Angewandte Stochastik

Dr. C.J. Luchsinger

1 Repetition Wahrscheinlichkeitstheorie (WT)

Bitte arbeiten Sie dieses Kapitel wenn notwendig *vor* Semesterbeginn durch. Es ist eine Zusammenfassung der aktuellsten Version von Kapitel 1-5 aus der Vorlesung "Einführung in die Wahrscheinlichkeitstheorie und Statistik". Die Bezeichnungen und Nummerierungen für Kapitel, Sektionen und Aussagen ist gleich belassen, um die Orientierung zu erleichtern; hingegen steht immer ein "WTS" (für **W**ahrscheinlichkeits**t**heorie und **S**tatistik) vor jedem Satz/Kapitel etc.

WTS-1 Wahrscheinlichkeit

Literatur WTS-Kapitel 1

* Cartoon Guide: Kapitel 3
* Krengel: § 1 und 2

WTS-1.4 Das Konzept der Wahrscheinlichkeit P

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.

WTS-1.4.1 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:

  * abzählbar unendliches Ω: zufällige Zahl aus $\Omega := \mathbb{N}$

  * überabzählbar unendliches Ω: Zeit bis Atom zerfällt ist im Intervall $\Omega := (0, \infty)$ und zufällig, zumindest in gewissen physikalischen Modellen. Unendlicher Münzwurf: $\Omega := \{K, Z\}^\infty$.

* 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
$A \backslash B$	A ohne B / ein Elementarereignis aus A tritt ein, aber nicht aus B
$A \subset B$	A ist Teilmenge von B / Wenn ein Elementarereignis aus A stattfindet, dann immer auch ein Elementarereignis aus B
$\emptyset$	leere Menge / unmögliches Ereignis
Ω	ganze Menge / sicheres Ereignis (etwas muss passieren)

WTS-1.4.2 σ-Algebra

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 WTS-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!

WTS-1.4.3 Wahrscheinlichkeit $P[.]$

WTS-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 WTS-Definition 1.2 c) z.B. auch $A_i = \emptyset, i \geq 3$ wählen!

Es gibt eine extrem praktische Interpretation von **Wahrscheinlichkeiten und Proportionen**: Wir haben 7.5 Millionen EinwohnerInnen in der Schweiz. Wenn man jetzt zufällig (jedeR EinwohnerIn hat gleiche Wahrscheinlichkeit, ausgewählt zu werden, also (1/7'500'000)) eineN EinwohnerIn herausgreift, so ist beispielsweise die Wahrscheinlichkeit, einen Basler herauszugreifen, genau gleich der Proportion der Basler an der Schweizerischen Bevölkerung. Wir werden diese Analogie in vielen Fällen einsetzen.

Aus WTS-Definition 1.2 lassen sich nützliche Eigenschaften ableiten, welche wir im folgenden Lemma zusammenfassen.

WTS-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. $P[\emptyset] = 0$

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

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

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

WTS-1.4.4 Wahrscheinlichkeitsraum

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

WTS-1.4.5 Bedingte Wahrscheinlichkeit $P[A|B]$

WTS-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.

WTS-1.4.6 Unabhängigkeit von Ereignissen

Das folgende Konzept sollte aus der Mittelschule (Gymnasium) bereits von den Wahrscheinlichkeitsbäumen / Ereignisbäumen her bekannt sein. Anschaulich und in der Modellierung bedeutet unabhängig, dass zwei Ereignisse von unabhängigen Mechanismen erzeugt wurden.

WTS-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 ist Unabhängigkeit von A und B gleichbedeutend mit*

$$P[A|B] = P[A] \qquad \text{("who cares" – Variante)}$$

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.

WTS-1.4.7 Formel von Bayes

Angenommen wir kennen $P[B|A]$, wollen aber eigentlich $P[A|B]$ wissen.

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[B|A]P[A]}{P[B|A]P[A] + P[B|A^c]P[A^c]}.$$

WTS-1.4.8 Formel von der totalen Wahrscheinlichkeit FTW

WTS-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 $A \in \mathcal{A}$:*

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

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

Wir fügen noch zwei miteinander eng verwandte Formeln hinzu:

WTS-Lemma 1.8 [$\lim P[A_n]$ bei ab- oder aufsteigenden Folgen] *Sei $A_1, A_2, \ldots$ eine aufsteigende Folge von Ereignissen, d.h. $A_1 \subseteq A_2 \subseteq A_3 \ldots$. Wir definieren in dem Fall*

$$A := \lim_{i \to \infty} A_i := \cup_{i \geq 1} A_i.$$

Dann gilt $P[A] = \lim_{i \to \infty} P[A_i]$.

Analog gilt: Sei $B_1, B_2, \ldots$ eine absteigende Folge von Ereignissen, d.h. $B_1 \supseteq B_2 \supseteq B_3 \ldots$. Wir definieren in dem Fall

$$B := \lim_{i \to \infty} B_i := \cap_{i \geq 1} B_i.$$

Dann gilt $P[B] = \lim_{i \to \infty} P[B_i]$.

WTS-2 Zufallsgrössen

Literatur WTS-Kapitel 2

* Cartoon Guide: Kapitel 4

* Krengel: 3.1 und 3.2 in § 3 und § 10 sowie 11.1, 11.2 und 11.3 in § 11.

WTS-2.1 Zufallsgrösse

WTS-Definition 2.1 [Zufallsgrösse 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 $\{\omega \in \Omega | X(\omega) \leq a\} \in \mathcal{A}$ für alle reellen a. Die geforderte Eigenschaft nennt man Messbarkeit.*

WTS-2.1 - Zugabe: Wir schmeissen die gesamte Wahrscheinlichkeit auf die reelle Achse

Es gilt bekanntlich $P[\Omega] = 1$. Durch eine Zufallsgrösse X auf $(\Omega, \mathcal{A}, P)$ wird diese ganze Wahrscheinlichkeit 1 auf die reelle Achse $\mathbb{R}$ "hinübergeworfen". Auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ können wir wieder einen Wahrscheinlichkeitsraum (vgl. WTS-1.4.4) definieren: Sei A eine beliebige Teilmenge von $\mathcal{B}(\mathbb{R})$. Wir definieren:

$$P_X[A] := P[\{\omega \in \Omega | X(\omega) \in A\}].$$

Wir erhalten also einen neuen Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$, welcher von X erzeugt wurde. $X : \Omega \to \mathbb{R}$ muss nicht surjektiv sein; das führt zu keinen Problemen! Man nennt P_X auch die Verteilung oder das Verteilungsgesetz der Zufallsgrösse X; häufig schreibt man einfach P statt P_X.

WTS-2.2 Verteilungsfunktion

WTS-Definition 2.2 [Verteilungsfunktion F] *Die Verteilungsfunktion F einer Zufallsgrösse X ist definiert als*

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

Man schreibt manchmal für die bessere Identifikation auch F_X statt nur F, falls man betonen will, dass die Verteilungsfunktion zu X gehört.

WTS-Lemma 2.3 [nützliche Eigenschaften von F] *Es gelten folgende Aussagen:*

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

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

WTS-2.3 Diskrete und stetige Zufallsgrösse

WTS-Definition 2.4 [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*

$$F_Y(a) := P[Y \leq a] = \int_{-\infty}^{a} f(u)du,$$

wobei f eine nichtnegative, integrierbare Funktion auf $\mathbb{R}$ ist. Diese Funktion f erfüllt

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

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:

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

WTS-2.4 Bivariate Verteilung

WTS-Definition 2.5 [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 wegen Eigenschaft c) von WTS-Definition 1.2:

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

Für *stetige Zufallsgrössen*:

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

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

2. X und Y seien diskrete Zufallsgrössen mit Werten $(x_i)_i$, $(y_j)_j$. Dann haben wir für beliebiges x_i:

$$P[X = x_i] = \sum_j P[X = x_i, Y = y_j]$$

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.

WTS-2.5 Unabhängigkeit von Zufallsgrössen

Anschaulich bedeutet für uns "X unabhängig von Y" einfach, dass X und Y von unterschiedlichen Mechanismen erzeugt wurden.

Mathematisch exakt führt man die Unabhängigkeit von Zufallsgrössen auf die Unabhängigkeit von Ereignissen zurück:

WTS-Definition 2.6 [Unabhängigkeit von Zufallsgrössen] *Zufallsgrössen auf einem gemeinsamen Wahrscheinlichkeitsraum* $X_1, X_2, \ldots, X_n$ *sind (gesamtheitlich) unabhängig voneinander, wenn*

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

für beliebige $B_1, \ldots, B_n$ *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**)

WTS-3 Erwartungswerte

Literatur WTS-Kapitel 3

* Cartoon Guide: Kapitel 4

* Krengel: 3.3 und 3.5 in § 3 und 11.4 in § 11

WTS-3.1 Erwartungswert und Varianz einer diskreten und stetigen Zufallsgrösse

WTS-Definition 3.1 [Erwartungswert einer diskreten und stetigen Zufallsgrösse] *Der Erwartungswert $E[X]$ einer Zufallsgrösse X ist definiert als*

$$E[X] := \begin{cases} \sum_{x_i} x_i P[X = x_i] & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{falls } X \text{ stetig.} \end{cases}$$

Sei $g(x)$ eine Funktion von $\mathbb{R}$ nach $\mathbb{R}$. Dann definieren wir:

$$E[g(X)] = \begin{cases} \sum_{x_i} g(x_i) P[X = x_i] & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{falls } X \text{ stetig.} \end{cases}$$

Diese Definitionen gelten, falls die Summe bzw. das Integral der Absolutwerte existiert. Dabei wird jeweils über den gesamten Wertebereich der Zufallsgrösse summiert respektive integriert.

WTS-Definition 3.2 [Varianz/Standardabweichung einer diskreten und stetigen Zufallsgrösse] *Mit $\mu_X := E[X]$ definieren wir die Varianz $V[X]$ einer Zufallsgrösse X als*

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

*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:*

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

WTS-Definition 3.3 [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*

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

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

WTS-3.2 Einige wichtige Rechenregeln

WTS-Lemma 3.4 [Absolutbetrag "vorher" und "nachher", Linearität, Indikatorfunktion] *Mit obiger Notation gelten (falls Summen und Integrale existieren):*

a) $|E[X]| \leq E[|X|]$.

b) Linearität des Erwartungswertes:

$$E[aX + bY] = aE[X] + bE[Y]$$

und damit unter anderem $E[b] = b$ *und* $E[0] = 0$ *(setze z.B.* $a = 0, Y = 1$*). Umgangssprachlich nennt man Teil b): "Konstanten herausziehen" und "Summen auseinanderziehen".*

c) Mit I Indikatorfunktion (d.h. $I_A(\omega) = 1$ *gdw* $\omega \in A$*, 0 sonst, wo* $A \in \mathcal{A}$*) gilt:*

$$E[I_A] = P[I_A = 1] = P[A].$$

Im Allgemeinen gilt nicht

$$E[g(X)] = g(E[X]);$$

aber es gilt die sogenannte Jensen-Ungleichung für konvexe g:

WTS-Lemma 3.5 [Jensen-Ungleichung] *Für konvexe g gilt:* $E[g(X)] \geq g(E[X])$.

WTS-Lemma 3.6 [Alternative Berechnung des Erwartungswerts] *Mit obiger Notation gelten (falls Summen und Integrale existieren):*
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:

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

WTS-Lemma 3.7 [elementare Rechenregeln der Varianz] *Mit obiger Notation gelten:*
a) $V[aX + b] = a^2 V[X]$ für a, b reelle Zahlen ("Konstante quadratisch raus").
b) $V[X] = E[X^2] - (E[X])^2$

WTS-Lemma 3.8 [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

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

("Produkte auseinandernehmen").

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").

WTS-3.4 Kovarianz und Korrelation

WTS-Definition 3.9 [Kovarianz, Korrelationskoeffizient] *Seien X, Y zwei Zufallsgrössen auf dem gleichen Wahrscheinlichkeitsraum mit $E[X^2] < \infty, E[Y^2] < \infty$.*

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

b) 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.

Es gilt:

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) $Cov(X, X) = V[X]$

e) Wenn $Cov(X, Y) = Cor(X, Y) = 0$, so nennen wir 2 Zufallsgrössen unkorreliert. 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.

f) Bei $Cov(X, Y) = 0$ gilt offenbar wegen a) auch $E[XY] = E[X]E[Y]$. Wir haben in WTS-Lemma 3.8 a) gesehen, dass wenn 2 Zufallsgrössen X, Y unabhängig voneinander sind, dann gilt auch $E[XY] = E[X]E[Y]$. Offenbar folgt also aus Unabhängigkeit Unkorreliertheit:

$$\text{unabhängig} \Rightarrow \text{unkorreliert}$$

g) 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 WTS-Definition 2.6, dass X und Y nicht unabhängig voneinander sind.

h) Wenn wir den Beweis von WTS-Lemma 3.8 b) nochmals durchgehen, sehen wir, dass für die Gültigkeit der Formel

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

die Unabhängigkeit der Zufallsgrössen nicht *zwingend* ist: Dank Bemerkung a) haben wir mit $Cov(X,Y)=0$ auch $E[XY]=E[X]E[Y]$ und damit *reicht bereits die Unkorreliertheit!*

WTS-Lemma 3.10 [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 WTS-Lemma 3.10 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.

WTS-3.5 Bedingte Verteilungen und Definition des bedingten Erwartungswertes gegeben $Y = y_0$

Um Verwechslungen zu vermeiden sei vorausgeschickt, dass eine allgemeine Definition bedingter Erwartungswerte sehr aufwendig ist. Wir bedingen hier lediglich auf ein *Ereignis (gegeben $Y = y_0$)* - wir bedingen hier nicht auf eine Zufallsgrösse ($E[X|Y]$, mehr dazu in einer allfälligen Vorlesung über Mass- und Wahrscheinlichkeitstheorie). Zudem behandeln wir hier nur die diskreten Zufallsgrössen und harmlose stetige Zufallsgrössen.

Wir müssen als Vorbereitung auf die bedingten Erwartungswerte noch bedingte Verteilungen einführen: Für bedingte Wahrscheinlichkeiten gilt

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

Wir können dann im Fall **diskreter Zufallsgrössen** folgendermassen fortfahren: Mit $A := \{X = x\}, B := \{Y = y_0\}$ ist $P[A|B]$ die Wahrscheinlichkeit, dass $X = x$ ist, wenn $Y = y_0$ ist (falls $X \coprod Y$, ist dies einfach $P[X = x]$):

$$P[X = x|Y = y_0] := \frac{P[\{X = x\} \cap \{Y = y_0\}]}{P[\{Y = y_0\}]} := \frac{P[\{\omega|X(\omega) = x\} \cap \{\omega|Y(\omega) = y_0\}]}{P[\{\omega|Y(\omega) = y_0\}]}.$$

Wir können jetzt noch x variieren und *erhalten damit die Verteilung von ganz* $X|Y = y_0$ (zur Erinnerung an WTS-1.4.5: bedingte Wahrscheinlichkeiten sind auch Wahrscheinlichkeiten).

Die Schwierigkeiten liegen im **stetigen Fall**: bei stetigen Zufallsgrössen X, Y gilt ja $P[X = x] = P[Y = y_0] = 0$, wenn x, y_0 konkrete, feste, reelle Zahlen sind. Damit hätten wir oben $0/0$, was nicht definiert ist. Da sich Wahrscheinlichkeitsfunktionen und Dichtefunktionen von der Bedeutung her entsprechen, ist man trotzdem versucht, bedingte Dichtefunktionen folgendermassen zu *definieren* (und in vielen Lehrbüchern geschieht dies auch diskussionslos):

$$f_{X|Y=y_0}(x) := \frac{f_{X,Y}(x, y_0)}{f_Y(y_0)}, \qquad (WTS - 3.2)$$

falls $f_Y(y_0) > 0$. (WTS-3.2) können wir als Definition stehen lassen (sie ist sinnvoll), wollen aber trotzdem mit einem Limesresultat noch begründen, dass dies die Verteilung von $X|Y = y_0$ angibt. Diese Begründung ist nicht Prüfungsstoff sondern Honours-Programm.

Wir repetieren zuerst WTS-Definition 2.4 und betrachten dann den Ausdruck

$$P[X \leq a|Y \in [y_0 - h, y_0 + h]] = \frac{P[X \leq a, Y \in [y_0 - h, y_0 + h]]}{P[Y \in [y_0 - h, y_0 + h]]}.$$

Jetzt sind Zähler und Nenner nicht mehr 0 - gehen aber in unteren Rechnungen gegen 0! Wir werden jetzt $h \searrow 0$ gehen lassen und benutzen zwei mal den Mittelwertsatz der Integralrechnung (Forster Analysis I, Satz 8, § 18). Weiter setzen wir voraus, dass die Dichten $f_X(x), f_Y(y)$ und $f_{X,Y}(x, y)$ stetige, beschränkte Funktionen sind (es geht auch mit weniger starken Forderungen, aber wir haben das immer erfüllt).

$$\begin{aligned}
\lim_{h \searrow 0} P[X \leq a|Y \in [y_0 - h, y_0 + h]] &= \lim_{h \searrow 0} \frac{P[X \leq a, Y \in [y_0 - h, y_0 + h]]}{P[Y \in [y_0 - h, y_0 + h]]} \\
&= \lim_{h \searrow 0} \frac{\int_{-\infty}^{a} \int_{y_0-h}^{y_0+h} f_{X,Y}(x, y) dy dx}{\int_{y_0-h}^{y_0+h} f_Y(y) dy} \\
&= \lim_{h \searrow 0} \frac{\int_{-\infty}^{a} 2h f_{X,Y}(x, \eta(x, h)) dx}{2h f_Y(\xi(h))} \\
&= \lim_{h \searrow 0} \frac{\int_{-\infty}^{a} f_{X,Y}(x, \eta(x, h)) dx}{f_Y(\xi(h))} \\
&= \frac{\int_{-\infty}^{a} f_{X,Y}(x, y_0) dx}{f_Y(y_0)} \\
&= \int_{-\infty}^{a} \frac{f_{X,Y}(x, y_0)}{f_Y(y_0)} dx
\end{aligned}$$

Also können wir wegen WTS-Definition 2.4 den Ausdruck (WTS-3.2) als bedingte Dichte von $X|Y = y_0$ auffassen. Damit können wir einen wichtigen Begriff einführen:

WTS-Definition 3.11 [bedingter Erwartungswert gegeben $Y = y_0$] *Wir definieren den bedingten Erwartungswert $E[X|Y = y_0]$ als:*

$$E[X|Y = y_0] := \begin{cases} \sum_{x_i} x_i P[X = x_i|Y = y_0] & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} x f_{X|Y=y_0}(x) dx & \text{falls } X \text{ stetig.} \end{cases}$$

Offenbar hängt dieser bedingte Erwartungswert von y_0 ab. Bedingte Wahrscheinlichkeiten sind selber auch Wahrscheinlichkeiten (ebenso Dichten!); wir haben also einfach einen gewöhnlichen Erwartungswert von X berechnet bedingt auf das Ereignis, dass $Y = y_0$ ist.

Das nachfolgende Resultat ist ein Pendant auf der Ebene der Erwartungswerte zu WTS-Lemma 1.7 (FTW) auf der Ebene der Wahrscheinlichkeiten:

WTS-Lemma 3.12 [Formel vom totalen Erwartungswert FTE] *Mit obigen Bezeichnungen gilt*

$$E[X] = \begin{cases} \sum_{y_j} E[X|Y = y_j]P[Y = y_j] & \text{falls } X, Y \text{ diskret} \\ \int_{-\infty}^{\infty} E[X|Y = y]f_Y(y)dy & \text{falls } X, Y \text{ stetig.} \end{cases}$$

WTS-4 Ausgewählte Verteilungen

WTS-4.2 Diskrete Verteilungen

WTS-4.2.1 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 $n=1$. Mit $0 \leq k \leq 1$ haben wir

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

WTS-4.2.2 Binomial Bin(n,p); Krengel § 2: 2.4

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). $E[Y]=np$ (WTS-Aufgabe 36 a)) und $V[Y]=np(1-p)$ (Bemerkung zu WTS-Lemma 3.8), **R-Befehl:** binom. $0 \leq k \leq n$:

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

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

WTS-4.2.3 Geometrisch Ge(p); Krengel § 2: 2.4

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.

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

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 (vgl. WTS-Aufgabenblatt 9)

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

"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.

WTS-4.2.4 Negativ Binomial NB(n,p); Krengel § 2: 2.4

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 WTS-4.2.3) und $V[W] = n(1-p)/p^2$ (folgt aus WTS-Lemma 3.8 b) und WTS-4.2.3), **R-Befehl:** nbinom. $w \geq n$

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

4.2.5 WTS-Poisson Po(λ); Krengel § 5: 5.4

Die Motivation für die Poissonverteilung kommt in Kapitel 8 dieser Vorlesung. 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.

WTS-4.3 Stetige Verteilungen

WTS-4.3.1 Uniform $U[a, b]$; Krengel § 10: 10.2

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 $E[U] = (a + b)/2$ und $V[U] = (b - a)^2/12$, **R-Befehl:** unif.

WTS-4.3.2 (Negativ-) Exponential Exp(λ); Krengel § 10: 10.2

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 $V[X] = 1/\lambda^2$ (WTS-Aufgabe 37), **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 (vgl. WTS-Aufgabenblatt 9)

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

"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!"

WTS-4.3.3 Gamma(n, λ); Krengel § 14: Anhang

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 WTS-4.3.2) und $V[Y] = n/\lambda^2$ (folgt aus WTS-Lemma 3.8 b) und WTS-4.3.2), **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 WTS-4.3.2 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).

WTS-4.3.4 Normal N(μ, σ^2); Krengel § 10: 10.2 und § 5: 5.2; Tabelle II hinten

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.

ACHTUNG: FÜR DIE FOLGENDEN BEIDEN RESULTATE SETZEN WIR NORMALVERTEILUNG VORAUS!

1. "Z-Transform" [Beweis auf WTS-Übungsblatt 9]: Wenn X eine $\mathcal{N}(\mu, \sigma^2)$-Verteilung hat, dann hat

$$\frac{X - \mu}{\sigma} \qquad \text{(Z – Transform)}$$

eine $\mathcal{N}(0, 1)$-Verteilung (welche in Büchern und Libraries von Statistik-Paketen abgelegt ist). $\mathcal{N}(0, 1)$ nennen wir "Standard-Normalverteilung".

2. Approximative Formeln für beliebige $\sigma > 0$: sei X eine $\mathcal{N}(0, \sigma^2)$-Zufallsgrösse. Dann gelten

$$P[|X| > \sigma] \doteq 33\%$$

und

$$P[|X| > 2\sigma] \doteq 5\%.$$

Allgemein: Eine Normalverteilung hat 66 % ihrer Wahrscheinlichkeit innerhalb 1 Standardabweichung vom Mittelwert entfernt und sogar 95 % ihrer Wahrscheinlichkeit innerhalb von 2 Standardabweichungen vom Mittelwert entfernt.

WTS-4.4 Zusammenhänge bei Verteilungen I: $\sum$ von *unabhängigen* Zufallsgrössen

WTS-4.4.1 Theoretische Grundlagen; Krengel § 11: 11.3

Wir werden uns in diesem Teil mit folgender Frage auseinandersetzen: "Wenn X und Y zwei unabhängige Zufallsgrössen sind, wie ist dann $X+Y$ verteilt"?

WTS-4.4.1.1 Diskreter Fall

$$P[X+Y=a] = \sum_i P[X=a-i]P[Y=i]$$

WTS-4.4.1.2 Stetiger Fall

$$P[X+Y \leq a] = \int_{-\infty}^{a} \int_{-\infty}^{\infty} f_X(x-y) f_Y(y) dy dx$$

Also ist $\int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$ die Dichte von $X+Y$ an der Stelle a.

WTS-4.4.2 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); vgl. 4.4.1.2.

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

WTS-5 $n \to \infty$ (Konvergenz, LLN, CLT)

Literatur WTS-Kapitel 5

* Cartoon Guide: Kapitel 5, Seite 106 in Kapitel 6
* Krengel: 3.6 in § 3, § 5 und § 12

WTS-5.1 Ungleichung von Bienayme-Tschebyschew

WTS-Satz 5.1 [Ungleichung von Bienayme-Tschebyschew] *Sei X eine Zufallsgrösse mit $E[|X|] < \infty$. Dann gilt für $\epsilon > 0$:*

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

auch

$$P[|X - \mu| \geq \epsilon] \leq \frac{1}{\epsilon^2} V[X]. \qquad (WTS - 5.1)$$

WTS-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?

WTS-5.2.2 LLN

Das Gesetz der grossen Zahlen ist streng genommen kein Theorem, sondern eine *Eigenschaft* einer Folge (das Theorem folgt dann in WTS-Theorem 5.3).

WTS-Definition 5.2 [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 WTS-Definition 5.2: 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).

Wir wissen noch nicht, wann eine Folge diesem Gesetz der grossen Zahlen genügt. Ein schönes Resultat dazu ist WTS-Theorem 5.3, welches *nicht die Existenz der Varianz fordert.*

WTS-Theorem 5.3 [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 $\epsilon > 0$:*

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

WTS-5.3 CLT (Central Limit Theorem; Zentraler Grenzwertsatz)

Fragestellung: Was geschieht mit dem Ausdruck

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}, \qquad (WTS-5.6)$$

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?

WTS-5.3.2 CLT

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

WTS-Theorem 5.4 [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 (WTS-5.6) konvergiert also in jedem Punkt gegen die Verteilungsfunktion der Standard-Normal-Verteilung.

Bemerkungen zu WTS-Theorem 5.4: 1. Wenn wir $\mu = 0, \sigma^2 = 1$ haben, sind wir in Situation WTS-5.3.1.

3. 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.

4. Ausdruck (WTS-5.6) erinnert zu Recht an die Z-Transform aus WTS-Kapitel 4. Betrachten wir dazu den Ausdruck

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

etwas genauer: mit

$$\sum_{k=1}^n X_k - n\mu$$

haben wir in einem ersten Schritt offenbar einfach *zentriert* (den Erwartungswert abgezogen - er ist jetzt 0), dann in einem zweiten Schritt mit $\sqrt{n}\sigma$ *normiert* und dabei

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

erhalten. Die Varianz ist jetzt 1. Zusammen nennt man diese beiden Schritte *standardisieren.*

And then a miracle occurs...,

denn das Ding hat am Schluss (bei $n = \infty$) eine Normalverteilung $\mathcal{N}(0,1)$. 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.

5. Ab welchem n "darf" man die Normalverteilung brauchen? Nichttriviale, allgemeine Fehler-Abschätzungen, welche nur von n abhängen und für alle Verteilungen gelten, existieren nicht. Wenn die X_k's aber Be(p)-Zufallsgrössen sind, so sagt eine Praktikerregel, dass $np(1-p) \geq 9$ gelten sollte, damit man mit gutem Gewissen die Normalverteilung benutzen darf. Seien Sie sich bewusst, dass dies von der konkreten Anwendung abhängt und mathematisch unpräzise ist (was wenn $np(1-p) = 8.9$?). Wenn p nahe bei 0 oder nahe bei 1 ist, wendet man besser WTS-Satz 5.6 an. Eine kleine Verbesserung der Approximation erreicht man im Fall von Summen diskreter Zufallsgrössen (also zum Beispiel Bernoulli/Binomial) mit der sogenannten Diskretisierungs-Korrektur (englisch: Continuity-Correction), siehe Cartoon-Guide Kapitel 5 oder Stahel 6.11 i.

WTS-5.4 Zusammenhänge bei Verteilungen II: Stetiges Analogon einer diskreten Verteilung und Limesverteilungen

In "WTS-4.4 Zusammenhänge bei Verteilungen I", haben wir uns gefragt, wie Summen von unabhängigen Zufallsgrössen verteilt sind. Jetzt wollen wir Paare von diskreten und stetigen Verteilungen finden, welche "etwas miteinander zu tun haben" (WTS-5.4.1 und WTS-5.4.2) und untersuchen, ob gewisse Verteilungen zu anderen Verteilungen werden, wenn wir *gezielt* einen Parameter gegen ∞ gehen lassen (WTS-5.4.3 und WTS-5.4.4).

WTS-5.4.1 Stetiges Analogon von Geometrisch ist Exponential

Bereits in der Analysis sieht man, dass geometrisches Wachstum $c^n, n \in \mathbb{N}$, und exponentielles Wachstum $c^t, t \in \mathbb{R}_+$, diskrete und stetige Pendants sind. Je nach Modellierungsgegenstand wird man das eine oder das andere wählen. Auf WTS-Aufgabenblatt 9 ist zu zeigen, dass geometrische und exponentielle Zufallsgrössen eine wichtige Eigenschaft teilen, nämlich die Gedächtnislosigkeit. Es drängt sich also fast auf, zu fragen, ob nicht ein tieferer Zusammenhang zwischen diesen beiden Verteilungen besteht. Das ist der Fall; wir präzisieren das in folgendem:

WTS-Satz 5.5 [Konvergenz der geometrischen- gegen die Exponential-Verteilung] *Sei T eine Exp(1)-Zufallsgrösse. Sei $X_n, n \geq 2$, eine Folge von Ge(1/n)-Zufallsgrössen, d.h. X_n habe eine Ge(1/n)-Verteilung. Dann gilt für $a \in \mathbb{R}_+$:*

$$P\left[\frac{1}{n}X_n \leq a\right] \longrightarrow P[T \leq a]$$

für $n \to \infty$.

Bemerkung zu WTS-Satz 5.5 Der Erwartungswert von T ist 1. Der Erwartungswert von X_n/n ist auch 1. Dies ist allgemein ein wichtiges Prinzip, wenn man solche Zusammenhänge formulieren will (vgl. auch WTS-5.4.3): man normiert und/oder zentriert mindestens derart, dass die Erwartungswerte der Limeszufallsgrösse (hier T) und der konvergierenden Zufallsgrössen (hier X_n) gleich sind (oder zumindest asymptotisch gleich werden).

WTS-5.4.2 Stetiges Analogon von Negativ Binomial ist Gamma

In WTS-Kapitel 4 haben wir gesehen, dass bei Unabhängigkeit der Summanden

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

und

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

Deshalb ist es nach WTS-5.4.1 nicht überraschend, dass das stetige Analogon von NB eine Gamma-Zufallsgrösse ist. Wir verzichten hier aus Zeitgründen auf Satz und entsprechenden Beweis.

WTS-5.4.3 Von Binomial zu Poisson

Wenn man die Wahrscheinlichkeitsfunktion (die Werte von $P[X = k], k \geq 0$) einer Binomial- und einer Poissonverteilung mit gleichem Erwartungswert vergleicht, so sieht man, dass der Unterschied immer kleiner wird, je grösser n in der Binomial-Verteilung ist:

WTS-Satz 5.6 [Konvergenz der Binomial- gegen die Poisson-Verteilung] *Sei X_n eine Folge von $Bin(n, p_n)$-Zufallsgrössen. Es gelte $\forall n$ genügend gross: $np_n = \lambda > 0$ (Erwartungswerte jeweils gleich). λ ist konstant! Dann gilt für $r \in \mathbb{N}_0$ fest:*

$$P[X_n = r] \longrightarrow e^{-\lambda} \frac{\lambda^r}{r!}$$

für $n \to \infty$. Damit haben wir im Limes eine Poisson-Verteilung mit Parameter λ.