Inhaltsverzeichnis

1 Einführung

Dieses Skript ist eine Einführung in die masstheoretische Wahrscheinlichkeitstheorie. Wir führen zuerst den Begriff der Verteilung ein und schauen uns dann an, was diskrete und stetige Verteilungen im masstheoretischen Sinn verbindet.

Es werden einige wenige Fakten über Integral wie Linearität, Satz der Monotonen Konvergenz,... verwendet. Dem Inhalt kann aber ohne weitere Vorkenntnisse gefolgt werden.

2 Verteilungen

Die Masstheorie bildet die Grundlage der modernen Wahrscheinlichkeitstheorie und geht auf Kolmogorov zurück.

Definition 2.1: σ-Algebra, Ereignisraum

Sei Ω eine beliebige Menge. Eine σ-Algebra ist eine Teilmenge der Potenzmenge $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ welche folgende Axiome erfüllt:

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
- $(A_i)_{i\in\mathbb{N}} \implies \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$

Das Tupel $(\Omega, \mathcal{A})$ wird *messbarer Raum* oder *Ereignisraum* genannt.

Die Definition einer σ-Algebra sieht der Definition einer Topologie sehr ähnlich. Im Unterschied zur Topologie sind in der Masstheorie die Mengenoperationen abgeschlossen unter *abzählbaren* Vereinigungen und Durchschnitten. (Dass $\mathcal{A}$ neben Vereinigungen auch abgeschlossen unter Durchschnitten ist folgt aus der De Morganschen Regel und den Axiomen). Der griechische Buchstabe σ steht hier für die Abzählbarkeit. Die Elemente der σ-Algebra werden *Ereignisse* genannt. Die σ-Algebra ist dann die Menge der (beobachtbaren) Ereignisse.

Beispiel 2.1. (Borel σ-algebra) Wir haben in der Serie gesehen, dass für eine Menge $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ eine eindeutig kleinste σ-Algebra $\sigma(\mathcal{E})$ existiert, welche $\mathcal{E}$ enthält. Sei nun $\Omega = \mathbb{R}$ und $\mathcal{E} := \mathcal{O}$ die Standardtopologie auf $\mathbb{R}$. Dann ist die Borel σ-Algebra definiert als $\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{O})$. Sei $a < b$, dann gilt insbesondere

$$(a, b] = \bigcap_{n\in\mathbb{N}} (a, b + \frac{1}{n}) \in \mathcal{B}(\mathbb{R}),$$

da $(a, b + \frac{1}{n}) \in \mathcal{O} \subseteq \mathcal{B}(\mathbb{R})$.

Das Konzept der Wahrscheinlichkeit wird nun auf einem Ereignisraum definiert.

Definition 2.2: (Wahrscheinlichkeits-)Mass

Sei $(\Omega, \mathcal{A})$ ein Ereignisraum. Eine Abbildung $\mu : \mathcal{A} \to [0, \infty]$ wird Mass genannt, falls

- $\mu(\varnothing) = 0$
- Für alle Folgen von disjunkten (beobachtbaren) Ereignissen $(A_i)_{i\in\mathbb{N}} \subseteq \mathcal{A}$, also $A_i \cap A_j = \varnothing$ für $i \neq j$, gilt

$$\mu\left[\bigsqcup_{i\in\mathbb{N}} A_i\right] = \sum_{i\in\mathbb{N}} \mu[A_i]. \quad (\sigma\text{-Additivität})$$

Falls das Mass μ zusätzlich *normiert* ist, also $\mu[\Omega] = 1$ gilt, so nennen wir das Mass auch *Wahrscheinlichkeitsmass* oder *(Wahrscheinlichkeits-)Verteilung*.

Wir definieren hier das Mass im allgemeinen Sinn, da wir später, wenn wir diskrete und stetige Verteilungen anschauen, das *Zählmass* sowie das *Lebesguemass* brauchen werden, welche nicht normiert sind. Wir repetieren kurz deren wichtigsten Eigenschaften, ohne auf genaue Details aus der Masstheorie (wie zum Beispiel der Existenz des Lebesguemasses) einzugehen.

Beispiel 2.2. (Zählmass) Sei Ω eine Menge. Sei $\mathcal{A} = \mathcal{P}(\Omega)$ die Potenzmenge (welche eine σ-Algebra ist). Dann ist die Abbildung

$$\nu : \mathcal{A} \to [0, \infty], \quad \nu(A) = |A| := \text{'Anzahl Elemente in A'}$$

ein Mass. Wir nennen μ das Zählmass (auf Ω). Es ist klar, dass ν ein Wahrscheinlichkeitsmass ist genau dann wenn $|\Omega| = 1$. Wir werden meistens $\varnothing \neq \Omega \subseteq \mathbb{R}$ abzählbar verwenden, sodass dann $\nu(\Omega) = \nu(\mathbb{R}) \in \mathbb{N} \cup \{\infty\}$. Integrale bezüglich dem Zählmass sind nichts anderes als Summen! Sei A eine abzählbare Menge und $f : \mathbb{R} \to \mathbb{R}$ eine Funktion. Dann gilt (vorausgesetzt die Summe ist absolut konvergent)

$$\int_A f(x) d\nu(x) = \sum_{x\in A} f(x) \tag{2.1}$$

Beispiel 2.3. (Lebesguemass) Das Lebesguemass λ ist das eindeutige Mass auf $\mathcal{B}(\mathbb{R})$ sodass für alle Intervalle $(a, b] \subseteq \mathbb{R}$ (welche gemäss Beispiel 2.1 in der Borel $\sigma-$Algebra enthalten sind) gilt

$$\lambda\left[(a, b]\right] = b - a.$$

Für Integrale bezüglich des Lebesgueintegrals können wir folgenden Fakt verwenden: Falls $f : \mathbb{R} \to \mathbb{R}$ Riemann-integrierbar, so ist f Lebesgue-integrierbar und es gilt

$$\int_{\mathbb{R}} f(x) d\lambda(x) = \int_{-\infty}^{\infty} f(x) dx. \tag{2.2}$$

Wir können also (in allen nützlichen Fällen) die Regeln aus der Analysis verwenden.

Definition 2.3: Wahrscheinlichkeitsraum

Sei $(\Omega, \mathcal{A})$ ein Ereignisraum und P ein Wahrscheinlichkeitsmass auf $\mathcal{A}$, so nennen wir das Tripel $(\Omega, \mathcal{A}, P)$ einen Wahrscheinlichkeitsraum.

Wir haben in den Übungsstunden viele Beispiele von Wahrscheinlichkeitsräumen gesehen. Dieses kurze Skript soll nur die Theorie zusammenfassen, weshalb wir an dieser Stelle keine weiteren Beispiele geben, sondern nur auf die Notizen aus den Übungsstunden verweisen.

Definition 2.4: Zufallsvariable

Sei $(\Omega, \mathcal{A})$ ein Ereignisraum. Ein Abbildung $X : \Omega \to \mathbb{R}$ wird Zufallsvariable genannt, falls für alle $B \in \mathcal{B}(\mathbb{R})$ gilt $X^{-1}(B) \in \mathcal{A}$.

Falls nun $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable ist, so ist insbesondere für ein beliebiges $x \in \mathbb{R}$ die Menge

$$A := \{X \leq x\} := \{\omega \in \Omega | X(\omega) \leq x\} = X^{-1}((-\infty, x]) \in \mathcal{A}$$

beobachtbar, das heisst wir können ihr eine Wahrscheinlichkeit $P[A]$ zuordnen. Analog können wir für eine beliebige Menge $B \in \mathcal{B}(\mathbb{R})$ benutzen, dass

$$A := \{X \in B\} := X^{-1}(B) \in \mathcal{A}$$

gilt, um der Menge A eine Wahrscheinlichkeit $P[A] = P[X^{-1}(B)]$ zuzuordnen.

Wir kommen nun zu der zentralen Aussage für, welche wir für die Definition der Verteilung verwenden.

Proposition 2.5: Pushforward, Bildmass

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable. Dann ist die Abbildung $\mu : \mathcal{B}(\mathbb{R}) \to [0, 1]$, $\mu[B] = P[X \in B]$ ein Wahrscheinlichkeitsmass.

Andere Notationen für μ sind $\mu_X, P_X, P \circ X^{-1}, X \circ P$. Wir verwenden die Notation μ_X für dieses Skript.

Beweis. Siehe Notizen Woche 2. Beachte, dass wir die Notation P_X benutzt haben und nur diskrete Zufallsvariablen betrachtet haben. Der Beweis für allgemeine Zufallsvariablen funktioniert aber genau gleich. □

Eine wichtige Eigenschaft des Bildmasses ist der Transformationssatz: Falls mindestens eines der beiden folgenden Integrale existiert, so gilt

$$\int_\Omega f(X)\ dP = \int_\mathbb{R} f\ d\mu_X \tag{2.3}$$

Siehe auch: Maran's Notizen [Theorem 8].

Definition 2.6: Verteilung

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable. Das induzierte Wahrscheinlichkeitsmass μ_X wird *Verteilung von* X (unter P) genannt.

Durch $(\Omega, \mathcal{A}, P)$ und eine Zufallsvariable X erhalten wir nun ein neues Tripel $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$, also einen neuen Wahrscheinlichkeitsraum. Dies erklärt, wieso das Bildmass auf Englisch auch *pushforward measure* genannt wird.

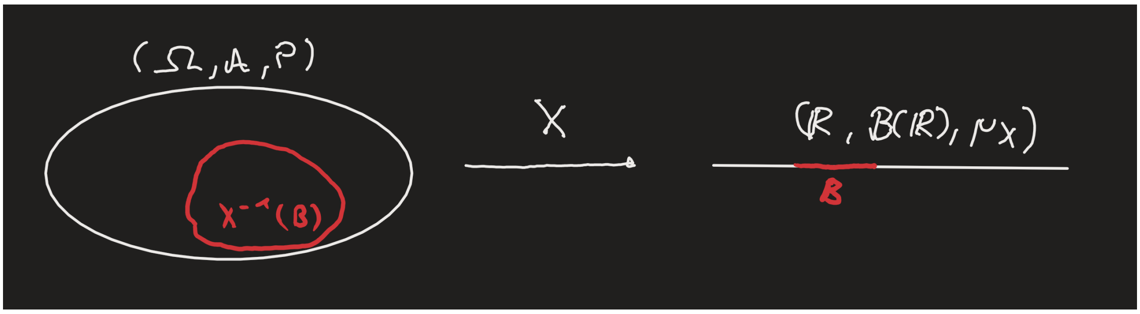

Abbildung 1: Aus dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ wird durch X mittels Bildmass ein neuer Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ erzeugt.

Oftmals interessieren wir uns nur für die Verteilung μ_X einer Zufallsvariablen, ohne überhaupt den Raum $(\Omega, \mathcal{A}, P)$ oder die Abbildung X zu definieren. In diesem Fall machen die mathematischen Objekte X und P nur Sinn, wenn sie gemeinsam vorkommen, also zum Beispiel in $P[X \in \{0, 1, 2\}] = \mu_X[\{0, 1, 2\}]$. Wenn wir also schreiben "Sei X eine Zufallsvariable mit Verteilung Bernoulli(0.3)", so meinen wir "Es existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine Zufallsvariable X sodass $P[X = 1] = 0.3$ und $P[X = 0] = 0.7$ (und 0 sonst)". Das ist aber äquivalent zu $\mu_X[\{1\}] = 0.3$ und $\mu_X[\{0\}] = 0.7$ (und $\mu_X[\mathbb{R} \backslash \{0, 1\}] = 0$). Wir sehen also, dass es in der Tat nicht nötig ist, $(\Omega, \mathcal{A}, P)$ und X genauer zu beschreiben!

Wir unterscheiden zwischen drei Arten von Verteilungen:

- diskrete Verteilungen (Kapitel 3 und 4.4)

- stetige Verteilungen (Kapitel 4)
- sonstige Verteilungen (Kapitel 5)

3 Diskrete Verteilungen I

Wir starten mit diskreten Verteilungen.

Definition 3.1: Diskrete Zufallsvariable und Verteilung

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Wir sagen, dass X diskret ist, falls eine abzählbare Menge $\mathcal{X} \subseteq \mathbb{R}$ existiert, sodass $P[X \in \mathcal{X}] = 1$.
Eine Verteilung μ heisst diskret, falls sie durch eine diskrete Zufallsvariable entstanden ist, also $\mu = \mu_X$ für eine diskrete Zufallsvariable X gilt.

Mithilfe der σ-Additivität können wir nun diskrete Masse sehr einfach charakterisieren: Da $\mu_X[\mathcal{X}^c] = P[X \in \mathcal{X}^c] = 0$, gilt für alle $B \in \mathcal{B}(\mathbb{R})$

$$\mu_X[B] = \mu[\underbrace{B \cap \mathcal{X}}_{\text{abzählbar}}] = \sum_{x \in B \cap \mathcal{X}} \mu_X[\{x\}].$$

Definition 3.2: Wahrscheinlichkeitsfunktion

Sei X eine diskrete Zufallsvariable mit Verteilung μ_X, dann definieren wir die Wahrscheinlichkeitsfunktion von X als

$$p_X : \mathcal{X} \to [0,1], x \mapsto \mu_X[\{x\}] = P[X = x].$$

Andere Bezeichnungen für p_X sind diskrete Dichte, pmf (probability mass function), Massefunktion oder Gewichtsfunktion. Für jede diskrete Verteilung finden wir eine Wahrscheinlichkeitsfunktion und umgekehrt. Diskrete Verteilungen werden fast ausschliesslich durch ihr Wahrscheinlichkeitsfunktion angegeben. Es gilt nun für das Zählmass ν

$$\|p_X\|_{L^1(\mathcal{X},\nu)} := \int_{\mathcal{X}} p_X(x) d\nu = \sum_{x \in \mathcal{X}} p_X(x) = \sum_{x \in \mathcal{X}} P[X = x] = 1 \tag{3.1}$$

Beispiel 3.1. Wir sagen, dass X Poissonverteilung hat mit Parameter $\lambda > 0$, falls für $x \in \mathbb{N}_0$

$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

Wir schreiben $X \sim Poisson(\lambda)$ was soviel bedeutet wie μ_X korrespondiert zur der genannten Wahrscheinlichkeitsfunktion p_X. In diesem Beispiel gilt nun $\mathcal{X} = \mathbb{N}_0$. Es fällt wiederum auf, dass weder $(\Omega, \mathcal{A}, P)$, noch X (als Abbildung) gegeben sind.

Es ist möglich (siehe Kapitel 4.2), dass es zwei (oder noch mehr) Wahrscheinlichkeitsmasse P, Q auf dem gleichen Ereignisraum $(\Omega, \mathcal{A})$ hat. Falls $P \neq Q$, also wenn es ein Ereignis $A \in \mathcal{A}$ mit $P[A] \neq Q[A]$ gibt, dann muss im Allgemeinen das von P induzierte Mass $B \to P[X \in B]$ und das von Q induzierte Mass $B \to Q[X \in B]$ nicht gleich sein. (Die Notation μ_X macht hier also keinen Sinn mehr; das ist der Vorteil der Notation $P \circ X^{-1}$ bzw. $Q \circ X^{-1}$).

Beispiel 3.2. Alice und Bob werfen je eine Münze. Dazu sei $\Omega = \{\text{Kopf}, \text{Zahl}\}$ (und $\mathcal{A} = \mathcal{P}(\Omega)$). Alice behauptet, dass ihr Münze fair ist, also $P[\{\text{Kopf}\}] = P[\{\text{Zahl}\}] = 0.5$. Bob findet heraus, dass seine Münze "Kopf" bevorzugt, indem $Q[\{\text{Kopf}\}] = 0.6$ und $Q[\{\text{Zahl}\}] = 0.4$. Es gilt also $P \neq Q$. Sei nun $X(\text{Kopf}) = 1$ und $X(\text{Zahl}) = 0$, dann gilt

$$p_X(0) = P[X = 0] = P[\{\text{Zahl}\}] = 0.5 = [\{\text{Kopf}\}] = P[X = 1] = p_X(1),$$

sowie

$$q_X(0) = Q[X = 0] = Q[\{\text{Zahl}\}] = 0.4, \quad q_X(1) = Q[X = 1] = Q[\{\text{Kopf}\}] = 0.6.$$

Insbesondere ist die Verteilung unter P nicht gleich der Verteilung unter Q. Um Missverständnisse zu vermeiden, geben wir in solchen Fällen an, unter welchem Mass welche Verteilung entstanden ist, also

$$X \overset{P}{\sim} Bernoulli(0.5) \quad \text{und} \quad X \overset{Q}{\sim} Bernoulli(0.6).$$

4 Absolut stetige Verteilungen

4.1 Realisierungen (Samples)

Neben der Theorie von Wahrscheinlichkeitsräumen und Verteilungen ist besonders mit Blick in Richtung Statistik auch die Anwendungsseite der Stochastik interessant. Während wir in der Theorie von einer Bernoulli(0.3) Verteilung μ_X sprechen, treffen wir in der Praxis eher etwas auf etwas von der Form

$$(0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0),$$

also auf 14 (unabhängige) *Realisierungen* einer Bernoulli(0.3) Verteilung.

In der Statistik befassen wir uns also genau mit dem Gegenteil von dem, was wir in der Wahrscheinlichkeitstheorie machen: Wir beginnen mit Realisierungen und versuchen Rückschlüsse auf unser Modell zu ziehen.

4.2 Absolute Stetigkeit

Beispiel 4.1. Angenommen wir messen während einer Woche (Montag bis Samstag) täglich die Anzahl Kunden, die zwischen 09:00 und 12:00 in einen Einkaufladen gehen. Wir erhalten folgendes Resultat:

$$(745, 692, 715, 1012, 557, 545).$$

Wir erhalten ein arithmetisches Mittel von 711 Kunden pro Tag. Wir überlegen uns folgende Modelle:

- $X \overset{P}{\sim} Bin(1000, 0.711)$
- $X \overset{Q}{\sim} Poisson(711)$

Beide Modelle erfüllen zwar $\mathbb{E}[X] = 711$ (ohne hier zu beschreiben was $\mathbb{E}$ überhaupt bedeutet), jedoch macht die erste Verteilung keinen Sinn für diese Daten: Es ist unmöglich (mit Wahrscheinlichkeit 0), dass eine binomialverteilte Zufallsvariable mit Parametern $n = 1000$ und $p \in [0, 1]$ den Wert 1012 annimmt! Dieses Sample kann also nicht aus einer Bin(1000,0.711) Verteilung entstanden sein.

Das Konzept, welches hier dahinter steckt, nennt sich absolute Stetigkeit: Seien P und Q zwei Wahrscheinlichkeitsmasse auf dem gleichen Ereignisraum $(\Omega, \mathcal{A})$. Angenommen wir nehmen ein beliebiges Sample $\omega \in \Omega$ erzeugt durch Q. Falls anhand dieses Samples ω die Möglichkeit besteht, dass wir mit Sicherheit ausschliessen können, dass es unter P erzeugt wurde, so ist Q nicht absolut stetig bezüglich P.

Im Beispiel 4.1 wäre also $X(\omega) = 1012$ ein Sample welches unter Q entstanden sein könnte (Q wählt ω aus), da $Q[X = 1012] = e^{-711}\frac{711^{1012}}{1012!} > 0$. Jedoch gilt $P[X = 1012] = 0$, sodass das Sample nicht durch P generiert worden sein kann. Also ist Q nicht absolut stetig bezüglich P.

Umgekehrt ist aber P absolut stetig bezüglich Q, da jeder Wert $x = X(\omega) \in \{0, 1, 2, \dots, 1000\}$, welcher durch eine Wahl von ω durch P generiert werden kann auch durch die Wahl von Q hätte generiert werden können. Wir können also nie mit Sicherheit sagen, dass tatsächlich P gewählt hat.

Im Beispiel 3.2 von Alice und Bob haben wir auch zwei verschiedene Wahrscheinlichkeitsmasse auf dem gleichen Raum $\Omega = \{\text{Kopf}, \text{Zahl}\}$. Sowohl bei $\omega = \text{Kopf}$ als auch bei $\omega = \text{Zahl}$ können wir nicht sagen, welche Münze geworfen wurde. P ist also absolut stetig bezüglich Q und Q ist absolut stetig bezüglich P. Wir sagen, dass P und Q äquivalent sind.

Definition 4.1: Absolute Stetigkeit

Sei $(\Omega, \mathcal{A})$ ein Ereignisraum und P, Q zwei (beliebige) Masse. Wir sagen Q ist *absolut stetig* bezüglich P und schreiben $Q \ll P$ falls

$$\forall A \in \mathcal{A} : P(A) = 0 \implies Q(A) = 0.$$

Die zwei Masse heissen *äquivalent*, wenn sowohl $P \ll Q$ als auch $Q \ll P$ gilt. Wir schreiben $P \approx Q$.

Falls $Q \ll P$, so sagen wir auch dass das Mass P das Mass Q *dominiert*. Es lässt sich leicht überprüfen, dass $\approx$ eine Äquivalenzrelation auf der Menge der Masse auf $(\Omega, \mathcal{A})$ definiert.

Beispiel 4.2. Sei $(\Omega, \mathcal{A})$ ein messbarer Raum ν ein Mass auf $(\Omega, \mathcal{A})$ Definiere für eine nicht-negative Funktion $f \in L^1(\Omega, \nu)$ und für alle $A \in \mathcal{A}$

$$\mu[A] := \int_\Omega f(\omega)\mathbb{1}_A(\omega)\ d\nu(\omega) \overset{Notation}{=} \int_A f\ d\nu.$$

Behauptung 1: μ ist ein Mass auf $\mathcal{A}$. Da per Definition $\mu[\Omega] = \int_\Omega f(\omega) d\nu(\omega)$ folgt daraus, dass μ ein Wahrscheinlichkeitsmass ist genau dann wenn $\|f\|_{L^1(\Omega,\nu)} = 1$.

Beweis. Wir zeigen die Axiome in Definiton 2.2.

- Da $f \geq 0$, ist $\mu[A] \geq 0$.
- $\mu[\varnothing] = \int_\Omega f(\omega)\mathbb{1}_\varnothing(\omega) d\nu(\omega) = \int_\Omega 0\ d\nu(\omega) = 0$
- Falls $(A_i)_{i\in\mathbb{N}} \subseteq \mathcal{A}$ disjunkt, dann gilt mit dem Satz der monotonen Konvergenz

$$\begin{aligned}
\mu\left[\bigsqcup_{i\in\mathbb{N}} A_i\right] &= \int_\Omega f(\omega)\mathbb{1}_{\bigsqcup_{i\in\mathbb{N}} A_i}(\omega) d\nu(\omega) \\
&= \int_\Omega \sum_{i\in\mathbb{N}} f(\omega)\mathbb{1}_{A_i}(\omega) d\nu(\omega) \\
&= \sum_{i\in\mathbb{N}} \int_\Omega f(\omega)\mathbb{1}_{A_i}(\omega) d\nu(\omega) \\
&= \sum_{i\in\mathbb{N}} \mu[A_i].
\end{aligned}$$

□

Behauptung 2: $\mu \ll \nu$.

Beweis. Sei $A \in \mathcal{A}$ mit $\nu[A] = 0$. Dann gilt $\mathbb{1}_A = 0$ ν-fast überall, sodass

$$\mu[\mathcal{A}] = \int_\Omega f(\omega)\mathbb{1}_A(\omega) d\nu(\omega) = \int_\Omega f(\omega) \cdot 0\ d\nu(x) = 0.$$

Also wird in der Tat μ von ν dominiert. □

Das Beispiel zeigt, dass wir für ein gegebenes Mass ν und eine nicht-negative Funktion f mit $\|f\|_{L^1(\Omega,\nu)} = 1$ eine Wahrscheinlichkeitsverteilung μ auf $(\Omega, \mathcal{A})$ konstruieren können. In Kapitel 4.4 schauen wir uns an, was passiert wenn wir für ν das Zählmass wählen und in Kapitel 4.5 wählen wir für ν das Lebesguemass.

4.3 Radon-Nikodym Theorem

Wir haben in Beispiel 4.2 gesehen, dass wir für ein Mass ν auf $(\Omega, \mathcal{A})$ und eine integrierbare, nicht-negative Funktion f eine absolut stetige Verteilung $\mu \ll \nu$ finden können durch

$$\mu[A] = \int_\Omega f(\omega)\mathbb{1}_A(\omega)\ d\nu(\omega), \quad \forall A \in \mathcal{A}. \tag{4.1}$$

Falls das Mass ν σ-endlich ist, so besagt das Radon-Nikodym, dass dies die Menge aller Verteilungen ist, welche durch ν dominiert werden.

Theorem 4.2: Radon-Nikodym

Sei $(\Omega, \mathcal{A}, \nu)$ ein Massraum mit ν σ-endlich. Sei μ ein Wahrscheinlichkeitsmass auf $\mathcal{A}$, dann sind äquivalent

- $\mu \ll \nu$
- Es existiert eine nicht-negative Funktion $f \in L^1(\Omega, \nu)$ mit $\|f\|_{L^1(\Omega,\nu)} = 1$, sodass μ von der Form 4.1 ist.

Die Funktion f wird auch Radon-Nikodym-Ableitung genannt und als $f = \frac{d\mu}{d\nu}$ geschrieben. Dies ist naheliegend unter dem Fakt, dass $\forall A \in \mathcal{A}$

$$\int_A d\mu = \mu[A] = \int_A f\ d\nu = \int_A \frac{d\mu}{d\nu}\ d\nu.$$

Zudem gilt für eine μ-integrierbare Funktion g

$$\int_\Omega g\ d\mu = \int_\Omega g\ \frac{d\mu}{d\nu}\ d\nu. \tag{4.2}$$

4.4 Diskrete Verteilungen II

Sei $\mathcal{X} \subseteq \mathbb{R}$ abzählbar und ν das Zählmass auf $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$. Dann ist $(\mathcal{X}, \mathcal{P}(\mathcal{X})\nu)$ ein messbarer Raum mit ν σ-endlich. Wir brauchen eine nicht-negative Funktion mit $\|f\|_{L^1(\mathcal{X},\nu)} = 1$. Dies ist gemäss (2.1) äquivalent zu

$$\sum_{x\in\mathcal{X}} f(x) = 1.$$

Das ist genau die Eigenschaft (3.1), welche Wahrscheinlichkeitsfunktionen (diskrete Dichten) definiert. Eine Wahrscheinlichkeitsfunktion ist also nichts anderes als eine Radon-Nikodym-Ableitung bezüglich dem Zählmass!

Beispiel 4.3. (Fortsetzung Beispiel 3.2) Wir schauen uns die Münze von Bob an, welche unter X folgende Verteilung hat:

$$\mathcal{X} = \{0,1\}, \quad \mu_X[\{0\}] = 0.4, \quad \mu_X[\{1\}] = 0.6.$$

Diese Verteilung ist absolut stetig bezüglich dem Zählmass auf $\mathcal{X} = \{0,1\}$. Gemäss Theorem 4.2 existiert $f \in L^1(\{0,1\},\nu)$ mit $f \geq 0$, sodass

$$\forall B \in \mathcal{P}(\{0,1\}): \ \mu_X[B] = \int_B f \ d\nu = \sum_{x \in B} f(x).$$

Wir sehen, dass $f(0) = 0.4$ und $f(1) = 0.6$ gelten muss; also ist $f = q_X$.

4.5 Stetige Verteilungen

Sei $(\Omega, \mathcal{A}, \nu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ die reellen Zahlen mit dem Lebesguemass λ, welches σ-endlich ist. Wir suchen eine nicht-negative Funktion f mit $\|f\|_{L^1(\mathbb{R},\lambda)} = 1$. Dann finden wir durch

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad \mu[B] = \int_B f(x) d\lambda(x)$$

eine Wahrscheinlichkeitsverteilung, welche absolut stetig bezüglich dem Lebesguemass ist.

Definition 4.3: Stetige Zufallsvariable und Verteilung

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Wir sagen, dass X *stetig* ist, falls eine nicht-negative Funktion f mit $\|f\|_{L^1(\mathbb{R},\lambda)} = 1$ existiert, sodass

$$\forall B \in \mathcal{B}(\mathbb{R}): \quad P[X \in B] =: \mu_X[B] = \int_B f(x) d\lambda(x)$$

Die entsprechende Verteilung $\mu_X = P \circ X^{-1}$ nennen wir dann stetig.

Gemäss dem Radon-Nikodym Theorem 4.2 ist dies äquivalent zu $\mu_X \ll \lambda$. Die Funktion f nennen wir die Dichte, Dichtefunktion oder pdf (probability density function) von X. Falls wir mehrere (stetige) Zufallsvariablen haben, schreiben wir auch f_X um klarzustellen, um welche Dichte es sich handelt. Die Dichte ist die Radon-Nikodym-Ableitung bezüglich Lebesguemass $f = \frac{d\mu_X}{d\lambda}$.

5 Sonstige Verteilungen

Nicht jede Verteilung ist absolut stetig bezüglich dem Zähl- oder Lebesguemass. Wir haben in den Serien die Cantor-Verteilung μ_C kennengelernt.

Beispiel 5.1. (Cantor Verteilung) Sei C die Cantormenge und μ_C die Cantorverteilung. Dann ist $\mu_C[\{x\}] = 0$ für alle $x \in \mathbb{R}$. Dementsprechend kann (wegen der σ-Additiviät von μ_C) keine abzählbare Menge $\mathcal{X}$ mit $\mu_C[\mathcal{X}] = 1$ existieren. Also ist die Cantorverteilung nicht diskret. Andererseits hat die Cantormenge Lebesguemass $\lambda[C] = 0$. Also gilt für jede nicht-negative Funktion $f \in L^1(\mathbb{R}, \lambda)$

$$\int_C f(x) d\lambda(x) = 0 \neq 1 = \mu_C[C].$$

Also kann die Cantorverteilung nicht absolut stetig bezüglich dem Lebesguemass sein. μ_C ist also auch keine stetige Verteilung.

6 Erwartungswerte

Der Vorteil von diskreten und stetigen Verteilungen ist, dass wir Integrale mit bekannten Methoden berechnen können, wie in (2.1) und (2.2) schon angedeutet wurde.

Definition 6.1: Erwartungswert

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Wir definieren den *Erwartungswert* von X als (falls das Integral existiert)

$$\mathbb{E}[X] := \int_\Omega X \, dP \tag{6.1}$$

Falls $g : \mathbb{R} \to \mathbb{R}$ eine messbare Abbildung ist, so ist $g(X)$ wieder ein Zufallsvariable und dementsprechen gilt in diesem Fall wieder (falls das Integral existiert)

$$\mathbb{E}[g(X)] = \int_\Omega g(X) \, dP.$$

6.1 Diskrete Erwartungswerte

Falls die Zufallsvariable X diskret ist, so existiert eine abzählbare Menge $\mathcal{X}$ sodass die Verteilung μ_X absolut stetig bezüglich dem Zählmass ν auf $\mathcal{X}$ ist. Also existiert eine Radon-Nikodym-Ableitung (nämlich die Wahrscheinlichkeitsfunktion) p_X sodass

$$\forall B \in \mathcal{P}(\mathcal{X}) : \quad \mu_X[B] = \int_B p_X(x) d\nu(x) = \sum_{x \in B} p_X(x).$$

Nehmen wir nun den Transformationssatz (2.3) des Bildmasses zu Hilfe erhalten wir einen einfachen Ausdruck für den Erwartungswert von X:

$$\mathbb{E}[X] = \int_\Omega X(\omega) dP(\omega) \overset{(2.3)}{=} \int_\mathcal{X} x \, d\mu_X(x) \overset{(4.2)}{=} \int_\mathcal{X} x \cdot p_X(x) d\nu(x) = \sum_{x \in \mathcal{X}} x \, p_X(x).$$

Dies ist der Ausdruck, den wir schon lange kennen und immer verwendet haben (ohne überhaupt zu wissen warum). Die Intuition hinter dem Erwartungswert diskreter Verteilungen ist der Schwerpunkt mehrerer unendlich kleinen Massenpunkte, so wie wir das aus der Physik kennen: “$x \cdot p_X(x)$ ” bedeutet “Ort des Gewichtes multipliziert mit der Masse des entsprechenden Gewichtes ”.

Falls wir wiederum eine Transformation $g(X)$ betrachten gilt

$$\mathbb{E}[g(X)] = \int_\Omega g(X)dP \overset{(2.3)}{=} \int_\mathcal{X} g \; d\mu_X \overset{(4.2)}{=} \int_\mathcal{X} g \cdot p_X \; d\nu = \sum_{x \in \mathcal{X}} g(x)p_X(x).$$

6.2 Stetige Erwartungswerte

Falls die Zufallsvariable X stetig ist, so ist die Verteilung μ_X absolut stetig bezüglich dem Lebesguemass λ. Wir finden also eine Radon-Nikodym-Ableitung (nämlich die Dichte) f_X sodass

$$\forall B \in \mathcal{B}(\mathbb{R}) : \quad \mu_X[B] = \int_B f_X(x)d\lambda(x).$$

Nehmen wir wiederum den Transformationssatz (2.3) des Bildmasses zu Hilfe erhalten wir einen einfachen Ausdruck für den Erwartungswert von X:

$$\mathbb{E}[X] = \int_\Omega X \; dP \overset{(2.3)}{=} \int_\mathbb{R} x \; d\mu_X(x) \overset{(4.2)}{=} \int_\mathbb{R} x \cdot f_X(x)d\lambda(x).$$

Dies ist der Ausdruck, den wir schon lange kennen und immer verwendet haben (ohne überhaupt zu wissen warum). Die Intuition hinter dem Erwartungswert stetiger Verteilung ist der Schwerpunkt eines Körpers, so wie wir das aus der Physik kennen: “$x \cdot f_X(x)$ ” bedeutet “Ort der Masse multipliziert mit der Dichte der Masse an diesem Ort ”.

Falls wir wiederum eine Transformation $g(X)$ betrachten gilt

$$\mathbb{E}[g(X)] = \int_\Omega g(X)dP \overset{(2.3)}{=} \int_\mathbb{R} g \; d\mu_X \overset{(4.2)}{=} \int_\mathbb{R} g \cdot f_X \; d\lambda.$$

7 Zusammenfassung

Wir haben gesehen, dass diskrete und stetige Verteilungen das gleiche Konzept der absoluten Stetigkeit erfüllen. Sie unterscheiden sich nur durch das dominierende Mass. Wir haben gesehen, woher die Definition des Erwartungswertes aus der Vorlesung kommt.

Eine Liste der wichtigsten diskreten und stetigen Verteilungen sowie einige derer Eigenschaften wie Erwartungswert und Varianz findest du hier.