

# Einführung in die Wahrscheinlichkeitstheorie und Statistik

Ulrich Bunke

## Inhaltsverzeichnis

<b>1</b>	<b>Endliche Wahrscheinlichkeitsräume</b>	<b>2</b>
1.1	Was ist Wahrscheinlichkeit . . . . .	2
1.2	Axiomatik für endliche $W$ -Räume . . . . .	7
1.3	Die Kategorie der endlichen $W$ -Räume . . . . .	10
1.4	Zufallsvariable und Erwartungswert . . . . .	15
1.5	Urnenmodelle - Kombinatorische Abzählresultate . . . . .	17
1.6	Wahrscheinlichkeitstheoretische Anwendungen . . . . .	19
1.7	Bedingte Wahrscheinlichkeit . . . . .	23
1.8	Unabhängigkeit . . . . .	33
<b>2</b>	<b>Nicht notwendig endliche <math>W</math>-Räume</b>	<b>40</b>
2.1	Abzählbare $W$ -Räume . . . . .	40
2.2	Allgemeine $W$ -Räume . . . . .	49
2.3	Konstruktion von $W$ -Maßen . . . . .	64
2.4	Sprechweisen für den Umgang mit Zufallsvariablen . . . . .	73
2.5	Dichtefunktionen . . . . .	75
2.6	$0 - 1$ - Gesetze . . . . .	78
<b>3</b>	<b>Analytische Grundlagen</b>	<b>89</b>

3.1	Abschätzungen . . . . .	89
3.2	Konvergenz von Folgen von Zufallsvariablen . . . . .	92
3.3	Gesetze der großen Zahl . . . . .	97
3.4	Verteilungsfunktionen . . . . .	104
3.5	$W$ -Maße auf Gruppen und reellen Vektorräumen . . . . .	111
3.6	Gaussche Maße . . . . .	123
3.7	Zentraler Grenzwertsatz . . . . .	131
<b>4</b>	<b>Schätztheorie</b>	<b>135</b>
4.1	Beschreibung des Problems und der grundlegenden Begriffe . . . . .	135
4.2	Schätzung der Erfolgswahrscheinlichkeit im Bernoulli Experiment . . . . .	144
4.3	Schätzung von Mittelwert und Varianz . . . . .	149
4.4	Taxiproblem . . . . .	153
4.5	Die Informationsungleichung . . . . .	157
4.6	$MLS$ und Konsistenz . . . . .	163
4.7	Lebensdauer . . . . .	169
4.8	Tests . . . . .	171
4.9	Mehr Theorie . . . . .	176

# 1 Endliche Wahrscheinlichkeitsräume

## 1.1 Was ist Wahrscheinlichkeit

betrachte Vorgang, Experiment, Frage mit ungewissen Ausgang

- Abstraktion durch den Begriff: *statistisches Experiment*

Gesamtheit der möglichen Ergebnisse oder Antworten

- Abstraktion durch den Begriff: *elementares Ereigniss*

Beispiele:

- Teste einen Halbleiterchip auf Funktion: elementare Ereignisse -  $\{ja, nein\}$
- Zähle Autos, die innerhalb von 20 min eine Stelle passieren: elementare Ereignisse -  $\mathbb{N}$
- Bestimme den Zeitpunkt, zu welchem ein ausgeleerter Abfallcontainer überläuft: elementare Ereignisse -  $[0, \infty)$

betrachten für die Vorlesung in der Regel sehr einfache Modellsituationen wie Würfeln, Kugeln ziehen oder Münzen werfen, um die Begriffe zu erklären

Ziel: möchte meine Information über das Auftreten oder Nichtauftreten eines Ereignisses bei einem statistischen Experiment quantitativ modellieren

- Abstraktion durch den Begriff: *Wahrscheinlichkeit*

**Beispiel 1.1.** Münzwurf:

elementare Ergebnisse:

- $z$  - Zahl
- $w$  - Wappen
- also Menge der elementaren Ereignisse:  $\{z, w\}$
- das Ergebnis  $z$  kann auftreten, muß aber nicht
- Symmetriebetrachtung: bin mir genauso unsicher über  $z$  und  $w$

□

**Beispiel 1.2.** Würfel

Elementare Ergebnisse:

- $\{1, 2, 3, 4, 5, 6\}$

zusammengesetzte Ergebnisse:

- es kommt eine gerade Zahl: modelliert durch die Teilmenge  $\{2, 4, 6\}$
- Symmetriebetrachtung: bin mir genauso unsicher über jedes elementare Ergebnis

- Symmetriebetrachtung: bin mir genau so sicher über das Auftreten eine geraden Zahl wie einer ungeraden

- plausibel: Das Ereignis  $\{2, 4, 6\}$  ist drei mal so wahrscheinlich wie  $\{2\}$

□

Allgemein: Experiment mit möglichem Ergebniss  $\omega$

möchte meine Sicherheit über  $\omega$  quantifizieren durch eine Zahl  $p \in [0, 1]$

-  $p = 0$  - ich bin mir sicher, daß  $\omega$  nicht kommt

-  $p = 1$  - bin mir sicher, daß  $\omega$  kommt

- Was bedeutet die Zahl  $p$ ?

### **Beispiel 1.3. subjektive Interpretation:**

Interpretation durch faire Wette:

- Gewinnrate  $r$  - ich bekomme beim Ergebnis  $w$  das  $r$ -fache meines Einsatzes, sonst nichts

- Empfände die Wette ab Rate  $r$  als fair.

- setze  $p \geq \frac{1}{pr}$ .

Problem: Was bedeutet "als fair empfinden" genau?

- mögliche Antwort: Ich kann das Spiel beliebig oft wiederholen und erwarte dabei, zumindest keinen Verlust zu machen.

Problem: Was bedeutet: "Ich erwarte ..." genau?

Problem: Nicht jedes Spiel ist wiederholbar.

□

### **Beispiel 1.4. Russisches Roulette**

Revolver mit 6er Trommel,

- eine Kugel

- Symmetriebetrachtung: Mit Wahrscheinlichkeit  $1/6$  sterbe ich.

- Was bedeutet das?

- Es gibt hier keine faire Gewinnrate.
- Dies ist ein "Experiment", welches nicht beliebig wiederholt werden kann!
- Kann man die Angst, dabei zu sterben, quantifizieren? Was bedeutet 1/6-der totalen Angst zu haben?
- Hier ist die Interpretation der Wahrscheinlichkeit problematisch.

### Beispiel 1.5. Interpretation durch Häufigkeit

- stelle mir vor, das Experiment mehrfach auszuführen

$$p = \frac{\text{Anzahl des Auftretens des Ereignisses}}{\text{Anzahl der Experimente}}$$

- In der Praxis ist das N-fache Ausführen insgesamt ein neues Experiment, welches durch ein neues Model beschrieben wird dessen Ergebnisraum die Folgen der Ausgänge der Einzelexperimente sind.
- es gibt eine mathematische Konstruktion, ausgehend von einem Modell für einen Versuch ein neues Model für  $N$  unabhängige Versuche zu konstruieren.
- man kann beweisen, daß fast sicher gilt:

$$\frac{\text{Anzahl des Auftretens des Ereignisses}}{N} \xrightarrow{N \rightarrow \infty} p$$

Problem: Es gibt Fälle, wo eine Wiederholung des Experiments undenkbar ist.

□

### Beispiel 1.6. Interpretation ist nicht absolut:

Impfschaden:

- Beobachtung: Bei 10 von  $10^6$  bisherigen Impfungen trat tödlicher Impfschaden ein
- Aus der Sicht des Imstoffherstellers der sich gegen die Schadensersatzforderung absichern will:
- Experiment ist wiederholbar
  - kann versuchen, einen fairen Preis für eine Schadensversicherung zu berechnen

Aus der Sicht eines Politikers, der die Impfung empfehlen will:

- Experiment ist wiederholbar
- kann das Schadensrisiko abwägen gegen die Möglichkeit der Steigerung seines Ansehen infolge der Durchsetzung der Impfung

Aus der Sicht des Impflings:

- Experiment ist nicht wiederholbar
- quantitative Beschreibung der Todesangst ist problematisch
- die Beobachtung des Impfschadens ist nur deshalb ein statistisches Experiment, weil der den schadenauslösende Faktor unbekannt oder bei der Studie einfach nicht erfaßt wurde
- Ich gewichte die Angst, an der Impfung oder an der Krankheit zu sterben, verschieden:
  - "Bill Gates ist Schuld" bzw. "Entscheidung Gottes"

Fazit: Die *Interpretation* der Wahrscheinlichkeit des Auftretens eines Ereignisses ist nicht absolut, sondern hängt von Standpunkt der Betrachtung und anderweitigen Interessen ab.

□

Wir werden in dieser Vorlesung Beispiele betrachten, in welchen die Interpretation der Wahrscheinlichkeit mehr oder weniger Konsens ist.

### **Wahrscheinlichkeitstheorie**

- Festlegung der Axiome
- Bildung von Begriffen
- Formulierung von Behauptungen
- mathematische Beweise
- wird als Teilgebiet der Mathematik verstanden

### **Modellbildung**

- Abbildung einer realen Situation durch die Auswahl eines mathematisches Modell, welches der Axiomatik genügt

- Begründung durch
- Plausibilität
- Symmetriebetrachtung
- in der Vorlesung vornehmlich, um die Begriffe mit der Alltagswelt zu verbinden und anschaulich zu machen

### **Statistik**

- Begründung der Auswahl des Modells durch Beobachtungen
- Schätzen von Parametern
- Auswahl des Modells aus einer Menge von Kandidaten durch Tests

## **1.2 Axiomatik für endliche W-Räume**

$\Omega$  - Menge der elementaren Ergebnisse

$\mathcal{P}_\Omega$  - Menge der zusammengesetzten Ergebnisse

zusammengesetztes Ereignis  $A \in \mathcal{P}_\Omega$  bedeutet: Elementarereignis liegt in  $A$

nennen eine Abbildung  $P : \mathcal{P}_\Omega \rightarrow [0, 1]$  eine *Mengenfunktion*

**Definition 1.7.** *Die Mengenfunktion heißt*

1. *additiv, wenn  $P(A \cup B) = P(A) + P(B)$  for alle  $A, B$  in  $\mathcal{P}_\Omega$  mit  $A \cap B = \emptyset$  gilt.*
2. *normiert, wenn  $P(\Omega) = 1$  gilt.*

für eine normierte additive Mengenfunktion  $P$  gelten

- $P(A^c) = 1 - P(A)$  (hier  $A^c := \Omega \setminus A$  - Komplement von  $A$ )
- $P(\emptyset) = 0$

$\Omega$  - endliche Menge

**Definition 1.8.** *Ein W-Maß auf  $\Omega$  ist eine additive und normierte Mengenfunktion.*

- Interpretation:  $P(A)$  heißt Wahrscheinlichkeit für das Auftreten von  $A$

sei  $P$  ein  $W$ -Maß auf einer endlichen Menge

**Definition 1.9.** *Die Funktion*

$$p : \Omega \rightarrow [0, 1], \quad \omega \mapsto P(\{\omega\}) \tag{1}$$

heißt *Wahrscheinlichkeitsvektor (W-Vektor)*.

W-Maße und W-Vektoren bestimmen sich gegenseitig eindeutig:

-  $p$  ist durch  $P$  bestimmt (siehe (1)):

-  $P$  ist durch  $p$  eindeutig bestimmt:  $P(A) = \sum_{\omega \in A} p(\omega)$  (wegen Additivität)

Eine Funktion  $p : \Omega \rightarrow [0, 1]$  ist ein W-Vektor genau dann wenn  $\sum_{\omega \in \Omega} p(\omega) = 1$  gilt.

Kann W-Maß also durch W-Vektor angeben:

- Betrachtung des W-Vektors führt zu Reduktion der notwendigen Information

-  $P$ :  $2^{|\Omega|}$ -viele Werte, sehr viele Bedingungen (Additivität)

-  $p$ :  $|\Omega|$ -viele Werte, eine Bedingung

**Definition 1.10.** *Ein endlicher Wahrscheinlichkeitsraum (endlicher W-Raum) ist ein Paar  $(\Omega, P)$  aus einer endlichen Menge  $\Omega$  und einem W-Maß  $P$ .*

**Beispiel 1.11.** Gleichverteilung

$\Omega$  - endliche Menge

W-Vektor:  $\text{const}_{|\Omega|^{-1}}$

-  $P(A) = \frac{|A|}{|\Omega|}$

W-theoretische Interpretation: alle elementaren Ereignisse treten mit der selben Wahrscheinlichkeit auf □



**Beispiel 1.12.** Dirac Maß

hier muß  $\Omega$  nicht endlich sein

- fixiere  $\omega \in \Omega$

-  $\delta_\omega : \mathcal{P}_\Omega \rightarrow [0, 1]$

-  $\delta_\omega(A) := \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$

Diracmaß bei  $\omega$

- ist additive und normierte Mengenfunktion

$W$ -theoretische Interpretation: Mit Sicherheit tritt Ergebnis  $\omega$  ein!

$W$ - Vektor:  $p(\omega') = \delta_{\omega, \omega'}$

□

**Beispiel 1.13. Modellbildung** für Münzwurf

reale Situation: werfe Münzen mit Ausgang Zahl oder Wappen

Prozess der Modellbildung besteht aus:

1. Wahl der Menge der elementaren Ereignissen

nehmen  $\Omega = \{z, w\}$  mit offensichtlicher Interpretation

2. Wahl des  $W$ -Maßes

- wir sehen keinen Grund warum Zahl oder Wappen bevorzugt sein sollte

- aus Symmetriegründen Gleichverteilung:

$$P(\{w\}) = P(\{z\}) = 1/2$$

Modell:  $(\{z, w\}, \text{Gleichverteilung})$

3. Interpretation:

- Würde ab einer Rate 2 auf  $w$  wetten

4. Realitätscheck:

- Wenn ich das Experiment  $N$  mal wiederhole, dann erwarte ich ungefähr  $N/2$ -viele Ausgänge "Zahl".

- das werden wie später genauer sagen können

### Simulation:

Benutzen in der Vorlesung  $R$

- kann bei <https://www.r-project.org/> runtergeladen werden

produziere Vektor von Zufallszahlen der Länge 100, gleichverteilt in  $[0, 2]$

bilde Vektor der ganzen Teile

Simulation des Münzwurfes  $z = 0, w = 1$

□

### Beispiel 1.14. Würfel:

1. Wahl der Menge der elementaren Ereignissen

-  $\Omega := \{1, 2, 3, 4, 5, 6\}$

2. Wahl des W-Maßes

- aus Symmetriegründen Gleichverteilung

$P(\{\omega\}) = 1/6$  für  $\omega$  in  $\Omega$

3. Interpretation:

- Würde ab einer Rate 6 auf 1 wetten

- Würde ab einer Rate 3 auf  $\{1, 2\}$  wetten

4. Realitätscheck:

Würfle  $N$ -mal und erstelle Histogramm der Ergebnisse

100 Würfe, erhöhe auf 1000 und 10000

□

## 1.3 Die Kategorie der endlichen W-Räume

$\Omega$  - Menge

$P : \mathcal{P}_\Omega \rightarrow [0, 1]$  - Mengenfunktion

$f : \Omega \rightarrow \Omega'$  - Abbildung von Mengen

**Definition 1.15.** Definiere die Mengenfunktion  $f_*P : \mathcal{P}_{\Omega'} \rightarrow [0, 1]$  durch

$$(f_*P)(A') := P(f^{-1}(A')) .$$

-  $f_*P$  heißt das Bild von  $P$  unter  $f$ .

**Lemma 1.16.** Wenn  $P$  eine additive (oder normierte) Mengenfunktion auf  $\Omega$  ist, dann ist  $f_*P$  eine additive (oder normierte) Mengenfunktion auf  $\Omega'$ .

*Beweis.* Setze  $P' := f_*P$

Normierung:  $P'(\Omega') = P(f^{-1}(\Omega')) = P(\Omega) = 1$

Additivität: sei  $A' \cap B' = \emptyset$

$$\begin{aligned} P'(A' \cup B') &= P(f^{-1}(A' \cup B')) = P(f^{-1}(A') \cup f^{-1}(B')) = P(f^{-1}(A')) + P(f^{-1}(B')) \\ &= P'(A') + P'(B') \end{aligned}$$

da  $f^{-1}(A' \cup B') = f^{-1}(A') \cup f^{-1}(B')$

und

$$f^{-1}(A') \cap f^{-1}(B') = f^{-1}(A' \cap B') = f^{-1}(\emptyset) = \emptyset$$

□

$(\Omega, P)$  - W-Raum

- wenn  $\Omega'$  endlich ist, erhalte neuen W-Raum  $(\Omega', f_*P)$

**Beispiel 1.17.** Gerade und ungerade Würfelergebnisse

$(\Omega, P)$  beschreibe Würfel

- Gleichverteilung auf  $\{1, 2, 3, 4, 5, 6\}$

$\Omega' := \{g, u\}$  - gerade, ungerade

$f : \Omega \rightarrow \Omega'$

$\omega$	1	2	3	4	5	6
$f(\omega)$	u	g	u	g	u	g

$$f^{-1}(\{g\}) = \{2, 4, 6\}$$

$$P'(\{g\}) = P(\{2, 4, 6\}) = 1/2$$

$(\Omega', P')$  beschreibt Würfelexperiment, wenn ich mich nur für die Parität des Ergebnisses interessiere

Wenn das ursprüngliche Modell den Realitätscheck (was immer das ist) bestanden hat, dann auch das abgeleitete. Umgekehrt, erfüllt das abgeleitete Modell den Realitätscheck nicht, dann müßte ich auch das ursprüngliche Modell verwerfen.

□

$(\Omega, P), (\Omega', P')$  - W-Räume

**Definition 1.18.** Ein Morphismus  $f : (\Omega, P) \rightarrow (\Omega', P')$  von W-Räumen ist eine Abbildung von Mengen  $f : \Omega \rightarrow \Omega'$  so daß  $f_*P = P'$  gilt.

erhalten Kategorie der W-Räume **WRaum**

- Objekte: W-Räume  $(\Omega, P)$
- Morphismen: Morphismen
- wir werden später unendliche Räume dazunehmen

haben Vergißfunktoren  $\mathbf{WRaum} \rightarrow \mathbf{Set}, (\Omega, P) \mapsto \Omega$

kategoriale Sprache liefert Begriffe wie :

- Isomorphismus, Monomorphismus, Epimorphismus,
- kartesisches Produkt, Koprodukt

$(\Omega, P)$  in **WRaum**

- $\text{Aut}_{\mathbf{WRaum}}(\Omega, P)$  - Gruppe der Automorphismen

**Beispiel 1.19.** Automorphismen der Gleichverteilung

$(\Omega, P)$  Gleichverteilung auf endlicher Menge

- $\Sigma_\Omega := \text{Aut}_{\mathbf{Set}}(\Omega)$  Gruppe der Bijektionen  $\Omega \rightarrow \Omega$
- z.B.:  $\Sigma_{\{1, \dots, n\}}$  - Gruppe der Permutationen von  $n$  Elementen (Kurznotation  $\Sigma_n$ )

**Lemma 1.20.** Der Vergißfunktork induziert einen Isomorphismus  $\text{Aut}_{\mathbf{WRaum}}(\Omega, P) \xrightarrow{\cong} \Sigma_{\Omega}$ .

*Beweis.* Vergißfunktork liefert Homomorphismus:  $\text{Aut}_{\mathbf{WRaum}}(\Omega, P) \rightarrow \Sigma_{\Omega}$

- Injektivität: klar

- Surjektivität:

- sei  $f$  in  $\Sigma_{\Omega}$

- müssen zeigen:  $f_*P = P$

—  $A$  in  $\mathcal{P}_{\Omega}$  bel.

—  $|f^{-1}(A)| = |A|$ , da  $f$  bijektiv

—  $(f_*P)(A) = P(f^{-1}(A)) = \frac{|f^{-1}(A)|}{|\Omega|} = \frac{|A|}{|\Omega|} = P(A)$

□

$\Omega$  - endliche Menge

$\mathcal{F} \subseteq \text{End}_{\mathbf{Set}}(\Omega)$  eine Menge von Selbstabbildungen von  $\Omega$

**Definition 1.21.** Ein  $W$ -Maß  $P$  auf  $\Omega$  heißt invariant unter  $\mathcal{F}$  falls  $f_*P = P$  gilt für alle  $f$  in  $\mathcal{F}$ .

$\Omega$  - eine Menge

-  $G$  - Untergruppe  $\text{Aut}_{\mathbf{Set}}(\Omega)$

**Definition 1.22.**  $G$  wirkt transitiv auf  $\Omega$  falls für je zwei  $\omega, \omega'$  in  $\Omega$  ein  $g$  in  $G$  mit  $g\omega = \omega'$  existiert .

- äquivalent:  $|G\omega| = |\Omega|$

**Beispiel 1.23.**  $\Omega = \{1, \dots, n\}$

folgende Gruppem wirken transitiv

-  $\Sigma_n$

-  $\langle (1, 2, \dots, n) \rangle$  (zyklische Permutationen)

$\langle (1, 2), (3, 4) \rangle$  wirkt nicht transitiv auf  $\{1, 2, 3, 4\}$

□

folgendes Lemma macht die Redeweise “aus Symmetriegründen gilt...” aus präzise

$\Omega$  - endliche Menge

**Lemma 1.24.** *Wenn  $G$  transitiv wirkt, dann ist die Gleichverteilung ist das einzige  $W$ -Maß auf  $\Omega$ , welches invariant unter  $G$  ist.*

*Beweis.*

sei  $P$  ein  $W$ -Maß, welches invariant unter  $G$  ist

- fixiere  $\omega_0$  in  $\Omega$
- zeige:  $W$ -Vektor ist konstant mit Wert  $P(\{\omega_0\})$
- wegen Normierung ist dieser Wert dann  $\frac{1}{|\Omega|}$
- sei  $\omega$  in  $\Omega$
- es gibt es ein  $g$  in  $G$  mit  $g\omega_0 = \omega$  (wegen Transitivität)
- $\sigma^{-1}(\{\omega\}) = \{\omega_0\}$
- $(\sigma_*P)(\{\omega\}) = P(\{\omega_0\})$

□

**Beispiel 1.25.** Betrachte

$G := \langle (1, 2), (3, 4) \rangle$

wirke auf  $\Omega = \{1, 2, 3, 4\}$  - nicht transitiv

$\Omega/G = \{[1], [3]\}$

- $P_0$  - Gleichverteilung ist invariant
- $P_1$  gegeben durch folgenden  $W$ -Vektor  $p$  ist invariant

$\omega$	1	2	3	4
$p(\omega)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$

$P_0 \neq P_1$

□

**Beispiel 1.26.** Linksschift

$$\Omega = \{0, 1, \dots, n\}$$

$$S : \Omega \rightarrow \Omega, \quad S(\omega) := \begin{cases} \omega - 1 & \omega > 0 \\ 0 & 0 \end{cases}$$

-  $S$  ist nicht invertierbar!

**Lemma 1.27.**  $\delta_0$  ist das einzige  $\{S\}$ -invariante  $W$ -Maß.

*Beweis.*

sei  $P$  invariantes  $W$ -Maß

-  $S_*P = P$  impliziert

-  $S_*^n P = P$

- aber  $S^n = \text{const}_0$

-  $S_*^n P = \delta_0$

- also  $P = \delta_0$

□

## 1.4 Zufallsvariable und Erwartungswert

$(\Omega, P)$  - endlicher  $W$ -Raum

**Definition 1.28.** Eine (numerische) Zufallsvariable ist eine Abbildung  $X : \Omega \rightarrow \mathbb{R}$ .

**Definition 1.29.** Der Erwartungswert von  $X$  ist durch

$$\mathbb{E}(X) := \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$$

definiert.

Die Varianz von  $X$  ist durch

$$\text{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2)$$

definiert.

Interpretation:

- $\mathbb{E}(X)$  - mittlerer Wert von  $X$  gewichtet durch  $P$
- $\text{Var}(X)$  - mittlere quadratische Abweichung von  $X$  von  $\mathbb{E}(X)$

**Bemerkung 1.30.** offensichtliche Rechenregeln für Erwartungswerte und Varianz:

1.  $\mathbb{E}(X) = \sum_{x \in \mathbb{R}} xP(\{X = x\})$
2.  $\mathbb{E}(X + X') = \mathbb{E}(X) + \mathbb{E}(X')$
3.  $\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$
4.  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$
5.  $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

□

zu 5.:

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}((X - \mathbb{E}(X))^2) \\
 &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\
 &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - \mathbb{E}(X)^2
 \end{aligned}$$

**Beispiel 1.31.** Sicheres Ereignis:

$(\Omega, \delta_\omega)$

- $\mathbb{E}(X) = X(\omega)$
- $\text{Var}(X) = 0$

□

**Beispiel 1.32.** ja-nein Experiment

- $\Omega = \{j, n\}$ ,  $p(j) = p$
- Gewinnrate  $r$
- Gewinn bei Einsatz von 1 beschrieben durch Zufallsvariable



$$X : \Omega \rightarrow \mathbb{R}, \quad X(\omega) := \begin{cases} r & j \\ 0 & n \end{cases}$$

- “Erwarteter” Gewinn:  $\mathbb{E}(X) = rp + 0(1 - p)$
- mögliche Definition von “fair”: also “erwartete Gewinn ist 1”
- daß liefert Bedingung  $r = p^{-1}$
- $\text{Var}(X) = r^2p - (rp)^2 = r^2p(1 - p)$
- $\text{Var}(X) \rightarrow 0$  für  $p \rightarrow 0, 1$
- Var mißt die “Streuung von  $X$  um  $\mathbb{E}(X)$ ”

□

**Beispiel 1.33.** Würfel:

$X : \Omega \rightarrow \mathbb{R}$  - Ergebnis selbst

$$\mathbb{E}(X) = \sum_{i=1}^6 i \frac{1}{6} = \frac{7}{2}$$

- Interpretation als mittlere Schrittweite beim Würfelspiel

$$\text{Var}(X) = 2.916667$$

□

## 1.5 Urnenmodelle - Kombinatorische Abzählresultate

habe endliche Grundmenge  $U$  (die Urne),

Anzahl der Elemente:  $u$

Anzahl der gezogenen Elemente:  $n$  Elemente

**Beispiel 1.34.**  $n$  mal Ziehen mit Zurücklegen und Reihenfolge

Menge der Ergebnisse:  $\Omega = U^n$ , Elemente -  $(u_1, \dots, u_n)$

Abzählen:

- 1. Zug  $u$  Möglichkeiten
- 2. Zug  $u$  Möglichkeiten

-  $\vdots$

$$|\Omega| = u^n$$

□

**Beispiel 1.35.**  $n$  mal Ziehen ohne Zurücklegen und mit Reihenfolge

$$\Omega = \{(u_1, \dots, u_n) \in U^n \mid u_i \neq u_j \text{ falls } i \neq j\}$$

Abzählen:

- 1. Zug  $u$  Möglichkeiten
- 2. Zug  $u - 1$  Möglichkeiten
- $\vdots$
- $n$ -ter Zug  $u - n + 1$  Möglichkeiten

-

$$|\Omega| = u(u - 1) \dots (u - n + 1) = \frac{u!}{(u - n)!} \quad (2)$$

□

**Beispiel 1.36.**  $n$  mal Ziehen ohne Zurücklegen und ohne Reihenfolge

sei  $\Omega'$  der Raum der Ergebnisse beim Ziehen ohne Zurücklegen mit Reihenfolge

- erhalte Raum  $\Omega$  durch Vergessen der Reihenfolge
- $\Sigma_n$  wirkt auf  $\Omega'$  frei durch Permutation:  $\sigma(u_1, \dots, u_n) := (u_{\sigma^{-1}(1)}, \dots, u_{\sigma^{-1}(n)})$
- $\Omega \cong \Omega' / \Sigma_n$ ,  $f : \Omega' \rightarrow \Omega$  - Projektion
- Anzahl der Elemente in den Fasern von  $f$ :  $|\Sigma_n| = n!$

$$|\Omega| = \frac{|\Omega'|}{|\Sigma_n|} = \frac{u!}{(n - u)!n!} = \binom{u}{n} \quad (3)$$

□

**Beispiel 1.37.**  $n$  mal Ziehen mit Zurücklegen ohne Reihenfolge

$(\Omega', P')$  - Raum für Ziehen mit Zurücklegen mit Reihenfolge

-  $\Omega$  entsteht durch Vergessen der Reihenfolge

- $\Sigma_n$  wirkt
- $\Omega = \Omega' / \Sigma_n$

-  $\pi : \Omega' \rightarrow \Omega$  Projektion

Problem: Wirkung nicht frei

- $(u, \dots, u)$  wird fixiert von  $\Sigma_n$ -  $[(u, \dots, u)]$  hat ein Element
- $[(u, u, u_3, \dots, u_n)]$  hat  $n!/2!$  Elemente (hier  $u, u_3, \dots, u_n$  paarweise Verschieden)
- Stabilisator ist  $\{e, (1, 2)\} \subseteq \Sigma_n$
- $n!/2 = |\Sigma_n| / |\text{Stabilisator}|$
- die Orbits/Fasern haben also nicht alle dieselbe Anzahl von Elementen

Trick zum Abzählen:

- wähle totale Ordnung auf  $U$
- gibt Bijektion  $U \cong \{1, \dots, u\}$
- in jeder Klasse gibts genau einen geordneten Repräsentanten
- $(u_1, \dots, u_n)$  mit  $u_1 \leq \dots \leq u_n$
- setze  $u'_i := u_i + i - 1$  in  $\{1, u + n - 1\}$
- erreiche damit strikte Ungleichung  $u'_1 < \dots < u'_n$
- $(u'_1, \dots, u'_n)$  bestimmt  $(u_1, \dots, u_n)$  eindeutig durch  $u_i = u'_i - i + 1$
- in der Tat:  $u_{i+1} - u_i = u'_{i+1} - (i + 1) + 1 - u'_i + i - 1 = u'_{i+1} - u'_i - 1 \geq 0$

Erhalte Bijektion  $\Omega \rightarrow \Omega''$  -  $\Omega''$  -Raum der Ergebnisse beim Ziehen ohne Zurücklegen ohne Reihenfolge aus  $u + n - 1$  Elementen

$$|\Omega| = \binom{u + n - 1}{n} \tag{4}$$

□

## 1.6 Wahrscheinlichkeitstheoretische Anwendungen

**Beispiel 1.38.** Lotto 6 as 49

- Elementarereignisse:  $\Omega$  - Ziehung von 6 Elementen ohne Zurücklegen ohne Reihenfolge
- $\Omega = \{A \in \mathcal{P}_{\{1, \dots, 49\}} \mid |A| = 6\}$
- $P$ : Gleichverteilung auf  $\Omega$  (aus Symmetriegründen)
- Anzahl:  $|\Omega| = \binom{49}{6}$

W-theoretisches Modell:  $(\Omega, P)$

Frage: Wie wahrscheinlich ist ein  $x$ -er mit  $x \in \{0, 1, 2, 3, 4, 5, 6\}$ ?

Tippen  $\{1, 2, 3, 4, 5, 6\}$  (alle Tipps gleich gut)

- definieren  $f : \Omega \rightarrow \Omega' := \{0, 1, 2, 3, 4, 5, 6\}$  durch:

- $f^{-1}(\omega) =$  Menge der  $x$ -er
- $f(\omega) = |\omega \cap \{1, 2, 3, 4, 5, 6\}|$

bestimmen Wahrscheinlichkeitsvektor von  $f_*P$

$$- p'(x) = \frac{|f^{-1}(x)|}{|\Omega|}$$

- bestimmen  $|f^{-1}(x)|$ :

- Teilen Urne auf in gute Elemente  $\{1, \dots, 6\}$  und Rest  $\{7, \dots, 49\}$
- für Urbild von  $x$ :  $x$  gute und  $6 - x$  restliche
- also  $x$  aus  $\{1, \dots, 6\}$  und  $6 - x$  aus  $\{7, \dots, 49\}$
- jeweils ohne Zurücklegen und ohne Reihenfolge, benutze (3)
- $|f^{-1}(x)| = \binom{6}{x} \binom{43}{6-x}$
- W-Vektor bei  $x$ :  $p(x) = \frac{\binom{6}{x} \binom{43}{6-x}}{\binom{49}{6}}$

Liste von  $p(x)$

□

**Beispiel 1.39.** Würfeln mit 4 Würfeln,

- Wie hoch ist die Wahrscheinlichkeit für das Auftreten von  $\{1, 2, 3, 4\}$ ?
- Wie hoch ist die Wahrscheinlichkeit für das Auftreten von vier verschiedenen Zahlen?

$$U = \{1, \dots, 6\}$$

- Menge der elementaren Ereignisse  $\Omega$ : - ziehe 4-Mal mit Zurücklegen mit Reihenfolge

-  $\Omega = U^4$

-  $|\Omega| = 6^4$

-  $P$ : Gleichverteilung auf  $\Omega$  aus Symmetriegründen

W-theoretisches Modell:  $(\Omega, P)$

Auftreten von  $\{1, 2, 3, 4\}$ :

- es gibt  $4!$  richtige Ziehungen

-  $P(\{1, 2, 3, 4\}) = \frac{4!}{6^4}$

Auftreten von vier verschiedenen Zahlen:

- entspricht 4 mal Ziehen aus 6er Urne ohne Zurücklegen mit Reihenfolge

- es geht hier nur um die Anzahl der Möglichkeiten

- es  $\frac{6!}{2!}$  viele richtige Fälle (siehe (2))

- also Wahrscheinlichkeit für "4 verschiedene Zahlen":  $\frac{6!}{2!} \cdot \frac{1}{6^4} = \frac{5}{18}$

□

**Beispiel 1.40.** Urne hat  $u$  Kugeln,

- davon  $r$  rote und  $s = u - r$  schwarze

- ziehe  $n$  Kugeln ohne Zurücklegen:

Bestimme Wahrscheinlichkeit für genau  $x$  rote Kugeln in der Stichprobe:

Elementarereignisse:

-  $\Omega$  für Ziehen von  $n$  ohne Zurücklegen ohne Reihenfolge

-  $|\Omega| = \binom{u}{n}$

-  $P$  - Gleichverteilung auf  $\Omega$  aus Symmetriegründen

W-theoretisches Modell:  $(\Omega, P)$

- setzen  $\Omega' = \{0, 1, \dots, s\}$

-  $f : \Omega \rightarrow \Omega'$ ,  $f(\omega) :=$  Anzahl der roten Kugeln in Zug  $\omega$

Aufgabe: bestimme Wahrscheinlichkeitsvektor von  $f_*P$

- bestimme Anzahl in der Faser über  $x$ :

- aus  $r$  roten  $x$  ziehen und aus  $s$  schwarzen  $(n-x)$  zu ziehen ohne Zurücklegen und ohne Reihenfolge

- benutze (3)

- Anzahl der Faser:  $\binom{r}{x} \binom{u-r}{n-x}$

insgesamt:

$$(f_*P)(\{x\}) = \frac{\binom{r}{x} \binom{u-r}{n-x}}{\binom{u}{n}} = \frac{\binom{r}{x} \binom{s}{n-x}}{\binom{r+s}{n}}$$

**Definition 1.41.** Das  $W$ -Maß auf  $\{0, \dots, n\}$  mit Vektor  $x \mapsto \frac{\binom{r}{x} \binom{u-r}{n-x}}{\binom{u}{n}}$  heißt hypergeometrisches Maß.

Erwartungswert:

- Zufallsvariable:  $X(i) := i$

- Beh:  $\mathbb{E}(X) = \frac{nr}{u}$

- benutzen  $\binom{a}{b} = \frac{a}{b} \binom{a-1}{b-1}$

-

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^n k \frac{\binom{r}{k} \binom{n-r}{n-k}}{\binom{u}{n}} \\ &= \sum_{k=1}^n k \frac{r}{k} \frac{\binom{r-1}{k-1} \binom{n-r}{n-k}}{\frac{u}{n} \binom{u-1}{n-1}} \\ &= \frac{nr}{u} \sum_{k=1}^n \frac{\binom{r-1}{k-1} \binom{n-r}{n-k}}{\binom{u-1}{n-1}} \\ &= \frac{nr}{u} \sum_{k=0}^{n-1} \frac{\binom{r-1}{k} \binom{n-r}{n-1-k}}{\binom{u-1}{n-1}} \\ &= \frac{nr}{u} \end{aligned}$$

-  $\sum = 1$  wegen Normierung der hypergeometrischen Verteilung: Parameter  $r - 1$   $n - 1$ ,  $u - 1$

- Varianz:  $\text{Var}(X) = \frac{nr}{u} \left(1 - \frac{n}{u}\right) \frac{u-n}{u-1}$

hypergeometrische Verteilung:  $u=3000$ ,  $n=100$ ,  $r=1000$

□

## 1.7 Bedingte Wahrscheinlichkeit

**Beispiel 1.42.** Würfeln:

$(\Omega, P) = (\{1, 2, 3, 4, 5, 6\}, \text{Gleichverteilung})$

- Wahrscheinlichkeit für jedes Elementarereignis ist  $1/6$

- würfle und bekomme zusätzlich Information, daß das Ergebnis gerade ist

- also  $\omega \in g := \{2, 4, 6\}$

- Wie wahrscheinlich ist dann  $\{2\}$ ? Notation:

-  $P(\{2\} | g)$

- aus Symmetriegründen Gleichverteilung

-  $P(\{2\} | g) = 1/3 = P(\{2\})P(g)$

□

$(\Omega, P)$  - W-Raum

- Partition in zwei Teile -  $(B, B^c)$

Annahme:  $P(B) \neq 0$

**Definition 1.43.**  $P(A | B) := \frac{P(A \cap B)}{P(B)}$  heißt die bedingte Wahrscheinlichkeit von  $A$  unter  $B$ .

Interpretation:

$P(A | B)$  ist die Wahrscheinlichkeit von  $A$  unter der Information  $B$

**Lemma 1.44.** Es gilt  $P(A) = P(A | B)P(B) + P(A | B^c)P(B^c)$

*Beweis.*

$$\begin{aligned}P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A | B)P(B) + P(A | B^c)P(B^c)\end{aligned}$$

□

$A, B \subseteq \Omega$

**Lemma 1.45** (Formel von Bayes). *Wenn  $P(A) \neq 0$  ist, dann gilt:*

$$P(B | A) = \frac{P(B)P(A | B)}{P(A)}$$

*Beweis.*  $P(B | A)P(A) = P(A \cap B) = P(A | B)P(B)$

□

man kann die Information über bedingte Wahrscheinlichkeiten benutzen, um W-theoretische Modelle zu konstruieren

**Beispiel 1.46.** Corona Test:

Experiment: Teste einen zufällig ausgewählten Kandidaten

Elementarereignis:  $(\text{Testergebnis}, \text{Gesundheitszustand})$

- Testergebnis in  $\{pos, neg\}$
- Gesundheit in  $\{g, k\}$
- $\Omega = \{(pos, g), (pos, k), (neg, g), (neg, k)\}$

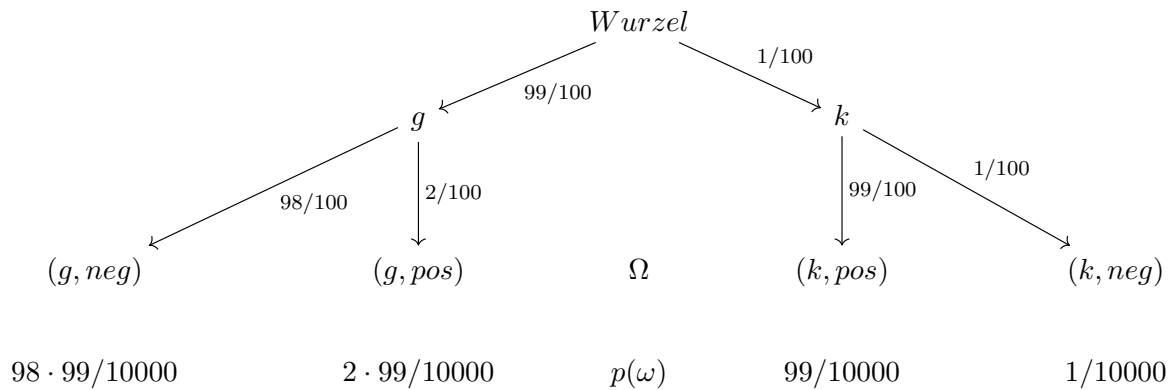
müssen nun W-Maß  $P$  bestimmen:

haben dazu folgende Daten:

- haben einen Anteil von Kranken von  $1/100$
- also ist die Wahrscheinlichkeit, einen kranken Kandidaten zu ziehen,  $1/100$
- haben einen Test der
  - einen Kranken mit  $99/100$  findet (falsch-negativ Wahrscheinlichkeit  $1/100$ )
  - bei einem gesunden mit  $2/100$  positiv ist (falsch-positiv Wahrscheinlichkeit  $2/100$ )



## W-Baum Diagram



teste zufällige Person:

- Zusatzinformation: Test positiv

Mit welcher Wahrscheinlichkeit ist die Person krank?

$$P(k | pos) = ?$$

Bayes Formel:

$$P(pos) = P(pos | k)P(k) + P(pos | g)P(g) = 99/100 \cdot 1/100 + 2/100 \cdot 99/100 = 297/1000$$

$$P(k | pos) = \frac{P(k)P(pos|k)}{P(pos)} = \frac{1/100 \cdot 99/100}{297/10000} = \frac{99}{297} = 0.333$$

Zahlenbeispiel  $x$  Anteil der Kranken an 10000, Wahrscheinlichkeit, bei positivem Test krank zu sein

□

$(\Omega, P)$  - endlicher Wahrscheinlichkeitsraum

- geben uns Partition von  $\Omega$  in mehrere Teilmengen vor

- beschreiben das durch Abbildung  $f : \Omega \rightarrow \Omega'$

-  $f(\omega)$  = Label des Elementes der Partition in welchem  $\omega$  liegt

- entspricht Partition in die Fasern von  $f$

**Beispiel 1.47.**  $B \subseteq \Omega$

-  $\Omega' := \{B, B^c\}$

-  $f : \Omega \rightarrow \Omega', \quad f(\omega) = \begin{cases} B & \omega \in B \\ B^c & \omega \notin B \end{cases}$

-  $f^{-1}(\{B\}) = B$

□

haben Wahrscheinlichkeitsmaß  $P' := f_*P$  auf  $\Omega'$

Problem:

Suchen für jedes  $\omega'$  in  $\Omega'$  ein W-Maß  $P_{\omega'}$  auf  $f^{-1}(\omega')$  so daß für jedes  $A$  in  $\mathcal{P}_\Omega$  gilt:

$$P(A) = \sum_{\omega' \in \Omega'} P_{\omega'}(A \cap f^{-1}(\omega')) P'(\{\omega'\}) \quad (5)$$

hier ist es sinnvoll, 0 auf  $\emptyset$  als W-Maß zu betrachten

**Bemerkung 1.48.**

Annahme: wir hätten Lösung des Problems:

$(P_{\omega'})_{\omega' \in \Omega'}$  - heißt auch "Disintegration" von  $P$  entlang  $f$

-  $X$  - Zufallsvariable

- bilden neue Zufallsvariable  $X'$  auf  $\Omega'$

-  $X'(\omega') := \mathbb{E}_{P_{\omega'}}(X|_{f^{-1}(\{\omega'\})}) = \mathbb{E}(X|\{f^{-1}(\omega')\})$

- faserweise Mittelwert

dann gilt:

-  $\mathbb{E}_P(X) = \mathbb{E}_{P'}(X')$

$$\begin{aligned}
\mathbb{E}_P(X) &= \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) \\
&= \sum_{\omega' \in \Omega'} \sum_{\omega \in f^{-1}(\{\omega'\})} X(\omega)P_{\omega'}(\{\omega\})P'(\{\omega'\}) \\
&= \sum_{\omega' \in \Omega'} X'(\omega')P'(\{\omega'\}) \\
&= \mathbb{E}_{P'}(X')
\end{aligned}$$

□

setzen: für  $A$  in  $f^{-1}(\{\omega'\})$ :

$$P_{\omega'}(A) := \begin{cases} \frac{P(A)}{P'(\{\omega'\})} & P'(\{\omega'\}) \neq 0 \\ \text{beliebige Wahl} & P'(\{\omega'\}) = 0 \end{cases}$$

**Lemma 1.49.**

1. Die Familie  $(P_{\omega'})_{\omega' \in \Omega'}$  löst das Problem.
2. Die Werte der Familie für  $\omega'$  in  $\Omega'$  mit  $P'(\{\omega'\}) \neq 0$  sind eindeutig bestimmt.

*Beweis.*

Zu 1.:

$$\begin{aligned}
P(A) &= P\left(\bigcup_{\omega' \in \Omega'} A \cap f^{-1}(\{\omega'\})\right) \\
&= \sum_{\omega' \in \Omega'} P(A \cap f^{-1}(\{\omega'\})) \\
&= \sum_{\omega' \in \Omega'} P_{\omega'}(A \cap f^{-1}(\{\omega'\}))P'(\{\omega'\})
\end{aligned}$$

Zu 2.:

sei  $A \subseteq f^{-1}(\omega')$

- erhalten:  $P(A) = P_{\omega'}(A)P'(\{\omega'\})$
- also notwendiger Weise  $P_{\omega'}(A) = \frac{P(A)}{P'(\{\omega'\})}$

□

Notation:

- es gilt  $P_{\omega'}(A \cap f^{-1}(\{\omega'\})) = P(A | f^{-1}(\{\omega'\}))$
- Formel (5) ist dann

$$P(A) = \sum_{\omega' \in \Omega'} P(A | f^{-1}(\{\omega'\}))P(f^{-1}(\{\omega'\})) \quad (6)$$

beschreibe Partition durch Angabe der Familie der Teilmengen

- $(B_i)_{i \in I}$  - Partition von  $\Omega$
- Formel (6) ist dann

$$P(A) = \sum_{j \in J} P(A | B_j)P(B_j)$$

$A \subseteq \Omega, P(A) > 0$

**Lemma 1.50** (Bayes Formel). *Für alle  $j$  in  $I$  gilt*

$$P(B_j | A) = \frac{P(B_j)P(A | B_j)}{\sum_{i \in I} P(A | B_i)P(B_i)}$$

*Beweis.*

$$\begin{aligned} \frac{P(B_j)P(A | B_j)}{\sum_{i \in I} P(A | B_i)P(B_i)} &= \frac{P(B_j)P(A | B_j)}{P(A)} \\ &= P(B_j)P(B_j \cap A) \\ &= P(B_j | A) \end{aligned}$$

□

**Beispiel 1.51.** Ziegenproblem:

3 Türen, hinter einer eine Ziege

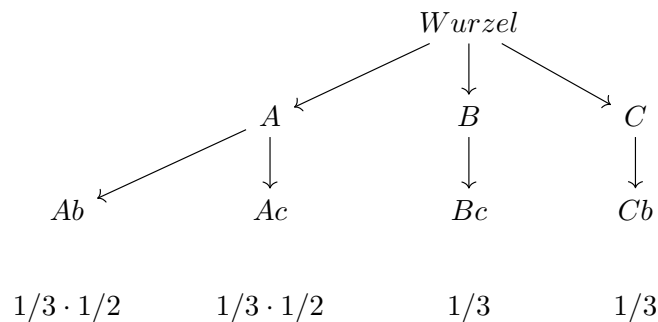
Spiel:

1. Spieler wählt eine Tür
2. Moderator öffnet eine Tür ohne Ziege
3. Spieler kann eine der verbleibenden Türen wählen

Wahrscheinlichkeitstheoretisches Modell

müssen uns klar werden, was die zufälligen Ereignisse sind:

- 1. Ort der Ziege:  $\{A, B, C\}$
  - Gleichverteilung
  - aus Symmetriegründen wähle ich zuerst  $A$  (nicht zufällig)
2. Moderatorenwahl:
- bei Ziege in  $A$  in  $\{b, c\}$  aus Symmetriegründen gleichverteilt,
  - bei Ziege in  $B$  notwendig  $c$
  - bei Ziege in  $C$  notwendig  $b$



Spielverlauf: Spieler wählt  $A$  und Moderator öffnet  $c$ :

- Spieler muß sich für eine der Türen  $A$  oder  $B$  entscheiden
- Wie ist dann die Wahrscheinlichkeit, die Ziege hinter  $A, B$  oder  $C$  zu finden?

Berechnen  $P(- | \{c\})$

$$P(\{c\}) = P(\{Ac\}) + P(\{Bc\}) = 1/6 + 1/3 = 1/2$$

- oder mit Bayes Formel:

$$P(\{c\}) = P(\{c\} | A)P(A) + P(\{c\} | B)P(B) + P(\{c\} | C)P(C) = 1/2 \cdot 1/3 + 1 \cdot 1/3 + 0 \cdot 1/3 = 1/2$$

$$P(A | \{c\}) = \frac{P(A \cap \{c\})}{P(\{c\})} = \frac{1/6}{1/2} = \frac{1}{3}$$

$$P(B | \{c\}) = \frac{P(B \cap \{c\})}{P(\{c\})} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

$$P(C | \{c\}) = \frac{P(C \cap \{c\})}{P(\{c\})} = \frac{0}{\frac{1}{2}} = 0$$

□

**Beispiel 1.52.** haben drei Produktionslinien für dasselbe Produkt

Linie	Ausstoß pro Zeit	Fehlerrate
1	500	0.05
2	200	0.1
3	100	0.02

Fehlerrate gibt den Wert  $P(\{def\}, \{i\})$  an

finde defektes Produkt:

mit welcher Wahrscheinlichkeit kommt es aus der Linie 1?

bilden W-theoretisches Modell:

$$- \Omega = \{1, 2, 3\} \times \{def, io\}$$

-  $P$  - noch nicht explizit bekannt

$P(i)$  - Wahrscheinlichkeit, daß das Produkt aus Linie  $i$  kommt

$$- P(1) = 5/8, P(2) = 2/8, P(3) = 1/8$$

Wahrscheinlichkeit, daß man ein defektes Produkt findet

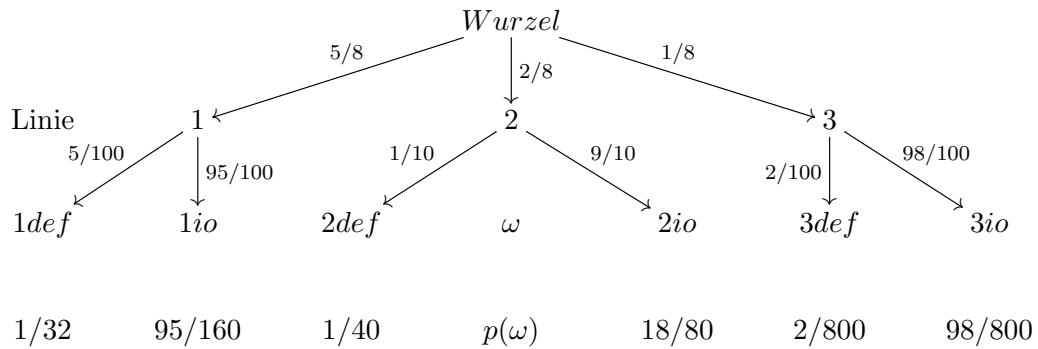
$$- P(def) = \sum_{i=1}^3 P(def | i) \cdot P(i) = 5/100 \cdot 5/8 + 2/8 \cdot 1/10 + 1/8 \cdot 2/100 = 47/800$$

Formel von Bayes anwenden:

$$- P(1 | def) = \frac{P(def|1) \cdot P(1)}{P(def)} = \frac{5/8 \cdot 5/100}{47/800} = \frac{25}{47}$$

Bestimmen ganzen Wahrscheinlichkeitsvektor  $p$ :

Baummodell:



□

Allgemein: Konstruktion von W-Räumen durch Baummodelle

gegeben

- $(\Omega_i)_{i=1,\dots,n}$  eine Familie endlicher Mengen
- bilden  $\Omega = \Omega_1 \times \dots \times \Omega_n$

gegeben:

- W-Maß  $P_1$  auf  $\Omega_1$
- Familie von W-Maßen  $(P_{2,\omega_1})_{\omega_1 \in \Omega_1}$  auf  $\Omega_2$
- Familie von W-Maßen  $(P_{3,(\omega_1,\omega_2)})_{(\omega_1,\omega_2) \in \Omega_1 \times \Omega_2}$  auf  $\Omega_3$
- ...
- Familie von W-Maßen  $(P_n, (\omega_1, \dots, \omega_{n-1}))_{(\omega_1, \dots, \omega_{n-1}) \in \Omega_1 \times \dots \times \Omega_{n-1}}$  auf  $\Omega_n$

definieren

$$p : \Omega \rightarrow [0, 1], \quad p((\omega_1, \dots, \omega_n)) = P_1(\{\omega_1\})P_{2,\omega_1}(\{\omega_2\}) \dots P_{n,(\omega_1, \dots, \omega_{n-1})}(\{\omega_n\}) \quad (7)$$

**Lemma 1.53.**

1.  $p$  ist ein Wahrscheinlichkeitsvektor.

2. Das von  $P$  bestimmte Maß ist eindeutig bestimmt durch: Für jedes  $l$  in  $(1, n)$  und jede Familie  $(\omega'_i)_{i=1, \dots, l}$  gilt

$$P(\{\omega_i = \omega'_i\} \mid \{\omega_1 = \omega'_1, \dots, \omega_{i-1} = \omega'_{i-1}\}) = P_{i, (\omega'_1, \dots, \omega'_{i-1})}(\{\omega'_i\}) .$$

*Beweis.*

$p$  hat nicht-negative Werte

zeigen

$$\sum_{\omega \in \Omega} p(\omega) = 1$$

$$\begin{aligned} \sum_{\omega \in \Omega} p(\omega) &= \sum_{(\omega_1, \dots, \omega_{n-1}) \in \Omega_1 \times \dots \times \Omega_{n-1}} \sum_{\omega_n \in \Omega_n} P_1(\{\omega_1\}) P_{2, \omega_1}(\{\omega_2\}) \dots P_{n, (\omega_1, \dots, \omega_{n-1})}(\{\omega_n\}) \\ &= \sum_{(\omega_1, \dots, \omega_{n-1}) \in \Omega_1 \times \dots \times \Omega_{n-1}} P_1(\{\omega_1\}) P_{2, \omega_1}(\{\omega_2\}) \dots P_{n-1, (\omega_1, \dots, \omega_{n-2})}(\{\omega_{n-1}\}) \\ &= \dots \\ &= \sum_{\omega_1 \in \Omega_1} P_1(\{\omega_1\}) \\ &= 1 \end{aligned}$$

zeigen: das so konstruierte Maß hat die Eigenschaft 2.

$$\begin{aligned} P(\{\omega_i = \omega'_i\} \mid \{\omega_1 = \omega'_1, \dots, \omega_{i-1} = \omega'_{i-1}\}) &= \frac{P(\{\omega_1 = \omega'_1, \dots, \omega_i = \omega'_i\})}{P(\{\omega_1 = \omega'_1, \dots, \omega_{i-1} = \omega'_{i-1}\})} \\ &= \frac{\sum_{(\omega_{i+1}, \dots, \omega_n) \in \Omega_{i+1} \times \dots \times \Omega_n} p(\omega'_1, \dots, \omega'_i, \omega_{i+1}, \dots, \omega_n)}{\sum_{(\omega_i, \dots, \omega_n) \in \Omega_i \times \dots \times \Omega_n} p(\omega'_1, \dots, \omega'_{i-1}, \omega_i, \dots, \omega_n)} \\ &= \frac{P_1(\{\omega'_1\}) P_{2, \omega'_1}(\{\omega'_2\}) \dots P_{i, (\omega'_1, \dots, \omega'_{i-1})}(\{\omega'_i\})}{P_1(\{\omega'_1\}) P_{2, \omega'_1}(\{\omega'_2\}) \dots P_{i-1, (\omega'_1, \dots, \omega'_{i-2})}(\{\omega'_{i-1}\})} \\ &= P_{i, (\omega'_1, \dots, \omega'_{i-1})}(\{\omega'_i\}) \end{aligned}$$

Eindeutigkeit:

sei  $P$  ein W-Maß



- zeigen durch Induktion:

$$P(A_1 \cap \dots \cap A_l) = P(A_1)P(A_2 | A_1) \dots P(A_l | A_1 \cap \dots \cap A_{l-1})$$

solange  $P(A_1 \cap \dots \cap A_l) \neq 0$

richtig für  $l = 1$

Schritt  $l - 1 \Rightarrow l$

$$\begin{aligned} P(A_1 \cap \dots \cap A_l) &= P(A_1 \cap \dots \cap A_{l-1})P(A_l | A_1 \cap \dots \cap A_{l-1}) \\ &= P(A_1)P(A_2 | A_1) \dots P(A_{l-1} | A_1 \cap \dots \cap A_{l-2})P(A_l | A_1 \cap \dots \cap A_{l-1}) \end{aligned}$$

anwenden auf  $A_i = \{\omega_i = \omega'_i\}$  liefert:

Wenn  $P$  die Bedingung 2 erfüllt, dann ist der Vektor von  $P$  durch (7) gegeben

□

## 1.8 Unabhängigkeit

$(\Omega, P)$  - W-Raum

$A, B \subseteq \Omega$  - Ereignisse

**Definition 1.54.**  $A$  und  $B$  heißen unabhängig wenn  $P(A \cap B) = P(A)P(B)$  gilt.

**Lemma 1.55.** Wenn  $A$  and  $B$  unabhängig sind, dann auch  $A^c$  und  $B^c$ :

*Beweis.*

$$\begin{aligned} P(A^c \cap B^c) &= P((A \cup B)^c) \\ &= 1 - P(A \cup B) \\ &= 1 - P(A) - P(B) + P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= (1 - P(A))(1 - P(B)) \\ &= P(A^c)P(B^c) \end{aligned}$$

□

**Beispiel 1.56.** Zwei Münzen werfen

zwei faire Münzen

Ereignisraum:  $\Omega = \{z, w\} \times \{z, w\}$

Gleichverteilung aus Symmetriegründen  $p = \text{const}_{1/4}$

$A =$  erste Münze Zahl  $= \{z\} \times \{z, w\}$

$B =$  zweite Münze Wappen  $= \{z, w\} \times \{w\}$

$A \cap B = \{(z, w)\}$

$P(A)P(B) = 1/2 \cdot 1/2 = 1/4 = P(A \cap B)$

$A$  und  $B$  sind unabhängig

□

$(\Omega, P)$  - W-Raum

$(B_i)_{i \in I}$  Familie in  $\mathcal{P}_\Omega$

**Definition 1.57.** Die Familie  $(B_i)_{i \in I}$  heißt unabhängig, falls für jede Teilmenge  $J \subseteq I$  gilt  $P(\bigcap_{j \in J} B_j) = \prod_{j \in J} P(B_j)$ .

das ist stärker als die Bedingung paarweise unabhängig zu sein

**Beispiel 1.58.** Werfe Münze zwei mal

$\Omega = \{z, w\} \times \{z, w\}$

$A = 1.$  Ergebnis ist  $z$

$B = 2.$  Ergebnis ist  $z$

$C =$  Ergebnisse gleich

Dann gilt:

-  $P(A) = 1/2$ ,  $P(B) = 1/2$  und  $P(C) = 1/2$

-  $P(A \cap B) = 1/4$ ,  $P(A \cap C) = 1/4$  und  $P(B \cap C) = 1/4$

- also Ereignisse paarweise unabhängig

-  $P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$

also ist  $(A, B, C)$  k eine unabhängige Familie □

$(\Omega_i, P_i)_{i=1, \dots, n}$  endliche Familie von W-Räumen

$$\Omega := \prod_{i=1}^n \Omega_i$$

- setzen  $P_{i, (\omega_1, \dots, \omega_{i-1})} := P_i$  (hängt also nicht von  $(\omega_1, \dots, \omega_{i-1})$  ab)

- erhalten nach Lemma 1.53 W-Maß  $P$  auf  $\Omega$

- Wahrscheinlichkeitsvektor

$$- p(\omega_1, \dots, \omega_n) = \prod_{i=1}^n P_i(\{\omega_i\})$$

- bestimmt W-Maß

**Definition 1.59.** Der W-Raum  $(\Omega, P)$  heißt Produkt der Familie von W-Räumen  $(\Omega_i, P_i)$ .

betrachte  $A_i \subseteq \Omega_i$  für  $i = 1, \dots, n$

$$\text{betrachte } B_i := \text{pr}_i^{-1}(A_i)$$

$i_1 < \dots < i_r$  - Teilfamilie

$$B_{i_j} \in \mathcal{P}_\Omega$$

- setzen  $B_i = \Omega_i$  für  $i \notin \{i_1, \dots, i_r\}$

$$P(B_{i_1} \cap \dots \cap B_{i_r}) = \sum_{(\omega_i) \in \prod_{i=1}^n \Omega_i, \omega_{i_j} \in B_{i_j} \forall j} \prod_{i=1}^n P_i(\{\omega_i\}) = \prod_{j=1}^r \sum_{\omega_{i_j} \in B_{i_j}} P_{i_j}(\omega_{i_j}) = \prod_{j=1}^r P(B_{i_j}) \quad (8)$$

- nutzen hier  $\sum_{\omega_i \in \Omega_i} P_i(\omega_i) = 1$  für alle  $i$  im Komplement der Teilfamilie

- die Familie  $(B_i)_{i=1, \dots, n}$  ist unabhängig

haben Morphismus

$$\text{pr}_i : (\Omega, P) \rightarrow (\Omega_i, P_i)$$

in der Tat  $\text{pr}_{i,*} P = P_i$

- da  $P(\text{pr}_i^{-1}(A_i)) = P_i(A_i)$  für alle  $A_i \subseteq \Omega_i$

$(\Omega, P)$  - W-Raum

$(f_i : \Omega \rightarrow \Omega_i)_{i \in I}$  Familie von Abbildungen

**Definition 1.60.** Die Familie heißt unabhängig, wenn für jede Wahl  $(A_i \subseteq \Omega_i)_{i \in I}$  die Familie  $(f_i^{-1}(A_i))_{i \in I}$  unabhängig ist.

$(\Omega_i, P_i)_{i \in I}$  Familie von W-Räumen

$(\Omega, P)$  -Produkt der Familie

**Lemma 1.61.** The Familie der Projektionen  $(\text{pr}_i : \Omega \rightarrow \Omega_i)_{i \in I}$  ist unabhängig.

*Beweis.* klar nach (8)

□

universelle Eigenschaft des Produktes:

$(\Omega_i, P_i)_{i \in I}$  endliche Familie von endlichen W-Räumen

$(\Omega, P)$  - Produkt

für W-Raum  $(T, Q)$  bezeichne

$$\prod_{i \in I}^{\text{un}} \text{Hom}_{\mathbf{WRaum}}((T, Q), (\Omega_i, P_i)) \subseteq \prod_{i \in I}^{\text{un}} \text{Hom}_{\mathbf{WRaum}}((T, Q), (\Omega_i, P_i))$$

der unabhängigen Familien

**Lemma 1.62.** Für jeden W-Raum  $(T, Q)$  ist die Abbildung

$$\text{Hom}_{\mathbf{WRaum}}((T, Q), (\Omega, P)) \xrightarrow{(\text{pr}_{i,*})_i} \prod_{i \in I}^{\text{un}} \text{Hom}_{\mathbf{WRaum}}((T, Q), (\Omega_i, P_i))$$

eine Bijektion.

*Beweis.*

injektiv:

- klar, weil die Menge  $\Omega$  das Produkt der Familie  $(\Omega_i)_{i \in I}$  ist

surjektiv:

$(f_i : (T, Q) \rightarrow (\Omega_i, P_i))_{i \in I}$  gegeben

-  $(f_i)_{i \in I}$  unabhängig

- definieren  $f : T \rightarrow \Omega$ ,  $f(t) := (f_i(\omega_i))_{i \in I}$

- zu zeigen:  $f_*Q = P$

$$\begin{aligned}(f_*Q)(\{(\omega_i)_i\}) &= Q(f^{-1}(\{(\omega_i)_i\})) \\ &= Q\left(\bigcap_{i \in I} f_i^{-1}(\{\omega_i\})\right) \\ &= \prod_{i \in I} Q(f_i^{-1}(\{\omega_i\})) \\ &= \prod_{i \in I} P_i(\{\omega_i\}) \\ &= P(\{(\omega_i)_i\})\end{aligned}$$

□

also ist das Produkt von Maßräumen **nicht** das kartesische Produkt in **WRaum**

$(\Omega, P)$  - W-Raum

-  $X, Y$  - Zufallsvariablen auf  $(\Omega, P)$

**Definition 1.63.** Die Kovarianz von  $X$  und  $Y$  ist durch

$$\text{Cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

definiert.

Beachte:  $\text{Cov}(X, X) = \text{Var}(X)$

**Lemma 1.64.** Wenn  $X$  und  $Y$  unabhängig sind, dann gilt

1.  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

2.  $\text{Cov}(X, Y) = 0$

*Beweis.*

zu 1.

-  $X$  und  $Y$  haben endlich viele Werte

-  $P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\})P(\{Y = y\})$  wegen Unabhängigkeit

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{z \in \mathbb{R}} zP(\{XY = z\}) \\ &= \sum_{z \in \mathbb{R}} \sum_{x, y \in \mathbb{R}, xy=z} zP(\{X = x\} \cap \{Y = y\}) \\ &= \sum_{x, y \in \mathbb{R}} xyP(\{X = x\})P(\{Y = y\}) \\ &= \sum_{x \in \mathbb{R}} xP(\{X = x\}) \sum_{y \in \mathbb{R}} yP(\{Y = y\}) \\ &= \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

zu 2.

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= 0\end{aligned}$$

□

beachte:  $\text{Cov}(X, Y) = 0$  impliziert nicht, daß  $X$  und  $Y$  unabhängig sind.

**Beispiel 1.65.**  $N$ -Mal Münze werfen

Bilden zunächst Modell  $(\Omega', P')$  für einfachen Münzwurf:

-  $\Omega' := \{z, w\}$

- Münze nicht notwendig fair

-  $p \in [0, 1]$  - Wahrscheinlichkeit für Zahl

-  $P'$  bestimmt durch  $p'(z) = p$  and  $p'(w) = 1 - p$

$N \in \mathbb{N}$

Plausibilität: Münzwürfe sind unabhängig

Model für  $N$ -fachen Wurf:  $(\Omega, P)$  ist  $N$ -faches Produkt von  $(\Omega', P')$  mit sich selbst

interessieren uns eigentlich nicht wirklich für den genauen Ausgang sondern Anzahl von  $z$

$f : \Omega \rightarrow \{0, 1, \dots, N\}$

-  $f((\omega_1, \dots, \omega_N)) := |\{i \in \{0, \dots, N\} \mid \omega_i = z\}|$

berechnen Vektor von  $f_*P$

- Vektor von  $P$ :  $p(\omega) = p^{f(\omega)}(1 - p)^{N-f(\omega)}$

- Anzahl von  $\omega$  mit  $f(\omega) = n$ : ist die Anzahl der  $n$ -Elementigen Teilmengen von  $\{1, \dots, N\}$  (nämlich die Stellen, an welchen  $\omega_i = z$  gilt)

-  $|\{\omega \in \Omega \mid f(\omega) = n\}| = \binom{N}{n}$

- Vektor von  $f_*P$ :  $n \mapsto \binom{N}{n} p^n (1 - p)^{N-n}$

- Check der Normierung:

$$\sum_{n=0}^N \binom{N}{n} p^n (1 - p)^{N-n} = (p + (1 - p))^N = 1 .$$

**Definition 1.66.** Das Maß auf  $\{0, \dots, N\}$  mit Vektor  $n \mapsto \binom{N}{n} p^n (1 - p)^{N-n}$  heißt die Binomialverteilung zu  $N$  und  $p$ .

Erwartungswert:  $\mathbb{E}(X) = Np$

$$\begin{aligned}
E(X) &= \sum_{n=0}^N n \binom{N}{n} p^n (1-p)^{N-n} \\
&= \sum_{n=0}^N n \frac{N}{n} \binom{N-1}{n-1} p^n (1-p)^{N-n} \\
&= pN \sum_{n=0}^{N-1} \binom{N-1}{n-1} p^{n-1} (1-p)^{N-1-n} \\
&= pN
\end{aligned}$$

- benutzen dazu die Normierung der Binomialverteilung für  $N-1$  und  $p$  um zu zeigen, daß die Summe 1 ist

Dichte der Binomialverteilung für  $N = 100$ ,  $p = 0.3$

Simulation: werfe Münze 1000 mal und zähle Zahl

- Versuche es 1000 mal
- Histogramm der Häufigkeiten

## 2 Nicht notwendig endliche $W$ -Räume

### 2.1 Abzählbare $W$ -Räume

**Beispiel 2.1.** Wartezeit:

- werfen Münze wiederholt solange bis das erste mal  $z$  kommt
- Typische Frage: Wie groß ist die Wahrscheinlichkeit, daß  $z$  beim 5. Wurf zum ersten Mal kommt?

Elementarereignisse:

- $\Omega := \prod_{\mathbb{N}} \{z, w\}$  - ist nicht endlich (nicht mal abzählbar)
- $P$  - geeignetes  $W$ -Maß auf  $\Omega$  - genauer später

uns interessiert aber im Augenblick nur das erste Auftreten von  $z$

- ist ein Element in  $\Omega' := \mathbb{N}$



- diese Menge ist abzählbar, aber nicht endlich
- $f : \Omega \rightarrow \Omega' - \omega \mapsto$  erstes Auftreten von  $z$
- $f((\omega_i)_i) := \min\{i \in \mathbb{N} \mid \omega_i = z\}$
- Beispiel:  $f((w, w, w, z, w, w, z, z, w, \dots)) = 3$
- erwarten, daß  $f_*P$  ist W-Maß auf  $\Omega'$
- Was für ein Objekt genau soll das sein.

□

betrachten also zunächst W-Maße auf abzählbaren Mengen

- wollen auf jeden Fall die folgenden Axiome behalten

1. Normierung:  $P(\Omega) = 1$
2. Additivität:  $P(A \cup B) = P(A) + P(B)$  for alle  $A, B$  in  $\mathcal{P}_\Omega$  mit  $A \cap B = \emptyset$ .

wenn man nur das fordert, gibt es pathologische Beispiele:

**Beispiel 2.2.** Ultrafilter

**Satz 2.3.** Auf  $\mathbb{N}$  gibt es eine additive und normierte Mengenfunktion  $P$  mit  $P(\{n\}) = 0$  für alle  $n$  in  $\mathbb{N}$ .

- wollen diese Situation später ausschließen
- führen aber erst einmal den Beweis des Satzes

$\Omega$  - Menge

**Definition 2.4.** Ein Filter auf  $\Omega$  ist eine Teilmenge  $F \subseteq \mathcal{P}_\Omega$  welche abgeschlossen unter der Bildung endlicher Durchschnitte und von Obermengen ist,  $X$  enthält, nicht aber  $\emptyset$ .

für  $\omega_0 \in \Omega$  ist  $F_{\omega_0} := \{A \in \mathcal{P}_\Omega \mid \omega_0 \in A\}$  ein Filter

- $\mathcal{F}$  - Menge der Filter auf  $\Omega$
- $\mathcal{F}$  ist partiell geordnet durch Inklusion

**Definition 2.5.** Maximale Filter heißen Ultrafilter.

Beispiel:  $F_{\omega_0}$  ist ein Ultrafilter

**Lemma 2.6.** Jeder Filter ist in einem Ultrafilter enthalten.

*Beweis.* benutzen Lemma von Zorn

- sei  $(F_i)_{i \in I}$  aufsteigende Kette von Filtern

- dann ist  $\bigcup_{i \in I} F_i$  ein Filter (nachrechnen!) und eine obere Schranke □

sei  $F$  Ultrafilter: dann gibt es zwei Fälle

1.  $\bigcap_{A \in F} A \neq \emptyset$

- dann ist  $\bigcap_{A \in F} A = \{\omega_0\}$  und  $F = F_{\omega_0}$

2.  $\bigcap_{A \in F} A = \emptyset$  - dann heißt  $F$  frei

**Lemma 2.7.** Wenn  $|\Omega| = \infty$ , dann gibt es freie Ultrafilter.

*Beweis.*

$F_0 :=$  Menge der Komplemente der endlichen Teilmengen von  $\Omega$

- rechnen nach:  $F_0$  ist Filter

-  $A^c$  endlich und  $A \subseteq B$  implizieren wegen  $B^c \subseteq A^c$  daß auch  $B^c$  endlich ist

-  $A^c, B^c$  endlich impliziert  $(A \cap B)^c = A^c \cup B^c$  daß endlich ist

-  $\Omega^c = \emptyset$  ist endlich

-  $\emptyset^c = \Omega$  ist nicht endlich ( $\emptyset$  also nicht in  $F_0$  enthalten)

- wähle nach Lemma 2.6 Ultrafilter  $F$ , der  $F_0$  enthält.

zeigen  $F$  ist frei:

-  $F$  enthält keine endlichen Teilmengen: angenommen,  $A \in F$  und  $|A| < \infty$

- dann  $A^c \in F_0 \subseteq F$

- dann  $\emptyset = A^c \cap A \in F$  ( $F$  ist abgeschlossen unter Bildung von Durchschnitten), ist falsch
- es gilt schon  $\bigcap_{A \in F_0} A = \emptyset$
- Annahme:  $\omega \in \bigcap_{A \in F_0} A$
- dann  $\Omega \setminus \{\omega\} \in F_0$
- also  $\omega \notin \bigcap_{A \in F_0} A$
- $\bigcap_{A \in F} A \subseteq \bigcap_{A \in F_0} A = \emptyset$
- also ist  $F$  frei □

sei  $F$  Ultrafilter auf  $\Omega$

- $A \subseteq \Omega$

**Lemma 2.8.** *Es gilt entweder  $A \in F$  oder  $A^c \in F$ .*

*Beweis.* indirekt:

Annahme:  $A \notin F$  und  $A^c \notin F$

- dann gilt für jedes  $B$  in  $F$ :  $A^c \cap B \neq \emptyset$  (weil sonst  $B \subseteq A$  und deshalb  $A \in F$  gelten würde)

- $F' :=$  Menge alle Obermengen von  $B \cap A^c$  mit  $B \in F$
- zeigen, daß  $F'$  ein Filter ist
- $F'$  ist abgeschlossen unter Obermengen (klar)
- $F'$  enthält nicht  $\emptyset$  (klar)
- abgeschlossen unter Durchschnitten und enthält nicht leere Menge
  - für  $B, C \in F$ :  $(B \cap A^c) \cap (C \cap A^c) = (B \cap C) \cap A^c$
  - da  $B \cap C \in F$  ist  $(B \cap A^c) \cap (C \cap A^c) \in F'$
- sei  $B \in F$ :
- wegen  $B \cap A^c \subseteq B$  auch  $B \in F'$

- schließen  $F \subseteq F'$
- wegen  $B \cap A^c \subseteq A^c$ , auch  $A^c \in F'$
- also  $F$  nicht maximal, Widerspruch □

fixiere Ultrafilter  $F$

- definiere Mengenfunktion  $P : \mathcal{P}_\Omega \rightarrow [0, 1]$

$$P(A) := \begin{cases} 1 & A \in F \\ 0 & A \notin F \end{cases}$$

erfüllt die Axiome

- $P(\Omega) = 1$  klar
- $P(A \cup B) = P(A) + P(B)$  falls  $A \cap B = \emptyset$  auch klar
- betrachte Fälle  $\omega_0 \in A$ ,  $\omega_0 \in B$ ,  $\omega_0 \notin A \cup B$

falls  $F$  nicht frei:  $F = F_{\omega_0}$  und  $P = \delta_{\omega_0}$

falls  $F$  frei: dann  $P(\{\omega\}) = 0$  für alle  $\omega$  in  $\Omega$  (da  $F$  keine endliche Teilmenge enthält)

- für  $\Omega = \mathbb{N}$  bekommen wir das gewünschte Beispiel □

$\Omega$  - abzählbare Menge,

- $P : \mathcal{P}_\Omega \rightarrow [0, 1]$  additive Mengenfunktion
- die folgende Bedingung soll die Pathologie ausschließen

**Definition 2.9.**  $P$  heißt  $\sigma$ -additiv, wenn für jedes  $A$  in  $\mathcal{P}_\Omega$  gilt:  $P(A) = \sum_{\omega \in A} P(\{\omega\})$

**Lemma 2.10.** Die folgenden Aussagen sind äquivalent:

1.  $P$  ist  $\sigma$ -additiv.
2. Für jede abzählbare paarweise disjunkte Familie  $(A_i)_{i \in I}$  in  $\mathcal{P}_\Omega$  gilt

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i) .$$

*Beweis.*

2  $\Rightarrow$  1.

-  $(\{\omega\})_{\omega \in \Omega}$  ist abzählbare paarweise disjunkte Familie in  $\mathcal{P}_\Omega$

-  $\bigcup_{\omega \in \Omega} \{\omega\} = \Omega$

- also  $\sum_{\omega \in \Omega} P(\{\omega\}) = P(\Omega) = 1$

1.  $\Rightarrow$  2.

- sei  $(A_i)_{i \in I}$  abzählbare paarweise disjunkte Familie

$$\sum_{i \in I} P(A_i) = \sum_{i \in I} \sum_{\omega \in A_i} P(\{\omega\}) = \sum_{\omega \in \bigcup_{i \in I} A_i} P(\{\omega\}) = P\left(\bigcup_{i \in I} A_i\right)$$

□

$\Omega$  - abzählbar

$P$  -  $\sigma$ -additiv:

-  $p : \Omega \rightarrow [0, 1]$  - W-Vektor:  $p(\omega) := P(\{\omega\})$

- erfüllt  $\sum_{\omega \in \Omega} p(\omega) = P(\Omega) = 1$

- diese Bedingung charakterisiert W-Vektoren

$p$  und  $P$  bestimmen sich eindeutig:

- gegeben  $p : \Omega \rightarrow [0, 1]$  mit  $\sum_{\omega \in \Omega} p(\omega) = 1$

- dann ist ein  $\sigma$ -additives W-Maß bestimmt durch

$$A \mapsto P(A) := \sum_{\omega \in A} p(\omega)$$

**Definition 2.11.** Ein abzählbarer W-Raum ist ein Paar  $(\Omega, P)$  aus einer abzählbaren Menge und einem  $\sigma$ -additiven W-Maß auf  $\Omega$ .

diskutieren Verallgemeinerung der Definition des Erwartungswertes:

- haben es jetzt mit unendlichen Summen zu tun

- Konvergenzproblem

$X : \Omega \rightarrow \mathbb{R}$  - numerische Zufallsvariable

einfacher Fall:  $X$  beschränkt

$$\mathbb{E}(X) := \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) \quad (9)$$

- ist definiert als unendliche, aber absolut konvergente Reihe

$$- |X(\omega)P(\{\omega\})| \leq \|X\|_{\infty}P(\{\omega\})$$

$$- \sum_{\omega \in \Omega} \|X\|_{\infty}P(\{\omega\}) = \|X\|_{\infty}$$

allgemeiner Fall:  $X$  nicht notwendig beschränkt:

- fordern:  $\sum_{\omega \in \Omega} |X(\omega)|P(\{\omega\})$  is konvergent (sagen:  $X$  ist summierbar)

- wenn  $X$  summierbar ist, dann ist  $\mathbb{E}(X)$  durch (9) definiert

- hat die üblichen Eigenschaften wie in (1.30)

wenn wenigstens der positive oder negative Teil von  $X$  summierbar ist, kann man  $E(X) \in [-\infty, \infty]$  definieren

**Beispiel 2.12.** Münzwurfbeispiel (Fortsetzung von 2.1):

Annahme:  $p \in (0, 1]$

- wir beginnen die Zählung der Würfe mit 0

beschreibe Wahrscheinlichkeit  $P'(\{n\})$  für erste Zahl bei  $n$

- Begründung später

-  $W$ -Vektor:  $n \mapsto (1-p)^n p$  (es kommt  $n$ -Mal Wappen und dann Zahl)

- Check der Normierung:

$$- \sum_{n \in \mathbb{N}} (1-p)^n p = \frac{p}{1-(1-p)} = 1 \quad (\text{für } p \in (0, 1))$$

— stimmt auch für  $p = 1$

**Definition 2.13.** Das  $W$ -Maß auf  $\mathbb{N}$  mit dem  $W$ -Vektor  $n \mapsto (1-p)^n p$  heißt geometrische Verteilung zum Parameter  $p$ .

$P(\text{gewinne irgendwann}) = 1$

geometrische Verteilung  $p = 1/10, x = 0 : 100$

Wahrscheinlichkeit, daß ich innerhalb von  $n$  Würfeln gewonnen habe.

-  $P([0, n]) = ?$

Code

- Wie oft muß ich mindestens werfen, damit ich mit Wahrscheinlichkeit 0.99 gewonnen habe

-  $\min_{n \in \mathbb{N}} P([0, n]) \geq 0.99$

**Definition 2.14.** Eine Zahl  $c$  heißt  $p$ -Quantil einer Zufallsvariablen  $X$ , falls  $P(\{X \leq c\}) \geq p$  und  $P(X \geq c) \geq 1 - p$  gilt.

suchen also das 0.99-Quantil von  $n$

Code

□

**Beispiel 2.15.** Poisson Verteilung zum Parameter  $\lambda > 0$

Router erhält Pakete zur Weiterleitung zu zufälligen Zeitpunkten

- er sammelt diese im Teitintervall  $[0, 1]$  und schickt sie dann als großes Paket weiter

- Interessieren uns für Anzahl der Pakete, die im Intervall ankommen.

- durchschnittlicher Paketeingang  $\lambda$

- Wieviele Pakete muß der Router zusammenfassen können um mit Wahrscheinlichkeit  $p$  nicht überlastet zu werden?

Elementarereignis: Anzahl der eingehenden Pakete im Einheitsintervall

-  $\Omega := \mathbb{N}$

- W-Maß  $P = ?$

- Heuristik:

- unterteilen  $[0, 1]$  in  $m$  gleichgroße Teilintervalle

- wenn  $m$  sehr groß ist, dann kann man davon ausgehen, daß im kleinen Intervall der Länge  $1/m$  höchstens ein Paket ankommt

- Wahrscheinlichkeit, daß Paket im Intervall der Länge  $1/m$  ankommt ist  $\lambda/m$
- Vergleich mit  $m$  Münzwürfen
- Interpretieren: z=Paket, w= nichts
- Wahrscheinlichkeit für Zahl:  $p = \lambda/m$
- $P(n) \sim \text{Binomialverteilung}(m, p, n) = \binom{m}{n} p^n (1-p)^{m-n}$

wollen  $m$  sehr groß machen

- benutzen  $(1 - \frac{\lambda}{m})^m \xrightarrow{m \rightarrow \infty} e^{-\lambda}$  und  $(1 - \frac{\lambda}{m})^{m-n} = (1 - \frac{\lambda}{m})^m (1 - \frac{\lambda}{m})^{-n} \rightarrow e^{-\lambda} \cdot 1$

$$\binom{m}{n} p^n (1-p)^{m-n} = \frac{m(m-1) \dots (m-n+1)}{n!} \frac{\lambda^n}{m^n} (1 - \frac{\lambda}{m})^{m-n}$$

$$\xrightarrow{m \rightarrow \infty} \frac{\lambda^n}{n!} e^{-\lambda}$$

$$p(n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

**Definition 2.16.** Was  $W$ -Maß auf  $\mathbb{N}$  mit  $W$ -Vektor  $n \mapsto p(n) = e^{-\lambda} \frac{\lambda^n}{n!}$  heißt *Poissonverteilung* (zum Parameter  $\lambda$ )

Check der Normierung:  $\sum_{n \in \mathbb{N}} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1$

Graph  $\lambda = 20, \omega = 0, \dots, 100$

Ab welchen  $N$  ist der Router mit Wahrscheinlichkeit kleiner  $q$  überlastet.

- Router überlastet, wenn  $n > N$ .
- Wollen also  $P(\{n > N\}) < q$  haben
- Also  $P(\{n \leq N\}) \geq p = 1 - q$

$\lambda = 20$ , 0.99-Quantil für Poissonverteilung

Also ab 31.

Wenn der Router mehr als 31 Pakete pro Zeiteinheit verarbeiten kann, dann wird er mit  $p = 0.99$  nicht überlastet



$\lambda = 20$ ,  $x/100$ -Quantil für Poissonverteilung

□

## 2.2 Allgemeine W-Räume

**Beispiel 2.17.** Gleichverteilung auf  $[0, 1]$

haben genaue Vorstellungen über Erwartungswerte:

- für stetige Zufallsvariable  $X$ :
- $\mathbb{E}(X) = \int_0^1 X(x)dx$  für stetige Zufallsvariable
- Riemann-Integral

betrachten nun W-Maß

- setze:  $p := P(\{1/2\})$
- $P(\{x\})$  sollte nicht von  $x$  abhängen aus Symmetriegründen
- also  $p = P(\{x\})$  für alle  $x$  in  $[0, 1]$
- $|[0, 1]| = \infty$
- mit unserer Definition 2.9 von  $\sigma$ -Additivität und Normierung ergibt sich
- $1 = \sum_{x \in [0,1]} p$
- kann nicht erfüllt werden

Fazit: mit den bisherigen Begrifflichkeiten kann man die Gleichverteilung auf  $[0, 1]$  nicht modellieren, obwohl wir für stetige Zufallsvariablen die Erwartungswerte kennen

Ausweg: Lemma 2.10 legt alternative Definition nahe

Für jede abzählbare paarweise disjunkte Familie  $(A_i)_{i \in I}$  in  $\mathcal{P}_\Omega$  gilt

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i) .$$

- damit bekommt man nicht sofort einen Widerspruch:

–  $[0, 1]$  ist ja nicht abzählbar (und damit keine abzählbare Vereinigung von Einpunktmen-  
gen)

□

**Beispiel 2.18.** Wiederholter Münzwurf

wir werfen eine Münze immer wieder (ohne apriori Begrenzung)

- Ereignisraum:  $\Omega = \prod_{\mathbb{N}}\{z, w\}$

- Menge der Folgen  $(\omega_i)_{i \in \mathbb{N}}$  mit  $\omega_i \in \{z, w\}$

zeigen im folgenden: Die Annahme, daß das Experiment durch  $\sigma$ -additives Maß  $P$  model-  
liert modelliert wird, führt zu Widersprüchen.

-  $P(\{\omega\})$  hängt nicht von  $\omega$  ab aus Symmetriegründen

– also  $P(\{\omega\}) = 0$  wegen Normierung

– benutzen also alternative Definition von  $\sigma$ -Additivität:

*Für jede abzählbare paarweise disjunkte Familie  $(A_i)_{i \in I}$  in  $\mathcal{P}_\Omega$  gilt*

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i) .$$

– dann  $P(A) = 0$  für jede abzählbare Menge  $A$

— kein Widerspruch zur Normierung, da  $\Omega$  nicht abzählbar

Zeigen aber weiteres Problem auf:

- Für  $n$  in  $\mathbb{N}$  sei  $T_n : \Omega \rightarrow \Omega$  die Abbildung, die das Ergebnis des  $n$ -Wurfs umdreht

– aus Symmetriegründen muß gelten  $T_{n,*}P = P$

definieren Äquivalenzrelation auf  $\Omega$

-  $(\omega_i)_i \sim (\omega'_i)_i$  falls  $|\{i \in \mathbb{N} \mid \omega_i \neq \omega'_i\}| < \infty$

– also  $(\omega_i)_i \sim (\omega'_i)_i$  falls die beiden Folgen in höchstens endlich vielen Stellen verschieden  
sind

- es gilt  $T_n(\omega) \sim \omega$

- $T_n$  ändert genau eine Stelle
- betrachten Projektion  $\pi : \Omega \rightarrow \Omega / \sim$
- Auswahlaxiom: es gibt einen Schnitt  $A \subseteq \Omega$
- $A$  enthält aus jeder Äquivalenzklasse genau ein Element
- $S :=$  Menge der endlichen Teilmengen von  $\mathbb{N}$
- $S$  ist abzählbar
- für  $I$  in  $S$  setze  $T_I = \prod_{i \in I} T_i$  (Komposition!)
- $T_I \omega \sim \omega$
- $T_I \omega$  und  $\omega$  unterscheiden sich nur an den endlich vielen Stellen, die durch  $I$  gegeben sind.
- $\Omega = (T_I(A))_{I \in S}$  ist abzählbare Partition von  $\Omega$
- die Familie  $(T_I(A))_{I \in S}$  überdeckt  $\Omega$ 
  - sei  $\omega$  in  $\Omega$
  - dann ist  $[\omega] \cap A = \{a\}$  für eindeutiges  $a \in A$
  - dann existiert  $I \in S$  mit  $T_I(a) = \omega$
  - also  $\omega \in \bigcup_{I \in S} T_I(A)$
- die Mitglieder der Familie  $(T_I(A))_{I \in S}$  sind paarweise disjunkt
  - $\omega \in T_I(A) \cap T_{I'}(A)$  heißt  $\omega = T_I(a)$  und  $\omega = T_{I'}(a')$
  - aber dann  $a \sim T_I(a) = T_{I'}(a') \sim a'$  und deshalb  $a = a'$ ,
  - wegen  $T_I(a) = T_{I'}(a)$  gilt dann  $I = I'$
- sei  $p = P(A)$
- dann  $P(T_I(A)) = P(T_I^{-1}(A)) = (T_{I,*}P)(A) = P(A) = p$
- wegen  $\sigma$ -Additivität von  $P$
- $1 = \sum_{I \in S} p$  - das ist unmöglich

Wir können also nicht fordern, daß

- $P$   $\sigma$ -additiv, invariant unter den Operationen  $T_{I,*}$  und gleichzeitig auf  $A$  definiert ist.
- wollen  $\sigma$ -Additivität
- für das Modell ist Invarianz nötig
- also muß man aufgeben, daß  $P$  auf ganz  $P_\Omega$  definiert ist
- wir schränken damit ein, welche Ereignisse wir betrachten wollen
- Beachte: Wir haben das Auswahlaxiom benutzt, um ein nicht zu betrachtendes Ereignis zu konstruieren □

$\Omega$  - Menge

-  $\mathcal{R} \subseteq \mathcal{P}_\Omega$

**Definition 2.19.**  $\mathcal{R}$  heißt Algebra, wenn

1.  $\Omega \in \mathcal{R}$
2. Mit  $B$  in  $\mathcal{R}$  ist auch  $B^c$  in  $\mathcal{R}$
3. Ist  $A, B$  in  $\mathcal{R}$ , dann auch  $A \cup B \in \mathcal{R}$

$\mathcal{R}$  soll also unter endlicher Anwendung der mengentheoretischen Operationen abgeschlossen sein

- sei  $\mathcal{A} \subseteq P_\Omega$

- vergrößern  $\mathcal{A}$  schrittweise durch wiederholte Anwendung der Operationen

—  $A \mapsto A^c$

—  $A, B \mapsto A \cup B$

- erhalten Algebra  $\mathcal{R}\langle \mathcal{A} \rangle$  - Algebra die von  $\mathcal{A}$  erzeugt wird

- man kann zeigen:  $\mathcal{R}\langle \mathcal{A} \rangle = \bigcap_{\mathcal{A} \subseteq \mathcal{R}'} \mathcal{R}'$

- Durchschnitt wird über alle Algebren genommen wird, welche  $\mathcal{A}$  enthalten

sei  $\mathcal{R}$  - Algebra

-  $P : \mathcal{R} \rightarrow [0, 1]$  - Mengenfunktion (jetzt nur auf  $\mathcal{R}$  definiert)

**Definition 2.20.**  $P$  ist additiv wenn gilt

1.  $P(\emptyset) = 0$
2.  $P(\Omega) = 1$
3.  $P(A \cup B) = P(A) + P(B)$  for alle  $A, B$  in  $\mathcal{R}$  mit  $A \cap B = \emptyset$ .

- die bisher betrachteten Beispiele waren auf  $\mathcal{R} = \mathcal{P}_\Omega$  definiert

**Definition 2.21.** Eine additive Mengenfunktion  $P$  auf  $\mathcal{R}$  ist  $\sigma$ -additiv, wenn für jede abzählbare paarweise disjunkte Familie  $(A_i)_{i \in I}$  in  $\mathcal{R}$  mit  $A := \bigcup_{i \in I} A_i \in \mathcal{R}$  gilt  $\sum_{i \in I} P(A_i) = P(A)$ .

sei  $P$  additive Mengenfunktion auf  $\mathcal{R}$

**Lemma 2.22.** Folgende Aussagen sind äquivalent:

1.  $P$  is  $\sigma$ -additiv
2. Für jede aufsteigende Familie  $A_0 \subseteq A_1 \subseteq \dots$  in  $\mathcal{R}$  mit  $A := \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{R}$  gilt  $\lim_{n \rightarrow \infty} P(A_n) = P(A)$ .
3. Für jede absteigende Familie  $A_0 \supseteq A_1 \supseteq \dots$  in  $\mathcal{R}$  mit  $A := \bigcap_{n \in \mathbb{N}} A_n \in \mathcal{R}$  gilt  $\lim_{n \rightarrow \infty} P(A_n) = P(A)$ .
4. Für jede absteigende Familie  $A_0 \supseteq A_1 \supseteq \dots$  in  $\mathcal{R}$  mit  $\emptyset = \bigcap_{n \in \mathbb{N}} A_n$  gilt  $\lim_{n \rightarrow \infty} P(A_n) = 0$ .

*Beweis.*

1  $\Rightarrow$  2

- gegeben:  $A_0 \subseteq A_1 \subseteq \dots$  in  $\mathcal{R}$  aufsteigend

-  $B_0 := A_0, B_i := A_i \setminus A_{i-1}$  für  $i \geq 1$

-  $(B_i)_{i \in \mathbb{N}}$  - paarweise disjunkt

-  $\bigcup_{i \in \mathbb{N}} B_i = A$

$$- \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} \sum_{i=0}^n P(B_i) = \sum_{i=0}^{\infty} P(B_i) = P(A)$$

— benutzen Additivität von  $P$  für erste Gleichung

— benutzen  $\sigma$ -Additivität von  $P$  für letzte Gleichung

2  $\Rightarrow$  3

- für absteigende Familie  $A_0 \supseteq A_1 \supseteq \dots$  betrachte aufsteigende Familie  $A_0^c \subseteq A_1^c \subseteq \dots$

$$- \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} (1 - P(A_n^c)) = 1 - \lim_{n \rightarrow \infty} P(A_n^c) = 1 - P(A^c) = P(A)$$

- benutzen 2. für vorletzte Gleichung

3  $\Rightarrow$  1

-  $(B_i)_{i \in \mathbb{N}}$  paarweise diskunkt

-  $(A_n)_{n \in \mathbb{N}} := ((\bigcup_{i=0}^n B_i)^c)_{n \in \mathbb{N}}$  - absteigende Familie

$$- \bigcap_{n \in \mathbb{N}} A_n = (\bigcup_{i \in \mathbb{N}} B_i)^c$$

$$\begin{aligned} P\left(\bigcup_{i \in \mathbb{N}} B_i\right) &= 1 - P\left(\left(\bigcup_{i \in \mathbb{N}} B_i\right)^c\right) \\ &\stackrel{3}{=} 1 - \lim_{n \rightarrow \infty} P(A_n) \\ &\stackrel{!}{=} 1 - \lim_{n \rightarrow \infty} \left(1 - \sum_{i=0}^n P(B_i)\right) \\ &= \sum_{i=0}^{\infty} P(B_i) \end{aligned}$$

- benutzen für ! endliche Additivität

3  $\Rightarrow$  4

- klar

4  $\Rightarrow$  3

- gegeben:  $A_0 \supseteq A_1 \supseteq \dots$  absteigende Familie in  $\mathcal{R}$  mit  $A := \bigcap_{n \in \mathbb{N}} A_n \in \mathcal{R}$

- setze  $A'_i := A_i \setminus A$

- dann  $A'_0 \supseteq A'_1 \supseteq \dots$  absteigende Familie in  $\mathcal{R}$  mit  $\emptyset = \bigcap_{n \in \mathbb{N}} A'_n$

$$\begin{aligned} \lim_{n \rightarrow \infty} P(A_n) &= \lim_{n \rightarrow \infty} P(A'_n) + P(A) \\ &\stackrel{4}{=} P(A) \end{aligned}$$

□

aus der Maßtheorie: der natürliche Definitionsbereich eines  $\sigma$ -additiven Maßes ist eine  $\sigma$ -Algebra

- Erinnerung
- $\Omega$  - Menge
- $\mathcal{B} \subseteq \mathcal{P}_\Omega$

**Definition 2.23.**  $\mathcal{B}$  ist eine  $\sigma$ -Algebra wenn gilt:

1.  $\mathcal{B}$  ist eine Algebra.
2. Ist  $(B_i)_{i \in I}$  eine abzählbare Familie in  $\mathcal{B}$ , dann ist auch  $\bigcup_{i \in I} B_i$  in  $\mathcal{B}$ .

eine  $\sigma$ -Algebra ist also eine Algebra, die unter abzählbarer Anwendung mengentheoretischer Operationen abgeschlossen ist

Erzeugung von  $\sigma$ -Algebren

- $\mathcal{A} \subseteq \mathcal{P}_\Omega$

**Lemma 2.24.**  $\sigma\langle \mathcal{A} \rangle := \bigcap_{\mathcal{A} \subseteq \mathcal{B}} \mathcal{B}$  ( $\mathcal{B}$  läuft durch alle  $\mathcal{A}$ -enthaltenden  $\sigma$ -Algebren) ist die kleinste  $\sigma$ -Algebra, die  $\mathcal{A}$  enthält

**Definition 2.25.** Ein meßbarer Raum ist ein Paar  $(\Omega, \mathcal{B})$  aus einer Menge und einer  $\sigma$ -Algebra  $\mathcal{B}$ .

**Beispiel 2.26.**  $\Omega$  - Menge

- $\Omega_{max} := (\Omega, \mathcal{P}_\Omega)$  ist meßbarer Raum
- $\Omega_{min} := (\Omega, \{\emptyset, \Omega\})$  ist meßbarer Raum

□

**Beispiel 2.27.** Borelsche  $\sigma$ -Algebra

$\Omega$  -topologischer Raum

$\mathcal{T}$  - Topologie - Menge der offenen Teilmengen

**Definition 2.28.** Die  $\sigma$ -Algebra  $\sigma\langle\mathcal{T}\rangle$  heißt die Borelsche  $\sigma$ -Algebra auf  $\Omega$

wir betrachten einen topologischen Raum standardmäßig mit der Borelschen  $\sigma$ -Algebra, wenn nichts anderes gesagt wird

□

$(\Omega, \mathcal{B}), (\Omega', \mathcal{B}')$  - meßbare Räume

**Definition 2.29.** Eine Abbildung  $f : \Omega \rightarrow \Omega'$  ist meßbar, wenn  $f^{-1}(B') \subseteq \mathcal{B}$  gilt.

- hier  $f^{-1}(B')$  - Menge der Urbilder  $f^{-1}(B')$  für  $B'$  in  $\mathcal{B}'$

- Komposition meßbarer Abbildung ist wieder meßbar

erhalten Kategorie **Meas** der meßbaren Räume und meßbaren Abbildungen

das folgende Lemma ist oft hilfreich, um Meßbarkeit nachzuweisen

-  $(\Omega, \mathcal{B}), (\Omega', \mathcal{B}')$  meßbare Räume

-  $f : \Omega \rightarrow \Omega'$

- Annahme:  $\mathcal{B}' = \sigma\langle\mathcal{A}'\rangle$  für  $\mathcal{A}' \subseteq \Omega'$

**Lemma 2.30.**  $f$  ist genau dann meßbar, wenn  $f^{-1}(\mathcal{A}') \subseteq \mathcal{B}$

*Beweis.*

Behaupten:  $\sigma\langle f^{-1}(\mathcal{A}')\rangle = f^{-1}(\sigma\langle\mathcal{A}'\rangle)$

-  $\subseteq$

-  $f^{-1}(-)$  ist mit mengentheoretischen Operationen verträglich

—  $f^{-1}(\sigma\langle\mathcal{A}'\rangle)$  ist  $\sigma$ -Algebra auf  $\Omega$

—  $f^{-1}(\mathcal{A}') \subseteq f^{-1}(\sigma\langle\mathcal{A}'\rangle)$



- damit  $\sigma\langle f^{-1}(\mathcal{A}') \rangle \subseteq f^{-1}(\sigma\langle \mathcal{A}' \rangle)$
- $\supseteq$
- $\mathcal{R} := \{A \in \sigma\langle \mathcal{A}' \rangle \mid f^{-1}(A) \in \sigma\langle f^{-1}(\mathcal{A}') \rangle\}$
- $\mathcal{R}$  ist  $\sigma$ -Algebra
- Übungsaufgabe!
- $\mathcal{A}' \subseteq \mathcal{R}$
- also  $\sigma\langle \mathcal{A}' \rangle \subseteq \mathcal{R}$
- also  $f^{-1}(\sigma\langle \mathcal{A}' \rangle) \subseteq f^{-1}(\mathcal{R}) \subseteq \sigma\langle f^{-1}(\mathcal{A}') \rangle$

eigentlicher Beweis:

$f$  meßbar

- dann  $f^{-1}(\mathcal{A}') \subseteq f^{-1}(\sigma\langle \mathcal{A}' \rangle) = f^{-1}(\mathcal{B}') \subseteq \mathcal{B}$

$f^{-1}(\mathcal{A}') \subseteq \mathcal{B}$

- $\mathcal{B}$  ist  $\sigma$ -Algebra
- wegen  $f^{-1}(\mathcal{A}') \subseteq \mathcal{B}$  gilt auch  $\sigma\langle f^{-1}(\mathcal{A}') \rangle \subseteq \mathcal{B}$
- also  $f^{-1}(\mathcal{B}') = f^{-1}(\sigma\langle \mathcal{A}' \rangle) = \sigma\langle f^{-1}(\mathcal{A}') \rangle \subseteq \mathcal{B}$

□

**Beispiel 2.31.** haben Vergißfunktoren  $F : \mathbf{Meas} \rightarrow \mathbf{Set}$

- $F(\Omega, \mathcal{B}) := \Omega$
- haben Adjunktionen:

$$(\Omega \mapsto \Omega_{max}) : \mathbf{Set} \rightleftarrows \mathbf{Meas} : F$$

$$F : \mathbf{Meas} \rightleftarrows \mathbf{Set} : (\Omega \mapsto \Omega_{min})$$

- Übungsaufgabe für Interessierte

□

$(\Omega, \mathcal{B})$  - meßbarer Raum

**Definition 2.32.** Eine numerische Zufallsvariable ist eine meßbare Abbildung  $X : \Omega \rightarrow \mathbb{R}$ .

- manchmal lassen wir auch Werte in  $[-\infty, \infty]$  zu

**Beispiel 2.33.**  $(\Omega, \mathcal{B})$  - meßbarer Raum

betrachte Abbildung  $X : \Omega \rightarrow \mathbb{R}$

- um zu zeigen, daß  $X$  meßbar ist reicht es, eine der folgenden Bedingungen zu prüfen:

1.  $X^{-1}((-\infty, c))$  ist meßbar für alle  $c$  in  $\mathbb{R}$
2.  $X^{-1}((-\infty, c])$  ist meßbar für alle  $c$  in  $\mathbb{R}$
3.  $X^{-1}((c, \infty))$  ist meßbar für alle  $c$  in  $\mathbb{R}$
4.  $X^{-1}([c, \infty))$  ist meßbar für alle  $c$  in  $\mathbb{R}$

- in der Tat erzeugen die betreffenden Familien von Intervallen die Borelsche  $\sigma$ -Algebra auf  $\mathbb{R}$

insbesondere: sei  $\Omega$  auch topologisch und  $\mathcal{B}$  die Borelsche  $\sigma$ -algebra

**Lemma 2.34.** Wenn  $X : \Omega \rightarrow \mathbb{R}$  stetig ist, dann ist  $X$  meßbar.

*Beweis.*  $f^{-1}((-\infty, c))$  is offen, also meßbar für alle  $c$  in  $\mathbb{R}$ . □

-  $(X_n)_{n \in \mathbb{N}}$  Folge von numerischen Zufallsvariablen

-  $X := \lim_{n \rightarrow \infty} X_n$  existiere punktweise

**Lemma 2.35.**  $X$  eine Zufallsvariable

*Beweis.* zu zeigen ist:  $X$  ist meßbar

- wurde in der Maßtheorie diskutiert □

□

### Beispiel 2.36. Kartesisches Produkt in Meas

Konstruktion:

- $(\Omega_i, \mathcal{B}_i)_{i \in I}$  - Familie in Meas
- $\Omega := \prod_{i \in I} \Omega_i$
- $\text{pr}_i : \Omega \rightarrow \Omega_i$
- $\mathcal{B} := \sigma(\bigcup_{i \in I} \text{pr}_i^{-1}(\mathcal{B}_i))$

klar ist:

- $\text{pr}_i : (\Omega, \mathcal{B}) \rightarrow (\Omega_i, \mathcal{B}_i)$  ist meßbar

**Lemma 2.37.**  $((\Omega, \mathcal{P}), (\text{pr}_i)_{i \in I})$  repräsentiert das kartesische Produkt in Meas.

*Beweis.*

$(T, \mathcal{T})$  in Meas beliebig

- müssen zeigen, daß folgende Abbildung ein Isomorphismus ist:

$$\text{Hom}_{\text{Meas}}((T, \mathcal{T}), (\Omega, \mathcal{B})) \xrightarrow{f \mapsto (\text{pr}_i \circ f)_i} \prod_{i \in I} \text{Hom}_{\text{Meas}}((T, \mathcal{T}), (\Omega_i, \mathcal{B}_i))$$

- injektiv: klar
- surjektiv:  $(g_i)_i \in \prod_{i \in I} \text{Hom}_{\text{Meas}}((T, \mathcal{T}), (\Omega_i, \mathcal{B}_i))$  gegeben
- definiere  $g : T \rightarrow \Omega$ ,  $g(t) = (g_i(t))_i$
- sehe  $g^{-1}(\text{pr}_i^{-1}(\mathcal{B}_i)) = g_i^{-1}(\mathcal{B}_i) \subseteq \mathcal{T}$
- daraus folgt Meßbarkeit von  $g$  mit Lemma 2.30 □

□

$(\Omega, \mathcal{B})$  - meßbarer Raum

**Definition 2.38.** Ein  $W$ -Maß auf  $(\Omega, \mathcal{B})$  ist eine  $\sigma$ -additive Mengenfunktion  $P : \mathcal{B} \rightarrow [0, 1]$ .

das ist unsere finale Definition eines  $W$ -Raumes

**Definition 2.39.** Ein  $W$ -Raum ist ein Tripel  $(\Omega, \mathcal{B}, P)$  aus einer Menge  $\Omega$ , einer  $\sigma$ -Algebra  $\mathcal{B}$ , und einem  $W$ -Maß  $P$  auf  $(\Omega, \mathcal{B})$ .

**Beispiel 2.40.** Bild von  $W$ -Maßen

$(\Omega, \mathcal{B}, P)$  -  $W$ -Raum

$f : (\Omega, \mathcal{B}) \rightarrow (\Omega', \mathcal{B}')$  - meßbar

-  $f_*P : \mathcal{B}' \rightarrow [0, 1]$  definiert durch  $(f_*P)(A') := P(f^{-1}(A'))$

dann ist  $f_*P$  ein  $W$ -Maß auf  $(\Omega', \mathcal{B}')$

- wohldefiniert, da  $f^{-1}(\mathcal{B}') \subseteq \mathcal{B}$

- genau deswegen betrachtet man meßbare Abbildungen

- Additivität und  $\sigma$ -Additivität klar, da  $f^{-1}$  mit mengentheoretischen Operationen verträglich ist

□

erweitern jetzt den Begriff des Erwartungswertes

$(\Omega, \mathcal{B}, P)$  -  $W$ -Raum

$X : \Omega \rightarrow \mathbb{R}$  - meßbar

1. Fall:  $X$  hat endlich viele Werte

- sagen, daß  $X$  einfach ist

-  $\{x\}$  ist meßbar in  $[0, \infty)$  (abgeschlossene Teilmenge)

-  $\{X = x\}$  ist meßbare Menge

$\mathbb{E}(X) := \sum_{x \in \mathbb{R}} xP(\{X = x\})$

- Eigenschaften:

- linear:  $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda \mathbb{E}(Y)$  für  $\lambda \in \mathbb{R}$

- positiv:  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$

- normiert:  $\mathbb{E}(1) = 1$

2. Fall:  $X$  nur meßbar, aber mit Werten in  $[0, \infty)$

setzen:  $\mathbb{E}(X) := \sup_{X' \leq X} \mathbb{E}(X')$  - Supremum über alle einfachen Zufallsvariablen  $X'$

- es kann  $\infty$  herauskommen!

3. Fall:  $X$  meßbar,  $\mathbb{E}(|X|) < \infty$  (wir sagen daß  $X$  integrierbar ist)

-  $X = X^+ + X^-$  - Zerlegung in positiven und negativen Teil

setzen:  $\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(-X^-)$

In der Notation der Maß- und Integrationstheorie:

$$\mathbb{E}(X) = \int_{\Omega} X dP .$$

- Integral von  $X$  bezüglich  $P$

- importieren einige Ergebnisse aus der Maß- und Integrationstheorie

-  $X$  - nicht negative Zufallsvariable

**Lemma 2.41.** *Es gibt eine monoton wachsende Folge einfacher Zufallsvariablen  $(X_n)_{n \in \mathbb{N}}$  mit  $\lim_{n \rightarrow \infty} X_n = X$ .*

*Beweis.* setze:

$$X_n := \sum_{j=0}^{4^n-1} \frac{j}{2^n} \chi_{\{X \in [\frac{j}{2^n}, \frac{j+1}{2^n})\}} + 2^n \chi_{\{X \in [2^n, \infty)\}}$$

□

$(X_n)_{n \in \mathbb{N}}$  - aufsteigende Folge nicht-negativer Zufallsvariablen

**Satz 2.42** (Monotone Konvergenz). *Es gilt  $\mathbb{E}(\lim_{n \in \mathbb{N}} X_n) = \lim_{n \in \mathbb{N}} \mathbb{E}(X_n)$*

*Beweis.* Maßtheorie

-  $\lim_{n \in \mathbb{N}} \mathbb{E}(X_n) \leq \mathbb{E}(\lim_{n \in \mathbb{N}} X_n)$  ist einfach

- für Gleichheit: benutze Lemma 2.41

□

- es können beide Seiten gleich  $\infty$  sein

**Satz 2.43.** Für integrierbare Zufallsvariablen  $X$  und  $Y$  und  $\lambda$  in  $\mathbb{R}$  gilt:

1.  $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda \mathbb{E}(Y)$

2.  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$

*Beweis.* Maßtheorie

- benutze Lemma 2.41 und Satz 2.42 □

$(X_i)_{i \in \mathbb{N}}$  Folge von Zufallsvariablen

-  $\lim_{i \in \mathbb{N}} X_i$  existiere punktweise (mit Werten in  $[-\infty, \infty]$ )

**Satz 2.44** (Majorisierte Konvergenz). Wenn es eine integrierbare Zufallsvariable  $X$  mit  $|X_i| \leq X$  für alle  $i$  in  $I$  gibt, dann ist  $\lim_{i \in I} X_i$  integrierbar und es gilt

$$\mathbb{E}(\lim_{i \in I} X_i) = \lim_{i \in I} \mathbb{E}(X_i) .$$

*Beweis.* Maßtheorie □

**Beispiel 2.45.** Verträglichkeit der neuen mit alten Definitionen im Fall abzählbarer Maßräume

-  $(\Omega, \mathcal{P}_\Omega, P)$  - abzählbarer Maßraum

-  $X : \Omega \rightarrow \mathbb{R}$  - Zufallsvariable

Beh:  $X$  integrierbar (im neuen Sinn) genau dann wenn  $\sum_{\omega \in \Omega} |X(\omega)|p(\omega) < \infty$

-  $X$  integrierbar

-  $I \subseteq \Omega$  endlich  $X'_I := \chi_I |X|$

-  $\sum_{\omega \in I} |X(\omega)|p(\omega) = \mathbb{E}(X'_I) \leq \mathbb{E}(|X|)$

-  $I$  bel.

— schließen  $\sum_{\omega \in \Omega} |X(\omega)|p(\omega) \leq \mathbb{E}(|X|) < \infty$

-  $\sum_{\omega \in \Omega} |X(\omega)|p(\omega) < \infty$

- $X'$  habe endlich viele Werte,  $0 \leq X' \leq |X|$
- $\mathbb{E}(X') = \sum_{\omega \in \Omega} X'(\omega)p(\omega) \leq \sum_{\omega \in \Omega} |X(\omega)|p(\omega) < \infty$
- $\mathbb{E}(|X|) = \sup_{X' \leq |X|} \mathbb{E}(X') \leq \sum_{\omega \in \Omega} |X(\omega)|p(\omega) < \infty$
- also  $X$  integrierbar

beide Ungleichungen zusammen zeigen:  $\mathbb{E}(|X|) = \sum_{\omega \in \Omega} |X(\omega)|p(\omega)$

Beh: Für integrierbares  $X$  gilt

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)p(\omega)$$

- schreiben  $X = X^+ - X^-$
- können annehmen, daß  $X \geq 0$
- haben oben gezeigt:  $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)p(\omega)$

□

$f : (\Omega', \mathcal{B}', P') \rightarrow (\Omega, \mathcal{B}, f_*P')$  - Morphismus von W-Räumen

-  $X : \Omega \rightarrow \mathbb{R}$  integrierbare Zufallsvariable

**Lemma 2.46** (Transformationsformel).  $f^*X$  is integrierbar und es gilt  $\mathbb{E}_{P'}(f^*X) = \mathbb{E}_{f_*P'}(X)$ .

*Beweis.* für Zufallsvariablen mit endlich vielen Werten gilt die Formel

mittels  $X = X^+ - X^-$  Reduktion auf Fall:  $X$  nicht negativ

Annahme:  $X \geq 0$

- wähle mit Lemma 2.41 Folge  $(X_n)_{n \in \mathbb{N}}$  mit:
- $X_n$  einfach für alle  $n$
- $(X_n)_{n \in \mathbb{N}}$  monoton wachsend
- $\lim_{n \rightarrow \infty} X_n = X$
- dann auch  $\lim_{n \rightarrow \infty} f^*X_n = f^*X$

benutzen Satz 2.42 über monotone Konvergenz:

$$\mathbb{E}_{P'}(f^*X) = \lim_{n \rightarrow \infty} \mathbb{E}_{P'}(f^*X_n) = \lim_{n \rightarrow \infty} \mathbb{E}_{f_*P'}(X_n) = \mathbb{E}_{f_*P'}(X)$$

□

**Bemerkung 2.47.**  $(\Omega, \mathcal{B}, P)$  - W-Raum

$X : \Omega \rightarrow \mathbb{R}$  - integrierbar

-  $X_*P$  is W-Maß auf  $\mathbb{R}$

-  $\mathbb{E}_P(X) = \mathbb{E}_{X_*P}(x)$

Fazit: um Erwartungswerte numerischer Zufallsvariablen zu bestimmen, muß man eigentlich nur die Funktion  $x$  auf  $\mathbb{R}$  integrieren, allerdings bez. aller möglichen Maße □

### 2.3 Konstruktion von W-Maßen

$(\Omega, \mathcal{B})$  meßbarer Raum

Problem: Wie kann man ein W-Maß auf  $(\Omega, \mathcal{B})$  angeben ?

- Angabe aller Werte nicht oft praktikabel
- Angabe durch W-Vektor i.a. nicht möglich

**Beispiel 2.48.** Gleichverteilung auf  $[0, 1]$

kennen  $P$  auf Intervallen:

- $P([a, b]) = b - a$
- Borelsche  $\sigma$ -Algebra enthält aber viel kompliziertere Mengen
- Gibt es ein W-Maß mit diesen Werten auf Intervallen?
- Wenn ja, wie viele?
- Stimmen die Erwartungswerte?
- siehe Beispiel 2.17

□

**Beispiel 2.49.** unendlicher Münzwurf,

$$\Omega = \prod_{\mathbb{N}} \{z, w\}$$



- auf welchen Teilmengen kennen wir das Maß?
- geben  $x = (\omega_0, \dots, \omega_n)$  vor
- also etwa  $x = (z, w, z, z, w)$
- $A(x) :=$  alle Folgen mit Anfang  $x$
- $P(A(x)) = p^{\text{Anzahl } z} (1-p)^{\text{Anzahl } w}$
- $\mathcal{A} :=$  Menge von Ereignissen dieser Art
- $\mathcal{A}$  ist keine  $\sigma$ -Algebra
- nicht abgeschlossen unter Vereinigungen
- $\{(z, z, z, \dots)\}$  ist nicht in  $\mathcal{A}$
- aber abzählbarer Durchschnitt von Elementen aus  $\mathcal{A}$  (der Ereignisse  $A(z), A(z, z), A(z, z, z), \dots$ )

Frage: Kann man die Funktion  $P : \mathcal{A} \rightarrow [0, 1]$  zu einem W-Maß auf  $\sigma(\mathcal{A})$  fortsetzen?

- Wenn ja, auf wie viele Weisen.
- Hat das Ergebnis die gewünschten Invarianzeigenschaften?
- siehe Beispiel 2.18

□

**Beispiel 2.50.**  $F$  freier Ultrafilter auf  $\mathbb{N}$

$P : \mathcal{P}_\Omega \rightarrow [0, 1]$  wie in Ex. 2.2

- ist schon auf ganz  $\mathcal{P}_\mathbb{N}$  definiert und nicht  $\sigma$ -additiv
- Fortsetzbarkeit also nicht immer gegeben
- brauchen Annahmen
- z.B. muß  $P(A \cup B) = P(A) + P(B)$  gelten für disjunkte  $A, B$  in  $\mathcal{A}$  mit  $A \cup B \in \mathcal{A}$ .
- ähnliches gilt für  $\sigma$ -Additivität

□

algebraisches Problem:

- Fortsetzung auf alle Mengen, die man mit endlich vielen Operationen  $A \mapsto A^c$  und  $A, B \mapsto A \cup B$  aus  $\mathcal{A}$  erzeugen kann.

- das algebraische Problem ist also das der Fortsetzung von  $P$  von  $\mathcal{A}$  auf  $\mathcal{R}\langle\mathcal{A}\rangle$  (die Algebra, die von  $\mathcal{A}$  erzeugt wird)

diskutieren analytisches Problem:

- Fortsetzung von  $P$  von einer Algebra  $\mathcal{R}$  auf  $\sigma\langle\mathcal{R}\rangle$

- also auf Mengen, die man durch abzählbar viele Mengenoperationen bilden kann

- involviert Approximationen

algebraisches Problem

einfachster Fall: Fortsetzung von einer Partition

$\mathcal{A} = (A_i)_{i \in I}$  - endliche Partition von  $\Omega$

- d.h.:  $(A_i)_{i \in I}$  ist paarweise diskunkt und  $\bigcup_{i \in I} A_i = \Omega$

- gebe  $P : \mathcal{A} \rightarrow [0, 1]$  vor

- einzige Bedingung  $\sum_{i \in I} P(A_i) = 1$

-  $\mathcal{R}\langle\mathcal{A}\rangle$  besteht genau aus den Mengen der Form

—  $A(J) := \bigcup_{j \in J} A_j$  mit  $J \subseteq I$

— für  $A \in \mathcal{R}\langle\mathcal{A}\rangle$  gilt  $A = A(J)$  für eindeutig bestimmtes  $J \subseteq I$

— nämlich:  $J = \{j \in I \mid A_j \subseteq A\}$

- setze  $P(A(J)) := \sum_{j \in J} P(A_j)$

- das definiert Fortsetzung von  $P$  zu additiver Mengenfunktion auf  $\mathcal{R}\langle\mathcal{A}\rangle$

- Check: Übungsaufgabe

— benutze:  $A(J) \cup A(J') = A(J \cup J')$ ,  $A(J) \cap A(J') = A(J \cap J')$

**Beispiel 2.51.**  $\Omega = \{z, w\}^{\mathbb{N}}$

fixieren  $N$  in  $\mathbb{N}$

$x = (\omega_1, \dots, \omega_N)$

- Partition  $\mathcal{A}_N$  bestehe aus den Mengen  $A(x)$  für alle Anfänge  $x$  der Länge  $N$

- also  $x \in \prod_{i=0}^N \{z, w\}$

- ist endliche Partition von  $\Omega$  durch  $2^N$  Elemente

setze wie oben  $P_N(A(x)) := p^{\text{Anzahl } z} (1-p)^{\text{Anzahl } w}$

- Check der Normierung:

$$\sum_{x \in \prod_{i=0}^N \{z, w\}} P_N(A(x)) = \sum_{n=0}^N \binom{N}{n} p^n (1-p)^{N-n} = (p + (1-p))^N = 1$$

durch Ausdehnung auf Algebra: erhalten Mengenfunktion  $P_N$  auf  $\mathcal{R}(\mathcal{A}_N)$

□

**Beispiel 2.52.** Lebesguemaß auf  $[0, 1)$

fixieren  $N$  in  $\mathbb{N}$

$\mathcal{A}_N$  - dyadische Partition durch Intervalle der Länge  $2^{-N}$

- Elemente:  $[\frac{i}{2^N}, \frac{i+1}{2^N})$

- genau  $2^N$  Intervalle

setze  $P([\frac{i}{2^N}, \frac{i+1}{2^N})) := 2^{-N}$

check der Normierung: Anzahl der Intervalle  $\cdot$  Maß der Intervalle =  $2^N \cdot 2^{-N} = 1$

durch Ausdehnung auf Algebra: erhalte Mengenfunktion  $P_N$  auf  $\mathcal{R}(\mathcal{A}_N)$

□

haben in den Beispielen für jedes  $N$  in  $\mathbb{N}$  eine Partition  $\mathcal{A}_N$  und eine entsprechende Ausdehnung  $P_N$  des Maßes

- wollen eigentlich ein Maß auf der Algebra, welches von allen diese Partitionen zusammen erzeugt wird.

- Konstruktion geht wie folgt:

Folgen von Verfeinerungen

-  $(\mathcal{A}_N)_{N \in \mathbb{N}}$  - Folge von endlichen Partitionen

- Notation:  $\mathcal{A}_N = (A_{N,i})_{i \in I_N}$

-  $\mathcal{A}_{N+1}$  verfeinere  $\mathcal{A}_N$ :

- d.h.: jedes Mitglied  $A_{N+1,j}$  von  $\mathcal{A}_{N+1}$  ist in (genau) einem Mitglied  $A_{N,i}$  von  $\mathcal{A}_N$  enthalten

- jedes Mitglied von  $\mathcal{A}_{N+1}$  ist Vereinigung einer wohlbestimmten endlichen Familie von Mitgliedern von  $\mathcal{A}_N$

- sei Folge  $(P_N : \mathcal{A}_N \rightarrow [0, 1])_{i \in I}$  gegeben

- mit Normierung  $\sum_{i \in I_N} P_N(A_{N,i}) = 1$  für alle  $N$

- fordern Verträglichkeitsbedingung:

$$- P_N(A_{N,i}) = \sum_{j \in I_{N+1}, A_{N+1,j} \subseteq A_{N,i}} P_{N+1}(A_{N+1,j})$$

-  $P_N$  dehnt sich wie oben aus zu normierter additiver Mengenfunktion  $P_N$  auf  $\mathcal{R}\langle \mathcal{A}_N \rangle$

-

- dann gilt für  $M \leq N$ :

—  $\mathcal{R}\langle \mathcal{A}_M \rangle \subseteq \mathcal{R}\langle \mathcal{A}_N \rangle$

—  $(P_N)|_{\mathcal{R}\langle \mathcal{A}_M \rangle} = P_M$

— check: Übungsaufgabe

— Induktion nach  $N - M$

— Beginn: Fall  $N = M + 1$  - benutze Verträglichkeit

definiere additive normierte Mengenfunktion  $P : \mathcal{R}\langle \bigcup_{N \in \mathbb{N}} \mathcal{A}_N \rangle \rightarrow [0, 1]$  wie folgt

- sei  $A$  in  $\mathcal{R}\langle \bigcup_{N \in \mathbb{N}} \mathcal{A}_N \rangle$

- es gibt  $N$  in  $\mathbb{N}$  so daß  $A \in \mathcal{R}\langle \mathcal{A}_N \rangle$

- setzen  $P(A) = P_N(A)$

— ist unabhängig von Wahl von  $N$

— checke: ist additiv und normiert

**Beispiel 2.53.** unendlicher Münzwurf (Fortsetzung von Beispiel 2.51)

$\mathcal{A}_{N+1}$  verfeinert  $\mathcal{A}_N$  (klar)

$(P_N)_{N \in \mathbb{N}}$  is verträglich:

-  $x = (\omega_1, \dots, \omega_N)$

-  $z(x)$  - Anzahl der  $z$  in  $x$ , analog  $w(x)$

$$\begin{aligned} P_{N+1}(A(x)) &= \sum_{i \in \{z, w\}} P_{N+1}(A(x, i)) = p^{z(x)+1}(1-p)^{w(x)} + p^{z(x)}(1-p)^{w(x)+1} \\ &= (p + (1-p))p^{z(x)}(1-p)^{w(x)} = p^{z(x)}(1-p)^{w(x)} = P_N(A(x)) \end{aligned}$$

$\mathcal{Z} := \mathcal{R}\langle \bigcup_{N \in \mathbb{N}} \mathcal{A}_N \rangle$  heißt Algebra der Zylindermengen

- erhalten additive normierte Mengenfunktion  $P : \mathcal{Z} \rightarrow [0, 1]$

□

**Beispiel 2.54.** Gleichverteilung auf  $[0, 1]$  (Fortsetzung von 2.52)

$\mathcal{A}_{N+1}$  verfeinert  $\mathcal{A}_N$  (klar)

$(P_N)_{N \in \mathbb{N}}$  is verträglich:

$\mathcal{D} := \mathcal{R}\langle \bigcup_{N \in \mathbb{N}} \mathcal{A}_N \rangle$  heißt dyadische Algebra

- erhalten additive normierte Mengenfunktion  $P : \mathcal{D} \rightarrow [0, 1]$

□

haben bisher algebraisches Fortsetzungsproblem betrachtet

- diskutieren jetzt analytisches Problem

$P : \mathcal{A} \rightarrow [0, 1]$  - additive Mengenfunktion auf einer Algebra  $\mathcal{A}$

**Satz 2.55** (Caratheódory). *Wenn  $P$   $\sigma$ -additiv ist, dann gibt es genau eine Fortsetzung von  $P$  zu einem  $W$ -Maß auf  $\sigma\langle \mathcal{A} \rangle$ .*

*Beweis.* Maßtheorie

Idee:

1. Schritt: setzen  $P$  zu  $\sigma$ -subadditiven äußeren Maß fort:

-  $\tilde{P} : \mathcal{P}_\Omega \rightarrow [0, 1]$

-  $\tilde{P}(A) := \inf \sum_{i \in I} P(A_i)$

- Infimum über abzählbare Familien  $(A_i)_{i \in I}$  in  $\mathcal{A}$  mit  $A \subseteq \bigcup_{i \in I} A_i$

2. Schritt: zeigen daß  $P := \tilde{P}|_{\sigma(\mathcal{A})}$  das gewünschte W-Maß ist

□

**Beispiel 2.56.** unendlicher Münzwurf: Fortsetzung von Beispiel 2.53

müssen  $\sigma$ -Additivität von  $P : \mathcal{Z} \rightarrow [0, 1]$  zeigen:

Hilfsdiskussion:

-  $\{z, w\}$  ist kompakter topologischer Raum

- Tychonov:  $\Omega := \prod_{\mathbb{N}} \{z, w\}$  ist kompakter topologischer Raum mit Produkttopologie

-  $\text{pr}_{[0, \dots, N]} : \Omega \rightarrow \prod_{i=0}^n \{z, w\}$  ist stetig (universelle Eigenschaft der Produkttopologie)

-  $\{x\} \in \prod_{i=0}^n \{z, w\}$  abgeschlossen

-  $A(x) = \text{pr}_{[0, \dots, N]}^{-1}(\{x\})$  ist abgeschlossen

alle Elemente in  $\mathcal{Z}$  sind endliche Vereinigungen von abgeschlossenen Teilmengen und damit auch abgeschlossen, also kompakt

allg. topologische Tatsache: sei  $(A_i)_{i \in I}$  absteigende Familie von kompakten Mengen in einem topologischen Raum

- dann:  $\bigcap_{i \in I} A_i = \emptyset$  impliziert: es gibt  $i$  in  $I$  mit  $A_i = \emptyset$

betrachten absteigende Familie  $(A_i)_{i \in I}$  in  $\mathcal{Z}$  mit  $\bigcap_{i \in I} A_i = \emptyset$

- nach Lemma 2.22 zu zeigen:  $\lim_{i \rightarrow \infty} P(A_i) = 0$

- das ist aber klar, da  $A_i = \emptyset$  for große  $i$  wegen allg. topologischem Fakt

Nach Satz 2.55 erhalten wir Fortsetzung von  $P$  zu W-Maß auf  $\sigma(\mathcal{Z})$  -  $\sigma$ -Algebra der Zylindermengen

- ist auch die Borelsche  $\sigma$ -Algebra auf  $\Omega$  (als kompakter Raum)

Im Fall  $p = 1/2$

- Invarianz unter  $T_i$  (Flip des  $i$ ten Ergebnisses:
- $P$  auf  $\mathcal{Z}$  ist  $T_i$ -invariant (Formel)
- wegen Eindeutigkeit in Satz 2.55 gilt  $T_{i,*}P = P$  für fortgesetztes W-Maß

können jetzt das Auftreten der geometrischen Verteilung in Beispiel 2.12 mikroskopisch begründen:

- betrachten Zufallsvariable:  $f : \Omega \rightarrow \mathbb{N}$
- $f((\omega_i)_i) = \min\{i \mid \omega_i = z\}$  - erstes Auftreten von Zahl
- $f$  ist meßbar:
- $f^{-1}(\{n\}) = A(\underbrace{(w, \dots, w)}_{n \text{ mal}})$  - Zählung beginnt bei 0
- berechnen  $f_*P$
- geben Vektor an
- $n \mapsto P(f^{-1}(\{n\})) = P_n(\underbrace{(w, \dots, w)}_{n \text{ mal}}) = (1-p)^n p$
- in der Tat: geometrische Verteilung zum Parameter  $p$

□

**Beispiel 2.57.** Gleichverteilung auf  $[0, 1]$ : Fortsetzung von Beispiel 2.52

- $P$  - ist  $\sigma$ -additiv auf  $\mathcal{D}$
- erhalten Fortsetzung zu W-Maß auf  $\sigma(\bigcup_N \mathcal{A}_N)$
- ist Lebesgue Maß auf Borelscher  $\sigma$ -Algebra
- Details wurden in der Maßtheorie behandelt

Erwartungswerte stetiger Zufallsvariablen:

- $X : [0, 1] \rightarrow \mathbb{R}$  stetig ist:
- dann gilt  $\mathbb{E}(X) = \int_0^1 X(t) dt$  (Riemannintegral)

□

**Beispiel 2.58.** Ermüdender Wanderer

Wanderer started by 0 wirft vor dem  $n$ -ten Schritt Münze

- bei  $z$  bleibt er stehen und bei  $w$  geht er einen  $2^{-n-1}$ -tel Schritt nach rechts

Ereignis: wo ist der Wanderer nach unendlich vielen Schritten:

interpretieren  $z$  als 0 und  $w$  als 1

-  $\Omega = \prod_{i=0}^{\infty} \{0, 1\}$

-  $f : \Omega \rightarrow \mathbb{R}, \quad f(\omega) := \sum_{i=0}^{\infty} \omega_i 2^{-i-1}$

- ist das eine Zufallsvariable?

-  $\text{pr}_i : \Omega \rightarrow \{0, 1\} \subseteq \mathbb{R}$  ist meßbar

-  $\cdot 2^{-i-1} : \mathbb{R} \rightarrow \mathbb{R}$  ist stetig und da mit meßbar

— also  $\Omega \rightarrow \mathbb{R}, \quad \omega \mapsto \omega_i 2^{-i-1}$  ist meßbar

- eine Summe meßbarer Funktionen ist meßbar

— hier ist das Argument:

—  $g, h : \Omega' \rightarrow \mathbb{R}$  - meßbar

—  $(g, h) : \Omega' \rightarrow \mathbb{R}^2$  - meßbar

— (universelle Eigenschaft des Produkts und Fakt, daß Produkt- $\sigma$ -Algebra auf  $\mathbb{R}^2$  die Borelsche  $\sigma$ -Algebra ist )

—  $+$  :  $\mathbb{R}^2 \rightarrow \mathbb{R}$  ist stetig und damit meßbar

— also ist  $g + h : \Omega' \rightarrow \mathbb{R}$  meßbar

- also ist  $\sum_{i=0}^N \omega_i 2^{-i-1}$  für jedes  $N$  in  $\mathbb{N}$  meßbar

-  $f$  ist punktwiser Grenzwert einer Folge von meßbaren Funktionen

— damit ist  $f$  meßbar

- müssen  $f_*P$  bestimmen:



$x = (x_0, \dots, x_N)$  - Anfang der Länge  $N$

- sei  $d(x) = \sum_{i=0}^N x_i 2^{-i-1}$

-  $2d(x)$  ist rationale Zahl mit der dyadischen Entwicklung  $x$

-  $f^{-1}([d(x), d(x) + 2^{-N+1}]) = A(x)$

- Oberer Randpunkt ist Bild von  $(x, 1, 1, 1, \dots)$

—  $P(\text{Punkt}) = 0$

-  $(f_*)P([d(x), d(x) + 2^{-N+1}]) = 2^{-N}$

-  $f_*P$  stimmt mit Lebesguemaß auf dyadischer Partition überein

-  $f_*P$  ist Lebesguemaß auf  $[0, 1]$

also  $f_*P$  is Lebesguemaß auf  $[-1/2, 1/2]$

- die Endposition des Wanderers wird durch die Gleichverteilung modelliert

Anwendung:

$\phi : [0, 1] \rightarrow \mathbb{R}$  stetig:

-  $\mathbb{E}_P(f^*\phi) = \mathbb{E}_{f_*P}\phi = \int_0^1 \phi(t) dt$

Simulation: 10000 mal 100 Schritte

Simulation von  $\int_0^1 \sin(\pi t)^2 dt = 1/2$

□

## 2.4 Sprechweisen für den Umgang mit Zufallsvariablen

$(\Omega, \mathcal{B}, P)$  - W-Raum

$(\Omega_i, \mathcal{B}_i)_{i \in I}$  - Familie meßbarer Räume

-  $(f_i)_{i \in I}$  Familie von meßbaren Abbildungen

- bilden  $(\Omega', \mathcal{B}') := \prod_{i \in I} (\Omega_i, \mathcal{B}_i)$  mit Produkt- $\sigma$ -Algebra

-  $(f_i)_i : \Omega \rightarrow \Omega'$  kanonische Abbildung (universelle Eigenschaft des Produktes in **Meas**)

-  $P' := ((f_i)_i)_*P$

- Achtung: das ist nicht notwendig ein Produktmaß
- falls die Familie  $(f_i)_{i \in I}$  nicht unabhängig ist
- für jede Zufallsvariable,  $X : \Omega' \rightarrow \mathbb{R}$  gilt :  $\mathbb{E}_P((f_i)_i)^* X = \mathbb{E}_{P'}(X)$
- wir brauchen also gar nicht mit  $(\Omega, \mathcal{B}, P)$  sondern nur mit  $(\Omega', \mathcal{B}', P')$  zu arbeiten

Sprechweise:

- sei  $X$  eine Zufallsvariable mit Verteilung  $P$  (Schreibweise  $X \sim P$ ) bedeutet:
- es gibt da einen W-Raum  $(\Omega', \mathcal{B}', P')$  und eine Zufallsvariable  $X : \Omega' \rightarrow \mathbb{R}$  mit  $X_* P' = P$
- kann für alle Fragen über  $X$  mit  $(\mathbb{R}, \mathcal{B}, P)$  arbeiten
- z.B.
  - $\mathbb{E}(g(X)) = \mathbb{E}(g)$
  - $\mathbb{E}(X) = \mathbb{E}(x)$
  - $\text{Var}(X) = \mathbb{E}((x - \mathbb{E}(x))^2)$

- sei  $X, Y$  zwei Zufallsvariablen mit gemeinsamer Verteilung  $P$  bedeutet:
- es gibt da einen W-Raum  $(\Omega', \mathcal{B}', P')$  und Zufallsvariablen  $(X, Y) : \Omega' \rightarrow \mathbb{R}^2$  mit  $(X, Y)_* P' = P$  auf  $\mathbb{R}^2$
- kann für alle Fragen über  $(X, Y)$  mit  $(\mathbb{R}^2, \mathcal{B}, P)$  arbeiten
  - hier benutzen wir:  $\mathcal{B}_{\mathbb{R}^2}$  (Borel) ist die Produkt  $\sigma$ -Algebra
- z.B.
  - $\mathbb{E}(g(X, Y)) = \mathbb{E}(g)$
  - $\mathbb{E}(X + Y) = \mathbb{E}(x + y)$
  - $\text{Cov}(X, Y) = \mathbb{E}((x - \mathbb{E}(x))(y - \mathbb{E}(y)))$

Warnung: es reicht i.a. nicht, die Randverteilungen  $P_X := X_* P'$  und  $P_Y = Y_* P'$  zu kennen

- außer wenn man zusätzlich weiss, daß  $X$  und  $Y$  unabhängig sind
- dann ist  $(X, Y)_* P$  das Produkt von  $P_X$  und  $P_Y$  auf  $\mathbb{R} \times \mathbb{R} \cong \mathbb{R}^2$

sei  $(X_i)_{i \in I}$  eine Familie von Zufallsvariablen mit gemeinsamer Verteilung  $P$  bedeutet:

- es gibt da einen W-Raum  $(\Omega', \mathcal{B}', P')$  und Zufallsvariablen  $(X_i : \Omega' \rightarrow \mathbb{R})_{i \in I}$  mit  $((X_i)_i)_* P' = P$  auf  $\prod_{i \in I} (\mathbb{R}, \mathcal{B})$

sei  $(X_i)_{i \in I}$  eine Familie unabhängiger von Zufallsvariablen mit Verteilungen  $(P_i)_{i \in I}$  bedeutet:

- es gibt da einen W-Raum  $(\Omega', \mathcal{B}', P')$  und Zufallsvariablen  $(X_i : \Omega' \rightarrow \mathbb{R})_{i \in I}$  so daß  $((X_i)_i)_* P' = P$  das Produktmaß der Familie  $((X_i)_* P')_{i \in I}$  ist.

in jedem Fall beschreibt  $(\Omega', \mathcal{B}', P')$  alle Details des Experiments die in der Regel implizit und in der Praxis meist unbekannt sind

- die Variablen mit ihrer Verteilung reflektieren die für uns interessanten Eigenschaften
- diese werden durch  $(\prod_{i \in I} (\mathbb{R}, \mathcal{B}), P)$  modelliert
- Fazit: Wir brauchen in der Praxis  $(\Omega', \mathcal{B}', P')$  nicht zu kennen, nur die Variablen und ihre gemeinsame Verteilung

## 2.5 Dichtefunktionen

$(\Omega, \mathcal{B}, \mu)$  - Maßraum mit Maß  $\mu$  (nicht notwendig normiert)

**Beispiel 2.59.** Beispiele:

-  $(\mathbb{R}, \mathcal{B}, \text{Leb})$ , Lebesguemaß

-  $(\mathbb{R}, \mathcal{P}_\Omega, |\cdot|)$  - Zählmaß □

-  $g : \Omega \rightarrow [0, \infty)$  - meßbar

-  $\int_\Omega g d\mu = 1$

- erhalten W-Maß  $g\mu$  durch

-  $A \mapsto P(A) := \int_\Omega \mathbf{1}_A g d\mu$

- Verifikationen: Maßtheorie

**Definition 2.60.** Die Funktion  $g$  heißt Dichte von  $P$  bezüglich  $\mu$ .

$\mu$  gegeben

- Welche W-Maße haben eine Dichte bezüglich  $\mu$ ?

**Definition 2.61.**  $P$  ist absolutstetig zu  $\mu$  wenn  $\mu(A) = 0$  auch  $P(A)$  impliziert.

wenn  $P$  Dichte bezüglich  $\mu$  hat:

- wenn  $\mu(A) = 0$
- dann  $\{\mathbf{1}_A g \neq 0\}$  ist  $\mu$ -Nullmenge
- also  $P(A) = \int_{\Omega} \mathbf{1}_A g d\mu = 0$
- also ist  $P$  absolutstetig zu  $\mu$

Reicht das als hinreichende Bedingung?

- nicht ganz aber fast

**Definition 2.62.**  $\mu$  ist  $\sigma$ -endlich, wenn es eine aufsteigende Folge  $(A_i)_{i \in \mathbb{N}}$  in  $\mathcal{B}$  gibt mit  $\mu(A_i) < \infty$  und  $\bigcup_{i \in \mathbb{N}} A_i = \Omega$ .

**Beispiel 2.63.** Beispiel:

Leb ist  $\sigma$ -endlich: setze  $A_i = [-2^i, 2^i]$

Zählmaß auf  $\mathbb{R}$  ist nicht  $\sigma$ -endlich

□

**Theorem 2.64** (Radon-Nikodym). Wenn  $\mu$  ein  $\sigma$ -endliches Maß ist und  $P$  absolutstetig zu  $\mu$  ist, dann existiert ein  $g$  mit  $P = g\mu$ . Wenn  $P = g'\mu$ , dann ist  $\mu(\{g \neq g'\}) = 0$

**Beispiel 2.65.** Gleichverteilte Zufallsvariable auf  $[a, b]$

$$P = \frac{\mathbf{1}_{[a,b]}}{b-a} \text{Leb}$$

$$- \mathbb{E}(X) = \frac{b-a}{2}$$

$$- \text{Var}(X) = \frac{(b-a)^2}{12}$$

$P$  hat keine Dichte bezüglich dem Zählmaß

- Begründung: Übungsaufgabe

□

Sei  $P$  ein  $W$ -Maß auf  $\mathbb{R}$ .

**Definition 2.66.** Die Zahlen

$$\mu_n := \mathbb{E}_P(x^n) = \int_{\mathbb{R}} x^n dP$$

für  $n \in \mathbb{N}$  heißen die Momente von  $P$  (sind definiert falls  $x^n$  integrierbar ist)

**Beispiel 2.67.** Exponentialverteilung mit Parameter  $\alpha$  in  $(0, \infty)$

$$\text{Exp}(\alpha) := \alpha e^{-\alpha x} \mathbf{1}_{[0, \infty)} \text{Leb}$$

$$\text{Normierung: } \int_0^\infty \alpha e^{-\alpha x} d\text{Leb} = 1$$

Berechnung der Momente von  $\text{Exp}(\alpha)$

$$(-\partial_\alpha)^n e^{-\alpha x} = x^n e^{-\alpha x}$$

$$\mu_n = \alpha \int_0^\infty x^n e^{-\alpha x} d\text{Leb} = \alpha (-\partial_\alpha)^n \int_0^\infty e^{-\alpha x} d\text{Leb} = \alpha (-\partial_\alpha)^n \frac{1}{\alpha} = \frac{n!}{\alpha^n}$$

□

### Dichte von $\text{Exp}(0.03)$

Anwendung:

Kernzerfall:

- beobachte radioaktiven Kern:
- Halbwertszeit  $T$
- mit welcher Wahrscheinlichkeit ist der Kern zur Zeit  $t$  noch nicht zerfallen:
- Zufallsvariable:  $t \in [0, \infty)$
- Ereignis:  $K_t :=$  kein Zerfall in  $[0, t]$
- Ansatz:  $P_\alpha(K_t) = e^{-t\alpha}$

Begründung des Ansatzes:

- für kleine  $\delta$ :

$$P(\text{Zerfall in } [t, t + \delta] \mid K_t) = \delta\alpha$$

- $\alpha$  noch zu bestimmen

- also

$$\frac{P(\text{Zerfall in } [0, t + \delta]) - P(\text{Zerfall in } [0, t])}{P(K_t)} = \delta\alpha$$

$$\frac{P(K_t) - P(K_{t+\delta})}{\delta} = \alpha P(K_t)$$

- $\delta \rightarrow 0$
- erhalten DGL:  $\partial_t P(K_t) = -\alpha P(K_t)$
- $P(K_t) = e^{-\alpha t}$
- Konstante durch Normierung:  $P(K_0) = 1$
- Bestimmung von  $\alpha$  aus Halbwertszeit:
  - $1/2 = P(K_T) = 1 - e^{-T\alpha}$
  - also  $\ln(1/2) = -T\alpha$ ,  $\alpha = -\frac{\ln(1/2)}{T}$
- Dichte:  $\partial_t P_\alpha(K_t) = \alpha e^{-\alpha t}$

□

**Beispiel 2.68.** Normalverteilte Zufallsvariable auf  $\mathbb{R}$  zum Erwartungswert  $\mu$  und Varianz  $\sigma^2$ :

$$\text{Norm}(\mu, \sigma) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{Leb}$$

- in der Tat:  $\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\text{Leb} = 1$
- $x - \mu \mapsto y$ :  $\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\text{Leb} = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} d\text{Leb}$
- $\mapsto z|\sigma$ :  $\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} d\text{Leb} = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} d\text{Leb} = 1$
- $\mathbb{E}(X) = \mu$
- $\text{Var}(X) = \sigma^2$

Dichte der Normalverteilung:  $\sigma = \sqrt{20}, \mu = 20$

haben später ganzes Kapitel über die Eigenschaften der Normalverteilung

□

## 2.6 0 – 1 - Gesetze

$(\Omega, \mathcal{B}, P)$  - W-Raum

$(A_i)_{i \in \mathbb{N}}$  - Folge in  $\mathcal{B}$

-  $A := \limsup_{i \rightarrow \infty} A_i = \bigcap_{k \in \mathbb{N}} \bigcup_{i \geq k} A_i$

-  $\omega \in A$  genau dann, wenn  $\omega \in A_i$  für unendlich viele  $i$

— die Menge der  $i \in \mathbb{N}$  mit  $\omega \in A_i$  ist unbeschränkt

**Lemma 2.69** (Borel-Cantelli).

1. Aus  $\sum_{i \in \mathbb{N}} P(A_i) < \infty$  folgt  $P(A) = 0$ .

2. Wenn  $(A_i)_{i \in \mathbb{N}}$  unabhängig ist, dann folgt aus  $\sum_{i \in \mathbb{N}} P(A_i) = \infty$  daß  $P(A) = 1$  gilt.

*Beweis.*

Zu 1.:

$k \mapsto \bigcup_{i \geq k} A_i$  ist absteigend

- nutzen Lemma 2.22.3

$$\begin{aligned} P(A) &= P\left(\bigcap_{k \in \mathbb{N}} \bigcup_{i \geq k} A_i\right) \\ &= \lim_{k \rightarrow \infty} P\left(\bigcup_{i \geq k} A_i\right) \\ &\leq \lim_{k \rightarrow \infty} \sum_{i=k}^{\infty} P(A_i) \\ &= 0 \end{aligned}$$

Zu 2.:

$m \mapsto \bigcap_{i=k}^m A_i^c$  ist absteigend

$$\begin{aligned}
P\left(\bigcap_{i \geq k} A_i^c\right) &\stackrel{\text{Lem. 2.22.3}}{=} \lim_{m \rightarrow \infty} P\left(\bigcap_{i=k}^m A_i^c\right) \\
&\stackrel{\text{Lem. 1.55}}{=} \lim_{m \rightarrow \infty} \prod_{i=k}^m P(A_i^c) \\
&= \lim_{m \rightarrow \infty} \prod_{i=k}^m (1 - P(A_i))
\end{aligned}$$

nutzen  $1 - x \leq e^{-x}$

$$\prod_{i=k}^m (1 - P(A_i)) \leq \prod_{i=k}^m e^{-P(A_i)} = e^{-\sum_{i=k}^m P(A_i)}$$

- geht gegen 0 für  $m \rightarrow \infty$

$$P(A) = 1 - P(A_c) = 1 - P\left(\bigcup_{k \in \mathbb{N}} \bigcap_{i \geq k} A_i^c\right) = 1 - \lim_{k \rightarrow \infty} P\left(\bigcap_{i \geq k} A_i^c\right) = 1 - 0 = 1$$

□

**Beispiel 2.70.** ziehen zufällig eine natürliche Zahlen  $n$  aus  $\mathbb{N}$

- W-Maß  $P$  auf  $\mathbb{N}$  noch unbekannt

- für jede Primzahl  $p$  betrachten wir das Ereignis  $A_p := \{n \mid p \text{ teilt } n\}$ :

-  $A_p \cup (1 + A_p) \cup \dots \cup (p - 1 + A_p) = \mathbb{N}$

- wollen aus Symmetriegründen:  $P(A_p) = 1/p$

Frage: Gibt es ein W-Maß auf  $\mathbb{N}$ , so daß  $(A_p)_{p \text{ prim}}$  unabhängig ist und  $P(A_p) = 1/p$  für alle Primzahlen  $p$  gilt.

$$\sum_{p \text{ prim}} 1/p = \infty$$

-  $A := \limsup_{p \rightarrow \infty} A_p$

-  $n \in A$  genau dann wenn  $n$  unendlich viele Primteiler hat



- also  $A = \emptyset$
- aber Borel-Cantelli sagt:  $P(A) = 1$ .

Das gesuchte Maß gibt es nicht.

□

$(\Omega, \mathcal{B}, P)$  - W-Raum

$(\mathcal{B}_n)_{n \in \mathbb{N}}$  - Familie von  $\sigma$ -Unteralgebren in  $\mathcal{B}$

-  $\mathcal{S}_n := \sigma(\bigcup_{m=n}^{\infty} \mathcal{B}_m)$

-  $\mathcal{S}_0 \supseteq \mathcal{S}_1 \supseteq \mathcal{S}_2 \supseteq \dots$  - absteigende Folge

**Definition 2.71.**  $\mathcal{S}_\infty := \bigcap_{n \in \mathbb{N}} \mathcal{S}_n$  heißt  $\sigma$ -Algebra der terminalen Ereignisse von  $(\mathcal{B}_n)_{n \in \mathbb{N}}$ .

**Beispiel 2.72.**  $\limsup_{n \rightarrow \infty} A_n$  ist terminal

genauer:

sei  $A_n \in \mathcal{B}_n$  für alle  $n \in \mathbb{N}$

-  $A = \limsup_{n \rightarrow \infty} A_n \in \mathcal{S}_\infty$

- in der Tat: für jedes  $l$  in  $\mathbb{N}$  gilt:

-  $A = \bigcap_{k \in \mathbb{N}} \bigcup_{n=k}^{\infty} A_n = \bigcap_{k=l}^{\infty} \bigcup_{n=k}^{\infty} A_n \in \mathcal{S}_l$

— da  $k \mapsto \bigcup_{n=k}^{\infty} A_n$  absteigend ist

□

**Beispiel 2.73.** Existenz von Grenzwerten

$(X_i)_{i \in \mathbb{N}}$  Folge von Zufallsvariablen

- setze  $\mathcal{B}_i := X_i^{-1}(\mathcal{B}_{\mathbb{R}})$

**Lemma 2.74.** Die Ereignisse

1.  $\{\limsup_{i \rightarrow \infty} X_i \leq c\}$
2.  $\{\liminf_{i \rightarrow \infty} X_i \geq c\}$
3.  $\{\lim_{i \rightarrow \infty} X_i \text{ existiert}\}$

4.  $\{\lim_{i \rightarrow \infty} X_i \text{ existiert und liegt in } [a, b]\}$  sind terminal

*Beweis.*

Zu 1.: für alle  $m$  gilt:

$$\{\limsup_{i \rightarrow \infty} X_i \leq c\} = \bigcap_{n \in \mathbb{N}} \bigcup_{k \in \mathbb{N}} \bigcap_{l=k}^{\infty} \{X_l \leq c + \frac{1}{n}\} = \bigcap_{n \in \mathbb{N}} \bigcup_{k=m}^{\infty} \bigcap_{l=k}^{\infty} \{X_l \leq c + \frac{1}{n}\}$$

- da  $k \mapsto \bigcap_{l=k}^{\infty} \{X_l \leq c + \frac{1}{n}\}$  aufsteigend ist

- rechte Seite in  $\mathcal{S}_m$

- also  $\limsup_{i \rightarrow \infty} \{X_i \leq c\} \in \mathcal{S}_{\infty}$

□

Zu 2.: analog

Zu 3.:

-  $\limsup_{i \rightarrow \infty} X_i$  ist meßbar bezüglich  $\mathcal{B}_{\infty}$  nach 1. ( $c$  variieren!)

-  $\liminf_{i \rightarrow \infty} X_i$  ist meßbar bezüglich  $\mathcal{B}_{\infty}$  nach 2.

- also  $\{\limsup_{i \rightarrow \infty} X_i = \liminf_{i \rightarrow \infty} X_i\}$  meßbar bezüglich  $\mathcal{B}_{\infty}$

Zu 4.:

- folgt aus 1.-3.

□

**Beispiel 2.75.** Spiel mit Siegwahrscheinlichkeit  $p$ , unendlich oft wiederholt

$$\Omega = \prod_{n \in \mathbb{N}} \{g, v\}$$

- W-Maß auf  $\{g, v\}$  durch  $p(g) = p$  bestimmt.

-  $\mathcal{B}$  -  $\sigma$ -Algebra der Zylindermengen

-  $P$  - Produktmaß

$A :=$  unendlich viele Siege

-  $\mathcal{B}_n := \text{pr}_n^{-1} \mathcal{P}_{\{g, v\}}$

- $(\mathcal{B}_n)_{n \in \mathbb{N}}$  ist unabhängig
- $A_n := \text{pr}_n^{-1}(\{g\})$
- $A = \limsup_{n \rightarrow \infty} A_n$  ist terminal

mit Borel Cantelli Lemma 2.69 können wir  $P(A)$  ausrechnen:

1. Fall:

- $p > 0$ :
- $P(A_n) = p > 0$
- $\sum_{n=0}^{\infty} P(A_n) = \infty$
- $(A_n)_{n \in \mathbb{N}}$  ist unabhängig
- also  $P(A) = 1$

2. Fall:

- $p = 0$
- $P(A_n) = 0$
- $\sum_{n=0}^{\infty} P(A_n) < \infty$
- $P(A) = 0$  □

daß hier die Werte 0 oder 1 herauskommen, ist kein Zufall

**Definition 2.76.**  $(\mathcal{B}_n)_{n \in \mathbb{N}}$  heißt unabhängig, wenn jede Familie  $(A_n)_{n \in \mathbb{N}}$  mit  $A_n \in \mathcal{B}_n$  unabhängig ist.

**Satz 2.77** (Kolmogorovsches 0 – 1 - Gesetz). Wenn  $(\mathcal{B}_n)_{n \in \mathbb{N}}$  unabhängig ist und  $A$  in  $\mathcal{S}_{\infty}$ , dann gilt  $P(A) \in \{0, 1\}$ .

*Beweis.* Beweis in 2.6 □

$(\Omega, \mathcal{B}, P)$  - W-Raum

$(A_n)_{n \in \mathbb{N}}$  - unabhängige Folge in  $\mathcal{B}$

**Satz 2.78** (0 – 1 - Gesetz von Borel). Für alle  $\omega$  in  $\Omega$  gilt:

$$P(\{\omega \in A_n \text{ für unendliche viele } n\}) \in \{0, 1\} .$$

*Beweis.* setze  $\mathcal{B}_n := \sigma\{A_n\}$

- $(\mathcal{B}_n)_{n \in \mathbb{N}}$  - unabhängig
- bilden  $\mathcal{S}_\infty$  für diese Familie
  
- klar ist  $A_n \in \mathcal{B}_n$
- $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=0}^{\infty} \bigcup_{m=n}^{\infty} A_m \in \mathcal{S}_\infty$  ist terminal
- $\limsup_{n \rightarrow \infty} A_n = \{\omega \in A_n \text{ für unendliche viele } n\}$
- wenden Satz 2.77 an

□

**Beispiel 2.79.** Konvergenz einer Folge unabhängiger Zufallsvariablen

sei  $(X_n)_{n \in \mathbb{N}}$  - unabhängige Folge von Zufallsvariablen

- $\mathcal{B}_n := X_n^{-1}(\text{Borel})$
- dann ist  $(\mathcal{B}_n)_{n \in \mathbb{N}}$  unabhängig

$A := \{\lim_{n \rightarrow \infty} X_n \text{ existiert}\}$  - terminal bezüglich  $(\mathcal{B}_n)_{n \in \mathbb{N}}$

- dann gilt  $P(A) \in \{0, 1\}$

Konkret es Beispiel: Würfeln:

- $X_n :=$  Ergebnis vom  $n$ ten Wurf
- $\lim_{n \rightarrow \infty} X_n$  existiert falls ab einer Stelle immer dasselbe Ergebnis herauskommt
- $P(A) = 0$

□

**Beispiel 2.80.** unendlich of würfeln:

$A := \{\text{es kommt nur endlich oft die } 6\}$

- $P(A) \in \{0, 1\}$  (in der Tat  $P(A) = 0$ )

$B := \{\text{für jedes } n \text{ kommt die } 6 \text{ mehr als } n\text{-mal hintereinander}\}$

-  $P(B) \in \{0, 1\}$  (in der Tat  $P(B) = 1$ )

□

*Beweis.* Beweis von Satz 2.77

Idee: zeigen:

$A$  ist unabhängig von sich selbst

-  $P(A) = P(A \cap A) = P(A)P(A)$  impliziert  $P(A) \in \{0, 1\}$

Achtung: der Beweis ist lang

Vorbereitungen:

**Definition 2.81.** Ein Dynkinsystem ist eine Teilmenge  $\mathcal{D}$  von  $\mathcal{P}_\Omega$  mit folgenden Eigenschaften:

1.  $\Omega \in \mathcal{D}$ .

2. Aus  $A \in \mathcal{D}$  folgt  $A^c \in \mathcal{D}$ .

3.  $\mathcal{D}$  is abgeschlossen unter abzählbaren paarweise disjunkten Vereinigungen.

- ein auch unter  $\cap$  abgeschlossenes Dynkinsystem ist eine  $\sigma$ -Algebra

- für eine Teilmenge  $\mathcal{E}$  von  $\mathcal{P}_\Omega$  ist

$$\delta\langle\mathcal{E}\rangle := \bigcap_{\mathcal{E} \subseteq \mathcal{D}} \mathcal{D}$$

(Durchschnitt über  $\mathcal{E}$ -enthaltende Dynkinsysteme) das kleinste  $\mathcal{E}$ -enthaltende Dynkinsystem

- sei  $\mathcal{E}$  eine Partition

- dann ist  $\delta\langle\mathcal{E}\rangle$  die Menge aller abzählbaren Vereinigungen aus Elementen der Partition und deren Komplemente

betrachten unabhängige Familie  $\mathbf{E} := (\mathcal{E}_i)_{i \in I}$  von Teilmengen von  $\mathcal{B}$

$$\mathcal{U}(\mathbf{E}) := \{D \in \mathcal{B} \mid (\{D\}, (\mathcal{E}_i)_{i \in I}) \text{ ist unabhängig}\}$$

**Lemma 2.82.**  $\mathcal{U}(\mathbf{E})$  ist ein Dynkin System.

*Beweis.*

$\Omega \in \mathcal{U}(\mathbf{E})$  - klar

$D \in \mathcal{U}(\mathbf{E}) \Rightarrow D^c \in \mathcal{U}(\mathbf{E})$

- sei  $J \subseteq I$  endlich

- betrachte  $E_j \in \mathcal{E}_j$  für alle  $j$

$$\begin{aligned}
 P(D^c \cap \bigcap_{j \in J} E_j) &= P\left(\bigcap_{j \in J} E_j\right) - P\left(D \cap \bigcap_{j \in J} E_j\right) \\
 &= \prod_{j \in J} P(E_j) - P(D) \prod_{j \in J} P(E_j) \\
 &= (1 - P(D)) \prod_{j \in J} P(E_j) \\
 &= P(D^c) \prod_{j \in J} P(E_j)
 \end{aligned}$$

$\mathcal{U}(\mathbf{E})$  abgeschlossen unter abzählbaren paarweise disjunkten Vereinigungen

- betrachte  $(D_i)_{i \in \mathbb{N}}$  - paarweise disjunkt in  $\mathcal{U}(\mathbf{E})$

- zeigen:  $\bigcup_{i \in \mathbb{N}} D_i \in \mathcal{U}(\mathbf{E})$

- sei  $J \subseteq I$  endlich

- betrachte  $E_j \in \mathcal{E}_j$  für alle  $j$

$$\begin{aligned}
 P\left(\bigcup_{i \in \mathbb{N}} D_i \cap \bigcap_{j \in J} E_j\right) &= P\left(\bigcup_{i \in \mathbb{N}} (D_i \cap \bigcap_{j \in J} E_j)\right) \\
 &= \sum_{i \in I} P(D_i \cap \bigcap_{j \in J} E_j) \\
 &= \sum_{i \in I} P(D_i) \prod_{j \in J} P(E_j) \\
 &= P\left(\bigcup_{i \in I} D_i\right) \prod_{j \in J} P(E_j)
 \end{aligned}$$

□

**Lemma 2.83.** Wenn  $(\mathcal{E}_i)_{i \in I}$  eine unabhängige Familie ist, so auch  $(\delta\langle \mathcal{E}_i \rangle)_{i \in I}$ .

*Beweis.* Vorbemerkung:

fixiere  $i_0 \in I$

- setze  $\mathcal{U} := \mathcal{U}((\mathcal{E}_i)_{I \setminus \{i_0\}})$

- dann  $\mathcal{E}_{i_0} \subseteq \mathcal{U}$

- also  $\delta\langle \mathcal{E}_{i_0} \rangle \subseteq \mathcal{U}$  da  $\mathcal{U}$  ein Dynkinsystem ist

- also ist  $(\delta\langle \mathcal{E}_{i_0} \rangle, (\mathcal{E}_i)_{I \setminus \{i_0\}})$  unabhängig

eigentlicher Beweis:

$J \subseteq I$  endlich

sei betrachte  $A_j \in \delta\langle \mathcal{E}_j \rangle$  für alle  $j$  in  $J$

- ordne  $J$

- sei  $J = \{j_1, \dots, j_n\}$

-  $(\mathcal{E}_{j_1}, \dots, \mathcal{E}_{j_n})$  sind unabhängig

-  $(\delta\langle \mathcal{E}_{j_1} \rangle, \mathcal{E}_{j_2}, \dots, \mathcal{E}_{j_n})$  sind unabhängig

-  $(\delta\langle \mathcal{E}_{j_1} \rangle, \delta\langle \mathcal{E}_{j_2} \rangle, \mathcal{E}_{j_3}, \dots, \mathcal{E}_{j_n})$  sind unabhängig

- endliche Induktion

-  $(\delta\langle \mathcal{E}_{j_1} \rangle, \dots, \delta\langle \mathcal{E}_{j_n} \rangle)$  sind unabhängig

-  $(A_j)_{j \in J}$  ist unabhängig

□

eine Familie  $\mathcal{E}$  heie  $\cap$ -abgeschlossen, wenn sie abgeschlossen unter abzählbaren Durchschnitten ist

$(\mathcal{E}_i)_{i \in I}$  Familie von Teilmengen von  $\mathcal{B}$

**Lemma 2.84.** Annahme:

1.  $\mathcal{E}_i$  ist  $\cap$ -abgeschlossen für alle  $i$  in  $I$ .

2.  $(\mathcal{E}_i)_{i \in I}$  ist unabhängig.

Dann ist  $(\sigma\langle \mathcal{E}_i \rangle)_{i \in I}$  unabhängig.

*Beweis.*

$\mathcal{E}$  ist  $\cap$ -abgeschlossen impliziert:  $\delta\langle \mathcal{E} \rangle = \sigma\langle \mathcal{E} \rangle$

Behauptung folgt aus Lemma 2.83 □

$(\mathcal{E}_i)_{i \in I}$  unabhängige Familie von  $\cap$ -abgeschlossenen Teilmengen von  $\mathcal{B}$

-  $(I_j)_{j \in J}$  - Partition von  $I$

-  $\mathcal{F}_j := \sigma\langle \bigcup_{i \in I_j} \mathcal{E}_i \rangle$

**Lemma 2.85** (Blockbildung). *Die Familie  $(\mathcal{F}_j)_{j \in J}$  ist unabhängig.*

*Beweis.*

$\tilde{\mathcal{F}}_j$  bestehe aus Mengen der Form  $E_{i_1} \cap \dots \cap E_{i_k}$  mit  $i_l \in I_j$  und  $E_{i_l} \in \mathcal{E}_{i_l}$

- dann ist  $(\tilde{\mathcal{F}}_j)_{j \in J}$  ist unabhängige Familie

- Nachweis: Übungsaufgabe

-  $\tilde{\mathcal{F}}_j$  ist  $\cap$ -abgeschlossen für jedes  $j$  nach Konstruktion

- nach Lemma 2.84 ist  $(\sigma\langle \tilde{\mathcal{F}}_j \rangle)_{j \in J}$  ist unabhängig

zeigen, daß  $\sigma\langle \tilde{\mathcal{F}}_j \rangle = \mathcal{F}_j$

- klar ist  $\tilde{\mathcal{F}}_j \subseteq \mathcal{F}_j$

- also  $\sigma\langle \tilde{\mathcal{F}}_j \rangle \subseteq \mathcal{F}_j$

- Erzeuger von  $\mathcal{F}_j$  sind in  $\tilde{\mathcal{F}}_j$

- also  $\mathcal{F}_j \subseteq \sigma\langle \tilde{\mathcal{F}}_j \rangle$

also ist  $(\mathcal{F}_j)_{j \in J}$  ist unabhängig □



Beweis des Satzes:

- bilden Dynkingsystem  $\mathcal{U} := \mathcal{U}(\{A\})$
  - Blockbildung bezüglich Partition  $\{1, \dots, n\}, \{n+1, \dots\}$
  - $\mathcal{S}_{n+1}$  ist unabhängig von  $\sigma\langle \mathcal{B}_1 \cup \dots \cup \mathcal{B}_n \rangle =: \mathcal{A}_n$
  - $A$  terminal, also  $A \in \mathcal{S}_{n+1}$  impliziert:  $\mathcal{A}_n \subseteq \mathcal{U}$  ( $\mathcal{A}_n$  ist unabhängig von  $\{A\}$ )
  - also  $\mathcal{A}_n \subseteq \mathcal{U}$  für alle  $n$
  - $\mathcal{A} := \bigcup_{n \in \mathbb{N}} \mathcal{A}_n$  ist Dynkingsystem
  - $\mathcal{A} := \bigcup_{n \in \mathbb{N}} \mathcal{A}_n$  ist  $\cap$ -abgeschlossen
  - $\sigma\langle \mathcal{A} \rangle = \delta\langle \mathcal{A} \rangle \subseteq \mathcal{U}$
  - $\mathcal{B}_n \subseteq \mathcal{A}$  für alle  $n$  in  $\mathbb{N}$
  - also  $\mathcal{S}_n \subseteq \sigma\langle \mathcal{A} \rangle$  für alle  $n$  in  $\mathbb{N}$
  - also auch  $A \in \sigma\langle \mathcal{A} \rangle \subseteq \mathcal{U}$
- das zeigt:  $A$  unabhängig von  $A$  □

### 3 Analytische Grundlagen

#### 3.1 Abschätzungen

betrachten eine Funktion  $g : \mathbb{R} \rightarrow \mathbb{R}$

- $g$  ist affin falls  $g(x) = ax + b$
- wenn  $g$  affin ist, dann gilt

$$\mathbb{E}(g(X)) = \mathbb{E}(aX + b) = a\mathbb{E}(X) + b = g(\mathbb{E}(X))$$

**Definition 3.1.**  $g$  ist konvex, wenn für jedes  $x$  in  $\mathbb{R}$  eine affine Abbildung  $l : \mathbb{R} \rightarrow \mathbb{R}$  existiert mit  $g \geq l$  und  $g(x) = l(x)$ .

$l$  heißt Stützfunktion von  $g$  bei  $x$

$X$  - numerische Zufallsvariable

**Lemma 3.2** (Jensen-Ungleichung). *Wenn  $g$  konvex und integrabel ist dann gilt*

$$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X)) .$$

*Beweis.*

sei  $l$  Stützfunktion von  $g$  bei  $\mathbb{E}(X)$

- also  $l$  affin,  $g \geq l$  und  $g(\mathbb{E}(X)) = l(\mathbb{E}(X))$

$$\begin{aligned} \mathbb{E}(g(X)) &\geq \mathbb{E}(l(X)) \\ &= l(\mathbb{E}(X)) \\ &= g(\mathbb{E}(X)) \end{aligned}$$

□

**Lemma 3.3** (verallg. Chebyshev-Ungleichung). *Sei  $g : \mathbb{R} \rightarrow [0, \infty]$  monoton wachsend. Für jedes  $c$  in  $\mathbb{R}$  mit  $g(c) > 0$  gilt dann*

$$P(\{X \geq c\}) \leq \frac{\mathbb{E}(g(X))}{g(c)} .$$

*Beweis.*

das  $g$  monoton wächst, gilt  $\mathbf{1}_{\{X \geq c\}} \leq \frac{g(X)}{g(c)}$

- wende  $\mathbb{E}$  an

Begründung der Ungleichung

-  $\omega$  in  $\Omega$

-  $X(\omega) \geq c : \frac{g(X(\omega))}{g(c)} \geq 1$  und  $\mathbf{1}_{\{X \geq c\}}(\omega) = 1$

-  $X(\omega) < c : \frac{g(X(\omega))}{g(c)} \geq 0$  und  $\mathbf{1}_{\{X \geq c\}}(\omega) = 0$

□

$X$  - numerische Zufallsvariable

- nehmen  $g(x) = \max\{x, 0\}$  - ist nicht-negativ und konvex

- Zufallsvariable  $|X|$

**Korollar 3.4** (Chebyshev-Ungleichung). Für  $c > 0$  gilt

$$P(\{|X| \geq c\}) \leq \frac{\mathbb{E}(|X|)}{c}.$$

- nehmen  $g(x) = \max\{0, x\}^2$

- ist nicht negativ und konvex

- Zufallsvariable  $|X - E(X)|$

-  $\mathbb{E}(|X - E(X)|^2) = \text{Var}(X)$

**Korollar 3.5.** Für  $c > 0$  gilt

$$P(\{|X - \mathbb{E}(X)| \geq c\}) \leq \frac{\text{Var}(X)}{c^2}.$$

benutzen oft folgende Sprechweise

-  $X$  - Zufallsvariable

-  $P$  Aussage über die Werte

-  $X$  erfüllt  $P$  fast sicher, falls  $\{\omega \in \Omega \mid P(X(\omega))\} = 1$

**Lemma 3.6.**

1. Wenn  $\mathbb{E}(|X|) = 0$ , dann  $X = 0$  fast sicher

2. Wenn  $\text{Var}(X) = 0$ , dann  $X = \mathbb{E}(X)$  fast sicher.

*Beweis.*

Zu 1.:

Lemma 2.22 bei !

Korollary 3.4 bei !!

$$\begin{aligned}P(\{X = 0\}) &= P\left(\bigcap_{n=1}^{\infty} \{|X| < \frac{1}{n}\}\right) \\&\stackrel{!}{=} \lim_{n \rightarrow \infty} P(\{|X| < \frac{1}{n}\}) \\&= 1 - \lim_{n \rightarrow \infty} P(\{|X| \geq \frac{1}{n}\}) \\&\stackrel{!!}{\geq} 1 - \lim_{n \rightarrow \infty} n\mathbb{E}(|X|) \\&= 1\end{aligned}$$

also  $P(\{X = 0\}) = 1$

Zu 2.:

Lemma 2.22 bei !

Korollar 3.5 be !!

$$\begin{aligned}P(\{X \neq \mathbb{E}(X)\}) &= P\left(\bigcup_{n \in \mathbb{N}} \{|X - \mathbb{E}(X)| \geq \frac{1}{n}\}\right) \\&\stackrel{!}{=} \lim_{n \rightarrow \infty} P(\{|X - \mathbb{E}(X)| \geq \frac{1}{n}\}) \\&\stackrel{!!}{\leq} \limsup_{n \rightarrow \infty} n^2 \text{Var}(X) \\&= 0\end{aligned}$$

also  $P(\{X = \mathbb{E}(X)\}) = 1$

□

### 3.2 Konvergenz von Folgen von Zufallsvariablen

$(X_i)_{i \in \mathbb{N}}$  - Folge von Zufallsvariablen

setzen

set opp  $\in \{\lim, \sup, \inf, \limsup, \liminf\}$

$$X := \begin{cases} \lim_{i \rightarrow \infty} X_i & \text{Wert existiert} \\ 0 & \text{sonst} \end{cases}$$

**Lemma 3.7.**  $X$  ist eine Zufallsvariable.

*Beweis.*

müssen zeigen daß  $X$  meßbar ist:

Einzelfallbetrachtung:

benutzen Beispiel 2.33

-  $\{X_i \leq c\}$  ist meßbar für  $i$  in  $\mathbb{N}$ ,  $c$  in  $\mathbb{R}$

$\sup_i X_i$  ist meßbar:

-  $\{\sup_i X_i \leq c\} = \bigcap_{i \in \mathbb{N}} \{X_i \leq c\}$

$\limsup_i X_i$  ist meßbar:

-  $\{\limsup_i X_i \leq c\} = \bigcup_{i \in \mathbb{N}} \bigcap_{k \geq i} \{X_k \leq c\}$

$$\{\lim_i X_i \text{ existiert}\} = \{\liminf_i X_i = \limsup_i X_i\}$$

- dazu

-  $F := (\liminf_i X_i, \limsup_i X_i)$  -  $\mathbb{R}^2$ -wertig, meßbar

-  $\{\liminf_i X_i = \limsup_i X_i\} = F^{-1}(\text{diag}(\mathbb{R}))$

$$\lim_{i \rightarrow \infty} X_i = \mathbf{1}_{\{\lim_{i \rightarrow \infty} X_i \text{ existiert}\}} \limsup_i X_i$$

□

$(X_i)_{i \in \mathbb{N}}$  - Folge von Zufallsvariablen

$X$  - Zufallsvariable

**Definition 3.8.** Die Folge  $(X_i)_{i \in \mathbb{N}}$  konvergiert in Wahrscheinlichkeit gegen  $X$ , falls für jedes  $\epsilon$  in  $(0, \infty)$  gilt

$$\lim_{i \rightarrow \infty} P(\{|X_i - X| \geq \epsilon\}) = 0 .$$

Kurzschreibweise:  $X_i \xrightarrow{i \rightarrow \infty} X$  (in  $P$ )

**Definition 3.9.** Die Folge  $(X_i)_{i \in \mathbb{N}}$  konvergiert fast sicher gegen  $X$ , falls

$$P(\{\lim_{i \rightarrow \infty} X_i = X\}) = 1 .$$

Kurzschreibweise:  $X_i \xrightarrow{i \rightarrow \infty} X$  (f.s)

**Definition 3.10.** Die Folge  $(X_i)_{i \in \mathbb{N}}$  konvergiert in schwach gegen  $X$ , falls für jedes stetige beschränkte  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  gilt  $\lim_{i \rightarrow \infty} E(\phi(X_i)) = E(\phi(X))$ .

Kurzschreibweise:  $X_i \xrightarrow{i \rightarrow \infty} X$  (schwach)

wenn die Erwartungswerte alle existieren:

**Definition 3.11.** Die Folge  $(X_i)_{i \in \mathbb{N}}$  konvergiert im Mittel gegen  $X$ , falls  $\lim_{i \rightarrow \infty} \mathbb{E}(|X_i - X|) = 0$ .

Kurzschreibweise:  $X_i \xrightarrow{i \rightarrow \infty} X$  (i.m.)

**Lemma 3.12.**

1. Fast sichere Konvergenz impliziert Konvergenz in Wahrscheinlichkeit
2. Wenn  $\sum_{i \in \mathbb{N}} P(\{|X_i - X| > \epsilon\}) < \infty$ <sup>1</sup>, dann gilt fast sichere Konvergenz.

*Beweis.*

Zu 1.:

gelte  $X_i \xrightarrow{i \rightarrow \infty} X$  (f.s.)

$$- 1 = P(\{\lim_{i \rightarrow \infty} X_i = X\})$$

$$- \{\lim_{i \rightarrow \infty} X_i = X\} = \bigcap_{l \in \mathbb{N}} \bigcup_{i \in \mathbb{N}} \bigcap_{k \geq i} \{|X_k - X| \leq 1/l\}$$

---

<sup>1</sup>das ist ein Verschärfung von Konvergenz in Wahrscheinlichkeit

- $1 = P(\bigcap_{l \in \mathbb{N}} \bigcup_{i \in \mathbb{N}} \bigcap_{k \geq i} \{|X_k - X| \leq 1/l\})$  für alle  $l \in \mathbb{N}$
- äquiv. zu  $1 = P(\bigcup_{i \in \mathbb{N}} \bigcap_{k \geq i} \{|X_k - X| \leq 1/l\})$  für alle  $l$  in  $\mathbb{N}$
- äquiv. zu  $\lim_{i \rightarrow \infty} P(\bigcap_{k \geq i} \{|X_k - X| \leq 1/l\}) = 1$
- impliziert  $\lim_{i \rightarrow \infty} P(\{|X_i - X| \leq 1/l\}) = 1$
- also  $X_i \xrightarrow{i \rightarrow \infty} X$  (in  $P$ )

Zu 2.:

Borel-Cantelli 2.69:

$$- P(\limsup_{i \rightarrow \infty} \{|X_i - X| > \epsilon\}) = 0$$

$$\begin{aligned}
 P(\{\lim_{i \rightarrow \infty} X_i = X\}) &= P(\bigcap_{n=0}^{\infty} \bigcup_{k=0}^{\infty} \bigcap_{i=k}^{\infty} \{|X_i - X| < 1/n\}) \\
 &= \lim_{n \rightarrow \infty} P(\bigcup_{k=0}^{\infty} \bigcap_{i=k}^{\infty} \{|X_i - X| < 1/n\}) \\
 &= 1 - \lim_{n \rightarrow \infty} P(\bigcup_{k=0}^{\infty} \bigcap_{i=k}^{\infty} \{|X_i - X| \geq 1/n\}) \\
 &= 1 - \lim_{n \rightarrow \infty} P(\limsup_{i \rightarrow \infty} \{|X_i - X| \geq 1/n\}) \\
 &= 1
 \end{aligned}$$

□

**Beispiel 3.13.** Konvergenz in  $P \not\Rightarrow$  Konvergenz fast sicher

$(X_n)_{n \in \mathbb{N}}$  - Folge unabhängiger Zufallsvariablen

-  $P_n$  - Verteilung von  $X_n$

$$- P_n := \frac{1}{n+1} \delta_1 + (1 - \frac{1}{n+1}) \delta_0$$

Beh:  $X_i \rightarrow 0$  in Wahrscheinlichkeit:

-  $(\Omega, \mathcal{B}, P)$  im Hintergrund

$$- P(\{|X_n - 0| \geq \epsilon\}) = P_n(\{|X_n| \geq \epsilon\}) = \frac{1}{n+1} \xrightarrow{n \rightarrow \infty} 0$$

-

Beh: es gilt nicht  $X_i \rightarrow 0$  fast sicher

haben  $\sum_{n \in \mathbb{N}} P(\{|X_n - 0| \geq 1/2\}) = \sum_{n \in \mathbb{N}} \frac{1}{n+1} = \infty$

- also nach Borel-Cantelli 2.69:

-  $(\{|X_n - 0| \geq 1/2\})_n$  ist Folge unabhängiger Ereignisse

-  $P(\limsup_n \{|X_n - 0| \geq 1/2\}) = 1$

- also  $P(\limsup_n \{|X_n - 0| < 1/2\}) = 0$

- für fast sicherer Konvergenz müßte aber 1 herauskommen

□

**Lemma 3.14.** *Konvergenz im Mittel  $\Rightarrow$  Konvergenz in Wahrscheinlichkeit*

*Beweis.*

gelte:  $X_n \xrightarrow{n \rightarrow \infty} X$  im Mittel

- es gilt für alle  $\epsilon > 0$ :

-  $\epsilon P(|X_n - X| \geq \epsilon) \leq \mathbb{E}(|X_n - X|)$

-  $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|) = 0$  impliziert:

-  $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$

- also  $X_n \xrightarrow{n \rightarrow \infty} X$  in Wahrscheinlichkeit

□

**Beispiel 3.15.** Konvergenz in Wahrscheinlichkeit  $\not\Rightarrow$  Konvergenz im Mittel

$\Omega := [-1, 1]$  mit  $1/2d\text{Leb}$  als W-Maß

$X_n := (n+1)e^{-(n+1)^2 x^2/2}$

- dann  $X_n \xrightarrow{n \rightarrow \infty} 0$  in Wahrscheinlichkeit

- das Intervall  $\{|X_n - 0| \geq \epsilon\}$  wird mit wachsendem  $n$  immer kürzer



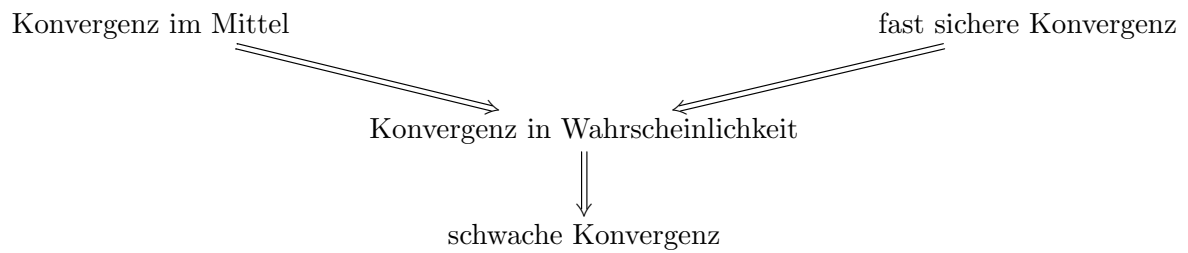
- nicht aber  $X_n \xrightarrow{n \rightarrow \infty} 0$  im Mittel:

-  $\mathbb{E}(|X_n - 0|) \xrightarrow{n \rightarrow \infty} \sqrt{2\pi}$

□

Übersicht über die Relationen zwischen den Konvergenzbegriffen:

- alle Implikationen echt



### 3.3 Gesetze der großen Zahl

betrachten Folge von Zufallsvariablen:  $(X_i)_{i \in \mathbb{N}}$

- Annahme: Erwartungswerte existieren und hängen nicht von  $i$  ab

- also ex.  $m$  in  $\mathbb{R}$  mit  $\mathbb{E}(X_i) = m$  für alle  $i$  in  $\mathbb{N}$

-  $S_n := \frac{1}{n} \sum_{i=0}^{n-1} X_i$

- Mittelwert der ersten  $n$  Folgenglieder

- das ist eine Zufallsvariable

- untersuchen das Verhalten von  $S_n$  für  $n \rightarrow \infty$

**Definition 3.16.** Die Folge  $(X_i)_{i \in \mathbb{N} \setminus \{0\}}$  erfüllt:

1. das schwache Gesetz der großen Zahl, falls für alle  $\epsilon$  in  $(0, \infty)$  gilt:

$$P(\{|S_n - m| \geq \epsilon\}) \xrightarrow{n \rightarrow \infty} 0 .$$

2. das starke Gesetz der großen Zahl, falls für alle  $\epsilon$  in  $(0, \infty)$

$$P(\{\limsup_{n \rightarrow \infty} |S_n - m| \leq \epsilon\}) = 1 .$$

**Bemerkung 3.17.**

schwache Gesetz der großen Zahl ist äquivalent zu:  $S_n \xrightarrow{n \rightarrow \infty} m$  (in  $P$ )

starkes Gesetz der großen Zahl ist äquivalent zu:  $S_n \xrightarrow{n \rightarrow \infty} m$  (f.s.) □

**Korollar 3.18.** *Stark  $\Rightarrow$  schwach*

*Beweis.* Lemma 3.12.1 □

zeigen hinreichendes Kriterium für schwaches Gesetz der großen Zahl

**Lemma 3.19.** *Annahme:*

1. Es gilt  $\sup_{i \in \mathbb{N}} \text{Var}(X_i) =: v < \infty$
2. Für alle  $i, j$  in  $\mathbb{N}$  mit  $i \neq j$  gilt  $\text{Cov}(X_i, X_j) = 0$ .

Dann erfüllt  $(X_i)_{i \in \mathbb{N}}$  das schwache Gesetz der großen Zahl.

Erläuterung zur Bedingung 2.:

- die Glieder der Folge paarweise unkorreliert
- Cov wegen 1 definiert

*Beweis.*

Schätzen Varianz von  $S_n$  ab:

- Aufspaltung der Summe in Terme mit  $i = j$  und  $i \neq j$ :
- Bedingung über Unkorreliertheit benutzen

$$\text{Var}(S_n) = \frac{1}{n^2} \mathbb{E} \left( \left( \sum_{i=0}^{n-1} (X_i - m) \right)^2 \right) \quad (10)$$

$$= \frac{1}{n^2} \sum_{i=0}^{n-1} \text{Var}(X_i) + 2 \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

$$\leq \frac{v}{n} \quad (11)$$

Lemma 3.5:

$$P(\{|S_n - m| \geq \epsilon\}) \leq \frac{v}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

das war zu zeigen

□

zeigen hinreichende Bedingung für schwaches Gesetz der großen Zahl für Folge von ZV ohne Annahme an Varianzen

$(X_i)_{i \in \mathbb{N}}$  - Folge von Zufallsvariablen

**Satz 3.20** ( $L^1$ -Version des schwachen Gesetzes der großen Zahl). *Annahmen:*

1.  $(X_i)_{i \in \mathbb{N}}$  ist paarweise unabhängig.
2.  $(X_i)_{i \in \mathbb{N}}$  ist identisch verteilt.
3.  $X_0$  ist integrierbar.

Dann erfüllt  $(X_i)_{i \in \mathbb{N}}$  das schwache Gesetz der großen Zahl.

*Beweis.*

setzen  $X_i^{\flat} := X_i \chi_{\{|X_i| \leq i^{1/4}\}}$

$X_i^{\sharp} := X_i - X_i^{\flat}$

$Y_n^* := \frac{1}{n} \sum_{i=0}^{n-1} (X_i^* - \mathbb{E}(X_i^*))$  mit  $*$  in  $\{\flat, \sharp\}$

zeigen:

-  $Y_n^{\flat} \xrightarrow{n \rightarrow \infty} 0$  (in  $P$ )

-  $Y_n^\sharp \xrightarrow{n \rightarrow \infty} 0$  (in  $P$ )

- daraus folgt Behauptung

$(X_i^b)_i$  ist paarweise unabhängig

-  $\text{Var}(X_i^b) \leq \mathbb{E}(X_i^{b,2}) \leq i^{1/2}$

-  $\text{Var}(Y_n^b) = \frac{1}{n^2} \sum_{i=0}^{n-1} \text{Var}(X_i^b) \leq \frac{1}{n^2} \sum_{i=0}^{n-1} i^{1/2} \leq \frac{1}{n^{1/2}}$

- Chebyshev:  $Y_n^b \xrightarrow{n \rightarrow \infty} 0$  (in Wahrscheinlichkeit)

monotone Konvergenz  $\mathbb{E}(|X_i^\sharp|) = \mathbb{E}(|X_1|) - \mathbb{E}(|X_1| \chi_{\{|X_1| \leq i^{1/4}\}}) \xrightarrow{i \rightarrow \infty} 0$

schließen daraus  $\mathbb{E}(|Y_n^\sharp|) \leq \frac{2}{n} \sum_{i=0}^{n-1} \mathbb{E}(X_i^\sharp) \xrightarrow{n \rightarrow \infty} 0$

□

zeigen nun hinreichendes Kriterium für starkes Gesetz der großen Zahl

- verschärfen Annahmen aus Lemma 3.19

**Satz 3.21.** *Annahme:*

1. Die Folge ist paarweise unabhängig und identisch verteilt.

2.  $X_0^2$  ist integrierbar.

Dann erfüllt  $(X_i)_{i \in \mathbb{N}}$  das starke Gesetz der großen Zahl.

*Beweis.*

wegen  $\mathbb{E}(X_0^2) < \infty$  ist  $X_0$  auch integrierbar

Reduktion auf nicht-negative Zufallsvariablen

Annahme: haben Satz unter Zusatzannahme:  $X_i \geq 0$  gezeigt:

- betrachten Folgen  $(X_i^\pm)_{i \in \mathbb{N}}$

- erfüllen auch die Voraussetzungen

- Unabhängigkeit bleibt erhalten, nicht aber Unkorreliertheit

- $S_n^\pm := \frac{1}{n} \sum_{i=0}^{n-1} X_i^\pm$
- $m^\pm := \mathbb{E}(X_0^\pm)$
- dann  $S_n^\pm \xrightarrow{n \rightarrow \infty} m^\pm$  fast sicher nach Annahme
- schießen:  $S_n = S_n^+ - S_n^- \xrightarrow{n \rightarrow \infty} m^+ - m^- = m$  fast sicher

ab hier Annahme:  $X_i \geq 0$  für alle  $i$  in  $\mathbb{N}$

$$m := E(X_1)$$

$$S_n := \frac{1}{n} \sum_{i=0}^{n-1} X_i$$

müssen zeigen:  $S_n \xrightarrow{n \rightarrow \infty} m$  (fast sicher)

$$v := \text{Var}(X_0)$$

- wegen paarweiser Unabhängigkeit aus (10):  $\text{Var}(S_n) = \frac{v}{n}$

wähle  $\epsilon > 0$

- setze  $k_n := \lfloor (1 + \epsilon)^{n+1} \rfloor$  (Gaußklammer, ganzer Teil)

- es gilt:  $k_n \geq \frac{1}{2}(1 + \epsilon)^{n+1}$  für alle  $n$

-  $k_n^{-1} \leq 2(1 + \epsilon)^{-(n+1)}$

-  $m := E(X_1)$

$$\begin{aligned}
 & \sum_{i=0}^{\infty} P(|S_{k_n} - m| \geq (1 + \epsilon)^{-(n+1)/4}) \\
 & \stackrel{\text{Chebyshev}}{\leq} \sum_{i=0}^{\infty} (1 + \epsilon)^{(n+1)/2} \text{Var}(S_{k_n}) \\
 & = \sum_{i=0}^{\infty} (1 + \epsilon)^{(n+1)/2} k_n^{-1} v \\
 & \leq 2v \sum_{i=0}^{\infty} (1 + \epsilon)^{-(n+1)/2} \\
 & < \infty
 \end{aligned}$$

nach Borel-Cantelli 2.69:

- es gilt  $S_{k_n} \xrightarrow{n \rightarrow \infty} m$  (fast sicher)
- das ist bisher nur eine Teilfolge von  $(S_n)_{n \in \mathbb{N}}$

$$\frac{k_{n+1}}{k_n} \leq \frac{(1+\epsilon)^{n+2} + 1}{(1+\epsilon)^{n+1} - 1} \xrightarrow{n \rightarrow \infty} 1 + \epsilon$$

- also für große  $n$ :  $k_{n+1} \leq (1 + 2\epsilon)k_n$

für  $l \in \{k_n, \dots, k_{n+1}\}$  gilt:

$$\frac{S_{k_n}}{1 + 2\epsilon} \leq k_n \frac{S_{k_n}}{k_{n+1}} \leq S_l \leq k_{n+1} \frac{S_{k_{n+1}}}{k_n} \leq (1 + 2\epsilon)S_{k_{n+1}}$$

- haben hier  $X_i \geq 0$  benutzt!

es gilt:  $(1 + 2\epsilon) - 1 = 2\epsilon \leq 2\epsilon(1 + 2\epsilon)$

also:  $1 - (1 + 2\epsilon)^{-1} \leq 2\epsilon$ ,

-  $1 - 2\epsilon \leq \frac{1}{1+2\epsilon}$

- ziehen  $m$  ab:

$$S_{k_n} - m - 2\epsilon S_{k_n} \leq S_l - m \leq S_{k_{n+1}} - m + 2\epsilon S_{k_{n+1}}$$

- zwei Fälle:

-  $S_l - m \geq 0$  - obere Abschätzung verwenden

$$- |S_l - m| \leq |S_{k_{n+1}} - m| + 2\epsilon S_{k_{n+1}}$$

-  $S_l - m < 0$  - untere Abschätzung verwenden

$$- |S_l - m| \leq |S_{k_n} - m| + 2\epsilon S_{k_n}$$

- zusammen:  $|S_l - m| \leq \max\{|S_{k_n} - m|, |S_{k_{n+1}} - m|\} + 2\epsilon \max\{S_{k_n}, S_{k_{n+1}}\}$

- die Ungleichung ! in folgender Kette gilt fast sicher

$$\begin{aligned} \limsup_{l \rightarrow \infty} |S_l - m| &\leq \limsup_{n \rightarrow \infty} |S_{k_n} - m| + 2\epsilon \limsup_{n \rightarrow \infty} |S_{k_n}| \\ &\stackrel{!}{\leq} 2\epsilon m \end{aligned}$$

- Setzen  $\epsilon_i := 1/(i + 1)$  für  $i \in \mathbb{N}$
- bilden Durchschnitt der fast sicheren Ereignisse:

$$\left\{ \limsup_{l \rightarrow \infty} |S_l - m| \leq \frac{2m}{i + 1} \right\}$$

schließen:

$\limsup_{l \rightarrow \infty} |S_l - m| = 0$  fast sicher □

### Beispiel 3.22. Monte-Carlo Simulation

$f : [0, 1] \rightarrow \mathbb{R}$  - meßbar

- wollen  $\int_0^1 f d\text{Leb}$  bestimmen

-  $(X_i)_{i \in I}$  auf  $[0, 1]$  gleichverteilte Zufallsvariablen, unabhängig

-  $\mathbb{E}(f(X_0)) = \int_0^1 f d\text{Leb}$

-  $S_n := \frac{1}{n} \sum_{i=0}^{n-1} f(X_i)$

starkes Gesetz der großen Zahl:

Es gilt:  $\lim_{n \rightarrow \infty} S_n = \int_0^1 f d\text{Leb}$  fast sicher

Fehlerdiskussion:

$v := \text{Var}(f(X)) = \int_0^1 f^2 d\text{Leb} - \left(\int_0^1 f d\text{Leb}\right)^2$

-  $\text{Var}(S_n) = \frac{v}{n}$

Chebyshev:

$P(|S_n - \mathbb{E}(f(X_0))| \geq \frac{\epsilon}{\sqrt{n}}) \leq \frac{v}{\epsilon^2}$

Genauigkeit bei  $n$  Stützstellen:  $\sim n^{-1/2}$

Simulation von  $\int_0^1 \sin(\pi t)^2 dt = 1/2$

Zum Vergleich naives Riemannintegral

Genauigkeit bei  $n$  Stützstellen:  $\sim n^{-2}$

Simulation vom Riemannintegral

□

### 3.4 Verteilungsfunktionen

$(\Omega, \mathcal{B}, P)$  - W-Raum

$X$  - numerische Zufallsvariable

**Definition 3.23.** Die Funktion

$$F : \mathbb{R} \rightarrow [0, 1], \quad F(t) = P(\{X \leq t\})$$

heißt Verteilungsfunktion der Zufallsvariable  $X$ .

nicht verwechseln mit der Verteilung von  $X$ : letztere ist das Maß  $X_*P$

- die Verteilung von  $X$  und die Verteilungsfunktion bestimmen sich jedoch wechselseitig eindeutig, siehe unten

**Lemma 3.24.**

1.  $F$  is monoton wachsend.
2.  $\lim_{t \rightarrow -\infty} F(t) = 0$
3.  $\lim_{t \rightarrow \infty} F(t) = 1$
4.  $F$  ist rechtsseitig stetig: Für alle  $t_0$  in  $\mathbb{R}$  gilt  $\lim_{t \downarrow t_0} F(t) = F(t_0)$ .
5.  $X_*P$  ist durch  $F$  eindeutig bestimmt

*Beweis.*

Zu 1.: Für  $t \leq t'$  gilt  $\{X \leq t\} \subseteq \{X \leq t'\}$ , also  $F(t) \leq F(t')$

für die restlichen Eigenschaften verwendet man die  $\sigma$ -Additivität von  $P$ :



Zu 2.:  $\lim_{t \rightarrow -\infty} F(t) = P(\bigcap_{t \in \mathbb{Z}} \{X \leq t\}) = P(\emptyset) = 0$

Zu 3.:  $\lim_{t \rightarrow \infty} F(t) = P(\bigcup_{t \in \mathbb{Z}} \{X \leq t\}) = P(\Omega) = 1$

Zu 4.:  $\lim_{t \downarrow t_0} F(t) = \lim_{n \rightarrow \infty} F(t_0 + \frac{1}{n}) = P(\bigcap_{n \in \mathbb{N} \setminus \{0\}} \{X \leq t_0 + \frac{1}{n}\}) = P(\{X \leq t_0\}) = F(t_0)$

- wegen der Monotonie von  $F$  von reicht es den Grenzwert von  $(F(t_0 + \frac{1}{n}))_{n \in \mathbb{N}_{>0}}$  zu betrachten

Zu 5.:

-  $X_*P$  bestimmt  $F$

-  $F(t) = (X_*P)(-\infty, t]$

-  $F$  bestimmt  $X_*P$

-  $X_*P$  ist durch die Werte auf den Teilmengen  $(s, t]$  eindeutig bestimmt.

- dazu Eindeutigkeitsaussage des Caratheódory Satzes 2.55 benutzen

- es gilt  $(X_*P)((s, t]) = F(t) - F(s)$

□

**Lemma 3.25.** *Sei  $F : \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion mit den Eigenschaften 1. ... 4.. Dann ist  $F$  die Verteilungsfunktion einer Zufallsvariablen.*

*Beweis.*

$X := \text{id}_{\mathbb{R}}$

- definieren Maß auf  $(\mathbb{R}, \text{Borel})$

- betrachten die Algebra  $\mathcal{R}$ , die von den Intervallen der Form  $(s, t]$  erzeugt wird

-  $t = \infty$  und  $s = -\infty$  zugelassen

- jedes Element von  $\mathcal{R}$  ist eine endliche Vereinigung von solchen Intervallen

- sei  $A = \bigcup_{i \in I} (s_i, t_i]$  ( $I$  endlich)

- setzen  $P(A) := \sum_{i \in I} F(t_i) - F(s_i)$

-  $P$  ist normiert

- offensichtlich additiv:

- beachte:  $(a, b] \cup (b, c] = (a, c]$  und  $(F(b) - F(a)) + (F(c) - F(b)) = F(c) - F(a)$

-  $P$  ist  $\sigma$ -additiv: benutzen rechtsseitige Stetigkeit

Idee:

-  $((s_i, t_i])_{i \in I}$  absteigend mit  $\bigcap_{i \in I} (s_i, t_i] = (s, t]$

- dann gilt:

—  $s_i = s$  für große  $i$  (sonst wäre  $s$  im Durchschnitt)

—  $t_i \downarrow t$

- also  $\lim_{i \rightarrow \infty} P(s_i, t_i] = \lim_{i \rightarrow \infty} F(t_i) - F(s) = F(t) - F(s)$  □

- wegen rechtsseitiger Stetigkeit bei  $t$

**Beispiel 3.26.** Verteilungsfunktion von  $\delta_x$

$$F(t) = \begin{cases} 0 & t < x \\ 1 & t \geq x \end{cases} \quad \square$$

**Beispiel 3.27.** Verteilungsfunktion der Gleichverteilung auf  $[0, 1]$

$$F(t) = \begin{cases} 0 & t < 0 \\ t & t \in [0, 1] \\ 1 & t > 1 \end{cases} \quad \square$$

### ***R* kennt Verteilungsfunktionen**

Exponentialverteilung:

Normalverteilung

$(X_i)_{i \in \mathbb{N}}$  - Folge von numerischen Zufallsvariablen

$X$  - eine Zufallsvariable

-  $F_i$  - Verteilungsfunktion von  $X_i$

-  $F$  - Verteilungsfunktion von  $X$

**Definition 3.28.**  $(X_i)_{i \in \mathbb{N}}$  konvergiert in Verteilung, falls jede Stetigkeitsstelle  $t$  von  $F$  gilt  $\lim_{n \rightarrow \infty} F_n(t) = F(t)$ .

**Lemma 3.29.** *Konvergenz in  $P \Rightarrow$  Konvergenz in Verteilung*

*Beweis.*

schreiben:  $F(t) = P(\{X \leq t\})$

Annahme:  $X_n \xrightarrow{n \rightarrow \infty} X$  in  $P$

wegen  $\{X_n \leq t\} \subseteq \{X \leq t + \epsilon\} \cup \{|X_n - X| \geq \epsilon\}$

-  $P(\{X_n \leq t\}) \leq P(\{X \leq t + \epsilon\}) + P(\{|X_n - X| \geq \epsilon\})$

analog

-  $P(\{X \leq t - \epsilon\}) \leq P(\{X_n \leq t\}) + P(\{|X_n - X| \geq \epsilon\})$

wegen Annahme and  $(X_n)_n$

-  $\limsup_{n \rightarrow \infty} F_n(t) \leq F(t + \epsilon)$

-  $\liminf_{n \rightarrow \infty} F_n(t) \geq F(t - \epsilon)$

sei nun  $t$  - Stetigkeitspunkt von  $F$

- bilden Grenzwert  $\epsilon \downarrow 0$

-  $\limsup_{n \rightarrow \infty} F_n(t) \leq F(t)$

-  $\liminf_{n \rightarrow \infty} F_n(t) \geq F(t)$

- also  $\lim_{n \rightarrow \infty} F_n(t) = F(t)$

□

**Beispiel 3.30.** *Konvergenz in Verteilung  $\not\Rightarrow$  Konvergenz in  $P$*

$(X_n)$  - konstante Folge von Zufallsvariablen

mit Verteilung  $P = \frac{1}{2}(\delta_{-1} + \delta_1)$

-  $Y = -X$

- Verteilung von  $Y$  ist Verteilung von  $X_n$

- also  $X_n \xrightarrow{n \rightarrow \infty} X$  in Verteilung

-  $P(\{|X_n - Y| \geq 1\}) = 1$  für alle  $n$

- also nicht  $X_n \xrightarrow{n \rightarrow \infty} Y$  in  $P$

□

Kurzschreibweise:  $X_i \xrightarrow{i \rightarrow \infty} X$  (i.V. )

zur Erinnerung: schwache Konvergenz Definition 3.10

geben Kriterium für schwache Konvergenz in Termen der Verteilungsfunktionen

betrachten  $(X_n)_{n \in \mathbb{N}}$  - Folge von Zufallsvariablen

$X$  - eine Zufallsvariable

-  $F_n$  - Verteilungsfunktion von  $X_n$

-  $F$  - Verteilungsfunktion von  $X$

**Lemma 3.31.** *Die folgenden Aussagen sind äquivalent.*

1. *Es gilt  $X_n \xrightarrow{n \rightarrow \infty} X$  schwach.*
2. *Es gilt  $X_n \xrightarrow{i \rightarrow \infty} X$  in Verteilung.*
3. *Es gilt  $\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$  für alle  $f \in C^2(\mathbb{R})$  mit*
  - (a)  *$f$  ist beschränkt*
  - (b)  *$f'$  ist beschränkt*
  - (c)  *$f''$  ist beschränkt und gleichmäßig stetig*

*Beweis.*

2.  $\Rightarrow$  1.

$f \in C_b(\mathbb{R})$  gegeben:

- zu zeigen:  $\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$

-  $F$  ist monoton

- Unstetigkeitsstelle von  $F$  ist Sprung

-  $F$  hat Werte in  $[0, 1]$

-  $F$  hat höchstens abzählbar viele Unstetigkeitsstellen

$\epsilon > 0$  gegeben

- wähle Stetigkeitsstellen von  $F$ :  $a < b$  mit  $F(a) < \epsilon$  und  $F(b) > 1 - \epsilon$
- wähle Stetigkeitsstellen von  $F$ :  $a = c_1 < c_2 < \dots < c_k = b$  so daß  $|f(x) - f(c_i)| \leq \epsilon$  für alle  $c_{i-1} \leq x \leq c_i$
- benutzen hier daß  $f|_{[a,b]}$  gleichmäßig stetig ist

$$\begin{aligned}\mathbb{E}(f(X_n)) &= \sum_{i=2}^k \mathbb{E}(f(X_n)\chi_{(c_{i-1}, c_i]}(X_n)) + \mathbb{E}(f(X_n)\chi_{(-\infty, a] \cup (b, \infty)}(X_n)) \\ &\leq \sum_{i=2}^k (f(c_i) + \epsilon)(F_n(c_i) - F_n(c_{i-1})) + \|f\|_\infty(F_n(a) + (1 - F_n(b)))\end{aligned}$$

analog

$$\sum_{i=2}^k f(c_i + \epsilon)(F(c_i) - F(c_{i-1})) \leq \mathbb{E}(f(X)\chi_{(a,b]}(X)) + 2\epsilon$$

also

$$\begin{aligned}\limsup_{n \rightarrow \infty} \mathbb{E}(f(X_n)) &\leq \mathbb{E}(f(X)\chi_{(a,b]}(X)) + 2\epsilon + \|f\|_\infty(F(a) + (1 - F(b))) \\ &\leq \mathbb{E}(f(X)) + 2\epsilon + 2\epsilon\|f\|_\infty\end{aligned}$$

analog  $f$  durch  $-f$  ersetzen

$$\liminf_{n \rightarrow \infty} \mathbb{E}(f(X_n)) \geq \mathbb{E}(f(X)) - 2\epsilon - 2\epsilon\|f\|_\infty$$

also

$$|\limsup_{n \rightarrow \infty} \mathbb{E}(f(X_n)) - \liminf_{n \rightarrow \infty} \mathbb{E}(f(X_n))| \leq 4\epsilon(1 + \|f\|_\infty)$$

-  $\epsilon$  beliebig impliziert Konvergenz

-  $\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$

1.  $\Rightarrow$  2.

sei  $c$  - Stetigkeitsstelle von  $F$

- wähle  $f : \mathbb{R} \rightarrow [0, 1]$  stetig mit  $\chi_{(-\infty, c]} \leq f \leq \chi_{(-\infty, c+\delta)}$
- für später kann zusätzlich annehmen:  $f', f''$  existieren, sind stetig und haben kompakten Träger

$$\begin{aligned} \limsup_{n \rightarrow \infty} F_n(c) &\leq \limsup_{n \rightarrow \infty} \mathbb{E}(f(X_n)) \\ &= \mathbb{E}(f(X)) \\ &\leq F(c + \delta) \end{aligned}$$

- $\delta$  - beliebig
- $F$  rechtsseitig stetig

$$\limsup_{n \rightarrow \infty} F_n(c) \leq F(c)$$

- wähle  $h : \mathbb{R} \rightarrow [0, 1]$  stetig mit  $\chi_{(-\infty, c-\delta]} \leq h \leq \chi_{(-\infty, c]}$
- für später kann zusätzlich annehmen:  $h', h''$  existieren, sind stetig und haben kompakten Träger

$$\begin{aligned} \liminf_{n \rightarrow \infty} F_n(c) &\geq \liminf_{n \rightarrow \infty} \mathbb{E}(h(X_n)) \\ &= \mathbb{E}(h(X)) \\ &\geq F(c - \delta) \end{aligned}$$

- $\delta$  - beliebig
- $F$  ist in  $c$  auch linksseitig stetig (da  $c$  Stetigkeitsstelle)

$$\liminf_{n \rightarrow \infty} F_n(c) \geq F(c)$$

Zusammen: Konvergenz und  $\lim_{n \rightarrow \infty} F_n(c) = F(c)$

1.  $\Rightarrow$  3. klar

3.  $\Rightarrow$  2.

- wie in 1.  $\Rightarrow$  2.

- benutze Zusatzbemerkung über Ableitungen der Testfunktionen  $f$  und  $h$

□

### 3.5 $W$ -Maße auf Gruppen und reellen Vektorräumen

$(\Omega, \mathcal{B})$  - meßbarer Raum

- bezeichnen die Menge der  $W$ -Maße mit  $P(\Omega, \mathcal{B})$ , oder kurz  $P(\Omega)$

sei  $((M, \mathcal{B}), 1, +)$  Monoid in **Meas**

-  $Q, P$  -  $W$ -Maße auf  $(M, \mathcal{B})$

- bilden  $P \times Q$  auf  $(M, \mathcal{B}) \times (M, \mathcal{B})$

-  $+$  :  $M \times M \rightarrow M$  meßbar

**Definition 3.32.** Die Faltung von  $P$  und  $Q$  ist durch  $P * Q := +_*(P \times Q)$  definiert.

**Lemma 3.33.**  $(P(M), *, \delta_e)$  ist ein Monoid.

*Beweis.*

müssen Assoziativitäts und Identitätsrelationen zeigen

- die folgen aus den Relationen für das Monoid  $((M, \mathcal{B}), 1, +)$

Assoziativität:

$$\begin{aligned}
(P * Q) * R &= +_*((P * Q) \times R) \\
&= +_*(+ \times \text{id}_M)_*(P \times Q \times R) \\
&= (+ \circ (+ \times \text{id}_M))_*(P \times Q \times R) \\
&= (+ \circ (\text{id}_M \times +))_*(P \times Q \times R) \\
&= +_*(\text{id}_M \times +)_*(P \times Q \times R) \\
&= +_*(P \times (Q * R)) \\
&= P * (Q * R)
\end{aligned}$$

Einheit:

- interpretiere Einheit von  $M$  als Abbildung  $e : * \rightarrow M$  mit  $e(*) = e$

- implizit benutze Identifikation  $\{*\} \times M \cong M$

-  $+ \circ (e \times \text{id}_M) = + \circ (\text{id}_M \times e) = \text{id}_M$

-  $\delta_e = e_* \delta_*$

$$\delta_e * P = +_*(\delta_e \times P) = +_*(e_* \delta_* \times P) = (+ \circ (e \times \text{id}_M))_*(\delta_* \times P) = \text{id}_* P = P$$

analog  $P * \delta_e = P$

□

**Beispiel 3.34.** Falten von Maßen mit Dichten

$P = f\text{Leb}$  und  $Q = g\text{Leb}$  auf  $\mathbb{R}$

Behauptung: es gilt  $P * Q = h\text{Leb}$  mit  $h(y) := \int_{\mathbb{R}} f(x)g(y-x)d\text{Leb}(x)$

Beweis:



Benutzen Translationsinvarianz von Leb im vorletzten und Fubini bei letzten Schritt

$$\begin{aligned}
 (P * Q)(A) &= +_*(P \times Q)(A) \\
 &= +_*(\text{pr}_1^* f \text{pr}_2^* g \text{Leb} \times \text{Leb})(A) \\
 &= \int_{\mathbb{R} \times \mathbb{R}} +^* \chi_A(x, z) f(x) g(z) d(\text{Leb} \times \text{Leb})(x, z) \\
 &= \int_{\mathbb{R} \times \mathbb{R}} \chi_A(x + z) f(x) g(z) d(\text{Leb} \times \text{Leb})(x, z) \\
 &= \int_{\mathbb{R} \times \mathbb{R}} \chi_A(y) f(x) g(y - x) d(\text{Leb} \times \text{Leb})(x, y) \\
 &= \int_{\mathbb{R}} \chi_A(y) \int_{\mathbb{R}} f(x) g(y - x) d\text{Leb}(x) d\text{Leb}(y)
 \end{aligned}$$

lesen daraus Formel für  $h$  ab

□

**Beispiel 3.35.** Mittelwert Bilden ist Falten und Skalieren

$(X_i)_{i=1, \dots, n}$  unabhängige Familie von Zufallsvariablen mit Verteilungen  $(P_i)_{i=1, \dots, n}$

- Verteilung von  $\sum_{i=1}^n X_i$  ist  $P_1 * \dots * P_n$

- hier ist Unabhängigkeit wichtig

- Verteilung von  $S_n := \frac{1}{n} \sum_{i=1}^n X_i$  ist  $(M_{\frac{1}{n}})_*(P_1 * \dots * P_n)$

-  $M_\lambda : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \lambda x$

**Mittelwerte der Gleichverteilung**

Konvergenz gegen  $\delta_{1/2}$  wegen Gesetz der Großen Zahl

- aber warum sieht das aus wie eine Gaußkurve?

- Erklärung durch den zentralen Grenzwertsatz

□

$V$  - reeller Vektorraum (endlich-dimensional)

- bilden  $S^n(V^*)$  - Vektorraum der symmetrischen  $n$ -Formen auf  $V$

$f : V \rightarrow \mathbb{R}$  - genügend oft diffbar

$df : V \rightarrow V^*$

betrachten höhere Ableitungen als Abbildungen

-  $d^2 f : V \rightarrow S^2(V^*)$  (wegen Symmetrie die partiellen Ableitung)

...

-  $d^n f : V \rightarrow S^n(V^*)$

-

$$P_n(V) := \{p \in C^\infty(\mathbb{R}) \mid (\forall \lambda \in \mathbb{R} \forall v \in V \mid f(\lambda v) = \lambda^n p(v))\}$$

ist Vektorraum der homogenen Polynome auf  $V$  vom Grad  $n$

- beachte die Glattheitsvoraussetzung:  $\|v\|^{2n+1}$  ist nicht in  $P_n(V)$

**Lemma 3.36.** Die Abbildung  $P^n(V) \rightarrow S^n(V^*)$  gegeben durch  $p \in P^n(V) \mapsto \frac{1}{n!} d^n p(0)$  ist ein linearer Isomorphismus mit dem Inversen  $s \in S^n(V^*) \mapsto (v \mapsto s(v, \dots, v))$ .

*Beweis.* Nachrechnen! □

Homogene Polynome auf  $V^*$  sind Elemente in  $S^n((V^*)^*) \cong S^n(V)$

betrachten W-Maß  $P$  auf einen endlich-dimensionalen reellem Vektorraum  $V$

**Definition 3.37.** Das  $n$ 'te-Moment von  $P$  ist die symmetrische  $n$ -Form  $\mu_n \in S^n(V)$  gegeben durch

$$\mu_n(l_1, \dots, l_n) := \mathbb{E}(l_1 \dots l_n)$$

für alle Familien  $(l_i)_{i=1}^n$  in  $V^*$  (falls der Erwartungswert existiert)

-  $\mu_0 = 1$  (Normierung)

-  $\mu_1(l) = \mathbb{E}(l)$

-  $\mu_2(l_1, l_2) = \mathbb{E}(l_1 l_2) = \text{Cov}(l_1, l_2) + \mathbb{E}(l_1)\mathbb{E}(l_2)$

- ...

wegen Lemma 3.36 fassen  $\mu_n$  oft auch als homogenes Polynom auf  $V^*$  auf:  $\mu_n(l) := \mathbb{E}(l^n)$

$G$  - topologische Gruppe

betrachten Menge  $\hat{G}$  der stetigen Homomorphismen  $\chi : G \rightarrow U(1)$

-  $\hat{G}$  hat wieder Gruppenstruktur  $(\chi \cdot \chi')(g) := \chi(g)\chi'(g)$

-  $\chi^{-1}(g) = \bar{\chi}(g)$

$G$  lokal-kompakt

-  $\hat{G}$  mit kompakt-offener Topologie ist wieder eine topologische Gruppe

- diese ist wieder lokal-kompakt

**Definition 3.38.** Für eine lokal-kompakte Gruppe  $G$  heißt die topologische Gruppe  $\hat{G}$  die duale Gruppe.

**Beispiel 3.39.** Beispiele für lokalkompakte abelsche topologische Gruppen und deren duale Gruppen

$\hat{\mathbb{Z}} \cong S^1$  -  $S^1 \ni z \mapsto (\chi_z : n \mapsto z^n)$

$\hat{S}^1 \cong \mathbb{Z}$  -  $\mathbb{Z} \ni n \mapsto (\chi_n : u \mapsto u^n)$

$\hat{\mathbb{R}} \cong \mathbb{R}$  -  $\mathbb{R} \ni \xi \mapsto (\chi_\xi : x \mapsto e^{-ix\xi})$

$\hat{V} \cong V^*$  -  $V^* \ni l \mapsto (\chi_l : v \mapsto e^{-il(v)})$  für endlich-dimensionalen reellen Vektorraum

$\hat{\mathbb{Z}/n\mathbb{Z}} \cong \mathbb{Z}/n\mathbb{Z}$  -  $\mathbb{Z}/n\mathbb{Z} \ni [k] \mapsto (\chi_{[k]} : [l] \mapsto e^{2\pi i \frac{kl}{n}})$

$\widehat{G_0 \times G_1} = \hat{G}_0 \times \hat{G}_1$

haben immer kanonische Abbildung  $G \rightarrow \hat{\hat{G}}, g \mapsto \hat{\chi}_g : (\hat{G} \ni \chi \mapsto \chi(g))$

- in den obigen Beispielen ist der folgende Satz offensichtlich:

**Satz 3.40** (Pontrjagin Dualität). Für eine lokalkompakte abelsche Gruppe ist die kanonische Abbildung  $G \rightarrow \hat{\hat{G}}$  ein Isomorphismus.

*Beweis.*

in den Beispielen durch Inspektion

allgemeiner Fall: Harmonische Analysis □

□

$P$  -  $W$ -Maß auf  $G$

**Definition 3.41.** Die Funktion

$$\hat{P} : \hat{G} \rightarrow \mathbb{C}, \quad \chi \mapsto \mathbb{E}(\chi)$$

heißt die charakteristische Funktion von  $P$

- benutzen das Symbol  $\hat{P}$  weil es wirklich die Fouriertransformation des Maßes  $P$  bezeichnet

offensichtliche Eigenschaften:

-  $\|\hat{P}\| \leq 1$  (da  $\|\chi\|_\infty = 1$ )

-  $\hat{P}(1) = 1$  (Normierung von  $P$ )

-  $\hat{P}(\chi^{-1}) = \overline{\hat{P}(\chi)}$  ( $P$  ist "reell")

-  $\hat{P}$  ist folgenstetig:

-  $(\chi_n)_{n \in \mathbb{N}}, \lim_{n \rightarrow \infty} \chi_n = \chi$  (in der kompakt-offenen Topologie)

— impliziert punktweise Konvergenz

— majorisierte Konvergenz mit Majorante 1

—  $\lim_{n \rightarrow \infty} \hat{P}(\chi_n) = \lim_{n \rightarrow \infty} \mathbb{E}(\chi_n) = \mathbb{E}(\chi) = \hat{P}(\chi)$

**Lemma 3.42.** Die Abbildung  $P \mapsto \hat{P}$  ist injektiv.

*Beweis.* Harmonische Analysis

□

Frage: Was ist das Bild? Antwort in der harmonischen Analysis!

**Definition 3.43.**  $f : \hat{G} \rightarrow \mathbb{C}$  heißt positiv semidefinit, falls für jede endliche Familie  $(\chi_i)_{i \in I}$  in  $\hat{G}$  die  $I \times I$ -Matrix  $(f(\chi_i^{-1} \chi_j))_{i, j \in I}$  positiv semidefinit ist.

**Lemma 3.44.**  $\hat{P}$  für ein  $W$ -Maß  $P$  ist positiv semidefinit.

*Beweis.*

sei  $(x_1, \dots, x_j)$  in  $\mathbb{C}^I$

$$\begin{aligned} \sum_{i,j \in I} \bar{x}_i x_j \hat{P}(\chi_i^{-1} \chi_j) &= \sum_{i,j \in I} \bar{x}_i x_j \mathbb{E}(\chi_i^{-1} \chi_j) \\ &= \mathbb{E} \left( \sum_{i,j \in I} \bar{x}_i x_j \bar{\chi}_i \chi_j \right) \\ &= \mathbb{E} \left( \left| \sum_{i \in I} x_i \chi_i \right|^2 \right) \\ &\geq 0 \end{aligned}$$

□

**Satz 3.45** (Bochner).  $f : \hat{G} \rightarrow \mathbb{C}$  ist charakteristische Funktion eines  $W$ -Maßes genau dann wenn  $f$  folgende Eigenschaften hat:

1.  $f$  ist stetig.
2.  $f(0) = 1$
3.  $f$  ist positiv semidefinit.

*Beweis.* Harmonische Analysis

□

Charakteristische Funktion und Translation

-  $g \in G$  induziert Translation um  $g$ :  $T_g : G \rightarrow G$ ,  $T_g(h) := gh$

$G$ - lokalkompakte abelsche topologische Gruppe

-  $g$  in  $G$

-  $P$ -  $W$ -Maß auf  $G$

**Lemma 3.46.**  $\widehat{T_{g,*}P}(\chi) = \chi(g)\hat{P}(\chi)$ .

*Beweis.*

haben Identität  $(T_g^* \chi)(-) = \chi(g-) = \chi(g)\chi(-)$  von Funktionen  $G \rightarrow \mathbb{C}$

- benutzen das bei !

- benutzen Transformationsformel bei !!

$$\begin{aligned} \widehat{T_{g,*}P}(\chi) &\stackrel{def}{=} \mathbb{E}_{T_{g,*}P}(\chi) \\ &\stackrel{!!}{=} \mathbb{E}_P(T_g^* \chi) \\ &\stackrel{!}{=} \mathbb{E}_P(\chi(g)\chi) \\ &= \chi(g)\mathbb{E}_P(\chi) \\ &\stackrel{def}{=} \chi(g)\hat{P}(\chi) \end{aligned}$$

□

Skalieren und charakteristische Funktion

$V$  -reeller Vektorraum,

$\lambda \in \mathbb{R}^*$

-  $M_\lambda : V \rightarrow V, \quad M_\lambda(v) := \lambda v$

$P$  - W-Maß auf  $V$

**Lemma 3.47.**  $\widehat{M_{\lambda,*}P} = M_\lambda^* \hat{P}$

*Beweis.*

benutzen

$$(M_\lambda^* \chi_l)(x) = e^{-il(\lambda x)} = e^{-i\lambda l(x)} = \chi_{\lambda l}(x)$$

$$\begin{aligned} \widehat{M_{\lambda,*}P}(l) &= \mathbb{E}_{M_{\lambda,*}P}(\chi_l) \\ &= \mathbb{E}_P(M_\lambda^* \chi_l) \\ &= \mathbb{E}_P(\chi_{\lambda l}) \\ &= (M_\lambda^* \hat{P})(l) \end{aligned}$$

□

Charakteristische Funktion und Produkte

-  $G, H$  topologische abelsche Gruppen

-  $P, Q$  - W-Maße auf  $G$  und  $H$

**Lemma 3.48.** *Es gilt  $\widehat{P \times Q} = \text{pr}_1^* \hat{P} \text{pr}_2^* \hat{Q}$*

*Beweis.*

$$\begin{aligned} \widehat{P \times Q}(\chi, \mu) &= \mathbb{E}_{P \times Q}(\text{pr}_G^* \chi \text{pr}_H^* \mu) \\ &= \mathbb{E}_P(\chi) \mathbb{E}_Q(\mu) \\ &= \hat{P}(\chi) \hat{Q}(\mu) \\ &= (\text{pr}_1^* \hat{P} \text{pr}_2^* \hat{Q})(\chi, \mu) \end{aligned}$$

□

Charakteristische Funktion und Homomorphismen

$A : G \rightarrow H$  Homomorphismus topologischer abelscher Gruppen

-  $\hat{A} : \hat{H} \rightarrow \hat{G}$  dualer Homomorphismus

-  $\hat{A}(\chi) := A^* \chi$

$P$  - W-Maß auf  $G$

**Lemma 3.49.** *Es gilt  $\widehat{A_* P} = \hat{A}^* \hat{P}$ .*

*Beweis.*  $\chi \in \hat{H}$

$$\widehat{A_* P}(\chi) = \mathbb{E}_{A_* P}(\chi) = \mathbb{E}_P(A^* \chi) = (\hat{A}^* P)(\chi)$$

□

**Beispiel 3.50.** Charakteristische Funktion der Gleichverteilung auf  $[a, b]$

- (als  $W$ -Maß auf  $\mathbb{R}$ )

$$\hat{P}(\xi) = \frac{1}{b-a} \int_a^b e^{-ix\xi} dx = \frac{e^{-ib\xi} - e^{ia\xi}}{i\xi(b-a)}$$

Ersetzen  $a, b$  durch  $a+c, b+c$

– Entspricht Translation der Gleichverteilung durch  $c$

– aus Formel ablesen:  $\widehat{T_{c,*}P}(\xi) = e^{-ic\xi} \hat{P}(\xi)$

— kompatibel mit Lemma 3.46

Ersetzen  $a, b$  durch  $a\lambda, b\lambda$

– Entspricht Skalierung der Gleichverteilung durch  $\lambda$

– aus Formel ablesen:  $\widehat{M_{\lambda,*}P}(\xi) = \hat{P}(\lambda\xi)$

— kompatibel mit Lemma 3.47

□

Charakteristische Funktion und Faltung

**Lemma 3.51.** *Es gilt  $\widehat{P*Q} = \hat{P}\hat{Q}$ .*

*Beweis.*  $+: G \times G \rightarrow G$

$\chi \in \hat{G}$

die Identität  $\chi(g+h) = \chi(g)\chi(h)$  ist äquivalent zu  $+^*\chi = \text{pr}_1\chi\text{pr}_2\chi$

$$\begin{aligned} \widehat{P*Q}(\chi) &= \mathbb{E}_{+^*(P \times Q)}(\chi) \\ &= \mathbb{E}_{P \times Q}(+^*\chi) \\ &= \mathbb{E}_{P \times Q}(\text{pr}_1\chi\text{pr}_2\chi) \\ &= \mathbb{E}_P(\chi)\mathbb{E}_Q(\chi) \\ &= \hat{P}(\chi)\hat{Q}(\chi) \end{aligned}$$

Fubini im vorletzten Schritt

□

**Beispiel 3.52.** Charakteristische Funktion und Momente

sei  $P$  - W-Maß auf  $V$



**Lemma 3.53.** Wenn  $\mu_n$  existiert, dann ist  $\hat{P}$   $n$ -mal stetig diffbar und es gilt für  $k = 0, \dots, n$

$$\mu_k = ((id)^k \hat{P})(0) .$$

*Beweis.*

-  $\hat{P}(l) = \mathbb{E}(e^{-il(-)})$

- leiten nach  $l$  ab

-  $id\hat{P}(l)(h) = \mathbb{E}(he^{-il})$

-  $id\hat{P}(0)(h) = \mathbb{E}(h)$

-  $\mu_1 = id\hat{P}(0)$

-  $(id)^2\hat{P}(l)(h_1, h_2) = \mathbb{E}(h_1h_2e^{-il})$

-  $(id)^2\hat{P}(0)(h_1, h_2) = \mathbb{E}(h_1h_2)$

-  $(id)^2\hat{P}(0) = \mu_2$

usw

□

□

**Beispiel 3.54.** Charakteristische Funktion der Exponentialverteilung

- (als  $W$ -Maß auf  $\mathbb{R}$ )

$$\hat{P}(\xi) = \alpha \int_0^\infty e^{-ix\xi} e^{-\alpha x} dx = \frac{\alpha}{\alpha + i\xi}$$

-  $\mu_n(P) = (id)^n \hat{P}(0) = \frac{n!}{\alpha^n}$

- paßt

□

**Beispiel 3.55.** Charakteristische Funktion der Normalverteilung

Rechnung:

$$\widehat{\text{Norm}(m, \sigma)}(\xi) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-ix\xi} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = e^{-mi\xi} e^{-\frac{\sigma^2\xi^2}{2}}$$

effektive Rechnung in kleinen Schritten:

- $\text{Norm}(0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{Leb}_{\mathbb{R}}$
- $\text{Norm}(m, \sigma) = T_{m,*} M_{\sigma,*} \text{Norm}(0, 1)$
- benutzen:  $M_{\sigma,*} \text{Leb}_{\mathbb{R}} = \frac{1}{\sigma} \text{Leb}_{\mathbb{R}}$

$$\begin{aligned}
 T_{m,*} M_{\sigma,*} \text{Norm}(0, 1) &= T_{m,*} M_{\sigma,*} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{Leb}_{\mathbb{R}} \right) \\
 &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} T_{m,*} M_{\sigma,*} \text{Leb}_{\mathbb{R}} \\
 &= \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-m)^2}{2\sigma^2}} T_{m,*} \text{Leb}_{\mathbb{R}} \\
 &= \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-m)^2}{2\sigma^2}} \text{Leb}_{\mathbb{R}}
 \end{aligned}$$

müssen also nur  $\widehat{\text{Norm}(0, 1)} = e^{-\frac{\xi^2}{2}}$  zeigen

$$\begin{aligned}
 \widehat{\text{Norm}(0, 1)}(\xi) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ix\xi} e^{-\frac{x^2}{2}} dx \\
 &= \frac{e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-i\xi)^2}{2}} dx \\
 &= \frac{e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \\
 &= e^{-\frac{\xi^2}{2}}
 \end{aligned}$$

in vorletzter Gleichung Cauchy Integralsatz angewendet

durch Berechnung der Ableitungen:

$$\mu_0 = 1$$

$$\mu_1 = m$$

$$\mu_2 = \sigma^2 + m^2$$

- $\mu_2 - \mu_1^2 = \text{Var}(x) = \sigma^2$
- paßt

□

### 3.6 Gaussche Maße

in diesem Kapitel alle Vektorräume endlich-dimensional über  $\mathbb{R}$

$V$  - Vektorraum

- es gibt ein bis auf Normierung eindeutiges Maß auf  $V$  welches translationsinvariant ist
- generische Notation  $\text{Leb}_V$

allgemein:  $b$  Bilinearform

- es gibt eindeutig bestimmte Dualitätsabbildung  $A_b : V \rightarrow V^*$  mit  $A_b(v)(w) = b(v, w) = A_b^*(w)(v)$
- wenn  $b$  nicht entartet (z.B. symmetrisch positiv definit) ist, dann ist  $A_b$  invertierbar
- $A_b^{-1} : V^* \rightarrow V \cong V^{**}$  bestimmt duale Bilinearform

$$b^*(\phi, \psi) := b(A_b^{-1}(\phi), A_b^{-1}(\psi)) = A_b(A_b^{-1}(\phi))(A_b^{-1}(\psi)) = A_b^{-1,*}(\phi)(\psi)$$

- also  $A_{b^*} = A_b^{-1,*}$

$f : V \rightarrow \mathbb{R}$  Funktion

**Definition 3.56.**  $f$  heißt quadratisch, falls  $f = q + l + c$  für eine homogen quadratische Funktion  $q$ , eine lineare Funktion  $l$  und eine Konstante  $c$  gilt.

$f$  bestimmt  $q, l, c$  eindeutig:

- $c = f(0)$
- $l = df(0)$
- $q = \frac{1}{2}d^2f(0)$
- schreiben  $b_f$  und  $A_f$  für  $b_q$  und  $A_{b_q}$

$A : V \rightarrow W$  Abbildung

**Definition 3.57.**  $A$  heißt affin wenn  $A = L + m$  für eine lineare Abbildung  $L : V \rightarrow W$  und Konstante  $m$  in  $W$ .

$A$  bestimmt  $L$  und  $m$  eindeutig:

- $m = A(0)$
- $L = dA(0)$

Wenn  $f : W \rightarrow \mathbb{R}$  quadratisch und  $A : V \rightarrow W$  affin ist, dann ist  $A^*q$  quadratisch.

berechnen:

$$f = q + l + c$$

$$A = L + m$$

$$\begin{aligned}(A^*f)(v) &= b_f(L(v) + m, L(v) + m) + l(L(v) + m) + c \\ &= b_f(L(v), L(v)) + 2b_f(m, L(v)) + l(L(v)) + b_f(m, m) + l(m) + c \\ &= [(L^*q) + (2L^*A_f(m) + L^*l) + (q(m) + l(m) + c)](v)\end{aligned}$$

$A^*f = L^*q + L^*(2A_f(m) + l) + (q(m) + l(m) + c)$  ist quadratisch

sei  $f$  quadratisch

- $n := \dim(V)$

**Lemma 3.58.** Wenn  $b_f$  positiv definit ist, dann gibt einen affinen Isomorphismus  $A : \mathbb{R}^n \rightarrow V$  derart, daß  $(A^*f)(x) = \lambda + \frac{1}{2}\|x\|^2$  gilt.

*Beweis.*  $f = q + l + c$

- suchen zunächst  $m$  so daß
- $f(x + m) = q(x) + \lambda$
- wegen  $f(x + m) = q(x) + 2A_f(m) + l(x) + q(m) + l(m) + c$
- $2A_f(m) = -l$

— setzen  $m := -\frac{1}{2}A_f^{-1}(l)$

- finde dann invertierbare Abbildung  $L : \mathbb{R}^n \rightarrow V$  derart daß  $(L^*q)(x) = \frac{1}{2}\|x\|^2$ .

setze  $A = T_m \circ L = L + m$

□

**Definition 3.59.** Eine Gaussche Dichte ist eine Funktion  $V \rightarrow \mathbb{R}$  der Form  $g(v) := e^{-f(v)}$  wobei  $f$  quadratisch mit positiv definitem  $b_f$  ist.

sei  $A : V \rightarrow W$  affin und injektiv,  $f$  eine Gaussche Dichte

**Lemma 3.60.**  $A^*f$  ist eine Gaussche Dichte

*Beweis.*

$A = L + m$

$-b_{A^*f} = b_f(L \otimes L)$  ist positiv definit

□

$g$  - Gaussche Dichte

**Lemma 3.61.** Das Maß  $g\text{Leb}_V$  hat alle Momente: Für jede endliche Folge  $(l_i)_{i=1,\dots,n}$  in  $V^*$  existiert  $\int_V \prod_{i=1}^n l_i g d\text{Leb}_V$ ,

*Beweis.* Lemma 3.58: nach geeigneter affiner Koordinatenwahl:

-  $V = \mathbb{R}^N$ ,  $\text{Leb}_V = \text{Leb}$ ,  $(A^*f)(x) = \lambda + \frac{1}{2}\|x\|^2$

-  $(|A^* \prod_{i=1}^n l_i|)(x) \leq C(1 + |x|)^n$

-  $(1 + \|x\|)^n e^{-\|x\|^2/2}$  bez.  $\text{Leb}$  integrierbar

□

sei  $g$  Gaussche Dichte

- man kann also  $\text{Leb}_V$  eindeutig normieren daß  $g\text{Leb}_V$  ein  $W$ -Maß ist

$P$  -  $W$ -Maß auf  $V$

**Definition 3.62.**  $P$  ist ein Gaussches Maß, falls es eine Dichte bezüglich  $\text{Leb}_V$  hat, welche eine Gaussche Dichte ist.

berechnen Momente für Gaussche Maße

$V$ - reeller Vektorraum

-  $f = q + l + c$  -quadratische Funktion, positiv definit

$P := e^{-f} d\text{Leb}_V$  Gaussches Maß

**Lemma 3.63.** Es gelten

$$\mu_1 = -\frac{1}{2} A_f^{-1,*} l$$

und

$$\mu_2 = \frac{1}{4} A_f^{-1,*} l^2 + \frac{1}{2} q^* = \mu_1^2 + \frac{1}{2} q^* .$$

Insbesondere ist  $P$  durch  $\mu_1$  und  $\mu_2$  eindeutig bestimmt.

*Beweis.*

für  $\xi$  in  $V$ ,  $t$  in  $\mathbb{R}$

-  $A_t(v) := v + t\xi$  - Translation

- Translationsinvarianz von  $\text{Leb}_V$

$$\int_V A_t^* g d\text{Leb}_V = \int_V g d(A_{t,*}\text{Leb}_V) = \int_V g d\text{Leb}_V$$

$$(\partial_t)_{|t=0} \int_V A_t^* g d\text{Leb}_V = 0$$

$$A_t^* f = f + t(2A_f(\xi) + l(\xi)) + t^2 q(\xi)$$

$$\partial_t(A_t^* e^{-f}) = -[2A_f(\xi) + l(\xi) + 2tq(\xi)] A_t^* e^{-f}$$

$$(\partial_t)_{|t=0}(A_t^* e^{-f}) = -(2A_f(\xi) + l(\xi)) e^{-f}$$

$$\begin{aligned}\partial_t^2(A_t^* e^{-f}) &= [[2A_f(\xi) + l(\xi) + 2tq(\xi)]^2 A_t^* e^{-f} - 2q(\xi)] A_t^* e^{-f} \\ (\partial_t^2)_{t=0}(A_t^* e^{-f}) &= [(2A_f(\xi) + l(\xi))^2 - 2q(\xi)] e^{-f}\end{aligned}$$

$$2\mathbb{E}(A_f(\xi)) + l(\xi) = 0$$

also

$$\mathbb{E}(A_f(\xi)) = -\frac{1}{2}l(\xi)$$

$$(2A_f(\xi) + l(\xi))^2 = 4A_f(\xi)^2 + 4A_f(\xi)l(\xi) + l(\xi)^2$$

$$0 = 4\mathbb{E}(A_f(\xi)^2) + 4l(\xi)\mathbb{E}(A_f(\xi)) + l(\xi)^2 - 2q(\xi) = 4\mathbb{E}(A_f(\xi)^2) - 2l(\xi)^2 + l(\xi)^2 - 2q(\xi) = 4\mathbb{E}(A_f(\xi)^2) - l(\xi)^2 - 2q(\xi)$$

also

$$\mathbb{E}(A_f(\xi)^2) = \frac{1}{2}q(\xi) + \frac{1}{4}l(\xi)^2$$

lesen ab:

$$A_f^* \mu_1 = -\frac{1}{2}l$$

$$A_f^* \mu_2 = \frac{1}{2}q + \frac{1}{4}l^2$$

-  $\mu_1$  und  $\mu_2$  bestimmen also Gaussches Maß eindeutig

$$\mu_1 = -\frac{1}{2}A_f^{-1,*}l$$

benutzen  $A_f^{-1,*}q = q^*$

$$\mu_2 = \frac{1}{4}A_f^{-1,*}l^2 + \frac{1}{2}q^* = \mu_1^2 + \frac{1}{2}q^*$$

□

Charakteristische Funktion

$P = e^{-f} \text{Leb}_V$  - Gaussches Maß

finde affine Abbildung  $A : V \rightarrow \mathbb{R}^n$  mit

$$f = A^{-1,*} \left( \frac{\|x\|^2}{2} + c \right)$$

-  $c$  bestimmt sich aus Normierung

$$- \int_V e^{-f} d\text{Leb}_V = \int_V A^* e^{-\left(\frac{\|x\|^2}{2} + c\right)} d\text{Leb}_V = s \int_V e^{-\left(\frac{\|x\|^2}{2} + c\right)} d\text{Leb}_{\mathbb{R}^n}(x) = s e^{-c} \sqrt{2\pi}^n$$

- mit  $s \text{Leb}_{\mathbb{R}^n}(x) = A_* \text{Leb}_V$

- also  $c = \ln(s \sqrt{2\pi}^n)$

Lemma 3.48

$$\widehat{A_* P} = \prod_{i=1}^n \widehat{\text{Norm}(0, 1)} = \prod_{i=1}^n e^{-\frac{\xi_i^2}{2}} = e^{-\frac{\|\xi\|^2}{2}}$$

Schreiben  $A = L + m = T_m \circ L$

$$- \widehat{A_* P}(\chi) = T_{m,*} \widehat{(L_* P)}(\chi) = \chi(m) \widehat{L_* P}$$

$$- \chi(m)^{-1} \widehat{A_* P}(\chi) = \widehat{L_* P}$$

Lemma 3.5

$$\widehat{P}(\chi) = \widehat{L}^{-1}(\chi)(m)^{-1} \widehat{A_* P}(\widehat{L}^{-1}(\chi))$$

- mit  $\chi(-) = e^{-il(-)}$  (Identifikation von  $\widehat{V}$  mit  $V^*$ )

$$\widehat{P}(l) = e^{il(L^{-1}(m))} e^{-\frac{\|L^*, -1(l)\|^2}{2}}$$

- beachte:  $L^* : \mathbb{R}^n \rightarrow V^*$  (mit  $\mathbb{R}^n = \mathbb{R}^{n,*}$ )

**Lemma 3.64.** *Ein Produkt von Gausschen Maßen ist ein Gaussches Maß.*

*Beweis.*  $P$  - Gaussches Maß auf  $V$



$Q$  - Gaussches Maß auf  $W$

$P \times Q$  ist Gaussches Maß auf  $V \times W$

$$P = e^{-f} \text{Leb}_V, Q = e^{-g} \text{Leb}_W$$

-  $\text{Leb}_{V \times W} = \text{Leb}_V \times \text{Leb}_W$  (bei geeigneter Normierung)

nach Fubini:

$$- P \times Q = e^{-\text{pr}_V^* f - \text{pr}_W^* g} \text{Leb}_{V \times W}$$

-  $\text{pr}_V^* f + \text{pr}_W^* g : V \oplus W \rightarrow \mathbb{R}$  ist quadratisch und positiv definit □

betrachten  $A : V \rightarrow W$  affin

$P$  Gaussches Maß auf  $V$

**Lemma 3.65.** *Wenn  $A$  surjektiv ist, dann ist  $A_* P$  ein Gaussches Maß auf  $W$ .*

*Beweis.*

1. Fall  $A$  ist Translation  $A(v) = v + m$

$$- P = e^{-f} \text{Leb}_V$$

$$- A_* P = e^{-A^{-1,*} f} P$$

-  $A^{-1,*} f$  ist quadratisch und positiv definit

2. Fall  $A$  is linear:

$$\text{sei } P = e^{-f} \text{Leb}_V$$

-  $b_f$  is positiv definite Bilinearform

- definiert Orthogonalität

$$- V = \ker(A) \oplus \ker(A)^\perp$$

- haben kanonischen Isomorphismus  $\iota : W \rightarrow \ker(A)^\perp$  mit  $A \circ \iota = \text{id}_W$

setze  $f_0 := f|_{\ker(A)}$  und  $f_1 := f|_{\ker(A)^\perp}$

$$g := \iota^* f_1$$

$$\text{Leb}_V = \text{Leb}_{\ker(A)} \times \iota_* \text{Leb}_W$$

$$- f = \text{pr}_{\ker(A)}^* f_0 + \text{pr}_{\ker(A)^\perp} f_1 - c$$

Fubini:

$$P = e^{-f} \text{Leb}_V = e^{-\text{pr}_{\ker(A)}^* f_0} \times e^{-\text{pr}_{\ker(A)^\perp} f_1} \iota_* \text{Leb}_W = e^{-\text{pr}_{\ker(A)}^* f_0} \times \iota_* e^{-g} \text{Leb}_W$$

$$A_* P = \int_{\ker(A)} e^{-\text{pr}_{\ker(A)}^* f_0} e^{-g} \text{Leb}_W$$

ist Gaussches Maß auf  $A$

□

**Korollar 3.66.** Die Faltung von Gausschen Maßen auf  $V$  ist wieder ein Gaussches Maß.

*Beweis.*  $P, Q$  - Gaussche Maße

-  $P \times Q$  - Gaussches Maß auf  $V \times V$

-  $+$  :  $V \times V \rightarrow V$  is affin und surjektiv

-  $+_*(P \times Q)$  ist Gaussches Maß nach Lemma 3.65

□

**Beispiel 3.67.** Berechnen  $\text{Norm}(m, \sigma) * \text{Norm}(m', \sigma')$

- berechnen charakteristische Funktion

$$-\widehat{\text{Norm}(m, \sigma) * \text{Norm}(m', \sigma')} = \widehat{\text{Norm}(m, \sigma)} \widehat{\text{Norm}(m', \sigma')}$$

$$\begin{aligned} \widehat{\text{Norm}(m, \sigma)} \widehat{\text{Norm}(m', \sigma')} &= e^{-mi\xi} e^{-\frac{\sigma^2 \xi^2}{2}} e^{-m'i\xi} e^{-\frac{\sigma'^2 \xi^2}{2}} \\ &= e^{-(m+m')\xi} e^{-\frac{(\sigma^2 + \sigma'^2) \xi^2}{2}} \\ &= \widehat{\text{Norm}(m + m', \sqrt{\sigma^2 + \sigma'^2})} \end{aligned}$$

also

$$\text{Norm}(m, \sigma) * \text{Norm}(m', \sigma') = \text{Norm}(m + m', \sqrt{\sigma^2 + \sigma'^2})$$

$X$  -  $\text{Norm}(m, \sigma)$ -verteilt

$X'$  -  $\text{Norm}(m', \sigma')$ -verteilt

- dann  $X + Y$  -  $\text{Norm}(m + m', \sqrt{\sigma^2 + \sigma'^2})$ -verteilt

beobachten:

$$M_{1/\sqrt{2},*}(\text{Norm}(0, \sigma) * \text{Norm}(0, \sigma)) = M_{1/2,*}\text{Norm}(0, \sqrt{2}\sigma) = \text{Norm}(0, \sigma)$$

analog

$$M_{1/\sqrt{n},*}(\underbrace{\text{Norm}(0, \sigma) * \dots * \text{Norm}(0, \sigma)}_{n \text{ mal}}) = \text{Norm}(0, \sigma)$$

also falls  $(X_i)_{i=1}^n$  unabhängig Norm(0,  $\sigma$ )-Verteilt

- dann  $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$  ist Norm(0,  $\sigma$ )-Verteilt

□

### 3.7 Zentraler Grenzwertsatz

**Bemerkung 3.68.** Beobachtung

$(X_i)_{i \in \mathbb{N}}$  Folge unabhängiger Zufallsvariablen mit Verteilung  $P$  mit Mittelwert 0

- charakteristische Funktion  $\hat{P}$

-  $T_n := \frac{1}{\sqrt{n}} \sum_{i=0}^n X_i$  hat Verteilung  $Q_n$  mit charakteristischer Funktion  $\hat{Q}_n = \hat{P}(\frac{\xi}{\sqrt{n}})^n$

- allgemein mit  $m = \mathbb{E}(X_0)$  und  $\mathbb{E}(X^2) = \sigma^2$

-  $\hat{P}(\xi) = 1 + m\xi + \frac{1}{2}\sigma^2\xi^2 + o(\xi^2)$

- wegen  $m = 0$

-  $\hat{Q}_n(\xi) = \hat{P}(\frac{\xi}{\sqrt{n}})^n = (1 + \frac{1}{2n}\sigma^2\xi^2 + o(\xi^2/n))^n$

- Grenzwert für  $n \rightarrow \infty$

-  $\lim_{n \rightarrow \infty} \hat{Q}_n(\xi) = e^{-\frac{\sigma^2\xi^2}{2}}$

— Grenzwert ist charakteristische Funktion von Norm(0,  $\sigma$ )

— Konvergenz würde aus  $\lim_{n \rightarrow \infty} T_n \rightarrow T$  (in Verteilung, schwach) mit  $T$  mit Verteilung Norm(0,  $\sigma$ ) folgen

Ziel dieses Kapitel: Beweis dieser Tatsache

Normierung von Zufallsvariablen

$X$  - Zufallsvariable

$\mathbb{E}(X) = m, \text{Var}(X) = \sigma^2$

**Definition 3.69.**  $X^* := \frac{1}{\sigma}(X - m)$  heißt die Normierung von  $X$

$$\mathbb{E}(X^*) = 0$$

$$\text{Var}(X^*) = 1$$

$(X_i)_{i \in \mathbb{N}}$  Folge von Zufallsvariablen

$$S_n^* := \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} X_i^*$$

Annahme: identisch verteilt, unabhängig

$$\mathbb{E}(S_n) = 0$$

$$\text{Var}(S_n^*) = \frac{1}{n} \sum_{i,j=0}^{n-1} \mathbb{E}(X_i^* X_j^*) = 1$$

Normalverteilung  $\mu = 0, \sigma = 1$

$X_i$  gleichverteilt in  $[0, 1]$

-  $\mathbb{E}(X_i) = 1/2$

-  $\text{Var}(X_i) = 1/12$

-  $X_i^* = \sqrt{12}(X_i - 1/2)$

- Simulation: 1000 mal Mittelwerte über 1000 Experimente

$X_i$  binomialverteilt,  $p = 0.4, n = 20$

$(X_i)_{i \in \mathbb{N}}$  Folge von Zufallsvariablen

$Y$  - Zufallsvariable, Norm(0, 1)-verteilt

**Satz 3.70.** Annahme:

1.  $(X_i)_{i \in \mathbb{N}}$  ist identisch verteilt.
2.  $(X_i)_{i \in \mathbb{N}}$  ist unabhängig.
3.  $\mathbb{E}(X_0)$  existiert.
4.  $\text{Var}(X_0)$  existiert und ist positiv.

Dann gilt  $\lim_{n \rightarrow \infty} S_n^* \rightarrow Y$  (in Verteilung/schwach).

*Beweis.*

wenn wir  $X_i$  durch  $X_i^*$  ersetzen, dann ändern wir  $S_n^*$  nicht

- nehmen oBdA an:  $\mathbb{E}(X_0) = 0, \mathbb{E}(X_0^2) = 1$

müssen zeigen:  $\lim_{n \rightarrow \infty} \mathbb{E}(f(S_n^*)) = \mathbb{E}(f(Y))$  für alle  $f \in C_b(\mathbb{R})$

- nach Lemma 3.31 können zusätzlich annehmen:  $f \in C^2, f'$  beschränkt,  $f''$  beschränkt und gleichmäßig stetig

- fixiere  $f$  mit diesen Eigenschaften

Taylor for  $f$  bei  $a$

-  $f(a+x) = f(a) + f'(a)x + \frac{1}{2}f''(a)x^2 + \frac{1}{2}x^2[f''(a+\theta x) - f''(a)]$

- mit  $\theta \in [0, 1]$  (Funktion von  $x$ )

-  $R(a, x) := \frac{1}{2}x^2[f''(a+\theta x) - f''(a)]$  - Rest

-  $|R(a, x)| \leq x^2 \|f''\|_\infty$  für alle  $(a, x)$

- sei  $\epsilon > 0$  gegeben

- wähle  $\delta > 0$  so daß  $|R(a, x)| \leq x^2 \epsilon$  für  $|x| \leq \delta$  and alle  $a$

- benutzen hier die gleichmäßige Stetigkeit von  $f''$

- dann

$$|R(a, x)| \leq x^2 [\epsilon \chi_{[-\delta, \delta]}(x) + \|f''\| \chi_{\mathbb{R} \setminus [-\delta, \delta]}(x)] \quad (12)$$

für alle  $(a, x)$

wähle Folge von Zufallsvariablen  $(Y_i)_{i \in \mathbb{N}}$

- unabhängig und unabhängig von  $(X_i)_{i \in \mathbb{N}}$

-  $N(0, 1)$ -verteilt

- setze  $T_n^* := \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} Y_i$

- dann  $T_n^*$  ist auch Norm(0, 1)-verteilt

genügt zu zeigen:  $\lim_{n \rightarrow \infty} \mathbb{E}(f(S_n^*)) - \mathbb{E}(f(T_n^*)) = 0$

setze  $X_{i,n} := \frac{X_i}{\sqrt{n}}$ ,  $Y_{i,n} := \frac{Y_i}{\sqrt{n}}$

setze  $W_{i,n} := \sum_{j=0}^{i-1} Y_{j,n} + \sum_{j=i+1}^{n-1} X_{j,n}$

- es gilt  $W_{i,n} + X_{i,n} = W_{i-1,n} + Y_{i-1,n}$

-  $f(S_n^*) - f(T_n^*) = \sum_{i=0}^{n-1} f(W_{i,n} + X_{i,n}) - f(W_{i,n} + Y_{i,n})$

-  $f(W_{i,n})$  und  $X_{i,n}$  sind unabhängig

-  $f(W_{i,n})$  und  $Y_{i,n}$  sind unabhängig

- benutzen Taylor Entwicklung von  $f$  tmit  $a := W_{i,n}$

$$\begin{aligned} f(S_n^*) - f(T_n^*) &= \sum_{i=0}^{n-1} f'(W_{i,n})(X_{i,n} - Y_{i,n}) + \sum_{i=0}^{n-1} \frac{1}{2} f''(W_{i,n})(X_{i,n}^2 - Y_{i,n}^2) \\ &+ \sum_{i=0}^{n-1} [R(W_{i,n}, X_{i,n}) - R(W_{i,n}, Y_{i,n})] \end{aligned}$$

wenden  $\mathbb{E}$  an:

$$\begin{aligned} |\mathbb{E}(f(S_n^*)) - \mathbb{E}(f(T_n^*))| &\leq \sum_{i=0}^{n-1} [\mathbb{E}(|R(W_{i,n}, X_{i,n})|) + \mathbb{E}(|R(W_{i,n}, Y_{i,n})|)] \\ &\stackrel{(12)}{\leq} \epsilon \sum_{i=0}^{n-1} [\mathbb{E}(X_{i,n}^2 \chi_{[-\delta, \delta]}(X_{i,n})) + \mathbb{E}(Y_{i,n}^2 \chi_{[-\delta, \delta]}(Y_{i,n}))] \\ &\quad + \|f''\| \sum_{i=0}^{n-1} [\mathbb{E}(X_{i,n}^2 \chi_{\mathbb{R} \setminus [-\delta, \delta]}(X_{i,n})) + \mathbb{E}(Y_{i,n}^2 \chi_{\mathbb{R} \setminus [-\delta, \delta]}(Y_{i,n}))] \\ &\leq 2\epsilon + 2\|f''\|_{\infty} \mathbb{E}(X_0 \chi_{\mathbb{R} \setminus [-\delta n, \delta n]}(X_0)) \\ &\xrightarrow{n \rightarrow \infty} 2\epsilon \end{aligned}$$

benutzen im ersten Schritt:

-  $\mathbb{E}(f'(W_{i,n})X_{i,n}) = \mathbb{E}(f'(W_{i,n}))\mathbb{E}(X_{i,n}) = 0$

-  $\mathbb{E}(f''(W_{i,n})X_{i,n}^2) = \mathbb{E}(f''(W_{i,n}))\mathbb{E}(X_{i,n}^2) = \mathbb{E}(f''(W_{i,n}))\mathbb{E}(Y_{i,n}^2) = \mathbb{E}(f''(W_{i,n})Y_{i,n}^2)$

benutzen in letzten Schritt  $\mathbb{E}(X_0^2) = 1$ ,

- also  $\mathbb{E}(X_{i,n}^2) = \frac{1}{n}$

- also  $\epsilon \sum_{i=0}^{n-1} \mathbb{E}(X_{i,n}^2 \chi_{(-\delta, \delta)}(X_{i,n})) \leq \epsilon$

- für ersten Summanden

$$\begin{aligned} \sum_{i=0}^{n-1} \mathbb{E}(X_{i,n}^2 \chi_{\mathbb{R} \setminus [-\delta, \delta]}(X_{i,n})) &= n \mathbb{E}(X_{0,n}^2 \chi_{\mathbb{R} \setminus [-\delta, \delta]}(X_{0,n})) \\ &= \mathbb{E}(X_0^2 \chi_{\mathbb{R} \setminus [-\delta n, \delta n]}(X_0)) \end{aligned}$$

- für zweiten n Summanden

—  $\lim_{n \rightarrow \infty} \mathbb{E}(X_0^2 \chi_{\mathbb{R} \setminus [-\delta n, \delta n]}(X_0)) = 0$

— wenden hier Satz über monotone Konvergenz

$\epsilon$  beliebig wählbar

- erhalten wie gewünscht  $|\mathbb{E}(f(S_n^*)) - \mathbb{E}(f(T_n^*))| \xrightarrow{n \rightarrow \infty} 0$

□

## 4 Schätztheorie

### 4.1 Beschreibung des Problems und der grundlegenden Begriffe

**Beispiel 4.1.** Reissnagel

werfen Reissnagel

- beobachten, ob er auf der Spitze oder dem Rücken liegen bleibt

Mit welcher Wahrscheinlichkeit  $p$  bleibt er auf dem Rücken?

Vermutung:  $p \approx \frac{\text{Anzahl der Rückenlagen}}{\text{Anzahl der Würfe}}$

- Wie gut ist die Schätzung?
- Wird sie besser, wenn man öfter wirft?
- Kann man systematisch zu der Vermutung kommen? □

**Beispiel 4.2.** Taxiproblem:

Taxis in einer Stadt sind fortlaufend nummeriert

- beobachte Taxis an einer belebten Straße und notiere Nummern

Wieviele zugelassene Taxis gibt es?

Schätzung: maximale beobachtete Nummer

- Ist diese Schätzung gut?
- Geht es besser?
- Sei max. notierte Nummer 1000. Kann man angeben, wie viele Taxis man beobachten muß, um mit Wahrscheinlichkeit 0.95 zu sagen, daß die gesuchte Zahl kleiner als 1010 ist? □

**Beispiel 4.3.** Karpfenteich

Möchte die Anzahl  $N$  der Karpfen im Teich schätzen

- setze  $n$  Goldfische ein
- die vertragen sich gut mit Karpfen und mischen sich drunter (Modellannahme)
- fange  $F$  Fische wieder raus
- davon sind  $f$  Goldfische

Vermutung:  $\frac{F}{N+n} \approx \frac{f}{n}$

- also  $\frac{N}{F} \approx \frac{n}{f} - \frac{n}{F}$

- also  $N \approx n\left(\frac{F}{f} - 1\right)$

- Wie gut ist diese Schätzung?



- Sei  $A$  gegeben. Wie groß muß ich  $F$  und  $n$  wählen damit  $N$  in  $[n(\frac{F}{f} - 1) - A, n(\frac{F}{f} - 1) + A]$  mit Wahrscheinlichkeit 0.95 liegt?

- hier ist  $f$  die Zufallsvariable

□

allgemeine Situation:

$(\Omega, \mathcal{B})$  meßbarer Raum: Beschreibt die elementaren Ereignisse in Hintergrund

- wir kennen aber noch nicht das  $W$ -Maß

- betrachten deshalb  $P : \Theta \rightarrow P(\Omega)$  - Abbildung

— Familie der infragekommenden  $W$ -Maße auf  $(\Omega, \mathcal{B})$

— schreiben  $P$  auch als Familie  $(P_\vartheta)_{\vartheta \in \Theta}$

**Definition 4.4.**  $(\Omega, \mathcal{B}, (P_\vartheta)_{\vartheta \in \Theta})$  heißt statistisches Modell.

**Beispiel 4.5.** Einen  $W$ -Raum  $(\Omega, \mathcal{B}, P)$  kann man als statistisches Modell betrachten

- mit  $\Theta = *$

□

$(\Omega', \mathcal{B}')$  meßbarer Raum: beschreibt beobachtbare Ergebnisse des Experiments

$X : \Omega \rightarrow \Omega'$  meßbare Abbildung

- beschreibt das Experiment

**Beispiel 4.6.**  $(\Omega', \mathcal{B}') = (\Omega, \mathcal{B})$  und  $X = \text{id}_\Omega$

□

**Definition 4.7.**  $X : (\Omega, \mathcal{B}) \rightarrow (\Omega', \mathcal{B}')$  heißt Statistik.

wollen Informationen über den Parameter  $\vartheta$  des vorliegende Maß  $P_\vartheta$  erhalten aus dem Ergebnis des Experimentes

-  $g : \Theta \rightarrow \Theta'$  - beschreibt uns interessierende Information

**Beispiel 4.8.**  $\Theta' = \Theta$  und  $g = \text{id}_\Theta$

□

wählen dazu  $\hat{g} : \Omega' \rightarrow \Theta'$

**Definition 4.9.** Die Funktion  $\hat{g}$  heißt (Punkt)Schätzer für  $g$ .

- bei Beobachtung von  $x$  in  $\Omega'$  ist  $\hat{g}(x)$  die Schätzung von  $g(\vartheta)$

**Bemerkung 4.10.** mögliche Interpretation:

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \Omega' \\ \uparrow P & & \downarrow \hat{g} \\ \Theta & \xrightarrow{g} & \Theta' \end{array}$$

verstehen  $P$  als verallgemeinerte “gewichtete mehrwertige” Abbildung nach  $\Omega$

- echte Abbildung  $\phi : \Theta \rightarrow \Omega$  ist Spezialfall:  $P_\vartheta = \delta_{\phi(\vartheta)}$

- suchen  $\hat{g}$  so daß Quadrat “möglichst gut” kommutiert □

**Beispiel 4.11.**

bisher keine Relation zwischen  $g$  und  $\hat{g}$  festgelegt

- man könnte einen Wert  $\vartheta_0$  in  $\Theta$  auswählen und  $\hat{g} := \text{const}_{g(\vartheta_0)}$  setzen

- unabhängig vom Ausgang des Experimentes schätzt man dann immer daß  $g(\vartheta_0)$  vorliegt

- das ist natürlich nicht sinnvoll, da man die Beobachtung nicht nutzt □

ein Punktschätzer liefert einen Schätzwert  $\hat{g}(x)$  in  $\Theta'$  für  $g(\vartheta)$  bei Beobachtung von  $x$

- ein Konfidenzbereich liefert ein Teilmenge  $C(x)$  von  $\Theta'$  in welcher wir  $g(\vartheta)$  vermuten bei Beobachtung von  $x$

**Definition 4.12.** Ein Konfidenzbereich ist eine Teilmenge  $C \subseteq \Omega' \times \Theta'$ .

- interpretieren  $C$  als Familie von Teilmengen  $(C(x))_{x \in \Omega'}$  von  $\Theta'$

- die Übersetzung wird durch folgende Gleichungen bestimmt:

- $C$  gegeben
- dann  $C(x) := \text{pr}_{\Omega'}^{-1}(\{x\}) \cap C$

- $(C(x))_{x \in \Omega'}$  gegeben
- dann  $C := \bigcup_{x \in \Omega'} \{x\} \times C(x)$

bei Beobachtung von  $x$  wird geschätzt, daß  $g(\vartheta) \in C(x)$  liegt

**Beispiel 4.13.** Konfidenzbereiche verallgemeinern Punktschätzer

$\hat{g}$  - Punktschätzer

- betrachten  $C := \text{Graph von } \hat{g} = \{(x, \hat{g}(x)) \mid x \in \Omega'\}$
- also  $C(x) = \{g(x)\}$

□

betrachten Maße für die Güte von Schätzern und Konfidenzbereichen

Gegeben:

- $g : \Theta \rightarrow \Theta'$
- $1 \geq \alpha > 0$  vorgegeben:  $1 - \alpha$  heißt Konfidenzniveau
- $C$  - Konfidenzbereich

**Definition 4.14.**  $C$  ist ein Konfidenzbereich zum Konfidenzniveau  $1 - \alpha$  falls

$$\inf_{\vartheta \in \Theta} P_{\vartheta}(\{g(\vartheta) \in C(X)\}) \geq 1 - \alpha$$

*gilt.*

- es kann passieren, daß die Teilmenge  $\{g(\vartheta) \in C(X)\}$  nicht in  $\mathcal{B}$  liegt
- dann interpretieren wir  $P_{\vartheta}(\{g(\vartheta) \in C(X)\})$  als  $\sup_{A \in \mathcal{B}, A \subseteq \{g(\vartheta) \in C(X)\}} P_{\vartheta}(A)$

Interpretation:

- Ich bin mir bei Beobachtung von  $x = X(\omega)$  sicher, daß  $g(\vartheta) \in C(x)$  mit Wahrscheinlichkeit  $1 - \alpha$  gilt.
- Hier ist  $X$  und damit das Intervall  $C(X)$  die zufällige Variable und  $\vartheta$  fest, aber unbekannt.
- wenn  $\alpha$  sehr klein ist, dann bin ich mir bei der Beobachtung von  $x$  sehr sicher, daß  $g(\vartheta) \in C(x)$  gilt

**Beispiel 4.15.**

$C = \Omega' \times \Theta'$  ist ein Konfidenzbereich zum Niveau 1

- $C(x) = \Theta'$  für alle  $x$  in  $\Omega'$
- es ist sicher daß  $g(\vartheta) \in \Theta'$
- aber liefert keine Information über  $g(\vartheta)$  □

Allg. Konstruktionsproblem für Konfidenzbereiche:

- je kleiner  $C(x)$  ist, um so mehr Information über  $g(\vartheta)$  erhält man
- Die Aufgabe wird es sein, zu vorgegebenen Konfidenzniveau einen möglichst kleinen Konfidenzbereich zu konstruieren

**Beispiel 4.16.** Allg. Konstruktionsprinzip von Konfidenzbereichen zum Niveau  $1 - \alpha$   
 $\alpha$  gegeben

- wähle für jedes  $\vartheta'$  in  $\Theta'$  Teilmenge  $A(\vartheta') \subseteq \Omega'$  so daß  $\inf_{\vartheta \in g^{-1}(\vartheta')} (X_* P_{\vartheta})(A(\vartheta')) \geq 1 - \alpha$
- definiere  $C := \bigcup_{\vartheta' \in \Theta'} A(\vartheta') \times \{\vartheta'\}$

- dann  $C(x) := C \cap \text{pr}_{\Omega'}^{-1}(\{x\})$

es gilt  $\vartheta' \in C(x) \Leftrightarrow (x, \vartheta') \in C \Leftrightarrow x \in A(\vartheta')$

- dann ist

$$\begin{aligned}
 \inf_{\vartheta \in \Theta} P_{\vartheta}(\{g(\vartheta) \in C(X)\}) &= \inf_{\vartheta' \in \Theta'} \inf_{\vartheta \in g^{-1}(\vartheta')} P_{\vartheta}(\{g(\vartheta) \in C(X)\}) \\
 &= \inf_{\vartheta' \in \Theta'} \inf_{\vartheta \in g^{-1}(\vartheta')} (X_* P_{\vartheta})(\{x \in \Omega' \mid \vartheta' \in C(x)\}) \\
 &= \inf_{\vartheta' \in \Theta'} \inf_{\vartheta \in g^{-1}(\vartheta')} (X_* P_{\vartheta})(\{x \in \Omega' \mid x \in A(\vartheta')\}) \\
 &\geq 1 - \alpha
 \end{aligned}$$

sollten die Teilmengen  $A(\vartheta)$  möglichst klein wählen

□

### Beispiel 4.17. Konfidenzintervalle

- nehme an daß  $\Theta' = \mathbb{R}$ :

- dann ist  $C(x)$  oft ein Intervall: Konfidenzintervall

- oft konstruiert man  $C(x)$  in der Form  $C(x) := [\hat{g}(x) - \epsilon_u(x), \hat{g}(x) + \epsilon_o(x)]$  für Punktschätzer  $\hat{g}$

□

Betrachten jetzt Maße für die Qualität von Schätzern

- nehme an daß  $\Theta' = \mathbb{R}$  und die Schätzer meßbar sind

- schreiben  $\mathbb{E}_{\vartheta} := \mathbb{E}_{P_{\vartheta}}$

- dann ist  $\hat{g} \circ X : \Omega \rightarrow \mathbb{R}$  numerische Zufallsvariable

**Definition 4.18.** Der Schätzer  $\hat{g}$  für  $g$  heißt erwartungstreu, falls  $\mathbb{E}_{\vartheta}(\hat{g}(X)) = g(\vartheta)$  für alle  $\vartheta$  in  $\Theta$  gilt.

Die Zahl  $\text{Bias}_{\vartheta}(\hat{g}) := \mathbb{E}_{\vartheta}(\hat{g}(X)) - g(\vartheta)$  heißt der Bias des Schätzers

Beispiel 4.28 zeigt: Es gibt nicht immer erwartungstreue Schätzer

**Definition 4.19.** Die mittlere quadratische Abweichung des Schätzers  $\hat{g}$  von  $g$  ist die Funktion

$$\Theta \ni \vartheta \mapsto R(\vartheta, \hat{g}) := \mathbb{E}_{\vartheta}((\hat{g}(X) - g(\vartheta))^2) .$$

man kann die mittlere quadratische Abweichung durch Varianz und Bias ausdrücken

$$\begin{aligned}
 R(\vartheta, \hat{g}) &= \mathbb{E}_{\vartheta}((\hat{g}(X) - g(\vartheta))^2) \\
 &= \mathbb{E}_{\vartheta}((\hat{g}(X) - \mathbb{E}_{\vartheta}(\hat{g}(X))) - (g(\vartheta) - \mathbb{E}_{\vartheta}(\hat{g}(X))))^2 \\
 &= \mathbb{E}_{\vartheta}((\hat{g}(X) - \mathbb{E}_{\vartheta}(\hat{g}(X)))^2) - 2\mathbb{E}_{\vartheta}((\hat{g}(X) - \mathbb{E}_{\vartheta}(\hat{g}(X)))(g(\vartheta) - \mathbb{E}_{\vartheta}(\hat{g}(X)))) + \mathbb{E}_{\vartheta}(g(\vartheta) - \mathbb{E}_{\vartheta}(\hat{g}(X)))^2 \\
 &= \text{Var}_{\vartheta}(\hat{g}(X)) + \text{Bias}_{\vartheta}(\hat{g})^2
 \end{aligned}$$

- wollen daß  $R(\vartheta, \hat{g})$  für alle  $\vartheta$  klein ist
- wollen also  $\text{Var}_{\vartheta}(\hat{g}(X))$  und  $|\text{Bias}_{\vartheta}(\hat{g})|$  klein haben

**Beispiel 4.20.**

man kann immer erreichen, daß  $R(\vartheta_0, \hat{g}) = 0$  für eine Wahl von  $\vartheta_0$ :

- setze  $\hat{g} = \text{const}_{g(\vartheta_0)}$
- das liefert aber keine Information und außerdem wird  $R(\vartheta, \hat{g})$  für  $\vartheta \neq \vartheta_0$  nicht klein sein □

$(\Omega, \mathcal{B}, (P_{\vartheta})_{\vartheta \in \Theta})$  - statistisches Modell

$X : (\Omega, \mathcal{B}) \rightarrow (\Omega', \mathcal{B}')$  - Statistik

-  $g : \Theta \rightarrow \Theta'$  - zu schätzender Parameter

**Definition 4.21.** *Ein erwartungstreuer Schätzer  $\hat{g} : \Omega' \rightarrow \Theta'$  für  $g$  heißt bester erwartungstreuer Schätzer, falls für jeden anderen erwartungstreuen Schätzer  $\hat{g}' : \Omega' \rightarrow \Theta'$  die Ungleichung  $\text{Var}_{\vartheta}(\hat{g}(X)) \leq \text{Var}_{\vartheta}(\hat{g}'(X))$  für alle  $\vartheta$  in  $\Theta$  gilt.*

- werden später diskutieren, wie man beste erwartungstreue Schätzer erkenne kann

oft hat man die Situation, daß man das Experiment mehrfach ausführen (oder anderweitig verfeinern) kann um damit die Qualität der Schätzung zu verbessern.

das führt zu:

- Familie von Statistiken  $(X_n)_{n \in \mathbb{N}}$
- $X_n : \Omega \rightarrow \Omega'_n$
- Familie  $(\hat{g}_n)_{n \in \mathbb{N}}$  von Schätzern
- $\hat{g}_n : \Omega'_n \rightarrow \Theta'$

**Definition 4.22.** Die Familie von Schätzern  $(\hat{g}_n)_{n \in \mathbb{N}}$  heißt konsistent, falls  $\hat{g}_n(X_n) \xrightarrow{n \rightarrow \infty} g(\vartheta)$  in Wahrscheinlichkeit gilt für alle  $\vartheta$  in  $\Theta$ .

- Konsistenz bedeutet also:

$$\lim_{n \rightarrow \infty} P_\vartheta(\{|\hat{g}_n(X) - g(\vartheta)| \geq \epsilon\}) = 0$$

für alle  $\epsilon$  in  $(0, \infty)$  und  $\vartheta$  in  $\Theta$

- wir nehmen an, daß auf  $(\Omega', \mathcal{B}')$  schon ein Maß  $\mu$  definiert sei
- wir nehmen an, daß  $X_* P_\vartheta = f_\vartheta \mu$  gilt für alle  $\vartheta$  in  $\Theta$  für eine Familie von Dichten  $(f_\vartheta)_{\vartheta \in \Theta}$

**Definition 4.23.** Die Funktion  $\Theta \times \Omega' : (\vartheta, x) \mapsto f_\vartheta(x)$  heißt Likelihoodfunktion.

- beachte, daß die Likelihoodfunktion  $f_\vartheta$  nur eindeutig ist auf Änderung auf einer  $\mu$ -Nullmenge bestimmt ist

betrachten Schätzer  $\hat{g} : \Omega' \rightarrow \Theta'$  für  $g$

**Bemerkung 4.24.**

liegt  $\vartheta$  vor und ich beobachte  $x$  so daß  $f_\vartheta(x)$  sehr klein ist, dann sollte die Beobachtung ein statistischer Ausreisser sein

- die Wahrscheinlichkeit von  $x$  bei Vorliegen von  $\vartheta$  sollte groß sein
- man sollte davon ausgehen, daß die Beobachtung von  $x$  kein Ausreisser ist und deshalb der Wert von  $\vartheta$  so war, daß  $f_\vartheta(x)$  recht groß ist
- die Betrachtung des Extremfalls führt zum Begriff des Maximum-Likelihood Schätzers  $\square$

**Definition 4.25.** Der Schätzer  $\hat{g}$  heißt Maximum-Likelihood Schätzer für  $g$ , wenn

$$\sup_{\vartheta \in \Theta} f_{\vartheta}(x) = \sup_{\vartheta \in g^{-1}(\hat{g}(x))} f_{\vartheta}(x)$$

gilt für fast (bez.  $\mu$ ) alle  $x$  in  $\Omega'$ .

**Bemerkung 4.26.** Abhängigkeit von  $\mu$

$\mu'$  - weiteres Maß auf  $\Omega'$  mit gleicher Maßklasse

- dann  $\mu = \rho\mu'$  mit  $\rho > 0$  ( $\mu$  fast überall)

- dann  $f'_{\vartheta} = \rho f_{\vartheta}$  (Likelihoodfunktion bezüglich  $\mu'$ )

-  $\rho$  unabhängig von  $\vartheta$

- wenn  $\hat{g}$  Maximum-Likelihoodschätzer für Wahl  $\mu'$  ist, dann auch für  $\mu$

Die Bedingung, Maximum-Likelihoodschätzer zu sein, hängt nur von der Maßklasse von  $\mu$  ab!

□

**Bemerkung 4.27.**

wenn  $\vartheta \mapsto f_{\vartheta}(x)$  für jedes feste  $x$  ein eindeutig bestimmtes Maximum hat, dann gibt es einen eindeutigen Maximum-Likelihood-Schätzer

- liefert Methode, sinnvolle Schätzer zu finden

- für jedes  $x$  definiere  $\hat{g}(x)$  als Bild unter  $g$  des Ortes des eindeutigen Maximums von  $f_{-}(x)$

□

## 4.2 Schätzung der Erfolgswahrscheinlichkeit im Bernoulli Experiment

Situation: Werfe Reisszwecke

- wollen aus Beobachtung Wahrscheinlichkeit  $p$  für Rückenlage bestimmen

bilden statistisches Modell:

-  $\Omega = \prod_{\mathbb{N}} \{r, s\}$  (Rücken, Spitze nach unten)

-  $\mathcal{B}$  -  $\sigma$ -Algebra der Zylindermengen



- $(P_p)_{p \in [0,1]}$
- $P_p$  ist unendliches Produkt der W-Maße auf  $\{r, s\}$  mit Parameter  $p(r) = p$
- $p \in \Theta = [0, 1]$

Experiment:

- werfe Reisszwecke  $n$ -Mal und zähle Anzahl der Rückenlagen
- $\Omega'_n = \{0, 1, \dots, n\}$
- $X_n : \Omega \rightarrow \Omega'_n$
- $X_n(\omega) := \#\{i \in \{0, \dots, n-1\} \mid \omega_i = r\}$
- $X_{n,*}P_p$  hat Dichte  $f_{n,p}(x) := \binom{n}{x} p^x (1-p)^{n-x}$  bezüglich Zählmaßes

die zu schätzende Größe:

- $\Theta' = \Theta$
- $g = \text{id}_{[0,1]}$

Versuchen, MLS  $\hat{g}_n$  für  $\vartheta$  zu finden

- wollen Maximum von  $f_{n,p}(x)$  bez.  $p$  bestimmen
- $\ln(f_{n,p}(x)) = x \ln p - (n-x) \ln(1-p) + \ln\left(\binom{n}{x}\right)$
- ableiten nach  $p$ :
- $\ln(f_{n,p}(x))' = \frac{x}{p} - \frac{n-x}{1-p}$
- $\ln(f_{n,p}(x))'' = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2} < 0$
- Nullstelle:
  - $(1-p)x = p(n-x)$
  - $x = pn$
- dort ist das Maximum wegen zweiter Ableitung  $< 0$

setze  $\hat{g}_n(x) := \frac{x}{n}$

ist der Maximum-Likelihood-Schätzer

Der Schätzer  $\hat{g}_n$  ist Erwartungstreu:

$$\begin{aligned}
 \mathbb{E}_p(\hat{g}_n(X)) &= E_{X_n, * P_p}(\hat{g}) \\
 &= \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} \frac{x}{n} \\
 &= \sum_{x=1}^n \binom{n-1}{x-1} p^x (1-p)^{n-x} \frac{n}{x} \frac{x}{n} \\
 &= p \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-1-x} \\
 &= p
 \end{aligned}$$

**Beispiel 4.28.** Es gibt nicht immer erwartungstreue Schätzer

- sei  $\hat{h} : \Omega_n \rightarrow \mathbb{R}$  Schätzer für  $h(p) = \sqrt{p}$

- dann

$$\mathbb{E}_p(\hat{h}(X_n)) = \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} \hat{h}(x)$$

- das ist ein Polynom in  $p$

-  $h(p)$  ist kein Polynom

- also kann nicht  $\mathbb{E}_p(\hat{h}(X)) = h(p)$  für alle  $p$  in  $[0, 1]$  gelten

□

Die Folge  $(\hat{g}_n)_{n \in \mathbb{N}}$  ist konsistent:

- betrachten die unabhängige Folge der Zufallsvariablen  $(X_i)_{i \in \mathbb{N}}$

$$- Y_i(\omega) = \begin{cases} 1 & \omega_i = r \\ 0 & \omega_i = s \end{cases}$$

- dann ist  $\hat{g}_n(X) = S_n = \frac{1}{n} \sum_{i=0}^{n-1} Y_i$

- die Mitglieder der Folge  $(Y_i)_{i \in \mathbb{N}}$  sind identisch verteilt und integrierbar

- starkes Gesetz der großen Zahl gilt:  $S_n \xrightarrow{n \rightarrow \infty} p$  fast sicher

— insbesondere  $\hat{g}_n(X_n) \xrightarrow{n \rightarrow \infty} p$  fast sicher

Konfidenzintervalle:

- betrachten  $\hat{g}_6$
- suchen Konfidenzbereich für Niveau 0,8
- benutzen Prinzip vom Beispiel 4.16

$x \backslash p$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	$C(x)$
0	1.00	0.53	0.26	0.12	0.05	0.02	0.00	0.00	0.00	0.00	0.00	[0, 0.2]
1	0.00	0.35	0.40	0.30	0.19	0.09	0.04	0.01	0.00	0.00	0.00	[0.1, 0.4]
2	0.00	0.10	0.25	0.32	0.31	0.23	0.14	0.06	0.02	0.00	0.00	[0.2, 0.6]
3	0.00	0.015	0.08	0.19	0.28	0.31	0.28	0.19	0.08	0.01	0.00	[0.3, 0.6]
4	0.00	0.00	0.02	0.06	0.14	0.23	0.31	0.32	0.25	0.10	0.00	[0.4, 0.8]
5	0.00	0.00	0.00	0.01	0.04	0.09	0.19	0.30	0.40	0.35	0.00	[0.5, 0.9]
6	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.12	0.26	0.53	1.00	[0.8, 1]

listen in der Tabelle  $P_p(\{\hat{g}_6(X) = x\})$

- streichen in jeder Spalte der Größe nach so viele Zahlen rot, daß die Summe der roten Zahlen größer als 0.8 ist
- das entspricht der Menge  $A(\vartheta')$  in 4.16
- Lesen den Konfidenzbereich  $C(x)$  als Menge der roten Einträge in der Zeile für  $x$

haben hier eigentlich mit  $\Theta = \{0, 1/10, 2/10, \dots, 9/10, 1\}$  gearbeitet

- checken also nur  $P_p(\{p \in C(x)\})$  nur für  $p = i/10, i = 0, \dots, 10$

Optimalität von  $\hat{g}_n$ :

- $\hat{g}_n$  ist bester erwartungstreuer Schätzer für  $p$

**Dieser Teil des Beispiels kann erst nach Kapitel 4.5 verstanden werden**

benutzen Fischer-Information  $I_n(p)$ :

$$f_{n,p}(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ bez. Zählmaß}$$

$$-\ln f_{n,p}(x) = \ln\left(\binom{n}{x}\right) + x \ln(p) + (n-x) \ln(1-p)$$

$$-(\ln f_{n,p}(x))' = \frac{x}{p} - \frac{n-x}{1-p} = \frac{(1-p)x - p(n-x)}{p(1-p)} = \frac{x-np}{p(1-p)}$$

$$\begin{aligned}
I_n(p) &= E_{X_{n,*}P_p}(((\ln f_{n,p}(x))')^2) \\
&= \mathbb{E}_{Binom(n,p)}\left(\left(\frac{x-np}{p(1-p)}\right)^2\right) \\
&= \frac{1}{p^2(1-p)^2} \mathbb{E}_{Binom(n,p)}(x^2 - 2npx + (np)^2) \\
&= \frac{np(1-p) + n^2p^2 - 2(np)^2 + (np)^2}{p^2(1-p)^2} \\
&= \frac{np(1-p)}{p^2(1-p)^2} \\
&= \frac{n}{p(1-p)}
\end{aligned}$$

- benutze hier:  $\mathbb{E}_{Binom(n,p)}(x) = np$  und  $\mathbb{E}_{Binom(n,p)}(x^2) = np(1-p) + n^2p^2$

$$\begin{aligned}
\text{Var}_{P_p}(\hat{g}_n(X)) &= \mathbb{E}_{Binom(n,p)}((\hat{g} - p)^2) \\
&= \mathbb{E}_{Binom(n,p)}((x/n - p)^2) \\
&= \mathbb{E}_{Binom(n,p)}\left(\frac{x^2}{n^2} - \frac{2xp}{n} + p^2\right) \\
&= \frac{(np(1-p) + n^2p^2)}{n^2} - 2p^2 + p^2 \\
&= \frac{p(1-p)}{n}
\end{aligned}$$

-  $\text{Var}_{P_p}(\hat{g}_n(X)) = \frac{1}{I_n(p)}$

-  $g'(p) = 1$

**Bemerkung 4.29.** Optimalität ist kein Wunder:

- in der Tat:  $X_{n,*}P_p$  hat Dichte  $e^{\hat{g}(x)n(\ln(p)-\ln(1-p))+n\ln(1-p)+\ln\binom{n}{x}}$  bezüglich Zählmass

- das ist eine exponentielle Statistik mit  $s(p) = n(\ln(p) - \ln(1-p))$  und  $b(p) = -n \ln(1-p)$  und  $c(x) = \ln\binom{n}{x}$

- insofern Optimalität kein Wunder

□

### 4.3 Schätzung von Mittelwert und Varianz

Situation:

- haben unabhängige Folge identisch verteilter Zufallsvariablen
- $(Y_i)_{i \in \mathbb{N}}$
- beobachten endliche viele Glieder und wollen daraus den Mittelwert und die Varianz der gemeinsamen Verteilung schätzen

statistisches Modell und Experimente

- $(\Omega, \mathcal{B})$
- $(Y_n : \Omega \rightarrow \mathbb{R})_{n \in \mathbb{N}}$  - Folge meßbarer Abbildung
- $\Theta = \{P \in \mathcal{P}(\Omega) \mid (Y_n)_{n \in \mathbb{N}} \text{ ist unabhängig, identisch verteilt mit Momenten bis zur Ordnung } 4\}$

Modell:  $(\Omega, \mathcal{B}, (P)_{P \in \Theta})$

$n$ -tes Experiment:

- $(\Omega'_n, \mathcal{B}'_n) = (\mathbb{R}^n, \text{Borel})$
- $X_n := (Y_0, \dots, Y_{n-1}) : \Omega \rightarrow \mathbb{R}^n$

zu schätzende Größen:

- $\mu : \Theta \rightarrow \mathbb{R}, \quad g(P) := \mathbb{E}_P(Y_0)$
- $v : \Theta \rightarrow \mathbb{R}, \quad v(P) = \text{Var}_P(Y_0)$
- man kann hier  $Y_0$  durch  $Y_i$  für  $i$  in  $\mathbb{N}$  beliebig ersetzen

wissen nichts über  $X_{n,*}P$  - können also Maximum-Likelihood-Methode nicht verwenden

**Beispiel 4.30.** betrachten speziellere Situation:

- unabhängige Familie normalverteilter Zufallsvariablen

$$(\Omega, \mathcal{B}) := \prod_{\mathbb{N}}(\mathbb{R}, \text{Borel})$$

$$\Theta = \mathbb{R} \times (0, \infty), (\mu, \sigma) \in \Theta$$

$$- P_{\mu, \sigma} := \prod_{\mathbb{N}} \text{Norm}(\mu, \sigma)$$

- Modell also  $(\Omega, \mathcal{B}, (P_{\mu, \sigma})_{\mu \in \mathbb{R}, \sigma \in (0, \infty)})$
- $Y_i$  - Projektion auf  $i$ te Komponente

wollen  $\mu$  und  $\sigma^2$  schätzen

- $X_n = (Y_0, \dots, Y_{n-1}) : \Omega \rightarrow \mathbb{R}^n$
- dann  $X_{n,*} P_{\mu, \sigma} = f_{n, \mu, \sigma} \text{Leb}$  - mehrdimensionale Normalverteilung
- $f_{n, \mu, \sigma} := \prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$
- suchen Maximum von  $f_{n, \mu, \sigma}(x_0, \dots, x_{n-1})$  bez.  $(\mu, \sigma)$  (für festes  $x_0, \dots, x_{n-1}$ )

$$0 = \partial_{\mu} \ln f_{n, \mu, \sigma} = -\frac{1}{\sigma^2} \sum_{i=0}^{n-1} (x_i - \mu)$$

$$\text{- gibt } \mu = \frac{1}{n} \sum_{i=0}^{n-1} x_i$$

$$0 = \partial_{\sigma} \ln f_{n, \mu, \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=0}^{n-1} (x_i - \mu)^2$$

$$\text{- gibt } \sigma^2 = \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \mu)^2$$

-

Check: diese Stelle ist ein Maximum

in dieser Situation:

$$\hat{\mu}_n(x) := \frac{1}{n} \sum_{i=0}^{n-1} x_i$$

$$\hat{v}_n(x) := \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \hat{\mu}_n(x))^2$$

sind MLH Schätzer für  $\mu$  und  $\sigma^2$

□

zurück zur allgemeinen Situation:

$$\text{- raten: } \hat{\mu}_n : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \hat{\mu}_n(x) := \frac{1}{n} \sum_{i=0}^{n-1} x_i$$

Erwartungstreue:

$$\mathbb{E}_P(\hat{\mu}_n(X_n)) = \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}_P(Y_i) = \mathbb{E}_P(Y_0) = \mu(P)$$

Konsistenz:

Starkes Gesetz der großen Zahl:

- für jedes  $P$  in  $\Theta$  gilt :  $\hat{\mu}_n(X_n) \xrightarrow{n \rightarrow \infty} \mu(P)$  fast sicher
- also ist  $(\hat{\mu}_n)_{n \in \mathbb{N}}$  konsistent

raten Schätzer für Varianz:

$$- \hat{v}_n : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \hat{v}_n(x) := \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \hat{\mu}_n)^2$$

Konsistenz:

$$\hat{v}_n = \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \hat{\mu}_n)^2 = \frac{1}{n} (\sum_{i=0}^{n-1} x_i^2 - 2\hat{\mu}_n \sum_{i=0}^{n-1} x_i + n\hat{\mu}_n^2) = \frac{1}{n} \sum_{i=0}^{n-1} x_i^2 - \hat{\mu}_n^2$$

- da  $P$  Momente bis zur Ordnung 4 hat, dann gilt starkes Gesetz der großen Zahl für  $(X_i^2)_{i \in \mathbb{N}}$
- $\hat{v}_n(X_n) \xrightarrow{n \rightarrow \infty} v(P)$  fast sicher
- also ist  $(\hat{v}_n)_{n \in \mathbb{N}}$  konsistent

Erwartungstreue:

- Vorbereitung:

$$- \text{schreibe } \hat{\mu}_n(X_n) - \mu(P) = \frac{1}{n} \sum_{i=0}^{n-1} (Y_i - \mu(P))$$

-  $(Y_i)_{i \in \mathbb{N}}$  - unkorreliert

$$- \mathbb{E}_P((\hat{\mu}_n(X_n) - \mu(P))^2) = \frac{1}{n} v(P)$$

$$\begin{aligned}
\mathbb{E}_P(\hat{v}_n(X_n)) &= \frac{1}{n} \mathbb{E}_P\left(\sum_{i=0}^{n-1} (Y_i - \hat{\mu}_n(X_n))^2\right) \\
&= \frac{1}{n} \mathbb{E}_P\left(\sum_{i=0}^{n-1} [(Y_i - \mu(P)) - (\hat{\mu}_n(X_n) - \mu(P))]^2\right) \\
&= \frac{1}{n} \mathbb{E}_P\left(\sum_{i=0}^{n-1} [(Y_i - \mu(P))^2 - 2(Y_i - \mu(P))(\hat{\mu}_n(X_n) - \mu(P)) + (\hat{\mu}_n(X_n) - \mu(P))^2]\right) \\
&= v(P) - 2\mathbb{E}_P((\hat{\mu}_n(X_n) - \mu(P))^2) + \mathbb{E}_P((\hat{\mu}_n(X_n) - \mu(P))^2) \\
&= v(P) - \frac{2}{n}v(P) + \frac{1}{n}v(P) \\
&= \left(1 - \frac{1}{n}\right)v(P)
\end{aligned}$$

der Schätzer  $\hat{v}_n$  ist also nicht erwartungstreu

Korrigieren:  $\hat{s}_n := \left(1 - \frac{1}{n}\right)^{-1} \hat{v}_n = \frac{1}{n-1} \sum_{i=0}^{n-1} (x_i - \hat{\mu}_n)^2$

- dann ist  $\hat{s}_n$  erwartungstreu

- die Familie  $(\hat{s}_n)_{n \in \mathbb{N}}$  ist immer noch konsistent (da  $\left(1 - \frac{1}{n}\right)^{-1} \xrightarrow{n \rightarrow \infty} 1$ )

da wir keine weitere Information über  $P$  haben, ist es im allgemeinen nicht möglich, Konfidenzintervalle anzugeben

- machen deshalb Zusatzannahme: betrachten Teilfamilie von  $\Theta$  mit  $\text{Var}_P(X_0) \leq v$

- dann  $\text{Var}(\hat{\mu}_n(X_n)) \leq \frac{v}{n}$

- Chebyshev:  $P(\{|\hat{\mu}_n(X_n) - \mu(P)| \geq \epsilon\}) \leq \frac{v}{n\epsilon^2}$

- schreiben  $x = (x_0, \dots, x_{n-1})$

- dann ist

$$C(x) := \left[ \hat{\mu}(x) - \sqrt{\frac{v}{n\alpha}}, \hat{\mu}(x) + \sqrt{\frac{v}{n\alpha}} \right]$$

ein Konfidenzintervall für  $\mu$  zum Konfidenzniveau  $1 - \alpha$



## 4.4 Taxiproblem

Situation:

habe Zufallsgenerator, welcher unabhängige gleichverteilte Zufallszahlen im Intervall  $[0, m]$  erzeugt

- beobachte die ersten  $n$  Ergebnisse und möchte  $m$  schätzen

$$(\Omega, \mathcal{B}) := \prod_{\mathbb{N}}(\mathbb{R}, \text{Borel})$$

$$\Theta = (0, \infty), m \in \Theta$$

$$- P_m = \prod_{\mathbb{N}} \frac{1}{m} \chi_{[0, m]} \text{Leb}$$

Modell:  $(\Omega, \mathcal{B}, (P_m)_{m \in \Theta})$

$Y_i : \Omega \rightarrow \mathbb{R}$  - ite Projektion

$n$ -tes Experiment

$$(\Omega'_n, \mathcal{B}'_n) = (\mathbb{R}^n, \text{Borel})$$

$$- X_n := (Y_0, \dots, Y_{n-1}) : \Omega \rightarrow \mathbb{R}^n$$

MLS für  $m$ :

$$- X_{n,*} P_m = \prod_{i=0}^{n-1} \frac{1}{m} \chi_{[0, m]}(x_i) \text{Leb}$$

$$- m \mapsto \prod_{i=0}^{n-1} \frac{1}{m} \chi_{[0, m]}(x_i) \text{ hat Maximum bez. } m \text{ in } \max\{x_0, \dots, x_{i-1}\}$$

- in der Tat

$$\prod_{i=0}^{n-1} \frac{1}{m} \chi_{[0, m]}(x_i) = \begin{cases} 0 & m < \max\{x_0, \dots, x_{i-1}\} \\ \frac{1}{m} & m \geq \max\{x_0, \dots, x_{i-1}\} \end{cases}$$

$$- \text{MLS für } m \text{ ist } \hat{m}_n(x_0, \dots, x_{n-1}) = \max\{x_0, \dots, x_{i-1}\}$$

anderer Schätzer für  $m$

$$- \text{wissen } \mathbb{E}_m(X_i) = \frac{1}{2}m$$

$$- \text{raten: } \hat{m}'_n(x_0, \dots, x_{n-1}) := \frac{2}{n} \sum_{i=0}^{n-1} x_i$$

untersuchen Eigenschaften

Konsistenz:

Starkes Gesetz der Großen Zahl gilt für  $\hat{m}'_n(X_n)$

-  $(\hat{m}'_n)_{n \in \mathbb{N}}$  ist konsistent

$$\begin{aligned} P_m(\{|\hat{m}'_n(X_n) - m| \geq \epsilon\}) &= P_m(\{\max\{Y_0, \dots, Y_{n-1}\} \leq m - \epsilon\}) \\ &= P_m(\{Y_0 \leq m - \epsilon, \dots, Y_n \leq m - \epsilon\}) \\ &= \left(\frac{m - \epsilon}{m}\right)^n \\ &\xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

-  $(\hat{m}_n)_{n \in \mathbb{N}}$  ist konsistent

Erwartungstreue:

$$\mathbb{E}_m(\hat{m}'_n(X_n)) = m$$

-  $\hat{m}'_n$  ist Erwartungstreu

bestimmen Dichte von  $\hat{m}_n(X_n) * P_m$  auf  $\mathbb{R}$

$$- P_m(\{\hat{m}_n(X_n) \leq x\}) = \left(\frac{x}{m}\right)^n$$

$$- \partial_x \left(\frac{x}{m}\right)^n = \frac{n}{m} \left(\frac{x}{m}\right)^{n-1}$$

$$\begin{aligned} \mathbb{E}_m(\hat{m}_n(X_n)) &= \int_0^m x \frac{n}{m} \left(\frac{x}{m}\right)^{n-1} dx \\ &= \frac{nm}{(n+1)} \end{aligned}$$

also ist  $\hat{m}_n$  nicht erwartungstreu

$$\text{Korrektur: } \hat{m}_n'' := \frac{n+1}{n} \hat{m}_n$$

- ist erwartungstreu und auch konsistent

welcher von  $\hat{m}'_n$  und  $\hat{m}_n''$  ist besser?

untersuchen dazu Varianz:

$$\begin{aligned}
 \text{Var}_m(\hat{m}'_n(X_n)) &= \text{Var}_m\left(\frac{2}{n} \sum_{i=0}^{n-1} Y_i\right) \\
 &= \frac{4}{n} \text{Var}_m(Y_0) \\
 &= \frac{4}{nm} \int_0^m \left(x - \frac{m}{2}\right)^2 dx \\
 &= \frac{4}{nm} \int_0^m \left(x^2 - mx + \frac{m^2}{4}\right) dx \\
 &= \frac{4}{nm} \left(\frac{m^3}{3} - \frac{m^3}{2} + \frac{m^3}{4}\right) \\
 &= \frac{m^2}{3n}
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}_m(\hat{m}_n(X_n)) &= \mathbb{E}_m(\hat{m}_n(X_n)^2) - \mathbb{E}_m(\hat{m}_n(X_n))^2 \\
 &= \int_0^m x^2 \frac{n}{m} \left(\frac{x}{m}\right)^{n-1} dx - \frac{m^2 n^2}{(n+1)^2} \\
 &= \frac{nm^2}{(n+2)} - \frac{m^2 n^2}{(n+1)^2} \\
 &= m^2 \frac{n(n+1)^2 - n^2(n+2)}{(n+1)^2(n+2)} \\
 &= m^2 \frac{n}{(n+1)^2(n+2)}
 \end{aligned}$$

also  $\text{Var}_m(\hat{m}''_n(X_n)) = \frac{(n+1)^2}{n^2} m^2 \frac{n}{(n+1)^2(n+2)} = m^2 \frac{1}{n(n+2)}$

- wegen  $n(n+2) \geq 3n$  gilt  $\text{Var}_m(\hat{m}''_n(X_n)) \leq \text{Var}_m(\hat{m}'_n(X_n))$
- der Schätzer  $\hat{m}''_n$  ist also "besser" als  $\hat{m}'_n$

Konfidenzintervall:

Suchen Intervall  $C(x)$  als Funktion von  $\hat{m}_n(x) := \max\{x_0, \dots, x_{n-1}\}$  zum Konfidenzniveau  $1 - \alpha$

- suchen zunächst  $m \mapsto \delta(m)$  derart daß  $P_m(\{\hat{m}(X_n) \in [m - \delta(m), m]\}) = 1 - \alpha$

$$\begin{aligned} P_m(\{\hat{m}(X_n) \in [m - \delta(m), m]\}) &= \int_{m-\delta(m)}^m \frac{nx^{n-1}}{m^n} dx \\ &= 1 - \left(\frac{m - \delta(m)}{m}\right)^n \end{aligned}$$

lösen nach  $\delta(m)$  gibt:  $\delta(m) = m(1 - \alpha^{1/n})$

Teilmenge  $C = \{(m, r) \mid m - \delta(m) \leq r \leq m\}$  von  $(0, \infty) \times \mathbb{R}$  ist Konfidenzbereich

Konfidenzbereich bei  $r$  ist dann durch  $C(r) = \{m \mid (m, r) \in C\}$  gegeben

-  $(m, r) \in C$  heißt  $m - \delta(m) \leq r \leq m$

- also  $C(r) = [r, r + \epsilon(r)]$  mit oberer Grenze bestimmt durch

-  $(r + \epsilon(r)) - \delta(r + \epsilon(r)) = r$

—  $\epsilon(r) - (r + \epsilon(r))(1 - \alpha^{1/n}) = 0$

—  $\alpha^{1/n}\epsilon(r) = r(1 - \alpha^{1/n})$

—  $\epsilon(r) = \frac{r(1 - \alpha^{1/n})}{\alpha^{1/n}}$

Konfidenzbereich zum Niveau  $1 - \alpha$  ist

$$C(r) = \left[r, \frac{r}{\alpha^{1/n}}\right]$$

- wird immer kleiner für  $n \rightarrow \infty$

Zahlenbeispiel

$\alpha = 0.05, r = 1$

wollen  $\alpha^{1/n} \geq 0.99$

-  $0.05^{1/n} \geq 0.99$

-  $n \geq \ln(0.05)/\ln(0.99) = 298$

- Damit ich  $m$  auf zwei Kommastellen mit Konfidenz 0.95 bestimmen kann muß ich ca 300 Beobachtungen machen.

## 4.5 Die Informationsungleichung

$(\Omega, \mathcal{B}, (P_\vartheta)_{\vartheta \in \Theta})$  - statistisches Modell

-  $X : (\Omega, \mathcal{B}) \rightarrow (\Omega', \mathcal{B}')$  - Statistik

-  $g : \Theta \rightarrow \Theta'$  - zu schätzende Funktion

Erinnerung: Definition 4.21 des Begriffs "bester erwartungstreuer Schätzer"

**Lemma 4.31.** *Ein bester erwartungstreuer Schätzer ist eindeutig.*

*Beweis.*

betrachten besten erwartungstreuen Schätzer  $\hat{g}$

- sei  $\hat{g}'$  weiterer bester erwartungstreuer Schätzer

- wissen:  $\text{Var}_\vartheta(\hat{g}(X)) = \text{Var}_\vartheta(\hat{g}'(X))$  für alle  $\vartheta$  in  $\Theta$

setzen

-  $\hat{g}'' := \frac{1}{2}(\hat{g} + \hat{g}')$  ist auch erwartungstreuer Schätzer

$$\begin{aligned} \text{Var}_\vartheta(\hat{g}''(X)) &= \frac{1}{4}(\text{Var}_\vartheta(\hat{g}(X)) + \text{Var}_\vartheta(\hat{g}'(X)) + 2\text{Cov}_\vartheta(\hat{g}(X), \hat{g}'(X))) \\ &\stackrel{!}{\leq} \frac{1}{4}(\text{Var}_\vartheta(\hat{g}'(X)) + \text{Var}_\vartheta(\hat{g}'(X)) + 2\text{Var}_\vartheta(\hat{g}(X))^{1/2}\text{Var}_\vartheta(\hat{g}'(X))^{1/2}) \\ &= \text{Var}_\vartheta(\hat{g}(X)) \\ &\leq \text{Var}_\vartheta(\hat{g}''(X)) \end{aligned}$$

- Cauchy-Schwarz bei !

- also gilt Gleichheit

- Cauchy-Schwarz (Diskussion des Gleichheitsfalls) :

$$\text{Cov}_\vartheta(\hat{g}(X), \hat{g}'(X)) = \mathbb{E}_{X^*P_\vartheta}((\hat{g}-g(\vartheta))(\hat{g}'-g(\vartheta))) = \mathbb{E}_{X^*P_\vartheta}((\hat{g}-g(\vartheta))^2)^{1/2}\mathbb{E}_\vartheta((\hat{g}'-g(\vartheta))^2)^{1/2}$$

- es gibt  $a(\vartheta)$  mit  $\hat{g} - g(\vartheta) = a(\vartheta)(\hat{g}' - g(\vartheta))$  ( $X_*P_\vartheta$ - fast sicher)
- also  $\hat{g} = a(\vartheta)(\hat{g}' - g(\vartheta)) + g(\vartheta) = a(\vartheta)\hat{g}' + b(\vartheta)$
- mit  $b(\vartheta) := g(\vartheta)(1 - a(\vartheta))$

bestimmen  $a$  und  $b$

$$\begin{aligned}\text{Var}_\vartheta(\hat{g}(X)) &= \text{Cov}_\vartheta(\hat{g}(X), \hat{g}'(X)) \\ &= \mathbb{E}_\vartheta(a(\vartheta)(\hat{g}'(X) - g(\vartheta))(\hat{g}'(X) - g(\vartheta))) \\ &= a(\vartheta)\text{Var}_\vartheta(\hat{g}'(X))\end{aligned}$$

impliziert

- $a(\vartheta) = 1$  und  $b(\vartheta) = 0$
- also  $\hat{g} = \hat{g}'$

□

es ergibt sich die Frage, wie man beste erwartungstreue Schätzer finden oder erkennen kann

- brauchen berechenbare untere Schranke für die Varianz erwartungstreuer Schätzer
- wenn ein erwartungstreuer Schätzer diese Schranke realisiert, dann ist er ein bester erwartungstreuer Schätzer

untere Schranke der Varianz von Schätzern durch Fisher Information

Annahmen:

- $\Theta$  - offenes Intervall in  $\mathbb{R}$
- $X_*P_\vartheta = f_\vartheta\mu$
- alle Ableitungen und Integrale unten mögen existieren (genaue Details möge der aufmerksame Leser selbst hinzufügen)

folgende Gleichungen gelten für alle  $\vartheta$  in  $\Theta$

- Normierung  $1 = \int_{\Omega'} f_\vartheta d\mu$

-  $\mathbb{E}_{X_*P_\vartheta}(\hat{g}) = g(\vartheta) + \text{Bias}_\vartheta(\hat{g})$

betrachten  $\mathcal{L}_\vartheta := \ln f_\vartheta$

- heißt oft “log-Likelihoodfunktion”

- ' - Ableitung nach  $\vartheta$

-  $0 = \left(\int_{\Omega'} f_\vartheta d\mu\right)' = \int_{\Omega'} f_\vartheta' d\mu = \int_{\Omega'} \frac{f_\vartheta'}{f_\vartheta} f_\vartheta d\mu = \mathbb{E}_{X_*P_\vartheta}(\mathcal{L}'_\vartheta)$

-  $g'(\vartheta) + \text{Bias}_\vartheta(\hat{g})' = \mathbb{E}_{X_*P_\vartheta}(\hat{g})' = \int_{\Omega'} \hat{g} f_\vartheta' d\mu = \mathbb{E}_{X_*P_\vartheta}(\hat{g}\mathcal{L}'_\vartheta) = \mathbb{E}_{X_*P_\vartheta}((\hat{g} - \mathbb{E}_{X_*P_\vartheta}(\hat{g}))\mathcal{L}'_\vartheta)$

Cauchy-Schwarz

$$(g'(\vartheta) + \text{Bias}_\vartheta(\hat{g})')^2 \leq \mathbb{E}_{X_*P_\vartheta}((\hat{g} - \mathbb{E}_{X_*P_\vartheta}(\hat{g}))^2) \mathbb{E}_{X_*P_\vartheta}(\mathcal{L}'_\vartheta{}^2)$$

**Definition 4.32.**  $I(\vartheta) := \mathbb{E}_{X_*P_\vartheta}(\mathcal{L}'_\vartheta{}^2)$  heißt Fisher Information.

-  $I(\vartheta)$  ist groß wenn  $X_*P_\vartheta$  stark von  $\vartheta$  abhängt

**Bemerkung 4.33.** Abhängigkeit der Fisher Information von  $\mu$ :

ersetzt man  $\mu$  durch  $h\mu$  und  $f_\vartheta$  durch  $\tilde{f} := f_\vartheta h^{-1}$

- dann  $\tilde{\mathcal{L}}_\vartheta = \mathcal{L}_\vartheta - \ln(h)$

-  $\mathcal{L}'_\vartheta = \tilde{\mathcal{L}}'_\vartheta$

-  $I(\vartheta)$  hängt nicht von der Wahl des Hintergrundmaßes  $\mu$  sondern nur seiner Maßklasse ab □

haben also Ungleichung

$$(g'(\vartheta) + \text{Bias}_\vartheta(\hat{g})')^2 \leq \text{Var}_\vartheta(\hat{g}(X))I(\vartheta)$$

betrachten Spezialfall daß  $\hat{g}$  erwartungstreu ist:

-  $\text{Bias}_\vartheta(\hat{g})' = 0$

-  $g'(\vartheta)^2 \leq \text{Var}_\vartheta(\hat{g}(X))I(\vartheta)$

**Korollar 4.34** (Fréchet, Cramér, Rao). *Informationsungleichung:*

$$\frac{g'(\vartheta)^2}{I(\vartheta)} \leq \text{Var}_\vartheta(\hat{g}(X))$$

-  $\frac{g'(\vartheta)^2}{I(\vartheta)}$  ist eine untere Schranke der Varianz von Schätzern für  $\vartheta$  auf Basis des Experiments  $X$

- die Varianz kann also im allgemeinen nicht durch gute Wahl von  $\hat{g}$  beliebig klein gemacht werden

- wenn  $X_*P_\vartheta$  praktisch nicht von  $\vartheta$  abhängt, dann ist  $\text{Var}_\vartheta(\hat{g}(X))$  sehr groß

- durch Beobachtung von  $X$  kann ich praktisch keine Information über  $g(\vartheta)$  gewinnen

**Beispiel 4.35.**  $N$  Unabhängige Wiederholung des Experiments

- liefert Fischer-Information  $I_N(\vartheta)$

Beh:  $I_N(\vartheta) = nI(\vartheta)$

$X^N := (X_n)_{n=1, \dots, N}$

- hat Werte in  $(\Omega'^N, \mathcal{B}'^N) := \prod_{n=1}^N (\Omega', \mathcal{B}')$

- nehmen Produktmaß  $\mu^N$

- unabhängig bedeutet:

-  $X_*^N P_\vartheta = \prod_{n=1}^N X_* P_\vartheta = (\prod_{n=1}^N \text{pr}_n^* f_\vartheta) \mu^N$

-  $f_\vartheta^N = \prod_{n=1}^N \text{pr}_n^* f_\vartheta$

-  $\mathcal{L}_\vartheta^N = \sum_{n=1}^N \text{pr}_n^* \mathcal{L}_\vartheta$

-  $(\mathcal{L}_\vartheta^{N'})^2 = \sum_{n=1}^N \text{pr}_n^* \mathcal{L}_\vartheta'^2 + \sum_{n,m=1}^N \text{pr}_n^* \mathcal{L}_\vartheta' \text{pr}_m^* \mathcal{L}_\vartheta'$

- wenden  $\mathbb{E}_{X_*^N P_\vartheta}$  an:

- wegen Unabhängigkeit und  $\mathbb{E}_{X_{n,*} P_\vartheta}(\mathcal{L}_\vartheta') = 0$  verschwinden gemischte Terme

-  $I_N(\vartheta) = \mathbb{E}_{X_*^N P_\vartheta}((\mathcal{L}_\vartheta^{N'})^2) = nI(\vartheta)$



wenn ich das Experiment wiederholen kann, dann habe ich eine Chance, eine Folge immer besserer Schätzer zu finden

□

Diskutieren den Gleichheitsfall der Informationsungleichung:

– Gleichheitsfall bei Cauchy-Schwarz

– für jedes  $\vartheta \in \Theta$  muß gelten:  $a(\vartheta)(\hat{g} - \mathbb{E}_{X_*P_\vartheta}(\hat{g})) = \mathcal{L}'_\vartheta$

–  $a(\vartheta)$  bestimmt sich aus  $a(\vartheta)^{-1} = \mathbb{E}_{X_*P_\vartheta}\left(\frac{\hat{g} - \mathbb{E}_{X_*P_\vartheta}(\hat{g})}{\mathcal{L}'_\vartheta}\right)$

–  $s(\vartheta)$  sei Stammfunktion von  $a(\vartheta)$

– dann

–  $\mathcal{L}_\vartheta(x) = s(\vartheta)\hat{g}(x) - b(\vartheta) + c(x)$  für geeignete Funktion  $b(\vartheta)$  und  $c(x)$

– in der Tat:

—  $\mathcal{L}_\vartheta(x)' = a(\vartheta)\hat{g}(x) - b'(\vartheta)$

– müssen  $b$  so wählen daß  $b(\vartheta)' = a(\vartheta)\mathbb{E}_{X_*P_\vartheta}(\hat{g})$

–  $f_\vartheta(x) = e^{s(\vartheta)\hat{g}(x) - b(\vartheta) + c(x)}$

–  $b$  bestimmt sich auch aus Normierung:  $b(\vartheta) = \ln\left(\int_\Omega e^{s(\vartheta)\hat{g}(x) + c(x)} d\mu(x)\right)$

$(\Omega, \mathcal{B}, (P_\vartheta)_{\vartheta \in \Theta})$  gegeben

$X : (\Omega, \mathcal{B}) \rightarrow (\Omega', \mathcal{B}')$  gegeben

–  $\mu$  - Maß auf  $(\Omega', \mathcal{B}')$

$\hat{g} : \Omega' \rightarrow \mathbb{R}$  gegeben

– setzen  $g(\vartheta) := \mathbb{E}_{X_*P_\vartheta}(\hat{g})$

– dann ist  $\hat{g}$  erwartungstreuer Schätzer für  $g$

**Definition 4.36.** Die Statistik  $X$  heißt exponentiell bezüglich  $\hat{g}$  wenn es Funktionen  $s(\vartheta) \in C^1(\Theta)$  ohne kritische Punkte und  $c : \Omega' \rightarrow \mathbb{R}$  (meßbar) gibt mit

$$X_*P_\vartheta = e^{s(\vartheta)\hat{g}(x) - b(\vartheta) + c(x)} \mu ,$$

wobei (notwendiger Weise)

$$b(\vartheta) = \ln\left(\int_{\Omega'} e^{s(\vartheta)\hat{g}(x)+c(x)} d\mu(x)\right)$$

gilt.

**Lemma 4.37.** Wenn die Statistik bezüglich  $\hat{g}$  exponentiell ist, dann ist für alle  $\vartheta$  in  $\Theta$

1.  $b'(\vartheta) = s'(\vartheta)g(\vartheta)$
2.  $g'(\vartheta) = s'(\vartheta)\mathbb{V}\text{ar}_{\vartheta}(\hat{g}(X))$
3.  $I(\vartheta) = g'(\vartheta)s'(\vartheta)$
4.  $\mathbb{V}\text{ar}_{\vartheta}(\hat{g}(X)) = \frac{g'(\vartheta)^2}{I(\vartheta)}$  und  $\hat{g}$  ist ein bester erwartungstreuer Schätzer für  $g$

*Beweis.*

$$\begin{aligned} 0 &= \left(\int_{\Omega'} e^{s(\vartheta)\hat{g}(x)-b(\vartheta)+c(x)} d\mu\right)' \\ &= s'(\vartheta) \int_{\Omega'} e^{s(\vartheta)\hat{g}(x)-b(\vartheta)+c(x)} \hat{g}(x) d\mu - b'(\vartheta) \int_{\Omega'} e^{s(\vartheta)\hat{g}(x)-b(\vartheta)+c(x)} d\mu \\ &= s'(\vartheta)\mathbb{E}_{\vartheta}(\hat{g}(X)) - b'(\vartheta) \\ &= s'(\vartheta)g(\vartheta) - b'(\vartheta) \end{aligned}$$

$$\begin{aligned} g'(\vartheta) &= \left(\int_{\Omega'} e^{s(\vartheta)\hat{g}(x)-b(\vartheta)+c(x)} \hat{g}(x) d\mu(x)\right)' \\ &= s'(\vartheta) \int_{\Omega'} e^{s(\vartheta)\hat{g}(x)-b(\vartheta)+c(x)} \hat{g}(x)^2 d\mu(x) - b'(\vartheta) \int_{\Omega'} e^{s(\vartheta)\hat{g}(x)-b(\vartheta)+c(x)} \hat{g}(x) d\mu(x) \\ &= s'(\vartheta)\mathbb{E}_{\vartheta}(\hat{g}(X)^2) - b'(\vartheta)g(\vartheta) \\ &= s'(\vartheta)(\mathbb{E}_{\vartheta}(\hat{g}(X)^2) - g(\vartheta)^2) \\ &= s'(\vartheta)\mathbb{V}\text{ar}_{\vartheta}(\hat{g}(X)) \end{aligned}$$

haben Gleichheitsfall in Informationsungleichung

$$\begin{aligned}
I(\vartheta)\text{Var}_{\vartheta}(\hat{g}(X)) &= g'(\vartheta)^2 \\
&= s'(\vartheta)\text{Var}_{\vartheta}(\hat{g}(X))g'(\vartheta)
\end{aligned}$$

4. ist klar

□

hier Beispiel 4.2 beenden

## 4.6 MLS und Konsistenz

wollen zeigen, daß MLS für wiederholte unabhängige Versuche eine konsistente Folge von Schätzern geben

Vorbereitung:

$(\Omega, \mathcal{B}, \mu)$  - Maßraum

$P = f\mu$  und  $Q = g\mu$  - W-Maße mit Dichten bez.  $\mu$

- wenn  $P$  zu  $Q$  absolutstetig ist

- d.h.  $Q(A) = 0 \Rightarrow P(A) = 0$

- dann nach Radon-Nikodym existiert  $\sigma : \Omega \rightarrow [0, \infty)$  (eind.  $Q$ -fast überall) mit  $f = g\sigma$

betrachten  $P$  und  $A$  auf  $(\Omega, \mathcal{B}, \mu)$  mit  $P = f\mu$  und  $Q = g\mu$

**Definition 4.38.** Die relative Entropie von  $P$  bezüglich  $Q$  wird durch

$$H(P; Q) := \mathbb{E}_P(\ln \sigma)$$

definiert, falls  $P$  zu  $Q$  absolutstetig ist und sonst als  $H(P; Q) := \infty$ .

**Lemma 4.39.** Die relative Entropie von  $P$  bezüglich  $Q$  ist wohldefiniert. Es gilt  $H(P; Q) \geq 0$  und  $H(P; Q) = 0$  genau dann wenn  $P = Q$ .

*Beweis.*

-  $\psi(s) := 1 - s + s \ln(s)$  auf  $[0, \infty)$  ist strikt konvex und hat Minimum 0 bei 1

-  $\psi' = \ln(s)$  - ist negativ für  $s < 1$  und positiv für  $s > 1$

-  $\psi(1) = 0$

1. Fall:  $P$  absolutstetig zu  $Q$

- schreiben  $f = \sigma g$

- wegen  $\psi \geq 0$  ist linke Seite definiert (in  $[0, \infty]$ )

-  $\mathbb{E}_Q(\psi(\sigma)) = \int_{\Omega} (1 - \sigma + \sigma \ln(\sigma)) g d\mu = \int_{\Omega} \ln(\sigma) f d\mu = H(P; Q)$

- benutzen hier Normierung von  $P$  und  $Q$  um die ersten beiden Terme loszuwerden

- schließen:  $H(P; Q) \geq 0$ .

-  $g, \sigma g = f$  sind  $\mu$ -integabel (Normierung)

-  $\psi(\sigma)g \geq 0$

- impliziert:  $[\sigma \ln(\sigma)]^-$  (negativer Teil) ist  $g\mu$ -integabel

-  $[\ln(\sigma)]^-$  ist  $f\mu$ -integabel

- also  $H(P; Q)$  wohldefiniert (aber möglicherweise  $\infty$ )

$H(P; Q) = 0$  genau wenn  $\psi(\sigma) = 0$   $Q$ -fast überall

- genau dann wenn  $\sigma = 1$   $Q$ -fast überall

- genau dann wenn  $f = g$   $Q$ -fast überall

- genau dann wenn also  $P = Q$

2. Fall:  $P$  nicht absolutstetig zu  $Q$  klar

□

**Beispiel 4.40.** Beziehung zwischen relativer Entropie und Fisher Information

sei  $\Theta$  intervall in  $\mathbb{R}$

-  $P = P_{\delta} = f_{\delta}\mu$

-  $Q = P_0 f_0 \mu$

- alles hänge 2x differenzierbar von  $\delta$  ab

Beh:

$$H(P_\delta; P_0)|_{\delta=0}'' = I(0)$$

- Fisher Information ist die zweite Ableitung der relativen Entropie bei 0

$$- \sigma_\delta = \frac{f_\delta}{f_0}$$

$$\begin{aligned} (f_\delta \ln(\frac{f_\delta}{f_0}))|_{\delta=0}'' &= (f_\delta' \ln(\frac{f_\delta}{f_0}) + f_\delta \frac{f_\delta'}{f_0} \frac{f_0}{f_\delta})|_{\delta=0}' \\ &= (f_\delta' (\ln(\frac{f_\delta}{f_0}) + 1))|_{\delta=0}' \\ &= (f_\delta'' (\ln(\frac{f_\delta}{f_0}) + 1) + \frac{f_\delta'^2}{f_0} \frac{f_0}{f_\delta})|_{\delta=0} \\ &= f_\delta'' + \frac{f_0'^2}{f_0} = f_0'' + \mathcal{L}_0'^2 f_0 \end{aligned}$$

$$(\partial_\delta^2)|_{\delta=0} H(P_\delta, P_0) = (\int_\Omega f_\delta \ln(\frac{f_\delta}{f_0}) d\mu)|_{\delta=0}'' = \int_\Omega (f_\delta'' + \mathcal{L}_0'^2 f_0) d\mu = I(0)$$

benutzen hier  $\int_\Omega f_\delta'' d\mu = (\int_\Omega f_\delta d\mu)'' = (1)'' = 0$

□

betrachten Model  $(\Omega_0, \mathcal{B}_0, (P_{0,\vartheta})_{\vartheta \in \Theta})$  für einfachen Versuch

$g : \Theta \rightarrow \mathbb{R}$  - zu schätzende Größe

Statistik  $Y : (\Omega_0, \mathcal{B}_0) \rightarrow (\Omega', \mathcal{B}')$

-  $\mu$  Maß

- nehmen an, daß  $Y_* P_{0,\vartheta} = f_\vartheta \mu$

bilden Model für mehrfache Wiederholung desselben Versuchs

- $(\Omega, \mathcal{B}, (P_\vartheta)_{\vartheta \in \Theta}) := \prod_{\mathbb{N}}(\Omega_0, \mathcal{B}_0, (P_{0,\vartheta})_{\vartheta \in \Theta})$
- Statistik für  $n$ -fache Wiederholung:
  - $(\Omega'_n, \mathcal{B}_n, \mu_n) := \prod_{i=0}^{n-1}(\Omega', \mathcal{B}', \mu)$
  - $X_n = (Y_0, \dots, Y_{n-1}) : \Omega \rightarrow \Omega'_n$
  - mit  $Y_i := Y \circ \text{pr}_i$
  - dann  $X_{n,*}P_\vartheta = f_{n,\vartheta}\mu_n$  mit  $f_{n,\vartheta} = \prod_{i=0}^{n-1} \text{pr}_i^* f_\vartheta$
- $\hat{g}_n : \Omega'_n \rightarrow \mathbb{R}$  - Schätzer für  $g$

**Satz 4.41.** *Annahmen:*

1.  $\vartheta \mapsto Y_*P_{0,\vartheta}$  ist injektiv
2.  $g(\Theta)$  ist offen in  $\mathbb{R}$
3.  $\hat{g}_n$  ist MLH-Schätzer für  $g$ .
4. *Monotone Parametrisierung:* Für  $\vartheta, \vartheta', \vartheta''$  in  $\Theta$  und  $x$  in  $\Omega'_n$  folgt aus  $g(\vartheta) < g(\vartheta'') < \hat{g}_n(x)$  daß  $f_{n,\vartheta}(x) < f_{n,\vartheta''}(x)$  bzw. aus  $\hat{g}_n(x) < g(\vartheta') < g(\vartheta)$  daß  $f_{n,\vartheta}(x) < f_{n,\vartheta'}(x)$  gilt.

Dann ist  $(\hat{g}_n)_{n \in \mathbb{N}}$  konsistent.

Die Annahme 4 ist eine Verschärfung der Bedingung daß  $\hat{g}_n$  MLH-Schätzer ist:

- $f_{n,\vartheta}(x)$  hat Maximum (bez.  $\vartheta$ , festes  $x$ ), wenn  $\hat{g}(x) = g(\vartheta)$
- die Bedingung besagt, daß wenn nun  $\vartheta$  variiert so daß  $g(\vartheta)$  oberhalb von  $\hat{g}(x)$  weiter wächst, oder unterhalb von  $\hat{g}(x)$  weiter fällt, dann fällt in jedem Fall auch  $f_{n,\vartheta}(x)$
- spezialfall:  $g = \text{id}$
- monotone Parametrierung besagt:  $f_{n,\vartheta}(x)$  hat Maximum in  $\vartheta = \hat{g}(x)$ , wächst unterhalb von  $\hat{g}(x)$  streng monoton und fällt oberhalb von  $\hat{g}(x)$  streng monoton

*Beweis.*

fixiere  $\vartheta$

- fixiere  $\epsilon > 0$

- wegen 2 kann  $\vartheta', \vartheta''$  in  $\Theta$  wählen so daß  $g(\vartheta) - \epsilon \leq g(\vartheta') < g(\vartheta) < g(\vartheta'') \leq g(\vartheta) + \epsilon$

- wegen 1 finde  $\delta > 0$  so daß

-  $\delta < H(Y_*P_{0,\vartheta}; Y_*P_{0,\vartheta'})$  und  $\delta < H(Y_*P_{0,\vartheta}; Y_*P_{0,\vartheta''})$ .

es gelten die folgenden Inklusionen von Teilmengen von  $\Omega'_n$ :

$$\{f_{n,\vartheta'} \leq f_{n,\vartheta} \geq f_{n,\vartheta''}\} \subseteq \{g(\vartheta') \leq \hat{g}_n \leq g(\vartheta'')\} \subseteq \{g(\vartheta) - \epsilon \leq \hat{g}_n \leq g(\vartheta) + \epsilon\}$$

- begründen erste Inklusion:

- wenn  $x \notin \{g(\vartheta') \leq \hat{g}_n \leq g(\vartheta'')\}$  dann:

—  $\hat{g}_n(x) < g(\vartheta') < g(\vartheta)$  und damit  $f_{n,\vartheta}(x) < f_{n,\vartheta'}(x)$

- oder

—  $g(\vartheta) < g(\vartheta'') < \hat{g}_n(x)$  und damit  $f_{n,\vartheta}(x) < f_{n,\vartheta''}(x)$

es gilt:

$$\left\{ \frac{1}{n} \ln \left( \frac{f_{n,\vartheta}}{f_{n,\vartheta'}} \right) > \delta \right\} \cap \left\{ \frac{1}{n} \ln \left( \frac{f_{n,\vartheta}}{f_{n,\vartheta''}} \right) > \delta \right\} \subset \{f_{n,\vartheta'}(x) \leq f_{n,\vartheta} \geq f_{n,\vartheta''}(x)\}$$

Annahme:  $H(Y_*P_{0,\vartheta}; Y_*P_{0,\vartheta'}) \neq \infty$

-  $\sigma := \frac{f_{\vartheta}}{f_{\vartheta'}}$

-  $\ln(\sigma(Y))$  ist integrierbar für  $P_{0,\vartheta}$

-  $(\ln(\sigma(Y_i)))_{i \in \mathbb{N}}$  ist unabhängige Folge von integrierbaren ZV

- erfüllt das schwache Gesetz der großen Zahl ( $L^1$ -Version, Satz 3.20):

$$\begin{aligned}
\frac{1}{n} \ln\left(\frac{f_{n,\vartheta}(X_n)}{f_{n,\vartheta'}(X_n)}\right) &= \frac{1}{n} \ln\left(\prod_{i=0}^{n-1} \sigma(Y_i)\right) \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \ln(\sigma(Y_i)) \\
&\xrightarrow{n \rightarrow \infty} \mathbb{E}_{\vartheta}(\ln \sigma(Y)) \quad (\text{in Wahrscheinlichkeit}) \\
&= H(Y_*P_{0,\vartheta}; Y_*P_{0,\vartheta'}) \\
&> \delta
\end{aligned}$$

analog

$$\frac{1}{n} \ln\left(\frac{f_{n,\vartheta}(X_n)}{f_{n,\vartheta'}(X_n)}\right) \xrightarrow{n \rightarrow \infty} H(Y_*P_{0,\vartheta}; Y_*P_{0,\vartheta'}) > \delta \quad (\text{in Wahrscheinlichkeit})$$

\*\*\*

- also  $\lim_{n \rightarrow \infty} P_{\vartheta}(\{\frac{1}{n} \ln(\frac{f_{n,\vartheta}(X_n)}{f_{n,\vartheta'}(X_n)}) > \delta\} \cap \{\frac{1}{n} \ln(\frac{f_{n,\vartheta'}(X_n)}{f_{n,\vartheta}(X_n)}) > \delta\}) = 1$
- also  $\lim_{n \rightarrow \infty} P(\{|\hat{g}_n(X) - g(\vartheta)| \leq \epsilon\}) = 1$

Annahme:  $H(Y_*P_{0,\vartheta}; Y_*P_{0,\vartheta'}) = \infty$ , aber  $Y_*P_{0,\vartheta}$  absolutstetig zu  $Y_*P_{0,\vartheta'}$

- $\sigma$  definiert
- für  $c > 0$  ist  $h_c := \ln(\sigma \wedge c)$   $Y_*P_{0,\vartheta}$ -integrierbar
- $\lim_{c \uparrow \infty} E_{Y_*P_{0,\vartheta}}(h_c) = \infty$  (Monotone Konvergenz)
- sei  $c_0$  so daß  $E_{Y_*P_{0,\vartheta}}(h_{c_0}) > \delta$
- $(h_{c_0}(Y_i))_{i \in \mathbb{N}}$  erfüllt das schwache Gesetz der großen Zahl ( $L^1$ -Version, Satz 3.20):
- $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} h_{c_0}(Y_i) \rightarrow E_{Y_*P_{0,\vartheta}}(h_{c_0}) > \delta$  (in Wahrscheinlichkeit)

$$P_{\vartheta}(\{\frac{1}{n} \ln(\frac{f_{n,\vartheta}(X_n)}{f_{n,\vartheta'}(X_n)}) > \delta\}) \geq P_{\vartheta}(\{\frac{1}{n} \sum_{i=0}^{n-1} h_{c_0}(Y_i) > \delta\}) \xrightarrow{n \rightarrow \infty} 1$$

- schliessen analog wie oben weiter bei \*\*\*



Annahme:  $H(Y_*P_{0,\vartheta}; Y_*P_{0,\vartheta'}) = \infty$ , aber  $Y_*P_{0,\vartheta}$  nicht absolutstetig zu  $Y_*P_{0,\vartheta'}$

- dann  $(Y_*P_{0,\vartheta})(\{f_{0,\vartheta'} = 0\}) =: a > 0$  (wegen nicht-Absolutstetigkeit)

- es gilt  $f_{0,\vartheta} > 0$   $Y_*P_{0,\vartheta}$ -fast überall

- also  $P_{\vartheta}(\{f_{n,\vartheta} \neq 0\}) = 1$

- auf  $\{f_{n,\vartheta} \neq 0\}$  gilt  $\ln\left(\frac{f_{n,\vartheta}(X_n)}{f_{n,\vartheta'}(X_n)}\right) \neq \infty$  wenn  $f_{0,\vartheta'}(Y_i) \neq 0$  für alle  $i$

-  $P_{\vartheta}(\{\frac{1}{n} \ln\left(\frac{f_{n,\vartheta}(X_n)}{f_{n,\vartheta'}(X_n)}\right) \neq \infty\}) = (1 - a)^n$

- also  $P_{\vartheta}(\{\frac{1}{n} \ln\left(\frac{f_{n,\vartheta}(X_n)}{f_{n,\vartheta'}(X_n)}\right) = \infty\}) = 1 - (1 - a)^n \xrightarrow{n \rightarrow \infty} 1$

- also auch  $P_{\vartheta}(\{\frac{1}{n} \ln\left(\frac{f_{n,\vartheta}(X_n)}{f_{n,\vartheta'}(X_n)}\right) > \delta\}) \xrightarrow{n \rightarrow \infty} 1$

- schließen analog wie oben weiter bei \*\*\*

□

## 4.7 Lebensdauer

Wollen Lebensdauer eines Produktes schätzen

Model für ein Exemplar:  $(\mathbb{R}, Borel, P_T)$  mit

-  $Y = \text{id}$  - Lebensdauer

-  $Q_T = \chi_{[0,\infty)}(x)Te^{-Tx}\text{Leb}$  - Exponentialverteilung

statistisches Modell für abzählbar viele Exemplare:

-  $(\Omega, \mathcal{B}, (P_T)_{T \in \Theta})$  mit

-  $(\Omega, \mathcal{B}, P_T) = \prod_{\mathbb{N}}(\mathbb{R}, Borel, Q_T)$

-  $\Theta = (0, \infty)$

$n$ -tes Experiment

- beobachte die Ausfallzeiten der ersten  $n$  Exemplare

-  $X_n := (\Omega, \mathcal{B}) \rightarrow (\mathbb{R}^n, Borel)$  - Projektion auf die ersten  $n$  Komponenten  $(Y_0, \dots, Y_{n-1})$

zu schätzende Größe:

-  $g : \Theta \rightarrow \mathbb{R}$

-  $g(T) := T^{-1}$  (mittlere Lebensdauer)

$$X_{n,*}P_T = T^n \prod_{i=0}^{n-1} \chi_{[0,\infty)}(x_i) e^{-x_i T} \text{Leb}_{\mathbb{R}^n} = e^{(\frac{1}{n} \sum_{i=0}^{n-1} x_i)(-nT) + n \ln(T)} \prod_{i=0}^{n-1} \chi_{[0,\infty)}(x_i) \text{Leb}_{\mathbb{R}^n}$$

- hat Struktur einer exponentiellen Familie bezüglich  $\mu := \prod_{i=0}^{n-1} \chi_{[0,\infty)}(x_i) \text{Leb}_{\mathbb{R}^n}$

$$- \hat{g}_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} x_i,$$

$$- s(T) = -nT,$$

$$- b(T) = -n \ln(T)$$

$$\begin{aligned} E_T(\hat{g}_n(X_n)) &= \int_{\mathbb{R}^n} \left( \frac{1}{n} \sum_{i=0}^{n-1} x_i \right) dX_{n,*}P_T \\ &= \int_{\mathbb{R}^n} \left( \frac{1}{n} \sum_{i=0}^{n-1} x_i \right) T^n \prod_{i=0}^{n-1} \chi_{[0,\infty)}(x_i) e^{-x_i T} \text{Leb}_{\mathbb{R}^n} \\ &= \int_0^\infty x T e^{-Tx} dx \\ &= (-x e^{-Tx})_0^\infty + \int_0^\infty e^{-Tx} dx \\ &= T^{-1} \\ &= g(T) \end{aligned}$$

also ist  $\hat{g}_n$  nach Lemma 4.37 bester erwartungstreuer Schätzer von  $g$

log-Likelihood-Funktion (bez.  $\mu$ ) ist  $f_{n,T}(x) = n \ln(T) - nT \hat{g}_n(x)$

$$- 0 = \partial_T f_{n,T}(x) = \frac{n}{T} - n \hat{g}_n(x)$$

- Nullstelle bei  $T = \hat{g}_n(x)$

- dort ist ein Maximum

- also ist  $\hat{g}_n$  ein MLS

also ist nach Satz 4.41 die Familie  $(\hat{g}_n)_{n \in \mathbb{N}}$  konsistent

- Voraussetzungen sind einfach zu prüfen
- alternativ kann man direkt mit dem Gesetz der großen Zahl schließen

## 4.8 Tests

betrachten spezielles Schätzproblem

$(\Omega, \mathcal{B}, (P_{\vartheta})_{\vartheta \in \Theta})$  - statistisches Modell

$X : (\Omega, \mathcal{B}) \rightarrow (\Omega', \mathcal{B}')$  - Statistik

$\Theta' := \{0, 1\}$  - der Zweipunktraum

-  $g : \Theta \rightarrow \Theta'$

-  $g^{-1}(0) =: H \subseteq \Theta$  - die Hypothese

— nehmen immer an:  $H \neq \emptyset$

-  $g^{-1}(1) =: K \subseteq \Theta$  - die Alternative

—  $H \cup K = \Theta, H \cap K = \emptyset$

-  $\hat{g} : \Omega' \rightarrow \Theta'$  - Schätzer für  $g$

-  $\hat{g}^{-1}(0) =: A$  - Annahmebereich

-  $\hat{g}^{-1}(1) =: V$  - Verwerfbereich

-  $A \cup V = \Omega', A \cap V = \emptyset$

Situation:

- haben statistisches Modell, kennen  $\vartheta$  nicht

- zerlegen  $\Theta$  in  $H$  und  $K$

- wollen schätzen ob  $\vartheta \in H$

- ob die Hypothese erfüllt ist

- beobachten dazu  $X$

- wenn  $X \in A$ , dann entscheiden wir, daß die Hypothese unterstützt wird

- wenn  $X \in V$ , dann verwerfen wir die Hypothese

**Definition 4.42.** Die Wahl von  $\hat{g}$  in dieser Situation heißt ein Test.

Fehler:

- Fehler 1. Art:  $\vartheta \in H$  und  $X \in V$
- Hypothese ist erfüllt, wird aber aufgrund der Beobachtung verworfen
  
- Fehler 2. Art:  $\vartheta \in K$  und  $X \in A$
- Hypothese ist nicht erfüllt, wird aber aufgrund der Beobachtung angenommen

Güte und Macht

**Definition 4.43.** Die Gütefunktion  $\beta : \Theta \rightarrow [0, 1]$  des Tests wird durch  $\beta(\vartheta) := P_{\vartheta}(\{X \in V\})$  definiert.

- es gilt  $\beta(\vartheta) = \mathbb{E}_{\vartheta}(\hat{g})$  (wenn man  $\hat{g}$  als  $\mathbb{R}$ -wertig betrachtet)

**Definition 4.44.** Das Niveau des Tests ist durch

$$\alpha = \sup_{\vartheta \in H} \beta(\vartheta)$$

gegeben

- $\alpha \leq 1$  wegen  $H \neq \emptyset$
- Wenn der Test das Niveau  $\alpha$  hat, dann ist die Wahrscheinlichkeit eines Fehlers 1. Art durch  $\alpha$  beschränkt.

**Definition 4.45.** Die Mächte des Tests ist die Funktion  $K \ni \vartheta \mapsto \beta(\vartheta) \in [0, 1]$ .

- Für  $\vartheta \in K$  ist  $1 - \beta(\vartheta)$  die Wahrscheinlichkeit eines Fehlers zweiter Art.

das allgemeine Problem ist:

- $\alpha$  gegeben
- Konstruiere Test zum Niveau  $\alpha$

– maximiere unter diesen die Macht des Tests (d.h. minimiere die Wahrscheinlichkeit eines Fehlers 2. Art)

**Bemerkung 4.46.**

das Infimum über  $\vartheta$  in  $K$  als Maß der Macht zu betrachten ist meist keine gute Idee:

- sei  $\Theta$  topologisch, zusammenhängend (z.B. ein Intervall) und  $\beta$  stetig
- dann ist  $\bar{H} \cap \bar{K} \neq \emptyset$
- dann ist  $\inf_{\vartheta \in K} \beta(\vartheta) \leq \inf_{\vartheta \in \bar{H} \cap \bar{K}} \beta(\vartheta) \leq \sup_{\vartheta \in H} \beta(\vartheta) = \alpha$
- wenn man das Niveau  $\alpha$  klein festlegt, dann kann die Macht nicht für alle  $\vartheta$  in  $K$  groß sein □

**Beispiel 4.47.** Münzwurf

im Spielkasino wird “der Münzwurf” gespielt

- das Casino möchte testen ob die Münze fair ist
- unfaire Münze als fair erkennen wäre fatal, soll mit großer Sicherheit vermieden werden:
- faire Münze als unfair zu verwerfen ist nur Verschwendung der Münze, aber nicht ganz so schlimm

Model:

$$\Omega = \prod_{\mathbb{N}} \{z, w\}$$

- $\mathcal{B}$  -  $\sigma$ -Algebra der Zylindermengen
- $\Theta = [0, 1], p \in \Theta$
- $P_p$  - Maß für Münze mit  $z$ -Wahrscheinlichkeit  $p$
- $(\Omega, \mathcal{B}, (P_p)_{p \in [0,1]})$

Fehler 1. Art: Münze wird als fair erkannt, ist es aber nicht

- Münze als fair erkannt = Hypothese verworfen
- also Hypothese = Münze ist unfair
- $H := \{p \neq 1/2\}$

— Alternative:  $K = \{p = 1/2\}$

$$g(p) := \begin{cases} 0 & p \neq 1/2 \\ 1 & p = 1/2 \end{cases}$$

Experiment: werfe Münze  $n$  mal und zähle Anzahl der Zahlen

-  $\Omega'_n = \{0, \dots, n\}$

-  $X_n : \Omega \rightarrow \Omega'_n$

-  $X_{n,*}P_p$  hat Dichte  $f_{n,p}(x) = \binom{n}{x}p^x(1-p)^{n-x}$  bezüglich dem Zählmaß

geben Niveau  $\alpha$  vor (klein)

- such Tests  $\hat{g} : \Omega'_n \rightarrow [0, 1]$  vom Niveau  $\alpha$

Triviale Wahl:

-  $\hat{g} = 1$  (also  $V = \emptyset$  verwerfe nie)

- dann hat der Test die Gütefunktion  $\beta = 0$ :

- Test hat Niveau 0 (also auch  $\leq \alpha$ )

- Test hat aber keine Macht:  $1 - \beta(p) = 1$

- würde alle Münzen wegwerfen

Ansatz für Verwerfbereich:  $V_k := [n/2 - k, n/2 + k]$

- Gütefunktion  $\beta_k(p) := P_p(\{X_n \in V_k\})$

- wenn  $p \neq 1/2$ , dann soll  $\beta_k(p) \leq \alpha$  gelten

-  $\beta_k(p)$  wächst mit  $p \uparrow 1/2$  und fällt mit  $p \downarrow 1/2$

### Plot

- wegen Stetigkeit ist unsere Bedingung also  $\beta_k(1/2) \leq \alpha$

- suchen also  $k$  derart, daß  $\beta_k(1/2) \leq \alpha$

- suchen  $k$  maximal mit dieser Eigenschaft

- dann wird Fehler zweiter Art minimal

— Gütefunktion wächst mit  $k$

- aus Symmetriegründen: wähle  $k \leq n/2$  maximal so daß  $P_{1/2}(\{X_n \leq n/2 - k - 1\}) \geq (1 - \alpha)/2$

-  $q = n/2 - k - 1$  ist minimal mit  $P_{1/2}(\{X_n \leq q\}) \geq (1 - \alpha)/2$

- also ist  $q$  das  $(1 - \alpha)/2$ -Quantil

$n = 100, 1000, 10000, \alpha = 0.01$ :

□

**Beispiel 4.48.** noch einmal die Münze

Situation: Jemand behaupt, daß die Münze fair ist.

- Ich glaube das nicht und will ihn widerlegen.

- Es wäre aber peinlich, dabei einen Fehler zu machen.

- Will möglichst vermeiden, daß Münze als unfair erkannt wird, aber doch fair ist.

benutzen dasselbe Modell wie in Beispiel 4.47

- Fehler erster Art: faire Münze als unfair erkannt

- Münze als unfair erkannt = Hypothese verworfen

also: Hypothese = Münze ist fair

-  $H = \{p = 1/2\}$

-  $K = \{p \neq 1/2\}$

machen dasselbe Experiment wie in Beispiel 4.47

- Ansatz für den Verwerfbereich:  $V_k := [0, n] \setminus [n/2 - k, n/2 + k]$

-  $k$  noch zu bestimmen

- Gütefunktion  $\beta_k(p) := P_p(\{X_n \in V_k\}) = 2P_p(X_n < n/2 - k)$  - aus Symmetriegründen

-  $\sup_{p \in H} \beta_k(p) = \beta_k(1/2)$  ist das Niveau des Tests

-  $\beta_k(p)$  fällt mit  $k$

- suchen  $k$  minimal (maximiere die Macht) mit  $\beta_k(1/2) \leq \alpha$
  - also mit  $P_{1/2}(\{X_n < n/2 - k\}) \leq \alpha/2$
  - sei  $q_\alpha$  das  $\alpha/2$ -Quantil, also minimal mit  $P_{1/2}(\{X_n \leq q_\alpha\}) \geq \alpha/2$
  - dann bestimmt sich  $k$  aus  $n/2 - k = q_\alpha - 1$
  - $k = n/2 - q_\alpha - 1$
  - $n/2 - k = q_\alpha - 1$
- $n = 100, 1000, 10000$

□

## 4.9 Mehr Theorie

$(\Omega, \mathcal{B}, (P_\vartheta)_{\vartheta \in \Theta})$  - statistisches Modell

$X : (\Omega, \mathcal{B}) \rightarrow (\Omega', \mathcal{B}')$  - Statistik

-  $\Theta = H \sqcup K$  -Hypothese und Alternative

Randomisierter Tests

**Definition 4.49.** Ein randomisierter Test ist eine meßbare Funktion  $\hat{g} : \Omega' \rightarrow [0, 1]$ .

Anwendung: Ist das Ergebnis des Experiments  $x$  in  $\Omega$ , dann führen wir ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit  $\hat{g}(x)$  durch und lehnen die Hypothese bei Erfolg ab

- erhalten den alten Begriff, wenn  $\hat{g}$  nur Werte in  $\{0, 1\}$  hat

Gütefunktion:  $\beta(\vartheta) = \mathbb{E}_\vartheta(\hat{g}(X))$

-  $\sup_{\vartheta \in H} \beta(\vartheta) =: \alpha$  ist das Niveau des Tests

-  $K \ni \vartheta \mapsto \beta(\vartheta)$  ist die Macht des Tests

$\Psi_\alpha := \{\hat{h} : \Omega \rightarrow [0, 1] \mid \hat{h} \text{ ist Test zum Niveau } \leq \alpha\}$  - Menge der Tests mit Niveau  $\leq \alpha$

**Definition 4.50.**  $\hat{g}$  ist ein gleichmäßig bester Test zu Niveau  $\alpha$ , wenn

$$\mathbb{E}_\vartheta(\hat{g}(X)) = \sup_{\hat{h} \in \Psi_\alpha} \mathbb{E}_\vartheta(\hat{h}(X))$$



für alle  $\vartheta$  in  $A$ .

Testen einfacher Hypothesen (Neyman-Pearson)

Annahmen:

- $\Theta = \{h, a\}$
- $H = \{h\}$  und  $K = \{a\}$

es gibt Maß  $\mu$  auf  $(\Omega', \mathcal{B}')$  mit:

- $X_*P_h = f_h\mu$
- $X_*P_a = f_a\mu$

geben  $k \in [0, \infty]$  und  $\gamma \in [0, 1]$  vor

$$\text{- setzen } \hat{g}(x) := \begin{cases} 1 & \frac{f_a(x)}{f_h(x)} > k \\ 0 & \frac{f_a(x)}{f_h(x)} < k \\ \gamma & \frac{f_a(x)}{f_h(x)} = k \end{cases}$$

- 0/0 tritt höchstens auf einer Nullmenge für  $X_*P_h$  und  $X_*P_a$  auf
- definieren  $\hat{g}$  dann als 0

die Wahl von  $\hat{g}$  definiert einen randomisierten Test

- $\hat{g}$  heißt auch Likelihoodquotiententest
- $E_h(\hat{g}(X)) =: \alpha$  ist das Niveau

**Lemma 4.51.**  $\hat{g}$  ist ein gleichmäßig bester Test zum Niveau  $\alpha$ .

*Beweis.*

sei  $\hat{h} \in \Psi_\alpha$

$$\text{- z.z.: } \mathbb{E}_a(\hat{h}(X)) \leq \mathbb{E}_a(\hat{g}(X))$$

Beh.:  $(\hat{g}(x) - \hat{h}(x))(f_a(x) - kf_h(x)) \geq 0$  für alle  $x$  in  $\Omega'$

1.  $f_a(x) - kf_h(x) > 0 \Rightarrow \hat{g}(x) = 1$  und also  $\hat{g}(x) - \hat{h}(x) \geq 0$

2.  $f_a(x) - kf_h(x) < 0 \Rightarrow \hat{g}(x) = 0$  und also  $\hat{g}(x) - \hat{h}(x) \leq 0$
3.  $f_a(x) - kf_h(x) = 0$  klar

aus Ungleichung folgt:

$$\int_{\Omega'} (\hat{g} - \hat{h}) f_a d\mu \geq k \int_{\Omega'} (\hat{g} - \hat{h}) f_h d\mu$$

$$\mathbb{E}_a(\hat{g}(X)) - \mathbb{E}_a(\hat{h}(X)) \geq k(\mathbb{E}_h(\hat{g}(X)) - \mathbb{E}_h(\hat{h}(X))) \geq k(\alpha - \alpha) \geq 0$$

- benutzen hier  $\mathbb{E}_h(\hat{g}(X)) = \alpha$  und  $\mathbb{E}_h(\hat{h}(X)) \leq \alpha$

$\mathbb{E}_a(\hat{g}(X)) - \mathbb{E}_a(\hat{h}(X)) \geq 0$  ist Behauptung

□

das folgende Lemma zeigt, daß es in der obigen Situation zu jedem vorgegeben Niveau  $\alpha$  einen besten (randomisierten) Test gibt

**Lemma 4.52.** *Zu jedem  $\alpha \in (0, 1)$  existieren  $k \in [0, \infty)$  und  $\gamma \in [0, 1]$  so daß  $\hat{g}$  das Niveau  $\alpha$  hat.*

*Beweis.*

betrachte  $T : \Omega' \rightarrow [0, \infty]$

$$T(x) := \begin{cases} \frac{f_a(x)}{f_h(x)} & \text{nicht } 0/0 \\ 0 & \text{sonst} \end{cases}$$

- es gilt:  $\mathbb{E}_h(\hat{g}(X)) = P_h(T(X) > k) + \gamma P_h(T(X) = k)$

setze  $k := \inf\{y \geq 0 \mid P_h(\{T(X) > y\}) \leq \alpha\}$

- müssen zeigen, daß die Menge der  $y$  nicht leer ist (weil sonst  $-\infty$  herauskommt)

-  $\mathbb{E}_h(T(X)) = \mathbb{E}_a(1) = 1 < \infty$

- also  $P_h(\{T(X) = \infty\}) = 0$

- also  $\lim_{y \rightarrow \infty} P_h(\{T(X) \geq y\}) = 0$  ( $\sigma$ -Additivität von  $P_h$ )

1. Fall:  $P_h(\{T(X) > k\}) = \alpha$

- setzen  $\gamma = 0$

2. Fall:  $P_h(\{T(X) > k\}) < \alpha$

- setzen  $\gamma := \frac{\alpha - P_h(\{T(X) > k\})}{P_h(\{T(X) = k\})}$

□

**Beispiel 4.53.** empfangen ein verauschtetes Signal oder reines Rauschen

- mehrere Messungen  $Y_1, \dots, Y_n$

- alle  $\text{Norm}(\mu, 1)$ -verteilt

dabei:

-  $\mu = 0$  - Rauschen

-  $\mu = 1$  - Signal

Wollen entscheiden, ob Signal vorliegt und vermeiden, fälschlicherweise ein Signal zu erkennen

- Hypothese:  $\mu = 0$

- Fehler erster Art: Hypothese verworfen (also Signal erkannt), obwohl keins vorliegt

- benutzen identische Statistik

-  $\Theta = \{0, 1\}$  - Wert von  $\mu$

- Dichten bezüglich  $\text{Leb}_{\mathbb{R}^n}$

$$- f_a(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i - 1)^2}{2}}$$

$$- f_h(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n x_i^2}{2}}$$

$$- \frac{f_a(x)}{f_h(x)} = e^{\sum_{i=1}^n x_i - n/2}$$

- hängt nur von  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  ab

$$- \frac{f_a(x)}{f_h(x)} = e^{n(\bar{x} - 1/2)}$$

- wächst monoton mit  $\bar{x}$
- suchen  $k$  und setzen an:

$$\hat{g}(x) = \hat{g}(\bar{x}) = \begin{cases} 1 & \bar{x} \geq k \\ 0 & \bar{x} < k \end{cases}$$

- brauchen keine Randomisierung, da  $\bar{x} = k$  nur auf Nullmenge eintritt
- suchen  $k$  so daß  $P_0(\{\bar{x} \geq k\}) = \alpha$ .
- unter der Hypothese ist  $\bar{x}$  ist Norm( $0, 1/\sqrt{n}$ ) verteilt
- $k$  ist also das  $1 - \alpha$ -Quantil für Norm( $0, 1/\sqrt{n}$ )
- $k\sqrt{n}$  ist das  $1 - \alpha$ -Quantil  $q_{1-\alpha}$  für Norm( $0, 1$ )
- also  $k = \sqrt{n}^{-1}q_{1-\alpha}$
- setzen also

$$\hat{g}(\bar{x}) := \begin{cases} 1 & \bar{x} \geq \sqrt{n}^{-1}q_{1-\alpha} \\ 0 & \bar{x} < \sqrt{n}^{-1}q_{1-\alpha} \end{cases}$$

- das ist der gleichmäßig beste Test zum Niveau  $\alpha$

□