Kapitel 4

Appendix: Sammlung von Fakten und Beispielen zu Zufallsvariablen

In diesem Teil des Appendix sammeln wir wesentliche Fakten und Beispiele. Am Ende des Tages ist Stochastik nicht nur abstrakte Maßtheorie, ihr müsst auch mit Beispielen rumrechnen könnne und ein Gefühl für Eigenschaften bekommen.

Allgemeine Definitionen für beliebige Zufallsvariablen:

- Für eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ schreibt man $X \sim F$, falls

$$\mathbb{P}(X \in (a,b]) = \mathbb{P}_X((a,b]) = F(b) - F(a).$$

  Wozu haben wir Maßtheorie betrieben? Um die Existenz von stochastischen Modellen (W-Raum und Zufallsvariable) zu beweisen! Wegen Carathéodory und der Borel-σ-Algebra gibt es für alle Verteilungsfunktionen Zufallsvariablen mit $X \sim F$.

- Wenn das Integral existiert, definiert man für $g : \mathbb{R} \to \overline{\mathbb{R}}$ messbar $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(X(\omega))\,\mathrm{d}\mathbb{P}(\omega)$ und es gilt aufgrund des Transformationssatzes $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\,\mathrm{d}\mathbb{P}_F(x)$. Einige Spezialfälle bekommen eigene Namen:
  - Erwartungswert für $g(x) = x$
  - k.tes Moment für $g(x) = x^k$
  - exponentielles Moment für $g(x) = e^{\lambda x}$

  Wozu haben wir Integrationstheorie betrieben? Um Erwartungswerte und ähnliches für beliebige Zufallsvariablen zu definieren! Ohne Integrationstheorie ginge das nur für diskrete Zufallsvariablen oder Zufallsvariablen mit Dichten.

4.1 Absolutstetige Zufallsvariablen

Definitionen und Fakten:

- $X \sim F$ heißt absolutstetig, falls F eine Dichte f hat. Es gilt dann

$$\mathbb{P}(X \in (a,b]) = \int_a^b f(x)\,\mathrm{d}x.$$

Weil $\{a\} = [a, a]$, gilt für absolutstetige Zufallsvariablen

$$\mathbb{P}(X = a) = \int_a^a f(x)dx = 0$$

und daher auch

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in (a, b]) = \mathbb{P}(X \in [a, b)) = \mathbb{P}(X \in (a, b)),$$

es gibt also keine Masse auf einpunktigen Mengen.

- Momente lassen sich einfach berechnen. Für messbare $g : \mathbb{R} \to \overline{\mathbb{R}}$ ist

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f(x)\,\mathrm{d}x,$$

wenn eine der Seiten (und damit die andere) definiert ist.

Wichtige Bemerkung für das Verständniss: Wenn wir irgendeine nichtnegative Funktion f mit $\int_{\mathbb{R}} f(x)\,\mathrm{d}x < \infty$ kennen, so bekommen wir mit $h(x) := Cf(x)$ eine Dichte, wobei $\frac{1}{C} = \int_{\mathbb{R}} f(x)\,\mathrm{d}x$ ist. Unsere Beispiele bestehen also im Prinzip aus den integrierbaren nichtnegativen Funktionen, die wir gut integrieren können. Das sind die Exponentialfunktion (gibt die Exponentialverteilung), Exponentialfunktion mit einem Quadrat (gibt die Normalverteilung), Varianten der Funktion $1/x^2$ (gibt die Cauchy Verteilung) oder konstante Funktionen auf kompakten Mengen (gibt die uniformen Verteilungen).

(A) Normalverteilung - $\mathcal{N}(\mu, \sigma^2)$

Die wichtigste Verteilung der Stochastik ist die Normalverteilung. Das liegt am zentralen Grenzwertsatz.

Vorteile:

- Alle Momente existierten und können berechnet werden.
- Die Verteilungsfunktion ist zwar nicht explizit, dennoch kann vieles ausgerechnet werden.
- Zwei Parameter μ und σ^2 geben viel Freiheit, die Verteilung an echte Daten anzupassen.

Nachteil:

- Verteilungsfunktion ist nicht explizit gegeben, Wahrscheinlichkeiten der Form $\mathbb{P}(X \in [a, b])$ müssen daher abgeschätzt werden.

Wichtige Charakteristiken auf einen Blick:

Parameter	$\mu \in \mathbb{R}$ (Verschiebung), $\sigma^2 > 0$ (Stauchung)
Wertebereich	$\mathbb{R}$
Verteilungsfunktion	$F_{\mu,\sigma^2}(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x$
Dichte	$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Grobe Verteilung der Masse	Viel Masse nah bei μ, extrem wenig bei $+\infty$ und $-\infty$.
Erwartungswert	$\mathbb{E}[X] = \mu$
Varianz	$\mathbb{V}[X] = \sigma^2$
Momentenerzeugende Funktion	existiert für alle $t \in \mathbb{R}$, $m_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

Zum rumspielen hier eine interaktive Graphik. Zu beachten ist: Viel Masse liegt dort, wo die Dichte groß ist. Das spiegelt sich dadurch wieder, dass dort die Verteilungsfunktion stark wächst. Andersrum: Wenig Masse liegt dort, wo die Verteilungsfunktion flach ist bzw. die Dichte klein ist.

(B) Exponentialverteilung - Exp(λ)

Wichtige Verteilung, besonders in der Anwendungen bei Markovprozessen.

Vorteile:

- Alle Momente existierten und können berechnet werden.
- Verteilungsfunktion und Dichte sind explizit und sehr einfach. Vieles kann berechnet werden.

Nachteil:

- Nur ein Parameter um Verteilung auf Daten anzupassen

Wichtige Charakteristiken auf einen Blick:

Parameter	$\lambda > 0$ (Stauchung)
Wertebereich	$[0, \infty)$
Verteilungsfunktion	$F_\lambda(t) = (1 - e^{-\lambda t})\mathbf{1}_{[0,\infty)}(t)$
Dichte	$f_\lambda(x) = \lambda e^{-\lambda}\mathbf{1}_{[0,\infty)}(x)$
Grobe Verteilung der Masse	Viel Masse nah bei 0, wenig bei $+\infty$
Erwartungswert	$\mathbb{E}[X] = \frac{1}{\lambda}$
Varianz	$\mathbb{V}[X] = \frac{1}{\lambda^2}$
Momentenerzeugende Funktion	existiert für alle $t < \lambda$, $m_X(t) = \frac{t}{t-\lambda}$

(C) Uniforme Verteilung - $\mathcal{U}([a, b])$

Modell für das einfachste Experiment, Ziehen aus einem Intervall ohne Bevorzugung verschiedener Bereiche.

Vorteile:

- Alle Momente existierten und können berechnet werden.
- Verteilungsfunktion und Dichte sind explizit und sehr einfach. Vieles kann berechnet werden.
- Man kann alle anderen Verteilungen aus $\mathcal{U}([0, 1])$ gewinnen, theoretisch reicht also die uniforme Verteilung für alles aus!

Nachteile:

- Gar kein Parameter um Verteilung auf Daten anzupassen
- Eher langweilig.

Wichtige Charakteristiken auf einen Blick:

Parameter	keine Parameter
Wertebereich	$[a,b]$
Verteilungsfunktion	$F(t) = \frac{t-a}{b-a} 1_{[a,b]}(t)$
Dichte	$f(x) = 1_{[a,b]}(x)$
Grobe Verteilung der Masse	uniform auf $[a,b]$ (daher der Name)
Erwartungswert	$\mathbb{E}[X] = \frac{a+b}{2}$
Varianz	$\mathbb{V}[X] = \frac{(b-a)^2}{12}$
Momentenerzeugende Funktion	existiert für alle $t \in \mathbb{R}$, $m_X(t) = \begin{cases} \frac{e^{bs}-e^{as}}{(b-s)s} & : s \neq 0 \\ 1 & : s = 0 \end{cases}$

(D) Cauchy Verteilung - Cauchy(s,t)

Die Cauchy-Verteilung ist eine sogenannte „heavz-tailed“ Verteilung, d. h. viel Masse liegt bei unendlich. Für uns hat die Verteilung eigentlich nur Nachteile, es geht so ziemlich alles schief.
Vorteile:

- Zwei Parameter s und t geben viel Freiheit, die Verteilung an echte Daten anzupassen.
- Verteilungsfunktion und Dichte sind explizit.

Nachteile

- Klassisches Beispiel für eine Verteilung, deren Erwartungswert nicht existiert.

Wichtige Charakteristiken auf einen Blick:

Parameter	$t \in \mathbb{R}$ (Verschiebung), $s > 0$ (Stauchung)
Wertebereich	$\mathbb{R}$
Verteilungsfunktion	$F_{s,t}(r) = \frac{1}{2} + \frac{1}{\pi} \arctan(\frac{r-t}{s})$
Dichte	$f_{s,t}(x) = \frac{1}{\pi} \frac{s}{s^2+(x-t)^2}$
Grobe Verteilung der Masse	viel Masse bei 0, aber auch viel bei $+\infty$ und $-\infty$
Erwartungswert	existiert nicht!
Varianz	existiert nicht!
Momentenerzeugende Funktion	existiert für kein $t \in \mathbb{R}$

4.2 Diskrete Zufallsvariablen

Definitionen und Fakten:

- $X \sim F$ heißt diskret, falls F eine diskrete Verteilungsfunktion ist (d. h. F ist stückweise konstant mit Sprüngen p_i an Stellen a_i). Man kann dann immer

  $$F(t) = \sum_{k=1}^{N} p_k \mathbf{1}_{[a_k,\infty)}(t), \quad t \in \mathbb{R},$$

  schreiben, wobei $N \in \mathbb{N} \cup \{+\infty\}$, $a_i \in \mathbb{R}$ für $i = 1, ..., N$ und $\sum_{k=1}^{N} p_k = 1$. Wir kennen in dem Fall diskreter Verteilungen auch das Maß $\mathbb{P}_F$ expliziert, es gilt

  $$\mathbb{P}_F = \sum_{k=1}^{N} p_k \delta_{a_k},$$

  für alle $A \in \mathcal{B}(\mathbb{R})$. Weil per Definition $\mathbb{P}(X \in A) = \mathbb{P}_F(A)$, folgt damit

  $$\mathbb{P}(X \in A) = \mathbb{P}_F(A) = \sum_{k=1, a_k \in A}^{N} p_k \quad \text{und insbesondere} \quad \mathbb{P}(X = a_k) = p_k.$$

  Wir sehen also, dass X ausschließlich die Werte $a_1, ..., a_N$ mit positiver Wahrscheinlichkeit annimmt, denn es gilt

  $$\mathbb{P}(X \notin \{a_1, ..., a_n\}) = 1 - \mathbb{P}(X \in \{a_1, ..., a_N\}) = 1 - 1 = 0.$$

  Wir nennen die p_k Wahrscheinlichkeiten von a_k, manchmal auch Wahrscheinlichkeitsgewichte. Im Gegensatz zu absolutstetigen Zufallsvariablen haben diskrete Zufallsvariablen also durchaus Masse auf einpunktigen Mengen, nähmlich gerade auf den Mengen $\{a_1\}, ...$

- Momente lassen sich einfach berechnen. Für messbare $g : \mathbb{R} \to \overline{\mathbb{R}}$ ist

  $$\mathbb{E}[g(X)] = \sum_{k=1}^{N} p_k g(a_k),$$

  wenn eine der Seiten (und damit die andere) definiert ist.

Wie bei absolutstetigen Verteilungen können wir uns aus den Analysis 1 Kenntnissen im Prinzip alle klassischen diskreten Verteilungen erraten. Es gibt zwei wesentliche Fälle:

- $N < \infty$: Man wähle irgendwelche $a_1, ..., a_N \in \mathbb{R}$ und denke sich irgendwelche $p_1, ..., p_N \geq 0$ aus, die sich zu 1 addieren.
- $N = +\infty$: Man wähle irgendeine Folge $a_1, a_2, ... \in \mathbb{R}$ und benutze eine bekannte konvergente Reihe $\sum_{k=1}^{\infty} b_k$ aus Analysis 1. Ist $\frac{1}{C} = \sum_{k=1}^{\infty} b_k$, so setzt man $p_k = Cb_k$. Schon haben wir eine diskrete Verteilung! Wenn wir jetzt an Analysis 1 denken, fallen uns nur wenige konvergente Reihen mit positiven Summanden ein. Das ist die sche Reihe (gibt die geometrische Verteilung) und die die Exponentialreihe (gibt die Poisson Verteilung). Das war es eigentlich schon.

(A) Bernoulli Verteilung - Ber(p)

Vorteile:

- Total einfach, Modell für den Münzwurf einer unfairen Münze.
- Alles, wirklich alles, lässt sich sofort ausrechnen.

Nachteil:

- Komplett langweilig!

Wichtige Charakteristiken auf einen Blick:

Parameter	$p \in [0, 1]$
Wertebereich	0 oder 1
Wahrscheinlichkeitsgewichte	$p_1 = p, p_0 = q$ wobei $q = 1 - p$
Grobe Verteilung der Masse	Anteil p der Masse auf der 1, Rest auf der 0
Erwartungswert	$\mathbb{E}[X] = p$
Varianz	$\mathbb{V}[X] = pq$
Momentenerzeugende Funktion	existiert für alle $t \in \mathbb{R}$, $m_X(t) = 1 - t + pe^t$

(B) Binomial Verteilung - Bin(n, p)

Vorteil:

- Modell für Ziehen mit Zurücklegen, klare Interpretation (p_k ist die Wahrscheinlichkeit, bei n Versuchen, k Treffer zu landen).
- Viele Formeln berechenbar. Es gibt einen einfachen Beweis für den zentralen Grenzwertsatz.

Nachteil:

- Teilweise fuzzelig zum Rechnen. Muss immer irgendwie die Binomialformel ins Spiel bringen.

Wichtige Charakteristiken auf einen Blick:

Parameter	$n \in \mathbb{N}, p \in [0, 1]$
Wertebereich	$\{0, ..., n\}$
Wahrscheinlichkeitsgewichte	$p_k = \binom{n}{k} p^k (1-p)^{n-k}$
Grobe Verteilung der Masse	viel Masse in der Mitte von $\{0, ..., n\}$, wenig bei 0 und n
Erwartungswert	$\mathbb{E}[X] = np$
Varianz	$\mathbb{V}[X] = npq$
Momentenerzeugende Funktion	existiert für alle $t \in \mathbb{R}$, $m_X(t) = (pe^t + 1 - p)^n$

(C) Poisson Verteilung Poi(λ)

Vorteile:

- Viele Formeln berechenbar, ist auch eine gute Reihe!

Wichtige Charakteristiken auf einen Blick:

Parameter	$\lambda > 0$
Wertebereich	$\mathbb{N}$
Wahrscheinlichkeitsgewichte	$p_k = e^{-\lambda}\frac{\lambda^k}{k!}$
Grobe Verteilung der Masse	viel Masse auf kleinen Zahlen, wenig bei $+\infty$.
Erwartungswert	$\mathbb{E}[X] = \lambda$
Varianz	$\mathbb{V}[X] = \lambda$
Momentenerzeugende Funktion	existiert für alle $t \in \mathbb{R}$, $m_X(t) = \exp(\lambda(e^t - 1))$

(D) Geometrische Verteilung Geo(p)

Vorteile:

- Viele Formeln berechenbar, ist auch eine gute Reihe!

Nachteil:

- Teilweise fuzzelig zum Rechnen. Viel Indexverschiebung als analog zur Substitution im Fall der Dichten.

Wichtige Charakteristiken auf einen Blick:

Parameter	$p > 0$
Wertebereich	$\mathbb{N}\backslash\{0\}$
Wahrscheinlichkeitsgewichte	$p_k = p(1-p)^{n-1}$
Grobe Verteilung der Masse	viel Masse auf kleinen Zahlen, wenig bei $+\infty$.
Erwartungswert	$\mathbb{E}[X] = \frac{1}{p}$
Varianz	$\mathbb{V}[X] = \frac{1-p}{p^2}$
Momentenerzeugende Funktion	existiert für alle $t \in \mathbb{R}$, $m_X(t) = \frac{pe^t}{1-(p-1)e^t}$