# Elementare Wahrscheinlichkeitstheorie und Statistik A

Dozent: Prof. Dr. R. Grübel

unkorrigierte Vorlesungsmitschrift
aus dem WS 2004 / 05
angefertigt von Florian Garbs

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

1. Vorlesung (14.10.2004)
§1 Ein mathematisches Modell für Zufallsexperimente

Stochastik: Wahrscheinlichkeitstheorie und Statistik.
Gegenstand: Zufallsexperimente (d.h. Ergebnis nicht durch Randbedingungen festgelegt).

Stichproben oder
Ergebnisraum: Eine Menge, die die möglichen Versuchsergebnisse enthält, $\Omega$.
Ereignisse: werden durch Teilmengen von $\Omega$ beschrieben.

Beispiel 1.1

a) Würfelwurf $\Omega = \{1, ..., 6\}$.
   Ereignis „gerade Zahl erscheint" wird beschrieben durch $A = \{2, 4, 6\}$.

b) Eine Probe radioaktives Material emittiert in einem bestimmten Zeitraum eine zufällige Anzahl von Partikeln. $\Omega = N_0 = \{0, 1, 2, ...\}$, $A = \{n \in N_0 : n \geq 10\}$ steht für „mindestens 10 Zerfälle".

c) Eine Faser der Länge 1 reißt unter Zug an einer zufälligen Stelle. Ergebnisraum für die Position des Bruchpunktes $\Omega = \{0, 1\}$. Das Ereignis, dass die Bruchstelle näher am Mittelpunkt der Faser, als am Endpunkt liegt entspricht der Teilmenge $\left(\frac{1}{4}, \frac{3}{4}\right)$.

Ereignisse mit nur einem Element, also von der Form $A = \{\omega\}$ mit einem $\omega \in \Omega$, nennt man Elementarereignisse.

Ereignisse lassen sich zu neuen Ereignissen kombinieren („Ereignisalgebra"). Auf der „Mengenseite" entspricht dies Mengenoperationen.
Das Ereignis „das durch A repräsentierte Ereignis tritt nicht ein" wird repräsentiert durch $A^C \left(= \{\omega \in \Omega : \omega \notin A\}\right)$.
„es kommt keine gerade Zahl": $\{2, 4, 6\}^C = \{1, 3, 5\}$.

$A \cap B$: A und B treten beide ein
$A \cup B$: A oder B (oder Beide) treten ein

Beispiel 1.2
In Bsp 1.1 a) ist $A = \{2, 4, 6\}$ das Ereignis „eine gerade Zahl kommt" und $B = \{1, 2, 3\}$ das Ereignis „eine Zahl $\leq 3$ kommt".
$A \cap B = \{2\}$ (ein Elementarereignis) steht für „eine gerade Zahl $\leq 3$ kommt".

Wie wird das Ereignis „A, B und C treten ein" dargestellt? $A \cap B \cap C$.
„genau eines": $A \cap B^C \cap C^C \cup A^C \cap B \cap C^C \cup A^C \cap B^C \cap C$ beachte: wir haben eine „disjunkte Vereinigung". Hierfür schreiben wir gelegentlich „+" statt „$\cup$".

Was ist nun „Wahrscheinlichkeit"?
Eigentlich außermathematische Frage, wie z.B. „was ist Raum?" .

Zwei nicht-identische, aber auch nicht-disjunkte Interpretationen:

- relative Häufigkeit, „frequentisch"
- Grad der Gemischtheit, „subjektivistisch" (insbesondere bei nicht wiederholbaren Experimenten sinnvoll)

Sei $N_n(A)$ die Anzahl der Experimente bei n-maliger Wiederholung, die denen A eingetreten ist. Würfelwurf, A gerade Zahl:

3 3 2 1 2 4 6 6 3 1

n=10, $N_n(A)=5$ <u>absolute Häufigkeit</u>

relative Häufigkeit $\frac{1}{n}N_n(A)$ (hier 0,5).

Rechenregeln für die relative Häufigkeit:

(1) $0 \le \frac{1}{n}N_n(A) \le 1,\ \frac{1}{n}N_n(\Omega)=1$

(2) $\frac{1}{n}N_n(A_1+\ldots+A_k)=\frac{1}{n}N_n(A_1)+\ldots+\frac{1}{n}N_n(A_k)$
für paarweise disjunkte Ereignisse $A_1,\ldots,A_k$

Jargon: „$A \to \frac{1}{n}N_n(A)$ ist Additiv". (n = fest)

<u>Definition 1.3 ( die Kolmegorov-Axiome)</u>

Gegeben: Ergebnisraum $\Omega$, System $\mathfrak{A}$ von Ereignissen (für uns normalerweise die Menge <u>aller</u> Teilmengen von $\Omega$, die Potenzmenge $\mathfrak{P}(\Omega)$.

Eine Wahrscheinlichkeit (oder besser ein Wahrscheinlichkeitsmaß) ist eine Abbildung P: $\mathfrak{A} \to [0,1]$ (dem Ereignis A wird die Wahrscheinlichkeit $P(A)$ zugeordnet) mit der Eigenschaft

(A1) $P(A) \ge 0$ , $P(\Omega)=1$

(A2) Für alle $A_1, A_2, \ldots \in \mathfrak{A}$ mit $A_i \cap A_j \neq \varnothing$ für $i \neq j$ (eine Folge von paarweise disjunkten Ereignissen) gilt:

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

Das Tripel ($\Omega$, $\mathfrak{A}$, P) nennt man Wahrscheinlichkeitsraum.
Erste Folgerung:

Satz 1.4

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

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

(III) $P(A^C)=1-P(A)$

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

(V) $P(A_1+...+A_k)=P(A_1)+...+P(A_k)$ (endliche Additivität)

Beweis:
(I) Verwendet man die $\sigma$-Additivität (A“) mit $A_j=\varnothing$ für alle $j\in\mathbb{N}$, so erhält man

$$P(\varnothing)=\sum_{j=1}^{\infty}P(\varnothing),$$

woraus wegen $P(\varnothing)\in\mathbb{R}$ die gewünschte Aussage $P(\varnothing)=0$ folgt.
(V) Verwendet man (A“) mit $A_j=\varnothing$ für $j>k$, so folgt wegen (I) die Behauptung.
(III) Wir benutzen die erste Hälfte von (A1) sowie die endliche Additivität (V) mit $A_1=A$, $A_2=A^C$ und erhalten

$$1=P(\Omega)=P(A+A^C)=P(A)+P(A^C),$$

also (III).
(IV) Mit dem zweiten Teil von (A1) und der endlichen Additivität folgt

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

also die gewünschte Monotonieeigenschaft.
(II) Die folgt aus der Monotonieeigenschaft (IV) und der ersten Hälfte von (A1).

q.e.d.

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)

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

$$P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1\cap A_2)$$

$$P(A_1\cup A_2\cup A_3)=P(A_1)+P(A_2)+P(A_3)-P(A_1\cap A_2)-P(A_1\cup A_3)$$
$$-P(A_2\cap A_3)+P(A_1\cap A_2\cap A_3)$$

2. Vorlesung (21.10.2004)

$\Omega$: Ergebnisraum, enthält die möglichen Ergebnisse der Zufallsexperimente
$\mathfrak{A}$: System der Ereignisse $A < \Omega$
P: $\mathfrak{A} \to [0,1]$: Abbildung, die jedem Ereignis A seine Wahrscheinlichkeit P(A) zuordnet

## §2 Laplace-Experiment, elementare Kombinatorik

Laplace-Experiment heißt: $\#\Omega < \infty$ (endlicher Ergebnisraum), $\mathfrak{A} = \mathfrak{P}(\Omega)$ (alle Teilmengen von $\Omega$ kommen als Ergebnis in Frage),
$P(A) = \frac{\#A}{\#\Omega}$ (= Anzahl der günstigen / Anzahl der Möglichkeiten insgesamt).

Beispiel 2.1
Ein Würfel wird 2-mal geworfen. Mit welcher Wahrscheinlichkeit erhält man die Augensumme 7 (bzw. 6)?
Wir nehmen an, das ein Laplace-Experiment über

$$\begin{aligned}\Omega &= \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} \\ &= \{(i,j) : i,j \in \{1,...,6\}\} \\ &= \{(1,1),...,(6,6)\}\end{aligned}$$

vorliegt.

$$A_7 = \{(6,1),(5,2),(4,3),(3,4),(2,5),(1,6)\}$$
$$A_6 = \{(5,1),(4,2),(3,3),(2,4),(1,5)\}$$

$$P(A_7) = \frac{\#A_7}{\#\Omega} = \frac{6}{36}, \quad P(A_6) = \frac{5}{36}$$

Berechnung von Wahrscheinlichkeiten läuft bei Laplace-Experimenten auf die Kunst des Zählens (Kombinatorik) hinaus.

Regel 1:
Lässt sich A auf B bijektiv abbilden, so gilt #A = #B.

Regel 2:
$\#(A \cup B) = \#A + \#B$, wenn A und B disjunkt sind.

Verallgemeinerung auf mehr als zwei: $\#(A_1 + ... + A_k) = \#A_1 + ... + \#A_k$

Nützlicher Spezialfall: Hat $C \subset A \times B$ die Eigenschaft $\#\{y \in B : (x,y) \in C\} = n$ für alle $x \in A$, so gilt $\#C = n\#A$.

Spezialfall vom Spezialfall: $\#(A \times B) = (\#A)(\#B)$
Wir schreiben abkürzend:

$$\begin{aligned}A^k &= A \times A \times ... \times A \\ &= \{(x_1,...,x_k) : x_i \in A\} \text{ für alle } i = 1 - k\end{aligned}$$

Klar: $\#\left(A^k\right) = \left(\#A\right)^k$.

Sei jetzt $M_n = \{1,...,n\}$. Dann nennen wir die Elemente von $M_n^k$, also die L-Tupel $\left(i_1,...,i_k\right)$ mit $i, j \in \{1,...,n\}$ für $j = 1,...,k$ $\boxed{\# M_n^k = n^k}$, <u>K-Permutationen von $M_n$ mit Wiederholung.</u>

<u>Interpretation:</u>

(I) Einer Menge von n Elementen kann man $n^k$ verschiedene Stichproben vom Umfang k entnehmen, wenn man zurücklegt und die Reihenfolge beachtet wird.

(II) Es gibt $n^k$ -Möglichkeiten, k verschiedene Objekte auf n Plätze zu verteilen, wieder bei Berücksichtigung der Reihenfolge und möglicher Mehrfachbesetzung.

<u>Zur Erinnerung:</u> $n! = n \cdot (n-1) \cdot ... \cdot 3 \cdot 2 \cdot 1$ (n-Fakultät)

$$\binom{n}{m} = \frac{n!}{(n-m)! - m!}$$ (n über k , Binominalkoeffizienten)

<u>Satz 2.2</u>

Für $1 \le k \le n$ gilt $\#\left\{\left(i_1,...,i_k\right) \in M_n^K : \begin{matrix} i, j \neq i_L \\ j \neq L \end{matrix}\right\} \frac{n!}{(n-k)!}$

<u>Verwende Regel 2:</u>

Es gibt n Möglichkeiten für $i_1$, bei gegebenem $i_1$ gibt es $n-1$ Möglichkeiten für $i_2$, bei gegebenen $i_1$ und $i_2$ gibt es $n-2$ Möglichkeiten für $i_3$, …, bei gegebenen $i_1,...,i_{k-1}$ gibt es $n-k+1$ Möglichkeiten für $i_k$.

Also $\#... = n(n-k+1) = \frac{n!}{(n-k)!}$.

<u>Interpretation:</u>

(I) Einer Menge von n Elementen kann man $\frac{n!}{(n-k)!}$ verschiedene Stichproben entnehmen, wenn die Reihenfolge beachtet wird, und <u>nicht</u> zurückgelegt wird.

(II) Es gibt $\frac{n!}{(n-k)!}$ Möglichkeiten, k Objekte unter Berücksichtigung der Reihenfolge auf n Plätze ohne Mehrfachbelegung zu verteilen.

Wichtiger Spezialfall: Es gibt $n!$ Permutationen einer Menge von n Elementen (mit $k = n$).

<u>Satz 2.3</u>

Für $1 \le k \le n$ gilt:

$$\#\left\{\left(i_1,...,i_k\right) \in M_n^k : i_1 < i_2 < ... < i_k\right\} = \binom{n}{k}$$

Zum Beweis:
Jedem Element dieser Menge lassen sich k! Elemente der Menge aus dem letzten Satz zuordnen (alle die, die man durch eine der k! möglichen Permutationen erhält).

Also (unbekannte Anzahl): $k! = \frac{n!}{(n-k)!}$.

Die Elemente dieser Menge werden K-Kombinationen von $M_n$ ohne Wiederholung genannt.

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

Interpretation:

(I) Es gibt $\binom{n}{k}$ Möglichkeiten, aus n Objekten k herauszugreifen (Stichprobe ohne zurücklegen und ohne Berücksichtigung der Reihenfolge).

(II) Es gibt $\binom{n}{k}$ Möglichkeiten, k Objekte ohne Mehrfachbesetzung auf n Plätze zu verteilen, wenn die Verteilungsreihenfolge nicht berücksichtigt wird.

Typische Anwendung: „6 aus 49“ , $\binom{49}{6}$ Möglichkeiten.

Beispiel 2.4
Eine (faire) Münze wird 10-mal geworfen. Mit welcher Wahrscheinlichkeit erscheint genau 7-mal Kopf? Wir gehen aus von einem Laplace-Experiment über $\Omega = \{0,1\}^{10}$, wobei 0 für Zahl und 1 für Kopf steht.
Beispielsweise bedeutet $\omega = (0,0,0,1,1,0,1,1,1,0)$, dass 3-mal Zahl, dann 2-mal Kopf usw. erscheint.
#A = Anzahl der Teilmenge vom Umfang 7 einer Menge von 10 Elementen.
(betrachte die Wurfmenge, bei denen eine 1 erscheint)

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\binom{10}{7}}{2^{10}} = \frac{120}{1024} = \frac{15}{128}.$$

Satz 2.5

$$\#\{(i_1,\dots,i_k) \in M_n^k : i_1 < i_2 < \dots < i_k\} = \binom{n+k-1}{k}$$

Zum Beweis betrachten wir die folgende Abbildung $M_n^k$ in $M_{n+k-1}^k$:

$$(i_1,\dots,i_k) \to (i_1, i_2+1, i_3+2, \dots, i_k+k-1).$$

Dies liefert eine bijektive Abbildung zwischen

$\{(i_1,\dots,i_k) \in M_n^k : i_1 \le i_2 \le \dots \le i_k\}$ und

$\{(i_1,\dots,i_k) \in M_{n+k-1}^k : i_1 < i_2 < \dots < i_k\}$

Nach Satz 2.4 hat die letztgenannte Menge $\binom{n+k-1}{k}$ Elemente.

Die Elemente der Menge aus Satz 2.5 werden K-Kombinationen von $M_n$ mit Wiederholung genannt.

Interpretation:

(I) Einer Menge von n Elementen kann man $\binom{n+k-1}{k}$ verschiedene Stichproben vom Umfang k entnehmen, wenn zurückgelegt wird und die Reihenfolge nicht beachtet wird.

(II) Es gibt $\binom{n+k-1}{k}$ Möglichkeiten k Objekte auf n Plätze zu verteilen, bei möglicher Mehrfachbesetzung, wenn die Verteilungsreihenfolge nicht berücksichtigt wird.

| | mit Wiederholung | ohne Wiederholung |
|---|---|---|
| Permutation<br>(Reihenfolge zählt) | $n^k$ | $\frac{n!}{(n-k)!}$ |
| Kombinationen<br>(Reihenfolge egal) | $\binom{n+k-1}{k}$ | $\binom{n}{k}$ |

## 3. Vorlesung (28.10.2004)

## Drei klassische Beispiele für Laplace-Experimente

### Beispiel 2.6 (Das Geburtstagsproblem)

In einem Raum befinden sich n Personen. Wie groß ist die Wahrscheinlichkeit, dass mindestens 2 Personen am gleichen Tag Geburtstag haben? $(n \le 365)$

Vereinfachende Annahmen:

- 29. Februar streichen
- keine Zwillings, etc.
- keine saisonalen Schwankungen

Dann ist ein Laplace-Experiment über $\Omega := \{(i_1, ..., n) : 1 \le ij \le 365, j = (1, ..., n)\}$ gerechtfertigt.

$ij = k$ bedeutet hier, dass Person j am k-ten Tag des Jahres Geburtstag hat.

$$A = \{(i_1, ..., i_n) \in \Omega : \exists l \neq j \text{ mit } i_l = ij\}$$

$$P(A) = \frac{\# A}{\# \Omega} \qquad \# \Omega = 365^n \qquad \# A = ?$$

Trick:

Betrachte $A^C = \{(i_1, ..., i_n) \in \Omega : ij \neq i_l \text{ für } j \neq l\}$

$$\# A^C = 365 * 364 * ... * (365 - n + 1) \quad (\text{beachte} (n \le 365))$$
$$= \frac{365!}{(365 - n)!}$$

$$P(A) = 1 - P(A^C) = 1 - \frac{365!}{(365 - n)! * 365!}$$
$$= 1 - \frac{365}{365} * \frac{364}{365} * \frac{363}{365} * ... * \frac{(365 - n + 1)}{365}$$

Offensichtlich monoton in n

Bemerkenswert: Ab $n = 23$ ist $P(A) > 0{,}5$. Bei $n = 50$: $P(A) \approx 0{,}97$.

Beispiel 2.7 Der zerstreute Postbote)

Ein Postbote verteilt n Briefe „zufällig“ auf n Briefkästen (einen pro Kasten).
Zu jeder der n Adressen gehört genau ein Brief. Wie viele Personen erhalten den für sie bestimmten Brief?
Wir nummerieren die Briefe und Briefkästen von 1 bis n, und zwar so, dass Brief i in Kasten i gehört.
Die möglichen Austeilungen entsprechen den Permutationen von $\{1,...,n\}$. Wir nehmen also ein Laplace-Experiment über $\Omega=\{(\omega_1,...,\omega_n):\omega_i\in\{1,...,n\},\omega_i\neq\omega_j \text{ für } i\neq j\}$ an.

A: # ist Null, d.h. **kein** Adressat erhält seinen Brief

$$A=\{(\omega_1,...,\omega_n)\in\Omega:\omega_i\neq i \text{ für } i=1,...,n\}$$

Wieder: $P(A)=\frac{\#A}{\#\Omega}$ $\qquad \#\Omega=n!$ $\qquad \#A=?$

$$A^C=\{(\omega_1,...,\omega_n)\in\Omega:\omega_i=i \text{ für ein } i\in\{1,...,n\}\}$$
$$=\bigcup_{i=1}^{n}A_i \text{ mit } A_i=\{\omega\in\Omega:\omega_i=i\}$$

Siebformel

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{k=1}^{n}(-1)^{k-1}\sum_{\substack{H\subset\{1,...,n\}\\ \#H=k}}P\left(\bigcap_{i\in H}A_i\right)$$
$$\left(P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1\cap A_2)\right)$$

$$\omega\in\bigcap_{i\in H}A_i\Leftrightarrow\omega_i=i \text{ für alle } i\in H.$$

Damit ist $\#\bigcap_{i\in H}A_i=(n-\#H)!$

k-Positionen liegen fest

Damit ist:
$$P(A)=1-\frac{1}{n!}\sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}(n-k)!$$
$$=1+\sum_{k=1}^{n}(-1)^k\frac{1}{k!}$$
$$=\sum_{k=0}^{n}\frac{(-1)^k}{k!}\xrightarrow{n\to\infty}\frac{1}{e}\approx 0,3679...$$

Beispiel 2.8 (Skat)
Typisch: Mit welcher Wahrscheinlichkeit liegen zwei Buben im Skat?
Laplace-Experiment.

$\Omega = \left\{ (\omega_1, \ldots, \omega_{32}) : \omega_i \in \{1, \ldots, 32\}, \text{ paarweise verschieden} \right\}$ (also $\#\Omega = 32!$)

Interpretation: Wir nummerieren die Karten mit den Zahlen 1 bis 32.
1 bis 4 seien die Buben. Die Austeilreihenfolge sei beispielsweise so, dass die 10. und die 11. Karte in den Skat kommen.
Ist A das Ereignis, dass zwei Buben im Skat landen,
so führt dies auf $A = \left\{ \omega \in \Omega : \omega_{10} \in \{1,2,3,4\}, \omega_{11} \in \{1,2,3,4\} \right\}$.

$$\#A = \underbrace{4}_{\text{Möglichkeiten für } \omega_{10}} * \underbrace{3}_{\text{Möglichkeiten für } \omega_{11} \text{ bei gg.}} * \underbrace{30 * 29 * 28 * \ldots * 3 * 2 * 1}_{\text{Möglichkeiten für } \omega_1 \text{ bei gg. } \omega_{10} \text{ und } \omega_{11}}$$

$$P(A) = \frac{\#A}{\#\Omega} = \frac{4*3*30!}{32!} = \frac{4*3}{32*31} = \frac{3}{248} \approx 0{,}012$$

In etwa einem Prozent der Spiele sind zwei Buben im Skat.

Es ist intuitiv klar, dass dieses Ergebnis von der Austeilreihenfolge unabhängig ist.
Fasst man alle $\omega$'s zusammen, die zu derselben Kartenverteilung führen, so erhält man als Ergebnisraum:

$$\tilde{\Omega} = \left\{ (H_1, H_2, H_3, S) : \begin{array}{r} H_1, H_2, H_3, S = \{1, \ldots, 32\} \\ \#H_1 = \#H_2 = \#H_3 = 10, \#S = 2 \\ H_1, H_2, H_3, S \text{ paarweise disjunkt} \end{array} \right\}$$

$\omega = (H_1, H_2, H_3, S)$ wird wie folgt interpretiert:
Spieler i erhält die Karten mit Nummern aus $H_i$, $i = 1,2,3$; $S = \text{Skat}$.

Wie sieht das Ereignis A (2 Buben im Skat) in diesem mathematischen Modell (also als Teilmenge von $\tilde{\Omega}$) aus?

$$A = \{(H_1, H_2, H_3, S) \in \tilde{\Omega} : S = \{1,2,3,4\}\}$$

$$\#\tilde{\Omega} = \binom{32}{10} \quad \binom{22}{10} \quad \binom{12}{10} \quad \binom{2}{2}$$

Anzahl d. Mglk. für Hand 1

Mglk. für Hand 2 bei gg. Hand 1

Mglk. Für Hand 3 bei gg. Hand 1&2

Skat bei gg. Hand 1&2&3

$$\#A = \binom{4}{2}\binom{30}{10}\binom{20}{10}\binom{10}{10}$$

$$P(A) = \frac{4!*30!*20!*10!*22!*10!*12!*10!*2!}{2!*2!*10!*20!*10!*10!*32!*22!*12!} = \frac{4!*30!}{2!*32!} = \frac{4*3}{32*31}$$

(passt zum anderen Ergebnis)

# Achtung!

Woher wissen wir eigentlich, dass ein Laplace-Experiment über $\tilde{\Omega}$ vorliegt?

Dies geht **nur**, wenn beim Übergang von $\Omega$ zu $\tilde{\Omega}$ jeweils gleichviele $\omega$'s zu einem $\tilde{\omega}$ zusammengefasst werden. (Dies ist hier der Fall: 10! 10! 10! 2! $\omega$'s werden zu einem $\tilde{\omega}$ zusammengefasst.)

Weiteres Beispiel (zeigt, dass man mit etwas Übung recht schnell mit $\tilde{\Omega}$ arbeiten kann):

D sei das Ereignis, dass Spieler 1 genau einen Buben erhält.

$$\#D = \binom{4}{1}\binom{28}{9}\binom{22}{10}\binom{12}{10}\binom{2}{2}$$

Damit ist $P(D) = ... = \frac{4*10*22*21*20}{32*31*30*29} = \frac{385}{899} \approx 0{,}4283$

$P(D \cap A) = ?$ Spieler 1 erhält einen Buben und zwei liegen im Skat.

$$\#(D \cap A) = \underbrace{\binom{4}{1}}_{\text{Der Bube für Spieler 1}} \underbrace{\binom{28}{9}}_{\text{Die „nicht-Buben“ für Spieler 1}} \underbrace{\binom{3}{2}}_{\text{Buben im Skat}} \binom{20}{10}\binom{10}{10}$$

$$P(D \cap A) = \frac{\#(D \cap A)}{\#\tilde{\Omega}} \ldots = \frac{5}{899} = 0,0056$$

4. Vorlesung (04.11.2004)

## §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 $N_n(A)-\text{mal}$ ein (absolute Häufigkeit). Unter dieser tritt auch B $N_n(A\cap B)-\text{mal}$ ein. Damit ergibt sich als relative Häufigkeit des Eintretens von B, unter denen die A liefern $\frac{N_n(A\cap B)}{N_n(A)}=\frac{\frac{1}{n}N_n(A\cap B)}{\frac{1}{n}N_n(A)}$ .
Der Zusammenhang relative Häufigkeit – Wahrscheinlichkeit motiviert nun:

Definition 3.1
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 3.2
In der im Beispiel 2.8 betrachteten Situation (Skat) sei A das Ereignis, dass sich 2 Buben im Skat befinden, D das Ereignis, dass Spieler 1 genau einen Buben erhält.

$$P(A\,|\,D)=\frac{P(A\cap D)}{P(A)}=\frac{5/899}{385/899}=\frac{5}{385}\approx 0{,}013$$

Dies ist etwas größer, als $P(A)\left(=3/248\approx 0{,}012\right)$.

Satz 3.3

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

(II) (Gesetz von der totalen Wahrscheinlichkeit)
Ist $A_1,\dots,A_n$ eine Ereignispartition von $\Omega$, d.h.
$A_1\cup\dots\cup A_n=\Omega,\ A_i\cap A_j=\varnothing$ für $i\neq j$, so gilt für alle Ereignisse B
$P(B)=\sum_{i=1}^{n}P(B\,|\,A_i)\,P(A_i)$. (Hierbei ist $P(A_i)=0$ zugelassen, der Summand wird dann als 0 interpretiert)

(III) (Formel von Bayes)
Es seien $A_1,\dots,A_n$ und B wie in (II) und $P(B)>0$. Dann gilt:

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

Beweis:

(I) Die rechte Seite wird zu

$$P(A_1)*\frac{P(A_2\cap A_1)}{P(A_1)}*\frac{P(A_3\cap A_1\cap A_2)}{P(A_1\cap A_2)}*\ldots*\frac{P(A_n\cap A_1\cap\ldots\cap A_{n-1})}{P(A_1\cap\ldots\cap A_{n-1})}$$

$$=P(A_1\cap A_2\cap\ldots\cap A_n)$$

(II)

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

$$=P\big(B\cap(A_1\cup\ldots\cup A_n)\big)=P(B\cap A_1\cup\ldots\cup B\cap A_n)$$

Disjunkte Mengen

$$=P(B\cap A_1)+\ldots+P(B\cap A_n)$$

$$=P(B\,|\,A_1)\,P(A_1)+\ldots+P(B\,|\,A_n)\,P(A_n)$$

(III)

Teil (II) liefert für den Nenner $P(B)$, also wird die rechte Seite zu

$$\frac{P(A_1)\,P(B\,|\,A_i)}{P(B)}=\frac{P(A_1)\dfrac{P(B\,|\,A_i)}{P(A_i)}}{P(B)}=P(A_i\,|\,B).$$

q.e.d

Beispiel 3.4

Ein bestimmter medizinischer Test ist zu 95% effektiv beim erkennen einer bestimmten Krankheit, liefert allerdings bei 1% der gesunden Personen „falschen Alarm“. Angenommen, 0,5% der Bevölkerung leiden unter dieser Krankheit. Mit welcher Wahrscheinlichkeit hat dann jemand diese Krankheit, wenn der Test dies behauptet?
Es sei A „die getestete Person hat die Krankheit“, B „der Test ist „positiv“ (zeigt das Vorliegend der Krankheit)“.

$P(A)=0{,}005$ $\quad$ $P(B\,|\,A)=0{,}95$ $\quad$ $P(B\,|\,A^C)=0{,}01$ $\quad$ $P(A\,|\,B)=?$

Die Bayes-Formel liefert:

$$P(A\,|\,B)=\frac{P(B\,|\,A)P(A)}{P(B\,|\,A)P(A)+P(B\,|\,A^C)P(A^C)}$$

$$=\frac{0{,}95*0{,}005}{0{,}95*0{,}005+0{,}01*(1-0{,}005)}$$

$$\approx 0{,}323$$

Einer der zentralen Begriffe der Stochastik ist der der (stochastischen) Unabhängigkeit.
Grob: B ist von A unabhängig, wenn die Information „A ist eingetreten“ die Wahrscheinlichkeit für B nicht ändert.

Definition 3.5
Zwei Ereignisse A und B heißen (stochastisch) unabhängig, wenn $P(A \cap B) = P(A) * P(B)$ gilt.

Eine Familie $\{A_i : i \in I\}$ im Ereignisraum heißt unabhängig, wenn

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i) \text{ (für alle endlichen Teilmengen J und I).}$$

Bei $I = \{1,2,3\}$:

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

Der Grund für die „schwierig aussehende“ Definition bei allgemeinen Ereignisfunktionen ist die Tatsache, dass aus der paarweisen Unabhängigkeit
$P(A_i \cap A_j) = P(A_i) * P(A_j)$, für alle $i, j \in I$ mit $i \neq j$ nicht die volle Unabhängigkeit folgt.

Beispiel 3.6
Der zweimalige Wurf einer fairen Münze führt auf ein Laplace-Experiment über $\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$, wobei wir 0 für Kopf und 1 für Wappen schreiben.

Wir betrachten:

$A_1 = \{(0,0),(0,1)\}$ „Kopf im 1. Wurf“

$A_2 = \{(0,0),(1,0)\}$ „Kopf im 2. Wurf“

$A_3 = \{(0,1),(1,0)\}$ „Ergebnisse verschieden“

$$P(A_1 \cap A_2) = P(\{0,0\}) = \frac{1}{4} = \frac{1}{2} * \frac{1}{2} = P(A_1) * P(A_2)$$
$$P(A_1 \cap A_3) = P(\{0,1\}) = P(A_1) * P(A_3)$$
$$P(A_2 \cap A_3) = P(\{1,0\}) = P(A_2) * P(A_3)$$

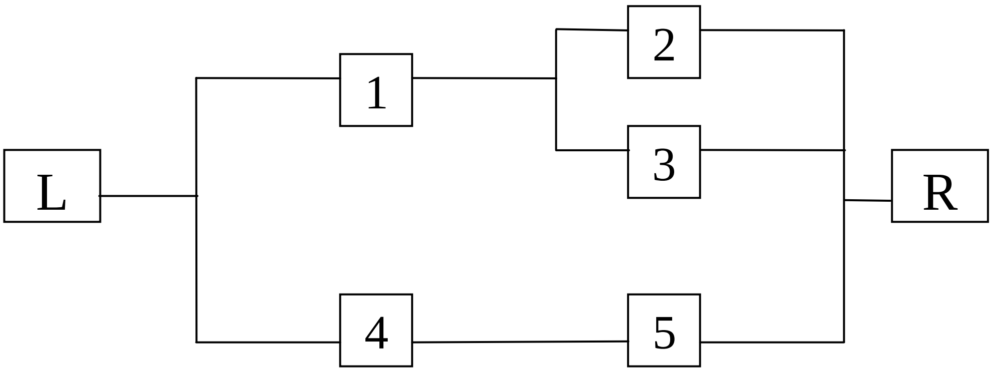
Das Ereignissystem $\{A_1, A_2, A_3\}$ ist also paarweise unabhängig, **aber:**

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

Beispiel 3.7

Eine typische Fragestellung der angewandten Wahrscheinlichkeitsrechnung beschäftigt sich mit dem Funktionieren von Innovationen.

Wir betrachten einen einfachen Fall:
Ein System bestehe aus 5 wie folgt angeordneten Komponenten:




Wir nehmen an, dass die Komponenten unabhängig voneinander und zwar jeweils mit der Wahrscheinlichkeit p funktionieren. Das Gesamtsystem funktioniert, wenn es einen Pfad funktionierender Komponenten von L nach R gibt.
Mit welcher Wahrscheinlichkeit funktioniert das Gesamtsystem?

Es sei $A_i$ das Ereignis, dass Komponente i funktioniert. B Das Ereignis, dass das Gesamtsystem funktioniert.

Dann gilt $B = B_1 \cup B_2$ mit

$B_1 = A_4 \cap A_5$ (der untere Pfad ist passierbar)

$B_2 = A_1 \cap (A_2 \cup A_3)$ (der obere Pfad ist passierbar)

$$P(B_1) = P(A_4 \cap A_5) = P(A_4) * P(A_5) = p^2 \quad \text{(Unabhängigkeit!)}$$

Paarweise Unabhängigkeit würde nicht reichen

$$\begin{aligned} P(B_2) &= P(A_1 \cap A_2 \cup A_1 \cap A_3) \\ &= P(A_1 \cap A_2) + P(A_1 \cap A_3) - P(A_1 \cap A_2 \cap A_3) \\ &= p^2 + p^2 - p^3 \end{aligned}$$

$$P(B) = P(B_1) + P(B_2) - P(B_1 \cap B_2)$$

$$\begin{aligned} P(B_1 \cap B_2) &= P(A_4 \cap A_5 \cap A_1 \cap A_2) + P(A_4 \cap A_5 \cap A_1 \cap A_3) - P(A_4 \cap A_5 \cap A_1 \cap A_2 \cap A_3) \\ &= 2p^4 - p^5 \end{aligned}$$

Insgesamt:

$$P(B) = p^2 + 2p^2 - p^3 - 2p^4 + p^5 = p^2 \left(3 - p - 2p^2 + p^3\right)$$

5. Vorlesung (11.11.2004)

## §4 Diskrete Zufallsgrößen

### 4.1 Allgemeines

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

Beispiel 4.1.1

Eine faire Münze wird 5-mal geworfen. Wie oft erschein „Kopf“?
Schreibt man „1“ für kopf und „0“ für Wappen, so erhält man ein Laplace-Experiment über $\Omega = \{(\omega_1, ..., \omega_5) : \omega_i \in [0,1]\} \quad \left(= \{0,1\}^5\right)$.

Mit dieser „Kodierung“ ist die Anzahl $X(\omega)$ der Kopfwürfe bei einem Ergebnis (Wurffolge) $\omega$ gegeben durch $X(\omega) = \omega_1 + ... + \omega_5 = \sum_{k=1}^{5} \omega_k$. $\quad (X : \Omega \to \mathbb{R})$

Die möglichen Werte sind 0,1,…,5.
Mit welcher Wahrscheinlichkeit treten diese Werte auf? $\quad P(X=0) = P(\{(0,0,0,0,0)\}) = 1/32$

Bei beliebigem Paar aus dem Wertebereich gibt es $\binom{5}{k}$ Möglichkeiten, die k Kopfwürfe auf die 5 möglichen Positionen zu verteilen, also $P(X=k) = \binom{5}{k} \cdot 2^5$ für $k = 0,1,...,5$.

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

Definition 4.1.2

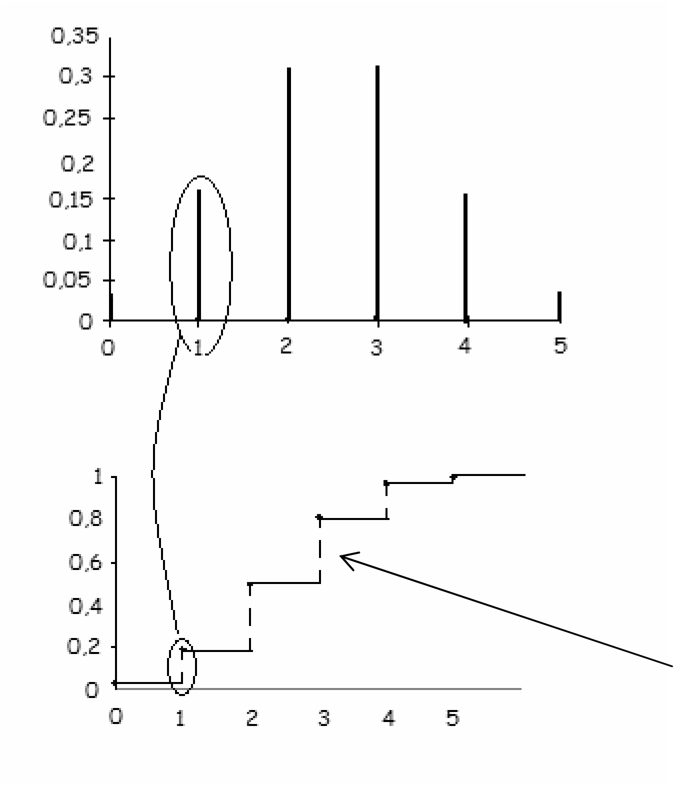
(I) Eine diskrete Zufallsvariable (ZV) ist eine Abbildung $X(\omega) \to \mathbb{R}$ mit endlich vielen oder abzählbar unendlich vielen Werten (z.B. $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$)

(II) Die Verteilung einer diskreten Zufallsvariablen ist die Abbildung $A \to P(X \in A) \quad \left(= P(\{\omega \in \Omega : X(\omega) \in A\})\right)$. Diese ist ein Wahrscheinlichkeitsmaß (s. Def. 1.3).

(III) Die Wahrscheinlichkeitsmassenfunktion (WMF) $p_x$ einer diskreten Zufallsvariablen X wird definiert durch $p_x : \mathbb{R} \to [0,1]$, $p_x(x) = P(X = x)$.

(IV) Die Verteilungsfunktion $F_x$ einer diskreten Zufallsvariablen wird definiert durch $F_x : \mathbb{R} \to [0,1]$, $F_x(x) = P(X \le x)$.

Massenfunktionen können beispielsweise durch ein Stabdiagramm illustriert werden. In der Situation des Beispiels erhält man:




Stabdiagramm für die WMF

Allgemein: $p_x(x) \geq 0$

$$\sum_x p_x(x) = 1$$

Verteilungsfunktion $F_x$

(isoton, schwach monoton wachsend)
Stetig von rechts, bei diskreter ZV von reinem Sprungtyp.

$F_x(0) = P(X \leq 0)$

Sprunghöhe in x ist

$P(X = x) = p_x(x)$

Bemerkung 4.1.3

(I) Zu jeder Zufallsvariablen gehört eine Verteilung, aber verschiedene ZV können durchaus dieselbe Verteilung haben (!). Beim 5-maligen Münzwurf erhält man beispielsweise auch für die Anzahl Y der Wappenwürfe die obige WMF (insbesondere haben X und Y dieselbe Verteilung), aber natürlich sind X und Y nicht gleich (hier gilt sogar $P(X = Y) = 0$).

(II) Da durch die Abbildung X in der Regel unterschiedlich viele Argumente $\omega$ zu einem Bildwert $X(\omega)$ zusammengefasst werden, sind in der Regel selbst bei einem Laplace-Experiment die Werte $P(X = x)$ nicht gleich groß.

## 4.2 Einige wichtige diskrete Verteilungen

### 4.2.1 (Binominalverteilung)

X heißt binominal verteilt mit Parametern n und p. Kurz: $X \sim \mathrm{Bin}(n, p)$, wenn gilt:

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

Die ZV X aus Beispiel 4.1.1 ist $\mathrm{Bin}\left(5, \tfrac{1}{2}\right)$-verteilt. Ganz allgemein erhält man $\mathrm{Bin}(n,p)$ als Verteilung der Anzahl X der Folge bei n unabhängigen Versuchsfolgen mit Erfolgswahrscheinlichkeit p:
Bezeichnet $A_i$ das Ereignis: „Erfolg im i-ten Versuch", so hat jede Sequenz $A_1 \cap A_2 \cap A_3^C \cap \ldots \cap A_{n-1}^C \cap A_n^C$ in der k-mal $A_i$ und $(n-k)$-mal $A_i^C$ steht, die Wahrscheinlichkeit $p^k(1-p)^{n-k}$.
Schließlich gibt es $\binom{n}{k}$ Möglichkeiten, die k Erfolge auf die n Positionen zu verteilen.
Wirft man beispielsweise einen Würfel 100x, so ist die Anzahl der Sechser $\mathrm{Bin}\left(100, \tfrac{1}{6}\right)$.

4.2.2 (Hypergeometrische Verteilung)

Eine ZV heißt hypergeometrisch verteilt mit Parametern M, m, n $(M, m, n \in \mathbb{N})$, wenn gilt:

$$P(X=k) = \frac{\binom{m}{k}\binom{M-m}{n-k}}{\binom{M}{n}}, \quad k = 0, \ldots, n.$$

Typische Fragestellung: Wahrscheinlichkeit für 2 Richtige beim Zahlenlotto „6 aus 49".

$$\frac{\binom{6}{2}\binom{49-6}{6-2}}{\binom{49}{6}}$$

Anzahl der möglichen Ziehungen

Die Hypergeometrische Verteilung taucht bei Stichproben ohne Zurücklegen auf: Enthält eine Urne M Kugeln, von denen m weiß und M-m schwarz sind, und werden n Kugeln zufällig und ohne (!) Zurücklegen entnommen, so ist X die Anzahl der weißen Kugeln in der Stichprobe:

Möglichkeiten für weiß

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

Möglichkeiten für schwarz

Anzahl der Stichproben

4.2.3 (Geometrische Verteilung)

Angenommen, ein Würfel wird solange geworfen, bis eine 6 erschein; X sei die hierfür nötige Wurfzahl.

$$P(X=n)=\underbrace{\left(\frac{5}{6}\right)}_{\text{Wahrscheinlichkeit für „nicht 6“}}^{n-1}\frac{1}{6}\qquad (n\in\mathbb{N}\setminus 0).$$

Allgemein: X heiß geometrisch verteilt mit Parameter p, wenn gilt:

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

Die geometrische Verteilung taucht immer dann auf, wenn man ein Zufallsexperiment solange wiederholt, bis erstmalig ein bestimmtes Ereignis A erscheint.
Der Parameter p ist die Wahrscheinlichkeit von A (Erfolgswahrscheinlichkeit).

**Achtung!**

Wenn man die Anzahl der Misserfolge zählt, erhält man eine andere Version der geometrischen Verteilung: $(1-p)^n\,p,\ \ n\in\mathbb{N}_0$

4.2.4 (Poisson Verteilung)

Die ZV X heißt Poisson-verteilt mit Parameter $\lambda(>0)$, wenn gilt

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!},\quad k\in\mathbb{N}_0.$$

Satz 4.3 („Das Gesetz der seltenen Ereignisse“)

Ist $(p_n)_{n\in\mathbb{N}}\subset[0,1]$ eine Nullfolge mit $\lim\limits_{n\to\infty} n\,p_n=\lambda$, so gilt für alle $k\in\mathbb{N}_0$:

$$\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^k=e^{-\lambda}\frac{\lambda^k}{k!}$$

Wahrscheinlichkeit für k Erfolge bei n Wiederholungen bei Erfolgswahrscheinlichkeit p.

Beweis:

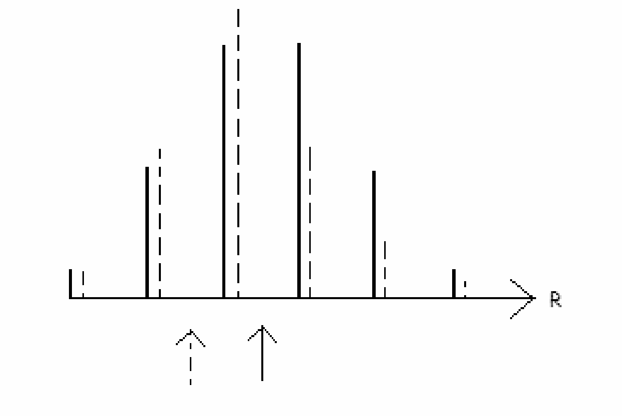
$$\binom{n}{k}p_n^k(1-p_n)^k=\underbrace{\frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{n^k}}_{=\frac{n}{n}\cdot\frac{n-1}{n}\cdot\ldots\cdot\frac{(n-k+1)}{n}=1}\cdot\underbrace{\frac{(n\,p_n)^k}{k!}}_{=\frac{\lambda^k}{k!}}\cdot\underbrace{\left(1-\frac{n\,p_n}{n}\right)^{n-k}}_{=\frac{\left(1-\frac{n\,p_n}{n}\right)^n=\left(1-\frac{\lambda}{n}\right)^n=e^{-\gamma}}{\left(1-\frac{n\,p_n}{n}\right)^k=1}}$$

Bekannt: $\lim\limits_{n\to\infty}\left(1+\frac{x}{n}\right)^n=e^x$

<u>6. Vorlesung (18.11.2004)</u>

## 4.4 Erwartungswerte und Momente

In diesem Abschnitt werden einige Verteilungsparameter eingeführt. Dies sind Zahlen, die bestimmte Aspekte einer Verteilung, wie beispielsweise die Lage oder die Streuung, beschreiben.



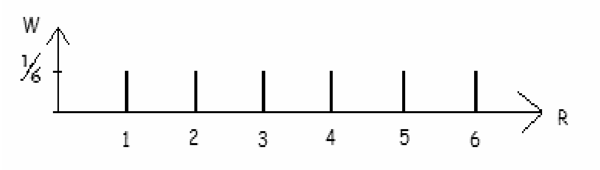

<u>Beispiel 4.5</u>
Sei X das Resultat beim Wurf eines fairen Würfels.
Der Mittelwert nach n Würfen ist

$$\frac{1}{n}\left(1\cdot\#(1)+2\cdot\#(2)+3\cdot\#(3)+...+6\cdot\#(6)\right),$$

wobei #(i) für die Anzahl der Würfe und Resultat i steht.

Für große n gilt $\frac{1}{n}\#(i)\approx P(X=i)$. Damit liegt der Mittelwert bei großem n in der Nähe von

$$1\cdot P(X=1)+2\cdot P(X=2)+...+6\cdot P(X=6)$$
$$=\frac{1}{6}(1+2+...+6)=\frac{21}{6}=3{,}5$$




<u>Definition 4.6</u>
Sei X eine diskrete Zufallsvariable mit WMF $p_x$. Dann wird der <u>Erwartungswert</u> $E(X)$ von X definiert durch

$$E(X)=\sum_{x\in\mathbb{R}} x\cdot P(X=x)\left(=\sum_{\substack{x\in\mathbb{R}\\ P(X=x)>0}} x\cdot P(X=x)\right)$$
$$=\sum_{x\in\mathbb{R}} x\cdot p_X(x)$$

Hierbei wird vorausgesetzt, dass die Reihe absolut konvergiert, d.h. $\sum|x|\,p_X(x)<\infty$.
Ist das nicht der Fall, so sagt man, dass der Erwartungswert von X nicht existiert.
Es bietet sich die folgende Analogie zu Mechanik an:
Platziert man die Massen $p_1, p_2, ...$ auf die Plätze $x_1, x_2, ...$, so ist $\sum x_i p_i$ die Lage des Schwerpunktes des Gesamtbildes.

<u>Beispiel 4.7</u>
$X\sim P(\lambda)$, $P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$, $k\in\mathbb{N}$. Gesucht EX.

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

Oft kann man die interessierende Zufallsvariable Y als Funktion g einer anderen ZV X schreiben: $\Omega \xrightarrow{x} \mathbb{R} \xrightarrow{g} \mathbb{R} \quad y := g(x)$.

<u>Satz 4.8</u>

Es gilt $Eg(x) = \sum_{x\in\mathbb{R}} g(x)\cdot p_X(x) = \sum_{x\in\mathbb{R}} g(x)\cdot P(X=x)$, vorausgesetzt, die Summen konvergieren absolut.

<u>Beweis:</u>

$$Eg(x) = \sum_{y\in\mathbb{R}} y\cdot P\big(g(x)=y\big) = \sum_{y\in\mathbb{R}} y\cdot P\big(x\in g^{-1}(y)\big) = \sum_{y\in\mathbb{R}} y \sum_{x\in g^{-1}(y)} P(X=x)$$
$$= \sum_{y\in\mathbb{R}} \sum_{x\in g^{-1}(y)} y\cdot P(X=x) = \sum_{x\in\mathbb{R}} g(x)\cdot P(X=x).$$

<u>Beispiel 4.9</u>

$X \sim P(\lambda)$. Gesucht ist $E\big(X(X-1)\big)$, verwende Satz 4.8 mit $g(x) = x(x-1)$.

$$EX(X-1) = \sum_{k=0}^{\infty} k\cdot(k-1)\cdot P(X=k) = \sum_{k=0}^{\infty} k\cdot(k-1)\cdot e^{-\lambda}\frac{\lambda^k}{k!}$$
$$= e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!} = e^{-\lambda}\cdot\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = e^{-\lambda}\cdot\lambda^2\cdot e^{\lambda} = \lambda^2$$

<u>Satz 4.10 (Rechenregeln)</u>

Sei X und Y diskrete ZV und $c\in\mathbb{R}$. Vorausgesetzt, dass die beteiligten Summen absolut konvergieren, gilt dann:

$E(X+Y) = E(X)+E(Y)$, $\qquad E(cX) = cE(X)$, $\qquad X\le Y \Rightarrow E(X)\le E(Y)$

(ohne Beweis)

Auch klar: $E(X_1+X_2+\ldots+X_n) = EX_1+EX_2+\ldots+EX_n$ (Folgt mit Induktion)

<u>Beispiel 4.11</u>

$X \sim Bin(n,p)$, gesucht: EX

$$P(X=k) = \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k},\; k=0,\ldots,n$$
$$EX = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n} k\cdot\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$
$$= n\cdot\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} = n\cdot\sum_{k=1}^{n}\binom{n-1}{k-1}p^k(1-p)^{n-k}$$
$$= n\cdot\sum_{k=0}^{n-1}\binom{n-1}{k}p^{k+1}(1-p)^{n-1-k} = n\cdot p\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-1-k}$$
$$EX = n\cdot p\text{, da } \sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-1-k} = 1$$

Die Summe über alle Werte der WMF zu $Bin(n-1,p)$

Alternativ unter Anwendung der Rechenregeln (Satz 4.10):
Es gilt $X = X_1 + ... + X_n$, wobei $X_i$ die Anzahl der Erfolge in der i-ten Versuchswiederholung ist.
$X_i \sim \mathrm{Bin}(1, p)$
Damit $EX_i = 0P(X_i = 0) + 1P(X_i = 1) = P(X_i = 1) = p$ und deshalb $EX = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} p = np$.

Allgemein
$X = 1_A \Rightarrow EX = 0P(\mathbb{1}_A = 0) + 1P(\mathbb{1}_A = 1) = P(\mathbb{1}_A = 1) = P(A)$
Die Einbettung $A \mapsto 1_A$ des Systems der Ereignisse in das System der ZV setzt P, was zu E führt.

Definition 4.12
Das k-te Moment von X ist $EX^k$ (vorausgesetzt $\sum_x |x|^k P(X = x) < \infty$)
Die Varianz von X ist $\mathrm{var}(X) = E(X - EX)^2$ (wieder unter der Voraussetzung, dass $EX^2 < \infty$, sonst sagt man, die Varianz existiert nicht)
Schließlich heißt $\sigma(X) = \sqrt{\mathrm{var}(X)}$ die Standartabweichung von X.
Die Varianz ist ein Maß für die Streuung einer ZV um ihren EW.

Rechenregel:

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

also $\boxed{\mathrm{var}(X) = E(X^2) - (EX)^2}$

Beispiel 4.13

$X \sim P(\lambda)$. Dann wissen wir $EX = \lambda$ (Bsp. 4.7), und $EX(X-1) = \lambda^2$ (Bsp. 4.9).
Frage $\mathrm{var}(X) = ?$

$$E(X^2) = E(X^2 - X) + EX = \lambda^2 + \lambda$$

$$(EX)^2 = \lambda^2$$

$$\Rightarrow \mathrm{var}(X) = E(X^2) - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Weitere Rechenregeln

Sei $c \in \mathbb{R}$, dann

$$\operatorname{var}(X+c) = E\big((X+c) - E(X+c)\big)^2 = E(X+c-EX-c)^2$$
$$= E(X-EX)^2 = \operatorname{var}(X)$$

$$\operatorname{var}(cX) = E(c^2X^2) - \big(E(cX)\big)^2 = c^2E(X^2) - c^2(EX)^2$$
$$= c^2\Big(E(X^2) - (EX)^2\Big) = c^2 \operatorname{var}(X)$$

Damit klar:

$\sigma(X+c) = \sigma(X)$, $\sigma(cX) = |c|\sigma(X)$

Tabelle mit den Erwartungswerten und Varianzen der im vorausgegangenen Unterabschnitt eingeführten Variablen:

| Verteilung | Parameter | EW | Var |
|---|---|---|---|
| Poisson | $\lambda$ | $\lambda$ | $\lambda$ |
| Binominal | n, p | $n \cdot p$ | $n \cdot p(1-p)$ |
| Geometrisch | p | $1/p$ | $\frac{1-p}{p^2}$ |
| Hypergeometrisch | M, m, n | $\frac{n \cdot m}{M}$ | $\frac{m \cdot n}{M} \cdot \frac{M-m}{M} \cdot \frac{M-n}{M-1}$ |

7. Vorlesung (25.11.2004)

## §5 Stetige Zufallsgrößen

### 5.1 Allgemeines

Viele Zufallsgrößen können ein Kontinuum von Werten annehmen: Lebensdauer, Messfehler, Proportionen, etc.

Definition 5.1.1

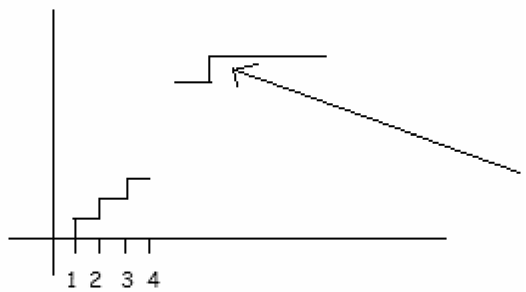
Eine ZV X heißt absolutstetig verteilt (kurz: stetig) mit Wahrscheinlichkeitsdichtefunktion $f_X$ (kurz: Dichte), wenn

$$\boxed{P(X \leq x) = \int_{-\infty}^{x} f_X(y)\,dy \qquad ,\forall x \in \mathbb{R}} \text{ gilt.}$$

Die Funktion $F_X : \mathbb{R} \to \mathbb{R},\ F_X = P(X \leq x)$, heißt die Verteilungsfunktion zu X.

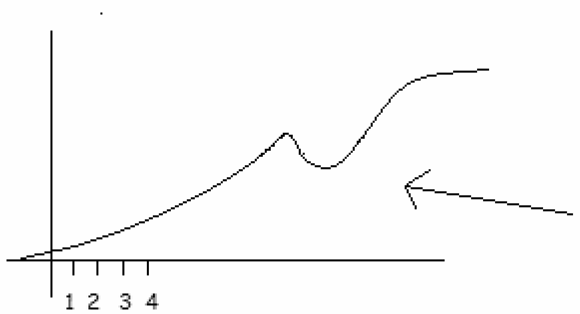
(Index X bei $f_X, F_X$ wird weggelassen, wenn X aus dem Zusammenhang hervorgeht.)

Im diskreten Fall ist $F_X$ von reinem Sprungtyp:




Sprung in x ist $P(X = x)$

Im stetigen Fall ist $F_X$ „glatt“: Grob ist $f_X$ die Ableitung von $F_X$, insbesondere also $F_X$ stetig.



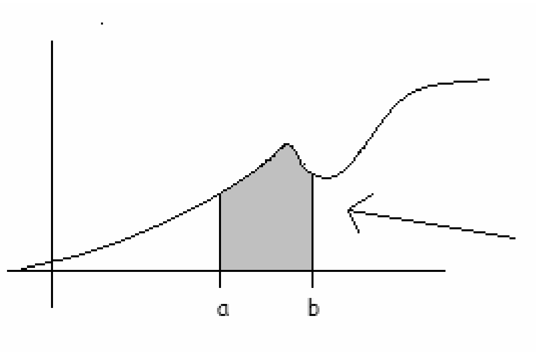

Keine Sprünge!

Insbesondere $P(X = x) = 0 \quad \forall x \in \mathbb{R}$ bei stetigen ZV.

Mit der Additivität $(-\infty < a < b < \infty)$

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= \int_{-\infty}^{b} f(y)\,dy - \int_{-\infty}^{a} f(y)\,dy = \int_{a}^{b} f(y)\,dy \\ &= P(a \leq X \leq b) = P(a < X < b) \end{aligned}$$

Die führt auf die folgende wichtige Interpretation von Wahrscheinlichkeitsdichte:




$\int_a^b f(y)\,dy = P(X \in [a,b])$

Die Wahrscheinlichkeit für $X \in A$ ist die Fläche unter f über A.

Eigenschaften von Dichten: $f \geq 0,\ \int_{-\infty}^{\infty} f(y)\,dy = 1 \quad (P(X \in \mathbb{R}))$

Analogie zur Massenfunktion: $p \geq 0,\ \sum_x p(x) = 1$

Im Gegensatz zur Massenfunktion können Dichten durchaus >1 sein.
Die Werte von f lassen sich als „Infinitesimale Wahrscheinlichkeit“ interpretieren als

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

Näherungsweise hat man also bei kleinen h: $P\left(x - \frac{h}{2} \leq X \leq x + \frac{h}{2}\right) \approx hf(x)$.

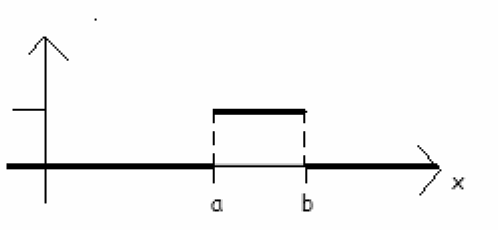
<u>5.2 Einige wichtige stetige Verteilungen</u>

<u>5.2.1 (Die Gleichverteilung)</u>

Eine ZV X heißt gleichverteilt auf dem Intervall $[a,b]$

(kurz: $\mathfrak{L}(X) = \text{unif}(a,b)$, $X \sim \text{unif}(a,b)$, wenn X die Dichtefunktion

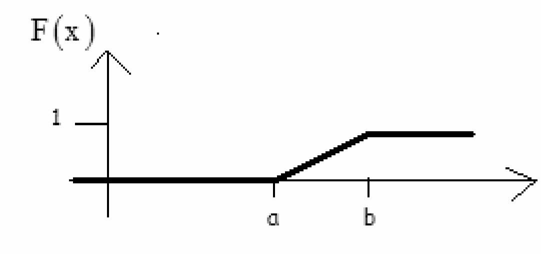
$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{sonst} \end{cases}$$




ist.

('<' oder '≤' ist egal, da die einzelnen Punkte die Wahrscheinlichkeit 0 haben)

Wie sieht die zugehörige Verteilungsfunktion aus? $F(x) = \int_{-\infty}^{x} f(y)\,dy$




unif $(a,b)$ kann als kontinuierliches Analogon zur Laplace-Verteilung (auf $(a,b)$) angesehen werden: alle infinitesimalen Wahrscheinlichkeiten haben denselben Wert.
unif $(0,1)$ spielt bei <u>stochastischen Simulationen</u> eine zentrale Rolle.

5.2.2 (Die Exponentialverteilung)

Die Exponentialverteilung mit Parameter $\lambda \ (>0)$ wird durch die Dichte

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

die sog. Verteilungsfunktion ist $F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

(taucht häufig im Zusammenhang mit Wartezeiten und Lebensdauer auf)

$P(X \geq x+y \mid X \geq y)$ Wahrscheinlichkeit dafür, dass das Bauteil noch weitere x Einheiten durchhält, wenn es das Alter >y hat.

$$= \frac{P(X \geq x+y \text{ und } X \geq y)}{P(X \geq y)} = \frac{1-F(x+y)}{1-F(y)} = \frac{1-\left(1-e^{-\lambda(x+y)}\right)}{1-\left(1-e^{-\lambda y}\right)} = e^{-\lambda x} = P(X \geq x)$$

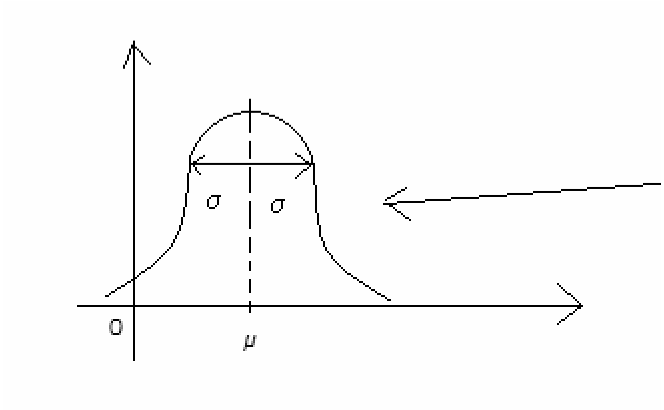
Kann gestrichen werden, da bereits in erster Bedingung enthalten.

Also: Das Alter spielt bei dieser Verteilung keine Rolle, „Gedächtnislosigkeit" der Exponentialverteilung. („Bauteil altert nicht")

5.2.3 (Die Normalverteilung)

Die wichtigste Verteilung ist die Normalverteilung mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, kurz: $N(\mu, \sigma^2)$, mit Dichtefunktion

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




Gaußsche Glockenkurve

Wendepunkte in $\mu \pm \sigma$

Taucht häufig bei Messfehlern auf (zentraler Grenzwertsatz)

Für die zugehörige Verteilungsfunktion gibt es keine einfache Formel. Für die Werte $\mu = 0$, $\sigma^2 = 1$ (man nennt $N(0,1)$ auch die <u>Standartverteilung</u>) sind die Werte der Verteilungsfunktion $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy$ verteilt.

Beispielsweise gibt $\Phi(1.645) \approx 0,95$

.

<u>Lemma 5.3</u>

Für alle $\beta \in \mathbb{R}, \alpha \neq 0$ gilt:

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

<u>Beweis:</u>

Sei $Y := \alpha X + \beta$.

Die zugehörige Verteilungsfunktion ist $F_Y(y) = P(Y \leq y) = P(\alpha X + \beta \leq y)$

$$= P\left(X \leq \frac{y-\beta}{\alpha}\right) = F_X\left(\frac{y-\beta}{\alpha}\right)$$

$$= \int_{-\infty}^{\frac{y-\beta}{\alpha}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right) dx$$

Substituiere $z = \alpha X + \beta$

Dichtefunktion zu $N(\alpha\mu + \beta, \alpha^2\sigma^2)$

$$= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\alpha^2\sigma^2}} \exp\left(\frac{-1}{2\alpha^2\sigma^2}\left(z - (\alpha\mu + \beta)\right)^2\right) dz$$

Die Verteilungsfunktion zu Y ist also die zu $N(\alpha\mu + \beta, \alpha^2\sigma^2)$ d.h. $Y \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$.

<u>Beispiel 5.4</u>

(I) Angenommen, es ist $P(1 < y < 1.5)$ für $Y \sim N\left(1, \frac{1}{4}\right)$ zu bestimmen.

Mit Lemma 5.3 folgt $2(Y-1) = 2Y - 2 \sim N(0,1)$,

also $P(1 < y < 1.5) = P(0 < 2(Y-1) < 1)$

$$= P(2(Y-1) \leq 1) - P(2(Y-1) \leq 0)$$

$$= \Phi(1) - \Phi(0) = 0.8413... - 0.5 = 0.3413...$$

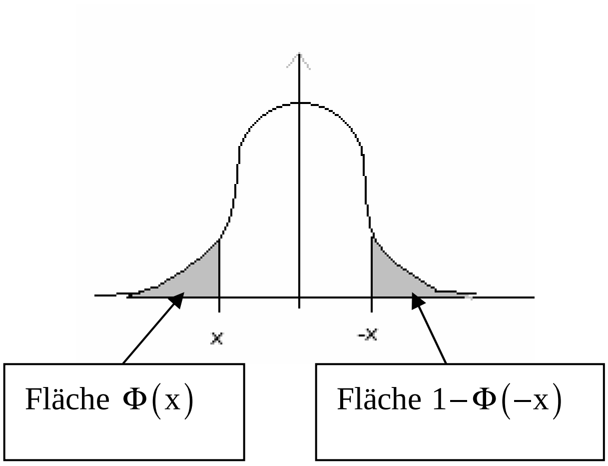
s. Tabelle

(II) Es sei Y wie in (I), nun soll ein Wert q mit der Eigenschaft $P(Y \leq q) : 0,95$ bestimmt werden.

Wegen $Y \leq y \Leftrightarrow 2(Y-1) \leq 2(q-1)$ und $P(2(Y-1) \leq 2(q-1)) = \Phi(2(q-1))$ und $\Phi(1.645) \approx 0,95$ führt dies auf $2(q-1) = 1.645$, also $q = 1.8225$.

Bei Standartparametern $\mu = 0, \tau^2 = 1$ ist die Glockenkurve symmetrisch um 0, dies ergibt

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




Insbesondere reicht eine Tafel für die Werte $x \geq 0$.

8. Vorlesung (02.12.2004)

absolutstetig verteilte Zufallsgröße X Dichte f heißt $P(X \in A) = \int\limits_A f(x)\,dx$

(Wahrscheinlichkeit als Fläche unter der Dichtefunktion)

5.3 Erwartungswert und Momente

Ist X eine stetige ZV mit Dichte f, so wird der Erwartungswert EX von X definiert durch

$$\boxed{EX = \int x\, f(x)\,dx},$$

vorausgesetzt, es gilt $\int |x|\, f(x)\,dx < \infty$ (sonst sagt man, dass der Erwartungswert nicht existiert).

Satz 5.5

$$\boxed{Eg(x) = \int g(x) f(x)\,dx}$$

Satz 5.6

Unter der Voraussetzung, dass die beteiligten Erwartungswerte existieren, gilt

$E(X+Y) = EX + EY, \quad E(cX) = cEX \qquad (c \in \mathbb{R},\ \text{linearität})$

$X \le Y \Rightarrow EX \le EY \quad (\text{Monotonie})$

Definition 5.7

Das k-te Moment einer (stetigen) ZV X ist $EX^k$, die Varianz $var(X) = E(X-EX)^2$, die Standartabweichung $\sigma(X) = \sqrt{var(X)}$.

Rechenregeln:

$$var(X) = EX^2 - (EX)^2$$
$$var(cX) = c^2\, var(X)$$
$$var(X+c) = var(X)$$

Beweis:

$$var(X) = E(X-EX)^2 = E\left(X^2 - 2XEX + (EX)^2\right)$$

(Linearität) $= EX^2 - 2EXEX + (EX)^2 = EX^2 - (EX)^2$

$$var(cX) = E(cX)^2 - (E(cX))^2 = E(c^2X^2)$$
$$= c^2E(X^2) - c^2(EX)^2 = c^2\left(EX^2 - (EX)^2\right) = c^2\, var(X)$$

<u>Beispiel 5.8</u>

(I) Ist X gleichverteilt auf dem Intervall $(0,1)$, so gilt:

$$f(x) = \begin{cases} 1 & ,0 < x < 1 \\ 0 & ,\text{sonst} \end{cases}$$

Also $EX = \int x\, f(x)\, dx = \int_0^1 x\, dx = \frac{1}{2}$.

$EX^2 = \int g(x) f(x)\, dx = \int_0^1 x^2\, dx = \frac{1}{3}$, also

$$\operatorname{var}(X) = EX^2 - (EX)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

$= Eg(x)$ mit $g(x) = \lambda^2$

(II) Im Falle $X \sim N(0,1)$ hat man $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, also

$EX = \int x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

$$\operatorname{var}(X) = EX^2 = \frac{1}{\sqrt{2\pi}} \int x^2 e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int x \left( x e^{-\frac{x^2}{2}} \right) dx$$

$$= \left[ -\frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} = \frac{1}{\sqrt{2\pi}} \int 1 \cdot \left( e^{-\frac{x^2}{2}} \right) dx = 0 + \frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2}{2}} dx = 1$$

Integral über die Dichtefkt. zu $N(0,1)$

Zur „Standartnormalverteilung" $N(0,1)$ gehören der Erwartungswert 0 und die Varianz 1.

Gilt allgemein $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, so folgt mit Lemma 5.2, dass $Y := \frac{X-\mu}{\sigma}$ die Verteilung $N(0,1)$ hat.

Also:

$$0 = EY = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{6}(EX - \mu)$$

damit $EX = \mu$

$$1 = \operatorname{var}(Y) = \operatorname{var}\left(\frac{X-\mu}{\sigma}\right)\frac{1}{\sigma^2}\operatorname{var}(X-\mu) = \frac{1}{\sigma^2}\operatorname{var}(X)$$

$$\boxed{\Rightarrow \operatorname{var}(X) = \sigma^2}$$

## $6 Mehrdimensionale Zufallsgrößen und Unabhängigkeit

Hat man mehrere von dem Ergebnis $\omega$ eines Zufallsexperiments abhängige Größen $X_1(\omega), X_2(\omega), \ldots, X_n(\omega) \in \mathbb{R}$, so kann man diese zu einem stochastischen Vektor X zusammenfassen: $X = (X_1, \ldots, X_n)$, $X : \Omega \to \mathbb{R}^n$, $\omega \to (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$.

Konkret: Einer von ihnen wird zufällig ausgewählt. Angenommen, man interessiert sich für die Matrikelnummer, die Körpergröße und das Gewicht:

$\omega$: ausgeählte Person

$X_1(\omega)$: Körpergröße

$X_2(\omega)$: Gewicht

$X_3(\omega)$: Matrikelnummer

Wir betrachten hauptsächlich $n = 2$, also Paare von ZV.

<u>Definition 6.1</u>

Die gemeinsame Verteilungsfunktion $F_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ zweier ZV X und Y (oder auch die Verteilungsfunktion $(X,Y)$) wird definiert durch:

$$F_{X,Y}(x,y) = P(X \le x \text{ und } Y \le y) = P(\{\omega \in \Omega : X(\omega) < x \cap Y(\omega) < y\})$$

Von zentraler Bedeutung sind stochastische Abhängigkeiten von ZV.
Was bedeutet „Unabhängigkeit" bei ZV?
(Bei Ereignissen: A, B unabhängig, wenn $P(A \cap B) = P(A)P(B)$)

Betrachte die Ereignisse $\{X \le x\}$ und $\{Y \le y\}$, $x, y \in \mathbb{R}$.

<u>Definition 6.2</u>

Es seinen X,Y ZV mit gemeinsamer Verteilungsfunktion $F_{X,Y}$. Weiter sei $F_X$ die Verteilungsfunktion zu X und $F_Y$ die Verteilungsfunktion zu Y. Dann heißen X und Y <u>stochastisch unabhängig</u>, wenn gilt:

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

Speziell bei diskreten bzw. stetigen ZV setzt man:

Definition 6.5

(I) Nehmen X und Y nur endlich viele oder höchstens abzählbar unendlich viele Werte an, so nennt man (X,Y) einen diskreten Zufallsvektor und $p_{X,Y}(x,y) := P(X=x, Y=y)$ die gemeinsame Massenfunktion von X und Y.

(II) Gibt es eine Funktion $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_+$ mit der Eigenschaft $F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)\,ds\,dt \quad \forall x,y \in \mathbb{R}$, so nennt man (X,Y) einen absolutstetig verteilten (kurz: stetig) Zufallsvektor und $f_{X,Y}$ eine gemeinsame Dichtefunktion von X und Y.

In Analogie zu den Sätzen 4.7 und 5.5 gilt nun:

Satz 6.4

Es seinen X,Y ZV und $g : \mathbb{R}^2 \to \mathbb{R}$ eine Funktion

(I) Ist (X,Y) ein diskreter Zufallsvektor mit Massenfunktion $p_{X,Y}$, so gilt $\boxed{Eg(X,Y) = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)}$.

(II) Ist (X,Y) ein stetiger Zufallsvektor mit Dichtefunktion $f_{X,Y}$, so gilt $\boxed{Eg(X,Y) = \int\int g(x,y) f_{X,Y}(x,y)\,dx\,dy}$.

Interpretation bei (II): Wahrscheinlichkeit als Volumina unter der durch die Dichte definierte Oberfläche.
Zusammenhang zur Unabhängigkeit:

Satz 6.5

(I) Ist (X,Y) diskret mit Massenfunktion $p_{X,Y}$, so sind X und Y genau dann unabhängig, wenn gilt: $p_{X,Y}(x,y) = p_X(x) p_Y(y)$ für alle $x, y \in \mathbb{R}$.
(Konkret: $P(X=x, Y=y) = P(X=x)P(Y=y)$)
Hierbei sind $p_X(x) = P(X=x) = \sum_y P(X=y, Y=y) = \sum_y p_{X,Y}(x,y)$ und $p_Y(y) = \sum_x p_{X,Y}(x,y)$ die Massenfunktion zu X bzw. Y.

(II) Sind X und Y stetig mit Dichtefunktion $f_X$ und $f_Y$, so sind sie genau dann unabhängig, wenn $f_{X,Y}$ mit $f_{X,Y}(x,y) := f_X(x) f_Y(y)$ eine Dichtefunktion zum Zufallsvektor (X,Y) ist.

9. Vorlesung (09.12.2004)

Satz 6.6 („Multiplikationsregel für Erwartungswerte“)
Für unabhängige ZV X und Y gilt:
$E(XY) = (EX)(EY)$ (Vorausgesetzt, die Erwartungswerte existieren)

Beweis für den stetigen Fall:

$$\begin{aligned} E(XY) &= \int\int xy\, f_{XY}(xy)\, dxdy \\ &= \int\int xy\, f_X(x) f_Y(y)\, dxdy \\ &= \int y\left[\int x\, f_X(x)\, dx\right] f_Y(y)\, dy = (EX)(EY), \text{ da } \left[\int x\, f_X(x)\, dx\right] = EX. \end{aligned}$$

Definition 6.7
Die Kovarianz $\operatorname{cov}(X,Y)$ zweier ZV X und Y wird definiert durch

$$\operatorname{cov}(X,Y) = E\big((X-EX)(Y-EY)\big).$$

Als Korrelationskoeffizienten $\rho(X,Y)$ bezeichnet man $\rho(X,Y) = \dfrac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}$.

(Wobei natürlich vorausgesetzt wird, dass X und Y eine endliche, von 0 verschiedene Varianz haben.)

Klar:

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

Insbesondere (Satz 6.6): X und Y unabhängig $\Rightarrow \operatorname{cov}(X,Y) = 0$. (X und Y sind unkorreliert)
Die Umkehrung gilt nicht!

cov und $\rho$ können als Maß einer linearen Abhängigkeit angesehen werden.

Cauchy-Schwarz-Ungleichung

$$\left|E(X-EX)(Y-EY)\right| \le \sqrt{\left(E(X-EX)^2\, E(Y-EY)^2\right)}$$

$$\left|\operatorname{cov}(X,Y)\right| \le \sqrt{\operatorname{var}(X)\operatorname{var}(Y)} - \sigma(X)\sigma(Y)$$

Insbesondere nimmt der Korrelationskoeffizient nur Werte im Intervall [-1,1] an.

Beispiel 6.8

Eine Urne enthält 2 weiße, 3 schwarze und 2 rote Kugeln. 3 Kugeln werden zufällig und ohne Zurücklegen entnommen. X sei die Anzahl der entnommenen weißen und Y die der schwarzen Kugeln.

$$P(X=k, Y=l) = \frac{\binom{2}{k}\binom{3}{l}\binom{7-2-3}{3-k-l}}{\binom{7}{3}}, \text{ für } k=0,1,2,\; l=0,1,2,3,\; k+l \in \{0,1,2,3\}$$

Bei diesem „kleinen" Beispiel die gemeinsame Massenfunktion komplett als Tabelle aufgeschrieben werden:

| y\x | 0 | 1 | 2 | |
|---|---|---|---|---|
| 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 | $1/35$ |
| | $10/35$ | $20/35$ | $5/35$ | 1 |

$$P(X=2) = P(X=2, Y=0) + P(X=2, Y=1) + P(X=2, Y=2) + P(X=2, Y=3) = 5/35$$

$$\qquad\qquad 2/35 \quad + \quad 3/35 \quad + \quad 0 \quad + \quad 0 \quad = 5/35$$

Als Zeilen- bzw. Spaltensummen erhält man die Massenfunktionen zu den einzelnen ZV (der Komponenten des Zufallsvektors).
Darüber hinaus lassen sich mit der gemeinsamen Massenfunktion Wahrscheinlichkeiten für Ereignisse berechnen, die sich auf <u>beide</u> ZV beziehen.

$$P(X=Y) = P(X=0, Y=0) + P(X=1, Y=1) + P(X=2, Y=2)$$

$$= \quad 0 \quad + \quad 12/35 \quad + \quad 0 \quad = 12/35$$

$$P(X=1, Y=1) = 12/35 \neq 20/35 \cdot 18/35 = P(X=1)\cdot P(Y=1)$$

Die gemeinsame Massenfunktion kann <u>nicht</u> als Produkt der marginalen Massenfunktionen geschrieben werden, X und Y sind also <u>nicht</u> unabhängig.

$$E(XY) = \sum_x \sum_y xy\, P(X = x, Y = y)$$
$$= \frac{1}{35}(1\cdot 1\cdot 12 + 2\cdot 1\cdot 3 + 1\cdot 2\cdot 6) = \frac{30}{35} = \frac{6}{7}$$

$$EX = 0\cdot\frac{10}{35} + 1\cdot\frac{20}{35} + 2\cdot\frac{5}{35} = \frac{6}{7}$$
$$EY = 0\cdot\frac{4}{35} + 1\cdot\frac{18}{35} + 2\cdot\frac{12}{35} + 3\cdot\frac{1}{35} = \frac{45}{35} = \frac{9}{7}$$

also $\operatorname{cov}(X,Y) = E(XY) - (EX)(EY) = \frac{6}{7} - \frac{6}{7}\cdot\frac{9}{7} = \frac{-12}{49} \approx -0{,}245$

(„schwach“ antilineare Abhängigkeit)

Beispiel 6.9

Wenn X und Y die gemeinsame Verteilungsfunktion

$$F_{XY}(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & , x, y \ge 0 \\ 0 & , \text{sonst} \end{cases}$$

haben., so gilt für alle $x \ge 0$

$$F_X(x) = P(X \le x) = \lim_{y\to 0} P(X \le x, Y \le y) = 1 - e^{-x}$$

analog für $F_Y(y) = P(Y \le y) = 1 - e^{-y}$.

Insbesondere sind X und Y exponentialverteilt unter Parameter 1. Sind X und Y unabhängig?
Für $x, y \ge 0$ gilt

$$F_{XY}(x,y) = 1 - e^{-x} - e^{-y} + e^{-x-y} = (1-e^{-x})(1-e^{-y}) = F_X(x)F_Y(y)$$

Da die Verteilungsfunktion als Produkt der marginalen Verteilungsfunktionen geschrieben werden kann, ist die Antwort „Ja“.
Gibt es eine gemeinsame Dichte?

$$\frac{\partial^2}{\partial x \partial y} F(x,y) = \frac{\partial}{\partial x}\left(e^{-y} + e^{-x-y}\right) = e^{-x-y}$$
$$\left(= f_X(x) f_Y(y)\right)$$

Beachte: Man kann die marginale Dichtefunktion auch durch Ausintegrieren der anderen Komponenten erhalten:

$$f_X(x) = \int_0^\infty f_{XY}(x,y)\,dy = \int_0^\infty e^{-x-y}\,dy = e^{-y}$$

Damit lassen sich Wahrscheinlichkeiten ausrechnen, die sich auf beide ZV beziehen.

$$P(X \le 2Y) = \iint\limits_{\left\{\substack{(x,y):\, x\le 2y \\ x,y\ge 0}\right\}} f_{XY}(x,y)\,dy\,dx$$

$$= \int_0^\infty \int_{x/2}^\infty e^{-x-y}\,dy\,dx = \frac{2}{3}$$

Als Kovarianz erhält man wegen der Unabhängigkeit den Wert 0. $\left( EXY = \int_0^\infty \int_0^\infty e^{-x-y}\,dx\,dy ... \right)$

Der Begriff der Kovarianz erlaubt auch eine wichtige Aussage über die Varianz einer Summe von ZV:

$$\begin{aligned}
\operatorname{var}(X+Y) &= E(X+Y)^2 - \left(E(X+Y)\right)^2 \\
&= E\left(X^2 + 2XY + Y^2\right) - (EX + EY)^2 \\
&= E\left(X^2\right) - (EX)^2 + E\left(Y^2\right) - (EY)^2 + 2\left(E(XY) - (EX)(EY)\right) \\
&= \quad \operatorname{var}(X) \quad + \quad \operatorname{var}(Y) \quad + \quad 2\operatorname{cov}(X,Y)
\end{aligned}$$

Insbesondere bei Unabhängigkeit: $\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y)$

<u>Satz 6.10 (Gleichheit von Bienaymé)</u>

Sind $X_1, ..., X_n$ unabhängige ZV mit endlicher Varianz, so gilt

$\operatorname{var}(X_1, ..., X_n) = \operatorname{var}(X_1) + ... + \operatorname{var}(X_n)$.

In Worten: Bei unabhängigen ZV „addieren sich die Varianzen“

Ganz allgemein hat man:

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

10. Vorlesung (16.12.2004)

Beispiel 6.11

Es sei $A_1,...,A_r$ eine Ereignispartition, also $A_i \cap A_j = \emptyset$ für $i \neq j$ (paarweise disjunkt), und $A_1 \cup ... \cup A_r = \Omega$ (ein Ereignis muss eintreten). Seien $p_1,...,p_r$ die zugehörigen Wahrscheinlichkeiten, also $p_1 + ... + p_r = 1$ (s. Satz 3.3 (II) „Gesetz von der totalen Wahrscheinlichkeit").

Es sei $X_k$ die Anzahl der Experimente bei n Wiederholungen, in denen $A_k$ eingetreten ist, dann ist $X = \begin{pmatrix} X_1 \\ | \\ X_r \end{pmatrix}$ ein r-dimensionaler, diskreter Zufallsvektor.

Mit den Methoden aus Abschnitt 2 erhält man:

$$P\left( X = \begin{pmatrix} X_1 \\ | \\ X_r \end{pmatrix} \right) = P\left(X_1 = k_1, X_2 = k_2, ..., X_r = k_r\right)$$

$$= \frac{n!}{k_1! \cdot k_2! \cdot ... \cdot k_r!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot ... \cdot p_r^{k_r}$$

Multinominalkoeffizient, auch: $\binom{n}{k_1, k_2, ..., k_r}$

(Es gibt $\binom{n}{k_1}$ Möglichkeiten, die $k_1$ Elemente vom Typ $A_1$ auf die n Positionen zu verteilen, sind diese festgelegt, so bleiben $\binom{n-k_1}{k_2}$ für die vom Typ $A_2, \ldots$

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}...\binom{n-k_1-...-k_{r-1}}{k_r} = \frac{n! \qquad (n-k_1)!...}{k_1(n-k_1)! \quad k_2!(n-k_1-k_2)!...}$$

für alle $k_1,...,k_r \in \mathbb{N}_0$ mit $k_1 + ... + k_r = n$. Man nennt diese Verteilung die Multinomialverteilung mit Parametern n und $p = (p_1,...,p_r)$. Aus Abschnitt 2.1 bekannt: Die Marginalverteilungen sind binominal, $X_k \sim \mathrm{Bin}(n, p_k)$.

Zur Notation: $\binom{n}{k} \xrightarrow[\text{würde notiert als}]{} \binom{n}{n, n-k}$

Numerisches Beispiel:

Wirft man einen fairen Würfel 10mal, so ist die Wahrscheinlichkeit dafür, dass 1 und 6 je dreimal, die anderen Augenzahlen je einmal auftreten,

$$\frac{10!}{3!1!1!1!1!3!}\left(\frac{1}{6}\right)^3 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \left(\frac{1}{6}\right)^3 = \frac{175}{104976} \approx 0,00167 .$$

Beispiel 6.12

Ein d-dimensionaler Zufallsvektor $X = \begin{pmatrix} X_1 \\ | \\ X_d \end{pmatrix}$ mit Dichte

$f(x_1, \dots, x_d) = (2\pi)^{-\frac{d}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left( \frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu) \right)$ heißt d-dimensional normalverteilt unter Parametern μ und $\Sigma_i$ . (hierbei $\mu \in \mathbb{R}^d$ und $\Sigma_i \in \mathbb{R}^{d\times d}$ positiv definit)

In Analogie zum eindimensionalen Fall ist $\mu = \begin{pmatrix} \mu_1 \\ | \\ \mu_d \end{pmatrix}$ der Erwartungswert zu X im Sinne von $EX_k = \mu_k$ für $k = 1, \dots, d$. $\Sigma$ ist die Kovarianzmatrix zu X, d.h. in Zeile i, Spalte j steht $(\Sigma)_{ij} = \operatorname{cov}(X_i, X_j)$ (auf der Diagonalen: $(\Sigma)_{ij} = \operatorname{cov}(X_i, X_i) = \operatorname{var}(X_i)$)

Ist die gemeinsame Verteilung von X und Y bekannt, so lässt sich (im Prinzip) die Verteilung von $\boxed{Z := g(X,Y)}$, durch $g: \mathbb{R}^2 \to \mathbb{R}$ gegeben, bestimmen.
Besonders wichtig:

$Z = X + Y$, also $g: \mathbb{R}^2 \to \mathbb{R}$, $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to x_1 + x_2$

Satz 6.13

(I) Sind X, Y unabhängige, diskrete ZV mit Massenfunktionen $p_X, p_Y$, so ist X+Y wieder eine diskrete ZV, mit Massenfunktion

$$\begin{aligned} p_{X+Y}(z) &= \sum_x p_X(x) p_Y(z-x) \\ &= \sum_y p_X(z-y) p_Y(y) \end{aligned}$$

(Terminologie: man nennt $p_{X+Y}$ die Faltung von $p_X$ und $p_Y$)

(II) Sind X, Y unabhängige, stetige ZV mit Dichtefunktion $f_X, f_Y$, so ist auch X+Y eine stetige ZV und $f_{X+Y}(z) = \int f_X(x) f_Y(z-x)\,dx$

$$= \int f_X(z-y) f_Y(y)\,dy$$

ist eine zugehörige Dichtefunktion. (Terminologie: Man nennt $f_{X+Y}$ die Faltung von $f_X$ und $f_Y$)

Zum Beweis im diskreten Fall:
Zerlegung nach dem Wert von X liefert

$$\begin{aligned} P(X+Y=z) &= \sum_x P(X+Y=z, X=x) \\ &= \sum_x P(Y=z-x, X=x) \\ \text{(Unabhängigkeit!)} \quad &= \sum_x P(Y=z-x) P(X=x) \\ &= \sum_x p_Y(z-x) p_X(x) \end{aligned}$$

Beispiel 6.14

(I) Es seien X, Y Poisson-verteilt mit Parametern $\lambda_X$ und $\lambda_Y$.

Welche Verteilung hat X+Y?

Für alle $n \in \mathbb{N}_0$ gilt:

$$
\begin{aligned}
p_{X+Y}(n) &= \sum_{k\in\mathbb{Z}} p_X(k)\,p_Y(n-k) = \sum_{k=0}^{n} p_X(k)\,p_Y(n-k) \\
&= \sum_{k=0}^{n} e^{-\lambda_X} \frac{(\lambda_X)^k}{k!}\, e^{-\lambda_Y} \frac{(\lambda_Y)^{n-k}}{(n-k)!} \\
&= e^{-\lambda_X-\lambda_Y} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} (\lambda_X)^k (\lambda_Y)^{n-k} \\
&= e^{-\lambda_X-\lambda_Y} \frac{(\lambda_X+\lambda_Y)^n}{n!} \quad \text{(binom. Lehrsatz)}
\end{aligned}
$$

Hieraus folgt X+Y ist Poisson-verteilt mit Parameter $\lambda = \lambda_X + \lambda_Y$.

(II) Sind X und Y unabhängig und $N(0,1)$-verteilt, so gilt:

$$
\begin{aligned}
f_{X+Y}(z) &= \int f_X(x)\,f_Y(z-x)\,dx \\
&= \frac{1}{2\pi} \int e^{\frac{x^2}{2}}\, e^{\frac{-(z-x)^2}{2}}\,dx \\
&= \frac{1}{2\pi} \int \exp\left(-\frac{1}{2}\left(2x^2+z^2-2zx\right)\right) dx
\end{aligned}
$$

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

$$
\begin{aligned}
&= \frac{1}{2\pi} \exp\left(-\frac{z^2}{4}\right) \int \exp\left(-\left(x-\frac{1}{2}z\right)^2\right) dx \\
&= \frac{1}{\sqrt{2\pi\cdot 2}} \exp\left(-\frac{1}{2\cdot 2} z^2\right) \frac{1}{\sqrt{2\pi}} \int e^{-\frac{y^2}{2}}\,dy
\end{aligned}
$$

$\frac{1}{\sqrt{2\pi}} \int e^{-\frac{y^2}{2}}\,dy$ ist Integral zur Dichtefunktion von $N(0,1)$.

Der Vorfaktor ist von der Form $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(z-\mu)^2}$ mit $\mu = 0, \sigma^2 = 2$.

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

Allgemein: $X \sim N(\mu_X, \sigma^2{}_X)$, $Y \sim N(\mu_Y, \sigma^2{}_Y)$ unabhängig

$$\Rightarrow X + Y \sim N\left(\mu_X + \mu_Y, \sigma^2{}_X + \sigma^2{}_Y\right)$$

(kann beispielsweise mit Lemma 5.2 bewiesen werden)

Verteilungsfkt.
zu Y zu X

Unter bestimmten Bedingungen an die Transformation $\Psi$ hat mit X auch $Y := \Psi(X)$ eine Dichte.

Konkret im 1-dim. Fall: Ist $\Psi$ streng monoton wachsend und diffbar., so folgt:

$$
\begin{aligned}
f_Y(y) &= \frac{\partial}{\partial y} P(Y \le y) = \frac{\partial}{\partial y} P(\Psi(X) \le y) = \frac{\partial}{\partial y} P\left(X \le \Psi^{-1}(y)\right) = \frac{\partial}{\partial y} F_X\left(\Psi^{-1}(x)\right) \\
&= f_X\left(\Psi^{-1}(y)\right)\left(\Psi^{-1}\right)'(y) \qquad \left(f_Y = \left(\Psi^{-1}\right)' f_X \circ \Psi^{-1}\right)
\end{aligned}
$$

Satz 6.15 (Transformationssatz für Wahrscheinlichkeitsdichten)
Es seien U und V offene Teilmengen von $\mathbb{R}^d$ und $\Psi : U \to V$ eine bijektive, stetig diffbare Abbildung mit $\det\left(\Psi^1(x)\right) \neq 0$ f.a. $x \in U$.
Ist dann X eine ZV mit Dichte $f_X$ und $P(X \in U) = 1$, so hat $Y := \Psi(X)$ die Dichte
$f_Y(y) = \left|\det\left(\left(\Psi^{-1}\right)'\right)(y)\right| f_X\left(\Psi^{-1}(y)\right),\ y \in V$.

Beispiel 6.16 (Transformation auf Polarkoordinaten)
Wir schreiben r, $\varphi$ anstelle von $y_1, y_2$.

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

$$\left|\det\left(\Psi^{-1}\right)'\begin{pmatrix} r \\ \varphi \end{pmatrix}\right| = r\cos^2(\varphi) - \left(-r\sin^2(\varphi)\right) = r$$

Seien $X_1, X_2$ unabhängig und $N(0,1)$-verteilt

$$f_{X_1X_2}(x_1, x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2) = \frac{1}{2\pi}\exp\left(-\frac{x_1^2 + x_2^2}{2}\right)$$

Damit ergibt sich als gemeinsame Dichte von R und $\Phi$

$$f_{R,\Phi\infty}(r,\varphi) = \frac{1}{2\pi} e^{-\frac{r^2}{2}},\ \text{für } 0 < r < \infty,\ 0 \leq \Phi < 2\pi$$

Ausintegrieren liefert:

$$f_R(r) = er^{-\frac{r^2}{2}} \qquad f_\Phi(\varphi) = \frac{1}{2\pi}$$

Insbesondere sind R und $\Phi$ unabhängig $\Phi \sim \text{unif}(0, 2\pi)$.

11. Vorlesung (13.01.2005)

## §7 Erzeugende Funktionen

Es geht um Transformationsmethoden, die in vielen Bereichen der (angewandten) Mathematik von großer Bedeutung sind und außerdem eine starke Verbindung zur Analysis herstellen.

Ist X eine Zufallsgröße mit Werten in $\mathbb{N}_0$, so heißt

$$g_X(s) = Es^X = \sum_{n=0}^{\infty} P(X=n)s^n \quad \text{(Potenzreihe)}$$

die Wahrscheinlichkeitserzeugende Funktion (der Verteilung).

Kennt man $g_X$, so lassen sich die Wahrscheinlichkeiten $P(X=n)$ über die Ableitungen von $g_X$ in 0 bestimmen.

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

Beispiele:

$g_X(s) = \frac{1}{2}(1+s)$

$P(X=0) = g_X(0) = \frac{1}{2}$

$P(X=1) = \frac{1}{2}$

$\sum_{n=0}^{\infty} 1 \cdot s^n = \frac{1}{1-s}$ (geom. Reihe)

Insbesondere gilt: Die Verteilung von X ist durch $g_X$ festgelegt.

Aus der Analysis sind Potenzreihen bekannt, insbesondere ist der Konvergenzradius von $g_X$ mindestens 1.

Außerdem gilt, wenn der Erwartungswert zu X existiert,

$$EX = \sum_{n=0}^{\infty} n \cdot P(X=n) = \frac{d}{ds}\left(\sum_{n=0}^{\infty} P(X=n)s^n\right)|_{s=1} = g'_X(1)$$

Analog:
$E(X(X-1)) = g''_X(1)$, insbesondere

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

Sind X und Y unabhängige, $\mathbb{N}_0$-wertige Zufallsgrößen, so gilt

$$\boxed{\begin{aligned} g_{X+Y}(s) = Es^{X+Y} &= E\left(s^X \cdot s^Y\right) \quad \text{(als Fkt. Von unabh. ZV wieder unabhängig)} \\ \text{(Multiplikationsregel)} &= \left(Es^X\right) \cdot \left(Es^Y\right) = g_X(s) \cdot g_Y(s) \end{aligned}}$$

Die Wahrscheinlichkeitserzeugende Funktion zur Summe von unabhängigen ZV ist das Produkt der einzelnen Wahrscheinlichkeitserzeugenden Funktionen.

Beispiel 7.1

Ist X Poisson-verteilt mit Parameter $\lambda$, so erhält man

$$
\begin{aligned}
g_X(s) &= \sum_{n=0}^{\infty} P(X=n)s^n \\
&= \sum_{n=0}^{\infty}\left(e^{-\lambda}\frac{\lambda^n}{n!}\right)s^n \\
&= e^{-\lambda}\left(\sum_{n=0}^{\infty}\frac{(\lambda s)^n}{n!}\right) = e^{-\lambda}\cdot e^{\lambda s} = e^{\lambda(s-1)}
\end{aligned}
$$

Exponential-Reihe

$\boxed{=1}$

$$g_X'(s) = \lambda\left(e^{\lambda(s-1)}\right)$$

$$g_X'(1) = \lambda \qquad EX = \lambda$$

$$\operatorname{var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Sind X und Y unabhängig und Poisson-verteilt mit Parametern $\lambda_X$ und $\lambda_Y$, so gilt

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

Dies ist die Wahrscheinlichkeitserzeugende Funktion zur Poisson-Verteilung mit Parameter $\lambda_X + \lambda_Y$.
Insgesamt haben wir also bewiesen:
Sind X und Y unabhängig und Poisson-verteilt, so ist auch die Summe Poisson-verteilt (siehe Beispiel 6.14).

Beispiel 7.2

Wie betrachten die Wiederholungen eines Zufallsexperimentes in dem ein bestimmtes Ereignis mit der Wahrscheinlichkeit p eintritt. Für jedes n sei $X_n = 1$, wenn dieses Ereignis in der n-ten Wiederholung eintritt, und 0 sonst. Auf diese Weise erhält man eine Folge $(X_n)_{n\in\mathbb{N}}$ von unabhängigen ZV mit $P(X=1)=p,\ P(X_n=0)=1-p$.
Problem: Wie lange dauert es, bis erstmals A zweimal hintereinander eintritt?
(Allgemeiner: Muster in zufälliger Zeichenkette)

$T := \inf\{n \geq 2 : X_{n-1} = X_n = 1\}$
Offensichtlich gilt $P(T=0) = P(T=1) = 0$.
Für $n \geq 3$ gilt $P(T=n) = P(X_1=0, T=n) + P(X_1=1, X_2=0, T=n)$,
sowie $P(T=2) = P(X_1=1, X_2=1) = p^2$.
Hilfsgröße: $T_i := \inf\{n \geq 2 : X_{n-1+i} = X_{n+i} = 1\} \quad , i = 1,2,...$

Offensichtlich hängt $T_i$ nur von $X_{i+1}, X_{i+2}, \ldots$ ab, ist also unabhängig von $X_1, \ldots, X_i$.
Außerdem ist diese Abhängigkeit dieselbe, wie die von T bezogen auf $X_1, X_2, \ldots$ d.h.

$$\boxed{P(T_1 = n) = P(T_2 = n) = P(T = n)}.$$

Damit

$$\boxed{P(T = n) = P(X_1 = 0, T_1 = n-1) + P(X_1 = 1, X_2 = 0, T_2 = n-2)}$$

$$\boxed{\begin{aligned} &P(X_1 = 0)P(T_1 = n-1) \text{ für } n \geq 3 \\ &= P(X_1 = 0)P(T = n-1) \\ &= (1-p)P(T = n-1) \end{aligned}}$$

Übergang zur Wahrscheinlichkeitserzeugenden Funktion:

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

Damit

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

Dies liefert beispielsweise

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

Weitergehende Resultate: Sind $1/a$ und $1/b$ die Nullstellen des Nenners von $g_T$ (bei $0 < p < 1$ sind diese verschieden), so liefert eine Partialbruchzerlegung (geht auch mit einem Computeralgebra-Programm):

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

Wartet man beispielsweise beim Würfelwurf, bis erstmals zweimal hintereinander eine 6 erscheint, so erhält man mit $p = 1/6$ für die Nennerfunktion $1-\frac{5}{6}s\left(1+\frac{1}{6}s\right)=-\frac{5}{36}s^2-\frac{5}{6}s+1$ die Nullstellen $1/a$ und $1/b$ mit $a=-3+\frac{9}{5}\sqrt{5}$ und $b=-3-\frac{9}{5}\sqrt{5}$.

Also $P(T=n)=\frac{1\left(a^{n-1}-b^{n-1}\right)}{36(a-b)}$ $\qquad ET = 42$

Was passiert, wenn die interessierenden Zufallsgrößen nicht Werte ausschließlich in $\mathbb{N}_0$ annehmen?

Betrachte die sog. <u>momenterzeugende</u> Funktion
$\varphi_X(t)=Ee^{eX}=\int e^{tx}f(x)\,dx$, wenn X eine Dichtefunktion hat.
(allgemeiner anwendbar, aber keine Potenzwerte mehr)

12. Vorlesung (20.01.2005)

Im Falle $P(X \in \mathbb{N}_0) = 1: g_X(s) = Es^X$

Allgemeiner: Die momenterzeugende Funktion $\varphi - \varphi_X$ :

$$\varphi_X(t) = Ee^{tX} = \begin{cases} \sum_x e^{tx} P(X = x) & \text{, im direkten Fall} \\ \int e^{tx} f_X(x)\,dx & \text{, wenn X Dichte von } f_X \text{ hat} \end{cases}$$

ACHTUNG: Muss nicht unbedingt außerhalb von $t = 0$ existieren.

$\varphi_X(0) = 1$ , $f_X(t) = \infty \quad \forall t \neq 0$ ist möglich

Existiert $f_X(t)$ in einer Nullumgebung, so lassen sich Erwartungswert und Ableitung (nach t) vertauschen.

$$\varphi_X'(t) = \frac{d}{dt}\left(Ee^{tX}\right) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$$

$$\varphi_X'(0) = EX \qquad \varphi_X''(t) = E\left(X^2 e^{tX}\right)$$

$$\varphi_X^{(k)}(t) = E\left(X^k e^{tX}\right) \text{, also } \varphi_X^{(k)}(0) = EX^k$$

Das k-te Moment ergibt sich als der Wert der k-ten Ableitung von $f_X$ in 0 – daher „momenterzeugende Funktion“.

Zusammenhang zu wahrscheinlichkeitserzeugender Funktion (bei $P(X \in \mathbb{N}_0) = 1$):

$$g_X(s) = Es^X = E\exp\left((\log s)X\right) = \varphi_X(\log s)$$

Bei $X \sim \text{Poisson}(\lambda)$ ist aus Beispiel 7.1 $g_X(s) = e^{\lambda(s-t)}$ bekannt, also

$\varphi_X(t) g_X(e^t) = \exp\left(\lambda\left(e^t - 1\right)\right)$.

Beispiel 7.3

a) Ist X exponentialverteilt mit Parameter $\lambda$, so erhält man

$$\varphi_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\,dx = \frac{\lambda}{\lambda - t} \quad \text{für } \lambda < 2$$

(Insbesondere: Der Definitionsbereich von $\varphi_X$ ist $(-\infty, \lambda)$, also nicht ganz $\mathbb{R}$, enthält aber ein Nullumgebung.

b) Ist X standartnormalverteilt, so erhält man

$$\varphi_X(t) = \int e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = e^{t^2/2} \underbrace{\left( \int \underbrace{\left( \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-t)^2}{2}} \right)}_{\text{Dichte zu } N(t,1)} dx \right)}_{=1} = e^{t^2/2} \quad , t \in \mathbb{R}$$

(Der Definitionsbereich ist her ganz $\mathbb{R}$ ).

Zusammenspiel mit affinen Transformationen : Zu $Y = aX + b$ gehört

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

Ist beispielsweise X normalverteilt mit Parametern μ und $\sigma^2$ so ist $Y := \frac{X-\mu}{\sigma}$ standartnormalverteilt (Beispiel 6.14 (ii)) also folgt mit Beispiel 7.3 (b)

$$\varphi_X(t) = e^{t\mu}\varphi_Y(\sigma t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \quad (\text{mit } X = \mu + \sigma Y).$$

Ganz analog wie bei wahrscheinlichkeitserzeugenden Funktionen erhält man für unabhängige ZV X und Y

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

Sind beispielsweise X und Y unabhängig mit $X \sim N\left(\mu_X, \sigma^2{}_X\right)$, $Y \sim N\left(\mu_Y, \sigma^2{}_Y\right)$, so folgt:

$$\begin{aligned}\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) &= \exp\left(\mu_X t + \frac{1}{2}\sigma^2{}_X t^2\right)\cdot\exp\left(\mu_Y t + \frac{1}{2}\sigma^2{}_Y t^2\right)\\ &= \exp\left(\left(\mu_X + \mu_Y\right)t + \frac{1}{2}\left(\sigma^2{}_X + \sigma^2{}_Y\right)t^2\right)\end{aligned}$$

Dies ist die momenterzeugende Funktion zu $N\left(\mu_X + \mu_Y, \sigma^2{}_X + \sigma^2{}_Y\right)$. Man sieht: Die Summe von unabhängigen, normalverteilten Zufallsgrößen ist wieder normalverteilt.
(geht auch über Faltungsformeln, aber mühsam)
Diese Aussage lässt sich auf mehr als 2 Summanden verallgemeinern:
Sind $X_1, \ldots, X_n$ unabhängig, so gilt:
$\varphi_{X_1+\ldots+X_n}(t) = \varphi_{X_1}(t) + \ldots + \varphi_{X_n}(t)$.
Dies kann sogar auf zufällige Summandenzahl verallgemeinert werden: Sind beispielsweise $X_1, X_2, \ldots$ unabhängig und identisch verteilt mit momenterzeugender Funktion $\varphi_X$ und weiter N eine $\mathbb{N}_0$-wertige Zufallsgröße mit wahrscheinlichkeitserzeugender Funktion $g_N$, N unabhängig von den X-Variablen, so gilt für die zufällige Summe $Y := \sum_{n=0}^{\infty} X_n$.

(Eine „leere Summe“ hat den Wert 0)

$$\varphi_Y(t) = E\exp(tY) = \sum_{n=0}^{\infty} E\left(\exp(tY)1_{\{N=n\}}\right)$$

Zerlegung nach dem Wert von N

$$\begin{aligned}&= \sum_{n=0}^{\infty} E\left(\left(\sum_{k=1}^{n} X_k\right)1_{\{N=n\}}\right)\\ &= \sum_{n=0}^{\infty} E\left(\exp\left(\sum_{k=1}^{n} X_k\right)\right)\cdot E\left(1_{\{N=n\}}\right)\\ &= \sum_{n=0}^{\infty} \varphi_X(t)^n P(N=n) = g_N\left(\varphi_X(t)\right)\end{aligned}$$

Beispiel 7.4

Ist N Poisson-verteilt mit Parameter $\lambda$ und sind $X_1, X_2, \dots$ unabhängig (auch von N) mit $P(X_i = 1) = p = 1 - P(X_i = 0)$ f.a. $i \in \mathbb{N}$, so gilt wegen

$\varphi_X(t) = Ee^{tX_1} = e^{t\cdot 1}p + e^{t\cdot 0}(1-p) = 1 - p + pe^t$ für die zufällige Summe $Y = \sum_{k=1}^{N} X_k$ :

$$\begin{aligned}\varphi_Y(t) &= g_N(\varphi_X(t)) \qquad \left(g_N(x) = e^{\lambda(x-1)}\right)\\ &= \exp\left(\lambda(\varphi_X(t) - 1)\right)\\ &= \exp\left(\lambda(1 - p + pe^t - 1)\right)\\ &= \exp\left(\lambda p(e^t - 1)\right).\end{aligned}$$

Dies ist die momenterzeugende Funktion zur Poisson-Verteilung mit Parameter $\lambda p$. Man kann dies als Ausdünnungseigenschaft der Poisson-Verteilung ansehen.

## §8 Grenzwertsätze

### 8.1 Ein Gesetz der großen Zahlen

Es sei X eine ZV mit Erwartungswert $\mu$ und Varianz $\sigma^2$ $(<\infty)$. Wir definieren Y durch

$$Y := \begin{cases} \varepsilon, & |X - \mu| \ge \varepsilon \\ 0, & \text{sonst} \end{cases}$$

Klar: $Y^2 \le (X - \mu)^2$

Hieraus folgt mit Satz 4.10 bzw. Satz 5.6

$$EY^2 \le E(X - \mu)^2 = \operatorname{var}(X) = \sigma^2 .$$

Da Y nun zwei Werte annehmen kann, erhält man

$$EY^2 = \varepsilon^2 P(Y = \varepsilon) + 0^2 P(Y = 0) = \varepsilon^2 P(|X - \mu| \ge \varepsilon).$$

Satz 8.1 (Die Ungleichung von Chebyshev)

$$P(|X - EX| \ge \varepsilon) \le \frac{1}{\varepsilon^2} \operatorname{var}(X) \quad \text{f.a. } \varepsilon > 0$$

Spezialfall: Mit $\varepsilon = 3\sigma$ erhält man die sog. Drei-Sigma-Regel:
Die Wahrscheinlichkeit dafür, dass der Wert einer ZV mehr als 3 Standartabweichungen vom Mittelwert entfernt ist, ist höchstens $1/9$. („10% bei Anwendern“)

Wir betrachten die folgende Situation:
Ein Zufallsexperiment wird n-mal unabhängig und ohne Veränderung der Nebenbedingungen wiederholt. Wir beobachten die ZV $X_1, \dots, X_n$ $\left(X_i \text{ in der } i-\text{ten Wiederholung}\right)$.

Der folgende Satz zeigt, dass unter schwachen Bedingungen der Mittelwert $\overline{X_n} := \frac{1}{n} \sum_{k=1}^{n} X_k$ der Beobachtungen mit $n \to \infty$ in einem bestimmten Sinne (Konvergenz in Wahrscheinlichkeit) gegen den Erwartungswert (eine Konstante!) konvergiert.

<u>Satz 8.2 (Das schwache Gesetz der großen Zahlen)</u>

Es sei $X_1, X_2, \dots$ eine Folge von unabhängigen und identisch verteilten ZV mit Erwartungswert μ und Varianz $\sigma^2 < \infty$. Dann gilt

$$\lim_{n\to\infty} P\left(\left|\overline{X_n} - \mu\right| \geq \varepsilon\right) = 0 \quad \text{f.a. } \varepsilon > 0$$

Der Beweis ist einfach: Es gilt

$$E\overline{X_n} = \frac{1}{n}\left(EX_1 + EX_2 + \dots + EX_n\right) = \mu$$

$$\begin{aligned} \operatorname{var}\left(\overline{X_n}\right) &= \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\operatorname{var}\left(X_1 + \dots + X_n\right) \\ &= \frac{1}{n^2}\left(\operatorname{var}\left(X_1\right) + \dots + \operatorname{var}\left(X_n\right)\right) \quad \text{(Bienaymé)} \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

$$P\left(\left|\overline{X_n} - \mu\right| \geq \varepsilon\right) = P\left(\left|\overline{X_n} - E\overline{X_n}\right| \geq \varepsilon\right) \underset{\substack{\uparrow \\ \boxed{\text{Chebyshev}}}}{\leq} \frac{1}{\varepsilon^2}\operatorname{var}\left(\overline{X_n}\right) = \frac{1}{\varepsilon^2}\frac{1}{n}\sigma^2 \xrightarrow[n\to\infty]{} 0.$$

q.e.d.

<u>13. Vorlesung (27.01.2005)</u>

Ein besonders wichtiger Spezialfall ergibt sich im Zusammenhang mit relativen Häufigkeiten: Ist A ein Ereignis mit Wahrscheinlichkeit p und setzt man

$$X_i = \begin{cases} 1, & A \text{ im } i-\text{ten Versuch eingetroffen} \\ 0, & \text{sonst} \end{cases}$$

(wird das Experiment unabhängig wiederholt und verändern sich die Rahmenbedingungen nicht, so können wie dir $X_i$'s als unabhängig und $\mathrm{Bin}(1,p)-\text{verteilt}$ ansehen)

Klar: $\mu = EX_i = p$, $\operatorname{var}(X_i) = p(1-p) < \infty$.

Das schwache Gesetz der großen Zahlen ist also anwendbar.

Interpretation:

$\overline{X_n} = \frac{1}{n}\overbrace{\left(\sum_{i=1}^{n} X_i\right)}^{\text{Wie oft ist das Ergebnis in n Versuchen eingetreten}}$ ist relative Häufigkeit des Eintretens von A in den ersten n Wiederholungen.

$P\left(\left|\overline{X_n} - p\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$ für alle $\varepsilon > 0$ zeigt nur, dass relative Häufigkeiten bei großen n mit großer Wahrscheinlichkeit in der Nähe der zugehörigen Wahrscheinlichkeit liegen.

Dies kann als (erste) Rechtfertigung des axiomatischen Aufbaus der Stochastik dienen.

## <u>8.2 Normalapproximation: der Zentrale Grenzwertsatz</u>

Nach dem schwachen Gesetz der großen Zahlen liegt $\overline{X_n}$ bei großen n in der Nähe von μ.

Nach 1000 Würfen eines fairen Würfels wird man also als mittleres Wurfergebnis (Augenzahl) etwa 3,5 ($EX_n$) erhalten, also etwa Augensumme 3500.

### <u>Satz 8.3 (Der Zentrale Grenzwertsatz)</u>

Zusätzlich zu den Voraussetzungen in Satz 8.2 sei $\sigma^2 > 0$. Dann gilt für alle $x \in \mathbb{R}$

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

Interpretation: $\frac{\sqrt{n}}{\sigma}\left(\overline{X_n} - \mu\right)$ ist bei großen n näherungsweise $N(0,1)$-verteilt.

Mit den Rechenregeln für Normalverteilungen (Lemma 5.2) erhält man die Approximationen

$\overline{X_n}$ ist näherungsweise $N\left(\mu, \sigma^2/n\right)$-verteilt.

$S_n := \sum_{i=1}^{n} X_i$ ist ebenfalls näherungsweise $N(\underbrace{n\mu}_{=ES_n}, \underbrace{n\sigma^2}_{=\operatorname{var}(S_n)})$-verteilt.

Mittelwert und Summe von unabhängigen, identisch verteilten ZV sind näherungsweise Normalverteilt, wobei für μ und $\sigma^2$ (also die Parameter der Normalverteilung) der Erwartungswert und die Varianz des Mittelwertes bzw. der Summe genommen werden.

<u>Beispiel 8.4</u>

In einem Zufallsexperiment tritt ein bestimmtes Ereignis mit Wahrscheinlichkeit p ein, $0 < p < 1$. Konkret hat man beispielsweise $p = \frac{1}{2}$ für „Kopf“ beim Wurf einer fairen Münze, $p = \frac{1}{6}$ beim Würfelwurf für das Ereignis „6 kommt“.

Ist $S_n$ die Anzahl der Erfolge nach n Wiederholungen, so gilt bekanntlich $S_n \sim \mathrm{Bin}(n, p)$ und der Zentrale Grenzwertsatz führt auf die Approximationen $N(np, np(1-p))$ bzw. $N(0,1)$ für die standardisierte Summenvariable $\frac{S_n - np}{\sqrt{np(1-p)}}$. Damit lässt sich beispielsweise eine näherungsweise Antwort auf Fragen von folgendem Typ geben:

Mit welcher Wahrscheinlichkeit erscheint beim 600maligen Wurf eines (fairen) Würfels mindestens 95mal und höchstens 105mal eine „6“?

Sei X diese Anzahl. Wegen $X \sim \mathrm{Bin}\left(600, \frac{1}{6}\right)$ gilt

$$P(95 \le X \le 105) = \sum_{k=95}^{105} \binom{600}{k} \left(\frac{1}{6}\right)^k \left(1 - \frac{1}{6}\right)^{600-k} = 0,453101...$$

Dies ist die exakte Antwort, die man beispielsweise mit einem Computeralgebraprogramm wie Maple leicht erhalten kann.

Mit $EX = 600 \cdot \frac{1}{6} = 100$ und $\mathrm{var}(X) = 600 \cdot \frac{1}{6}\left(1 - \frac{1}{6}\right) = \frac{500}{6}$ ergibt die Normalapproximation

$$\begin{aligned}
P(95 \le X \le 105) &= P(-5 \le X - EX \le 5) \\
&= P\left(\frac{-5}{\sqrt{500/6}} \le \frac{X - EX}{\sqrt{\mathrm{var}(X)}} \le \left(\frac{5}{\sqrt{500/6}}\right)\right) \\
&\approx \Phi(0,5477) - \Phi(-0,5477) \quad \text{denn die standardisierte Variable} \\
&\qquad \text{ist näherungsweise } N(0,1)\text{-verteilt.} \\
&= \Phi(0,5477) - (1 - \Phi(0,5477)) \\
&= 2\Phi(0,5477) - 1 \approx 2 \cdot 0,7088 - 1 \\
&= 0,4176
\end{aligned}$$

Dies kann mit der <u>Stetigkeitskorrektur</u> verbessert werden:

Die Verteilungsfunktion von X ist auf Intervallen der Länge 1 konstant. Wählt man den jeweiligen Intervallmittelpunkt, so erhält man

$$\begin{aligned}
P(95 \le X \le 105) &= P(94,5 \le X \le 105,5) \\
&= P(-5,5 \le X - EX \le 5,5) \\
\ldots &\approx 2\Phi(0,65025) - 1 \approx 0,4532
\end{aligned}$$

Ein „richtiger“ Beweis liegt etwas außerhalb unserer Möglichkeiten. Wir begnügen uns mit dem Nachweis der Konvergenz der momenterzeugenden Funktion, setzen also insbesondere voraus, dass $Ee^{tX_1} < \infty$ für $|t| \le \delta$ mit einem $\delta > 0$ gilt.

Zunächst eine einfache Transformation: $Y_n := \frac{1}{6}(X_n - \mu)$.

Dies liefert eine Folge $(Y_n)_{n\in\mathbb{N}}$ von unabhängigen, identisch verteilten ZV mit Erwartungswert 0 und Varianz 1.
Klar:

$$\boxed{\sqrt{n}\,\frac{\overline{X_n}-\mu}{\sigma} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k}$$

$$\varphi_{\sqrt{n}\frac{\overline{X_n}-\mu}{\sigma}}(t) = \varphi_{\frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k}(t)$$

$$= \varphi_{\sum_{k=1}^{n} Y_k}\left(\frac{t}{\sqrt{n}}\right) \text{ (Rechenregel für momenterzeugende Funktionen)}$$

$$= \left(\varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

(Verhalten von momenterzeugenden Funktionen bei der Summe von unabhängigen ZV)

$$= \left(\underbrace{\varphi_{Y_1}(0)}_{=1} + \frac{t}{\sqrt{n}}\underbrace{\varphi'_{Y_1}(0)}_{=EY_1=0} + \frac{1}{2}\frac{t^2}{n}\left(\underbrace{\varphi''_{Y_1}(0)}_{=EY_1^2=1} + t\cdot o(1)\right)\right)^n$$

Vernachlässigbarer Restterm

$$= \left(1 + \frac{t^2}{2n}\left(1 + \cancel{o(1)}\right)\right)^n \xrightarrow[n\to\infty]{} e^{t^2/2} \quad (t > 0 \text{ fest})$$

Dies ist die momenterzeugende Funktion zu $N(0,1)$.

<u>Beispiel 8.5</u>

Die Lebensdauern $X_1, \dots, X_{10}$ von 10 Glühlampen seien unabhängig und exponential verteilt mit Erwartungswert 100 (Stunden). Gesucht ist ein Wert z mit der Eigenschaft, dass die Gesamtbrenndauer $S = X_1 + \dots + X_{10}$ mit Wahrscheinlichkeit 0,9 den Wert z übersteigt.

Es gilt $ES = 1000$, $\operatorname{var}(S) = 10^5$, also ist $\frac{S-1000}{\sqrt{10^5}}$ näherungsweise $N(0,1)$-verteilt.

Gesucht ist ein Wert z mit $P(S \ge z) = 0{,}9$.

$$P(S \ge z) = P\left(\frac{S-10^3}{\sqrt{10^5}} \ge \frac{z-10^3}{\sqrt{10^5}}\right)$$

$$= 1 - P\left(\frac{S-10^3}{\sqrt{10^5}} \le \frac{z-10^3}{\sqrt{10^5}}\right) \approx 1 - \Phi\left(\frac{z-10^3}{\sqrt{10^5}}\right) = \Phi\left(\frac{10^3-z}{\sqrt{10^5}}\right)$$

Tabelle: $= \Phi(1{,}282) \approx 0{,}9$

Man wird auf $\frac{10^3-z}{\sqrt{10^5}} = 0{,}9$ geführt, dies liefert $z = 10^3 - 1{,}282\cdot\sqrt{10^5} = 594{,}6$ Stunden.

14. Vorlesung (03.02.2005)

§9 Stochastik und Algorithmen

Wichtig für Anwendungen, aktuell in der Forschung: Algorithmen, die (selbstgemachten) Zufall als Hilfsmittel verwenden.
Klassisches Beispiel: Quicksort.
Gegeben ist ein n - Tupel $(a_1,...,a_n)$ von (paarweise verschiedenen) reellen zahlen; diese sollen aufsteigend geordnet werden.

(1) Wähle k zufällig und gleichverteilt (!) aus $\{1,...,n\}$

(2) Bestimme durch Vergleich von $a_k$ mit $a_1,...,a_{k-1},a_{k+1},...,a_n$ die Tupel
$a^{(L)}=\left(a_1^{(L)},...,a_j^{(L)}\right)$ , $a^{(R)}=\left(a_1^{(R)},...,a_{n-j-L}^{(R)}\right)$

(3) Sortiere $a^{(L)}$ zu $b^{(L)}$ und $a^{(R)}$ zu $b^{(R)}$

(4) Füge zusammen:
$\left(b_1,...,b_{n-j-L}\right)=\left(b_1^{(L)},...,b_j^{(L)},a_k,b_1^{(R)},...,b_{a-j-L}^{(R)}\right)$.

Dies ist ein „Teile und herrsche"-Algorithmus: durch „Splitter" (auch Pivot) $a_k$ wird das Ausgangsproblem $a\to b$ reduziert auf die beiden (kleineren) Teilprobleme
$a^{(L)}\to b^{(L)}$, $a^{(R)}\to b^{(R)}$.
Man spricht von einem „randomisierten" Algorithmus, weil in (1) ein zufallsabhängiger Schritt eingebaut ist.
Als Konsequenz ist die Laufzeit eine Zufallsgröße (!).

Worst-case-Analyse: Was passiert schlimmstenfalls?
Maßgeblich für die Laufzeit ist die Zahl der Vergleiche. Wird jedes Mal der kleinste Wert gewählt, so ist die (L)-menge jeweils leer, und die (R)-Menge verringert sich nur um ein Element.

$$(n-1)+(n-2)+(n-3)+...+3+2+1=\frac{1}{2}n(n-1) \quad \approx n^2$$

Allerdings ist dieses Ereignis extrem unwahrscheinlich.

Average-case-Analyse: Was passiert im Mittel?

Satz 9.1
Es sei $X_n$ die von Quicksort benötigte Anzahl von Vergleichen bei Inputumfang n. Dann gilt $\boxed{EX_n=2(n+1)H_n-4n}$, wobei $H_n:=\sum_{k=1}^{n}\frac{1}{k}$ die n-te harmonische Zahl bezeichnet.

Beweis:
Sei $f(n)=EX_n$. Der Schritt (3) liefert

$$f(n)=(n-1)+\frac{1}{n}\sum_{k=1}^{n}\left(f(k-L)+f(n-k)\right)=(n-1)+\frac{2}{n}\sum_{k=0}^{n-1}f(k)$$

Klar: $f(0) = f(1) = 0$

Diese Rekursion können wir auflösen:

$$n\,f(n) = n(n-1) + 2\sum_{k=1}^{n-1} f(k)$$

$$(n-1)\,f(n-1) = (n-1)(n-2) + 2\sum_{k=1}^{n-2} f(k)$$

also

$$n \cdot f(n) - (n-1)\cdot f(n-1) = 2(n-1) + 2\cdot f(n-1)$$

$$n \cdot f(n) = (n+1)\cdot f(n-1) + 2(n-1)$$

Für $g(n) = \frac{1}{n+1} f(n)$ bedeutet dies:

$$g(n) = g(n-1) + \frac{2(n-1)}{n(n+1)}$$

$$= g(n-2) + \frac{2(n-1)}{n(n+1)} + \frac{2(n-2)}{(n-1)n}$$

$$\ldots\ldots$$

$$= 2\sum_{i=1}^{n} \frac{i-1}{i(i+1)} = 2\sum_{i=1}^{n}\left(\frac{2}{i+1} - \frac{1}{i}\right) = 2H_n - \frac{4n}{n+1}$$

und somit

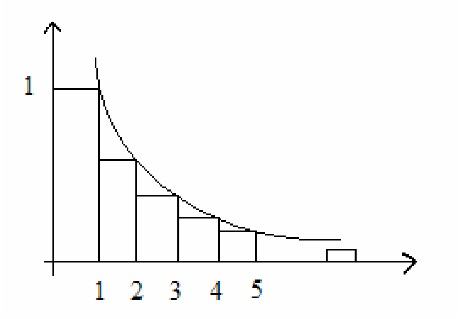
$$f(n) = (n+1)\cdot g(n) = 2(n+1)H_n - 4n$$

q.e.d.

Wie verhält sich $EX_n$ mit $n \to \infty$?

Vergleicht man die Funktion $x \to \frac{1}{x}$ mit geeigneten Treppenfunktionen, so erhält man

$\log n \le H_n \le (\log n) + 1$ f.a. $n \in \mathbb{N}$,




$$\int_1^n \frac{1}{x}\,dx \ge H_n - 1$$

$$\boxed{\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \ldots}$$

Insbesondere erhält man:

$$\frac{H_n}{\log n} \xrightarrow[n\to\infty]{} 1, \text{ also } \lim_{n\to\infty} \frac{EX_n}{n \log n} = 2$$

Also: im Mittel braucht man nur etwa $2n\log n$ Vergleiche. Dies ist viel besser, als $\frac{1}{2}n^2$, und die Größenordnung selbst ist in der Klasse der Vergleichsbasierten Sortieralgorithmen optimal.