Einführung in die
Wahrscheinlichkeitstheorie und Statistik

Dr. C.J. Luchsinger

2 Zufallsgrössen

Literatur Kapitel 2

* Statistik in Cartoons: Kapitel 4
* Krengel: 3.1 und 3.2 in § 3 und (Honours Program) § 10 sowie 11.1, 11.2 und 11.3 in § 11.
* Storrer: 37, 38, 39, 40, 41

2.1 Zufallsgrösse

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.*

Bemerkungen zu Definition 2.1:

1. Zufallsgrössen nennt man auch Zufallsvariablen.

2. Der erste Teil dieser Definition ist einfach zu begreifen. Der zweite Teil wird uns in dieser Vorlesung nicht weiter beschäftigen (in Vlsg WT hingegen schon) und ist auch nicht prüfungsrelevant: in 1.4.2 haben wir gesagt, dass wir bei endlichem Ω einfach die Potenzmenge für $\mathcal{A}$ nehmen. Falls $\Omega = \mathbb{R}$, nehmen wir einfach die Borel-σ-Algebra. Sie enthält alle für unsere Vorlesungen benötigten Teilmengen von $\mathbb{R}$; insbesondere wird bei uns $\{\omega \in \Omega | X(\omega) \leq a\} \in \mathcal{A}$ für alle reellen a immer erfüllt sein. Als Kontrast: auf Übungsblatt 4 ist eine *einfache* Situation zu konstruieren, in der die Messbarkeit nicht erfüllt ist. Man kann dies einfach erreichen, in dem wir $\mathcal{A}$ nicht so reichhaltig ausstatten.

Beispiele:

1. Würfel: $\Omega = \{1,2,3,4,5,6\}$. X soll die Zufallsgrösse sein, welche die Augenzahl angibt. Wir wählen $X(i) = i$ für alle $1 \leq i \leq 6$.

2. Münzwurf: Ich gewinne 1.- falls es Kopf (k) gibt und verliere 1.- falls es Zahl (z) gibt. X_1 ist mein Konto nach dem *ersten* Münzwurf: $X_1(k) = 1, X_1(z) = -1$.

2'. Multipler Münzwurf: Werfe gleiche Münze 10 mal, unabhängig voneinander (siehe 2.5 für exakte Definition): X_i ist Gewinn/Verlust (engl.: P/L=Profit/Loss) im i'ten Wurf.

$$Y_{10} := \sum_{i=1}^{10} X_i$$

ist mein Konto zum Zeitpunkt 10. Übrigens: gute Wahl ist

$$\Omega := \{(k,k,k,k,k,k,k,k,k,k),(k,k,k,k,k,k,k,k,k,z),\ldots,(z,z,z,z,z,z,z,z,z,z)\}$$

mit $2^{10} = 1024$ Elementen! Mit $\omega_1 := (k,k,k,k,k,k,k,k,k,k)$ hätte ich $Y_{10}(\omega_1) = X_1(\omega_1) + X_2(\omega_1) + \ldots + X_{10}(\omega_1) = 1 + 1 + \ldots + 1 = 10$.

3. Tierversuch mit einer Maus: $\Omega := \{S, D\}$ ("survives, dies"). $U_1(S) = 1$ falls Maus überlebt, $U_1(D) = 0$ falls Maus stirbt. U_1 ist die Anzahl Mäuse, welche nach Versuch noch leben (0 oder 1).

3'. Mehrere Mäuse: gleiches Experiment mit 5 Mäusen, unabhängige Versuche. U_i ist Anzahl überlebende Mäuse im i'ten Versuch (0 oder 1).

$$W := \sum_{i=1}^{5} U_i$$

ist die Anzahl noch lebende Mäuse, sobald wir alle 5 Experimente abgeschlossen haben.

Wir werden jetzt das bisher erreichte mit dem ersten Kapitel kombinieren und führen das "P" ein. Angenommen, die Überlebenswahrscheinlichkeit ist 10 %, so können wir

$$P[U_1 = 1]$$

in Beispiel 3 berechnen; die Antwort ist:

Mathematisch sauber ist dies

$$P[U_1 = 1] := P[\{\omega | U_1(\omega) = 1\}] = P[\{S\}] = 0.1,$$

weil P auf Teilmengen von Ω (hier $\{S\}$) operiert. Wir müssen aber nicht immer so in die Niederungen der Definitionen absteigen.

Betrachten wir ein Beispiel, welches ein bisschen komplizierter ist (3'): Mit Überlebenswahrscheinlichkeit 10 % haben wir

$$P[W = 4] = 0.00045;$$

dies ist

$$\binom{5}{4} 0.1^4 0.9^1.$$

Mehr dazu drei Seiten später

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. 1.4.4) definieren: Sei A ein beliebiges Element 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.

Wie werden jetzt die folgenden Situationen gemäss Definition 2.1 mathematisch exakt modelliert?

1. X soll die Anzahl "Kopf" sein im fairen Münzwurfspiel, wenn 3 mal geworfen wird.

Wie sind $(\Omega, \mathcal{A}, P)$?

Geben Sie P_X an; wo nimmt es welche Werte $\neq 0$ an?

***

2. X soll das Quadrat der Augenzahl sein im fairen Wurf eines Würfels.

Wie sind $(\Omega, \mathcal{A}, P)$?

Geben Sie P_X an; wo nimmt es welche Werte $\neq 0$ an?

Geben Sie hier elegant X an.

Wie ist $P_X[[8, 17]]\,(= P[\{\omega \in \Omega | X(\omega) \in [8, 17]\}])$?

***

3. U sei eine sogenannte Uniform-verteilte Zufallsgrösse auf $[0, 1]$ (später mehr dazu).

$(\Omega, \mathcal{A}) = ([0, 1], \mathcal{B}([0, 1]))$

$P_U[[a, b]] := b - a$, falls $0 \leq a \leq b \leq 1$.

$U(\omega) = \omega$.

Bevor wir uns der Verteilungs*funktion* zuwenden, wollen wir vom Publikum wissen, welche Verteilungen bereits bekannt sind (es kommen im Verlauf der Vorlesung immer wieder neue Verteilungen vor; ausführliche Besprechung in Kapitel 4).

Schema:

* Name
* Wahrscheinlichkeitsfunktionen (diskret) bzw. Dichten (stetig) mit Graph
* Bedeutung, Motivation, Modell für ...

diskret (Bernoulli, Binomial, eine weitere):

stetig; vgl. Cartoon Guide p. 63 (Uniform, Normalverteilung, eine weitere):

Die kommenden 5 Seiten sind zentral und für jedeN DozentIn schwierig zu präsentieren und für die StudentInnen entsprechend schwer zu verstehen. Eine gute Strategie für die StudentInnen ist es, die Resultate vorerst einfach zu akzeptieren, die Beispiele sorgfältig anzuschauen und Übungen dazu gut zu lösen.

2.2 Verteilungsfunktion

Definition 2.2 [Verteilungsfunktion F; engl: Cumulative Distribution Function CDF] *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.

Beispiele von beiden letzten Seiten:

a) Bernoulli Be(p) und danach noch Bin(n, p)

b) Geometrisch Ge(p)

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 2.3 zusammenfassen:

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.*

Beweis von Lemma 2.3: a) Wir werden nur den zweiten Teil beweisen, der erste Teil geht analog und ist als kleine, freiwillige Übung empfohlen (zuerst für die Erstsemestrigen Besprechung von "lim"). Betrachten wir zuerst die ganz spezielle Folge $a \to \infty$, welche monoton gegen ∞ konvergiert "entlang n". Wir können dann wegen Lemma 1.8 (1. Fall mit aufsteigender Folge von Ereignissen) folgern:

$$\lim_{n\to\infty} F(n) := \lim_{n\to\infty} P[X \leq n] = P[\cup_{n=1}^{\infty}\{X \leq n\}] = P[\Omega] = 1.$$

Wir haben jetzt aber eine ganz spezielle Folge genommen. Unter $\lim_{a\to\infty}$ ist aber eine beliebige Folge gemeint, welche gegen ∞ konvergiert. Aber angenommen, jemand nimmt eine andere Folge a_n, welche gegen ∞ konvergiert. Für gegebenes $\epsilon > 0$ habe ich aber ein N, sodass $|1 - F(n)| < \epsilon$ für alle $n \geq N$. Wenn die andere Folge a_n auch gegen ∞ konvergiert, wird sie irgendwann definitiv $> N$ sein. Dann wird aber wegen Lemma 1.3 d) auch gelten $1 - \epsilon \leq F(N) \leq F(a_n)$. Da ϵ beliebig war, ist somit dieser zweite Teil bewiesen.

b) Sei $a_1 \leq a_2$. Wir müssen zeigen, dass $F(a_1) \leq F(a_2)$. Dies ist aber gleichbedeutend mit $P[X \leq a_1] \leq P[X \leq a_2]$. Dies gilt aber wegen Lemma 1.3 d). Also ist F monoton wachsend.

Für die Rechtsstetigkeit muss gelten: $\lim_{h\searrow 0} F(a + h) = F(a)$. Aber für $h > 0$ gilt:

$$P[X \leq a + h] = P[X \leq a] + P[a < X \leq a + h].$$

Zu zeigen bleibt also, dass $\lim_{h\searrow 0} P[a < X \leq a + h] = 0$. Dies kann man mit den gleichen Argumenten wie in Teil a) als freiwillige Übung beweisen.

$\square$

2.3 Diskrete und (absolut) stetige Zufallsgrösse

Wir werden jetzt die Integralrechnung einsetzen. Sie wird in der Vorlesung Infinitesimalrechnung I gegen Ende Semester exakt behandelt. Die Kenntnisse aus dem Gymnasium reichen jedoch für unsere Bedürfnisse. Wer die Differential- und Integralrechnung im Gymnasium nicht kennengelernt hat, lehne sich bitte für die Übungen an StudentInnen an, welche diese Konzepte bereits kennen.

Definition 2.4 [Diskrete und (absolut) 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 (absolut) 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 werden in der Vlsg WT die Definition von (absolut) stetigen Zufallsgrössen noch präzisieren: das obige Integral ist nämlich in der WT nicht ein Riemann-Integral, sondern ein Lebesgue-Integral (Verallgemeinerung des Riemann-Integrals), welches Sie im dritten oder fünften Semester vor der Vlsg WT kennenlernen werden. Wir schreiben in der Vlsg WTS "stetige Zufallsgrösse" statt "absolut stetige Zufallsgrösse".

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:

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

Betrachten Sie dazu neben der Normalverteilung auch Beispiel c) in 2.2.

Weshalb betrachtet man die **Verteilungsfunktion** F einer Zufallsgrösse? 5 Gründe:

1. Wir können im Fall von stetigen Zufallsgrössen mit Hilfe von $F' = f$ einfach die Dichte berechnen (siehe auch 2. Grund).
2. Wie in 2.6 (weiter unten) ersichtlich, wird man mit Hilfe der Verteilungsfunktion einfach die Verteilung (insbesondere die Dichte) von transformierten Zufallsgrössen berechnen können (z.B. von X^2).
3. In Kapitel 5 werden wir mit Hilfe der Verteilungsfunktion den CLT (Central Limit Theorem; Zentraler Grenzwertsatz) formulieren können.
4. In der Statistik kann man zum Beispiel mit sogenannten QQ-Plots (Quantil vs Quantil-Plots) Verteilungen miteinander vergleichen.
5. Eher theoretischer Grund: Die Verteilungsfunktion existiert immer (diskret und stetig). Ohne Aufsummieren bzw. Aufintegrieren haben wir im diskreten Fall die Wahrscheinlichkeitsfunktion $P[X = x_i]$ und im stetigen Fall die Dichte $f(x)$. Diese beiden Objekte (Wahrscheinlichkeitsfunktionen und Dichten) entsprechen sich zwar, wir haben aber von der Analysis her andere Gebilde (Summen vs Integrale). Die Verteilungsfunktionen hingegen haben in beiden Fällen die gleiche Bedeutung und Definition:

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

Lose Bemerkung: Falls es in einer Aufgabe oder Frage heisst "Wie ist die Verteilung von X?", so ist diese Frage ungenau. Antworten können dann sein:

1. Die Dichte ist...
2. Die Verteilungsfunktion sieht folgendermassen aus...
3. Die Verteilung ist $\mathcal{N}(1, 4)$.

2.4 Bivariate Verteilung

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.

<u>Turnübungen</u>

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

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

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:

$$\begin{aligned}
P[X = x_i] := P[\{\omega | X(\omega) = x_i\}] &= P[\{\omega | X(\omega) = x_i\} \cap \Omega] \\
&= P[\{\omega | X(\omega) = x_i\} \cap (\cup_j \{\omega | Y(\omega) = y_j\})] \\
&= P[\cup_j (\{\omega | X(\omega) = x_i\} \cap \{\omega | Y(\omega) = y_j\})] \\
&= \sum_j P[\{\omega | X(\omega) = x_i\} \cap \{\omega | Y(\omega) = y_j\}] \\
&=: \sum_j P[X = x_i, Y = y_j].
\end{aligned}$$

Stetig erhält man analog

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

Der Beweis hierfür (nur für StudentInnen der dritten Semester und mehr als kleine freiwillige Hausaufgabe) geht über eine Ableitung nach x von:

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

Die kleine Rechnung ist dabei einfach; man muss die Schritte aber begründen können (mehr in Vlsg WT).

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

2.5 Unabhängigkeit von Zufallsgrössen

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

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

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):

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

Teilweise Beweise der Äquivalenzen - vollständiger Beweis in Vlsg WT

$$(I) \Leftrightarrow (II) \Leftrightarrow (III);$$

<u>$(I) \Rightarrow (II)$ (In der Klasse):</u>

$\underline{(III) \Rightarrow (II)}$:

$$
\begin{aligned}
F_{X,Y}(a,b) &:= \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(u,v)dvdu = \int_{-\infty}^{a}\int_{-\infty}^{b} f_X(u)f_Y(v)dvdu \\
&= \int_{-\infty}^{a} f_X(u)du \int_{-\infty}^{b} f_Y(v)dv = F_X(a)F_Y(b).
\end{aligned}
$$

$\square$

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**)

2.6 Transformation von (stetigen) Zufallsgrössen [Krengel 11.2]

Wir werden manchmal (meist stetige) Zufallsgrössen transformieren müssen. Zum Beispiel haben wir eine Zufallsgrösse X, welche normalverteilt ist, und wir fragen uns nach der Dichte von X^2. Elementare Kenntnisse aus der Integralrechnung lassen vermuten, dass die Lösung nicht einfach $[f_X(x)]^2$ ist.

Meist sind die Rechnungen relativ einfach. Wir können die Verteilungsfunktion gewinnbringend einsetzen:

Beispiel 1: X habe eine $U[0,1]$-Verteilung. Wie ist die Verteilung (genauer die Dichtefunktion) von X^2?

Beispiel 2: $X_1, \ldots, X_n$ seien iid Exp(λ)-verteilt. Wie ist die Verteilung (genauer die Dichtefunktion) von

$$Y := \min_i \{X_i | 1 \leq i \leq n\}?$$

Überlegen Sie sich zuerst, was Y für ein Tier ist. Eine analoge Aufgabe mit Maximum ist in den Übungen zu lösen.

Beispiel 3: Allgemeine Formel (stetig) im Fall von differenzierbaren, streng monoton wachsenden Transformationen g.

Im diskreten Fall sind in erster Linie gute Kenntnisse in Buchhaltung und eine entsprechend exakte Arbeitsweise hilfreich. Gute, allgemeine Formeln wie in Beispiel 3 gibt es dort nicht.