



Prof. Dr. Dr. Stefan Schäffler
nihil mathematicis inlustrius

MaPr

Institut für Mathematik und
Datenverarbeitung
Fakultät für Elektrotechnik
und Informationstechnik
85577 Neubiberg
Tel.: 089/6004 2633
Fax: 089/6004 2615
stefan.schaeffler@unibw-muenchen.de

Mathematik 4: Stochastik

Herbst 2004

Kapitel 11

Maß und Integration

11.1 Maße

Wir betrachten eine beliebige nichtleere Basismenge Ω . Eine Menge $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ wird als Mengensystem (über Ω) bezeichnet, wobei $\mathcal{P}$ die Potenzmenge (Menge aller Teilmengen) von Ω darstellt. Mit $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ wird eine Erweiterung der Menge aller reellen Zahlen definiert. Die algebraische Struktur von $\mathbb{R}$ wird folgendermaßen auf $\bar{\mathbb{R}}$ erweitert: Für alle $a \in \mathbb{R}$ gilt:

$$a + (\pm\infty) = (\pm\infty) + a = (\pm\infty) + (\pm\infty) = (\pm\infty), \quad +\infty - (-\infty) = +\infty,$$

$$a \cdot (\pm\infty) = (\pm\infty) \cdot a = \begin{cases} (\pm\infty), & \text{für } a > 0, \\ 0, & \text{für } a = 0, \\ (\mp\infty), & \text{für } a < 0, \end{cases}$$

$$(\pm\infty) \cdot (\pm\infty) = +\infty, \quad (\pm\infty) \cdot (\mp\infty) = -\infty, \quad \frac{a}{\pm\infty} = 0.$$

Somit ist $\bar{\mathbb{R}}$ kein Körper. Die Vorzeichen bei $\pm\infty$ dürfen bei den obigen Formeln nicht kombiniert werden, denn der Ausdruck $+\infty - (+\infty)$ ist nicht definiert. Die Bedeutung der Festlegung $0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = 0$ wird später deutlich. Vorsicht ist allerdings bei den Grenzwertsätzen geboten:

$$\lim_{x \rightarrow +\infty} \left(x \cdot \frac{1}{x} \right) \neq (+\infty) \cdot 0 = 0.$$

Ergänzt man die Ordnungsstruktur von $\mathbb{R}$ durch $-\infty < a, a < +\infty$ für alle $a \in \mathbb{R}$ und $-\infty < +\infty$, so ist $(\bar{\mathbb{R}}, \leq)$ eine geordnete Menge. Wir vereinbaren unter Verzicht auf die entsprechenden Grenzwertsätze, dass die Folge $\{n\}$, $n \in \mathbb{N}$, den Grenzwert $+\infty \in \bar{\mathbb{R}}$ besitzt. Für „ $+\infty$ “ schreiben wir oft „ ∞ “.

In Analogie zur Berechnung von Volumina in der Geometrie versucht man, Mengen aus einem Mengensystem $\mathcal{F}$ über Ω Maße (Volumina) zuzuordnen. Zu diesem Zweck zeichnet man spezielle Funktionen aus.

Definition 11.1 ((σ -endliches) Maß) Sei $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, $\emptyset \in \mathcal{F}$. Eine Funktion $\mu: \mathcal{F} \rightarrow \mathbb{R}$ heißt Maß auf $\mathcal{F}$, falls die folgenden Bedingungen erfüllt sind:

(M1) $\mu(A) \geq 0$ für alle $A \in \mathcal{F}$,

(M2) $\mu(\emptyset) = 0$,

(M3) Für jede Folge $\{A_i\}, i \in \mathbb{N}$, paarweise disjunkter Mengen mit $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, und $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ gilt:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad (\sigma\text{-Additivität}).$$

Besitzen für eine Folge $\{B_i\}, i \in \mathbb{N}$, mit $B_i \subseteq B_{i+1}$, $B_i \in \mathcal{F}$ und $\bigcup_{i=1}^{\infty} B_i = \Omega$ alle Mengen B_i ein endliches Maß, so wird μ als σ -endlich bezeichnet.

Es wäre naheliegend, Maße auf der Potenzmenge von Ω zu betrachten. Allerdings ist diese Vorgehensweise problematisch, da es zum Beispiel nicht möglich ist, ein Maß μ auf der Potenzmenge des $\mathbb{R}^3$ mit $\mu([0, 1]^3) = 1$ zu finden, das nur von der Struktur der jeweiligen Menge abhängt und nicht von Drehungen oder Verschiebungen. Daher hat man sich im allgemeinen mit speziellen Mengensystemen über Ω (Teilmengen der Potenzmenge) zu begnügen. Dies führt auf den Begriff der σ -Algebra.

Definition 11.2 (σ -Algebra) Ein Mengensystem $\mathcal{S} \subseteq \mathcal{P}(\Omega)$ heißt σ -Algebra über Ω , falls die folgenden Axiome erfüllt sind:

(S1) $\Omega \in \mathcal{S}$,

(S2) Aus $A \in \mathcal{S}$ folgt $A^c := \Omega \setminus A \in \mathcal{S}$,

(S3) Aus $A_i \in \mathcal{S}, i \in \mathbb{N}$, folgt $\bigcup_{i=1}^{\infty} A_i \in \mathcal{S}$.

Die folgende Eigenschaft von σ -Algebren ist wichtig.

Satz und Definition 11.3 (Durchschnittsstabilität von σ -Algebren) Sei I eine beliebige nichtleere Menge und $\mathcal{S}_i$ für jedes $i \in I$ eine σ -Algebra über Ω , so ist auch $\bigcap_{i \in I} \mathcal{S}_i$ eine σ -Algebra über Ω . Diese Eigenschaft wird Durchschnittsstabilität von σ -Algebren genannt.

Wir können also von erzeugten σ -Algebren sprechen.

Definition 11.4 (erzeugte σ -Algebra) Sei $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ und sei Σ die Menge aller σ -Algebren über Ω , die $\mathcal{F}$ enthalten, dann wird die σ -Algebra $\sigma(\mathcal{F}) := \bigcap_{\mathcal{S} \in \Sigma} \mathcal{S}$ als die von $\mathcal{F}$ erzeugte σ -Algebra bezeichnet.

Für $\Omega = \mathbb{R}^n$, $n \in \mathbb{N}$, betrachten wir die σ -Algebra

$$\mathcal{B}^n = \sigma(\{B \subseteq \mathbb{R}^n, B \text{ offen}\}).$$

$\mathcal{B}^n$ wird als Borelsche σ -Algebra bezeichnet. Alle für die Praxis wichtigen Teilmengen des $\mathbb{R}^n$ (neben den offenen auch alle abgeschlossenen und kompakten Teilmengen) sind in $\mathcal{B}^n$ enthalten. Auf dieser σ -Algebra läßt sich nun ein spezielles Maß λ festlegen, das unter allen Maßen μ auf $\mathcal{B}^n$ das einzige Maß mit $\mu([0, 1[\times \dots \times [0, 1[) = 1$ ist, deren Wert nicht von Drehungen oder Verschiebungen der Mengen abhängt. Dieses Maß heißt Lebesgue-Borel-Maß. Sei nun μ ein Maß auf einer σ -Algebra $\mathcal{S}$ über Ω , so heißt jede Menge $A \in \mathcal{S}$ mit $\mu(A) = 0$ eine μ -Nullmenge. Es ist nun naheliegend, jeder Teilmenge $B \subseteq A$ einer μ -Nullmenge ebenfalls das Maß $\mu(B) = 0$ zuzuordnen. Allerdings ist nicht gewährleistet, dass für jedes $B \subseteq A$ auch $B \in \mathcal{S}$ gilt. Das führt zum Begriff der Vervollständigung und des vollständigen Maßes.

Definition 11.5 (vollständiges Maß, Vervollständigung) Ein Maß μ auf einer σ -Algebra $\mathcal{S}$ über Ω heißt vollständig, falls jede Teilmenge einer μ -Nullmenge zu $\mathcal{S}$ gehört und damit eine μ -Nullmenge ist. Die σ -Algebra

$$\mathcal{S}_0 := \{A \cup N; A \in \mathcal{S}, N \text{ Teilmenge einer } \mu\text{-Nullmenge}\}$$

heißt μ -Vervollständigung von $\mathcal{S}$. Mit $\mu_0(A \cup N) := \mu(A)$ ist μ_0 ein vollständiges Maß auf $\mathcal{S}_0$.

Die Mengen der σ -Algebra $\mathcal{B}_0^n$ heißen Lebesgue-messbare Mengen. Das Maß λ_0^n auf $\mathcal{B}_0^n$ heißt Lebesgue-Maß. Die zugehörigen Nullmengen heißen Lebesguesche Nullmengen.

Betrachtet man eine Funktion $F: \mathbb{R} \rightarrow \mathbb{R}$ mit folgenden Eigenschaften:

- F ist monoton steigend,
- F ist stetig von links,

so ist durch

$$\mu^F([a, b[\cap \mathbb{R}) := \begin{cases} F(b) - F(a), & \text{falls } -\infty < a < b < \infty \\ \lim_{b \rightarrow \infty} F(b) - F(a), & \text{falls } -\infty < a < b = \infty \\ F(b) - \lim_{a \rightarrow -\infty} F(a), & \text{falls } -\infty = a < b < \infty \\ \lim_{b \rightarrow \infty} F(b) - \lim_{a \rightarrow -\infty} F(a), & \text{falls } -\infty = a, b = \infty \\ 0, & \text{falls } a \geq b \end{cases}$$

ein eindeutiges Maß μ^F auf $\mathcal{B}$ definiert. Dieser Sachverhalt führt zu folgender Definition.

Definition 11.6 (maßerzeugende Funktion, $\Omega = \mathbb{R}$) Eine monoton steigende Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$, die stetig von links ist, heißt maßerzeugende Funktion.

Das Maß μ^F heißt Lebesgue-Borel-Stieltjes-Maß. Das vollständige Maß μ_0^F auf der μ^F -Vervollständigung $\mathcal{B}_0^F$ von $\mathcal{B}$ heißt Lebesgue-Stieltjes-Maß. Die Mengen $A \in \mathcal{B}_0^F$ heißen Lebesgue-Stieltjes-messbar. Durch analoge Vorgehensweise lassen sich maßerzeugende Funktionen auf $\Omega = \mathbb{R}^n$ definieren. Wir wollen darauf aber nicht näher eingehen.

Ist $\mathcal{S}$ eine σ -Algebra über Ω , so bezeichnen wir im folgenden das Paar $(\Omega, \mathcal{S})$ als Messraum. Ist μ ein Maß auf $\mathcal{S}$, so heißt das Tripel $(\Omega, \mathcal{S}, \mu)$ Maßraum. Nun untersuchen wir spezielle Funktionen zwischen zwei Grundmengen $\Omega_1, \Omega_2 \neq \emptyset$.

Definition 11.7 (messbare Abbildung) Seien $(\Omega_1, \mathcal{S}_1)$ und $(\Omega_2, \mathcal{S}_2)$ zwei Messräume.

Eine Abbildung $T : \Omega_1 \rightarrow \Omega_2$ mit $T^{-1}(A') := \{x \in \Omega_1; T(x) \in A'\} \in \mathcal{S}_1$ für alle $A' \in \mathcal{S}_2$ heißt $\mathcal{S}_1$ - $\mathcal{S}_2$ -messbar.

Messbare Abbildungen spielen in der Wahrscheinlichkeitstheorie bei der Definition von Zufallsvariablen eine wichtige Rolle. Der folgende Satz zeigt, dass für den Nachweis der Messbarkeit einer Abbildung nicht immer das Urbild $T^{-1}(A')$ für alle Mengen $A' \in \mathcal{S}_2$ untersucht werden muss.

Satz 11.8 (Messbarkeit bei einer erzeugten σ -Algebra $\mathcal{S}_2$) Seien $(\Omega_1, \mathcal{S}_1)$ und $(\Omega_2, \mathcal{S}_2)$ zwei Messräume, wobei $\mathcal{S}_2 = \sigma(\mathcal{F})$ von einem Mengensystem $\mathcal{F}$ erzeugt ist. Die Abbildung $T : \Omega_1 \rightarrow \Omega_2$ ist genau dann $\mathcal{S}_1$ - $\mathcal{S}_2$ -messbar, falls $T^{-1}(A') \in \mathcal{S}_1$ für alle $A' \in \mathcal{F}$.

Sind drei Messräume $(\Omega_1, \mathcal{S}_1)$, $(\Omega_2, \mathcal{S}_2)$, $(\Omega_3, \mathcal{S}_3)$ und zwei Abbildungen

$T_1 : \Omega_1 \rightarrow \Omega_2$, T_1 $\mathcal{S}_1$ - $\mathcal{S}_2$ -messbar,

$T_2 : \Omega_2 \rightarrow \Omega_3$, T_2 $\mathcal{S}_2$ - $\mathcal{S}_3$ -messbar,

gegeben, so ist die Abbildung

$T_2 \circ T_1 : \Omega_1 \rightarrow \Omega_3$, $\omega \mapsto T_2(T_1(\omega))$, $\mathcal{S}_1$ - $\mathcal{S}_3$ -messbar.

Satz und Definition 11.9 (Bildmaß) Seien $(\Omega_1, \mathcal{S}_1, \mu_1)$ ein Maßraum, $(\Omega_2, \mathcal{S}_2)$ ein Messraum und $T : \Omega_1 \rightarrow \Omega_2$ $\mathcal{S}_1$ - $\mathcal{S}_2$ -messbar, so ist durch

$$\mu_2(A') := \mu_1(T^{-1}(A')), \quad A' \in \mathcal{S}_2,$$

ein Maß μ_2 auf $\mathcal{S}_2$ definiert.

Das Maß μ_2 wird als Bildmaß von μ_1 bezeichnet (Schreibweise: $\mu_2 = T(\mu_1)$).

11.2 Integration

Um Zufallsgrößen analysieren zu können, benötigt man einen Integralbegriff. Daher soll im folgenden kurz die Integrationstheorie für messbare Abbildungen zusammengefasst werden. Zunächst betrachten wir die Integration einer speziellen Klasse von Funktionen.

Definition 11.10 (elementare Funktion) Sei $(\Omega, \mathcal{S})$ ein Messraum. Eine $\mathcal{S}$ - $\mathcal{B}$ -messbare Funktion $e : \Omega \rightarrow \mathbb{R}$ heißt elementare Funktion, falls sie nur endlich viele verschiedene Funktionswerte annimmt.

Eine spezielle elementare Funktion ist die Indikatorfunktion

$$I_A : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{sonst} \end{cases},$$

die anzeigt, ob ω Element einer Menge $A \in \mathcal{S}$ ist. Mit Hilfe von Indikatorfunktionen lassen sich die elementaren Funktionen darstellen.

Satz 11.11 (Darstellung elementarer Funktionen) Sei $(\Omega, \mathcal{S})$ ein Maßraum. Ist $e : \Omega \rightarrow \mathbb{R}$ eine elementare Funktion, so existieren eine natürliche Zahl n , paarweise disjunkte Mengen $A_1, \dots, A_n \in \mathcal{S}$ und reelle Zahlen $\alpha_1, \dots, \alpha_n$ mit:

$$e = \sum_{i=1}^n \alpha_i I_{A_i}, \quad \sum_{i=1}^n A_i = \Omega.$$

Die eben betrachtete Darstellung von e heißt eine Normaldarstellung von e . Sind alle α_i paarweise verschieden, so spricht man von einer kürzesten Normaldarstellung von e . Kürzeste Normaldarstellungen sind eindeutig. Aus der Normaldarstellung elementarer Funktionen folgt sofort: Summe, Differenz und Produkt elementarer Funktionen sind elementare Funktionen. Für alle $c \in \mathbb{R}$ ist auch $c \cdot e$ eine elementare Funktion, wenn e eine elementare Funktion ist.

Nun betrachten wir nichtnegative elementare Funktionen auf einem Maßraum $(\Omega, \mathcal{S}, \mu)$ und definieren das (μ) -Integral dieser Funktionen.

Definition 11.12 ((μ) -Integral nichtnegativer elementarer Funktionen)

Sei $(\Omega, \mathcal{S}, \mu)$ ein Maßraum und $e : \Omega \rightarrow \mathbb{R}_0^+$, $e = \sum_{i=1}^n \alpha_i I_{A_i}$, $\alpha_i \geq 0$, $i = 1, \dots, n$, eine nichtnegative elementare Funktion in Normaldarstellung, so wird

$$\int e \, d\mu := \int_{\Omega} e \, d\mu := \sum_{i=1}^n \alpha_i \cdot \mu(A_i)$$

als (μ) -Integral von e über Ω bezeichnet.

Sei nun E die Menge aller nichtnegativen elementaren Funktionen auf $(\Omega, \mathcal{S}, \mu)$, so erhalten wir die folgenden Eigenschaften:

- $\int I_A \, d\mu = \mu(A)$ für alle $A \in \mathcal{S}$.
- $\int (\alpha e) \, d\mu = \alpha \int e \, d\mu$ für alle $e \in E$, $\alpha \in \mathbb{R}_0^+$.
- $\int (u + v) \, d\mu = \int u \, d\mu + \int v \, d\mu$ für alle $u, v \in E$.
- Ist $u(\omega) \leq v(\omega)$ für alle $\omega \in \Omega$, so ist $\int u \, d\mu \leq \int v \, d\mu$ für alle $u, v \in E$.

Wählen wir $\Omega = \mathbb{R}^n$, $\mathcal{S} = \mathcal{B}^n$, $\mu = \lambda^n$ und $f : \Omega \rightarrow \mathbb{R}_0^+$, $x \mapsto 0$, so erhalten wir

$$\int f \, d\lambda^n = \int 0 \, d\lambda^n = 0 \cdot \lambda^n(\mathbb{R}^n) = 0 \cdot \infty = 0.$$

Unsere Vereinbarung $0 \cdot \infty = 0$ erlaubt uns somit, das (λ^n) -Integral über die Nullfunktion zu berechnen.

Betrachtet man die Menge $\bar{\mathbb{R}}$ der um $\{\pm\infty\}$ erweiterten reellen Zahlen, so bildet die Menge $\bar{\mathcal{B}} := \{A \in \mathcal{P}(\bar{\mathbb{R}}); A \cap \mathbb{R} \in \mathcal{B}\}$ eine σ -Algebra über $\bar{\mathbb{R}}$. Um nun den Integralbegriff auf eine größere Klasse von Funktionen fortzusetzen, benötigen wir die folgende Definition.

Definition 11.13 (numerische Funktion) Eine auf einer nichtleeren Menge $A \subseteq \Omega$ definierte Funktion $f : A \rightarrow \bar{\mathbb{R}}$ heißt numerische Funktion.

Nun betrachten wir nichtnegative numerische Funktionen, die als Grenzwert einer Folge elementarer Funktionen gegeben sind.

Satz 11.14 (Grenzwerte spezieller Folgen elementarer Funktionen)

Seien $(\Omega, \mathcal{S})$ ein Messraum und $f : \Omega \rightarrow \bar{\mathbb{R}}_0^+$ eine nichtnegative, $\mathcal{S}$ - $\bar{\mathcal{B}}$ -messbare numerische Funktion, so gibt es eine monoton steigende Folge $\{e_n\}$, $n \in \mathbb{N}$, von nichtnegativen elementaren Funktionen $e_n : \Omega \rightarrow \bar{\mathbb{R}}_0^+$, $n \in \mathbb{N}$, die punktweise gegen f konvergiert. Wir schreiben dafür: $e_n \uparrow f$.

Beispiel: $\Omega = \mathbb{R}$, $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto e^x$.

$$e_n : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto 0 \cdot I_{]-\infty, \ln(\frac{1}{2^n})[}(x) + \sum_{k=2}^{n2^n} \frac{k-1}{2^n} I_{[\ln(\frac{k-1}{2^n}), \ln(\frac{k}{2^n})[}(x) + \\ + n I_{[\ln(n), \infty[}(x).$$

Nach diesen Vorbereitungen sind wir in der Lage, die (μ) -Integration auf eine spezielle Klasse von Funktionen in naheliegender Weise fortzusetzen.

Definition 11.15 ((μ) -Integrale) Seien $(\Omega, \mathcal{S}, \mu)$ ein Maßraum und $f : \Omega \rightarrow \bar{\mathbb{R}}_0^+$ eine $\mathcal{S}$ - $\bar{\mathcal{B}}$ -messbare, nichtnegative numerische Funktion. Sei ferner $\{e_n\}$, $n \in \mathbb{N}$, eine monoton steigende Folge nichtnegativer elementarer Funktionen $e_n : \Omega \rightarrow \bar{\mathbb{R}}_0^+$, $n \in \mathbb{N}$, mit $e_n \uparrow f$, so definieren wir durch

$$\int f d\mu := \int_{\Omega} f d\mu := \lim_{n \rightarrow \infty} \int e_n d\mu$$

das (μ) -Integral von f über Ω .

Wir werden nun in einem letzten Schritt die Klasse der integrierbaren Funktionen erweitern. Dazu dient die folgende Definition.

Definition 11.16 (Positivteil, Negativteil einer numerischen Funktion)

Seien $(\Omega, \mathcal{S})$ ein Messraum und $f : \Omega \rightarrow \mathbb{R}$ eine $\mathcal{S}$ - $\bar{\mathcal{B}}$ -messbare numerische Funktion, so wird die Funktion

$$f^+ : \Omega \rightarrow \bar{\mathbb{R}}_0^+, \omega \mapsto \begin{cases} f(\omega), & \text{falls } f(\omega) \geq 0 \\ 0, & \text{sonst} \end{cases}$$

Positivteil von f und die Funktion

$$f^- : \Omega \rightarrow \bar{\mathbb{R}}_0^+, \omega \mapsto \begin{cases} -f(\omega), & \text{falls } f(\omega) \leq 0 \\ 0, & \text{sonst} \end{cases}$$

Negativteil von f genannt.

Die folgenden Eigenschaften von f^+ und f^- sind wichtig:

- $f^+(\omega) \geq 0, f^-(\omega) \geq 0$ für alle $\omega \in \Omega$.
- f^+ und f^- sind $\mathcal{S}$ - $\bar{\mathcal{B}}$ -messbare numerische Funktionen.
- $f = f^+ - f^-$.

Mit Hilfe des Positiv- und Negativteils einer messbaren numerischen Funktion $f : \Omega \rightarrow \mathbb{R}$ können wir das (μ_-) -Integral auf messbare numerische Funktionen erweitern.

Definition 11.17 ((μ_-) integrierbar, (μ_-) quasiintegrierbar, (μ_-) Integral)

Seien $(\Omega, \mathcal{S}, \mu)$ ein Maßraum und $f : \Omega \rightarrow \mathbb{R}$ eine $\mathcal{S}$ - $\bar{\mathcal{B}}$ -messbare numerische Funktion.

f heißt (μ_-) integrierbar, falls $\int f^+ d\mu < \infty$ und $\int f^- d\mu < \infty$.

f heißt (μ_-) quasiintegrierbar, falls $\int f^+ d\mu < \infty$ oder $\int f^- d\mu < \infty$.

Ist f (μ_-) quasiintegrierbar, so ist durch

$$\int f d\mu := \int_{\Omega} f d\mu := \int f^+ d\mu - \int f^- d\mu$$

das (μ_-) Integral von f über Ω definiert.

Als (μ_-) Integral über einer Menge $A \in \mathcal{S}$ definieren wir für (μ_-) quasiintegrierbares $f \cdot I_A$:

$$\int_A f d\mu := \int f \cdot I_A d\mu.$$

Betrachtet man speziell den Maßraum $(\mathbb{R}^n, \mathcal{B}^n, \lambda^n)$, so wird das (λ^n_-) Integral als Lebesgue-Integral bezeichnet. Ist f (λ^n_-) integrierbar, so heißt f Lebesgue-integrierbar. Ist ein Maß μ^F durch eine maßerzeugende Funktion $F : \mathbb{R}^n \rightarrow \mathbb{R}$

gegeben, so wird das $(\mu^F\text{-})$ Integral als Lebesgue-Stieltjes-Integral bezeichnet und in der Form

$$\int f dF := \int f d\mu^F$$

geschrieben. Lebesgue-Stieltjes-Integrale besitzen die wichtige Eigenschaft, dass sie häufig durch Riemann-Integrale berechnet werden können.

Abschliessend betrachten wir Eigenschaften der $(\mu\text{-})$ Integration:

Sind f und g $(\mu\text{-})$ integrierbar, so gilt für alle $A \in \mathcal{S}$:

$$\int_A (f + g) d\mu = \int_A f d\mu + \int_A g d\mu.$$

Ist f $(\mu\text{-})$ quasiintegrierbar, so gilt für alle $c \in \mathbb{R}$ und alle $A \in \mathcal{S}$:

$$\int_A (c \cdot f) d\mu = c \int_A f d\mu.$$

Kapitel 12

Wahrscheinlichkeitstheorie

12.1 Zufallsvariable und Verteilungen

In der Wahrscheinlichkeitstheorie werden Methoden zur Beschreibung und Analyse von Zufallsexperimenten (Experimente mit nicht vorhersehbarem Ausgang) bereitgestellt. Der umgangssprachliche Begriff 'Zufallsexperiment' wird durch einen Maßraum $(\Omega, \mathcal{S}, P)$ mit der Eigenschaft $P(\Omega) = 1$ mathematisch präzisiert. Wir definieren daher:

Definition 12.1 (Wahrscheinlichkeitsraum) Ein Maßraum $(\Omega, \mathcal{S}, P)$ mit $P(\Omega) = 1$ wird als Wahrscheinlichkeitsraum bezeichnet. Die Punkte $\omega \in \Omega$ heißen Ergebnisse, die Mengen $A \in \mathcal{S}$ Ereignisse. Das Maß P wird als Wahrscheinlichkeitsmaß bezeichnet. Für alle Ereignisse A wird $P(A)$ die Wahrscheinlichkeit von A genannt.

Wir werden im folgenden davon ausgehen, dass ein Zufallsexperiment durch einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ gegeben ist. Es ist in der Praxis oft nicht leicht, ein verbal formuliertes Zufallsexperiment durch einen Wahrscheinlichkeitsraum zu modellieren - insbesondere dann, wenn das Experiment ungenau formuliert ist. Die Elemente der Menge Ω stellen die möglichen Ergebnisse des Zufallsexperimentes dar.

Roulette

Beim Roulette wäre etwa $\Omega = \{0, \dots, 36\}$. Häufig interessiert man sich weniger für die Frage, mit welcher Wahrscheinlichkeit ein spezielles Ergebnis eines Zufallsexperimentes eintritt, sondern dafür, mit welcher Wahrscheinlichkeit das Ergebnis Element einer speziellen Teilmenge von Ω ist. Die in Frage kommenden Teilmengen werden Ereignisse genannt und in der σ -Algebra $\mathcal{S}$ zusammengefasst. Die Wahrscheinlichkeit, dass das Ergebnis eines durch $(\Omega, \mathcal{S}, P)$ repräsentierten Zufallsexperiments Element der Menge $A \in \mathcal{S}$ ist, ist durch $P(A)$ gegeben. Einen Roulettespieler, der auf 'gerade Zahl' setzt, interessiert es zum Beispiel nicht, mit

welcher Wahrscheinlichkeit ein Ergebnis eintritt, sondern mit welcher Wahrscheinlichkeit das Ergebnis Element der Menge $\{2, 4, 6, \dots, 36\}$ ist. Gilt für ein $\omega \in \Omega$ auch $\{\omega\} \in \mathcal{S}$, so spricht man von einem Elementarereignis. Das Ereignis $\emptyset$ heißt unmögliches Ereignis, das Ereignis Ω heißt sicheres Ereignis. Da $P(\emptyset) = 0$ und $P(\Omega) = 1$, bezeichnet man eine Menge $A \in \mathcal{S}$, für die $P(A) = 0$ gilt (also eine P -Nullmenge) als (P -)fast unmögliches Ereignis, eine Menge $B \in \mathcal{S}$, für die $P(B) = 1$ gilt, als (P -)fast sicheres Ereignis. Es ist wichtig festzuhalten, dass die definierenden Eigenschaften eines Maßes mit den intuitiv einsichtigen Eigenschaften von Wahrscheinlichkeiten übereinstimmen. Für das Roulette ist es sicher sinnvoll, den folgenden Wahrscheinlichkeitsraum zu betrachten:

$\Omega = \{0, 1, 2, \dots, 36\}$, $\mathcal{S} = \mathcal{P}(\Omega)$, $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$,

$$A \mapsto \begin{cases} 0, & \text{falls } A = \emptyset \\ \frac{|A|}{37}, & \text{falls } A \neq \emptyset \end{cases}, \text{ wobei } |A| \text{ die Anzahl der Elemente von } A \text{ bezeichnet.}$$

Scheibenschießen

Nun betrachten wir das Schießen mit einem Gewehr auf eine kreisförmige Schießscheibe mit Radius $r = \frac{1}{\sqrt{\pi}}$ und dem Mittelpunkt $\mathbf{m} = (0, 0)^T$. Wir nehmen an, dass bei jedem Schuss die Scheibe getroffen wird. Als Ergebnis eines Schusses erhalten wir einen Punkt $\omega = (\omega^1, \omega^2)^T \in \Omega := K_{\frac{1}{\sqrt{\pi}}, 0} := \{\mathbf{x} \in \mathbb{R}^2; \|\mathbf{x}\|_2 \leq \frac{1}{\sqrt{\pi}}\}$. Wir wählen $\{A \cap K_{\frac{1}{\sqrt{\pi}}, 0}; A \in \mathcal{B}^2\}$ als σ -Algebra und $P = \lambda^2|_{\mathcal{S}}$ als Wahrscheinlichkeitsmaß auf $\mathcal{S}$. Da der Schütze bei jedem Schuss umso mehr Punkte (Ringe) erhält, je kleiner der Abstand seines Schusses zum Mittelpunkt der Schießscheibe ist, interessiert als Ergebnis in erster Linie dieser Abstand zum Mittelpunkt.

Man betrachtet also eine Funktion

$$d : \Omega \rightarrow [0, \frac{1}{\sqrt{\pi}}] =: \Omega', \omega \mapsto \|\omega\|_2.$$

Kann man nun mit Hilfe der Funktion d und des Wahrscheinlichkeitsraumes $(\Omega, \mathcal{S}, P)$ jeder Menge $A \in \mathcal{S}' := \{B \cap [0, \frac{1}{\sqrt{\pi}}]; B \in \mathcal{B}\}$ eine Wahrscheinlichkeit zuordnen? Dies ist genau dann möglich, wenn d $\mathcal{S}$ - $\mathcal{S}'$ -messbar ist. Daher definieren wir:

Definition 12.2 (n -dimensionale reelle, numerische) Zufallsvariable)

Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathcal{S}')$ ein Messraum, dann heißt eine $\mathcal{S}$ - $\mathcal{S}'$ -messbare Funktion $X : \Omega \rightarrow \Omega'$ Zufallsvariable.

Ist $\Omega' = \mathbb{R}^n$, $n \in \mathbb{N}$, und $\mathcal{S}' = \mathcal{B}^n$, so wird X als n -dimensionale reelle Zufallsvariable bezeichnet. Ist $\Omega' = \bar{\mathbb{R}}$ und $\mathcal{S}' = \bar{\mathcal{B}}$, so wird X als numerische Zufallsvariable bezeichnet. Eine eindimensionale reelle Zufallsvariable wird reelle Zufallsvariable genannt.

Als geeignetes Wahrscheinlichkeitsmaß P' auf $\mathcal{S}'$ ergibt sich das Bildmaß von X . Somit erhalten wir für unser obiges Beispiel $P'(A') = P(d^{-1}(A'))$ für alle $A' \in \mathcal{S}'$. Die Tatsache, dass $\lambda^2(\{\omega\}) = 0$ für alle $\omega \in K_{\frac{1}{\sqrt{\pi}}, 0}$ verdeutlicht den Sinn der Verwendung von Ereignissen $A \in \mathcal{S}$.

Definition 12.3 (Verteilung einer Zufallsvariablen) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{S}')$ ein Messraum und $X : \Omega \rightarrow \Omega'$ eine Zufallsvariable, dann wird das Bildmaß P_X von X Verteilung von X genannt.

Nach unserer Interpretation von Wahrscheinlichkeitsräumen ist der Wert $X(\omega)$ einer Zufallsvariablen an der Stelle ω vom Ergebnis eines Zufallsexperimentes abhängig. Wir fragen danach, welcher Wert von X 'zu erwarten' ist.

Definition 12.4 (Erwartungswert einer numerischen Zufallsvariablen) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und X eine (P) -quasiintegrierbare numerische Zufallsvariable $X : \Omega \rightarrow \bar{\mathbb{R}}$, dann wird durch

$$\mathbb{E}(X) := \int X \, dP$$

der Erwartungswert von X definiert.

Um eine Vorstellung vom Begriff des Erwartungswertes zu bekommen, betrachten wir die folgende reelle Zufallsvariable auf $(\Omega, \mathcal{S}, P)$: Seien $A_1, \dots, A_n$ paarweise disjunkte Mengen aus $\mathcal{S}$ mit $\sum_{i=1}^n A_i = \Omega$ und $\alpha_1, \dots, \alpha_n$ nichtnegative reelle Zahlen, dann ist

$$X : \Omega \rightarrow \mathbb{R}, \omega \mapsto \sum_{i=1}^n \alpha_i I_{A_i}(\omega)$$

eine reelle Zufallsvariable. Für den Erwartungswert von X erhalten wir

$$\mathbb{E}(X) = \sum_{i=1}^n \alpha_i P(A_i).$$

Der Erwartungswert ist in diesem Fall also eine gewichtete Summe der möglichen Werte von X , wobei die Gewichte gerade die Wahrscheinlichkeiten für das Auftreten dieser Werte sind. Gilt $P(A_i) = \frac{1}{n}$ für alle $i = 1, \dots, n$, so erhalten wir als Erwartungswert das arithmetische Mittel der Werte von X .

Da wir auch an Erwartungswerten von speziellen Funktionen von X interessiert sind, benötigen wir den folgenden Satz.

Satz 12.5 ($\mathcal{S}$ - $\mathcal{B}$ -Messbarkeit stetiger Funktionen reeller Zufallsvariablen) Seien X eine reelle Zufallsvariable definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine stetige Funktion, dann ist $g \circ X : \Omega \rightarrow \mathbb{R}, \omega \mapsto g(X(\omega))$ eine reelle Zufallsvariable auf $(\Omega, \mathcal{S}, P)$.

Somit folgt sofort, dass für eine reelle Zufallsvariable X auf $(\Omega, \mathcal{S}, P)$ und für jedes $k \in \mathbb{N}$ und jedes $\alpha \in \mathbb{R}$ auch $(X - \alpha)^k$ und $|X - \alpha|^k$ reelle Zufallsvariablen auf $(\Omega, \mathcal{S}, P)$ sind. Dies ermöglicht die folgenden Definitionen.

Definition 12.6 (zentrierte (absolute) Momente k -ter Ordnung) Sei X eine auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ definierte reelle Zufallsvariable, dann heißt $\mathbb{E}(|X - \alpha|^k)$, $k \in \mathbb{N}$, das in α zentrierte absolute Moment k -ter Ordnung von X . Ist $(X - \alpha)^k$ (P) -quasiintegrierbar, so heißt $\mathbb{E}((X - \alpha)^k)$ das in α zentrierte Moment k -ter Ordnung. Ist $\alpha = 0$, so spricht man nur von absoluten Momenten bzw. Momenten k -ter Ordnung.

Besonders interessant ist der Fall $k = 2$.

Definition 12.7 (Varianz einer reellen Zufallsvariablen) Sei X eine auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ definierte, (P) -integrierbare reelle Zufallsvariable, dann heißt

$$\mathbb{V}(X) := \int (X - \mathbb{E}(X))^2 dP$$

die Varianz von X .

Die Zahl $\sigma = \sqrt{\mathbb{V}(X)}$ wird als Streuung oder Standardabweichung von X bezeichnet. Oft schreibt man daher σ^2 für $\mathbb{V}(X)$. Die Varianz ist ein Maß für die zu erwartende Abweichung von X und $\mathbb{E}(X)$.

Standardisierung

Für eine reelle Zufallsvariable X mit Streuung $0 < \sigma < \infty$ folgt sofort, dass die reelle Zufallsvariable

$$Y := \frac{X - \mathbb{E}(X)}{\sigma}$$

den Erwartungswert $\mathbb{E}(Y) = 0$ und die Varianz $\mathbb{V}(Y) = 1$ besitzt. Den Übergang von X zu Y bezeichnet man als 'Standardisierung' von X . Den Erwartungswert einer n -dimensionalen reellen Zufallsvariablen definiert man durch komponentenweise Bildung des Erwartungswertes.

Im folgenden betrachten wir einige wichtige Begriffe der elementaren Wahrscheinlichkeitstheorie. Ausgangspunkt ist der Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ und zwei Mengen $A, B \in \mathcal{S}$ mit $P(B) > 0$. Auf $\mathcal{S}$ definieren wir nun ein Wahrscheinlichkeitsmaß $P^B : \mathcal{S} \rightarrow [0, 1]$ durch $A \mapsto \frac{P(A \cap B)}{P(B)}$. Durch den Übergang von P zu P^B erhält die Menge B das Wahrscheinlichkeitsmaß 1. Wir interpretieren $P^B(A)$ als die Wahrscheinlichkeit von A unter der Bedingung, dass das Ereignis B (P^B -)fast sicher eintritt.

Formel von der totalen Wahrscheinlichkeit

Betrachtet man nun eine Partition $\{D_i \subset \Omega; i \in \mathbb{N}\}$ von Ω , sodass für alle $i \in \mathbb{N}$ $D_i \in \mathcal{S}$ und $P(D_i) > 0$ gilt, so lässt sich sehr leicht die folgende 'Formel von der totalen Wahrscheinlichkeit' nachweisen:

$$P(A) = \sum_{i=1}^{\infty} P(D_i) \cdot P^{D_i}(A) \quad \text{für alle } A \in \mathcal{S}.$$

Satz von Bayes

Gilt zusätzlich $P(A) > 0$, so folgt aus

$$P^A(D_i) = \frac{P(D_i \cap A)}{P(A)} = \frac{P^{D_i}(A) \cdot P(D_i)}{P(A)}, \quad i \in \mathbb{N},$$

der 'Satz von Bayes':

$$P^A(D_i) = \frac{P^{D_i}(A) \cdot P(D_i)}{\sum_{j=1}^{\infty} P(D_j) \cdot P^{D_j}(A)} \quad \text{für alle } i \in \mathbb{N}.$$

Analoge Formeln ergeben sich natürlich für eine endliche Partition $\{D_i \subset \Omega; i = 1, \dots, n\}$ von Ω .

Dichte

Im folgenden betrachten wir einen Messraum $(\Omega, \mathcal{S})$, ein Wahrscheinlichkeitsmaß P auf $\mathcal{S}$ und ein Maß μ auf $\mathcal{S}$. Es soll zunächst die Frage untersucht werden, unter welchen Voraussetzungen das Wahrscheinlichkeitsmaß P in der folgenden Art und Weise durch das Maß μ dargestellt werden kann:

Es existiert eine nichtnegative, $\mathcal{S}$ - $\bar{\mathcal{B}}$ -messbare numerische Funktion $f : \Omega \rightarrow \bar{\mathbb{R}}$ mit

$$P(A) = \int_A f d\mu \quad \text{für alle } A \in \mathcal{S}. \quad (1.1)$$

Eine nichtnegative, $\mathcal{S}$ - $\bar{\mathcal{B}}$ -messbare numerische Funktion f , die die Bedingung (1.1) erfüllt, wird Dichte(funktion) des Wahrscheinlichkeitsmaßes P bezüglich μ genannt. Man sagt auch, dass P bezüglich μ eine Dichte f besitzt.

Satz 12.8 (Beziehung zwischen (P) - und (μ) -Nullmengen) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, μ ein Maß auf $\mathcal{S}$ und f eine Dichte von P bezüglich μ , dann gilt für alle $A \in \mathcal{S}$ mit $\mu(A) = 0$: $P(A) = 0$.

Die folgende Definition resultiert aus dem eben betrachteten Satz.

Definition 12.9 (absolute Stetigkeit von P bez. μ) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und μ ein Maß auf $\mathcal{S}$. P heißt absolutstetig bezüglich μ , falls für alle $A \in \mathcal{S}$ mit $\mu(A) = 0$ gilt: $P(A) = 0$.

Wie der folgende Satz zeigt, ist die absolute Stetigkeit bezüglich eines σ -endlichen Maes μ das entscheidende Kriterium fur die Existenz einer Dichte.

Satz 12.10 (Radon-Nikodym) *Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und μ ein σ -endliches Ma auf $\mathcal{S}$, dann besitzt P genau dann eine Dichte bezuglich μ , wenn P absolutstetig bezuglich μ ist.*

Nun betrachten wir eine spezielle Klasse von Wahrscheinlichkeitsmaen. Mit $|A|$ wird die Anzahl der Elemente (Mchtigkeit) von A bezeichnet.

Definition 12.11 (diskretes Wahrscheinlichkeitsma, disk. Zufallsvar.)
Sei $(\mathbb{R}^n, \mathcal{B}^n, P)$, $n \in \mathbb{N}$, ein Wahrscheinlichkeitsraum. Das Wahrscheinlichkeitsma P heit diskret, falls eine Menge $B \in \mathcal{B}^n$ mit $|B| \leq |\mathbb{N}|$ und $P(B) = 1$ existiert. Eine m -dimensionale reelle Zufallsvariable X definiert auf $(\mathbb{R}^n, \mathcal{B}^n, P)$, $m \in \mathbb{N}$, heit diskret, falls das Bildma P_X von X ein diskretes Wahrscheinlichkeitsma auf $(\mathbb{R}^m, \mathcal{B}^m)$ ist.

Da fur $m = n$ das Bildma P_X der Zufallsvariable $X : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $x \mapsto x$, gleich P ist, wird oft der Begriff Verteilung statt Wahrscheinlichkeitsma verwendet. Um nun mit Hilfe des Satzes von Radon-Nikodym diskrete Verteilungen (Wahrscheinlichkeitsmae) durch Dichtefunktionen darstellen zu konnen, bentigen wir ein spezielles Ma.

Definition 12.12 (Zhlma) *Das auf einer σ -Algebra $\mathcal{S}$ uber Ω definierte Ma*

$$\zeta : \mathcal{S} \rightarrow \bar{\mathbb{R}}, A \mapsto \begin{cases} |A|, & \text{falls } |A| \text{ endlich ist} \\ \infty, & \text{sonst} \end{cases}$$

wird als das Zhlma auf $\mathcal{S}$ bezeichnet.

Sei nun $(\mathbb{R}^n, \mathcal{B}^n, P)$ ein Wahrscheinlichkeitsraum und P eine diskrete Verteilung auf $\mathcal{B}^n$ mit $P(B) = 1$ fur ein $B \in \mathcal{B}^n$ und $|B| \leq |\mathbb{N}|$, dann gilt fur alle $C \in \mathcal{B}^n$:

$$P(C) = P(C \cap B) + P(C \cap B^c) = P(C \cap B).$$

Somit genugt es, den Wahrscheinlichkeitsraum $(B, \mathcal{B}_B^n, P)$ mit $\mathcal{B}_B^n := \{C \cap B; C \in \mathcal{B}^n\} = \mathcal{P}(B)$ zu betrachten. Da ζ ein σ -endliches Ma auf $\mathcal{P}(B)$ ist und $\zeta(A) = 0$ genau dann gilt, wenn $A = \emptyset$, ist jedes Wahrscheinlichkeitsma auf $\mathcal{P}(B)$ absolutstetig bezuglich ζ . Somit existiert zu jedem Wahrscheinlichkeitsma P auf $\mathcal{P}(B)$

eine Dichte $f : B \rightarrow \mathbb{R}_0^+$ mit

$$P(A) = \int_A f d\zeta = \sum_{\omega \in A} f(\omega) = \sum_{\omega \in A} P(\{\omega\}) \text{ für alle } A \in \mathcal{P}(B).$$

Es lässt sich also jede diskrete Verteilung auf $\mathcal{B}^n$ durch eine Folge $\{p_j\}$, $j \in \mathbb{N}_0$, nichtnegativer reeller Zahlen mit $\sum_{j=0}^{\infty} p_j = 1$ darstellen.

Poisson-Verteilung

Ist

$$p_j = e^{-\lambda} \frac{\lambda^j}{j!}, \quad j \in \mathbb{N}_0, \quad \lambda > 0,$$

so spricht man von einer Poisson-Verteilung mit Parameter λ (D. Poisson (1781-1840)).

Gleichverteilung und Laplace-Experiment

Ist

$$p_j = \frac{1}{k+1}, \text{ für } j = 0, \dots, k, \text{ und } p_j = 0 \text{ für } j > k, \quad k \in \mathbb{N}_0,$$

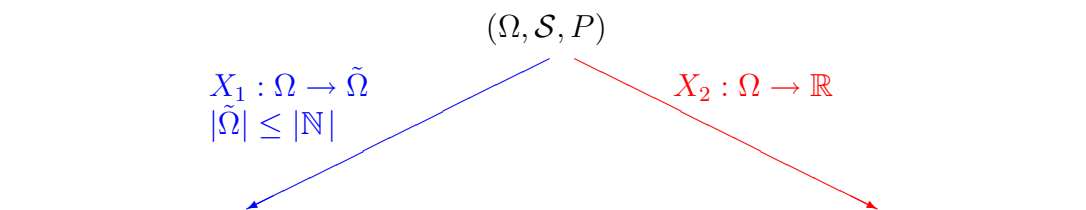
so wird diese Verteilung Gleichverteilung genannt. Ein Zufallsexperiment, das durch einen Wahrscheinlichkeitsraum mit Gleichverteilung repräsentiert wird, heißt nach P. S. de Laplace (1749-1827) Laplace-Experiment.

Binomial-Verteilung und Bernoulli-Experiment

Wählt man $p \in \mathbb{R}$, $0 < p < 1$, und $B = \{0, 1, 2, \dots, s\}$, $s \in \mathbb{N}$, so wird (mit $\binom{s}{j} := \frac{s!}{(s-j)!j!}$) die durch

$$p_j = \binom{s}{j} p^j (1-p)^{s-j} \text{ für } j = 0, \dots, s, \text{ und } p_j = 0 \text{ für } j > s,$$

gegebene Verteilung Binomial-Verteilung $B(s, p)$ mit Parameter s, p genannt. Ein Zufallsexperiment, das durch einen Wahrscheinlichkeitsraum mit Binomial-Verteilung mit Parameter s, p repräsentiert wird, heißt nach J. Bernoulli (1654-1705) Bernoulli-Experiment mit Parameter s, p . Ein Bernoulli-Experiment kann folgendermaßen interpretiert werden: Man betrachtet ein Zufallsexperiment, bei dem es nur zwei mögliche Ergebnisse gibt, nämlich mit Wahrscheinlichkeit p das Ergebnis 'T' (Treffer) und mit Wahrscheinlichkeit $(1-p)$ das Ergebnis 'N' (Niete). Dieses Experiment führen wir s -mal durch, ohne dass sich die Ergebnisse gegenseitig beeinflussen. Die Wahrscheinlichkeit, dass nach diesen s Versuchen genau j Treffer auftreten, ist gegeben durch $\binom{s}{j} p^j (1-p)^{s-j}$, $0 \leq j \leq s$, $s \in \mathbb{N}$. Somit wird die s -malige Durchführung unseres Experimentes durch ein Bernoulli-Experiment beschrieben, falls die Ergebnisse sich nicht gegenseitig beeinflussen. Für sehr grosse s und sehr kleine p ist es möglich, eine Binomial-Verteilung durch die wesentlich einfacher zu berechnende Poisson-Verteilung mit Parameter $\lambda = s \cdot p$ zu approximieren.

Zusammenfassung:

$$(\tilde{\Omega}, \mathcal{P}(\tilde{\Omega}), P_{X_1})$$

Dichte:

$$f : \tilde{\Omega} \rightarrow [0, 1], \tilde{\omega} \mapsto P_{X_1}(\{\tilde{\omega}\})$$

$$P_{X_1}(A) = \sum_{\tilde{\omega} \in A} f(\tilde{\omega})$$

Erwartungswert für $\tilde{\Omega} \subset \mathbb{R}$:

$$\mathbb{E}(X_1) = \sum_{\tilde{\omega} \in \tilde{\Omega}} \tilde{\omega} \cdot f(\tilde{\omega})$$

Allgemein: Sei $g : \tilde{\Omega} \rightarrow \mathbb{R}$:

$$\mathbb{E}(g(X_1)) = \sum_{\tilde{\omega} \in \tilde{\Omega}} g(\tilde{\omega}) \cdot f(\tilde{\omega})$$

Beispiele:

- Poisson-Verteilung

$$\tilde{\Omega} = \mathbb{N}_0, \lambda > 0$$

$$f(j) = e^{-\lambda} \frac{\lambda^j}{j!}$$

- Binomial-Verteilung

$$\tilde{\Omega} = \{0, 1, \dots, s\}, p \in]0, 1[$$

$$f(j) = \binom{s}{j} p^j (1-p)^{s-j},$$

„genau j Treffer bei s Versuchen, Trefferwahrscheinlichkeit gleich p, keine gegenseitige Beeinflussung.“

$$(\mathbb{R}, \mathcal{B}, P_{X_2})$$

Annahme: Es existiert bezüglich λ eine stetige Dichte $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$:

$$P_{X_2}(A) = \int_A f(x) dx$$

Erwartungswert:

$$\mathbb{E}(X_2) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Allgemein: Sei $g : \mathbb{R} \rightarrow \mathbb{R}$ und $\mathcal{B}$ - $\mathcal{B}$ -messbar, so gilt:

$$\mathbb{E}(g(X_2)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$$

Beispiele:

- Normalverteilung

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mathbb{E}(X_2) = \mu \in \mathbb{R}, \mathbb{V}(X_2) = \sigma^2 > 0$$

- Rayleigh-Verteilung, $\lambda > 0$:

$$f(x) = \frac{2x}{\lambda} e^{-\frac{x^2}{\lambda}}, x > 0, (0 \text{ sonst})$$

$$\mathbb{E}(X_2) = \frac{1}{2} \sqrt{\lambda\pi}, \mathbb{V}(X_2) = (1 - \frac{\pi}{4})\lambda$$

- Cauchy-Verteilung, $\lambda > 0$:

$$f(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x - \mu)^2}, \mu \in \mathbb{R}.$$

Stetige Dichten für $n > 1$:

Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, eine stetige Funktion mit folgenden Eigenschaften:

- $f(x) \geq 0$ für alle $x \in \mathbb{R}^n$,
- $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x) dx = 1$,

dann ist auch $\int f d\lambda^n = 1$ und wir können die Funktion f als Dichte eines Wahrscheinlichkeitsmaßes bezüglich λ^n auffassen. Nun betrachten wir für jeden Vektor $\mu \in \mathbb{R}^n$ und für jede positiv definite Matrix $\Sigma \in \mathbb{R}^{n,n}$ die Funktion

$$\nu_{\mu,\Sigma} : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \cdot \exp \left(-\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2} \right).$$

Offensichtlich ist $\nu_{\mu,\Sigma}(x) > 0$ für alle $\mu, x \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n,n}$, Σ positiv definit. Aus der Analysis (Substitutionsregel, Satz von Fubini) ist das Folgende bekannt:

$$\int_{\mathbb{R}^n} \exp \left(-\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2} \right) dx = \sqrt{(2\pi)^n \det(\Sigma)}$$

für alle $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n,n}$, Σ positiv definit. Somit können wir $\nu_{\mu,\Sigma}$ als Dichte eines Wahrscheinlichkeitsmaßes bezüglich λ^n auffassen.

Definition 12.13 (Normalverteilung) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $\mu \in \mathbb{R}^n$, $n \in \mathbb{N}$, und $\Sigma \in \mathbb{R}^{n,n}$, Σ positiv definit. Die Zufallsvariable $X_{\mu,\Sigma} : \Omega \rightarrow \mathbb{R}^n$ heißt $\mathcal{N}(\mu, \Sigma)$ normalverteilt, falls ihr Bildmaß $P_{X_{\mu,\Sigma}}$ bezüglich λ^n die folgende Dichte besitzt:

$$\nu_{\mu,\Sigma} : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \cdot \exp \left(-\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2} \right).$$

Häufig werden spezielle Wahrscheinlichkeitsmaße durch folgende Funktionen dargestellt:

Definition 12.14 (Verteilungsfunktion) Sei $(\mathbb{R}^n, \mathcal{B}^n, P)$, $n \in \mathbb{N}$, ein Wahrscheinlichkeitsraum. Die Funktion

$$F : \mathbb{R}^n \rightarrow [0, 1], (x_1, \dots, x_n)^T \mapsto P([-\infty, x_1[\times \dots \times]-\infty, x_n[)$$

wird als Verteilungsfunktion von P bezeichnet. Die Verteilungsfunktion des Bildmaßes P_X einer n -dimensionalen reellen Zufallsvariable $X : \Omega \rightarrow \mathbb{R}^n$, $n \in \mathbb{N}$, wird auch Verteilungsfunktion von X genannt.

Um die Parameter μ und Σ einer Normalverteilung interpretieren zu können, benötigen wir die folgende Definition.

Definition 12.15 (Covarianz, unkorreliert) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$, $Y : \Omega \rightarrow \mathbb{R}$ zwei reelle, (P) -integrierbare Zufallsvariable mit (P) -integrierbarem Produkt $X \cdot Y$, dann heißt

$$\mathbb{K}(X, Y) := \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

die Covarianz von X und Y . X und Y heißen unkorreliert, falls $\mathbb{K}(X, Y) = 0$.

Besitzen die Zufallsvariablen X bzw. Y zudem endliche Varianzen $\mathbb{V}(X)$ bzw. $\mathbb{V}(Y)$, so wird die Größe

$$\rho(X, Y) := \frac{\mathbb{K}(X, Y)}{\sqrt{\mathbb{V}(X) \cdot \mathbb{V}(Y)}}$$

Korrelationskoeffizient von X und Y genannt.

Normalverteilte Zufallsvariablen spielen in der Wahrscheinlichkeitstheorie eine bedeutende Rolle, auf die wir im Zusammenhang mit dem zentralen Grenzwertsatz noch zu sprechen kommen. Zunächst fassen wir einige Eigenschaften einer $\mathcal{N}(\mu, \Sigma)$ normalverteilten Zufallsvariablen $X_{\mu, \Sigma}$ zusammen. Dazu fassen wir die Funktion $X_{\mu, \Sigma} : \Omega \rightarrow \mathbb{R}^n$ als Abbildung

$$\omega \mapsto (X_{\mu, \Sigma}^1(\omega), \dots, X_{\mu, \Sigma}^n(\omega))^T$$

auf. Jede Funktion $X_{\mu, \Sigma}^i : \Omega \rightarrow \mathbb{R}$, $i = 1, \dots, n$, ist eine reelle Zufallsvariable. Definiert man

$$\mathbb{E}(X_{\mu, \Sigma}) := (\mathbb{E}(X_{\mu, \Sigma}^1), \dots, \mathbb{E}(X_{\mu, \Sigma}^n))^T,$$

so erhält man

$$\mathbb{E}(X_{\mu, \Sigma}) = \mu.$$

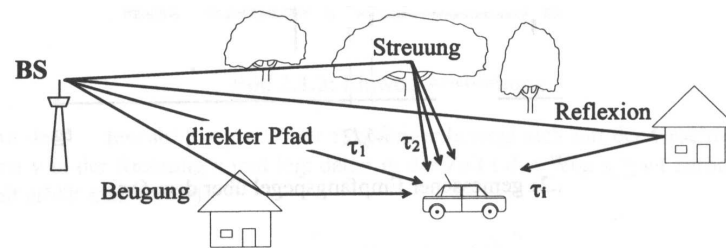
Ferner gilt mit $\Sigma = (\sigma_{i,j})_{i,j=1,\dots,n}$:

$$\mathbb{K}(X_{\mu, \Sigma}^i, X_{\mu, \Sigma}^j) = \sigma_{i,j}, \quad i, j = 1, \dots, n.$$

Daher heißt Σ die Covarianzmatrix von $X_{\mu, \Sigma}$.

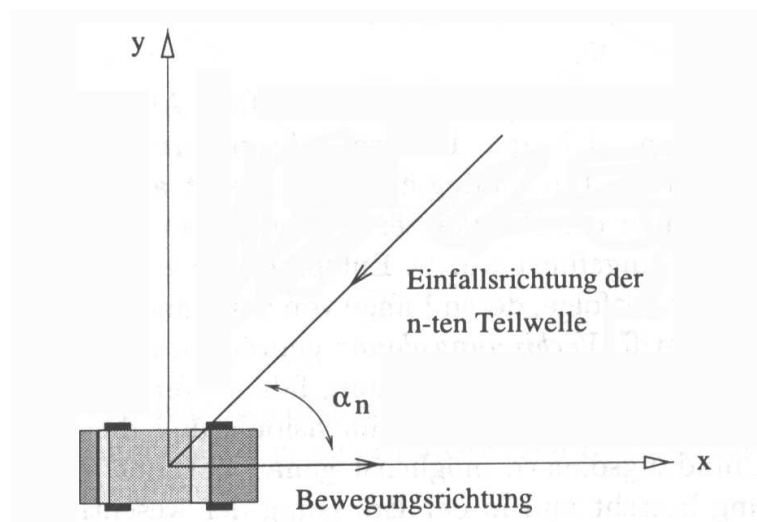
Beispiel: Mobilfunk

Beim Mobilfunk gelangen die vom Sender abgestrahlten elektromagnetischen Wellen aufgrund von Hindernissen häufig nicht (nur) über den direkten Weg zur Empfangsantenne. Vielmehr setzen sich die empfangenen Wellen vorwiegend aus einer Vielzahl von reflektierten, gebeugten und gestreuten Wellen zusammen. Ein typisches Beispiel für diese so genannte Mehrwegeausbreitung ist in der folgenden Abbildung für den terrestrischen Mobilfunkkanal dargestellt. Bedingt durch die Mehrwege-



ausbreitung besteht das Empfangssignal aus einer Überlagerung von unterschiedlich stark gedämpften, verzögerten und phasenverschobenen Versionen des Sendesignals, welche sich gegenseitig auslöschen beziehungsweise verstärken. Daher kann je nach Phasenlage der einzelnen Teilwellen die Überlagerung konstruktiv oder destruktiv erfolgen. Außerdem kommt es durch die Mehrwegeausbreitung bei der digitalen Datenübertragung zu einer Veränderung der Impulsform bis hin zu mehreren, einzeln trennbaren Impulsen. Dieser Effekt wird als Impulsdispersion bezeichnet. Die Größe der Impulsdispersion hängt von den Laufzeitdifferenzen und den Amplitudenverhältnissen der Teilwellen ab.

Neben der Mehrwegeausbreitung wirkt sich auch der Dopplereffekt negativ auf die Übertragungseigenschaften des Mobilfunkkanals aus. Durch die Bewegung der Mobilstation kommt es aufgrund des Dopplereffektes zu einer Frequenzverschiebung der einzelnen Teilwellen. Der Einfallswinkel - definiert durch die Einfallsrichtung der Teilwelle und die Bewegungsrichtung der Mobilstation - bestimmt die Dopplerfrequenz (Frequenzverschiebung) einer Teilwelle. Bedingt durch den Dopplereffekt



erfährt das Spektrum eines Sendesignals auf dem Übertragungsweg eine Frequenzverbreiterung. Dieser Effekt wird als Frequenzdispersion bezeichnet. Die Größe der Frequenzdispersion hängt im wesentlichen von der maximalen Dopplerfrequenz und den Amplituden der einzelnen empfangenen Teilwellen ab. Im Zeitbereich äußert sich der Dopplereffekt dadurch, dass die Impulsantwort des Kanals zeitvariant wird.

Moderne Methoden zur stochastischen Modellierung von Mobilfunkkanälen zeichnen sich insbesondere dadurch aus, dass sie in der Lage sind, nicht nur die Wahrscheinlichkeitsdichten der Amplitude und Phase zu modellieren, sondern auch Aussagen über die Pegelunterschreitungsraten und die mittlere Fadingdauer bereitzustellen, die für Bündelfehler verantwortlich sind. Da zudem eine additive Überlagerung mit stochastisch zu modellierenden Störungen zu berücksichtigen ist, wird das Empfangssignal somit durch die Verknüpfung verschiedener stochastischer Prozesse als sehr komplexer stochastischer Prozess mit parameterabhängiger Kovarianzstruktur modelliert. Das vollständige stochastische Kanalmodell erhält man schließlich durch Berücksichtigung des Demodulators, was im allgemeinen einer komplizierten stochastischen Integration entspricht.

Letztendlich erhält man als Ergebnis des Empfängers die Realisierung einer mischverteilten Zufallsvariablen mit parameterabhängiger Kovarianzstruktur.

Auf der Basis eines Wahrscheinlichkeitsraumes $(\Omega, \mathcal{S}, P)$ haben wir für $A, B \in \mathcal{S}$ und $P(B) > 0$ durch $P^B(A) = \frac{P(A \cap B)}{P(B)}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{S}$ eingeführt. Wir interpretierten $P^B(A)$ als die Wahrscheinlichkeit von A unter der Bedingung, dass B (P^B -)fast sicher eintritt. Nun stellt sich die Frage, wann diese Bedingung die Wahrscheinlichkeit für A nicht ändert, wann also $P^B(A) = P(A)$ gilt. Wir erhalten:

$$P^B(A) = P(A) \iff P(A \cap B) = P(A) \cdot P(B).$$

Definition 12.16 (stochastisch unabhängige Ereignisse) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n \in \mathcal{S}$, $n \in \mathbb{N}$, dann heißen die Ereignisse $A_1, \dots, A_n$ stochastisch unabhängig, falls für alle $k \in \mathbb{N}$, $k \leq n$, und für alle $i_j \in \mathbb{N}$, $1 \leq j \leq k$, mit $1 \leq i_1 < \dots < i_k \leq n$ gilt:

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}).$$

Die stochastische Unabhängigkeit einer Menge $\{A_i \in \mathcal{S}; i \in I\}$, $I \neq \emptyset$, von Ereignissen führt man auf die stochastische Unabhängigkeit ihrer endlichen Teilmengen zurück.

Definition 12.17 (stochastische Unabhängigkeit von Ereignissen) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\{A_i \in \mathcal{S}; i \in I\}$, $I \neq \emptyset$, eine Menge von Ereignissen, dann heißen diese Ereignisse stochastisch unabhängig, falls $A_{i_1}, \dots, A_{i_n}$ für jedes $n \in \mathbb{N}$ mit $n \leq |I|$ und für jede Menge $\{i_1, \dots, i_n\} \subseteq I$ stochastisch unabhängig sind.

Um stochastisch unabhängige Zufallsvariable definieren zu können, wird zunächst die stochastische Unabhängigkeit von Mengensystemen betrachtet.

Definition 12.18 (stochastische Unabhängigkeit von Mengensystemen)

Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\{\mathcal{F}_i \subseteq \mathcal{S}; i \in I\}$, $I \neq \emptyset$, eine Menge von Mengensystemen über Ω , dann heißen diese Mengensysteme stochastisch unabhängig, falls für jedes $n \in \mathbb{N}$ mit $n \leq |I|$ und für jedes $\{i_1, \dots, i_n\} \subseteq I$ die n Ereignisse $A_{i_1}, \dots, A_{i_n}$ für beliebige $A_{i_k} \in \mathcal{F}_{i_k}$, $k = 1, \dots, n$, stochastisch unabhängig sind.

Nun betrachten wir einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$, einen Messraum $(\Omega', \mathcal{S}')$ und eine Zufallsvariable $X : \Omega \rightarrow \Omega'$. Mit $\mathcal{F}$ bezeichnen wir die Menge aller σ -Algebren über Ω , für die gilt: X ist $\mathcal{C}$ - $\mathcal{S}'$ -messbar genau dann, wenn $\mathcal{C} \in \mathcal{F}$. Die Menge $\sigma(X) := \bigcap_{\mathcal{C} \in \mathcal{F}} \mathcal{C}$ ist ebenfalls eine σ -Algebra und wird die von X erzeugte σ -Algebra genannt. Unter allen σ -Algebren $\mathcal{A}$ über Ω ist $\sigma(X)$ die kleinste, für die X $\mathcal{A}$ - $\mathcal{S}'$ -messbar ist. Somit sind wir in der Lage, die stochastische Unabhängigkeit von Zufallsvariablen in naheliegender Weise durch die stochastische Unabhängigkeit von speziellen Mengensystemen zu definieren.

Definition 12.19 (stochastische Unabhängigkeit von Zufallsvariablen)

Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{S}')$ ein Messraum und $\{X_i : \Omega \rightarrow \Omega'; i \in I\}$, $I \neq \emptyset$, eine Menge von Zufallsvariablen, dann heißen diese Zufallsvariablen stochastisch unabhängig, falls die Mengensysteme $\{\sigma(X_i); i \in I\}$ stochastisch unabhängig sind.

Die stochastische Unabhängigkeit von Zufallsvariablen ist ein zentraler Begriff der Wahrscheinlichkeitstheorie und im wesentlichen Bestandteil der Modellierung zu untersuchender Vorgänge. Für reelle, normalverteilte Zufallsvariablen sind die Begriffe „unkorreliert“ und „stochastisch unabhängig“ äquivalent. Ansonsten sind stochastisch unabhängige, integrierbare Zufallsvariablen immer paarweise unkorreliert.

Beispiele:

- Zwei Zufallsvariablen

Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega_1, \mathcal{S}_1)$ und $(\Omega_2, \mathcal{S}_2)$ Messräume und $X_1 : \Omega \rightarrow \Omega_1$, $X_2 : \Omega \rightarrow \Omega_2$ zwei stochastisch unabhängige Zufallsvariablen, so gilt für alle $A_1 \in \mathcal{S}_1$ und alle $A_2 \in \mathcal{S}_2$:

$$P(X_1^{-1}(A_1) \cap X_2^{-1}(A_2)) = P_{X_1}(A_1) \cdot P_{X_2}(A_2).$$

- n-dim. Normalverteilung:

Betrachtet man die λ^n -Dichte

$$\nu_{\mu, \Sigma} : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \cdot \exp \left(-\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2} \right).$$

der n -dim. Normalverteilung und setzt man voraus, dass je zwei Komponenten des zugrundegelegten n -dim. Zufallsvektors unkorreliert sind, so erhält man mit $\Sigma = (\sigma_{i,j})_{i,j=1,\dots,n}$:

$$\sigma_{i,j} = \begin{cases} \sigma_i^2 & \text{falls } i = j \\ 0 & \text{sonst} \end{cases}.$$

Die Dichte ergibt sich dann zu

$$\begin{aligned} \nu_{\mu, \Sigma}(x_1, \dots, x_n) &= \frac{1}{\sqrt{(2\pi)^n \sigma_1^2 \cdot \dots \cdot \sigma_n^2}} \cdot \exp \left(-\sum_{i=1}^n \frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{(2\pi)\sigma_i^2}} \cdot \exp \left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right). \\ &= \prod_{i=1}^n f_i(x_i), \end{aligned}$$

wobei f_i die λ -Dichte einer eindimensionalen Normalverteilung mit Erwartungswert μ_i und Varianz σ_i^2 darstellt (abgekürzt: $\mathcal{N}(\mu_i, \sigma_i^2)$ normalverteilt) und die Verteilung der i -ten Komponente des betrachteten Zufallsvektors repräsentiert. Diese Produktformel für Dichten ist charakteristisch für stochastische Unabhängigkeit.

Im folgenden betrachten wir Folgen reeller Zufallsvariablen. Da eine Folge von reellen Zufallsvariablen eine Funktionenfolge ist, definiert man - wie in der Analysis (z.B. gleichmäßige- und punktweise Konvergenz) - auch in der Wahrscheinlichkeitstheorie verschiedene Konvergenzbegriffe.

Definition 12.20 (Konvergenzbegriffe für Folgen reeller Zufallsvariable)

Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $\{X_i\}$, $i \in \mathbb{N}$, eine Folge reeller Zufallsvariable $X_i : \Omega \rightarrow \mathbb{R}$, und $X : \Omega \rightarrow \mathbb{R}$ ebenfalls eine reelle Zufallsvariable, dann konvergiert $\{X_i\}$, $i \in \mathbb{N}$, definitionsgemäß

- stochastisch gegen X genau dann, wenn für alle $\epsilon > 0$

$$\lim_{i \rightarrow \infty} P(\{\omega \in \Omega; |X_i(\omega) - X(\omega)| < \epsilon\}) = 1,$$

- mit Wahrscheinlichkeit 1 gegen X genau dann, wenn

$$P\left(\left\{\omega \in \Omega; \lim_{i \rightarrow \infty} X_i(\omega) = X(\omega)\right\}\right) = 1,$$

- in Verteilung gegen X genau dann, wenn die Folge der Verteilungsfunktionen $\{F_{X_i}\}$ von $\{X_i\}$ an jeder Stelle $x \in \mathbb{R}$, an der die Verteilungsfunktion F_X von X stetig ist, gegen F_X an der Stelle x konvergiert:

$$\lim_{i \rightarrow \infty} F_{X_i}(x) = F_X(x).$$

Die stochastische Konvergenz von $\{X_i\}$, $i \in \mathbb{N}$, gegen X wird oft durch

$$(P-) \lim_{i \rightarrow \infty} X_i = X, \text{ st-} \lim_{i \rightarrow \infty} X_i = X \text{ oder } X_i \rightarrow X \text{ nach Wahrscheinlichkeit}$$

dargestellt. Die Konvergenz mit Wahrscheinlichkeit 1 von $\{X_i\}$, $i \in \mathbb{N}$, gegen X heißt auch (P) -fast sichere Konvergenz und wird durch $X_i \rightarrow X$ (P) -f.s. dargestellt. Die Konvergenz nach Verteilung wird auch als schwache Konvergenz bezeichnet. Die folgenden Implikationen gelten.

schwache Konvergenz	$\Leftarrow$	stochastische Konvergenz	$\Leftarrow$	Konvergenz mit Wahrscheinlichkeit 1
------------------------	--------------	-----------------------------	--------------	--

Ausgehend von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ betrachten wir spezielle Folgen $\{X_i\}$, $i \in \mathbb{N}$, von reellen Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$, $i \in \mathbb{N}$, deren Quadrate $X_i^2 : \Omega \rightarrow \mathbb{R}$, $\omega \mapsto X_i^2(\omega)$ für alle $i \in \mathbb{N}$ (P) -integrierbar sind. Wegen

$$\begin{aligned} \int_{\Omega} |X_i| dP &= \int_{\{\omega \in \Omega; |X_i(\omega)| \leq 1\}} |X_i| dP + \int_{\{\omega \in \Omega; |X_i(\omega)| > 1\}} |X_i| dP \\ &\leq 1 + \int_{\{\omega \in \Omega; |X_i(\omega)| > 1\}} |X_i| dP \leq 1 + \int_{\Omega} X_i^2 dP \quad \text{für alle } i \in \mathbb{N} \end{aligned}$$

besitzen die Zufallsvariablen $X_i, i \in \mathbb{N}$, endliche Erwartungswerte. Wir vereinbaren, dass für die Folge $\{X_i\}, i \in \mathbb{N}$, genau dann der zentrale Grenzwertsatz gilt, wenn die Folge $\{T_i\}, i \in \mathbb{N}$, standardisierter reeller Zufallsvariablen

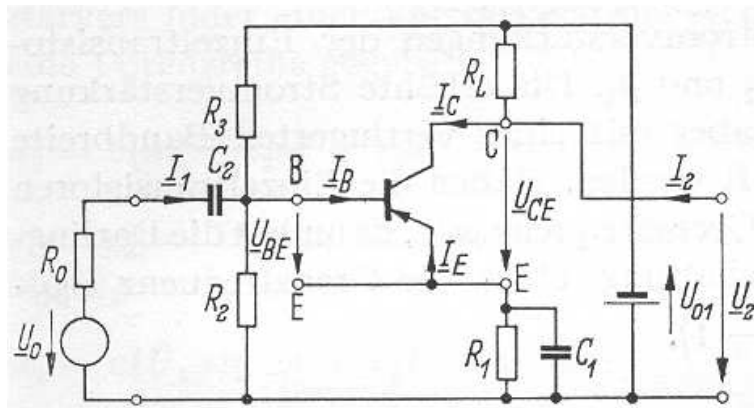
$$T_i : \Omega \rightarrow \mathbb{R}, \omega \mapsto \frac{\sum_{j=1}^i (X_j - \mathbb{E}(X_j))}{\sqrt{\mathbb{V}(\sum_{j=1}^i X_j)}}, \quad i \in \mathbb{N},$$

in Verteilung gegen eine $\mathcal{N}(0, 1)$ normalverteilte Zufallsvariable konvergiert.

Satz 12.21 (Ein zentraler Grenzwertsatz) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\{X_i\}, i \in \mathbb{N}$, eine Folge stochastisch unabhängiger, identisch verteilter (d.h. $P_{X_i} = P_{X_j}$ für alle $i, j \in \mathbb{N}$) reeller Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$ mit $0 < \mathbb{V}(X_i) < \infty$ für alle $i \in \mathbb{N}$, dann gilt für $\{X_i\}, i \in \mathbb{N}$, der zentrale Grenzwertsatz.

Beispiel: Schaltkreissimulation

Bei der numerischen Analyse elektronischer Schaltkreise ist es heute unumgänglich, das elektronische Rauschen der einzelnen Bauteile durch stochastische Modelle zu erfassen. Bereits bei dem einfachen Beispiel einer Transistorverstärkerstufe



ist es nicht mehr ratsam, jedes Bauteil einzeln zu modellieren. Man verwendet zentrale Grenzwertsätze zur theoretischen Begründung eines stochastischen Gesamtmodells.

Zentrale Grenzwertsätze sind der Grund für die grosse Bedeutung der $\mathcal{N}(0, 1)$ Normalverteilung, deren Verteilungsfunktion Φ wir im folgenden tabellieren, da sie nicht analytisch berechenbar ist.

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^u \exp\left(-\frac{t^2}{2}\right) dt = 1 - \Phi(-u), \quad \text{z.B. } \Phi(1.32) = 0.90658.$$

u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52791	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56750	0.57143	0.57535
0.2	0.57926	0.58317	0.58707	0.59096	0.59484	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68438	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84135	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90148
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92786	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95819	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Der nun folgende Satz stellt ein sehr hilfreiches Resultat zur Abschätzung von Wahrscheinlichkeiten dar.

Satz 12.22 (Ungleichung von Chebyshev-Markov) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine reelle Zufallsvariable, dann gilt für jedes Paar reeller Zahlen $\epsilon > 0$, $\kappa > 0$ die folgende Ungleichung von Chebyshev-Markov

$$P(\{\omega \in \Omega; |X(\omega)| < \epsilon\}) \geq 1 - \frac{1}{\epsilon^\kappa} \int |X|^\kappa dP.$$

Beispiel: n-fache Wiederholung

Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\{X_i\}$, $i \in \mathbb{N}$, eine Folge stochastisch unabhängiger, identisch verteilter (d.h. $P_{X_i} = P_{X_j}$ für alle $i, j \in \mathbb{N}$) (P-)integrierbarer Zufallsvariablen für alle $i \in \mathbb{N}$, so gilt mit dem obigen Satz angewendet auf $Y = \sum_{i=1}^n \frac{X_i}{n} - \mu$ mit $\mu = \mathbb{E}(X_i)$ und $\kappa = 2$:

$$P\left(\left\{\omega \in \Omega; \left|\sum_{i=1}^n \frac{X_i(\omega)}{n} - \mu\right| < \epsilon\right\}\right) \geq 1 - \frac{1}{\epsilon^2} \frac{\mathbb{V}(X_i)}{n}.$$

Also gilt für endliche Varianz:

$$\text{st-} \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{X_i}{n} = \mu.$$

12.2 Bedingte Erwartungen

Ausgehend von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ betrachten wir nun eine numerische, (P-)integrierbare Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ und die Abbildung $Y : \Omega \rightarrow \mathbb{R}$, $\omega \mapsto \mathbb{E}(X)$. Diese Abbildung ist für jede σ -Algebra $\mathcal{G}$ über Ω $\mathcal{G}$ - $\bar{\mathcal{B}}$ -messbar. Somit ist $\{\emptyset, \Omega\}$ die kleinste aller σ -Algebren $\mathcal{G}$ über Ω , für die Y $\mathcal{G}$ - $\bar{\mathcal{B}}$ -messbar ist, und es gilt:

$$\int_A Y dP = \int_A X dP \quad \text{für alle } A \in \{\emptyset, \Omega\}.$$

Ist nun X ebenfalls $\{\emptyset, \Omega\}$ - $\bar{\mathcal{B}}$ -messbar, so existiert eine (P-)Nullmenge $N \in \mathcal{S}$ mit

$$X(\omega) = \mathbb{E}(X) = Y(\omega) \quad \text{für alle } \omega \in N^c.$$

Eine andere Schreibweise dafür lautet:

$$X = \mathbb{E}(X) = Y \quad (P\text{-}) \text{ fast sicher.}$$

Daher ist der Übergang von X zu Y mit keinerlei Informationsverlust (bezüglich der Verteilung von X) verbunden. Ist aber X nicht $\{\emptyset, \Omega\}$ - $\bar{\mathcal{B}}$ -messbar, so haben wir beim Übergang von X zu Y Information über die Verteilung von X verloren. Diesen Informationsverlust wollen wir folgendermaßen quantifizieren: Ist $\mathcal{F}$ die Menge aller σ -Algebren über Ω , sodass X für alle $\mathcal{C} \in \mathcal{F}$ $\mathcal{C}$ - $\bar{\mathcal{B}}$ -messbar ist, so ist

$$\sigma(X) := \bigcap_{\mathcal{C} \in \mathcal{F}} \mathcal{C}$$

die kleinste σ -Algebra über Ω , für die X $\mathcal{C}$ - $\bar{\mathcal{B}}$ -messbar ist. Selbstverständlich gilt $\{\emptyset, \Omega\} \subseteq \sigma(X)$. Ist nun $\{\emptyset, \Omega\} = \sigma(X)$, so haben wir keine Information verloren. Der Informationsverlust beim Übergang von X zu Y (bezüglich der Verteilung von X) wächst mit der Anzahl der Mengen $A \in \sigma(X)$, $A \neq \{\emptyset, \Omega\}$. Wir verwenden also σ -Algebren (in unserem Beispiel $\{\emptyset, \Omega\}$ und $\sigma(X)$), um den Informationsverlust zu charakterisieren.

Im folgenden betrachten wir den umgekehrten Weg, indem wir uns einen Informationsverlust in Form einer σ -Algebra $\mathcal{C} \subseteq \mathcal{S}$ über Ω vorgeben und eine $\mathcal{C}$ - $\bar{\mathcal{B}}$ -messbare numerische Zufallsvariable $Y : \Omega \rightarrow \bar{\mathbb{R}}$ betrachten, für die gilt:

$$\int_A Y dP = \int_A X dP \quad \text{für alle } A \in \mathcal{C}.$$

Satz und Definition 12.23 (bedingte Erwartung) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \bar{\mathbb{R}}$ eine (P) -integrierbare numerische Zufallsvariable, dann existiert zu jeder σ -Algebra $\mathcal{C} \subseteq \mathcal{S}$ über Ω eine $\mathcal{C}$ - $\bar{\mathcal{B}}$ -messbare numerische Zufallsvariable $Y : \Omega \rightarrow \bar{\mathbb{R}}$ mit:

$$\int_A Y dP = \int_A X dP \quad \text{für alle } A \in \mathcal{C}.$$

Für zwei $\mathcal{C}$ - $\bar{\mathcal{B}}$ -messbare numerische Zufallsvariable $Y, Y_1 : \Omega \rightarrow \bar{\mathbb{R}}$, die die obige Gleichung erfüllen, gilt

$$Y = Y_1 \quad (P)\text{-fast sicher.}$$

Jede $\mathcal{C}$ - $\bar{\mathcal{B}}$ -messbare numerische Zufallsvariable $Z : \Omega \rightarrow \bar{\mathbb{R}}$, für die (P) -fast sicher $Z = Y$ gilt, wird mit $\mathbb{E}(X|\mathcal{C})$ bezeichnet und heißt die bedingte Erwartung von X unter $\mathcal{C}$. Daher ist die Zufallsvariable $\mathbb{E}(X|\mathcal{C})$ nicht eindeutig festgelegt. Ein fest gewähltes Z nennt man Version von $\mathbb{E}(X|\mathcal{C})$.

Ausgehend von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ und einem Messraum $(\Omega', \mathcal{S}')$ betrachten wir eine (P) -integrierbare numerische Zufallsvariable $X : \Omega \rightarrow \bar{\mathbb{R}}$ und n Zufallsvariablen $Z_1, \dots, Z_n : \Omega \rightarrow \Omega'$.

Mit $\sigma(Z_1, \dots, Z_n)$ bezeichnen wir die kleinste unter allen σ -Algebren $\mathcal{C}$ über Ω , für

die die Zufallsvariablen $Z_1, \dots, Z_n$ $\mathcal{C}$ - $\mathcal{S}'$ -messbar sind. Interessiert man sich nun für die bedingte Erwartung $\mathbb{E}(X|\sigma(Z_1, \dots, Z_n))$, so schreibt man dafür $\mathbb{E}(X|Z_1, \dots, Z_n)$. Eigenschaften bedingter Erwartungen:

- a) $\mathbb{E}(\mathbb{E}(X|\mathcal{C})) = \mathbb{E}(X)$,
- b) Ist X $\mathcal{C}$ - $\bar{\mathcal{B}}$ -messbar, so folgt $\mathbb{E}(X|\mathcal{C}) = X$ (P -) fast sicher,
- c) Ist $X(\omega) = \alpha \in \mathbb{R}$ für alle $\omega \in \Omega$, so gilt $\mathbb{E}(X|\mathcal{C}) = \alpha$ (P -) fast sicher.

Ist nun $Z : \Omega \rightarrow \bar{\mathbb{R}}$ eine (P -)integrierbare numerische Zufallsvariable, so erhalten wir:

- a) Für alle $\alpha, \beta \in \mathbb{R}$ mit $\alpha X + \beta Z : \Omega \rightarrow \bar{\mathbb{R}}$ (d.h. nicht „ $\infty - \infty$ “) gilt:

$$\mathbb{E}(\alpha X + \beta Z|\mathcal{C}) = \alpha \mathbb{E}(X|\mathcal{C}) + \beta \mathbb{E}(Z|\mathcal{C}) \quad (P\text{-}) \text{ fast sicher,}$$

- b) Sei nun $X \leq Z$ (P -) fast sicher, dann gilt:

$$\mathbb{E}(X|\mathcal{C}) \leq \mathbb{E}(Z|\mathcal{C}) \quad (P\text{-}) \text{ fast sicher.}$$

Besonders wichtig ist die in der folgenden Definition festgelegte Klasse von bedingten Erwartungen.

Definition 12.24 (bedingte Wahrscheinlichkeit) Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und A eine beliebige Menge aus der σ -Algebra $\mathcal{S}$, dann wird für jede σ -Algebra $\mathcal{C} \subseteq \mathcal{S}$ die bedingte Erwartung $\mathbb{E}(I_A|\mathcal{C})$ als bedingte Wahrscheinlichkeit bezeichnet und in der Form $P(A|\mathcal{C})$ dargestellt.

Seien nun $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{S}')$ ein Messraum und $Y : \Omega \rightarrow \Omega'$ eine Zufallsvariable derart, dass es ein $\omega' \in \Omega'$ gibt mit $P(B) > 0$, wobei $B := \{\omega \in \Omega; Y(\omega) = \omega'\}$, so gilt für jede Menge $A \in \mathcal{S}$ und für alle reelle Versionen von $P(A|Y)$:

$$P(A|Y)(\omega) = P^B(A) \text{ für alle } \omega \in B.$$

Diese Gleichung rechtfertigt den in Definition 2.24 gewählten Namen. Zudem ist es nun prinzipiell möglich, Wahrscheinlichkeiten der Form $P^B(A)$ für Mengen $B := \{\omega \in \Omega; Y(\omega) = \omega'\}$ mit $P(B) = 0$ zu untersuchen:

$$P^B(A) := P(A|Y = \omega') := P(A|Y)(\omega) \text{ für alle } \omega \in B.$$

Analog schreiben wir für bedingte Erwartungen:

$$\mathbb{E}(A|Y = \omega') := \mathbb{E}(A|Y)(\omega) \text{ für alle } \omega \in B.$$

Dass diese Festlegungen sinnvoll sind, wird im Rahmen der Wahrscheinlichkeitstheorie bewiesen.

Kapitel 13

Mathematische Statistik

13.1 Einführung

Hat man ein Zufallsexperiment durch einen Wahrscheinlichkeitsraum modelliert, so stellt die Wahrscheinlichkeitstheorie Hilfsmittel bereit, um bei bekanntem Wahrscheinlichkeitsraum Aussagen über den Ablauf des zugrundeliegenden Zufallsexperimentes machen zu können. Die mathematische Statistik behandelt die folgende Problemstellung: Das zu modellierende Zufallsexperiment wird zunächst durch einen unvollständigen Wahrscheinlichkeitsraum beschrieben. Bei dieser Beschreibung werden die Grundmenge Ω , die σ -Algebra $\mathcal{S}$ und eine Menge von Wahrscheinlichkeitsmaßen auf $\mathcal{S}$ festgelegt. Dabei wird die Menge der in Frage kommenden Wahrscheinlichkeitsmaße häufig durch einen Parameter θ aus einem Parameterraum Θ dargestellt. Die Menge aller eindimensionalen Normalverteilungen kann zum Beispiel durch den Parameterraum

$$\theta = (\mu, \sigma^2)^T \in \mathbb{R} \times \mathbb{R}^+ = \Theta$$

dargestellt werden. Um nun zu einer vollständigen mathematischen Beschreibung unseres Zufallsexperimentes zu kommen, müssen wir uns für ein Wahrscheinlichkeitsmaß P aus der Menge der in Frage kommenden Wahrscheinlichkeitsmaße entscheiden. Ein wesentliches Kriterium der mathematischen Statistik besteht nun darin, dass eine Entscheidung über die Wahl des Wahrscheinlichkeitsmaßes beziehungsweise über die Verkleinerung der Menge aller in Frage kommenden Wahrscheinlichkeitsmaße von Ergebnissen des Zufallsexperimentes abhängt. Somit kann die Problemstellung der mathematischen Statistik als Umkehrung der Problemstellung der Wahrscheinlichkeitstheorie aufgefasst werden. Dabei werden die Ergebnisse des Zufallsexperimentes häufig nicht unmittelbar für die zu treffende Entscheidung verwendet, sondern die von diesen Ergebnissen abhängigen Werte einer Zufallsvariablen X definiert auf Ω . Wir erhalten somit die folgende Ausgangssituation:

Gegeben ist ein Tripel $(\Omega, \mathcal{S}, \mathcal{W})$ bestehend aus einer Grundmenge Ω , einer σ -Algebra $\mathcal{S}$ und einer Menge von Wahrscheinlichkeitsmaßen $\mathcal{W}$ auf $\mathcal{S}$. Ist diese Menge $\mathcal{W}$ durch einen Parameter $\theta \in \Theta$ beschrieben, so schreiben wir $(\Omega, \mathcal{S}, \mathcal{W}_{\theta \in \Theta})$. Ferner sind ein Messraum $(\Psi, \mathcal{G})$, eine Zufallsvariable $X : \Omega \rightarrow \Psi$ und der Funktionswert $X(\hat{\omega})$ der Zufallsvariable X für mindestens ein beobachtetes Ergebnis $\hat{\omega} \in \Omega$ des

zugrundegelegten Zufallsexperimentes gegeben. Basierend auf $X(\hat{\omega})$ soll nun unter verschiedenen weiteren Vorgaben eine Entscheidung für die Wahl des Wahrscheinlichkeitsmaßes $P \in \mathcal{W}$ auf $\mathcal{S}$ ermöglicht oder zumindest vereinfacht werden. Wir werden die folgenden Problemstellungen untersuchen:

Punktschätzung

Unter der Annahme, dass unsere Menge $\mathcal{W}$ von Wahrscheinlichkeitsmaßen durch einen Parameter $\theta \in \Theta$ dargestellt ist, soll für eine nichtleere Menge Γ und eine vorgegebene Funktion $\gamma : \Theta \rightarrow \Gamma$ ein Funktionswert $\hat{\gamma}$ von γ ermittelt werden. Durch die Wahl von $\hat{\gamma}$ wird $\mathcal{W}_{\theta \in \Theta}$ auf $\mathcal{W}_{\theta \in \{\tau \in \Theta; \gamma(\tau) = \hat{\gamma}\}}$ reduziert. Für die Menge aller eindimensionalen Normalverteilungen mit

$$\theta = (\mu, \sigma^2)^T \in \mathbb{R} \times \mathbb{R}^+ = \Theta$$

könnte die Funktion γ zum Beispiel in der Projektion auf die erste Komponente bestehen:

$$\gamma : \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R} (= \Gamma), \quad (\mu, \sigma^2) \mapsto \mu.$$

Man interessiert sich also nur für die Festlegung des Erwartungswertes.

Bereichsschätzung

Im Gegensatz zur Punktschätzung begnügt man sich bei der Bereichsschätzung damit, die Menge $\mathcal{W}_{\theta \in \Theta}$ der möglichen Wahrscheinlichkeitsmaße durch die Wahl einer speziellen Teilmenge $B \subset \Gamma$ auf die Menge $\mathcal{W}_{\theta \in \{\tau \in \Theta; \gamma(\tau) \in B\}}$ zu reduzieren.

Testtheorie

In der Testtheorie betrachtet man eine Partition $\mathcal{W}_0, \mathcal{W}_1$ unserer Menge von Wahrscheinlichkeitsmaßen $\mathcal{W}$. Das Ergebnis eines Tests ist eine Entscheidung darüber, ob die Menge der zur Diskussion stehenden Wahrscheinlichkeitsmaße $\mathcal{W}$ auf $\mathcal{W}_0$ oder $\mathcal{W}_1$ reduziert wird.

Durch die Zufallsvariable $X : \Omega \rightarrow \Psi$ erhalten wir zu jedem $P \in \mathcal{W}$ ein Wahrscheinlichkeitsmaß P_X auf $\mathcal{G}$, das Bildmaß von X . Die zu $\mathcal{W}$ gehörige Menge aller Bildmaße P_X auf $\mathcal{G}$ bezeichnen wir mit $P_{X, \mathcal{W}}$ beziehungsweise mit $P_{X, \mathcal{W}_{\theta \in \Theta}}$, falls $\mathcal{W}$ durch einen Parameter $\theta \in \Theta$ dargestellt wird. Da wir nicht die Beobachtung $\hat{\omega} \in \Omega$ als Basis unserer Überlegungen gewählt haben, sondern $X(\hat{\omega})$, ist es sinnvoll, Aussagen über Wahrscheinlichkeitsmaße $P \in \mathcal{W}$ (auf $\mathcal{S}$) auf Aussagen über Wahrscheinlichkeitsmaße $P_X \in P_{X, \mathcal{W}}$ zu verlagern.

Definition 13.1 (Stichprobe(nraum), stat. Raum, Realisierung) Seien $(\Omega, \mathcal{S})$ ein Messraum, $\mathcal{W}$ eine Menge von Wahrscheinlichkeitsmaßen auf $\mathcal{S}$ und $(\Omega, \mathcal{S}, \mathcal{W})$ die unvollständige Beschreibung eines Zufallsexperimentes gemäß der obigen Motivation. Seien ferner $\hat{\omega} \in \Omega$ ein Ergebnis dieses Zufallsexperimentes, $(\Psi, \mathcal{G})$ ein Messraum und $X : \Omega \rightarrow \Psi$ eine Zufallsvariable (also $\mathcal{S}$ - $\mathcal{G}$ -messbar), dann heißt die Menge Ψ Stichprobenraum, und der Wert $\bar{x} = X(\hat{\omega})$ Stichprobe oder Realisierung von X . Ist nun $P_{X, \mathcal{W}}$ die Menge aller Bildmaße von X in Abhängigkeit von $\mathcal{W}$, so heißt das Tupel $(\Psi, \mathcal{G}, P_{X, \mathcal{W}})$ statistischer Raum.

Im folgenden Beispiel sollen die bisherigen Begriffe eingeübt und dabei auch auf einen zentralen Begriff der mathematischen Statistik hingeführt werden.

Beispiel: Herstellung von Glühbirnen

Eine Firma stellt Glühbirnen her, wobei jede Glühbirne mit einer festen Wahrscheinlichkeit $\theta \in]0, 1[$ defekt ist. In einem großen Lager werden M Glühbirnen aufbewahrt. Ein Kunde bestellt K ($K < M$) Glühbirnen, die aus dem Lager entnommen und ausgeliefert werden. Bei der Verwendung jeder dieser K Glühbirnen wird vom Kunden notiert, ob sie defekt ($\hat{= 1}$) ist oder nicht ($\hat{= 0}$). Die Firma hätte gerne aufgrund der Erfahrungen des Kunden eine Schätzung, mit welcher Wahrscheinlichkeit eine Glühbirne defekt ist. Das im Lager befindliche M -Tupel von Glühbirnen modellieren wir durch einen binären Vektor $\omega \in \Omega = \{0, 1\}^M$, wobei wir uns die M Glühbirnen als durchnummeriert vorstellen und $\omega_i = 1$ bedeutet, dass die i -te Glühbirne defekt ist (also ist für $\omega_i = 0$ die i -te Glühbirne in Ordnung). Als σ -Algebra $\mathcal{S}$ auf $\{0, 1\}^M$ können wir die Potenzmenge $\mathcal{P}(\{0, 1\}^M)$ verwenden. Unter der Annahme, dass die Zustände der einzelnen Glühbirnen stochastisch unabhängig sind, erhalten wir als mögliche Wahrscheinlichkeitsmaße:

$$\mathcal{W}_{\theta \in]0, 1[}(\omega) = \prod_{i=1}^M \theta^{\omega_i} (1 - \theta)^{(1-\omega_i)}.$$

Nun steht uns allerdings kein Ergebnis dieses Zufallsexperimentes unmittelbar zur Verfügung, da wir ja nur die K ausgelieferten Glühbirnen beobachten können. Seien nun $1 \leq i_1 < \dots < i_K \leq M$ die Nummern derjenigen K Glühbirnen, die ausgeliefert wurden, so können wir mit $\Psi = \{0, 1\}^K$ und $\mathcal{G} = \mathcal{P}(\{0, 1\}^K)$ die Zufallsvariable

$$X : \Omega \rightarrow \Psi, \quad \omega \mapsto (\omega_{i_1}, \dots, \omega_{i_K}) =: x = (x_1, \dots, x_K)$$

angeben. Für die entsprechenden Bildmaße erhalten wir:

$$P_{X, \mathcal{W}_{\theta \in]0, 1[}}(x) = \prod_{i=1}^K \theta^{x_i} (1 - \theta)^{(1-x_i)}$$

Eine Realisierung unserer Zufallsvariablen X ist somit durch einen binären Vektor $\bar{x} \in \Psi$ gegeben. Dieser Vektor ist die einzige Information, die wir zur Schätzung

von θ zur Verfügung haben. Da aber

$$P_{X, \mathcal{W}_{\theta \in]0,1[}}(x) = \prod_{i=1}^K \theta^{x_i} (1-\theta)^{(1-x_i)} = \theta^{\sum_{i=1}^K x_i} (1-\theta)^{K - \sum_{i=1}^K x_i},$$

scheint es zu genügen, sich statt $x \in \{0,1\}^K$ nur die Zahl $\sum_{i=1}^K x_i$ zu merken. Die entscheidende Frage lautet nun: Sind alle Informationen, die in $x \in \{0,1\}^K$ über den unbekannten Parameter θ enthalten sind, auch in der Zahl $\sum_{i=1}^K x_i$ enthalten? In diesem Fall nennt man die Abbildung

$$T : \Psi = \{0,1\}^K \rightarrow \Omega_T = \{0,1,\dots,K\}, \quad x \mapsto \sum_{i=1}^K x_i$$

eine suffiziente Statistik für θ .

Rein intuitiv würde man einen Schätzwert $\hat{\theta}$ für θ folgendermaßen berechnen:

$$\hat{\theta} = \frac{\sum_{i=1}^K \bar{x}_i}{K}.$$

Wie könnte man nun entscheiden, ob alle Informationen über θ bereits in der Abbildung T enthalten sind? Zur Beantwortung dieser Frage stellen wir zunächst fest, dass T $\mathcal{P}(\{0,1\}^K)$ - $\mathcal{P}(\{0,1,\dots,K\})$ -messbar ist. Somit können wir für jede Menge

$$D_t := \{x \in \{0,1\}^K; T(x) = t\}, \quad t \in \{0,1,\dots,K\},$$

die bedingten Wahrscheinlichkeiten

$$P_{X, \mathcal{W}_{\theta}}^{D_t}(x) = \frac{P_{X, \mathcal{W}_{\theta}}(x \cap D_t)}{P_{X, \mathcal{W}_{\theta}}(D_t)}$$

für $\theta \in]0,1[$ berechnen. Es gilt:

$$P_{X, \mathcal{W}_{\theta}}^{D_t}(x) = \begin{cases} 0 & \text{für alle } x \text{ mit } \sum_{i=1}^K x_i \neq t \\ \frac{\theta^t (1-\theta)^{(K-t)}}{\binom{K}{t} \theta^t (1-\theta)^{(K-t)}} = \frac{1}{\binom{K}{t}} & \text{für alle } x \text{ mit } \sum_{i=1}^K x_i = t \end{cases}.$$

Entscheidend ist nun die Tatsache, dass die verschiedenen Werte $P_{X, \mathcal{W}_{\theta}}^{D_t}(x)$ für alle $x \in \{0,1\}^K$ und alle $t \in \{0,1,\dots,K\}$ nicht mehr von θ abhängen. Diesen Sachverhalt interpretieren wir dahingehend, dass die Beobachtung $T(x)$ hinreichend (suffizient) dafür ist, jede Information über θ zu erhalten, die man aus der Stichprobe $\bar{x} = X(\hat{\omega})$ entnehmen kann.

13.2 Punkt- und Bereichsschätzung

In diesem Abschnitt gehen wir von einem statistischen Raum $(\Psi, \mathcal{G}, P_{X, \mathcal{W}_{\theta \in \Theta}})$ aus und betrachten die Aufgabe, aufgrund einer Stichprobe $\bar{x} \in \Psi$ eine Schätzung für $\gamma(\theta)$ mit $\gamma : \Theta \rightarrow \Gamma$ zu berechnen und Kriterien für die Güte dieser Schätzung anzugeben. Motiviert durch das obige Beispiel definieren wir:

Definition 13.2 ((suffiziente)Statistik) Seien $(\Psi, \mathcal{G}, P_{X, \mathcal{W}_{\theta \in \Theta}})$ ein statistischer Raum, $(\Omega_T, \mathcal{S}_T)$ ein Messraum und $T : \Psi \rightarrow \Omega_T$ eine $\mathcal{G}$ - $\mathcal{S}_T$ -messbare Abbildung, dann wird die Abbildung T als Statistik bezeichnet. Die Statistik T heißt *suffizient* für $\theta \in \Theta$, falls es für jedes $A \in \mathcal{G}$ und für jedes $\theta \in \Theta$ eine Version von $P_{X, \mathcal{W}_{\theta}}(A|T)$ gibt, die nicht von θ abhängt. In diesem Fall heißt die von der Statistik T erzeugte σ -Algebra $\sigma(T) \in \mathcal{G}$ *suffizient* für $\theta \in \Theta$.

Das Auffinden suffizienter Statistiken ist im allgemeinen keine einfache Aufgabe. Im folgenden betrachten wir einen wichtigen Spezialfall:

Exponentialfamilien:

Seien $(\Psi, \mathcal{G}, P_{X, \mathcal{W}_{\theta \in \Theta}})$ ein statistischer Raum, μ ein σ -endliches Maß auf $\mathcal{G}$ und $T_1, \dots, T_n : \Psi \rightarrow \mathbb{R}$ $\mathcal{G}$ - $\mathcal{B}$ -messbare Funktionen. Seien ferner $C : \Theta \rightarrow \mathbb{R}^+$, $\xi_1, \dots, \xi_n : \Theta \rightarrow \mathbb{R}$ reelle Abbildungen und $h : \Psi \rightarrow \mathbb{R}_0^+$ $\mathcal{G}$ - $\mathcal{B}$ -messbar, dann heißt die Familie $P_{X, \mathcal{W}_{\theta \in \Theta}}$ von Wahrscheinlichkeitsmaßen eine Exponentialfamilie, wenn für jedes $\theta \in \Theta$ durch

$$f_{X, \theta} : \Psi \rightarrow \mathbb{R}_0^+, \quad x \mapsto C(\theta) h(x) \exp \left(\sum_{j=1}^n \xi_j(\theta) T_j(x) \right)$$

eine μ -Dichte von $P_{X, \mathcal{W}_{\theta}}$ gegeben ist.

Entscheidend ist nun, dass bei Exponentialfamilien die Statistik

$$T : \Psi \rightarrow \mathbb{R}^n, \quad x \mapsto (T_1(x), \dots, T_n(x))^T$$

suffizient ist für $\theta \in \Theta$. Ist nun $P_{X, \mathcal{W}_{\theta \in \Theta}}$ eine Exponentialfamilie mit entsprechender suffizienter Statistik T für $\theta \in \Theta$ und betrachtet man statt der Zufallsvariable X einen Messraum $(\Psi^p, \sigma(\{A_1 \times \dots \times A_p; A_i \in \mathcal{G}, i = 1, \dots, p\}))$ und eine Zufallsvariable

$$\hat{X} : \Omega \rightarrow \Psi^p, \quad \omega \mapsto (X_1(\omega), \dots, X_p(\omega))^T, \quad p \in \mathbb{N},$$

wobei $X_1, \dots, X_p$ stochastisch unabhängige Zufallsvariablen sind mit

$$P_{X, \mathcal{W}_{\theta}} = P_{X_i, \mathcal{W}_{\theta}} \text{ für alle } \theta \in \Theta \text{ und alle } i \in \{1, \dots, p\},$$

so erhält man einen neuen statistischen Raum $(\Psi^p, \sigma(\{A_1 \times \dots \times A_p; A_i \in \mathcal{G}, i = 1, \dots, p\}), P_{\hat{X}, \mathcal{W}_{\theta \in \Theta}})$ und eine suffiziente Statistik

$$\hat{T} : \Psi^p \rightarrow \mathbb{R}^n, \quad \hat{x} \mapsto \left(\sum_{i=1}^p T_1(\hat{x}_i), \dots, \sum_{i=1}^p T_n(\hat{x}_i) \right)^T$$

für $\theta \in \Theta$, wobei T für $i = 1, \dots, p$ die suffiziente Statistik der Exponentialfamilie $P_{X_i, \mathcal{W}_{\theta \in \Theta}} = P_{X, \mathcal{W}_{\theta \in \Theta}}$ für $\theta \in \Theta$ ist.

Beispiel 1: Glühbirnen

Seien $\Psi = \{0, 1\}$, $\mathcal{G} = \mathcal{P}(\{0, 1\})$ und $P_{X, \mathcal{W}_{\theta}}$ für jedes $\theta \in]0, 1[$ gegeben durch

$$P_{X, \mathcal{W}_{\theta}}(x) = f_{X, \theta}(x) = \theta^x (1 - \theta)^{1-x} = (1 - \theta) \exp \left(x \ln \left(\frac{\theta}{1 - \theta} \right) \right),$$

so ist $T : \{0, 1\} \rightarrow \mathbb{R}, x \mapsto x$ eine suffiziente Statistik für $\theta \in]0, 1[$. Mit

$$\hat{X} : \Omega \rightarrow \Psi^K, \quad \omega \mapsto (X_1(\omega), \dots, X_K(\omega))^T,$$

wobei $X_1, \dots, X_K$ stochastisch unabhängige Zufallsvariablen sind, für die gilt:

$$P_{X, \mathcal{W}_{\theta}} = P_{X_i, \mathcal{W}_{\theta}} \text{ für alle } \theta \in]0, 1[\text{ und alle } i \in \{1, \dots, K\},$$

erhalten wir eine suffiziente Statistik

$$\hat{T} : \{0, 1\}^K \rightarrow \mathbb{R}, \hat{x} \mapsto \sum_{i=1}^K \hat{x}_i$$

für $\theta \in]0, 1[$. Es gilt:

$$P_{\hat{X}, \mathcal{W}_{\theta}}(\hat{x}) = \prod_{i=1}^K \theta^{\hat{x}_i} (1 - \theta)^{(1 - \hat{x}_i)} = \theta^{\sum_{i=1}^K \hat{x}_i} (1 - \theta)^{K - \sum_{i=1}^K \hat{x}_i}.$$

Beispiel 2: Normalverteilung

Seien $\Psi = \mathbb{R}$, $\mathcal{G} = \mathcal{B}$ und $P_{X, \mathcal{W}_{\theta}}$ für jedes $\theta \in \mathbb{R} \times \mathbb{R}^+$ gegeben durch die Lebesgue-Dichte

$$\begin{aligned} f_{X, \theta}(x) &= \frac{1}{\sqrt{2\pi\theta_2}} \exp \left(-\frac{(x - \theta_1)^2}{2\theta_2} \right) = \\ &= \frac{1}{\sqrt{2\pi\theta_2}} \exp \left(-\frac{\theta_1^2}{2\theta_2} \right) \exp \left(\frac{\theta_1}{\theta_2} x - \frac{1}{2\theta_2} x^2 \right) = \\ &= C(\theta) h(x) \exp \left(\sum_{j=1}^2 \xi_j(\theta) T_j(x) \right), \end{aligned}$$

so ist $T : \mathbb{R} \rightarrow \mathbb{R} \times \mathbb{R}_0^+, x \mapsto (x, x^2)^T$ eine suffiziente Statistik für $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$. Mit

$$\hat{X} : \Omega \rightarrow \mathbb{R}^p, \quad \omega \mapsto (X_1(\omega), \dots, X_p(\omega))^T,$$

wobei $X_1, \dots, X_p$ stochastisch unabhängige Zufallsvariablen sind, für die gilt:

$$P_{X, \mathcal{W}_{\theta}} = P_{X_i, \mathcal{W}_{\theta}} \text{ für alle } \theta \in \mathbb{R} \times \mathbb{R}^+ \text{ und alle } i \in \{1, \dots, p\},$$

erhalten wir eine suffiziente Statistik

$$\hat{T} : \mathbb{R}^p \rightarrow \mathbb{R} \times \mathbb{R}_0^+, \hat{x} \mapsto \left(\sum_{i=1}^p \hat{x}_i, \sum_{i=1}^p \hat{x}_i^2 \right)^T$$

für $\theta \in \mathbb{R} \times \mathbb{R}^+$, wobei $P_{\hat{X}, \mathcal{W}_\theta}$ eine p -dimensionale $\mathcal{N}((\mu, \dots, \mu)^T, \sigma^2 E_p)$ Normalverteilung darstellt (E_p bezeichnet die p -dimensionale Einheitsmatrix). Somit muss man sich statt des p -dimensionalen reellen Vektors $\hat{x}$ nur die zwei reellen Zahlen $\left(\sum_{i=1}^p \hat{x}_i, \sum_{i=1}^p \hat{x}_i^2\right)$ merken, ohne Informationen über θ zu verlieren. Dies gilt überraschenderweise für alle $p \in \mathbb{N}$.

Nach diesen Vorbereitungen sind wir nun in der Lage, ausgehend von einem statistischen Raum $(\Psi, \mathcal{G}, P_{X, \mathcal{W}_{\theta \in \Theta}})$ und basierend auf einer Realisierung $\bar{x} \in \Psi$ die Aufgabe zu untersuchen, einen Schätzer $\gamma(\theta)$ für $\gamma : \Theta \rightarrow \Gamma$ anzugeben. Wir untersuchen den Spezialfall $\Gamma = \mathbb{R}$. Sei nun $g : \Psi \rightarrow \mathbb{R}$ eine $\mathcal{G}$ - $\mathcal{B}$ -messbare Funktion, so nennt man g eine Schätzfunktion für $\gamma(\theta)$ und $g(\bar{x})$ einen Schätzer für $\gamma(\theta)$. Die Schätzfunktion g heißt erwartungstreu, falls

$$\mathbb{E}_\theta(g) := \int g dP_{X, \mathcal{W}_\theta} = \gamma(\theta) \text{ für alle } \theta \in \Theta.$$

Es kann passieren, dass es für $\gamma(\theta)$ keine erwartungstreue Schätzfunktion gibt; in diesem Fall wird $\gamma(\theta)$ als nicht schätzbar bezeichnet. Ansonsten heißt $\gamma(\theta)$ schätzbar. Wir gehen im folgenden immer von schätzbaren $\gamma(\theta)$ aus und wollen nun überlegen, wie man zwei gegebene erwartungstreue Schätzfunktionen g_1 und g_2 für $\gamma(\theta)$ vergleichen kann. Ein Maß für die Abweichung einer Zufallsvariable von ihrem Erwartungswert ist die Varianz dieser Zufallsvariable. Da der Erwartungswert einer erwartungstreuen Schätzfunktion gerade die zu schätzende Größe ist, ist g_1 besser als g_2 , falls die Varianz von g_1 kleiner ist als die Varianz von g_2 . Der folgende Satz beschreibt nun die Möglichkeit, mit Hilfe suffizienter Statistiken zu besseren erwartungstreuen Schätzfunktionen zu kommen.

Satz 13.3 (Rao-Blackwell) Seien $(\Psi, \mathcal{G}, P_{X, \mathcal{W}_{\theta \in \Theta}})$ ein statistischer Raum, $(\Omega_T, \mathcal{S}_T)$ ein Messraum, $\gamma : \Theta \rightarrow \mathbb{R}$ eine Abbildung und $g : \Psi \rightarrow \mathbb{R}$ eine erwartungstreue Schätzfunktion für $\gamma(\theta)$. Sei ferner $T : \Psi \rightarrow \Omega_T$ eine suffiziente Statistik für $\theta \in \Theta$, so ist die Schätzfunktion

$$h : \Psi \rightarrow \mathbb{R}, \quad x \mapsto \mathbb{E}_\theta(g|T = T(x))$$

ebenfalls erwartungstreu (wobei für die bedingte Erwartung $\mathbb{E}_\theta$ das Wahrscheinlichkeitsmaß $P_{X, \mathcal{W}_\theta}$ herangezogen wird) und es gilt:

$$\mathbb{V}_\theta(g) \geq \mathbb{V}_\theta(h).$$

Beispiel: Glühbirnen

Kehren wir nun zu unserem Beispiel über die Herstellung von Glühbirnen mit dem statistischen Raum $(\{0, 1\}^K, \mathcal{P}(\{0, 1\}^K), P_{X, \mathcal{W}_{\theta \in [0, 1]}})$ zurück und betrachten zwei Schätzfunktionen

$$g : \{0, 1\}^K \rightarrow \mathbb{R}, x \mapsto x_1$$

und mit $T : \{0, 1\}^K \rightarrow \{0, \dots, K\}$, $x \mapsto \sum_{i=1}^K x_i$:

$$h : \{0, 1\}^K \rightarrow \mathbb{R}, \quad x \mapsto \mathbb{E}_\theta(g|T = T(x))$$

Nach einfacher Rechnung erhält man:

$$h(x) = \frac{\sum_{i=1}^K x_i}{K}.$$

Für die Varianzen folgt somit:

$$\mathbb{V}_\theta(g) = \theta(1 - \theta) \geq \mathbb{V}_\theta(h) = \frac{\theta(1 - \theta)}{K}.$$

Eines der wichtigsten Verfahren zur Punktschätzung ist das Maximum-Likelihood-Verfahren. Ausgangspunkt ist ein statistischer Raum $(\Psi, \mathcal{G}, P_{X, \mathcal{W}_{\theta \in \Theta}})$, eine Stichprobe $\bar{x}$, ein σ -endliches Maß μ auf $\mathcal{G}$ und die Forderung, dass $P_{X, \mathcal{W}_\theta}$ für jedes $\theta \in \Theta$ absolutstetig bezüglich μ ist mit der Dichte $f_{X, \theta} : \Psi \rightarrow \mathbb{R}_0^+$. Ist nun $\gamma : \Theta \rightarrow \Gamma$ bijektiv, so berechnet man ein $\hat{\theta} \in \Theta$ mit:

$$f_{X, \hat{\theta}}(\bar{x}) \geq f_{X, \theta}(\bar{x}) \text{ für alle } \theta \in \Theta$$

und verwendet den Wert $\gamma(\hat{\theta})$ als Maximum-Likelihood-Schätzer für $\gamma(\theta)$. Maximum-Likelihood-Schätzer müssen nicht existieren. Die Berechnung von $\hat{\theta}$ führt auf das Gebiet der mathematischen Optimierung. Die statistische Analyse von Maximum-Likelihood-Schätzern ist ein äusserst schwieriges Problem.

Beispiel 1: Glühbirnen

Für

$$f_{X, \theta}(x) = \theta^{\sum_{i=1}^K x_i} (1 - \theta)^{K - \sum_{i=1}^K x_i}$$

ergibt sich

$$\hat{\theta} = \frac{\sum_{i=1}^K \bar{x}_i}{K}$$

Beispiel 2: Normalverteilung

Sei

$$f_{X, \theta}(x) = \prod_{i=1}^p \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x_i - \theta_1)^2}{2\theta_2}\right),$$

so ergibt sich:

$$\hat{\theta}_1 = \frac{\sum_{i=1}^p \bar{x}_i}{p} \quad \hat{\theta}_2 = \frac{\sum_{i=1}^p (\bar{x}_i - \hat{\theta}_1)^2}{p}.$$

Ist die zugehörige Schätzfunktion erwartungstreu?

Die Berechnung von Bereichsschätzern wollen wir an einem Beispiel demonstrieren.

Beispiel: Normalverteilung

Sei

$$f_{X,\theta}(x) = \prod_{i=1}^p \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_i - \theta)^2}{2}\right).$$

Gesucht ist ein $a \in \mathbb{R}_0^+$ derart, dass für $0 \leq c \leq 1$ gilt:

$$P_{X,\mathcal{W}_\theta} \left(\left\{ x \in \Psi; \theta \in \left[\frac{1}{p} \sum_{i=1}^p x_i - a, \frac{1}{p} \sum_{i=1}^p x_i + a \right] \right\} \right) \geq c$$

für alle $\theta \in \mathbb{R}$. Dies ist äquivalent zu

$$P_{X,\mathcal{W}_\theta} \left(\left\{ x \in \Psi; -a\sqrt{p} \leq \frac{1}{\sqrt{p}} \sum_{i=1}^p x_i - \sqrt{p}\theta \leq a\sqrt{p} \right\} \right) \geq c.$$

Da aber die Zufallsvariable

$$Z_p : \Psi \rightarrow \mathbb{R}, x \mapsto \frac{1}{\sqrt{p}} \sum_{i=1}^p x_i - \sqrt{p}\theta$$

$\mathcal{N}(0,1)$ normalverteilt ist für alle $p \in \mathbb{N}$, erhalten wir für a :

$$\Phi(a\sqrt{p}) - \Phi(-a\sqrt{p}) \geq c$$

beziehungsweise

$$\Phi(a\sqrt{p}) \geq \frac{1+c}{2} \quad (\text{siehe Tabelle Seite 254}).$$

Es ergibt sich mit der Realisierung $\bar{x}$ und mit dem aus der obigen Ungleichung bestimmten a :

$$\theta \in \left[\frac{1}{p} \sum_{i=1}^p \bar{x}_i - a, \frac{1}{p} \sum_{i=1}^p \bar{x}_i + a \right].$$

13.3 Testtheorie

Die Testtheorie ist ein wichtiges und umfangreiches Teilgebiet der mathematischen Statistik; daher können wir nur auf Grundideen eingehen. Ausgangspunkt ist ein statistischer Raum $(\Psi, \mathcal{G}, P_{X,\mathcal{W}_{\theta \in \Theta}})$ und eine Partition Θ_0, Θ_1 von Θ (also: $\Theta_0, \Theta_1 \neq \emptyset$, $\Theta_0 \cap \Theta_1 = \emptyset$ und $\Theta_0 \cup \Theta_1 = \Theta$). Basierend auf einer Stichprobe $\bar{x} \in \Psi$ soll nun entschieden werden, ob Θ auf Θ_1 oder Θ_2 reduziert wird. In der Testtheorie wird diese Fragestellung durch die Entscheidung zwischen einer Nullhypothese

$$H_0 : \theta \in \Theta_0$$

und einer Gegenhypothese

$$H_1 : \theta \in \Theta_1$$

formuliert. Zu bestimmen ist also eine $\mathcal{G}\text{-}\mathcal{P}(\{0, 1\})$ -messbare Entscheidungsfunktion $\delta : \Psi \rightarrow \{0, 1\}$, die jeder möglichen Stichprobe $\bar{x}$ eine Entscheidung für Θ_0 (also: $\delta(\bar{x}) = 0$) oder für Θ_1 (also: $\delta(\bar{x}) = 1$) zuordnet. Ein Test ist natürlich dann festgelegt, wenn eine Menge $\bar{K} \subseteq \Theta$ mit

$$\bar{K} = \{x \in \Psi; \delta(x) = 1\}.$$

festgelegt ist. Um nun die Menge $\bar{K}$ bestimmen zu können, benötigt man eine Vorstellung, wie man die Güte eines Tests quantifizieren kann. Dazu betrachtet man den Fehler 1. Art

$$P_{X, \mathcal{W}_\theta}(\{x \in \Psi; x \in \bar{K}\}), \quad \theta \in \Theta_0,$$

und verbunden damit das Testniveau α :

$$\alpha := \sup_{\theta \in \Theta_0} P_{X, \mathcal{W}_\theta}(\{x \in \Psi; x \in \bar{K}\}).$$

Die Zahl $\alpha \in [0, 1]$ gibt die kleinste obere Schranke für die Wahrscheinlichkeit an, dass man sich für Θ_1 entscheidet, obwohl $\theta \in \Theta_0$ gilt. Es ist üblich, nur Tests mit dem selben Testniveau zu vergleichen. Ein Test ϕ_1 mit dem Testniveau α ist besser als ein Test ϕ_2 mit dem selben Testniveau, falls

$$P_{X, \mathcal{W}_\theta}(\{x \in \Psi; \delta_1(x) = 1\}) \geq P_{X, \mathcal{W}_\theta}(\{x \in \Psi; \delta_2(x) = 1\}) \text{ für alle } \theta \in \Theta_1.$$

Mit anderen Worten: Der Test ϕ_1 ist besser als der Test ϕ_2 , falls der Fehler 2. Art für ϕ_1 für jedes $\theta \in \Theta_1$ kleiner ist als der Fehler 2. Art für ϕ_2 :

$$P_{X, \mathcal{W}_\theta}(\{x \in \Psi; \delta_1(x) = 0\}) \leq P_{X, \mathcal{W}_\theta}(\{x \in \Psi; \delta_2(x) = 0\}) \text{ für alle } \theta \in \Theta_1.$$

Praktisch besteht nun die Vorgehensweise darin, zu einem gegebenen statistischen Raum $(\Psi, \mathcal{G}, P_{X, \mathcal{W}_{\theta \in \Theta}})$ und zu einem gewählten Testniveau α unter allen Tests vom Testniveau α denjenigen Test zu finden (also die Menge $\bar{K}$), der für alle $\theta \in \Theta_1$ den kleinsten Fehler 2. Art besitzt. Diesen Test nennt man gleichmäßig besten Niveau- α -Test. Natürlich stellen sich in diesem Zusammenhang wichtige Fragen, die allerdings im Rahmen dieser Vorlesung nicht diskutiert werden können:

- Gibt es zu einem gegebenen statistischen Raum und zu einem Testniveau α überhaupt einen gleichmäßig besten Test?
- Wenn es diesen gleichmäßig besten Test gibt, wie kann man dann die entsprechende Menge $\bar{K}$ finden?

Bevor wir für einen Spezialfall diese Fragen beantworten, soll auf einen wichtigen Punkt hingewiesen werden. Bei der Suche nach einem Testverfahren wird die kleinste obere Schranke für den Fehler 1. Art gewählt, während man mit dem entsprechenden Fehler 2. Art leben muss. Hier ist also eine Asymmetrie zwischen den

Hypothesen H_0 und H_1 erkennbar. Diese Asymmetrie ist gewollt, da sie häufig den praktischen Gegebenheiten entspricht. Soll zum Beispiel aufgrund von Messungen überprüft werden, ob es gefährliche Wechselwirkungen zwischen zwei Medikamenten gibt (Hypothese H_0 : Ja, Hypothese H_1 : Nein), so ist der Fehler 1. Art, also die Entscheidung, dass es diese Wechselwirkungen nicht gibt, obwohl sie existieren, viel gefährlicher als der Fehler 2. Art, also die Entscheidung, dass diese Wechselwirkungen vorhanden sind, obwohl sie nicht existieren. Doch nun zum bereits erwähnten Spezialfall:

Satz 13.4 ((gleichmäßig) beste Tests bei zweipunktigem Θ) Seien $(\Psi, \mathcal{G}, P_{X, \mathcal{W}_{\theta \in \{\theta_0, \theta_1\}}})$ ein statistischer Raum, μ ein σ -endliches Maß auf $\mathcal{G}$ und $P_{X, \mathcal{W}_{\theta_0}}$ sowie $P_{X, \mathcal{W}_{\theta_1}}$ absolutstetig bezüglich μ mit den Dichten $f_{X, \theta_0} : \Psi \rightarrow \mathbb{R}_0^+$ beziehungsweise $f_{X, \theta_1} : \Psi \rightarrow \mathbb{R}_0^+$, dann ist für jedes $k \in \mathbb{R}_0^+$ der Test

$$\delta : \Psi \rightarrow \{0, 1\}, x \mapsto \begin{cases} 1 & \text{für alle } x \text{ mit } f_{X, \theta_1}(x) > k f_{X, \theta_0}(x) \\ 0 & \text{für alle } x \text{ mit } f_{X, \theta_1}(x) \leq k f_{X, \theta_0}(x) \end{cases}$$

unter allen Tests vom Testniveau

$$\alpha = P_{X, \mathcal{W}_{\theta_0}}(\{x \in \Psi; f_{X, \theta_1}(x) > k f_{X, \theta_0}(x)\})$$

der (gleichmäßig) beste.

Beispiel: Glühbirnen

Kehren wir nun zu unserem Beispiel über die Herstellung von Glühbirnen mit dem statistischen Raum $(\{0, 1\}^K, \mathcal{P}(\{0, 1\}^K), P_{X, \mathcal{W}_{\theta \in [0.2, 0.8]}})$, mit $k=1$ und mit

$$f_{X, 0.2}(x) = 0.2^{\sum_{i=1}^K x_i} (1 - 0.2)^{K - \sum_{i=1}^K x_i}, \quad f_{X, 0.8}(x) = 0.8^{\sum_{i=1}^K x_i} (1 - 0.8)^{K - \sum_{i=1}^K x_i}$$

zurück, dann ist der Test

$$\delta : \Psi \rightarrow \{0, 1\}, x \mapsto \begin{cases} 1 & \text{für alle } x \text{ mit } \sum_{i=1}^K x_i > \frac{K}{2} \\ 0 & \text{für alle } x \text{ mit } \sum_{i=1}^K x_i \leq \frac{K}{2} \end{cases}$$

bester unter allen Tests mit Testniveau

$$\alpha = P_{X, \mathcal{W}_{0.2}} \left(\left\{ x \in \Psi; \sum_{i=1}^K x_i > \frac{K}{2} \right\} \right).$$

Für $K = 10$ erhalten wir mit $\alpha = 0.00642$ den Fehler 2. Art: 0.03279.

Für $K = 20$ erhalten wir mit $\alpha = 0.00056$ den Fehler 2. Art: 0.00259.