Statistische Methoden

Dr. C.J. Luchsinger

1 Repetition: Wahrscheinlichkeitstheorie

Literatur Kapitel 1

* Cartoon Guide: Kapitel 3 - 5
* Krengel: § 1 - 3, 5, 10 - 12
* Stahel: Kapitel 4 - 6

1.1 Wahrscheinlichkeit

Wir werden die naive Vorstellung von Wahrscheinlichkeit und Zufall formalisieren, in die Sprache der Mathematik übersetzen. Als erstes müssen wir, ohne die Wahrscheinlichkeiten anzuschauen, Versuchsausgänge formalisieren.

1.1.1 Ereignisraum (auch Menge der Versuchsausgänge)

* Sei Ω eine nichtleere Menge.

* Sie steht für die Menge der Versuchsausgänge; wir nennen sie auch Ereignisraum. Es findet jeweils in einem Experiment genau ein sogenanntes Elementarereignis statt, z.B. $\omega_1 \in \Omega$ oder $\omega_2 \in \Omega$ etc.

* Bei einem Würfel wird man sinnvollerweise mit $\Omega := \{1, 2, 3, 4, 5, 6\}$ die Menge der möglichen Anzahl Augen wählen.

* Bei Münzwurf wird man sinnvollerweise $\Omega := \{K, Z\}$ wählen. $k-$facher Münzwurf: $\Omega := \{K, Z\}^k$.

* Die Menge Ω kann von endlicher oder unendlicher Kardinalität sein; sogar überabzählbar unendlich.

* Beim Würfel können wir fortfahren mit: A ist mit $A := \{2, 4, 6\}$ die Menge der geraden Zahlen und $B := \{4, 5, 6\}$ die Menge der Zahlen streng grösser 3.

* A bzw. B nennen wir ein Ereignis. Ereignisse sind allgemein Teilmengen von Ω. Dies ist verwirrend. Es kann ja (Ereignis A) nicht gleichzeitig $2, 4$ und 6 geworfen werden. Es gilt nach wie vor, dass genau ein sogenanntes Elementarereignis stattfindet. Wenn dieses ω_1 z.B. gleich 2 ist, so sagen wir: A ist eingetreten - es ist ja auch eine gerade Zahl geworfen worden; eine reicht!

Wir fassen die mengentheoretischen Ausdrücke und ihre Bedeutung für die Wahrscheinlichkeitstheorie in folgender Tabelle zusammen:

Symbol	**Mengentheorie / Bedeutung für die WT**
Ω	Menge / Ereignisraum, Menge der Versuchsausgänge
ω	Element von Ω / Elementarereignis, Versuchsausgang
A	Teilmenge von Ω / Ereignis; falls $\omega \in A$, sagt man, dass das Ereignis A eingetreten ist
A^c	Komplement von A / kein Elementarereignis aus A findet statt
$A \cap B$	Schnittmenge von A und B / ein Elementarereignis aus A und B findet statt
$A \cup B$	Vereinigung von A und B / ein Elementarereignis aus A oder B findet statt
$A \subset B$	A ist Teilmenge von B / Wenn ein Elementarereignis aus A stattfindet, dann immer auch ein Elementarereignis aus B
ϕ	leere Menge / unmögliches Ereignis
Ω	ganze Menge / sicheres Ereignis (etwas muss passieren)
$\limsup A_n$	$\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k$ / Ereignis, bestehend im Eintreten von unendlich vielen der Ereignisse $A_1, A_2, \ldots$.
$\liminf A_n$	$\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k$ / Ereignis, bestehend im Eintreten aller Ereignisse $A_1, A_2, \ldots$ mit eventueller Ausnahme einer endlichen Anzahl.

1.1.2 σ-Algebren, Hausaufgabe für mathematische Pharisäer

Nicht Prüfungsstoff!

Wir wollen den Ereignissen (z.B. A aus Ω) später eine Wahrscheinlichkeit ($P[A]$) zuordnen. Wenn wir mehrere Ereignisse vorgegeben haben, wollen wir auch die Wahrscheinlichkeiten von deren Vereinigungen, Durchschnitten oder Komplementen angeben können. An die Menge der Teilmengen von Ω, welche wir untersuchen, stellen wir also ein paar wenige Bedingungen:

Definition 1.1 [σ-Algebra] *Ein Teilmengensystem $\mathcal{A}$ von Ω heisst σ-Algebra, wenn folgende 3 Bedingungen erfüllt sind:*

a) $\Omega \in \mathcal{A}$

b) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

c) $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \cup_{n \geq 1} A_n \in \mathcal{A}$.

Falls $|\Omega| = n < \infty$, so hat die Potenzmenge von Ω bekanntlich Kardinalität 2^n, ist also wiederum endlich. Wir werden im Fall $|\Omega| = n < \infty$ in dieser Vorlesung einfach als $\mathcal{A}$ die Potenzmenge von Ω wählen.

Nur nebenbei sei erwähnt, dass man z.B. im Fall $\Omega = \mathbb{R}$ etliche Klippen überwinden muss(te), um sowohl diese Konzepte wie auch die dazugehörigen Wahrscheinlichkeitsmasse $P[.]$ korrekt und ohne Widersprüche zu definieren. Dies wird in der Masstheorie behandelt. Wir gehen hier nicht weiter darauf ein, sondern präsentieren im Eilzugtempo: Wenn wir $\Omega = \mathbb{R}$ wählen, dann ist die sogenannte *Borel-σ-Algebra* $\mathcal{B}(\mathbb{R})$ die kleinste σ-Algebra auf $\mathbb{R}$, welche z.B. alle geschlossenen Intervalle enthält. Dieses Teil-Mengensystem enthält auch alle halboffenen und alle offenen Intervalle, auch "fast alle" sonstigen Teilmengen von $\mathbb{R}$. Man muss aktiv und gezielt Mengen konstruieren, welche in dieser σ-Algebra *nicht* enthalten sind!

1.1.3 Wahrscheinlichkeit $P[.]$

Definition 1.2 [Wahrscheinlichkeit P] *Eine Wahrscheinlichkeit P ist eine reellwertige Funktion auf den Mengen aus $\mathcal{A}$. Dabei müssen folgende 3 Bedingungen erfüllt sein:*

a) $A \in \mathcal{A} \Rightarrow 0 \leq P[A] \leq 1$

b) $P[\Omega] = 1$

c) Sei $\{A_i\}_{i=1}^{\infty}$ *eine abzählbare Folge von disjunkten Mengen aus* $\mathcal{A}$*, dann muss gelten:*

$$P[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i].$$

Man darf in Definition 1.2 c) z.B. auch $A_i = \phi, i \geq 3$ wählen!

Aus dieser Definition lassen sich nützliche Eigenschaften ableiten, welche wir im folgenden Lemma zusammenfassen.

Lemma 1.3 [nützliche Eigenschaften von P] *Mit* $A, B \in \mathcal{A}$ *gelten folgende Aussagen:*

a) Prinzip der Gegenwahrscheinlichkeit: $P[A] = 1 - P[A^c]$*; v.a.* $P[\phi] = 0$

b) Sei $\{A_i\}_{i=1}^{\infty}$ *eine abzählbare Folge von Mengen aus* $\mathcal{A}$*, dann muss gelten:*

$$P[\cup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} P[A_i].$$

c) $A \subset B \Rightarrow P[B] = P[A] + P[B \backslash A]$

d) $A \subset B \Rightarrow P[A] \leq P[B]$

e) $P[A \cup B] = P[A] + P[B] - P[A \cap B]$*.*

1.1.4 Wahrscheinlichkeitsraum

Definition 1.4 [Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$] *Wir nennen ein Tripel der Form* $(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum.*

1.1.5 Bedingte Wahrscheinlichkeit $P[A|B]$

Definition 1.5 [Bedingte Wahrscheinlichkeit $P[A|B]$]

$$P[A|B] := \frac{P[A \cap B]}{P[B]},$$

falls $P[B] > 0$. Man nennt $P[A|B]$ die bedingte Wahrscheinlichkeit von A gegeben B.

Es gilt:

$$P[A|B]P[B] = P[A \cap B] = P[B|A]P[A].$$

Der Leser / die Leserin zeige: $P[.|B]$ ist selber auch eine Wahrscheinlichkeit:

Sei $B \subset A$. Zeigen Sie, dass dann gilt: $P[A|B] = 1$. Warum muss dies so sein (anschaulich)?

Wir leiten über zu 1.1.6; untersuchen Sie $P[A|B]$, falls $P[A \cap B] = P[A]P[B]$.

1.1.6 Unabhängigkeit von Ereignissen

Definition 1.6 [Unabhängigkeit von Ereignissen] *Ereignisse A und B sind (paarweise) unabhängig voneinander, wenn $P[A \cap B] = P[A]P[B]$. Falls $P[B] > 0$ resp. $P[A] > 0$, so folgt aus der Unabhängigkeit von A und B, dass $P[A|B] = P[A]$ resp. $P[B|A] = P[B]$. Bei mehr als 2 involvierten Ereignissen ist Vorsicht am Platz. Beginnen wir zuerst mit der Definition: Die Ereignisse $A_1, \ldots, A_n$ sind in der Gesamtheit unabhängig voneinander, wenn für beliebige Auswahl $k = 1, \ldots, n$ und $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ gilt:*

$$P[A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}] = P[A_{i_1}]P[A_{i_2}] \ldots P[A_{i_k}].$$

Aus der paarweisen Unabhängigkeit folgt jedoch nicht zwingend die (gesamtheitliche) Unabhängigkeit! Entwickeln Sie aus der folgenden Ausgangslage ein Gegenbeispiel dazu: Sei $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Alle elementaren Ereignisse seien gleichwahrscheinlich (1/4). Untersuchen Sie die Ereignisse $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$ und $C = \{\omega_1, \omega_4\}$ unter diesem Gesichtspunkt.

1.1.7 Formel von der totalen Wahrscheinlichkeit FTW

Lemma 1.7 [Formel von der totalen Wahrscheinlichkeit FTW] *$B_1, B_2, \ldots$ sei eine Partition von Ω (die B_i's sind disjunkt und $\cup_{i=1}^{\infty} B_i = \Omega$). Weiter sei für alle $B_i, i \geq 1$, $P[B_i] > 0$ erfüllt. Dann gilt für jedes $A \in \mathcal{A}$:*

$$P[A] = \sum_{i=1}^{\infty} P[A|B_i]P[B_i]. \tag{FTW}$$

Wie man am Beweis unten sofort sieht, gilt ein analoges Resultat auch für eine endliche Partition.

Beweis von Lemma 1.7:

$\square$

1.2 Zufallsgrössen

Definition 1.8 [Zufallsgrösse (Zufallsvariable) X auf $(\Omega, \mathcal{A}, P)$] *Eine Zufallsgrösse auf* $(\Omega, \mathcal{A}, P)$ *ist eine Funktion* $X : \Omega \to \mathbb{R}$ *mit der Eigenschaft, dass* $\{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{A}$ *für alle reellen* x. *Die geforderte Eigenschaft nennt man Messbarkeit.*

Bevor wir uns der Verteilungs*funktion* zuwenden, wollen wir vom Publikum wissen, welche Verteilungen bereits bekannt sind (ausführliche Besprechung in 1.4).

1. Name
2. Wahrscheinlichkeitsfunktion (diskret) bzw. Dichten (stetig)
3. Bedeutung, Motivation, Modell für ...

diskret:

stetig:

1.2.1 Verteilungsfunktion

Definition 1.9 [Verteilungsfunktion F] *Die Verteilungsfunktion F einer Zufallsgrösse X ist definiert als*

$$F(x) := P[\{\omega | X(\omega) \leq x\}].$$

Man schreibt manchmal für die bessere Identifikation auch F_X statt nur F, falls man betonen will, dass die Verteilungsfunktion zu X gehört. Des weiteren schreiben wir von jetzt an auch:

$$P[X \leq x] := P[\{\omega | X(\omega) \leq x\}].$$

Beispiele von beiden letzten Seiten:

a)

b)

c) Versuchen Sie auch zu skizzieren: Uniform auf $[-1, 0.5]$

d) und $\mathcal{N}(2, 4)$

Welche Eigenschaften besitzen offenbar Verteilungsfunktionen?

Die letzten 4 Beispiele weisen auf einige wichtige Eigenschaften von Verteilungsfunktionen hin, welche wir in folgendem Lemma 1.10 zusammenfassen:

Lemma 1.10 [nützliche Eigenschaften von F] *Es gelten folgende Aussagen:*

a) $\lim_{x\to-\infty} F(x) = 0$ *und* $\lim_{x\to\infty} F(x) = 1$

b) $F(x)$ *ist monoton wachsend in* x *und rechtsstetig.*

1.2.2 Diskrete und stetige Zufallsgrösse

Definition 1.11 [Diskrete und stetige Zufallsgrösse] *Wenn eine Zufallsgrösse* X *nur Werte auf einer abzählbaren Teilmenge von* $\mathbb{R}$ *annehmen kann, nennen wir sie diskret. Wir nennen eine Zufallsgrösse* Y *stetig, wenn Ihre Verteilungsfunktion* F_Y *sich darstellen lässt als*

$$F_Y(x) = \int_{-\infty}^{x} f(u)du,$$

wobei f *eine nichtnegative, integrierbare Funktion auf* $\mathbb{R}$ *ist. Wir nennen* f *Dichte oder Dichtefunktion der Zufallsgrösse* Y *(eventuell mit* f_Y *bezeichnet). Salopp gilt: diskret ist auf einzelnen Punkten und stetig ist auf durchgezogenen Intervallen.*

Wir wollen im Folgenden 4 Eigenschaften von **stetigen Zufallsgrössen** herausarbeiten und am Beispiel der Normalverteilung veranschaulichen.

1. Für $a < b$ haben wir:

$$P[a < X \leq b] = P[X \leq b] - P[X \leq a] = F(b) - F(a) = \int_a^b f(x)dx.$$

2. Für stetige Zufallsgrössen gilt: $P[X = x_0] = \int_{x_0}^{x_0} f(x)dx = 0.$

3. Wegen 2. gilt insbesondere für $a < b$:

$$\begin{aligned} P[a \leq X \leq b] &= P[a < X \leq b] = P[a < X < b] = P[a \leq X < b] \\ &= P[X \leq b] - P[X \leq a] = F(b) - F(a) = \int_a^b f(x)dx. \end{aligned}$$

4. Wenn F differenzierbar ist (mit gesundem Menschenverstand auch in den meisten anderen Fällen) gilt:

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

1.2.3 Bivariate Verteilung

Definition 1.12 [gemeinsame Verteilungsfunktion von X und Y] *Seien X und Y zwei Zufallsgrössen auf dem gleichen Wahrscheinlichkeitsraum. Die gemeinsame Verteilungsfunktion wird dann definiert als*

$$F_{X,Y}(x, y) := P[\{\omega | X(\omega) \leq x\} \cap \{\omega | Y(\omega) \leq y\}];$$

kurz schreibt man hierfür auch $P[X \leq x, Y \leq y]$. Das Komma "," ist dabei als "und" zu lesen.

Turnübungen

1. Für *diskrete Zufallsgrössen* gilt dann wegen Eigenschaft c) von Definition 1.2:

$$F_{X,Y}(x, y) = \sum_{x_i \leq x} \sum_{y_i \leq y} P[X = x_i, Y = y_i];$$

Für *stetige Zufallsgrössen*:

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) dv du,$$

wobei eine bivariate Zufallsgrösse als stetig bezeichnet wird, wenn eine solche Darstellung mit Dichte $f_{X,Y}$ existiert.

2. Diskret gilt wegen Eigenschaft c) von Definition 1.2:

$$P[X = x] = \sum_{y_i} P[X = x, Y = y_i].$$

Stetig erhält man analog

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy.$$

3. Bei multivariaten Situationen der Dimension 3 und mehr verfährt man analog.

1.2.4 Unabhängigkeit von Zufallsgrössen

Die Unabhängigkeit von Zufallsgrössen führt man auf die Unabhängigkeit von Ereignissen zurück:

Definition 1.13 [Unabhängigkeit von Zufallsgrössen] *Zufallsgrössen auf einem gemeinsamen Wahrscheinlichkeitsraum $X_1, X_2, \dots, X_n$ sind (gesamtheitlich) unabhängig voneinander, wenn*

$$P[X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n] = P[X_1 \in B_1]P[X_2 \in B_2] \dots P[X_n \in B_n] \qquad (I)$$

für beliebige $B_1, \dots, B_n$ aus der Borel-σ-Algebra $\mathcal{B}(\mathbb{R})$.

Für den diskreten Fall mag dies unproblematisch sein, da die einzelnen Werte, welche die diskreten Zufallsgrössen annehmen können, sicher in der Borel-σ-Algebra $\mathcal{B}(\mathbb{R})$ enthalten sind. Im stetigen Fall ist (I) aber ein wenig umständlich. Gott sei Dank gibt es äquivalente Bedingungen unter Benutzung der Verteilungsfunktionen resp. Dichten: X und Y sind genau dann unabhängig voneinander, wenn für beliebige $a, b \in \mathbb{R}$ gilt (analog für mehr als 2 Zufallsgrössen):

$$F_{X,Y}(a, b) = F_X(a)F_Y(b), \qquad (II)$$

oder (wieder äquivalent):

$$f_{X,Y}(a, b) = f_X(a)f_Y(b). \qquad (III)$$

Jargon:

1. Das Wort "gesamtheitlich" lassen wir in Zukunft weg: Wenn nicht anders angegeben, ist immer die gesamtheitliche Unabhängigkeit gemeint (bei Ereignissen wie auch bei Zufallsgrössen).

2. Bei 2 Zufallsgrössen vereinbaren wir weiter das Symbol: $X \coprod Y$, wenn X und Y unabhängig sind.

3. unabhängig, identisch verteilt = independent and identically distributed (**iid**)

1.3 Erwartungswerte

1.3.1 Erwartungswert und Varianz einer diskreten und stetigen Zufallsgrösse

Definition 1.14 [Erwartungswert einer diskreten und stetigen Zufallsgrösse, auch Mittelwert genannt] *Der Erwartungswert $E[X]$ einer Zufallsgrösse X ist definiert als*

$$E[X] := \begin{cases} \sum_x xP[X=x] & \text{falls } X \text{ diskret} \\ \int_x xf(x)dx & \text{falls } X \text{ stetig,} \end{cases}$$

falls die Summe bzw. das Integral existiert. Dabei wird über den gesamten Wertebereich der Zufallsgrösse summiert respektive integriert.

Beispiele I

1. Berechnen Sie in der Stunde den Erwartungswert der Anzahl Augen beim Wurf eines perfekten Würfels. Überlegen Sie sich zuerst, was es geben sollte.

2. Berechnen Sie in der Stunde den Erwartungswert einer $U[-2,1]$-Zufallsgrösse. Überlegen Sie sich zuerst, was es geben sollte.

Definition 1.15 [Varianz/Standardabweichung einer diskreten und stetigen Zufallsgrösse] *Mit $\mu_X := E[X]$ definieren wir die Varianz $V[X]$ einer Zufallsgrösse X als*

$$V[X] := E[(X-\mu_X)^2] = \begin{cases} \sum_x (x-\mu_X)^2 P[X=x] & \text{falls } X \text{ diskret} \\ \int_x (x-\mu_X)^2 f(x)dx & \text{falls } X \text{ stetig.} \end{cases}$$

*Dabei wird auch hier über den gesamten Wertebereich der Zufallsgrösse summiert respektive integriert. Die Standardabweichung sd (***S***tandard* **D***eviation) ist die Wurzel aus der Varianz:*

$$sd[X] := \sqrt{V[X]}.$$

Beispiele II

3. Berechnen Sie die Varianz einer Be(p)-Zufallsgrösse, wo $p \in (0,1)$:

4. Berechnen Sie die Varianz einer $U[0,1]$-Zufallsgrösse:

5. Was vermuten Sie: wie wird die Varianz einer $U[0,3]$-Zufallsgrösse sein (Auflösung nach Lemma 1.18)?

Definition 1.16 [Erwartungswert bei 2 involvierten Zufallsgrössen] *Der Erwartungswert $E[g(X,Y)]$ (z.B. $g(x,y) = xy$) von 2 Zufallsgrössen X und Y auf dem gleichen Wahrscheinlichkeitsraum ist definiert als*

$$E[g(X,Y)] := \begin{cases} \sum_x \sum_y g(x,y) P[X=x, Y=y] & \text{falls } X, Y \text{ diskret} \\ \int_x \int_y g(x,y) f(x,y) dy dx & \text{falls } X, Y \text{ stetig.} \end{cases}$$

Dabei wird auch hier über den gesamten Wertebereich der Zufallsgrössen summiert respektive integriert. Analog verfährt man bei mehr als 2 Zufallsgrössen.

Hausaufgabe: rechnen Sie zu Definition 1.16 noch nachfolgendes diskretes Beispiel durch. Es ist deshalb sehr wichtig, weil sehr viel der bisherigen Theorie darin vorkommt!

Aufgabe: Sei $\Omega := \{\omega_1, \omega_2, \omega_3\}$. P sei derart, dass $P[\{\omega_1\}] = 0.5, P[\{\omega_2\}] = 0.3$ und $P[\{\omega_3\}] = 0.2$. $X(\omega_1) = 3, X(\omega_2) = 4, X(\omega_3) = 5; Y(\omega_1) = 2, Y(\omega_2) = 2, Y(\omega_3) = 7$. Gesucht ist nun $E[X^2Y + 1/Y]$.

Lösung: Wie ist g? $g(a, b) = a^2b + 1/b$.

Welche Werte nimmt (X, Y) an? $(3, 2), (4, 2), (5, 7)$.

Mit welchen Wahrscheinlichkeiten? $0.5, 0.3, 0.2$; d.h. z.B.
$P[X = 3, Y = 2] := P[\{\omega | X(\omega) = 3, Y(\omega) = 2\}] = P[\{\omega_1\}] = 0.5$.

In den Summen in Definition 1.16 kann man auch die Wertekombination $(3, 7)$ suchen und sogar berücksichtigen! Man kann sogar alle 3 mal 3 Kombinationen nehmen (ohne Unterscheidung, dass Y 2 mal den gleichen Wert annimmt). Jedoch kommen nur 3 Fälle mit positiver Wahrscheinlichkeit vor ($(3, 2), (4, 2), (5, 7)$), der Rest wird dann in unterer Summe mit 0 multipliziert.

Wir berechnen jetzt die Summe aus Definition 1.16, wobei wir nur noch die Kombinationen nehmen, welche positive Wahrscheinlichkeit haben:

$$E[X^2Y + 1/Y] = (18 + 0.5)0.5 + (32 + 0.5)0.3 + (175 + 1/7)0.2 \doteq 54.03.$$

1.3.2 Einige wichtige Rechenregeln

Lemma 1.17 [elementare Rechenregeln des Erwartungswerts] *Mit obiger Notation gelten (falls Summen und Integrale existieren):*
a) Sei $g(x)$ eine Funktion von $\mathbb{R}$ nach $\mathbb{R}$. Dann gilt:

$$E[g(X)] = \begin{cases} \sum_x g(x)P[X = x] & \text{falls } X \text{ diskret} \\ \int_x g(x)f(x)dx & \text{falls } X \text{ stetig.} \end{cases}$$

b) $|E[X]| \leq E[|X|]$.
c) Linearität des Erwartungswertes: $E[aX + bY] = aE[X] + bE[Y]$ ($\Rightarrow E[b] = b$ und $E[0] = 0$ (setze z.B. $a = 0, Y = 1$))

d) Im Allgemeinen gilt nicht $E[g(X)] = g(E[X])$; aber es gilt die sogenannte Jensen-Ungleichung für konvexe g: $E[g(X)] \geq g(E[X])$ (siehe z.B. Lemma 1.18 b) wo $g(x) = x^2$).
e) Falls X Werte auf $\{0, 1, 2, 3, \ldots\}$ annimmt, dann gilt:

$$E[X] = \sum_{n \geq 1} P[X \geq n].$$

Falls X eine stetige Zufallsgrösse auf $(0, \infty)$ mit Verteilungsfunktion $F(t)$ ist, dann gilt analog:

$$E[X] = \int_0^\infty (1 - F(t))dt = \int_0^\infty P[X \geq t]dt.$$

f) Mit I Indikatorfunktion (d.h. $I(A)(\omega) = 1$ gdw $\omega \in A$, 0 sonst, wo $A \in \mathcal{A}$) gilt: $E[I(A)] = P[I(A) = 1] = P[A]$.

Sei X eine stetige Zufallsgrösse auf $(0, \infty)$. Versuchen Sie mit Hilfe von Lemma 1.17 d) herauszufinden, ob in diesem Fall $E[1/X] \leq 1/E[X]$ oder $E[1/X] \geq 1/E[X]$ gilt.

Lemma 1.18 [elementare Rechenregeln der Varianz] *Mit obiger Notation gelten:*
a) $V[aX + b] = a^2 V[X]$ für a, b reelle Zahlen.
b) $V[X] = E[X^2] - (E[X])^2$

Bemerkungen zu Lemma 1.18: 1. Wegen Lemma 1.18 a) folgt mit $a \in \mathbb{R}$:

$$sd[aX] = |a| sd[X];$$

im Gegensatz zur Varianz kann man bei der Standardabweichung einen konstanten Faktor einfach herausnehmen (Absolutbetrag!).

2. Wir haben uns in Beispiel 5 gefragt, wie wohl die Varianz einer $U[0, 3]$-Zufallsgrösse sein muss. Sei X eine $U[0, 1]$-Zufallsgrösse. Man kann einfach via Verteilungsfunktion zeigen, dass dann $Y := 3X$ eine $U[0, 3]$-Zufallsgrösse ist. Deshalb muss die Varianz von Y wegen Lemma 1.18 a) 9 mal grösser sein als diejenige von X (also $9/12 = 3/4$). Überprüfen Sie, dass $3X$ eine $U[0, 3]$-Zufallsgrösse ist.

Lemma 1.19 [Unabhängigkeit von Zufallsgrössen und Erwartungswerte] *Mit obiger Notation gelten (falls Summen und Integrale existieren):*

a) Bei 2 Zufallsgrössen X und Y auf dem gleichen Wahrscheinlichkeitsraum gilt im Fall von $X \coprod Y$, dass

$$E[XY] = E[X]E[Y].$$

b) Sei $(X_i)_{i=1}^n$ eine Folge von unabhängigen Zufallsgrössen (die gleiche Verteilung wird nicht gefordert!). Dann gilt:

$$V[\sum_{i=1}^n X_i] = \sum_{i=1}^n V[X_i]$$

("Varianz der Summe ist die Summe der Varianzen").

Bemerkung zu Lemma 1.19: Wir haben in Beispiel 3 gesehen, dass die Varianz einer Be(p)-Zufallsgrösse gleich $p(1-p)$ ist. Die Bin(n, p)-Zufallsgrösse ist ja eine Summe von n Be(p)-Zufallsgrössen (sogar unabhängig!). Wegen Lemma 1.19 b) muss deshalb die Varianz einer Bin(n, p)-Zufallsgrösse gleich $np(1-p)$ sein.

1.3.3 Kovarianz und Korrelation

Wir haben in Lemma 1.19 a) gesehen, dass wenn 2 Zufallsgrössen X, Y unabhängig voneinander sind, dann gilt

$$E[XY] = E[X]E[Y].$$

Diese Eigenschaft ist zum Beispiel in der Finanzmathematik (Portfolio-Management, dort v.a. der Mean-Variance-Ansatz von Markovitz) so wichtig, dass man ihr einen eigenen Namen gegeben hat:

Definition 1.20 [unkorreliert, Kovarianz, Korrelationskoeffizient] *Seien* X, Y *zwei Zufallsgrössen auf dem gleichen Wahrscheinlichkeitsraum mit* $E[X^2] < \infty, E[Y^2] < \infty$.

a) Wir nennen X *und* Y *unkorreliert, wenn* $E[XY] = E[X]E[Y]$.

b) Den Ausdruck $Cov(X, Y) := E[(X - E[X])(Y - E[Y])]$ *nennen wir Kovarianz von* X *und* Y.

c) Falls $V[X] > 0, V[Y] > 0$, *so bezeichnen wir*

$$Cor(X, Y) := \frac{Cov(X, Y)}{\sqrt{V[X]V[Y]}}$$

als Korrelationskoeffizient von X *und* Y.

Rechenregeln:

a) $Cov(X, Y) := E[(X - E[X])(Y - E[Y])] = E[X(Y - E[Y])] = E[(X - E[X])Y] = E[XY] - E[X]E[Y]$

b) $Cov(aX + b, cY + d) = acCov(X, Y)$ für $a, b, c, d \in \mathbb{R}$

c) $|Cor(aX+b, cY+d)| = |Cor(X, Y)|$ für $a, b, c, d \in \mathbb{R}$ (*Skaleninvarianz* dank Normierung!)

d) Das folgende Gegenbeispiel zeigt, dass Unabhängigkeit von X und Y stärker ist als Unkorreliertheit. Sei $P[(X, Y) = (-1, 1)] = 1/3, P[(X, Y) = (0, -2)] = 1/3, P[(X, Y) = (1, 1)] = 1/3$. Zeigen Sie: $Cov(X, Y) = 0$; zeigen Sie mit Definition 1.13, dass X und Y nicht unabhängig voneinander sind.

Bei Unkorreliertheit gilt offenbar wegen Rechenregel a) $Cov(X, Y) = Cor(X, Y) = 0$. Wenn $Cor(X, Y) > 0$ sagen wir, X und Y seien positiv korreliert und wenn $Cor(X, Y) < 0$ sagen wir, X und Y seien negativ korreliert.

Im folgenden Lemma werden wir im Teil b) ein Resultat präsentieren, welches viel zum Verständnis der Korrelationen beiträgt.

Lemma 1.21 [Der Korrelationskoeffizient als Mass der *linearen* Abhängigkeit] *Seien X, Y zwei Zufallsgrössen auf dem gleichen Wahrscheinlichkeitsraum mit $E[X^2] < \infty, E[Y^2] < \infty$. Dann gelten:*

a) $|Cor(X, Y)| \leq 1$

b) $|Cor(X, Y)| = 1 \Leftrightarrow \exists\, a, b \in \mathbb{R} : X(\omega) = a + bY(\omega)\ \forall\, \omega \in \Omega \backslash N, P[N] = 0.$

Bemerkungen zu Lemma 1.21 1. In b) finden wir den Ausdruck "$\Omega \backslash N, P[N] = 0$". Dies ist rein technisch (fast sicher, bis auf eine Nullmenge N). Wenn wir eine (endliche) Stichprobe haben, ist es sogar ohne diesen Zusatz gültig.

2. Man kann es nicht oft genug betonen: es geht in b) nur um *lineare* Abhängigkeit zwischen X und Y. Wenn man Rechenregel d) zu Definition 1.20 anschaut, sieht man, dass dort die Korrelation 0 ist. Das heisst, dass es keinerlei *lineare* Abhängigkeit gibt. Aber Y ist 100 % von X abhängig!

3. Typische Bilder und deren Korrelationen:

1.4 Ausgewählte Verteilungen

* diskret: Bernoulli, Binomial, Geometrisch, Negativ-Binomial, Poisson
* stetig: Uniform, (Negativ-)Exponential, Gamma, Normal, Cauchy, χ^2, F, t
* Wertebereich, Verteilung (Wahrscheinlichkeitsfunktion / Dichte), E, V, R-Befehl, Einsatz

* Man beachte, dass sowohl bei diskreten wie auch bei stetigen Zufallsgrössen die Wahrscheinlichkeitsfunktionen respektive Dichten fast immer nach dem gleichen Schema aufgebaut sind:

Normierungskonstante mal informativer Teil.

1.4.1 Diskrete Verteilungen

1.4.1.1 Bernoulli Be(p)

X kann 2 Werte annehmen: 0 und 1 (alternativ auch -1 und $+1$). $P[X=1]=p$ (Erfolg) und $P[X=0]=1-p$ (Misserfolg). $E[X]=p$ und $V[X]=p(1-p)$. **R-Befehl:** binom mit $n=1$.

$$P[X=x]=p^x(1-p)^{1-x}.$$

1.4.1.2 Binomial Bin(n,p)

Seien X_i, $1 \leq i \leq n$, n iid Be(p)-Zufallsgrössen. Sei $Y := \sum_{i=1}^n X_i$. Dann hat Y per Definitionem die Binomialverteilung mit Parametern n und p; Bin(n,p). $E[Y]=np$ und $V[Y]=np(1-p)$ (Bemerkung zu Lemma 1.19), **R-Befehl:** binom. $0 \leq y \leq n$:

$$P[Y=y]=\binom{n}{y}p^y(1-p)^{n-y}.$$

Einsatz: Anzahl Erfolge bei n unabhängigen Versuchen mit Erfolgswahrscheinlichkeit p.

1.4.1.3 Geometrisch Ge(p)

Seien wieder X_i, $i \geq 1$, iid Be(p)-Zufallsgrössen. Wir interessieren uns für den Zeitpunkt des ersten Erfolgs: $Z := \min\{i|X_i = 1\}$. Z ist eine Zufallsgrösse auf den natürlichen Zahlen ohne die Null. Z hat per Definitionem eine geometrische Verteilung Ge(p) mit Parameter p. Es gilt: $E[Z] = 1/p$ und $V[Z] = (1-p)/p^2$, **R-Befehl:** geom.

$$P[Z = z] = p(1-p)^{z-1}.$$

Die Ge(p) hat die für die angewandte Stochastik zentral wichtige Eigenschaft, dass sie die einzige diskrete Zufallsgrösse ist, welche *gedächtnislos* ist: mit $n > m > 0$ gilt hier

$$P[Z > n|Z > m] = P[Z > (n-m)].$$

"Gegeben, es hat schon $m = 1000$ Würfe ohne 6 (=Erfolg) gegeben, wie gross ist die Wahrscheinlichkeit, dass es sogar insgesamt mindestens $n = 1004$ Würfe ohne 6 geben wird? Also dies ist nur noch von der Differenz $n - m = 4$ abhängig. Wie lange es bereits keine 6 gegeben hat, ist egal!"

Es sei noch erwähnt, dass die geometrische Verteilung in gewissen Lehrbüchern so definiert wird, dass man die Anzahl Misserfolge bis zum ersten Erfolg zählt. Dann nimmt die Ge(p) Werte auf den natürlichen Zahlen *inklusive* die 0 an. Die Resultate sind analog aber leicht komplizierter - das selbe gilt auch für eine alternative Definition der NB(n,p) nachfolgend.

1.4.1.4 Negativ Binomial NB(n,p)

Seien Z_i, $1 \leq i \leq n$, n iid Ge(p)-Zufallsgrössen. Sei $W := \sum_{i=1}^{n} Z_i$. Dann hat W per Definitionem die Negativ-Binomialverteilung mit Parametern n und p; NB(n,p). Die NB(n,p)-Verteilung beschreibt die Zeit, bis es n Erfolge gegeben hat. $E[W] = n/p$ (folgt aus Linearität des Erwartungswerts und 1.4.1.3) und $V[W] = n(1-p)/p^2$ (folgt aus Lemma 1.19 b) und 1.4.1.3), **R-Befehl:** nbinom. $w \geq n$

$$P[W = w] = \binom{w-1}{n-1} p^n (1-p)^{w-n}.$$

1.4.1.5 Poisson Po(λ)

Die Motivation für die Poissonverteilung folgt in der Vorlesung Angewandte Stochastik (FS 2013). Eine Zufallsgrösse V ist poissonsch, wenn Sie Werte auf den natürlichen Zahlen inklusive 0 annimmt und zwar mit folgenden Wahrscheinlichkeiten:

$$P[V = v] = e^{-\lambda} \frac{\lambda^v}{v!}.$$

Es gilt: $E[V] = V[V] = \lambda$, **R-Befehl:** pois.

1.4.2 Stetige Verteilungen

1.4.2.1 Uniform $U[a, b]$

Die einfachste stetige Verteilung ist die Uniform-Verteilung: Eine Zufallsgrösse U ist per Definitionem auf dem Intervall $[a, b]$ uniform verteilt, wenn U folgende Dichtefunktion hat:

$$f(u) = (b - a)^{-1},$$

wobei dann natürlich $a \leq u \leq b$ zu gelten hat. Ausserhalb von $[a, b]$ ist die Dichte gleich null. Es gilt $E[U] = (a + b)/2$ (2. Beispiel von 1.3.1) und $V[U] = (b - a)^2/12$ (Beispiel 4 in 1.3.1), **R-Befehl:** unif.

1.4.2.2 (Negativ-) Exponential Exp(λ)

Eine Zufallsgrösse X mit Dichtefunktion

$$f(x) = \lambda e^{-\lambda x}, \;\; x \geq 0,$$

heisst exponentialverteilt mit Parameter λ; Exp(λ). Die gleiche Wahl des Parameters wie bei der Poissonverteilung ist *nicht* zufällig! Es gilt $E[X] = 1/\lambda$ und $V[X] = 1/\lambda^2$, **R-Befehl:** exp. Modell für: radioaktiver Zerfall, "wann geht eine Glühbirne kaputt?", Zwischenzeit bei der Ankunft von KundInnen in einem Geschäft und vieles mehr; mehr dazu in der Vorlesung "Angewandte Stochastik".

Die Exp(λ) hat die für die angewandte Stochastik zentral wichtige Eigenschaft, dass sie die einzige stetige Zufallsgrösse ist, welche *gedächtnislos* ist: mit $t > s > 0$ gilt hier

$$P[X > t | X > s] = P[X > (t - s)].$$

"Gegeben, es hat schon $s = 1000$ Sekunden keinen Atomzerfall gegeben, wie gross ist die Wahrscheinlichkeit, dass es sogar insgesamt mindestens $t = 1004$ Sekunden keinen Atomzerfall geben wird? Also dies ist nur noch von der Differenz $t - s = 4$ Sekunden abhängig. Wie lange es bereits keinen Atomzerfall gegeben hat, ist egal!"

1.4.2.3 Gamma(n, λ)

Seien X_i, $1 \leq i \leq n$, n iid Exp(λ)-Zufallsgrössen. Sei $Y := \sum_{i=1}^n X_i$. Dann hat Y per Definitionem die Gammaverteilung mit Parametern n und λ; Gamma(n,λ). $E[Y] = n/\lambda$ (folgt aus Linearität des Erwartungswerts und 1.4.2.2) und $V[Y] = n/\lambda^2$ (folgt aus Lemma 1.19 b) und 1.4.2.2), **R-Befehl:** gamma.

$$f(y) = \frac{y^{n-1}e^{-\lambda y}\lambda^n}{\Gamma(n)},\ y \geq 0.$$

Mit $n = 1$ wird dies in der Tat zu einer Exponentialverteilung. In Anbetracht von 1.4.2.2 ist das ein Modell für "wann zerfällt das n-te Atom?", geht die n-te Glühbirne kaputt, kommt der n-te Kunde. In der Statistik hat man mit der Gamma-Verteilung auch ein relativ flexibles Modell, um Einkommensverteilungen zu modellieren (2 frei wählbare Parameter n, λ).

1.4.2.4 Normal $\mathcal{N}(\mu, \sigma^2)$

Wegen des zentralen Grenzwertsatzes ist die Normalverteilung sehr wichtig: Mit Dichtefunktion

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in \mathbb{R},$$

gilt $E[X] = \mu$ und $V[X] = \sigma^2$, **R-Befehl:** norm. Ist X eine $\mathcal{N}(\mu, \sigma^2)$-Zufallsgrösse, so ist

$$\frac{X - \mu}{\sigma}$$

eine $\mathcal{N}(0, 1)$-Zufallsgrösse.

1.4.2.5 Cauchy

Die Cauchy-Verteilung ist gleichzeitig die t_1-Verteilung (siehe 1.4.2.8 weiter unten). Neben praktischen Anwendungen ist sie für die Theorie der Statistik und Wahrscheinlichkeitsrechnung gut als Verteilung mit viel Gewicht in den Enden (Langschwänzigkeit, heavy tails) im Vergleich zur Normalverteilung. Es gilt insbesondere $E[|X|] = \infty$, wie man aus unterer Dichte sofort sieht:

$$f(x) = \frac{d}{\pi(d^2 + (x-m)^2)}, \; x \in \mathbb{R}.$$

Dabei ist m der Median und d ein Skalenparameter. **R-Befehl:** cauchy.

1.4.2.6 χ^2

Diese Verteilung ist in der Statistik zentral wichtig und verdankt ihre Existenz weitgehend dem zentralen Grenzwertsatz, insbesondere, dass man in Modellen der Datenanalyse Fehlerterme normalverteilt modelliert. Wenn $(X_i)_{i=1}^n$ iid $\mathcal{N}(0,1)$-verteilt sind, dann ist

$$\sum_{i=1}^n X_i^2$$

χ_n^2-verteilt (sprich "Chiquadrat mit n Freiheitsgraden" (degree of freedom, df)). Wenn $(Y_i)_{i=1}^n$ iid $\mathcal{N}(\mu, \sigma^2)$-verteilt sind, dann hat wegen der Z-Transformation

$$\sum_{i=1}^n \frac{(Y_i - \mu)^2}{\sigma^2}$$

die χ_n^2-Verteilung. Ein wenig überraschend: Mit

$$S^2 := \sum_{i=1}^n (Y_i - \overline{Y})^2,$$

wo $\overline{Y} := \sum_{i=1}^n Y_i/n$, hat S^2/σ^2 die χ_{n-1}^2-Verteilung. Eine Bauernregel sagt: ein df geht verloren, pro Parameter, den man schätzt (hier μ wird mit $\overline{Y}$ geschätzt). Die Dichte der χ_n^2-Verteilung ist

$$f(x) := \frac{x^{n/2-1} e^{-x/2}}{\Gamma(n/2) 2^{n/2}}, \; x \geq 0.$$

$E[\chi^2_n] = n; V[\chi^2_n] = 2n$. Wenn man die χ^2_n-Verteilung mit der Gamma-Verteilung vergleicht, sieht man, dass jede zweite χ^2_n-Verteilung eine Gamma-Verteilung mit Parameter $\lambda = 0.5$ ist, wenn man n erhöht. **R-Befehl:** chisq

1.4.2.7 F

Seien U und V zwei unabhängige, χ^2_m-, bzw. χ^2_n-verteilte Zufallsgrössen. Dann ist der Ausdruck

$$W := \frac{U/m}{V/n}$$

F-Verteilt mit Parametern m, n: $F_{m,n}$. Diese Zufallsgrössen kommen in der Statistik vor, wenn man testen will, ob die Varianzen von 2 unabhängigen Stichproben gleich sind oder nicht (in viel komplexeren statistischen Untersuchungen kommt $F_{m,n}$ auch (relativ unerwartet) vor).

$$f(w) = \frac{\Gamma[(m+n)/2](m/n)^{m/2}}{\Gamma(m/2)\Gamma(n/2)} \frac{w^{m/2-1}}{(1+mw/n)^{(m+n)/2}},\ w \geq 0,$$

ist die Dichtefunktion von W. Es gilt $E[W] = n/(n-2)$ (falls $n > 2$), $V[W] = [2n^2(m+n-2)]/[m(n-2)^2(n-4)]$. **R-Befehl:** f.

1.4.2.8 t

Y habe eine $\mathcal{N}(0,1)$-Verteilung. Z habe eine χ^2_n-Verteilung. $Y \coprod Z$. Dann hat

$$T_n := \frac{Y}{\sqrt{Z/n}}$$

eine Student-t-Verteilung mit n Freiheitsgraden oder kurz t_n-Verteilung. Die Dichte der t_n-Verteilung ist

$$f(x) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\Gamma(n/2)} \frac{1}{(1+x^2/n)^{(n+1)/2}},\ x \in \mathbb{R}.$$

Es gilt $E[t_n] = 0$ (falls $n > 1$ und falls $n = 1$ existiert der Erwartungswert nicht), $V[t_n] = n/(n-2)$ sobald $n > 2$. **R-Befehl:** t. Man sieht sofort, dass "$\sqrt{F_{1,n}} = |t_n|$". Des weiteren: sind X_1, X_2 zwei iid $\mathcal{N}(0,1)$-Zufallgrössen, so ist X_1/X_2 ebenfalls t_1- (und damit Cauchy-) verteilt.

1.4.3 Summen von Binomial, Geometrisch, Poisson, Exponential, Normal

Die Summanden in den folgenden Summen müssen immer unabhängig sein. Wenn nicht anders erwähnt, fordern wir auch die gleiche Verteilung. Mit kompakten Formeln wie

$$\sum_1^n Be(p) = Bin(n,p)$$

weiter unten ist eigentlich gemeint: Seien $X_1, \ldots, X_n$ iid Be(p)-verteilt. Dann hat $Y := \sum_1^n X_i$ eine Bin(n,p)-Verteilung.

* Summe von Bernoulli ist Binomial

$$\sum_1^n Be(p) = Bin(n,p)$$

* Grosser Bruder hiervon: Summe von Binomial ist Binomial (p muss immer gleich sein!)

$$\sum_1^n Bin(n_i,p) = Bin(\sum_1^n n_i, p)$$

* Summe von Geometrisch ist Negativ Binomial

$$\sum_1^n Ge(p) = NB(n,p)$$

* Summe von Poisson ist Poisson (Summanden müssen nicht mal identisch verteilt sein)

$$\sum_{i=1}^n Po(\lambda_i) = Po(\sum_{i=1}^n \lambda_i)$$

* Summe von Exponential ist Gamma

$$\sum_1^n Exp(\lambda) = Gamma(n,\lambda)$$

* Summe von Normal ist Normal (Summanden müssen nicht mal identisch verteilt sein)

$$\sum_{i=1}^n \mathcal{N}(\mu_i, \sigma_i^2) = \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$$

* Weitere Summen (z.B. von X^2) kommen bei den Definitionen von χ^2, F und t-Verteilungen vor.

1.5 $n \to \infty$ (Konvergenz, LLN, CLT)

1.5.1 Ungleichung von Bienayme-Tschebyschew

Wir beginnen mit einer wichtigen Ungleichung, welche auch ohne "$n \to \infty$" grosse Bedeutung hat.

Satz 1.22 [Ungleichung von Bienayme-Tschebyschew] *Sei X eine Zufallsgrösse mit $E[|X|] < \infty$. Dann gilt für $\epsilon > 0$:*

$$P[|X| \geq \epsilon] \leq \frac{1}{\epsilon} E[|X|].$$

Der praktische Vorteil der Ungleichung von Bienayme-Tschebyschew ist der, dass mit Erwartungswerten (Linearität) einfacher gerechnet werden kann als mit Wahrscheinlichkeiten.

Obwohl in der Ungleichung von Bienayme-Tschebyschew ein $\epsilon > 0$ vorkommt, muss darauf hingewiesen werden, dass dieses ϵ in den Anwendungen nicht umgedingt "ziemlich klein" sein wird. Im Gegenteil wird man mit dieser Ungleichung eventuell die Variabilität einer Zufallsgrösse unter Kontrolle haben wollen. Dazu müssen wir diese Ungleichung ein wenig umformen: X ist ziemlich allgemein gewählt. Es gilt sicher

$$P[|X - \mu| \geq \epsilon] \leq \frac{1}{\epsilon} E[|X - \mu|],$$

wo μ beliebig und konstant. Also auch $\mu := E[X]$. Weiter gilt auch (falls $E[X^2] < \infty$)

$$P[|X - \mu|^2 \geq \epsilon^2] \leq \frac{1}{\epsilon^2} E[|X - \mu|^2],$$

oder analog

$$P[|X - \mu| \geq \epsilon] \leq \frac{1}{\epsilon^2} V[X].$$

Damit haben wir aber folgendes Resultat: Wir haben eine Abschätzung für die Wahrscheinlichkeit erhalten, dass die Zufallsgrösse X Werte annimmt, die mehr als ϵ vom Erwartungswert entfernt sind. Dieses Resultat ist zudem *verteilungsfrei* hergeleitet worden und gilt somit für alle Zufallsgrössen mit endlicher Varianz!

1.5.2 LLN (Law of Large Numbers; Gesetz der grossen Zahlen)

Fragestellung: Was geschieht mit dem Ausdruck

$$\frac{1}{n}\sum_{k=1}^{n} X_k$$

falls $n \to \infty$ und welche Voraussetzungen werden wir sinnvollerweise machen?

Das Gesetz der grossen Zahlen ist streng genommen kein Theorem, sondern eine *Eigenschaft* einer Folge (das Theorem folgt dann in Theorem 1.24).

Definition 1.23 [Gesetz der grossen Zahlen] *Sei $X_i, i \geq 1$, eine Folge von identisch verteilten Zufallsgrössen mit Erwartungswert μ, sodass $E[|X_1|] < \infty$. Wir sagen, dass die Folge von Zufallsgrössen $X_i, i \geq 1$, dem Gesetz der grossen Zahlen genügt, wenn für jedes $\epsilon > 0$ gilt:*

$$\lim_{n\to\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right] = 0.$$

Bemerkungen zu Definition 1.23: 1. Diese Definition ist eigentlich ein Spezialfall des Gesetzes der grossen Zahlen. Es ist nur das sogenannte *schwache* Gesetz der grossen Zahlen im Fall von identisch verteilten Zufallsgrössen und die Konvergenz ist bei uns ausschliesslich gegen eine konstante Zahl (den Erwartungswert). Dies reicht jedoch für diese Vorlesung.

2. Anschaulich sagt die Definition, dass das arithmetische Mittel immer näher und näher zum Erwartungswert der Zufallsgrösse kommt.

Wir wissen noch nicht, wann eine Folge diesem Gesetz der grossen Zahlen genügt. Ein schönes Resultat dazu ist Theorem 1.24, welches *nicht die Existenz der Varianz fordert.*

Theorem 1.24 [Satz von Kolmogoroff] *Sei $X_i, i \geq 1$, eine Folge von paarweise unabhängigen, identisch verteilten Zufallsgrössen mit Erwartungswert μ und $E[|X_1|] < \infty$. Dann genügt diese Folge dem Gesetz der grossen Zahlen; es gilt also für jedes $\epsilon > 0$:*

$$\lim_{n\to\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right] = 0.$$

1.5.3 CLT (Central Limit Theorem; Zentraler Grenzwertsatz)

Fragestellung: Was geschieht mit dem Ausdruck

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}, \tag{1.1}$$

einfacherer Fall ($\mu = 0$, $\sigma = 1$)

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k,$$

falls $n \to \infty$ und welche Voraussetzungen werden wir sinnvollerweise machen?

Der zentrale Grenzwertsatz (englisch Central Limit Theorem CLT) unterstreicht die grosse Bedeutung der Normalverteilung:

Theorem 1.25 [Zentraler Grenzwertsatz] *Sei $X_k, k \geq 1$, eine Folge von iid Zufallsgrössen mit $E[X_1] =: \mu$ und $0 < V[X_1] =: \sigma^2 < \infty$. Dann gilt für $a \in \mathbb{R}$ beliebig:*

$$P\left[\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leq a\right] \longrightarrow P[\mathcal{N}(0,1) \leq a]$$

wenn $n \to \infty$.

Mit "$P[\mathcal{N}(0,1) \leq a]$" meinen wir die Wahrscheinlichkeit, dass eine Standard-Normal-verteilte Zufallsgrösse Werte kleiner oder gleich a annimmt.
Die Verteilungsfunktion der zentrierten und normierten Summe (1.1) konvergiert also in jedem Punkt gegen die Verteilungsfunktion der Standard-Normal-Verteilung.

Bemerkungen zu Theorem 1.25: 1. Wegen Lemma 1.17 c) und Lemma 1.18 a) ist es *nicht überraschend*, dass die Normalverteilung (die Limesverteilung) Erwartungswert 0 und Varianz 1 hat: wir ziehen ja den Erwartungswert $n\mu$ ab (Zentrierung) und normieren mit der Standardabweichung $\sqrt{n}\sigma$. Das *Überraschende* am CLT ist, dass mit wenigen Voraussetzungen (v.a. keine über die Verteilung der X_k's) immer eine Normalverteilung als Limesverteilung resultiert.

2. Vorsicht: Man darf die Zentrierung und Normierung nicht "auf die andere Seite" nehmen in der Form: "$\sum_{k=1}^{n} X_k$ konvergiert mit $n \to \infty$ in Verteilung gegen eine Zufallsgrösse mit

Verteilung $\mathcal{N}(n\mu, n\sigma^2)$". Dies ist ja nicht stabil, das darf man in der Analysis bei der Konvergenz von Folgen ja auch nicht. In der Statistik wird dies jedoch als approximative Methode benutzt.