaben ohne Gewähr

7.1 Einige Formeln

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cap B) = P(A) + P(B) - P(A \cup B)$
$P(A \cap B) = P(B) - P(A^c \cap B)$
$P(X = r) = P(X > r - 1) - P(X > r)$
$var(aX) = a^2 var(X)$
$var(a + X) = var(X)$

7.2 Eselsbrücken

Zum Rechnen: $\cap \hat{=} *$ *und* $\cup \hat{=} +$, *mit* $\cap$*-Rechnung geht vor* $\cup$*-Rechnung*
Siebformel:
$P(A_1 \cup \cdots \cup A_n) = \sum \bigcap (\textit{ungerade Mengen}) - \sum \bigcap (\textit{gerade Mengen})$
$\sigma_x = \sqrt{var(X)}$
Polarisationstrick: $a * b = \frac{1}{4} * ((a + b)^2 - (a - b)^2)$
$var(X) = E(X - EX)^2 = E(X^2) - (EX)^2 = EX(X - 1) + EX - (EX)^2$
$A \triangle B = (A \cap B^c) \cup (A^c \cap B) =?(A \cap B)^c?$
De Morgan: $\bigcap A_i = (\bigcup A_i^c)^c$
$log(1 - x) = -\sum_{n=1}^{\infty} \frac{x^n}{n} \xrightarrow{x \to 0} -x + O(x^2)$
$E1_A = P(A)$
$P(X + Y = z) = \sum_{l=\dots}^{z} P(X = l, Y = z - l)$ *(Faltung !)*
$P^{X|X+Y=z}(\{x\}) = P(X = x|X + Y = z) = \frac{P(X=x, X+Y=z)}{P(X+Y=z)}$
$EX = \hat{p}_x(z)'$
$var\ X = \hat{p}_x(z)''$
$\phi(x|\mu, \sigma^2) = \frac{1}{2*\pi*\sigma^2} * exp(-\frac{1}{2*\sigma^2} * (x - \mu)^2)$

Aufgabe: Gesucht ist die Anzahl der Möglichkeiten "MIFRIFI" anders zu ordnen.
Es gibt 7 Elemente ⇒ 7! Permutationen.
Davon: $1 * M, 3 * I, 2 * F, 1 * R$
Also lautet die Lösung: $\frac{7!}{1!*3!*2!*1!} = 420$

8 Zeichenerklärung und Schlagwortregister

Erklärung für einige Sonderzeichen:

$\#$	*Mächtigkeit*
$\exists$	*es gibt*
∞	*unendlich*
z.B.	*zum Beispiel*
$\rightsquigarrow$	*siehe, vergleiche*
$\notin$	*ist kein Element von*
d.h.	*daß heißt*
$\downarrow$	*strebt von oben gegen*
$\square$	*was zu beweisen war*
$\surd$	*logisch, trivial, abgehakt*
$\approx$	*ungefähr*
etc.	*und so weiter*
bzw.	*beziehungsweise*
$\Leftrightarrow$	*äquivalent*
$\Rightarrow$	*daraus folgt*
$\equiv$	*identisch, Kongruent*
ZV	*Zufallsvariable*
ZG	*Zufallsgröße*
i.a.	*im Allgemeinen*
s.	*siehe*
$P \star Q$	*Faltung von P und Q*
g.a.n.	*general abstract nonsens*
g.ä.h.n.	*general abstract boring nonsens (Trusch-Version)*
(☺)	*Das is´n smiley,du ...*
$\forall$	*für alle*
$\subset, \subseteq$	*ist Teilmenge von*
$\mapsto$	*wird abgebildet auf*
$\rightarrow$	*geht über*
$\cup$	*vereinigt*
$\cap$	*geschnitten*
oBdA	*ohne Beschränkung der Allgemeinheit*
$\sim$	*konjugiert zu*
LORA	*Links Offen Rechts Abgeschlossen*
ROLA	*Rechts Offen Links Abgeschlossen*

$z.T.$	*zu Zeigen*
$\vee$	*oder*
$\wedge$	*und*
$\neg$	*nicht*
$\lceil X \rceil$	*oberer Grenzwert, die kleinste ganze Zahl, größer oder gleich X*
$\lfloor X \rfloor$	*unterer Grenzwert, die größte ganze Zahl, kleiner oder gleich X*
$\triangle$	*symmetrische Differenz*

Index