Elementare Wahrscheinlichkeitstheorie und Statistik A

Dozent: Prof. Dr. R. Grübel

unkorrigierte Vorlesungsmitschrift
aus dem **WS 2006 / 07**
angefertigt von Florian Garbs

Bei etwaigen Fehlern, würde ich mich über eine Benachrichtigung per
E-mail (sos_1981@yahoo.de) freuen.

www.sos-1981.de

Inhaltsverzeichnis

- Stochastik: Sammelbegriff für Wahrscheinlichkeitstheorie und mathematische Statistik (randomisierte Algorithmen)

„Mathematik des Zufalls"

§1 Ein mathematisches Modell für Zufallsexperimente
Ergebnis – Ereignis – Wahrscheinlichkeit

Zufallsexperiment: Ergebnis nicht durch Randbedingungen festgelegt.
Der Ergebnisraum Ω ist eine nicht-leere Menge, die die möglichen Resultate des Zufallsexperiments enthält.

Beispiel 1.1

Würfelwurf: $\Omega = \{1,...,6\}$

Radioaktiver Zerfall:

a) Anzahl der Emissionen in einem bestimmten Zeitintervall $\Omega = \mathrm{N}_0 = \{0,1,2,...\}$

b) Zeitpunkt des ersten Zerfalls: $\Omega = (\Omega, \infty]$

Ereignisse: Eine Aussage über das Ergebnis des Experiments wird „übersetzt" in die Teilmenge von Ω, für die die Aussage richtig ist: beim Würfelwurf wird das Ereignis „es kommt eine gerade Zahl" beschrieben durch $A = \{2,4,6\} \subset \Omega$.

Die 1 ist nicht dabei

Beim radioaktiven Zerfall: kein „Tick" im Zeitintervall $[0,1]$: $(1,\infty]$

Diese Kombinationen haben ein „mengenalgebraisches Gegenstück"

1. A tritt nicht ein wird durch $A^C \left(= \{\omega \in \Omega : \omega \notin A\}\right)$ dargestellt.
   $A^C = \{\omega \in \Omega : \omega \notin \{2,4,6\}\} = \{1,3,5\}$ „Komplementärereignis"
2. A und B treten beide ein: $A \cap B$
3. A tritt ein oder B tritt ein oder beide: $A \cup B$

Beispiel 1.2

Ein Würfel wird zweimal geworfen.

$\Omega = \{(i,j) : i,j \in \{1,...,6\}\} = \{1,2,...,6\}^2$

Ergebnis des 2. Wurfs

Ergebnis des 1. Wurfs

A: „Pasch", beide Resultate gleich $= \{(1,1),(2,2),...,(6,6)\}$

B: Augensumme ≤ 5: $B = \{(1,1),(2,1),(1,2),(3,1),(3,2),(2,2),(4,1),(1,4),...\}$

$A \cap B = \{(1,1),(2,2)\}$

Typisches Beispiel mit mehr als zwei Ereignissen:
„Genau eines der drei Ereignisse A, B, C tritt ein"

$$\left(A \cap B^C \cap C^C\right) \cup \left(A^C \cap B \cap C^C\right) \cup \left(A^C \cap B^C \cap C\right)$$

A tritt ein, B und C nicht

Vereinbarung: sind A und B disjunkt, so schreibt man auch $A+B$ anstelle von $A \cup B$.

Was ist Wahrscheinlichkeit?

Wurf einer fairen Münze: Was bedeutet es, wenn man sagt, dass „Kopf“ mit Wahrscheinlichkeit $1/2$ kommt?

1. Häufigkeitsauffassung (*Frequentisten*): Bei einer großen Zahl von Wiederholungen kommt ungefähr in etwa der Hälfte der Fälle „Kopf“. (relative Häufigkeit)
2. Glaubensauffassung (*Subjektisten*): Die Wahrscheinlichkeit gibt, auf einer Skala von 0 bis 1, den Grad meines Glaubens an das Eintreten des Ereignisses wieder. Geht auch bei nicht-wiederholbaren Experimenten. Kann über Wetten formalisiert werden.

Bezeichnet $N_n(A)$ die Anzahl der Versuchswiederholungen (bei insgesamt n), in denen A eingetreten ist (die absolute Häufigkeit), so ist $\frac{1}{n}N_n(A)$ die relative Häufigkeit.

Klar: Sind A und B disjunkt, so gilt: $\frac{1}{n}N_n(A \cup B) = \frac{1}{n}N_n(A) + \frac{1}{n}N_n(B)$

„relative Häufigkeiten sind additiv“:

Auch klar: $\frac{1}{n}N_n(\varnothing) = 0$.

Hierdurch werden die so genannten Kolmegorov-Axiome motiviert:

Definition 1.3 (die Kolmegorov-Axiome)

Eine Wahrscheinlichkeit (oder besser ein Wahrscheinlichkeitsmaß) ist eine Abbildung P die jedem Ereignis A (eine Teilmenge von Ω) eine reelle Zahl zuordnet, und dabei folgenden Regeln genügt:

(A1) $\mathrm{P}(A) \geq 0$, $\mathrm{P}(\Omega) = 1$

(A2) Sind $A_1, A_2, \ldots$ paarweise disjunkte Ereignisse (d.h. $A_i \cap A_j = \varnothing \;\; \textit{für} \;\; i \neq j$), so gilt:

$$\boxed{P\left(\bigcup_{\mathrm{i}=1}^{\infty} A_{\mathrm{i}}\right) = \sum_{\mathrm{i}=1}^{\infty} P(A_{\mathrm{i}})} \;(\sigma\text{-Additivität})$$

Analogie zu Länge, Flächeninhalt, Volumen:

Erste Folgerungen:

Satz 1.4

(I) $P(\varnothing) = 0$

(II) $P(A) \leq 1$

(III) $P(A^{\mathrm{C}}) = 1 - P(A)$

(IV) $A \subset B \Rightarrow P(A) \leq P(B)$

(V) $P(A_1 + \ldots + A_k) = P(A_1) + \ldots + P(A_k)$

(endliche Additivität, Ereignisse müssen disjunkt sein)

Typischer Beweisschritt: Wenn man die endliche Additivität hat, also insbesondere

$$P(A \cup B) = P(A) + P(B)$$

bei $A \cup B = \varnothing$, dann folgt mit $A = B = \varnothing$ $(\varnothing \cap \varnothing = \varnothing)$

$$P(\varnothing) = P(\varnothing) + P(\varnothing)$$

Hieraus folgt:

$P(\varnothing) = 0$, also Aussage (I).

Satz 1.5

(I) (Boolesche Ungleichung):

$$P(A_1 \cup ... \cup A_k) \leq P(A_1) + ... + P(A_k)$$

(Wichtig hierbei: es wird vorausgesetzt, dass $A_1,..,A_k$ paarweise disjunkt sind)

(II) (Formel von Sylvester-Poincaré, Siebformel)

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{H \subset \{1,...,n\} \\ \#H=K}} P\left(\bigcap_{i \in H} A_i\right)$$

(#A ist die Anzahl der Elemente von A)

Bei $n = 2$: *bei* $H = \{1\}$ *bei* $H = \{2\}$ *bei* $H = \{1,2\}$

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

Teilmengen H vom Umfang 1 von $\{1,2\}$ sind $\{1\}$ und $\{2\}$

Teilmengen H vom Umfang 2 von $\{1,2\}$ ist $\{1,2\}$

Beweis für diesen Spezialfall:

$$A \cup B = A + A^C \cap B$$

$$B = A \cap B + A^C \cap B$$

$$P(A \cup B) = P(A) + P(A^C \cap B)$$

$$P(B) = P(A \cap B) + P(A^C \cap B)$$

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

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - \mathrm{P}(\mathrm{A} \cap \mathrm{C}) + \mathrm{P}(\mathrm{A} \cap \mathrm{B} \cap \mathrm{C})$$

§2 Laplace-Experimente

Ω: Ergebnisraum (die „Resultate“), das Wahrscheinlichkeitsmaß P ordnet den Ereignissen A (die als Teilmengen von Ω aufgefasst werden) deren Wahrscheinlichkeit $P(A)$ zu. Bei Laplace-Experimenten ist Ω endlich und es gilt $P(A) = \frac{\# A}{\# \Omega} \; \frac{Anzahl\ der\ günstigen\ Fälle}{Anzahl\ der\ möglichen\ Fälle}$.

Beispiel:
Ein Würfel wird 2 mal geworfen.

$\Omega = \{(i, j) : i, j \in \{1, ..., 6\}\} = \{1, 2, ..., 6\}^2$

1. Wurf 2. Wurf

Annahme: Das Laplace-Experiment über Ω ist das richtige Modell
A_k : *Augensumme k*

$A_6 = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}, \qquad P(A_6) = \frac{5}{36}$

$A_7 = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}, \qquad P(A_7) = \frac{6}{36} \left(= \frac{1}{6} \right)$

Insbesondere gilt: $P(A_6) < P(A_7)$

Klar: Auszählen per hand führt nicht weit!
Die systematische Kunst des Zählens nennt man <u>Kombinatorik</u>.

<u>Grundbegriffe:</u>
Ausgangspunkt sind zwei endliche Mengen $A = \{a_1, ..., a_m\}$ und $B = \{b_1, ..., b_n\}$, also $\# A = m, \quad \# B = n$.

Wir betrachten Funktionen $f : A \to B$
(für die Menge dieser Funktionen schreibt man auch B^A

x	$a_1 \; a_2 \; ------ \; a_n$
$f(x)$	$f(a_1)--------$

Wie viele solche Funktionen gibt es?
n Möglichkeiten für $f(a_1)$
n Möglichkeiten für $f(a_2)$
……
n Möglichkeiten für $f(a_m)$

$\#(B^A) = \overbrace{n \cdot n \cdot ... \cdot n}^{m-mal} = n^m$

$\#(B^A) = (\# B)^{\# A}$

Man nennt die Elemente von B^A <u>Permutationen mit Wiederholung</u>

Standartinterpretationen:

(1) m Kugeln werden auf n Urnen verteilt; Mehrfachbelegung erlaubt
(2) Entnahme einer Stichprobe „mit Zurücklegen“

Was passiert, wenn man keine Mehrfachbelegung zulässt, bzw. nicht zurücklegt?
Für die Funktionen bedeutet dies: $f(a_i) \neq f(a_j)$ bei $i \neq j$.

$\#\left\{f \in B^A : f \ \ injektiv\right\} = ?$ (Macht nur Sinn bei $m \leq n$)

n Möglichkeiten für $f(a_1)$ Permutationen ohne Wiederholung
$n-1$ Möglichkeiten für $f(a_2)$
…
$n-m+1$ Möglichkeiten für $f(a_m)$

Anzahl: $n(n-1)\cdot(n-2)\cdot\ldots\cdot(n-m+1) = \frac{n!}{(n-m)!}$

Interpretationen: keine Mehrfachbelegung bzw. ohne Zurücklegen.

Spricht man einfach von Permutationen, so ist in der Regel <u>ohne</u> Wiederholung gemeint. Zu einer Menge mit n Elementen gibt es $n!$ <u>bijektive Selbstabbildungen</u>.

Beim Mischen eines Kartenspiels mit 32 bzw. 52 karten gibt es $32!$ bzw. $52!$ Möglichkeiten.
„Gut gemischt“ heißt: Laplace-Experiment über

$$\Omega = \left\{(i_1,\ldots,i_{32}) : i_j \neq i_l \ \ f\ddot{u}r\ l \neq j,\ i_j \in \{1,\ldots,32\}\right\}$$

$32! \approx 2,6 \cdot 10^{35}$ (eine große Zahl)

Was passiert, wenn man die Verteilungsreihenfolge ignoriert (bzw. die Reihenfolge der Entnahme)?

Alle $m!$ bijektiven Selbstabbildungen von A führen auf dasselbe Ergebnis, also:
Anzahl der <u>Kombinationen ohne Wiederholung</u> ist

$$\frac{n!}{(n-m)!}\Big/ m! = \binom{n}{m}$$ (Binominalkoeffizient)

Einfaches Beispiel: 2 Kugeln auf 3 Urnen (keine Mehrfachbelegung)

$$m = 2,\ n = 3 \qquad \frac{n!}{(n-m)!} = \frac{3 \cdot 2 \cdot 1}{1} = 6$$

	1	2	
	1		2
		1	2
	2	1	
	2		1
		2	1
Urne	1	2	3

Was passiert, wenn die Reihenfolge nicht berücksichtigt (die Kugeln ihre Identität verlieren)?
Wie viele bijektive Selbstabbildungen (Permutationen) der Menge $A = \{1, 2\}$ gibt es?

$$\binom{n}{m} = \frac{6}{2} = 3$$

Für eine injektive Abbildung f ist $f(A)$ eine Teilmenge vom Umfang $\# A$ von *B*.
Also: wichtiger Spezialfall:

Eine Menge mit *n* Elementen hat $\binom{n}{m}$ Teilmengen von Umfang *m*.

(Bei 3 Teilnehmern gibt es $\binom{83}{5}$ Möglichkeiten, eine AG mit 5 Mitgliedern auszuwählen…)

Beispiel:
Eine Münze wird 10mal geworfen. Mit welcher Wahrscheinlichkeit erhält man 7mal Kopf?

1. Schritt: Finde ein geeignetes Modell. Naheliegend (aber nicht beweisbar „im math. Sinn“) ist das Laplace-Experiment über $\Omega = \{0,1\}^{10} = \{(\omega_1, ..., \omega_{10}) : w_i \in \{0,1\} \textit{ für alle } i\}$.

2. Schritt: Identifiziere das interessierende Ereignis als Teilmenge von Ω.

$$A = \{(\omega_1, ..., \omega_{10}) \in \Omega : \#\{1 \le i \le 10 : \omega_i = 1\} = 7\}$$
$$= \{\omega \in \Omega : \omega_1 + ... + \omega_{10} = 7\}$$

3. Schritt: Zählen $\#\Omega = 2^{10} = 1024$ $\qquad \# A = ?$

Jedes $\omega \in \Omega$ ist eindeutig durch die Positionen der Einsen festgelegt. Diese wiederum bilden eine Teilmenge vom Umfang 7 einer Menge von 10 Elementen.

Also $\# A = \binom{10}{7}$.

Schlussakkord: $P(A) = \frac{\# A}{\# \Omega} = \frac{1}{1024} \cdot 120 = \frac{15}{128}$.

(bei Ω: Permutationen mit Wiederholung, bei *A*: Kombinationen ohne Wiederholung)

Beim Übergang von Permutationen zu Kombinationen werden Funktionen zusammengefasst, die sich über eine Permutation der Argumente ergeben.

$$\{(i_1,...,i_m): i_j \in \{1,...,n\}, i_1 < i_2 < ... < i_m\} = \binom{n}{m}$$

Gibt die Indizes der Elemente der Grundmenge an, in aufsteigender Reihenfolge, die in der Teilmenge landen.

Kombinationen <u>mit Wiederholung</u>:

$$\#\{(i_1,...,i_m): i_j \in \{1,...,n\}, i_1 \le i_2 \le ... \le i_m\} = \binom{n+m-1}{m}$$

(Bei $n=3$ Urnen und $m=2$ Kugeln: $\binom{4}{2} = \frac{4!}{2!\cdot 2!} = 6$)

<u>Beweis:</u>

Urne 1	Urne 2	Urne 3
oo		
	oo	
		oo
o	o	
o		o
	o	o

<u>Idee</u>: Zwei Mengen C und D haben dieselbe Mächtigkeit, wenn es eine bijektive Abbildung φ von C nach D gibt.

Betrachte: $C := \{(i_1,...,i_m): alle\ in\ \{1,...,n\},\ i_1 \le i_2 \le ... \le i_m\}$

$D := \{(i_1,...,i_m): alle\ in\ \{1,...,n+m-1\},\ i_1 > i_2 > ... > i_m\}$

Bekannt: $\#D = \binom{n+m-1}{m}$.

Also: der Beweis ist fertig, wenn man eine bijektive Abbildung $\varphi: C \to D$ findet.

$(i_1,...,i_m) \mapsto (i_1, i_2+1, i_3+2, ..., i_m+m-1)$ injektiv! surjektiv!

	mit Wiederholung	ohne Wiederholung
Permutation (Reihenfolge zählt)	n^m	$\frac{n!}{(n-m)!}$
Kombinationen (Reihenfolge egal)	$\binom{n+m-1}{m}$	$\binom{n}{m}$

Permutationen mit Wiederholung: $\{(i_1,...,i_m): i_j \in \{1,...,n\} \text{ für } j=1,...,n\}$
$= M_n^m$ mit $M_n := \{1,...,n\} \quad (=[n])$

Permutationen ohne Wiederholung: $\{(i_1,...,i_m) \in M_n^m : \; i_j \neq i_l \quad \text{für} \quad j \neq l\}$

Kombinationen mit Wiederholung: $\{(i_1,...,i_m) \in M_n^m : \; i_1 \leq ... \leq i_m\}$

Kombinationen ohne Wiederholung: $\{(i_1,...,i_m) \in M_n^m : \; i_1 < ... < i_m\}$

Anwendungen

1. Das Geburtstagsproblem: n Personen in einem Raum: mit welcher Wahrscheinlichkeit haben (mindestens) 2 am gleichen Tag Geburtstag? $(n < 365)$

   Vereinfachende Annahmen:
   - 29. Februar streichen
   - keine Zwillings, etc.
   - keine saisonalen Schwankungen

   Unter diesen Vorraussetzungen ist es plausibel, dass ein Laplace-Experiment vorliegt über

   $$\Omega = \{(i_1,...,i_n): i_j \in \{1,...,365\}\}$$

   Interpretation: $i_j = k$ bedeutet, dass Person j am Tag k Geburtstag hat.
   Es geht um $A = \{(i_1,...,i_n) \in \Omega : i_j = i_l \text{ für ein } j \neq l\}$

   Trick: $1 - P(A^C) = 1 - \frac{\# A^C}{\# \Omega}$
   $$A^C = \{(i_1,...,i_n) \in \Omega : i_j \neq i_l \;\; \textit{für alle} \;\; j \neq l\}$$
   Wir erkennen zwei Einträge in der Tabelle:
   $$P(A) = 1 - \frac{365!}{(365-n)!} = \frac{1}{365^n}$$
   Mit zunehmendem n wird dies größer, ab 23 ist die Wahrscheinlichkeit größer als $1/2$.

2. Der zerstreute Postbote: Ein Postbote verteilt n Briefe auf n Kästen, einen pro Kasten. Zu jeder der n Adressen gehört genau ein Brief.
   Mit welcher Wahrscheinlichkeit bekommt keiner der Adressaten den richtigen Brief?

   Modell ist ein Laplace-Experiment über
   $$\Omega = \{(i_1,...,i_n): i_j \in \{1,...,n\}, \; i_j \neq i_k \;\; \textit{für} \;\; j \neq k\}$$
   (In Worten: alle Permutationen sind gleich wahrscheinlich)

Interpretation: $i_j = k$ bedeutet, dass in Kasten j der Brief für Person k landet.

1	2	3	4	5	6	7	8	9
4	3	7	8	1	2	9	5	6

$A = \{\omega \in \Omega :\ \omega_i \neq i \text{ für } i = 1,\ldots,n\}$

Dies ist die Menge der fixpunktfreien Permutationen.

$$A^C = \{\omega \in \Omega :\ \exists i \in \{1,\ldots,n\} : \omega_i = i\} = \bigcup_{i=1}^{n} B_i \quad mit \quad B_i = \{\omega \in \Omega : \omega_i = i\}$$

Verwende Siebformel:

$$P\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\substack{J \subset \{1,\ldots,n\} \\ \# J = k}} \left(P \bigcap_{i \in J} B_i \right)$$

$\omega \in \bigcap_{i \in J} B_i \Leftrightarrow \omega_i \quad für \quad alle \quad i \in J$

In Worten: k Positionen von ω liegen fest $(\# J = k)$, die übrigen $n-k$ sind frei.

also: $\# \bigcap_{i \in J} B_i = (n-k)!$.

Damit ist: $$P\left(A^C\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\substack{J \subset \{1,\ldots,n\} \\ \# J = k}} \left(\frac{(n-k)!}{n!} \right).$$

Die Anzahl der Teilmengen vom Umfang k einer Menge von n Elementen ist $\binom{n}{k}$

(Kombinationen ohne Widerholung). In der Summe hängt der Summand nicht vom Summationsindex ab, also

$$\boxed{\frac{n!}{(n-k)!\,k!}}$$

$$P(A) = 1 - P\left(A^C\right) = 1 - \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!}$$

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

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

Aus der Analysis bekannt: $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$ (Exponentialreihe)

Bezeichnet also p_n die Wahrscheinlichkeit dafür, dass bei n Briefen „alle im falschen Kasten landen" (oder beim Wichteln: keines der n Kinder erhält sein eigenes Geschenk zurück), so folgt

$$\lim_{x \to \infty} p_n = \frac{1}{e} \approx 0{,}3679$$

(Da eine alternierende Reihe vorliegt, ist die Konvergenz sehr schnell)

Bemerkenswert: Man hat implizit die Anzahl der fixpunktfreien Permutationen einer Menge von n Elementen bestimmt.

§3 Bedingte Wahrscheinlichkeit und Unabhängigkeit

Es seien *A* und *B* Ereignisse in einem Zufallsexperiment.
Welche Wahrscheinlichkeit hat *B*, wenn bekannt ist, dass *A* eintritt?

Bei *n* Wiederholungen tritt *A* $N_n(A)-$mal ein (absolute Häufigkeit), unter diesen tritt *B* $N_n(A\cap B)-$mal ein. Damit ergibt sich als relative Häufigkeit des Eintretens von *B*, unter denen die *A* liefern $\frac{N_n(A\cap B)}{N_n(A)} = \frac{\frac{1}{n}N_n(A\cap B)}{\frac{1}{n}N_n(A)}$.

Wir „haben etwas mit relativen Häufigkeiten zu tun“. Dies motiviert die folgende

Definition:
Es sei *A* ein Ereignis mit $P(A)>0$. Die bedingte Wahrscheinlichkeit von *B* unter *A* wird definiert durch $P(B\,|\,A) := \frac{P(A\cap B)}{P(A)}$..

Beispiel: 2facher Würfelwurf, B: Pasch (2 gleiche Ergebnisse)
A: Augensumme 8

$$\Omega=\{1,...,6\}^2,\ B=\{(i,j)\in\Omega : i=j\}=\{(1,1),...,(6,6)\}$$
$$A=\{(i,j)\in\Omega : i+j=8\}=\{(2,6),...,(6,2)\}$$

$$P(B\,|\,A)=\frac{P(A\cap B)}{P(A)}=\frac{\frac{1}{36}\#(A\cap B)}{\frac{1}{36}\#A}=\frac{1}{5}$$

Satz

(I) (Multiplikationsregel)
Sind $A_1,...,A_n$ Ereignisse mit $P(A_1,...,A_n)>0$, so gilt:
$$P(A_1\cap...\cap A_n)=P(A_1)\cdot P(A_2\,|\,A_1)\cdot P(A_3\,|\,A_1\cap A_2)\cdot...\cdot P(A_n\,|\,A_1\cap...\cap A_{n-1})$$

Beweis:
$$P(A_1\cap...\cap A_n)=P(A_1)\cdot P(A_2\,|\,A_1)\cdot P(A_3\,|\,A_1\cap A_2)\cdot...\cdot P(A_n\,|\,A_1\cap...\cap A_{n-1})$$
$$=P(A_1)\cdot\frac{P(A_2\cap A_1)}{P(A_1)}\cdot\frac{P(A_3\cap A_1\cap A_2)}{P(A_1\cap A_2)}\cdot...\cdot\frac{P(A_n\cap A_1\cap...\cap A_{n-1})}{P(A_1\cap...\cap A_{n-1})}$$
$$=P(A_1\cap A_2\cap...\cap A_n)$$

(II) (Gesetz von der totalen Wahrscheinlichkeit)

Ist $A_1,...,A_n$ eine Ereignispartition von Ω, d.h.

$A_i \cap A_j = \varnothing$ für $i \neq j$, $\bigcup_{i=1}^{n} A_i = \Omega$ so gilt für alle Ereignisse B

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)\, P(A_i).$$

Beweis:

Rechte Seite $= \sum_{k=1}^{n} \frac{P(B \cap A_k)}{P(A_k)} P(A_k) = \sum_{k=1}^{n} P(B \cap A_k) = P\left(\bigcup_{k=1}^{n} (B \cap A_k) \right)$

Da $B \cap A_1,...,B \cap A_n$ paarweise disjunkt sind. (Additivität)

$= P\left(B \cap \left(\bigcup_{k=1}^{n} A_k \right) \right)$ Distributivität

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

Satz (Mit Formel zu bedingten Wahrscheinlichkeiten)

Multiplikationsregel

Gesetz von der totalen Wahrscheinlichkeit: $\overbrace{A_1 \cup ... \cup A_n}^{\text{eines der } A_i\text{'s MUSS eintreten}} = \Omega$, $\underbrace{A_i \cap A_j = \varnothing \text{ für } i \neq j}_{\text{höchstens ein Ereignis kann eintreten}}$

Dann ist

$$P(B) = \sum_{i:1}^{n} P(B \mid A_i) \cdot P(A_i)$$

(Vereinbarung dabei: $P(B \mid A_i) \cdot P(A_i) = 0$ bei $P(A_i) = 0$)

(III) Die Formel von Bayes:

$A_1,...,A_n$ wie in (b). Dann gilt:

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{ji:1}^{n} P(B \mid A_j) \cdot P(A_j)}$$

Beweis: Im Nenner steht $P(B)$ (Gesetz von der totalen Wahrscheinlichkeit)

Also: rechte Seite $= \frac{\frac{P(B \cap A_i)}{P(A_i)} \cdot P(A_i)}{P(B)} = P(A_i \mid B).$

q.e.d.

Beispiel:

Ein bestimmter Test ist zu 95% effektiv beim Erkennen einer bestimmten Krankheit, liefert allerdings bei 1% der gesunden Personen einen falschen Alarm.
Angenommen, 0,5% der Bevölkerung leiden unter dieser Krankheit.
Mit welcher Wahrscheinlichkeit hat jemand diese Krankheit, wenn der Test dies behauptet?

A: getestete Person hat die Krankheit
B: Test „positiv" (zeigt Krankheit an)

Die obigen Annahmen übersetzen sich wie folgt:

$$P(B \mid A) = 0{,}95 \qquad P\left(B \mid A^C\right) = 0{,}01 \qquad P(A) = 0{,}005$$

Worum geht es? $P(A \mid B) = ?$

Bayes liefert:

$$\begin{aligned} P(A \mid B) &= \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P\left(B \mid A^C\right)P\left(A^C\right)} \\ &= \frac{0{,}95 \cdot 0{,}005}{0{,}95 \cdot 0{,}005 + 0{,}01 \cdot 0{,}995} \\ &\approx 0{,}323... \end{aligned}$$

Auf den ersten Blick ist das ein überraschend niedriger Wert.
Diesen Rechnungen liegen bestimmte Annahmen zugrunde,
beispielsweise die, dass die getestete Person zufällig aus der Gesamtbevölkerung (und nicht aus einer Risikogruppe) ausgewählt wird.

(„Der Hund, der Eier legt" als Buchempfehlung in diesem Zusammenhang)

Naheliegend: A ist von B unabhängig, wenn die Information, dass B eingetreten ist, die Wahrscheinlichkeit von A nicht verändert, also
Schwierigkeiten mit $P(B) = 0$ vermeidet.

$$\underbrace{P(A \mid B)}_{\frac{P(A \cap B)}{P(B)}} = P(A)$$

Definition:
Zwei Ereignisse A und B heißen unabhängig, wenn $P(A \cap B) = P(A) \cdot P(B)$ gilt.

Eine Familie $\{A_i : i \in I\}$ heißt unabhängig, wenn $P\left(\bigcap_{i \in H} A_i\right) = \prod_{i \in H} P(A_i)$ für alle endlichen Teilmengen H von I. $\left(\prod_{i=1}^{n} a_i = a_1 \cdot ... \cdot a_n\right)$

Warum diese diffizile Untersuchung?

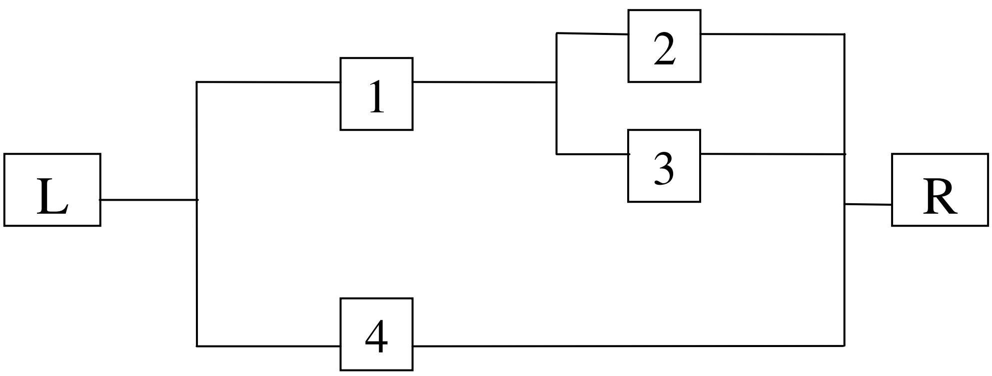
Dafür, dass aus der paarweisen Unabhängigkeit nicht die volle Unabhängigkeit im Sinne der obigen Definition folgt

Beispiel:
Eine faire Münze wird zweimal geworfen:
1: Kopf
0: Wappen
Laplace-Experiment über $\Omega=\{(0,0),(0,1),(1,0),(1,1)\}$
$A_1=\{(0,0),(0,1)\}$ beim 1. Mal kommt Wappen
$A_2=\{(0,0),(1,0)\}$ beim 2. Mal kommt Wappen
$A_3=\{(0,1),(1,0)\}$ Resultate verschieden

$$P(A_1\cap A_2)=P(\{(0,0)\})=\tfrac{1}{4}$$
$$P(A_1)\cdot P(A_2)=\frac{\#A_1}{\#\Omega}\cdot\frac{\#A_2}{\#\Omega}=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$$

$$P(A_2 \mid A_3)=P(A_2)\cdot P(A_3)=\tfrac{1}{4}$$
$$P(A_1 \mid A_3)=P(A_1)\cdot P(A_3)=\tfrac{1}{4}$$
Die Ereignisse A_1,A_2,A_3 sind also paarweise unabhängig.
$$P(A_1\cap A_2\cap A_3)=P(\emptyset)\quad\neq\quad\frac{1}{8}=P(A_1)\cdot P(A_2)\cdot P(A_3)$$

Beispiel:
(Funktionieren von Netzwerken)



Mit welcher Wahrscheinlichkeit kann ein Strom von L nach R fließen, wenn die Schalter unabhängig voneinander, jeweils mit Wahrscheinlichkeit p geschlossen sind?

A_i: Komponente *i* funktioniert (Schalter *i* geschlossen)
B: Gesamtsystem funktioniert (Strom kann fließen)

B_1: A_4 (unterer Pfand passierbar)
B_2: $A_1\cap(A_2\cup A_3)$ (oberer Pfad passierbar)

Klar: $B = B_1 \cup B_2$

$$\begin{aligned}
P\left(A_1 \cap \left(A_2 \cup A_3\right)\right) &= P\left(\left(A_1 \cap A_2\right) \cup \left(A_1 \cap A_3\right)\right) \\
&= P\left(A_1 \cap A_2\right) + P\left(A_1 \cap A_3\right) - P\left(A_1 \cap A_2 \cap A_3\right) \\
\textit{unabh.} \qquad &= P\left(A_1\right) \cdot P\left(A_2\right) + P\left(A_1\right) \cdot P\left(A_3\right) - P\left(A_1\right) \cdot P\left(A_2\right) \cdot P\left(A_3\right) \\
&= p \cdot p + p \cdot p - p \cdot p \cdot p = 2p^2 - p^3
\end{aligned}$$

$$\begin{aligned}
P(B) &= P\left(B_1 \cup B_2\right) = P\left(B_1\right) + P\left(B_2\right) - P\left(B_1 \cap B_2\right) \\
&= P\left(A_4\right) + P\left(A_1 \cap \left(A_2 \cup A_3\right)\right) - P\left(A_4 \left(A_1 \cap \left(A_2 \cup A_3\right)\right)\right) \\
&= p + 2p^2 - p^3 - \underbrace{P\left(A_4 \cap A_1 \cap A_2 \cup A_4 \cap A_1 \cap A_3\right)}_{\substack{P\left(A_4 \cap A_1 \cap A_2\right) + P\left(A_4 \cap A_1 \cap A_3\right) \\ -P\left(A_4 \cap A_1 \cap A_2 \cap A_3\right) = 2p^3 - p^4}}
\end{aligned}$$

$$P(B) = p + 2p^2 - 3p^3 + p^4$$

$4 Diskrete Zufallsgrößen

4.1 Allgemeines

Oft interessiert man sich nicht für das Ergebnis ω eines Zufallsexperimentes, sondern nur für einen hiervon abhängigen Wert $X(\omega)$.

Beispiel:
Eine (faire) Münze wird fünfmal geworfen. Wie oft erscheint „Kopf"?
Mit „1" für Kopf, „0" sonst hat man ein Laplace-Experiment über

$$\Omega = \{(\omega_1,...,\omega_5) : \omega_i \in \{0,1\}\} \quad \left(= \{0,1\}^5\right)$$

Mit dieser Kodierung hat man $X(\omega) = \omega_1 + ... + \omega_5$.
Welche Werte kann X annehmen? 0, 1, 2, 3, 4, 5

Interessant: mit welcher Wahrscheinlichkeit werden dieser Werte angenommen?

$$P(X = 0) = P(\{\omega \in \Omega : X(\omega) = 0\}) = P(\{0,0,0,0,0\}) = 1/2^5 = 1/32$$

Allgemein:

$$P(X = k) = \binom{5}{k} \cdot \left(\frac{1}{2}\right)^5$$

Begründung: Jede 0-1-Folge der Länge 5 hat Wahrscheinlichkeit $\left(\frac{1}{2}\right)^5$.

Es gibt $\binom{5}{k}$ Möglichkeiten, die k Kopf-Würfe auf die 5 möglichen Positionen zu verteilen.

$\boxed{k = 2}$

$$\binom{5}{2} = \frac{5!}{2!3!} = \frac{5 \cdot 4}{2} = 10$$

k	0	1	2	3	4	5
$P(X = k)$	$1/32$	$5/32$	$10/32$	$10/32$	$5/32$	$1/32$

Ω: *Ergebnisraum*
Ereignisse $a \subset \Omega$
wird eine Wahrscheinlichkeit $P(A)$ *zugeordnet*

Definition:

Eine diskrete Zufallsvariable (ZV) ist eine Abb. $X : \Omega \to \mathbb{R}$, die nur endlich viele oder abzählbar unendlich viele Werte annimmt.
(z.B. $Bild(X) = \{0,1,...,5\}$ oder $Bild(X) = \mathbb{N}_0$ $(= \{0,1,2,...\})$)

Die Verteilung von X ist die Abbildung $\mathbb{R} \supset A \mapsto P(X \in A)$ $\left(= P\left(\{\omega \in \Omega : X(\omega) \in A\}\right)\right)$.

Die Verteilung ist ein Wahrscheinlichkeitsmaß. Dieses wird bei diskreten ZV beschrieben durch die Wahrscheinlichkeitsmassenfunktion

$$p_X : \mathbb{R} \to [0,1],\ p_X(x) = P(X = x) \quad \forall x \in \mathbb{R},$$

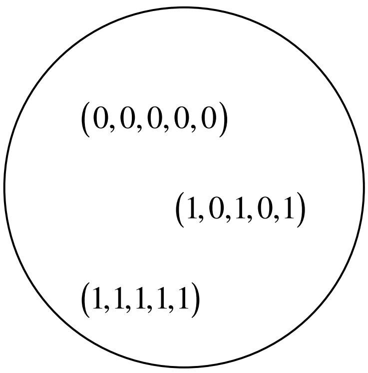
aber auch durch die Verteilungsfunktion

$$F_X : \mathbb{R} \to [0,1],\ F_X(x) = P(X \le x) \quad \forall x \in \mathbb{R}$$

Beim 5fachen Münzwurf:

$$\Omega = \{0,1\}^5, \qquad X(\omega) = \sum_{i=1}^{n} \omega_i \quad \textit{für } \omega = (\omega_1, ..., \omega_n)$$

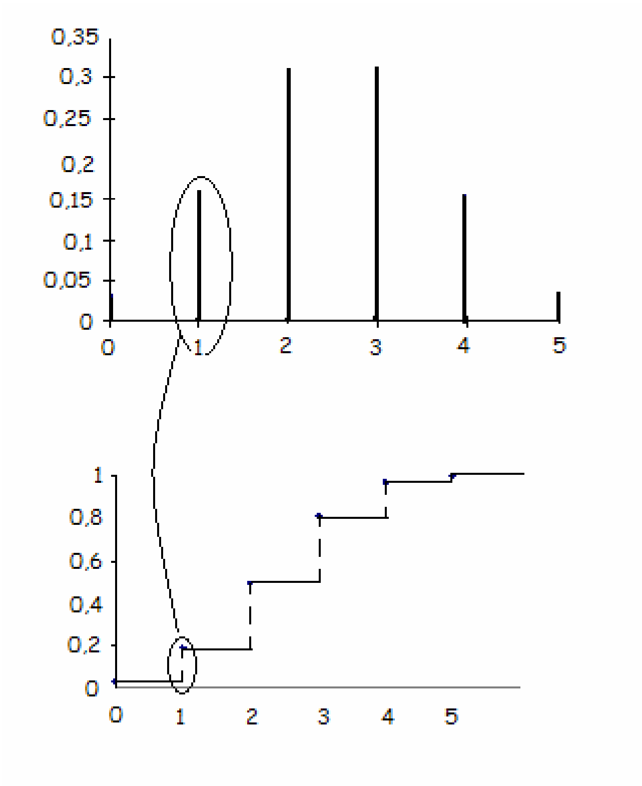
(Anzahl der Kopf-Würfe)



$$\xrightarrow{X} \mathbb{R}$$

k	0	1	2	3	4	5
$P(X = k)$	$1/32$	$5/32$	$10/32$	$10/32$	$5/32$	$1/32$

Massenfunktionen können durch Stabdiagramme illustriert werden:



Anwendung: Mit welcher Wahrscheinlichkeit bekommt man 2mal Kopf?

$$\begin{aligned} P(X \le 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= p_X(0) + p_X(1) + p_X(2) \\ &= \frac{1}{32} + \frac{5}{32} + \frac{10}{32} = \frac{1}{2} \end{aligned}$$

Wie sieht die sog. Verteilungsfunktion aus?
Bei einer diskreten ZV ist F_X von „reinem Sprungtyp“: hat Sprünge und ist zwischen diesen konstant.

Man kann aus F_X die Funktion p_X zurückgewinnen:
Auch durch die Verteilungsfunktion ist die Verteilung der ZV festgelegt.

$\boxed{X \text{ interessiert weniger als die Abbildung; die Verteilung ist wichtig}}$.

Unterschiedliche ZVen können durchaus dieselbe Verteilung haben!
Bezeichnet beispielsweise *Y* die Anzahl der „Wappenwürfe“ beim 5fachen Münzwurf hat dieselbe Verteilung wie *X* (die Anzahl der „Kopfwürfe“), aber *X* und *Y* sind natürlich nicht gleich (es gilt hier sogar $P(X=Y)=0$).

Einige Spezielle Verteilungen:

<u>Binomialverteilungen</u>: *X* heißt <u>binomialverteilt</u> mit Parametern *n* und *p* (wobei $n \in \mathbb{N},\ p \in [0,1]$), wenn gilt:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad , \textit{für } k = 0,\dots,n$$

(Nebenrechnung:

$$\begin{aligned} \sum_{k=0}^{n} P(X=k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \left(p + (1-p)\right)^n = 1 \end{aligned}$$

)

Im Münzwurf-Beispiel hat *X* die Binomialverteilung mit $n = 5$ und $p = \frac{1}{2}$.

Wir schreiben kurz: $X \sim Bin(n, p)$, hier $X \sim Bin\left(5, \frac{1}{2}\right)$.

Allgemeiner:
Wiederholt man ein Zufallsexperiment, in dem ein Ereignis A die Wahrscheinlichkeit p hat, n-mal, so ist die Anzahl X der Experimente, in denen A eintritt, $Bin(n, p)$-verteilt.
(Anzahl „Erfolge" bei n Wiederholungen und Erfolgswahrscheinlichkeit p)

Begründung: jede feste Reihenfolge von k Erfolgen und $n-k$ Misserfolgen hat Wahrscheinlichkeit $p^k (1-p)^{n-k}$.

7-maliger Würfelwurf: 1 1 6 4 1 6 3
M M E M M E M

Die Reihenfolge

M M E M M E M

hat die Wahrscheinlichkeit

$$\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}\cdot\frac{5}{6}=\underbrace{\left(\frac{1}{6}\right)}_{p}^{\overbrace{2}^{k}}\cdot\underbrace{\left(\frac{5}{6}\right)}_{(1-p)}^{\overbrace{5}^{n-k}}$$

Eine andere Reihenfolge, die auch auf 2 Erfolge führt, wäre

E M M M E M M

Wie viele solcher Reihenfolgen gibt es? $\binom{7}{2}$

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

Hypergeometrische Verteilung: X heißt hypergeometrisch verteilt mit Parametern M, m und n, wenn gilt:

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

Anwendung:
Eine Urne enthält M Kugeln, von denen m weiß sind. n Kugeln werden ohne Zurücklegen entnommen. X sei die Anzahl der weißen Kugeln in der Stichprobe.

Begründung: Es gibt $\binom{M}{n}$ mögliche Stichproben (LaPlace-Experiment; dieser Wert kommt also in den Nenner)

Anzahl der „günstigen“? Es gibt $\binom{m}{k}$ Möglichkeiten, die k weißen Kugeln in der Stichprobe aus den insgesamt m weißen Kugeln auszuwählen und $\binom{M-m}{n-k}$ Möglichkeiten, die $n-k$ nichtweißen Kugeln in der Stichprobe aus den insgesamt $M-m$ nichtweißen Kugeln auszuwählen.

<u>Geometrische Verteilung</u>: X heißt <u>geometrisch verteilt</u> mit Parameter p $(\in (0,1))$, wenn gilt:

$$P(X=k)=(1-p)^{k-1}\,p \quad \textit{für alle } k\in\mathbb{N}.$$

[Nebenrechnung:

$$\sum_{k=1}^{\infty}P(X=k)=\sum_{k=1}^{\infty}(1-p)^{k-1}p$$
$$=p\cdot\underbrace{\sum_{k=0}^{\infty}(1-p)^{k}}_{\frac{1}{1-(1-p)}}=1$$

]

Anwendung: Ein Zufallsexperiment, in dem ein bestimmtes Ereignis die Wahrscheinlichkeit p hat, wird so oft wiederholt, bis erstmalig dieses Ereignis eintritt. X sei die Nr. der Wiederholung, in der dies geschieht. Dann gilt $X\sim Geo(p)$.

Begründung:

$$X=k \Leftrightarrow \textit{Versuch } 1...k-1 \textit{ liefern Misserfolg}$$
$$\textit{Versuch } k \textit{ liefert Erfolg}$$

Konkret beim Würfelwurf: Wann kommt die erste 6?

Antwort: Das wie man nicht. Immerhin: Es geht um eine ZV X mit $X\sim Geo\left(\frac{1}{6}\right)$.

Interessant hier: Der Wertebereich ist nicht mehr endlich!

(X kann beliebig groß werden – allerdings sind die Wahrscheinlichkeiten $P(X=k)$ bei großen k sehr klein)

Zählt man die Anzahl der Misserfolge, so erhält man eine verschobene geometrische Verteilung, die auch häufig als geometrische Verteilung bezeichnet wird.

<u>Poisson-Verteilung</u>: X heißt <u>Poison-verteilt</u>, mit Parameter $\lambda(>0)$, wenn gilt:

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!} \text{ für alle } k\in\mathbb{N}_0 \quad (=\{0,1,2,....\})$$

[Nebenrechnung: $\sum_{k=0}^{\infty}P(X=k)=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=1$]

Satz: Gesetz der seltenen Ereignisse

Ist $(p_n)_{n\in\mathbb{N}}$ eine Nullfolge mit der Eigenschaft $\lim\limits_{n\to\infty} n\cdot p_n = \lambda \quad (\in (0,\infty))$,
so gilt für alle $k \in \mathbb{N}_0$

$$\underbrace{\lim_{n\to\infty}\binom{n}{k} p_n^k (1-p)^{n-k}}_{P(X=k) bei X\sim Bin(n,p_n)} = \underbrace{e^{-\lambda}\frac{\lambda^k}{k!}}_{P(X=k)\ bei\ X\sim Poisson(\lambda)}$$

Erwartungswert und Momente

Erinnerung: Die Wahrscheinlichkeitsmassenfunktion p_X zu einer diskreten ZV X ist

$$p_X(x) = P(X = x)$$

(Illustration beispielsweise durch Stabdiagramm Würfelwurf)

$$p_X(x) = \begin{cases} \frac{1}{6} & , x\in\{1,\dots,6\} \\ 0 & , sonst \end{cases}$$

Es geht nun um Kennzahlen die beispielsweise die Lage und Streuung der Verteilung von X beschreiben.

Definition:
Der Erwartungswert zu einer diskreten ZV X mit Massenfunktion p_X wird definiert durch

$$EX := \sum_x x p_X(x) \qquad \left(= \sum_x P(X = x)\right)$$

(Die Summe ist endlich oder abzählbar ∞, da $P(X = x)$ nur für endlich viele oder abzählbar ∞ viele x von 0 verschieden ist)
Mathematischer Vorbehalt: Die Summe muss absolut konvergieren, d.h. man fordert $\sum |x|\, p_X(x) < \infty$ und sagt sonst, dass der Erwartungswert nicht existiert.

(Der Würfelwurf (EW = 3,5) zeigt, dass der Erwartungswert nicht unbedingt ein Wert ist, den man erwarten würde.)

Analogie zum Schwerpunkt in der Mechanik: Der Erwartungswert beschreibt die Lage der Verteilung.

Beispiel:
Ist X poisson-verteilt mit Parameter λ, so erhält man:

$$EX = \sum_{k=0}^{\infty} k\cdot e^{-\lambda}\cdot\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1+1}}{(k-1)!} = \lambda e^{-\lambda}\underbrace{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}}_{=e^{\lambda}} = \lambda$$

Der Erwartungswert der Poisson-Verteilung ist gerade der Parameter der Verteilung.

Wichtige Erweiterungen:
Ist $g:\mathbb{R}\to\mathbb{R}$ eine Funktion, so ist $Y := g(x)$ wieder eine diskrete ZV.

Satz:

$$\boxed{E\,g(x)=\sum_x g(x)P(X=x)}$$

(Wieder unter de Vorbehalt, dass die Reihe absolut konvergiert).
Abwärtskompatibel? Ja, denn mit $g(x)=x$ (Identität) erhält man die alte Formel.

Beispiel:
Wieder: X Poisson-verteilt Mit $g(x)=x(x-1)$

$$E\big(X(X-1)\big)=\sum_{\substack{k=\cancel{0}\\ 2}}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!}$$

$$=e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2+2}}{(k-2)!}=\lambda^2 e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=\lambda^2$$

Zentrale Eigenschaften des Erwatungswertes

Zentrale Eigenschaften des Erwatungswertes sind Linearität und Isotonie („schwach monoton wachsend")
Die sind auch wichtige Rechenregeln für den Umgang mit „E"

Satz:
Seien X, Y diskrete ZVen, $c\in\mathbb{R}$. Dann gilt:
(unter der Vorraussetzung, dass die beteiligten Ewerte existieren):

$$E(X+Y)=(EX)+(EY),\qquad E(cX)=c(EX)$$
$$X\le Y\ \Rightarrow\ EX\le EY$$

Klar: Mit Induktion überträgt sich die erste Regel auf

$$E(X_1+X_2+\ldots+X_n)=EX_1+EX_2+\ldots+EX_n$$

Beispiel:
Bezeichne X die Anzahl der Erfolge bei n Versuchswiederholungen und Erfolgswahrscheinlichkeit p, so gilt $X\sim Bin(n,p)$.

$$
\begin{aligned}
EX &= \sum_{k=\underset{1}{\cancel{0}}}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= n \cdot p \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= n \cdot p \underbrace{\sum_{k=0}^{n-1} \overbrace{\frac{(n-1)!}{k!(n-1-k)!}}^{\binom{n-1}{k}} p^{k-1} (1-p)^{n-1-k}}_{=1} \\
&= n \cdot p
\end{aligned}
$$

(Begründung: $\sum_{k=0}^{m} \binom{m}{k} a^k b^{m-k} = (a+b)^m$)

Also: zu $Bin(n, p)$ gehört der Erwartungswert $n \cdot p$.

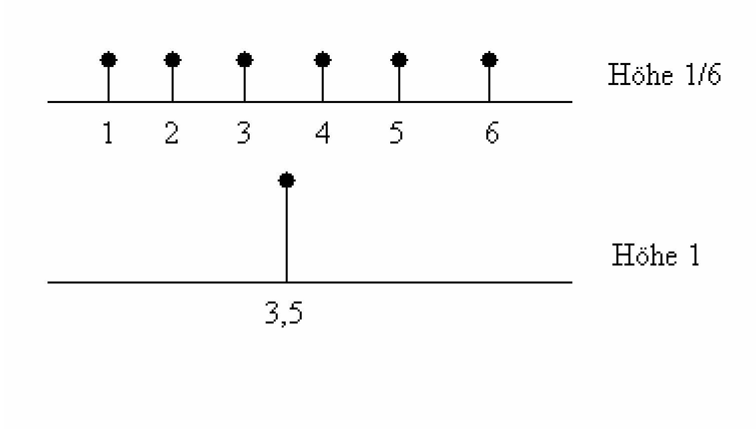
Alternativ: Schreibe $X_i = 1$, wenn im *i*-ten Durchgang Erfolg einritt, $X_i = 0$ sonst.

Klar: $X = X_1 + ... + X_n$.

$$EX_i = \sum_x xP(X_i = x) = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = p$$

Also: $EX = \underbrace{p + p + ... + p}_{n-mal} = n \cdot p$

Soweit zu Lageparametern.



Beides sind W'Massenfunktionen

oben zum Würfelwurf, unten zu einer konstanten ZV, $P(X = 3{,}5) = 1$.

In beiden Fällen erhält man denselben Erwartungswert (nämlich 3,5).

Also: wir brauchen weitere Verteilungsparameter, die beispielsweise auch die Streuung (Variabilität einer Verteilung beschreiben können.

Definition:

Das *k*-te Moment zu *X* ist EX^k.
(Insbesondere: Als erstes Moment erhält man den EWert)

Die Varianz ist $\operatorname{var}(X) = E(X - EX)^2$
(der Erwartungswert der quadrierten Abweichung vom EWert)

Die Standartabweichung ist die Wurzel aus der Varianz, also
$$\sigma(X) = \sqrt{\operatorname{var}(X)}$$
(Im Gegensatz zu Varianz hat die Standartabweihung dieselbe Dimension wie *X*.)

Rechenregel:

$$\begin{aligned}\operatorname{var}(X) &= E\left(X^2 - 2XEX + (EX)^2\right)\\ &= E\left(X^2\right) - E\left((-2EX)X\right) + E\left((EX)^2\right)\\ &= E\left(X^2\right) - 2(EX)(EX) + E\left(X^2\right) = E\left(X^2\right) - (EX)^2\end{aligned}$$

Wie zieht man Konstanten raus? Sei also $c \in \mathbb{R}$.

$$\begin{aligned}\operatorname{var}(cX) &= E\left(\underbrace{(cX)^2}_{c^2X^2}\right) - \left(\underbrace{E(cX)}_{cEX}\right)^2\\ &= c^2E\left(X^2\right) - c^2(EX)^2\\ &= c^2\left(E\left(X^2\right) - (EX)^2\right) = c^2 \operatorname{var}(X)\end{aligned}$$

Analog (selbst ist die Person): $\operatorname{var}(c + X) = \operatorname{var}(X)$

Beispiel:
Ist *X* Poisson-verteilt mit Parameter λ, so gilt nach früheren Rechnungen
$$EX = \lambda$$
$$E\left(X(X-1)\right) = \lambda^2$$

Also $E\left(X^2\right) = E\left(X(X-1)\right) + EX = \lambda^2 + \lambda$ (das zweite Moment)
$$\operatorname{var}(X) = E\left(X^2\right) - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Bei Poisson-Verteilungen ist der Parameter auch gleich der Varianz.

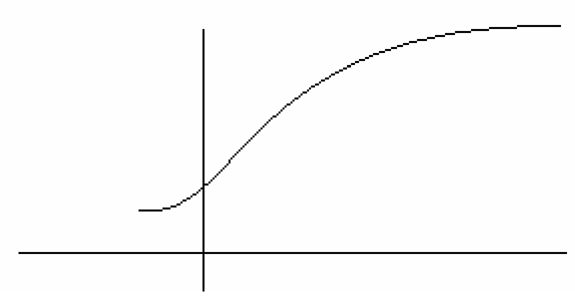
§5 Stetige Zufallsgrößen

Messfehler, Lebensdauer, etc. sind in der Regel nicht diskret.

Definition:
Eine ZV X heißt absolut stetig verteilt (kurz: stetig) mit Wahrscheinlichkeitsdichtefunktion (kurz: Dichte) f_X, wenn gilt:

$$\boxed{P(X \le x) = \int_{-\infty}^{\infty} f_X(y)\ dy \qquad , x \in \mathbb{R}}$$

Außerdem nennt man $F_X : \mathbb{R} \to [0,1], F_X(x) := P(X \le x)$, die Verteilungsfunktion zu X.



Grob: Die Dichte ist die Ableitung der Verteilungsfunktion.

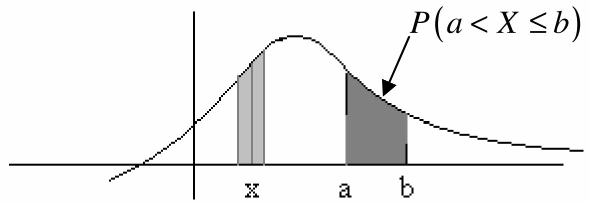
Bei stetigen ZV ist F_X immer stetig, hat insbesondere keine Sprünge (alle Sprünge haben die Höhe 0):
Ist X stetig, so gilt $P(X = x) = 0$

(Zur Erinnerung: Bei diskreten ZV ist die Verteilungsfunktion von „reinem“ Sprungtyp“)

Mit der Additivität von Wahrscheinlichkeitsmaßen folgt:

$$\begin{aligned} P(a < X \le b) &= P(X \le b) - P(X < a) \\ &= F_X(b) - F_X(a) \\ &= P(a \le X \le b) \\ &= P(a < X < b) \end{aligned}$$

Illustration:



$$\begin{aligned} F_X(b) - F_X(a) &= \int_{-\infty}^{b} f_X(y)\ dy - \int_{-\infty}^{a} f_X(y)\ dy \\ &= \int_{a}^{b} f_X(y)\ dy \end{aligned}$$

Wahrscheinlichkeiten ergeben sich als Fläche unter der Dichtefunktion.

Konsequenz: $\int_{-\infty}^{\infty} f_X(y)\ dy = 1 \quad , f_X(y) \ge 0$

wieder grob: Dichte $\Leftrightarrow$ Massenfunktion

$$\lim_{h \to 0} \frac{1}{h} P\left(x - \frac{h}{2} \le X \le x + \frac{h}{2}\right) = \lim_{h \to 0} \frac{1}{h} \int_{x-\frac{h}{2}}^{x+\frac{h}{2}} f(y)\ dy = f(x)$$, wenn f in x stetig ist.

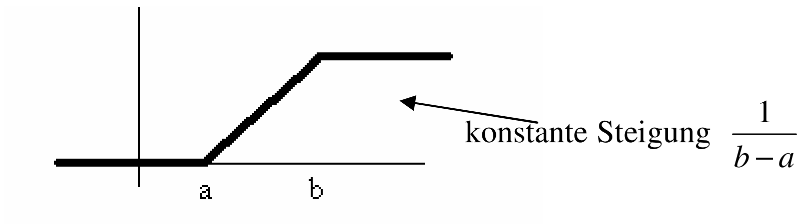
Dichten können als „infinitesimale Wahrscheinlichkeiten“ interpretiert werden.
$f(x)$ kann, im Gegensatz zu $p(x)$ bei Massenfunktionen, einen Wert >1 annehmen.

Einige wichtige Verteilungen:

Gleichverteilung:
X heißt gleichverteilt auf dem Intervall (a,b), wobei $-\infty < a < b < \infty$, kurz: $X \sim unif\,(a,b)$, wenn *X* die Dichtefunktion

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0\,, & sonst \end{cases}$$

Verteilungsfunktion:



Als Formel:

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$

Formale Randbemerkung: Einzelne Punkte spielen bei Dichten keine Rolle; auch

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \boxed{\le} x \boxed{\le} b \\ 0\,, & sonst \end{cases}$$

würde auf $X \sim unif\,(a,b)$ führen.
(Gleichverteilungen können als Analogon zu Laplace-Experimenten betrachtet werden.)

Exponentialverteilung:
X heißt <u>exponentialverteilt</u> mit Parameter λ (wobei $\lambda > 0$ beliebig), wenn

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & sonst \end{cases}$$

Verteilungsfunktion: $F(x) = 0$ für $x \le 0,$ bei $x > 0$:

$$F(x) = \int_{-\infty}^{\infty} f(y)\; dy = \int_{0}^{x} \lambda e^{-\lambda x}\; dy = -e^{-\lambda y}\Big|_0^x = 1 - e^{-\lambda x}$$

Ist *X* mit Parameter λ exponentialverteilt (kurz: $X \sim Exp(\lambda)$) dann gilt:

$$P(X \geq x+y \mid X \geq x) = \frac{P(\{X \geq x+y\} \cap \{X \geq x\})}{P(X \geq x)} = \frac{P(X \geq x+y)}{P(X \geq x)} = \frac{1-\overbrace{P(X \leq x+y)}^{F(x+y)}}{1-P(X \leq x)} = \frac{1-F(x+y)}{1-F(x)}$$

$$= \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda y} = 1-F(y) = P(X \leq y)$$

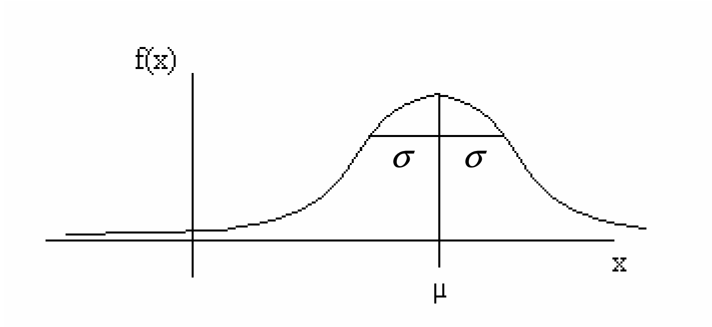
Exponentialverteilungen tauchen als Lebensdauerverteilung von nichtalternden Komponenten auf; auch bei Wartezeiten.
(stetiges Analogon zur Geometrischen Verteilung)

Normalverteilung:
X heißt normalverteilt mit Parametern μ und σ^2 (wobei $\mu \in \mathbb{R}$, $\sigma^2 > 0$), wenn

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

gilt.

Illustration:



μ beschreibt die Lage
σ die Breite der Kurve
(die Kurve ist symmetrisch um μ)
Im Falle $\mu = 0, \sigma^2 = 1$ spricht man von der Standartnormalverteilung $N(0,1)$

(allgemeine Notation: $X \sim N(\mu, \sigma^2)$

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

Es gibt keine Formel für Φ, die diese Funktion über die üblichen (transzendenten) Funktionen ausdrückt. Mit numerischen Methoden erhält man natürlich beliebig genaue Approximationen für den Wert von $\Phi(x)$ bei gegebenem Wert für *x*.

Die Werte von Φ sind vertafelt (und es gibt Computerprogramme). Was passiert, wenn man $\mu \neq 0$ *oder* $\sigma^2 \neq 1$ hat?

Lemma 5.2

$$X \sim N(\mu, \sigma^2) \Rightarrow \alpha X + \beta \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$$

In Worten: Affine Transformationen führen nicht aus der Klasse der Normalverteilungen heraus.
(Beweis: Betrachte die Verteilungsfunktion, führe Substitution aus.)

Beispiel:

a) Angenommen, es ist $P(1<Y<1.5)$ für $Y \sim N\left(1, \frac{1}{4}\right)$ gefragt. Das Lemma liefert $X := 2(Y-1)$ normalverteilt ist mit $\mu = 0, \sigma^2 = 1$:
$2Y-2$
$\alpha = 2 \quad \beta = -2$
$\mu = 1 \quad \sigma^2 = \frac{1}{4}$

In Worten: Die Transformation $y \mapsto 2(y-t)$ macht aus einer $N\left(1, \frac{1}{4}\right)$-verteilten ZV eine Standartnormalverteilte.

Damit:

$$\begin{aligned} P(1<Y<1.5) &= P\left(0 < \overbrace{2(Y-1)}^{X} < 1\right) \\ &= P(X \leq 1) - P(X \leq 0) \\ &= \Phi(1) - \Phi(0) \qquad da\, N \sim N(0,1) \\ &\approx 0,8513 - 0,5 \\ &\approx 0,3413 \end{aligned}$$

Umgang mit Normalverteilungen: $X \sim N(\mu, \sigma^2) \Rightarrow Y := \alpha X + \beta$
$\sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$

Reduziert Probleme auf den Fall $\mu = 0,\ \sigma^2 = 1$

Konkret; $X \sim (\mu, \sigma^2) \Rightarrow Y := \dfrac{X-\mu}{\sigma} \sim N(0,1)$

Begründung:

$$\begin{aligned} P(Y \leq y) &= P(X \leq \mu + \sigma y) && x = \mu + \sigma z \\ &= \int_{-\infty}^{\mu+\sigma y} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^2\right) dx && z := \frac{x-\mu}{\sigma} \end{aligned}$$

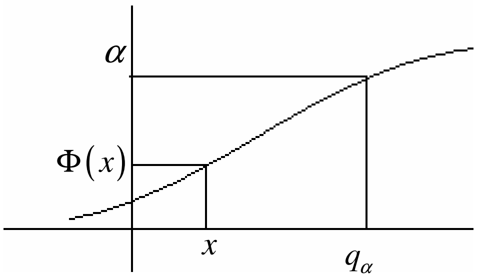
$$[dx = \sigma dz,\ -\infty < x \leq \mu + \sigma y \Leftrightarrow -\infty < z \leq y]$$

$$= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2} dz = \Phi(y)$$

Wobei Φ die Verteilungsfunktion zu $N(0,1)$.

Es geht auch „andersrum“: Als α -Quantil zur Verteilung von X bezeichnet man einen Wert q_α mit $P(X \le q_\alpha) = \alpha$.

Illustration:



Damit: bei $\alpha = 0{,}95$: $q_\alpha \approx 1{,}645$

Wie verfährt man, wenn man das α -Quantil zu $X \sim N(\mu, \sigma^2)$ bei $\mu \ne 0,\ \sigma^2 \ne 1$ braucht?

Gesucht ist ein Wert $\widetilde{q_\alpha}$ mit $P\left(X \le \widetilde{q_\alpha}\right) = \alpha$.

$$P\left(X \le \widetilde{q_\alpha}\right) = P\left(\underbrace{\frac{X-\mu}{\sigma}}_{N(0,1)} \le \frac{\widetilde{q_\alpha}-\mu}{\sigma}\right) = \Phi\left(\frac{\widetilde{q_\alpha}-\mu}{\sigma}\right) \overset{!}{=} \alpha$$

Bezeichnet $\widetilde{q_\alpha}$ das α -Quantil zu $N(0,1)$, so erhält an aus dem Ansatz $\frac{\widetilde{q_\alpha}-\mu}{\sigma} = \widetilde{q_\alpha}$,

also $\boxed{\widetilde{q_\alpha} = \mu + \sigma q_\alpha}$.

Konkret: $X \sim N\left(1, \frac{1}{4}\right)$, $\alpha = 0{,}95$: $\mu = 1$, $\sigma = \frac{1}{2}$, $\alpha = 0{,}95$, $q_\alpha = 1{,}645$, also

$$\widetilde{q_\alpha} = 1 + \tfrac{1}{2} \cdot 1{,}645 \approx 1{,}8225$$

Tabellen enthalten $\Phi(x)$ in der Regel nur für $x \ge 0$. Was macht man bei $x < 0$?

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

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

Zusammengefasst: $\Phi(-x) = 1 - \Phi(x)$

Erwartungswert und Momente

Angenommen, X ist eine (stetige) ZV mit Dichtefunktion f.
Was ist der Erwartungswert zu X?

Erinnerung: Währe X diskret, mit Massenfunktion $p(x) = P(X = x)$, so hätte man
$EX = \sum_x xp(x)$

$$\boxed{Eg(X) = \sum_x g(x) P(X = x)}$$

Kann man hieraus deduzieren, wie die Definition von EX im stetigen Fall aussehen sollte?
Idee: Monotonie $(X \le Y \Rightarrow EX \le EY)$ sollte erhalten bleiben.
(h klein)
$\lceil x \rceil = \min\{k \in \mathbb{Z} : k \le x\}$
$\lfloor x \rfloor = \max\{k \in \mathbb{Z} : k \ge x\}$
$\lfloor x \rfloor \le x \le \lceil x \rceil$

$$h\left\lfloor \frac{X}{h} \right\rfloor \le X \le h\left\lceil \frac{X}{h} \right\rceil$$

(für $h = 10^{-n}$, $n \in \mathbb{N}$ erhält man die Anfangsstücke der Dezimalentwicklung)

Soll Monotonie erhalten bleiben, so bräuchte man $\underbrace{E\left(h\left\lfloor \frac{X}{h} \right\rfloor\right)}_{Diskret} \le E(X) \le \underbrace{E\left(h\left\lceil \frac{X}{h} \right\rceil\right)}_{Diskret} \quad \forall h$

Da links und rechts diskrete ZV stehen, können wir EX durch bekannte Größen einschachteln:

$$\begin{aligned} E\left(h\left\lfloor \frac{X}{h} \right\rfloor\right) &= \sum_{k\in\mathbb{Z}} h \cdot k \cdot P\left(\left\lfloor \frac{X}{h} \right\rfloor = k\right) \\ &= \sum_{k\in\mathbb{Z}} h \cdot k \int_{hk}^{(k+1)h} f(x)\, dx \\ &= \sum_{k\in\mathbb{Z}} \int_{hk}^{(k+1)h} h\left\lfloor \frac{x}{h} \right\rfloor f(x)\, dx \\ &= \int_{-\infty}^{\infty} \overbrace{h\left\lfloor \frac{x}{h} \right\rfloor}^{\le x} f(x)\, dx \\ &\le \int xf(x)\,dx \le \int h\left\lceil \frac{x}{h} \right\rceil f(x)\,dx = \ldots = E\left(h\left\lceil \frac{X}{h} \right\rceil\right) \end{aligned}$$

Definition:
$EX := \int xf(x)dx$, vorausgesetzt, es gilt $EX := \int |x| f(x)dx < \infty$. Sonst sagt man, dass der Erwartungswert zu X nicht existiert,

Satz:
$Eg(X) := \int g(x) f(x) dx < \infty$, wieder vorausgesetzt, dass $Eg(X) := \int |g(x)| f(x) dx < \infty$ gilt.

Analogie zum diskreten Fall: ersetze Massenfunktion durch Dichte. Summe durch Integral.

	X diskret mit Massenfunktion	X stetig mit Dichtefunktion
EX	$\sum x \cdot p(x)$	$\int x \cdot f(x)\, dx$
$Eg(X)$	$\sum g(x) \cdot p(x)$	$\int g(x) \cdot f(x)\, dx$

Satz:
Linearität und Monotonie gelten weiterhin.
$E(X+Y) = EX + EY, \quad E(cX) = cEX, \quad X \le Y \Rightarrow EX \le EY$

Definition:
Das k-te Moment zu X ist EX^k,
$\operatorname{var}(X) = E(X - EX)^2, \quad \sigma(X) = \sqrt{\operatorname{var}(X)}$

Rechenregeln, wie z.B. $\operatorname{var}(X) = E(X^2) - (EX)^2$ gelten weiterhin.

Beispiele:

a) Ist X gleichverteilt auf dem Intervall $(0,1)$, also $X \sim unif(0,1)$, so gilt:

$$EX = \int x \cdot f(x)\, dx = \int_0^1 x \cdot 1\, dx = \left[\tfrac{1}{2} x^2\right]_0^1 = \frac{1}{2}$$

$$\operatorname{var}(X) = \underbrace{E(X^2)}_{Eg(X)} - \underbrace{(EX)^2}_{=1/4} = \int_0^1 x^2 \cdot 1\, dx - \frac{1}{4} = \frac{1}{12}$$

b) $X \sim N(0,1) \quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$

$$EX = \int x \cdot f(x)\, dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 0 \quad \text{(aus Symmetriegründen)}$$

Sicherheitshalber überprüfen, ob $\int |x| \cdot f(x)\, dx < \infty$ gilt:

$$\int_a^b x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \left(-\frac{1}{\sqrt{2\pi}}\right) \int_a^b \left(-x \cdot e^{-x^2/2}\right) dx$$

$$= -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Big|_a^b = \left(e^{-a^2/2} - e^{-b^2/2}\right) \frac{1}{\sqrt{2\pi}}$$

$$\xrightarrow[a,b \to \infty]{} 0$$

$$E|X| = \int |x| f(x)\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x| e^{-x^2/2}\, dx$$
$$= \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2}\, dx = -\sqrt{\frac{2}{\pi}}\, e^{-x^2/2} \Big|_0^{\infty} = \sqrt{\frac{2}{\pi}}$$

$$EX^2 = \frac{1}{\sqrt{2\pi}} \int \underset{\substack{\downarrow \\ 1}}{x} \cdot \underbrace{x e^{-x^2/2}}_{\substack{\uparrow \\ -e^{-x^2/2}}}\, dx \qquad \text{partielle Integration}$$
$$= \frac{1}{\sqrt{2\pi}}\, x\left(-e^{-x^2/2}\right)\Big|_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1 \cdot \left(-e^{-x^2/2}\right) dx$$
$$= 0 + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx$$
$$= 1$$

(Ist *f* eine Wahrscheinlichkeitsdichte, so gilt $\int f(x)\, dx = 1$)

Insgesamt: $EX^2 = 1$. Damit $\operatorname{var}(X) = EX^2 - (EX)^2 = 1 - 0^2 = 1$.

Allgemeiner: Bei $X \sim N(\mu, \sigma^2)$ erhält man mit $Y := \frac{X-\mu}{\sigma} \sim N(0,1)$:

$$E\left(\frac{X-\mu}{\sigma}\right) = EY = 0$$

Nutzt man die Linearität aus, so folgt

$$0 = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma} E(X-\mu) = \frac{1}{\sigma}(EX - \mu), \text{ also } EX = \sigma \cdot 0 + \mu = \mu.$$

Analog:

$$\operatorname{var}(X) = \sigma^2 \operatorname{var}\left(\frac{X}{\sigma}\right)$$
$$= \sigma^2 \operatorname{var}\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \sigma^2 \operatorname{var}(Y) = \sigma^2$$

Die Parameter μ und σ^2 sind gerade Erwartungswerte und Varianz zu *X*.

§6 Zufallsvektoren und Unabhängigkeit

Bisher: Eindimensionale Zufallsgrößen $X:\Omega\to\mathbb{R}$.
Im diskreten und im stetigen Fall wird die Verteilung von *X* durch die zugehörige Verteilungsfunktion

$$F:\mathbb{R}\to\mathbb{R},\qquad F(x)=P(X\le x)$$

beschrieben.

Wir betrachten nun Zufallsvektoren, also mehrdimensionale Zufallsgrößen

$X:\Omega\to\mathbb{R}^n,\quad X=\begin{pmatrix}X_1\\ \dots\\ X_n\end{pmatrix}$, wobei die X_i eindimensionale Zufallsgrößen sind.

Beispiel:
Ein Zuhörer wird zufällig ausgewählt.
$\Omega=$Menge der Zuhörer $\qquad \omega\in\Omega:$der zufällig ausgewählte Zuhörer

$X_1(\omega)=$Körpergröße
$X_2(\omega)=$Gewicht
$X_3(\omega)=$Matrikelnummer

$X(\omega)=\begin{pmatrix}X_1(\omega)\\ X_2(\omega)\\ X_3(\omega)\end{pmatrix}$ ist ein 3-dim. Zufallsvektor.

Typische Fragestellung: Hängen die Komponenten „irgendwie" voneinander ab?

Wir betrachten vorläufig nur 2-dim. Vektoren $\begin{pmatrix}X\\ Y\end{pmatrix}$ mit Komponenten *X, Y*

Definition:

a) $F_{X,Y}:\mathbb{R}^2\leftarrow\mathbb{R}$ mit
$F_{X,Y}(x,y):=P(X\le x,Y\le y)$ für alle $x,y\in\mathbb{R}$
Heißt gemeinsame Verteilungsfunktion von *X* und *Y* (oder auch Verteilungsfunktion des Zufallsvektors $\begin{pmatrix}X\\ Y\end{pmatrix}$).

b) Nehmen *X* und *Y* nur endlich oder abzählbar unendlich viele Werte an, so nennt man $p_{X,Y}:\mathbb{R}^2\to[0,1]$ mit $p_{X,Y}(x,y):=P(X=x,Y=y)$ die gemeinsame Massenfunktion zu X und Y (oder auch Massenfunktion zu $\begin{pmatrix}X\\ Y\end{pmatrix}$).

c) Gibt es eine Funktion $f_{X,Y}:\mathbb{R}^2\to\mathbb{R}_+$ mit

$$F_X(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{X,Y}(s,t)\ dt\,ds\quad\forall x,y\in\mathbb{R}$$

So heißt $f_{X,Y}$ gemeinsame Dichtefunktion zu *X* und *Y*
(oder auch Dichtefunktion zu $\begin{pmatrix}X\\ Y\end{pmatrix}$).

Warum sollte man *X* und *Y* unabhängig nennen?
Erinnerung: Ereignisse *A* und *B* heißen unabhängig, wenn $P(A \cap B) = P(A) \cdot P(B)$ gilt.

$$F_{X,Y}(x,y) = P(\{X \le x\} \cap \{Y \le y\}) \overset{!}{=} P(X \le x) P(Y \le y) = F_X(x) F_Y(y)$$

Definition:
Sind X,Y ZV mit Verteilungsfunktion F_X, F_Y und gemeinsamer Verteilungsfunktion $F_{X,Y}$, so heißen diese (stochastisch) unabhängig, wenn gilt:

$$F_{X,Y}(x,y) = F_X(x) F_Y(y) \quad \text{für alle } x, y \in \mathbb{R}.$$

„Übertragung" auf den diskreten und stetigen (Spezial-) Fall:

Satz:

a) Sind *X* und *Y* diskret, so sind sie genau dann unabhängig, wenn gilt
$$\underbrace{p_{X,Y}(x,y)}_{P(X=x,Y=y)} = \underbrace{p_X(x)}_{P(X=x)} \underbrace{p_Y(y)}_{P(Y=y)} \qquad \textit{für alle } x, y \in \mathbb{R}$$

b) Ist $f_{X,Y}$ eine gemeinsame Dichte von *X* und *Y*, so sind *X* und *Y* genau dann unabhängig, wenn
$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \qquad \textit{für alle } x, y \in \mathbb{R}$$
gilt.
(technische Marginale: Da Dichten nicht eindeutig sind, bezieht sich dies auf geeignete Versionen).

Grob: Unabhängig bedeutet „Faktorisierung" der gemeinsamen Verteilungsfunktion / Massenfunktion / Dichtefunktion

Wichtige Frage: Wie erhält man aus der gemeinsamen Verteilungsfunktion bzw. Massenfunktion bzw. Dichtefunktion die entsprechende Funktion für die Komponenten?

Satz:

a) X, Y ZV mit gemeinsamer Verteilungsfunktion $F_{X,Y}$.
Dann ist $F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y)$ die Verteilungsfunktion zu *X*,
$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x,y)$ die Verteilungsfunktion zu *Y*.

b) Im diskreten Fall ist $p_X(x) = \sum_y p_{X,Y}(x,y)$ (man erhält die „marginale Massenfunktion" durch „Aussummieren" der anderen Komponente.)
Analog: $p_Y(y) = \sum_x p_{X,Y}(x,y)$

c) Ist $f_{X,Y}$ gemeinsame Dichtefunktion zu *X* und *Y*, so ist $f_X(x) = \int f_{X,Y}(x,y)\, dy$ Dichtefunktion zu *X* und $f_Y(y) = \int f_{X,Y}(x,y)\, dx$ Dichtefunktion zu *Y*.

Schließlich: Wie berechnet man Erwartungswerte von Funktionen von X und Y?
Erinnerung: $f_Y(y) = \int f_{X,Y}(x,y)\, dx$ im diskreten Fall, $Eg(X) = \int g(x) f_X(x)\, dx$ im stetigen Fall.

Satz:

$$\boxed{Eg(X,Y) = \sum_y \sum_x g(x,y)\, p_{X,Y}(x,y)}$$

im diskreten Fall.

$$\boxed{Eg(X,Y) = \int\int g(x,y)\, f_{X,Y}(x,y)\; dx\, dy}$$

wenn X, Y gemeinsame Dichte $f_{X,Y}$ haben.

Beispiel:
Eine Urne enthält zwei weiße, drei schwarze und zwei rote Kugeln. Drei Kugeln werden zufällig und ohne Zurücklegen entnommen.

X: Anzahl der entnommenen weißen Kugeln.
Y: Anzahl der entnommenen schwarzen Kugeln.

Wie sieht die gemeinsame Massenfunktion aus?

Abschnitt2: $P(X=k, Y=l) = \dfrac{\binom{2}{k}\binom{3}{l}\binom{7-2-3}{3-k-l}}{\binom{7}{3}}$

Tabelle:

Y \ X	0	1	2	$P(Y=y)$ Massenfkt zu Y
0	0	$2/35$	$2/35$	$4/35$
1	$3/35$	$12/35$	$3/35$	$18/35$
2	$6/35$	$6/35$	0	$12/35$
3	$1/35$	0	0	$1/35$
$P(X=x)$ Massenfkt zu X	$10/35$	$20/35$	$5/35$	$35/35$ Gesamtsumme muss 1 sein

Intuitiv ist klar, dass X und Y nicht unabhängig sein können.
Beweis?
Angenommen, X und Y wären unabhängig. Dann müsste gelten:

$$p_{X,Y}(k,l) = p_X(k)\,p_Y(l) \qquad \text{für alle } k,l$$

Nun gilt für $k=0,\ l=0$.

$p_{X,Y}(0,0)=0, \quad p_X(0)=\frac{10}{35},\ p_Y(0)=\frac{4}{35}$

und damit $p_{X,Y}(0,0) \neq p_X(0)\,p_Y(0)$

Wichtig ist, dass man mit der gemeinsamen Verteilung Wahrscheinlichkeiten und Erwartungswette ausrechnen kann, die sich auf **beide** Zufallsgrößen beziehen:

$$\underbrace{P(X+Y=3)}_{\substack{\text{(keine rote Kugel in}\\ \text{der Stichprobe)}}} = \tfrac{1}{35}+\tfrac{6}{35}+\tfrac{3}{35}=\tfrac{10}{35}$$

$$P(X>Y)=\frac{2+2+3}{35}=\tfrac{7}{35}$$

$$E(X\cdot Y)=\sum_{k=0}^{2}\sum_{l=0}^{3} k\cdot l\cdot p_{X,Y}(k,l)$$
$$=\tfrac{6}{7}$$

<u>*X, Y* Zufallsvariablen</u>

Die gemeinsame Verteilungsfunktion $F_{X,Y}$ ist $F_{X,Y}(x,y)=P(X\le x, Y\le y)$
Analog: gemeinsame Massenfunktion $f_{X,Y}(x,y)=P(X=x,Y=y)$ im diskreten Fall, gemeinsame Dichtefunktion im stetigen Fall:

$Eg(X,Y)=\sum_x\sum_y g(x,y)\cdot P_{X,Y}(x,y)$ im diskreten Fall
$Eg(X,Y)=\int\int g(x,y)\,f_{X,Y}(x,y)\,dx\,dy$ im stetigen Fall.

Aus $F_{X,Y}$ erhält man die marginale Verteilungsfunktion F_X, F_Y zu X und Y durch:

$$F_X(x)=\lim_{y\to\infty}F_{X,Y}(x,y), \quad F_Y(y)=\lim_{x\to\infty}F_{X,Y}(x,y)$$

- bei Massenfunktion: $P_X(x)=\sum_y P_{X,Y}(x,y)$ (Aussummieren der anderen Komponente)
- bei Dichtefunktion: $f_X(x)=\int f_{X,Y}(x,y)$ (Ausintegrieren der anderen Komponente)

Unabhängigkeit: $F_{X,Y}(x,y)=F_X(x)\cdot F_Y(y)$

$$P_{X,Y}(x,y)=P(X=x,Y=y)=P(\{X=x\}\cap\{Y=y\})$$
$$=P(X=x)\cdot P(Y=y)=P_X(x)\cdot P_Y(y)$$

(Analog bei Dichten)

Satz: (Multiplikationsregel)
Sind X und Y unabhängig, und existieren die Erwartungswerte,
so gilt: $E(X \cdot Y) = (EX) \cdot (EY)$.

Beweis: (im stetigen Fall)

$$
\begin{aligned}
E(X \cdot Y) &= Eg(X,Y) \\
&= \int\int g(x,y) \cdot f_{X,Y}(x,y)\, dx\, dy \\
&= \int\int x \cdot y \cdot f_X(x) \cdot f_Y(y)\, dx\, dy \\
&= \int \left(y \cdot f_Y(y) \underbrace{\left(\int x \cdot f_X(x)\, dx \right)}_{=EX} dy \right) \\
&= (EX) \cdot \left(\int y \cdot f_Y(y)\, dy \right) = (EX) \cdot (EY)
\end{aligned}
$$

Definition:
Die Kovarianz $\operatorname{cov}(X,Y)$ wird definiert durch $\operatorname{cov}(X,Y) = E\left((X - EX)(Y - EY)\right)$
(unter der Vorraussetzung, dass die beteiligten Erwartungswerte existieren)
Sind die Streuungen $\sigma(X)$ $\left(= \sqrt{\operatorname{var}(X)} = \sqrt{E(X - EX)^2} \right)$, $\sigma(Y)$ echt größer 0, so nennt man $\rho(X,Y) = \dfrac{\operatorname{cov}(X,Y)}{\sigma(X) \cdot \sigma(Y)}$ den Korrelationskoeffizienten von X und Y.
(mit der Cauchy-Schwarz-Ungleichung kann man zeigen, dass $-1 \le \rho(X,Y) \le 1$ gilt)

Beide messen die (lineare) Abhängigkeit; $\rho(X,Y)$ ist im Gegensatz zu $\operatorname{cov}(X,Y)$ dimensionslos.

Rechenregel:

$$
\begin{aligned}
\operatorname{cov}(X,Y) &= E\left(XY - (EX)Y\right)Y - X(EY) + (EX)(EY) \\
&= E(XY) - (EX)(EY) - (EX)(EY) + (EX)(EY) \\
&= E(XY) - (EX)(EY)
\end{aligned}
$$

Damit klar:
Wenn X und Y unabhängig sind, so ist $\operatorname{cov}(X,Y) = 0$ und damit natürlich auch $\rho(X,Y) = 0$.

Bei $\rho(X,Y) = 0$ (oder allgemeiner $\operatorname{cov}(X,Y) = 0$) nennt man X und Y unkorreliert. Also: unabhängige ZV sind unkorreliert.

ACHTUNG: Die Umkehrung gilt nicht!

Sind $X_1,...,X_n$ Zufallsvariablen, so gilt

$$\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = E\left(\left(\left(\sum_{i=1}^{n} X_i\right) - \underbrace{E\left(\sum_{i=1}^{n} X_i\right)}_{=\sum_{i=1}^{n} EX_i}\right)^2\right)$$

$$= E\left(\left(\sum_{i=1}^{n}(X_i - EX_i)\right)^2\right)$$

$$= E\left(\sum_{i=1}^{n}(X_i - EX_i)^2 + \sum_{\substack{i,j=1\\ i\neq j}}(X_i - EX_i)(X_j - EX_j)\right)$$

$$= \sum_{i=1}^{n} E(X_i - EX_i)^2 + \sum_{\substack{i,j=1\\ i\neq j}}(X_i - EX_i)(X_j - EX_j)$$

Also:

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

Satz: (Gleichheit von Bienaymé)

Wenn $X_1,...,X_n$ unabhängig sind, so gilt:

$$\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{var}(X_i)$$

Beispiel:

Wenn *X* und *Y* die gemeinsame Verteilungsfunktion

$$F_{X,Y}(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & , x, y \geq 0 \\ 0 & , \text{sonst} \end{cases}$$ haben, so gilt:

$F_X(x) = \lim_{y \to x} F_{X,Y}(x,y) = 1 - e^{-x}$, $d.h.\ X \sim Exp(1)$

Analog $F_Y(y) = 1 - e^{-y}$, also auch $X \sim Exp(1)$

Damit: $F_X(x) F_Y(y) = (1 - e^{-x})(1 - e^{-y}) = 1 - e^{-x} - e^{-y} + e^{-x-y} = F_{X,Y}$

Also sind *X* und *Y* unabhängig.

Als gemeinsame Dichte ergibt sich damit $f_{X,Y}(x,y) = f_X(x) f_y(y) = e^{-x-y}$.

(Kontrolle: $\frac{\partial}{\partial x}\frac{\partial}{\partial y} F_{X,Y}(x,y) = \frac{\partial}{\partial x}\left(e^{-y} - e^{-x-y}\right) = e^{-x-y}$)

Typische Fragestellung: $P(X \le 2Y) = ?$

$$\begin{aligned}
P(X \le Y) &= \iint\limits_{\{x,y|x\le 2y\}} f_{X,Y}(x,y)\,dx\,dy \\
&= \int_0^\infty \int_{x/2}^\infty e^{-x-y}\,dx\,dy \\
&= \int_0^\infty e^{-x} \underbrace{\left(\int_{x/2}^\infty e^{-y}\,dy \right)}_{\left[-e^{-y}\right]_{x/2}^\infty} dx = \int_0^\infty \underbrace{e^{-x} e^{-x/2}}_{=e^{-3x/2}}\,dx \\
&= \left[-\frac{2}{3} e^{-3x/2} \right]_0^\infty = \frac{2}{3}
\end{aligned}$$

Wir besprechen noch 2 wichtige Familien von mehrdimensionalen Verteilungen:

Beispiel:
Es sei $A_1,...,A_n$ eine Ereignispartition in einem Zufallsexperiment ($A_1,...,A_n$ disjunkt, $A_1 \cup ... \cup A_n = \Omega$) Es sei p_i die Wahrscheinlichkeit von A_i (Würfelwurf: $\Omega = \{1,...,6\}$, $A_i = \{i\}$, $p_i = 1/6$). Wir wiederholen *n*-Mal. Es sei X_i die Anzahl der Versuche mit Ergebnis in A_i.
(Konkret beim Würfelwurf: 4, 4, 2, 1, 4, 6, 6, 3, 2, 1 $n = 10$)

Der Zufallsvektor *X* ist sechsdimensional:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{pmatrix}, \text{ hier: } \begin{pmatrix} 2 \\ 2 \\ 1 \\ 3 \\ 0 \\ 2 \end{pmatrix} \qquad \left(\text{Kontrolle: } \sum_{i=1}^n X_i \text{ muss gleich } n \text{ sein!}\right)$$

Allgemein erhält man die Massenfunktion:

$$P\left(X = \begin{pmatrix} k_1 \\ ... \\ k_r \end{pmatrix} \right) = \underbrace{\frac{n!}{k_1! \cdot k_2! \cdot ... \cdot k_r!}}_{=\binom{n}{k_1,...,k_r} \text{Multinomialkoeffizienten (vgl. Vorl. "Diskrete Strukturen")}} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot ... \cdot p_r^{k_r}$$

Begründung:
Wie viele Möglichkeiten gibt es, die k_1-Ereignisse vom Typ1 auf die n Positionen zu verteilen? $\binom{n}{k_1}$. Es bleiben $n-k_1$ Positionen übrig. Möglichkeiten k_2 vom Typ2 auf diese zu verteilen? $\binom{n-k_1}{k_2}$. Für Typ3: $\binom{n-k_1-k_2}{k_3}$,…… usw.

$$\binom{n}{k_1}\cdot\binom{n-k_1}{k_2}\cdot\binom{n-k_1-k_2}{k_3}\cdot\ldots\cdot\binom{n-k_1-\ldots-k_{r-1}}{k_r}$$

$$=\frac{n!}{k_1!\cancel{(n-k_1)!}}\cdot\frac{\cancel{(n-k_1)!}}{k_2!\cancel{(n-k_1-k_2)!}}\cdot\frac{\cancel{(n-k_1-k_2)!}}{k_3!\cancel{(n-k_1-k_2-k_3)!}}\cdot\ldots\cdot\frac{\cancel{(n-k_1-\ldots-k_{r-1})!}}{k_r!\underbrace{(n-k_1-\ldots-k_r)!}_{0!=1}}$$

$$=\frac{n!}{k_1!\cdot\ldots\cdot k_r!}$$

Man nennt diese Verteilung die Multinomialverteilung mit Parametern n (Anzahl (Wdh.)) und $(p_1,\ldots,p_r)$ (die Einzelnen)
Zufallsvektoren

$$X=\begin{pmatrix}X_1\\ \ldots\\ X_n\end{pmatrix}$$

Eine wichtige mehrdimensionale, diskrete Verteilung ist die Multinomialverteilung mit Parametern n (Anzahl der Versuchswiederholungen) und

$p=\begin{pmatrix}p_1\\ \ldots\\ p_d\end{pmatrix}$ (die einzelnen Erfolgswahrscheinlichkeiten).

Für $k_1,\ldots,k_d\in\mathbb{N}_0$ $\left(=\{0,1,2,\ldots\}\right)$ mit $k_1+\ldots+k_d=n$ gilt

$$P\left(X=\begin{pmatrix}k_1\\ \ldots\\ k_d\end{pmatrix}\right)=\frac{n!}{k_1!\cdot\ldots\cdot k_d!}p_1^{k_1}\cdot\ldots\cdot p_d^{k_d}$$

Konkret: Ein fairer Würfel wird 10 mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass 1 und 6 je dreimal, ale anderen Ergebnisse jeweils einmal vorkommen?

$n=10$,
$d=6$
$p_1,\ldots,p_6=1/6$

$$\frac{10!}{3!\cdot1!\cdot1!\cdot1!\cdot1!\cdot3!}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^3=\frac{175}{104976}\approx0{,}00167$$

Eine wichtige, mehrdimensionale, stetige Verteilung ist die (d-dimensionale) Normalverteilung $N_d(\mu,\Sigma)$ mit Parametern μ und Σ: Die zugehörige Dichtefunktion ist

$$\boxed{f(x \mid \mu,\Sigma) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2} \underbrace{(x-\mu)^t}_{\text{Zeilenvektor der Länge } d} \overbrace{\Sigma^{-1}}^{d\times d} \underbrace{(x-\mu)}_{\text{Spaltenvektor der Länge } d}\right)}$$

(Erinnerung: bei $d=1$ hat man $f(x \mid \mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$)

.

Σ ist eine $d\times d$ -Matrix, μ ein d-dimensionaler Vektor, A^t ist die transponierte Matrix zu A.
Interpretation der Parameter:

$$\mu = \begin{pmatrix} \mu_1 \\ \dots \\ \mu_d \end{pmatrix} \quad mit \quad \begin{matrix} \mu_1 = EX_1 \\ \\ \mu_d = EX_d \end{matrix}$$

, also $\mu = EX$ ist der Erwartungswertvektor zu X.

$\Sigma = (\Sigma)_{i,j=1}^d$, wobei in Zeile i und Spalte j $\operatorname{cov}(X_i, X_j)$ steht:
Σ ist die (Ko-)Varianzmatrix zu X.

Wie sieht die Verteilung der Summe von unabhängigen ZV aus?
Im diskreten Fall haben wir eine ZV X mit Massenfunktion p_X sowie eine ZV Y mit Massenfunktion p_Y. Wie sieht die Massenfunktion p_{X+Y} zu $X+Y$ aus?

$$\begin{aligned} p_{X+Y}(z) &= P(X+Y=z) = \sum_x P(X+Y=z, X=x) \\ &= \sum_x P(X=x, Y=z-x) \\ &= \sum_x \underbrace{P(X=x)}_{p_X(x)} \cdot \underbrace{P(Y=z-x)}_{p_Y(z-x)} \end{aligned}$$

Also $\boxed{p_{X+Y}(z) = \sum_x p_X(x) p_Y(z-x) \quad \left(= \sum_y p_y(y)\, p_X(z-y)\right)}$

Man nennt p_{X+Y} die Faltung von p_X und p_Y.
Analog: Sind X und Y stetig und Wahrscheinlichkeitsdichten f_X, f_Y, so hat $X+Y$ die Dichte

$$\boxed{f_{X+Y}(z) = \int f_X(x)\, f_Y(z-x)\, dx = \int f_Y(y) f_X(z-y)\, dy}$$

Beispiel: Sind X und Y unabhängig und Poisson-verteilt mit Parametern λ_X und λ_Y so erhält man $p_{X+Y}(k) = \sum_j p_X(j)\, p_Y(k-j)$

$$= \boxed{\frac{1}{k!}} \sum_{j=0}^{k} \left(\boxed{e^{-\lambda_X}} \frac{(\lambda_X)^j}{j!} \right) \left(\boxed{e^{-\lambda_Y}} \frac{(\lambda_Y)^{k-j}}{(k-j)!} \right) \boxed{k!}$$

$$= \frac{1}{k!} e^{-(\lambda_X + \lambda_Y)} \boxed{\sum_{j=0}^{k} \binom{k}{j}} \lambda_X^j \lambda_Y^{k-j}$$

$$= e^{-(\lambda_X + \lambda_Y)} \frac{(\lambda_X + \lambda_Y)^k}{k!} \qquad (k \in \mathbb{N}_0)$$

Dies ist die Massenfunktion zur Poisson-Verteilung mit Parameter $\lambda_X + \lambda_Y$:

$$\boxed{X, Y \text{ unabh., } X \sim Po(\lambda_X), Y \sim Po(\lambda_Y) \Rightarrow X + Y \sim Po(\lambda_X + \lambda_Y)}$$

(Die Faltung von Poisson-Verteilungen ist wieder eine Poisson-Verteilung)

$E(X+Y) = EX + EY \quad , X \sim Po(\lambda) \quad \Rightarrow \quad EX = \lambda$

Beispiel:
X, Y unabhängig, $X \sim N(0,1)$, $Y \sim N(0,1)$

$$f_{X+Y}(z) = \int f_X(x) f_Y(z-x)\, dx \qquad\qquad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \forall x \in \mathbb{R}$$

$$= \frac{1}{2\pi} \int \exp \underbrace{\left(-\frac{1}{2} x^2 - \frac{1}{2} (z-x)^2 \right)}_{-\frac{1}{2}\underbrace{\left(x^2+z^2-2xz+x^2\right)}_{2\left(x-\frac{1}{2}z\right)^2+\frac{1}{2}z^2}} dx$$

$$= \frac{1}{2\pi} e^{-z^2/4} \int \exp\left(-\left(x - \frac{1}{2} z \right)^2 \right) dx$$

Substitution: $y = \left(x - \frac{1}{2} z \right) \cdot \sqrt{2} \qquad dx = \frac{1}{\sqrt{2}}$

$$= \frac{1}{2\pi} e^{-z^2/4} \frac{2\pi}{\sqrt{2}} \int \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$

$$= \frac{1}{2\sqrt{\pi}} e^{-z^2/4}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (z-\mu)^2 \right)$$

mit $\mu = 0,\ \sigma^2 = 2$

Also $X + Y \sim N(0,2)$.
Allgemein gilt:

$$\boxed{X, Y \text{ unabh., } X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2) \Rightarrow X + Y \sim (\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)}$$

Die Faltung von Normalverteilungen ist wieder eine Normalverteilung.

Satz: (Transformationsformel für Wahrscheinlichkeitsdichten)
Es seien U und V offene Teilmengen von $\mathbb{R}^d$ und $\Psi : U \to V$ eine bijektive, stetig differenzierbare Abbildung mit $\det \Psi'(x) \neq 0 \ \forall x \in U$
(Ψ' ist hierbei die Matrix der partiellen Ableitungen)
Ist X ein Zufallsvektor mit Dichte f_X, so hat der Zufallsvektor $Y := \Psi(X)$ Dichte

$$\boxed{f_Y(y) = \left|\det\left(\Psi^{-1}\right)'(y)\right| f_X\left(\Psi^{-1}(y)\right)}$$

Beispiel:

Wir betrachten bei $d = 2$ die Transformation Ψ auf Polarkoordinaten: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} r \\ \vartheta \end{pmatrix}$.

$U = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 : x_1 < 0 \text{ oder } x_2 \neq 0 \right\}$ die geschlitzte Ebene. $V = (0,\infty) \times (0, 2\pi)$.

$$\Psi^{-1}\left(\begin{pmatrix} r \\ \vartheta \end{pmatrix}\right) = \begin{pmatrix} r \cdot \cos(\vartheta) \\ r \cdot \sin(\vartheta) \end{pmatrix}$$

Matrix der partiellen Ableitung:

$$\Psi^{-1}\left(\begin{pmatrix} r \\ \vartheta \end{pmatrix}\right) = \begin{pmatrix} \cos(\vartheta) & -r\sin(\vartheta) \\ \sin(\vartheta) & r\cos(\vartheta) \end{pmatrix}$$

$$\begin{aligned} \det\left(\Psi^{-1}\left(\begin{pmatrix} r \\ \vartheta \end{pmatrix}\right)\right) &= \cos(\vartheta)\, r\cos(\vartheta) + \sin(\vartheta)\, r\sin(\vartheta) \\ &= r \end{aligned}$$

Es seien nun X_1, X_2 unabhängig und $N(0,1)$-verteilt.

$$\begin{pmatrix} R \\ \Theta \end{pmatrix} = \Psi\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \qquad f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(x_1^2 + x_2^2\right)\right)$$

(Produkt der einzelnen Dichten wegen Unabhängigkeit)

$$f_{R,\Theta}(r,\vartheta) = r\frac{1}{2\pi}\exp\left(-\frac{1}{2}r^2\right) \quad \text{für } r > 0,\ 0 < \vartheta < 2\pi$$

Marginalverteilungen?

$$f_R(r) = \int f_{R,\Theta}(r,\vartheta)\, d\vartheta = r \cdot e^{-r^2/2}$$

$$f_\Theta(\vartheta) = \int_0^\infty r\frac{1}{2\pi}e^{-r^2/2}\, dr = \frac{1}{2\pi} \qquad \text{für } 0 < \vartheta < 2\pi \quad \text{Gleichverteilung}$$

§7 Erzeugende Funktionen

Ist X eine Zufallsgröße mit Werten in $\mathbb{N}_0 \; (=\{0,1,2,...\})$, so heißt

$$g_X(s) = \sum_{k=0}^{\infty} P(X=k)s^k$$

die wahrscheinlichkeitserzeugende Funktion zu X (oder: zur Verteilung von X); im Grunde die Potenzreihen zur Wahrscheinlichkeitsmassenfunktion.

Beispiel für eine Transformation:
Aus der Folge $\left(P(X=k)\right)_{k\in\mathbb{N}_0}$ wird die Funktion g_X.
Kommt man zurück? Konvergenzradius ≥ 1.
$g_X(0) = P(X=0) \quad g_X'(0) = P(X=1) \quad$ etc.

$$\boxed{P(X=n) = \frac{1}{n!}\frac{d^n}{ds^n}\, g_X(s)|_{s=0}}$$

Im Falle $EX < \infty$ hat man

$$\boxed{\begin{aligned} EX &= \sum_{k=1}^{\infty} k\,P(X=k) = \sum_{k=1}^{\infty} P(X=k)\frac{d}{ds}s^k\,|_{s=1} \\ &= \frac{d}{ds}\left(\sum_{k=0}^{\infty} P(X=k)s^k\right)|_{s=1} \\ &= g_X'(1) \end{aligned}}$$

Analog: $g_X''(1) = \sum_{k=0}^{\infty} k(k-1)P(X=k) = E\left(X(X-1)\right)$

Damit:

$$\boxed{\begin{aligned} \operatorname{var}(X) &= E(X^2) - (EX)^2 = E\left(X(X-1)\right) + EX - (EX)^2 \\ &= g_X''(1) + g_X'(1) - \left(g_X'(1)\right)^2 \end{aligned}}$$

Mit diesen Funktionen lassen sich also Erwartungswerte, Varianzen, … bestimmen.
Außerdem: Sind X und Y unabhängig, mit erz. Funktionen g_X und g_Y, so erhält man als erz. Funktion der Summe

$$\begin{aligned} g_{X+Y}(s) &= E\left(s^{X+Y}\right) = E\left(s^X \cdot s^Y\right) \quad (\text{mit } X \text{ und } Y \text{ sind auch } s^X \text{ und } s^Y \text{ unabhängig}) \\ &= \left(Es^X\right)\cdot\left(Es^Y\right) \; (\text{Multiplikationsregel für Erwartungswerte}) \\ &= g_X(s)\cdot g_Y(s) \end{aligned}$$

Beispiel:
Ist X poisson-verteilt mit Parameter λ, so gilt

$$g_X(s) = \sum_{k=0}^{\infty} P(X=k)s^k = \sum_{k=0}^{\infty}\left(e^{-\lambda}\frac{\boxed{\lambda^k}}{k!}\right)\boxed{s^k}$$
$$= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda s)^k}{k!} = e^{-\lambda}e^{\lambda s} = e^{\lambda(s-1)}$$

$$g_X^{'}(s) = \lambda e^{\lambda(s-1)} \qquad g_X^{''}(s) = \lambda^2 e^{\lambda(s-1)}$$

$$EX = g_X^{'}(1) = \lambda, \quad \operatorname{var}(X) = g_X^{''}(1) + g_X^{'}(1) - \left(g_X^{'}(1)\right)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Sind X, X unabhängig, $X \sim Po(\lambda)$, $Y \sim Po(\mu)$, so gilt

$$g_{X+Y}(s) = g_X(s)\cdot g_Y(s) = e^{\lambda(s-1)}e^{\mu(s-1)}$$
$$= e^{(\lambda+\mu)(s-1)}$$

Da dies die erz. Funktion zu $Po(\lambda+\mu)$ ist, folgt hieraus, dass $X+Y$ wieder poisson-verteilt ist, und zwar mit Parameter $\lambda+\mu$. (man vergleiche mit der Faltungsformel aus §6)

Ein Einfachst<u>beispiel</u> zur Mustererkennung

Wir betrachten eine zufällige Folge von 0en und 1en, wobei die Werte in den einzelnen Positionen unabhängig sind, und mit Wahrscheinlichkeit p eine 1 kommt. (mit W. $1-p$ eine 0).
(Münzwurf: $(0,0,1,1,...)$, $p = 1/2$)
Wir lange dauert es, bis erstmalig ein bestimmtes „Muster" hier konkret „1,1" auftaucht?
(Im Beispiel: Im 4. Versuch ist das Muster das 1. Mal komplett aufgetreten.)

Sei X_n der Wert in Position n; die ZV X_n, $n \in \mathbb{N}$, sind also unabhängig mit $P(X_n = 1) = p = 1 - P(X_n = 0)$. Es geht um

$$T := \inf\{n \geq 2 : X_{n-1} = X_n = 1\}$$

Kann man die erzeugende Funktion zu T bestimmen?

Klar: $P(T=0) = P(T=1) = 0, \quad P(T=2) = p^2$
Bei $n \geq 3$ zerlegen wir nach dem Wert von X_1:

$$P(T=n) = P(X_1 = 0, T = n) + P\left(X_1 = 1, \underbrace{X_2 = 0}_{\substack{muss\ sein,\\ sonst\ T=2}}, T = n\right)$$
$$= P(X_1=0)P(T=n-1) + P(X_1=1)P(X_2=0)P(T=n-2)$$

(Idee hierbei: bei 0 fängt man im Grunde von Vorne an)

Kurz: $\boxed{P(T=n)=(1-p)\ P(T=n-1)+p\ (1-p)\ P(T=n-2)}$

$$\begin{aligned}
g_T(s) &= P(T=0)\,s^0+P(T=1)\,s^1+P(T=2)\,s^2+\sum_{n=3}^{\infty}P(T=n)\,s^n \\
&= p^2s^2+\sum_{n=3}^{\infty}\big((1-p)\,P(T=n-1)+p(1-p)\,P(T=n-2)\big)\,s^n \\
&= p^2s^2+(1-p)\,s\underbrace{\sum_{n03}^{\infty}P(T=n-1)\,s^{n-1}}_{\sum_{n=0}^{\infty}P(T=n)s^n}+p(1-p)\underbrace{\sum_{n=3}^{\infty}P(T=n-2)\,s^{n-2}}_{\sum_{n=0}^{\infty}P(T=n)s^n-P(T=0)s^0}
\end{aligned}$$

Damit haben wir eine Gleichung für die unbekannte Funktion g_T erhalten:

$$g_T(s)=p^2s^2+(1-p)\,s\ g_T(s)+p(1-p)\,s^2g_T(s)$$

und damit

$$g_T(s)=\frac{p^2s^2}{1-(1-p)\,s(1+ps)}$$

Und damit

$$ET=g_T^{'}(1)=\frac{1+p}{p^2}$$

Darüber hinaus: Sind a^{-1} und b^{-1} die Nullstellen des Nenners, so liefert eine Partialbruchzerlegung

$$\begin{aligned}
g_T(s) &= p^2s^2\frac{1}{(1-as)(1-bs)}=\frac{p^2s^2}{a-b}\left(\frac{a}{1-as}-\frac{b}{1-bs}\right) \\
&= \sum_{n=0}^{\infty}\frac{p^2}{a-b}\left(a^{n+1}-b^{n+1}\right)s^{n+2} \\
&\left(=\sum_{k=0}^{\infty}P(T=k)\,s^k\right)
\end{aligned}$$

Durch Koeffizientenvergleich erhält man

$$P(T=n)=\frac{p^2}{a-b}\left(a^{-1}-b^{n-1}\right)\qquad \text{für } n\geq 2$$

Konkret: Wann erhält man beim Würfelwurf das erste Mal zwei Sechsen hintereinander?

$p=\frac{1}{6}$ $\qquad ET=42$ $\qquad$ Nullstellen: $a=-3+\frac{9}{\sqrt{5}}$ $\quad b=-3-\frac{9}{\sqrt{5}}$

Einzelne Wahrscheinlichkeiten können beispielsweise mit Maple ausgerechnet werden.

Wiederholung:

Bei $P(X \in N_0) = 1$:

$$g_X(S) = \sum_{k=0}^{\infty} s^k P(X = k) \quad (= ES^X)$$

Vorteile:

- $EX = g_X^{'}(1)$

  $E(X(X-1)) = g_X^{''}(1)$ *usw.*
- Sind *X*, *Y* unabhängig, so gilt $g_{X+Y}(s) = g_X(s) + g_Y(s)$

Aus der schwierigen Operation der Faltung wird die (einfache)Multiplikation von Funktionen.

Was macht man nun bei allgemeinen ZV?
Die Momenterzeugende Funktion zu *X* (bzw. zur Verteilung von *X*)

$$\varphi_X(t) = Ee^{tX} = \begin{cases} \sum\limits_x e^{tX} \; P(X = x), & \text{bei diskretem } X \\ \int e^{tX} \; f_X(x)\, dx, & \text{wenn } X \text{ Dichte } f_X \text{ hat} \end{cases}$$

Gute Nachricht: Ist allgemein anwendbar.
Schlechte Nachricht: Bei manchen ZV erhält man

$$\varphi_X(t) = \begin{cases} 1, & t = 0 \\ \infty, & t \neq 0 \end{cases} \quad \text{(Man kommt dann nicht zurück)}$$

Gilt jedoch: $\exists\, \delta > 0 \; \forall t, |t| \leq \delta \; \varphi_X^{(t)} < \infty$
(In Worten: Die Momenterzeugende Funktion existiert in einer Nullumgebung)

Wir setzen dies ab jetzt voraus, man kann dann schnell mit diesen Transformationen arbeiten (kommt insbesondere wieder zurück).

Insbesondere

1. Ableitung: $\frac{d}{dt}\varphi_X(t) = \frac{d}{dt}\left(Ee^{tX}\right) \overset{!}{=} E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$

   Also $\varphi_X^{'}(0) = EX$, allgemeiner: $\varphi_X^{(k)}(0) = EX^k$

2. Ableitung: $\frac{d^2}{dt^2}\varphi_X(t) = \frac{d^2}{dt^2}\left(Ee^{tX}\right) \overset{!}{=} E\left(\frac{d^2}{dt^2}e^{tX}\right) = E\left(X^2 e^{tX}\right)$

(daher momenterzeugende Funktion)

Außerdem: Sind *X*, *Y* unabhängig, so gilt:

$$\varphi_{X+Y}(t) = Ee^{t(X+Y)} = E\left(e^{tX}e^{tY}\right) = \left(Ee^{tX}\right)\cdot\left(Ee^{tY}\right) = \varphi_X(t)\cdot\varphi_Y(t)$$

Wie bei wahrscheinlichkeitserzeugenden Funktionen: Faltung wird zur Multiplikation. Bei allgemeinen (reellwertigen) ZV wird das Verhalten unter affinen Transformationen $X \mapsto y : ax+b$ interessant.
(bei ganzzahligen erhält man unter Umständen eine ZV die nicht mehr ganzzahlig ist).

Wie wirkt sich dies auf momenterzeugende Funktionen aus?

Sei $Y = aX + b$. Dann gilt

$$\varphi_Y(t) = Ee^{tY} = Ee^{t(aX+b)} = E\left(e^{tb} \cdot e^{(at)X}\right) = e^{tb}\varphi_X(at)$$

Beispiele:

a) $X \sim Exp(\lambda)$ (Exponentialverteilt)

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & sonst \end{cases}$$

$$\varphi_X(t) = \int e^{tX} f_X(x)\, dx = \int_0^\infty e^{tX} \lambda e^{-\lambda x}\, dx$$

$$= \lambda \int_0^\infty e^{(t-\lambda)x}\, dx$$

$$= \left[\frac{\lambda}{t-\lambda} e^{x(t-\lambda)}\right]_0^\infty \qquad \boxed{e^\infty = \infty \cdot e^{-\infty} = 0}$$

$$= \begin{cases} \dfrac{\lambda}{\lambda - t}, & \text{für } t < \lambda \\ \infty, & \text{sonst} \end{cases}$$

Also: φ_X existiert nur auf $(-\infty, \lambda)$. Dieses Intervall enthält eine Nullumgebung; man kann also hier sinnvoll mit momenterzeugenden Funktionen arbeiten.

$$\lambda(\lambda - t)^{-1} \text{ abgeleitetet } \lambda(\lambda - t)^{-2}$$

Allgemein:

$$\frac{k\lambda}{(\lambda - t)^{-(k+1)}}$$

b) $X \sim N(0,1)$

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int e^{t\lambda x} e^{-x^2/2}\, dx = e^{x^2/2}$$

Wie sieht die momenterzeugende Funktion bei allgemeinen Normalverteilungen aus?
Bekannt: $X \sim N(0,1) \Rightarrow Y = \mu + \sigma X \sim N(\mu, \sigma^2)$

Obige Rechenregel liefert:

$$\varphi_Y(t) = e^{t\mu}\varphi_X(\sigma t) = e^{t\mu}e^{(\sigma t)^2/2}$$
$$= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

Damit: Sind X, Y unabhängig mit $X \sim (\mu_X, \sigma_X^2)$, $Y \sim (\mu_Y, \sigma_Y^2)$, so hat man

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = \exp\left(\mu_X t + \frac{1}{2}\sigma_X^2 t^2\right)\exp\left(\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2\right)$$
$$= \exp\left((\mu_X \mu_Y)t + \frac{1}{2}\left(\sigma_X^2 + \sigma_Y^2\right)t^2\right)$$

Dies ist die momenterzeugende Funktion zu $N\left(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2\right)$, d.h. man kommt zurück.

$$X + Y \sim N\left(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2\right)$$

§8 Grenzwertsätze

Zur Vorbereitung behandeln wir eine berühmte Ungleichung:
Sei *X* eine ZV mit Erwartungswert μ und Varianz $\sigma^2 < \infty$.

Definiere eine neue ZV *Y* durch

$$Y := \begin{cases} \varepsilon \,, |x-\mu| \geq \varepsilon \\ 0 \,, sonst \end{cases}$$

Klar: Dann gilt

$$y^2 \leq (x-\mu)^2 ,$$
$$EY^2 \leq E(x-\mu)^2$$

Wir betrachten die beiden Seiten

$$EY^2 = \varepsilon^2 P(Y = \varepsilon) + \underbrace{\cancel{\sigma^2 P(Y=0)}}_{=0}$$
$$= \varepsilon^2 P(|x-\mu| \geq \varepsilon)$$

$$E(x-\mu)^2 = \sigma^2$$

Satz (Ungleichung von Chebyshev)

$$\boxed{P(|x-\mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \operatorname{var}(X)}$$

Diese Ungleichung ist Grundlage der **DreiSigmaRegel**.
Die Wahrscheinlichkeit, dass eine ZV mehr als drei Standardabweichungen von ihrem Mittelwert entfernt ist, ist höchstens etwa 10%.

Mit $\varepsilon = 3\sigma$ erhält man

$$P(|x-\mu| \geq 3\sigma) \leq \frac{1}{3\sigma^2} \sigma^2 = \frac{1}{9}$$

Hiermit lässt sich folgender, wichtiger Grenzwertsatz beweisen:

Satz (Das schwache Gesetz der großen Zahlen)

Es sei X_1, X_2, X_3 eine Folge von unabhängigen und identisch verteilten ZV mit Eigenwert μ und Varianz $\sigma^2 < \infty$.
Dann gilt für den Mittelwert

$$\overline{X}_n := \frac{1}{n} \sum_{k=1}^{n} X_k$$

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

(In groben Worten: Bei großem Stichprobenumfang ist der Mittelwert um den Erwartungswert herum konzentriert)

Spezialfall X_k zeigt an, ob in der *k*-ten Wiederholung eines Zufallsexperimentes ein bestimmtes Ereignis *A*, welches die Wahrscheinlichkeit *p* hat, eintritt $(X_k = 1)$ oder nicht $(X_k = 0)$.

$$EX_k = p$$

$\overline{X_n}$: relative Häufigkeit von *A* in den ersten *n* Versuchen. Der Satz zeigt, dass bei großen *n* mit großer Wahrscheinlichkeit die relative Häufigkeit in der Nähe der Wahrscheinlichkeit liegt.

$X_1, X_2, \ldots$ sind unabhängige ZV, alle mit derselben Verteilung;

$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ ist der <u>Mittelwert</u> der ersten *n* Variablen,

$s_n := \sum_{i=1}^{n} X_i \ \left(= n\overline{X_n}\right)$ ist die Partialsumme der ersten *n* Variablen.

Wir setzen generell voraus, dass zu den X_i das zweite Moment existiert $\left(EX_i^2 < \infty\right)$, und schreiben $\mu := EX_i$ für den Erwartungswert sowie $\sigma^2 := \operatorname{var}(X_i)$ für die Varianz der Variablen. Es sei auch $\sigma^2 > 0$ (Erinnerung: die Chebyshev-Ungleichung $P\left(|X - EX| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\operatorname{var}(X)$ impliziert $P(X = EX) = 1$ bei $\sigma^2 = 0$, *X* ist also bei $\sigma^2 = 0$ „degeneriert“).

Das schwache Gesetz der großen Zahlen besagt in dieser Situation $P\left(\left|\overline{X_n} - \mu\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$ der Mittelwert der Daten ist also bei großem Stichprobenumfang immer stärker um den Erwartungswert herum konzentriert.

Ist X_i die Augenzahl beim *i*-ten Wurf eines fairen Würfels, so ist also die mittlere Augenzahl $\overline{X_n}$ nach *n* Würfen mit wachsendem n immer stärker um 3,5 herum konzentriert.

Frage: Lässt sich etwas über die Differenz $\overline{X_n} - \mu$ aussagen, insbesondere über deren Größenordnung?

<u>Satz (Der Zentrale Grenzwertsatz)</u>

Für alle $x \in \mathbb{R}$ gilt

$$\boxed{P\left(\frac{\sqrt{n}}{\sigma}\left(\overline{X_n} - \mu\right) \leq x\right) \xrightarrow[n\to\infty]{} \Phi(x) \quad \left(= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy\right)}$$

In Worten: Die standardisierte Abweichung vom Erwartungswert ist asymptotisch Standardnormalverteilt (Φ ist die Verteilungsfunktion zu $N(0,1)$). Warum „standardisiert“?

Die Variable $Z_n := \frac{\sqrt{n}}{\sigma}\left(\overline{X_n} - \mu\right)$ hat Erwartungswert 0

$$EZ_n = \frac{\sqrt{n}}{\sigma}\left(E\overline{X_n} - \mu\right) = 0$$

und Varianz 1

$$\operatorname{var}(Z_n) = \frac{1}{\sigma^2}\left(\overline{X_n}\right) = \frac{n}{\sigma^2}\frac{1}{n^2}\overbrace{\operatorname{var}\left(\sum_{i=1}^n X_i\right)}^{\substack{\sum_{i=1}^n \operatorname{var}(X_i),\\ Bienaym\acute{e}}} = 1$$

Anstelle eines Beweises eine Argumentation mit momenterzeugenden Funktionen:
Wir zeigen, dass die momenterzeugende Funktion zu Z_n mit $n \to \infty$ gegen die momenterzeugende Funktion zu $N(0,1)$, also $\varphi(t) = e^{t^2/2}$ mit $n \to \infty$ konvergiert.

$$\varphi_n(t) = Ee^{tZ_n} = E\exp\left(\frac{t\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)\right)$$

$$= \exp\left(\frac{t\sqrt{n}}{\sigma}\mu\right)E\exp\left(\overbrace{\left(\frac{t}{\sigma\sqrt{n}}\right)}^{\tilde{t}}\sum_{i=1}^n X_i\right)$$

$$= \exp\left(\frac{t\sqrt{n}}{\sigma}\mu\right)\varphi_{\sum_{i=1}^n X_i}\left(\left(\frac{t}{\sigma\sqrt{n}}\right)\right)$$

$$= \exp\left(\frac{t\sqrt{n}}{\sigma}\mu\right)\prod_{i=1}^n \varphi_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

(Die „meF“ zur Summe ist das Produkt der einzelnen meF’s, da $X_1,...,X_n$ unabhängig sind.)

$$= \exp\left(\frac{t\sqrt{n}}{\sigma}\mu\right)\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n$$

$$\varphi_{X_1}(t) = \varphi_{X_1}(0) + \varphi_{X_1}'(0)t + \frac{1}{2}\varphi_{X_1}''(0)t^2 + \text{Rest}$$

Im vorliegenden Fall

$$\varphi_{X_1}(t) = 1 + \mu t$$

$$\varphi_{X_1}'(o) = EX_1$$

$$\varphi_{X_1}''(o) = EX_1^2 = \sigma^2 + \mu^2$$

Oben eingesetzt:

$$\varphi_n(t) = \exp\left(-\frac{t\sqrt{n}}{\sigma}\mu\right)\left(1+\mu\frac{t}{\sigma\sqrt{n}}+\frac{\sigma^2+\mu^2}{2}\frac{t^2}{\sigma^2 n}+\text{Rest}\right)^n$$

$$= \exp\left(-\frac{t\sqrt{n}}{\sigma}\mu + n\cdot\log\left(1+\mu\frac{t}{\sigma\sqrt{n}}+\frac{\sigma^2+\mu^2}{2}\frac{t^2}{\sigma^2 n}+\text{Rest}\right)\right)$$

Um weiter zu kommen, benötigen wir das Verhalten von $\ln(1+x)$ für kleine x.

Taylor-Entwicklung von $n \to \ln(1+x)$ um $n \to \infty$

$$\ln(1+x) = 0 + \underbrace{\frac{d}{dx}\ln(1+x)|_{x=0}}_{\frac{1}{1+x}} \cdot x + \frac{1}{2}\underbrace{\frac{d^2}{dx^2}\ln(1+x)|_{x=0}}_{\frac{-1}{(1+x)^2}} \cdot x^2 + \text{Rest}$$

$$= x - \frac{1}{2}x^2 + \text{Rest}$$

Damit:

$$\varphi_n(t) = \exp\left(-\frac{t\sqrt{n}}{\sigma}\mu + n\left(\mu\frac{t}{\sigma\sqrt{n}}+\frac{\sigma^2+\mu^2}{2}\frac{t^2}{\sigma^2 n}-\frac{1}{2}\left(\mu\frac{t}{\sigma\sqrt{n}}+\frac{\sigma^2+\mu^2}{2}\frac{t^2}{\sigma^2 n}\right)^2\right)+\text{Rest}\right)$$

Wie sieht der teil aus der μ enthält:

$$\underbrace{-\frac{t\sqrt{n}}{\sigma}+\frac{t\sqrt{n}}{\sigma}}_{=0} - \frac{1}{2}\cdot 2\cdot \cancel{n}\frac{\mu t}{\sigma\sqrt{n}}\frac{\sigma^2}{2}\frac{t^2}{\sigma^2\cancel{n}} \to 0$$

Wenn wir tatsächlich einfachheitshalber $\mu = 0$ annehmen, so erhält man

$$\varphi_n(t) = \exp\left(\cancel{n}\frac{\cancel{\sigma^2}}{2}\frac{t^2}{\cancel{\sigma^2}\cancel{n}}+\text{Rest}\right) \to e^{t^2/2}$$

q.e.d.

Beispiel:

Mit welcher Wahrscheinlichkeit erscheint bei 600maligem Wurf eines fairen Würfels mindesten 95 und höchstens 105-mal eine 6?

$X_1,...,X_{600}$ sind $Bin\left(1,\frac{1}{6}\right)$, $S_{600} = \sum_{i=1}^{600} X_1$ ist die Anzahl der Sechsen nach 600 Durchgängen.

In dieser konkreten Situation kennt man die Verteilung von S_{600}, kann also die gewünschte Wahrscheinlichkeit „richtig" ausrechnen:

S_{600} ist binomialverteilt mit Parametern $n=600$ und $p=1/6$:

$$P(95\le S_{600}\le 105)=\sum_{k=95}^{105}\binom{600}{k}\left(\frac{1}{6}\right)^k\left(1-\frac{1}{6}\right)^{600-k}$$

$$=0,453101 \quad \text{(Computer-Programm)}$$

Normalapproximation:

$$P(95\le S_{600}\le 105)=P\left(\frac{95}{600}\le \overline{X}_{600}\le\frac{105}{600}\right)$$

$$=P\left(-0,5477\le\left(\frac{\overline{X}_{600}-\frac{100}{600}}{\sqrt{\frac{500}{6}}}\right)\cdot\sqrt{600}\le 0,5477\right)$$

(Nebenrechnung:

$$\frac{\frac{105}{600}-\frac{100}{600}}{\sqrt{\frac{500}{6}}}\sqrt{600}=\frac{5\sqrt{6}\sqrt{6}}{600\underbrace{\sqrt{500}}_{\sqrt{5}\cdot 10}}\cdot 10=\frac{\sqrt{5}}{100} \quad \text{(falsch. Wir nachgereicht))}$$

Mit der Normalapproximation erhält man $\approx\Phi(0,5477)-\Phi(-0,5477)=0,4176...$

<u>Der zentrale Grenzwertsatz (ZGWS)</u>

Sind $X_1,X_2,...$ unabhängige Zufallsvariablen mit derselben Verteilung und $0<r^2\ \left(=\operatorname{var}(X_1)\right)<\infty$, so gilt mit

$$\mu=EX_1\ ,S_n=\sum_{i=1}^{n}X_i \quad (n\text{-te Partialsumme})$$

$$\overline{X}_n=\frac{1}{n}S_n \quad (\text{Mittelwert der ersten } n)$$

$$\lim_{n\to\infty}P\left(\frac{\sqrt{n}\left(\overline{X}n-\mu\right)}{\sigma}\le x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-y^2/2}\,dy \quad \left(=\Phi(x)\right)$$

Klar:

$$P\left(\frac{\sqrt{n}\left(\overline{X}n-\mu\right)}{\sigma}\le x\right)=P\left(\underbrace{\frac{S_n-n\mu}{\sqrt{n}\sigma}}_{\substack{\text{standartisierte Partialsumme } S_n^* \\ ES_n^*=0,\ \operatorname{var}(S_n^*)=1}}\le x\right)$$

Φ ist die Verteilungsfunktion zu $N(0,1)$, der Standartnormalverteilung.

In Worten: „Standardisierte Partialsummen von unabhängigen, identisch verteilten Zufallsvariablen mit positiver, endlicher Varianz sind asymptotisch normal" ZGWS erklärt die Allgegenwart der Gaußschen Glockenkurve.
Standardisieren heißt: Erwartungswert subtrahieren, dann durch Standartabweichung teilen.

Ein wichtiger Spezialfall:
Ein Zufallsexperiment wird (unabhängig, und ohne Veränderung der Randbedingungen) wiederholt ausgeführt. Das Ereignis A in diesem Experiment hab die Wahrscheinlichkeit p $0 < p < 1$. Wie oft tritt A bei n Wiederholungen ein?

Wir setzen:

$$X_i = \begin{cases} 1, & A \text{ tritt in der } i\text{-ten Wdhlg ein} \\ 0, & \text{sonst} \end{cases}$$

Dann sind $X_1, X_2, \ldots$ unabhängig mit

$$P(X_i = 1) = p \qquad P(X_i = 0) = 1 - p \qquad \text{für alle } i \in \mathbb{N},$$

(insbesondere haben alle dieselbe Verteilung) und es gilt

$$EX_1 = \sum_x xP(X_1 = x) = 1 \cdot p + 0 \cdot (1-p) = p \quad (= \mu)$$

$$EX_1^2 = \sum_x x^2 P(X_1 = x) = 1^2 \cdot p + 0 \cdot (1-p) = p$$

$$\sigma^2 = \operatorname{var}(X_1) = (EX_1^2) - (EX_1)^2 = p - p^2 = p(1-p)$$

Klar:
$S_n = X_1 + \ldots + X_n$ ist die interessierende Anzahl der Experimente (unter den ersten n), bei denen das Ereignis A eingetreten ist. ZGWS liefert

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x \right) = \Phi(x) \qquad \forall x \in \mathbb{R},$$

insbesondere

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b \right) = \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \, dx$$

Dieser Spezialfall des ZGWS ist der Satz von de Moivre-Laplace

Beispiel:
Wie viele Sechsen erhält man, wenn man einen (fairen) Würfel 600mal wirft?
„Im Prinzip" ist jeder Wert $0, 1, \ldots, 599, 600$ möglich.

Stattdessen:
Wie groß ist die Wahrscheinlichkeit, dass mindestens 95-mal und höchstens 105-mal eine 6 erscheint?

Wir wenden de Moivre-Laplace an:
Das wiederholt ausgeführte zufallsexperiment ist der Würfelwurf, es geht um A: Eine Sechs erscheint. Man hat also $n = 600,\ p = \frac{1}{6}$.

Entscheidend ist nun, dass man die Schranken für S_n auf die standardisierte Version umrechnet:

$$95 \le S_n \le 105 \quad \Leftrightarrow \quad -5 \le S_n - \overbrace{np}^{100} \le 5$$
$$\Leftrightarrow -\sqrt{\frac{3}{10}} \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \sqrt{\frac{3}{10}}$$

Betrachtet man $n = 600$ als groß, so erhält man

$$\begin{aligned} P(95 \le S_n \le 105) &= P\left(-\sqrt{0,3} \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \sqrt{0,3} \right) \\ &\approx \Phi\left(\sqrt{0,3}\right) - \Phi\left(-\sqrt{0,3}\right) \\ &= 2\Phi\left(\sqrt{0,3}\right) - 1 \\ &= \Phi(0,5477) = 0,7089 \\ &= 0,4176 \end{aligned}$$

$$\boxed{\Phi(x) = 1 - \Phi(-x)}$$

Mit $S_{600} \sim Bin\left(600, \frac{1}{6}\right)$ ergibt sich der exakte Wert $P(95 \le S_{600} \le 105) = 0,453$.

<u>Stetigkeitskorrektur:</u> $P(95 \le S_{600} \le 105) = P(94,9 \le S_{600} \le 105,5)$
$$\approx 2 \cdot \Phi(0,6023) - 1 = 0,4532$$