Kapitel 4

Allgemeine Likelihoodtheorie

Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Modell mit dem Stichprobenraum $(E, \mathscr{E})$. Die Wahrscheinlichkeitsverteilung der Stichprobe X unter dem Wahrscheinlichkeitsmaß $\mathbb{P}_\vartheta$ ist gegeben durch

$$\mathbb{P}_\vartheta^X(B) := \mathbb{P}_\vartheta(X \in B), \quad B \in \mathscr{E}.$$

Im Fall, daß man stochastische Prozesse studiert, ist der Stichprobenraum E in der Regel ein Funktionenraum, z.B. der Raum $\mathbb{C}\big([0,T]\big)$ aller stetigen Funktionen auf dem Beobachtungsintervall $[0,T]$.
In diesem allgemeinen Fall gibt es keine ausgezeichnete „gleichmäßige Verteilung“ wie das Lebesguemaß im $\mathbb{R}^n$, bezüglich der man Dichten $f_\vartheta^X(x)$ bilden kann. Einen Ausweg bietet das Theorem von Radon-Nikodym aus der Maßtheorie, das in bestimmten Fällen die Existenz von Funktionen sichert, die die Rolle von Dichten im klassischen Fall übernehmen.

4.1 Das Theorem von Radon-Nikodym

Wir beginnen mit einem fundamentalen Satz über die Darstellung eines Maßes μ als Integral über eine Funktion f bezüglich eines anderen Maßes ν, dem *Satz von Radon-Nikodym.*

Definition 4.1. Es seien μ und ν zwei σ-finite Maße auf einem meßbaren Raum $(F, \mathscr{F})$. Man sagt, *μ sei absolutstetig bezüglich* ν, falls für jedes $A \in \mathscr{F}$ mit $\nu(A) = 0$ auch $\mu(A) = 0$ gilt.
Symbolisch schreibt man dafür $\mu \ll \nu$. Offenbar folgt aus $\mu \ll \nu$ und $\nu \ll \lambda$ auch $\mu \ll \lambda$ (Transitivität). Gilt sowohl $\mu \ll \nu$ als auch $\nu \ll \mu$, so heißen μ und ν *äquivalent*, im Zeichen: $\mu \equiv \nu$.

Beispiel 4.1. Es sei $(F, \mathscr{F})$ ein meßbarer Raum. Ist f eine $\mathscr{F}$-meßbare reellwertige, nichtnegative Funktion auf F, so wird durch

$$\mu(A) := \int_A f(x)\nu(\,dx), \quad A \in \mathscr{F}$$

ein σ-finites Maß μ auf $\mathscr{F}$ definiert mit $\mu \ll \nu$.

Aussage 4.1 (Satz von Radon-Nikodym). *Es seien μ und ν zwei σ-finite Maße auf $(F, \mathscr{F})$. Ist $\mu \ll \nu$, d.h., folgt für alle $A \in \mathscr{F}$ mit $\nu(A) = 0$ auch $\mu(A) = 0$, so existiert eine nichtnegative Funktion f auf F mit folgenden Eigenschaften:*

1. *f ist $\mathscr{F}$ -meßbar,*
2. $\mu(A) = \int_A f(x)\nu(\,dx) \quad \forall A \in \mathscr{F}$.

Die Funktion f ist ν-f.ü. eindeutig bestimmt, d.h., für jede $\mathscr{F}$-meßbare Funktion g mit $\int_A f\,d\nu = \int_A g\,d\nu$, $\forall A \in \mathscr{F}$, gilt $f = g$ ν-f.ü.

Beweis:
Siehe Bauer, H. (1992) Maß- und Integrationstheorie, de Gruyter-Verlag, Kapitel 17. □

Die Funktion f aus dem Satz von Radon-Nikodym heißt *Radon-Nikodym-Ableitung von μ nach ν* und wird mit $\dfrac{d\mu}{d\nu}$ bezeichnet. Damit kann man die Eigenschaft 2. in der Aussage 4.1 schreiben als:

$$\mu(A) = \int_A \frac{d\mu}{d\nu}(x)\,\nu(\,dx), \quad A \in \mathscr{F}. \tag{4.1}$$

Eigenschaften der Radon-Nikodym-Ableitung

Es gilt (x ist hier stets Element aus F)

$$\text{Falls } \mu \ll \nu, \quad \text{so gilt} \quad \frac{d\mu}{d\nu}(x) > 0 \quad \mu\text{-f.ü.} \tag{4.2}$$

$$\text{Falls } \mu \ll \nu \text{ und } \nu \ll \lambda, \quad \text{so haben wir} \quad \frac{d\mu}{d\lambda}(x) = \frac{d\mu}{d\nu}(x) \cdot \frac{d\nu}{d\lambda}(x) \quad \lambda\text{-f.ü.} \tag{4.3}$$

$$\text{Gilt } \nu \equiv \mu, \quad \text{so ist} \quad \frac{d\nu}{d\mu}(x) = \left(\frac{d\mu}{d\nu}(x)\right)^{-1} \quad \mu\text{- und } \nu\text{-f.ü.} \tag{4.4}$$

Beweis:

1. $\mu\left(\{x : \frac{d\mu}{d\nu}(x) = 0\}\right) = \int_{\{\frac{d\mu}{d\nu}=0\}} \frac{d\mu}{d\nu}\, d\nu = 0$

2. Wegen $\nu(A) = \int_A \frac{d\nu}{d\lambda}(x)\lambda(\,dx), \quad A \in \mathscr{F},$ gilt für jede nichtnegative Funktion g : $\int_F g(x)\nu(\,dx) = \int_F g(x)\frac{d\nu}{d\lambda}(x)\lambda(\,dx).$

   (Man überlege sich die Gleichung für Indikatorfunktionen, für Linearkombinationen von Indikatorfunktionen und approximiere die erwähnten g durch monotone Folgen solcher Linearkombinationen. Danach wende man den Satz über monotone Konvergenz an.)
   Speziell für $g = \frac{d\mu}{d\nu}\mathbf{1}_A$ gilt: $\mu(A) = \int_A \frac{d\mu}{d\nu}(x)\nu(\,dx) = \int_A \frac{d\mu}{d\nu}(x)\frac{d\nu}{d\lambda}(x)\lambda(\,dx).$

   Die Behauptung folgt auf Grund der λ-fast sicheren Eindeutigkeit der Radon-Nikodym-Ableitung.

3. folgt aus 2. mit $\lambda = \mu$: $\quad \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\nu} = 1 \quad \mu$- und ν -f.ü. □

Bemerkung:
Wir nennen die Vorgehensweise im Punkt 2. die „Approximationsmethode“ und werden sie später noch mehrfach verwenden.

Zum Begriff „Absolutstetigkeit von Maßen und Funktionen“

Aussage 4.2. *Es gilt folgende Äquivalenz für je zwei endliche Maße μ und ν auf einem meßbaren Raum $(\Omega, \mathscr{A})$:*

$$\mu \ll \nu, \text{ d.h. } \nu(A) = 0 \Longrightarrow \mu(A) = 0 \quad \forall A \in \mathscr{A}$$

ist äquivalent mit

$$\forall\, \epsilon > 0 : \exists\, \delta > 0 : \forall\, A \text{ mit } \mu(A) < \delta \text{ folgt } \mu(A) < \epsilon.$$

Beweis:
Hinlänglichkeit der Bedingung: Gilt $\nu(A) = 0$, so ist $\nu(A) < \delta$ für alle $\delta > 0$, also auch $\mu(A) < \epsilon$ für alle $\epsilon > 0$, somit gilt $\mu(A) = 0$.

Notwendigkeit: Angenommen, die Bedingung gilt nicht. Dann gibt es ein $\epsilon > 0$ und eine Folge $A_n \in \mathscr{A}$ mit $\nu(A_n) \leq 2^{-n}$ und $\mu(A_n) \geq \epsilon$. Wir setzen $A = \limsup_{n\to\infty} A_n$. Dann gilt $\nu(A) = 0$ (Borel-Cantelli). Andererseits ist $\mu(A_n) = \displaystyle\int_\Omega \mathbf{1}_{A_n}\, d\mu$ und

$$\mu(A) = \int_\Omega \limsup_{n\to\infty} \mathbf{1}_{A_n}\, d\mu \geq \limsup_{n\to\infty} \int_\Omega \mathbf{1}_{A_n}\, d\mu \geq \epsilon \qquad \text{(Fatousches Lemma)}$$

Widerspruch zur Annahme. □

Definition 4.2. Eine Verteilungsfunktion F auf $\mathbb{R}$ heißt *absolutstetig bezüglich dem Lebesguemaß* λ auf $\mathbb{R}$, falls das von ihr erzeugte Wahrscheinlichkeitsmaß $\mathbb{P}_F$ absolutstetig bezüglich dem Lebesguemaß λ ist.

In diesem Fall gibt es auf Grund des Radon-Nikodym-Theorems eine Borelmeßbare Funktion f mit

$$\mathbb{P}_F(A) = \int_A f\, d\lambda, \quad A \in \mathscr{B}, \quad \text{insbesondere gilt} \quad F(x) = \int_{(-\infty,x]} f(s)\, ds, \quad x \in \mathbb{R}\,.$$

Nicht jede stetige Funktion F ist absolutstetig (z.B. die Cantorsche Funktion).
Eine Verteilungsfunktion F ist absolutstetig genau dann, wenn für alle $\epsilon > 0$ ein $\delta > 0$ existiert, so daß für jede endliche Menge $\{[\alpha_k, \beta_k],\ k = 1, 2, \ldots, m\}$ paarweise disjunkter Teilintervalle mit $\sum_{k=1}^m (\beta_k - \alpha_k) < \delta$ die Ungleichung $\sum_{k=1}^m F(\beta_k) - F(\alpha_k) < \epsilon$ gilt.

Ist F absolutstetig, so ist F Lebesgue-fast überall differenzierbar, und es gilt

$$\frac{dF}{dx} = f(x) \qquad \lambda\text{ -f.ü.},$$

d.h. die Ableitung von F ist λ -f.ü. gleich der Radon-Nikodym-Ableitung von $\mathbb{P}_F$ nach λ.
Außerdem gilt somit für jede absolutstetige Verteilungsfunktion F

$$F(x) = \int_{(-\infty,x]} \frac{dF}{ds}\, ds, \quad \text{d.h. } F \text{ hat eine Dichte } f(s) = \frac{dF}{ds}.$$

Literatur: Natanson, I.P., Theorie der Funktionen einer reellen Veränderlichen, Akademie-Verlag Berlin, 1961

Beispiel 4.2. Es seien F und G zwei absolutstetige Verteilungsfunktionen auf $\mathbb{R}$ mit den Dichten f bzw. g bezüglich dem Lebesguemaß λ. Es bezeichnen $\mathbb{P}_F$ und $\mathbb{P}_G$ die durch F bzw. G erzeugten Wahrscheinlichkeitsmaße auf $\mathscr{B}$, d.h. es gilt

$$\mathbb{P}_F\,(\,(a,b]\,) = F(b) - F(a) = \int_a^b f(s)\, ds \quad \text{und}$$

$$\mathbb{P}_G\,(\,(a,b]\,) = G(b) - G(a) = \int_a^b g(s)\, ds$$

für alle a, b mit $a < b$. Dann haben wir die

Aussage 4.3. *a)* $\mathbb{P}_F(B) = 0 \Longleftrightarrow \lambda\big(\{f > 0\} \cap B\big) = 0$ *(für alle Borelmengen B aus $\mathbb{R}$). Analoges gilt für $\mathbb{P}_G$ und g anstelle von $\mathbb{P}_F$ und f.*

b) $\mathbb{P}_F \ll \mathbb{P}_G \Longleftrightarrow \{x \in \mathbb{R} \mid f(x) > 0\} \overset{\lambda\text{-f.ü.}}{\subseteq} \{x \in \mathbb{R} \mid g(x) > 0\}$ *im Sinne von*

$$\lambda\big(\{f > 0\} \setminus \{g > 0\}\big) = 0. \tag{4.5}$$

c) In diesem Fall ist

$$\frac{d\,\mathbb{P}_F}{d\,\mathbb{P}_G}(x) = \frac{f(x)}{g(x)} \quad \text{für } \mathbb{P}_G\text{-fast alle } x \in \mathbb{R}. \tag{4.6}$$

Beweis:

a) Es gilt

$$\mathbb{P}_F(B) = 0 \Longleftrightarrow \int_B f(s)\,ds = 0 \Longleftrightarrow \int_{B\cap\{f>0\}} f(s)\,ds = 0 \Longleftrightarrow \lambda\big(B \cap \{f > 0\}\big) = 0.$$

b) Es gelte $\mathbb{P}_F \ll \mathbb{P}_G$. Dann gilt $\mathbb{P}_G\big(\{g = 0\}\big) = 0$ und folglich $\mathbb{P}_F\big(\{g = 0\}\big) = 0$, somit haben wir mit a)

$$\lambda\big(\{g = 0\} \cap \{f > 0\}\big) = 0, \quad \text{d.h.} \quad \lambda\big(\{f > 0\} \setminus \{g > 0\}\big) = 0.$$

Gilt dagegen (4.5) und ist $\mathbb{P}_G(B) = 0$, so haben wir (siehe a)) $\lambda\big(B \cap \{g > 0\}\big) = 0$ und somit wegen

$$\{f > 0\} \cap B = \Big(\{f > 0\} \cap \{g > 0\} \cap B\Big) \cup \Big(\{f > 0\} \cap \{g = 0\} \cap B\Big)$$

auch

$$\lambda\big(\{f > 0\} \cap B\big) \le \lambda\big(\{g > 0\} \cap B\big) + \lambda\big(\{f > 0\} \setminus \{g > 0\}\big) = 0.$$

Das bedeutet $\mathbb{P}_F(B) = 0$.

c) Weiterhin gilt (mit $\frac{0}{0} := 0$)

$$\mathbb{P}_F(B) = \int_B f\,d\lambda = \int_{B\cap\{f>0\}} f\,d\lambda = \int_{B\cap\{f>0\}} \frac{f(s)}{g(s)} g(s)\,ds. \text{ (wegen } \{f > 0\} \subseteq \{g > 0\}\ \lambda\text{-f.ü.)}$$

Bekanntlich gilt für alle nichtnegativen Borelmeßbaren Funktionen h:

$$\int_{\mathbb{R}} h(x)\,\mathbb{P}_G(\,dx) = \int_{\mathbb{R}} h(x)g(x)\,dx$$

Folglich ist

$$\mathbb{P}_F(B) = \int_{B\cap\{f>0\}} \frac{f(s)}{g(s)}\, \mathbb{P}_G(\,ds) = \int_B \frac{f(s)}{g(s)}\, \mathbb{P}_G(\,ds),$$

$$\text{d.h.}\quad \frac{d\,\mathbb{P}_F}{d\,\mathbb{P}_G}(s) = \frac{f(s)}{g(s)} \qquad \mathbb{P}_G\text{-f.s. bzw. } \lambda\text{-f.ü. auf } \{g>0\}.$$

□

Beispiel 4.3. Es sei $\mathbb{P}$ eine diskrete Wahrscheinlichkeitsverteilung auf $E = \{e_1, e_2, \ldots, e_m, \ldots\}$ mit

$$\mathbb{P}\big(\{e_k\}\big) =: p_k \geq 0, \quad \sum_{k=1}^{\infty} p_k = 1, \quad \mathscr{E} = \wp(E).$$

Es sei weiterhin Q ein σ-finites Maß auf $\mathscr{E}$. Dann kann man die Absolutstetigkeit an Hand der Einzelwahrscheinlichkeiten prüfen:

Aussage 4.4. *Es gilt*

a) $\mathbb{P} \ll Q \Longleftrightarrow \{e_k \,:\, p_k > 0\} \subseteq \{e_k \,:\, q_k > 0\}$

b) $\dfrac{d\,\mathbb{P}}{dQ}(e_k) = \dfrac{p_k}{q_k}$ *für alle* k *mit* $q_k > 0$.
(Für e_k *mit* $q_k = 0$ *kann* $\frac{d\,\mathbb{P}}{dQ}(e_k)$ *beliebig gewählt werden.)*

Beweis:
Es gilt

$$\mathbb{P}\big(\{e_k\}\big) = \frac{p_k}{q_k} q_k; \quad P(B) = \sum_{k\,:\,e_k\in B} \frac{p_k}{q_k} q_k, \quad B \subseteq E.$$

□

Q bezeichne das *Zählmaß auf* E, d.h. es gelte

$$Q\big(\{e_k\}\big) = 1,\ \forall\, k, \quad Q(B) = \sum_{k\,:\,e_k\in B} \mathbf{1},\ B \subseteq E.$$

Damit ist $\mathbb{P} \ll Q$ $\quad (Q(B) = 0$ impliziert $B = \emptyset)$ und es gilt

$$\frac{d\,\mathbb{P}}{dQ}(e_k) = p_k, \quad k \geq 1.$$

4.2 Likelihood-Funktionen für dominierte statistische Räume

Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ mit $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ ein statistischer Raum.

Definition 4.3. Gilt für ein σ-finites Maß μ auf $(\Omega, \mathscr{A})$ die Beziehung $\mathbb{P}_\vartheta \ll \mu$ für alle $\vartheta \in \Theta$, so heißt die Familie $\mathscr{P}$ *durch das Maß* μ *dominiert* und μ ein *dominierendes Maß für* $\mathscr{P}$.

Bemerkung:
Auf der reellen Achse $\mathbb{R}$ oder im $\mathbb{R}^n$ ist häufig das Lebesguemaß ein dominierendes Maß für eine gegebene Familie $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ von Wahrscheinlichkeitsverteilungen, z.B. für die Familie aller eindimensionalen Normalverteilungen $\mathcal{N}(\mu, \sigma^2)$ mit $\vartheta = (\mu, \sigma^2)^T \in \mathbb{R} \times (0, \infty) =: \Theta$. Jede Familie $\mathbb{P}_\vartheta$, $\vartheta \in \Theta$ diskreter Verteilungen auf einer höchstens abzählbar unendlichen Menge $\{a_1, a_2, \ldots, a_m, \ldots\}$ ist dominiert durch das sogenannte „Zählmaß“, definiert durch $\mu(\{a_m\}) \equiv 1$.

Aussage 4.5. *Ist* μ *ein dominierendes Maß für* $\mathscr{P}$, *so gibt es ein zu* μ *äquivalentes Wahrscheinlichkeitsmaß* μ_0, *das* $\mathscr{P}$ *ebenfalls dominiert.*

Beweis:
Da μ σ-finit ist, gibt es nämlich eine Zerlegung von Ω in meßbare Mengen $Z_k, \quad k \geq 1$, mit $\mu(Z_k) \in (0, \infty)$. Wir setzen

$$\mu_0(A) = \sum_{k=1}^{\infty} 2^{-k} \, \frac{\mu(A \cap Z_k)}{\mu(Z_k)}, \qquad A \in \mathscr{A}.$$

$\square$

Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Experiment, dominiert durch ein σ-finites Maß μ. Für jedes $\vartheta \in \Theta$ definieren wir durch

$$L(\vartheta\,; \mathscr{A}) := \frac{d\,\mathbb{P}_\vartheta}{d\mu}$$

eine Zufallsgröße auf $(\Omega, \mathscr{A})$. Sie ist

1. meßbar bezüglich $\mathscr{A}$ und
2. es gilt $\mathbb{P}_\vartheta(A) = \displaystyle\int_A L(\vartheta\,; \mathscr{A})(\omega)\mu(\,d\omega)$ für alle $A \in \mathscr{A}$.

Definition 4.4. Als Funktion von ϑ heißt $\vartheta \longrightarrow L(\vartheta\,;\,\mathscr{A})(\omega)$ die *Likelihoodfunktion der Familie $\mathscr{P}$ bezüglich μ oder einfach von $\mathscr{P}$ bezüglich μ.*

Beispiel 4.4. Im Fall der Familie der Normalverteilungen $\mathcal{N}(\mu, \sigma^2)$, $(\mu \in \mathbb{R},\ \sigma^2 > 0)$ auf $\mathbb{R}$ erhält man für $\vartheta = (\mu, \sigma^2)^T$

$$L(\vartheta\,;\,\mathscr{B})(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\omega-\mu)^2}{2\sigma^2}\right\}, \quad \omega \in \mathbb{R}$$

für das dominierende Lebesguemaß λ.

Beispiel 4.5. Ist Ω eine abzählbare Menge, $\Omega = \{\omega_k : k \geq 1\}$ und $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ eine durch die Einzelwahrscheinlichkeiten $p_k(\vartheta) = \mathbb{P}_\vartheta(\{\omega_k\})$ definierte Wahrscheinlichkeitsverteilung auf $(\Omega, \wp(\Omega))$, so ist wie bereits erwähnt, das Zählmaß μ, definiert durch $\mu(\{\omega_k\}) \equiv 1$, ein die Familie $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ dominierendes Maß. Es gilt

$$L\big(\vartheta\,;\,\wp(\Omega)\big)(\omega_k) = \frac{p_k(\vartheta)}{\mu(\{\omega_k\})} = p_k(\vartheta), \quad k \geq 1.$$

4.3 Stochastische Likelihoodfunktionen

Definition 4.5. Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistischer Raum mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und es sei $\mathscr{H}$ eine Teil-σ-Algebra von $\mathscr{A}$. Es gelte $\mathbb{P}_\vartheta\big|_{\mathscr{H}} \ll \mu$ für ein σ-finites Maß μ auf $\mathscr{H}$.
($\mathbb{P}_\vartheta\big|_{\mathscr{H}}$ bezeichnet die Einschränkung von $\mathbb{P}_\vartheta$ auf die Teil-σ-Algebra $\mathscr{H}$ von $\mathscr{A}$.)
Dann heißt

$$L(\vartheta\,;\,\mathscr{H})(\omega) := \frac{d\,\mathbb{P}_\vartheta\big|_{\mathscr{H}}}{d\mu}(\omega),$$

die *stochastische Likelihoodfunktion von $\mathscr{P}$ bezüglich μ und $\mathscr{H}$.*
Ist $\mathscr{H} = \mathscr{A}$, so handelt es sich um die in Abschnitt 4.2 definierte Likelihoodfunktion von $\mathscr{P}$ bezüglich μ.

Nach Definition ist $\vartheta \longrightarrow L(\vartheta\,;\,\mathscr{H})$ eine Funktion, deren Werte Zufallsgrößen über $(\Omega, \mathscr{A})$ sind, mit folgenden zwei Eigenschaften:

1. Für jedes $\vartheta \in \Theta$ ist $L(\vartheta\,;\,\mathscr{H})$ ist eine $\mathscr{H}$-meßbare Zufallsgröße,
2. $\int_H L(\vartheta\,;\,\mathscr{H})\,d\mu = \mathbb{P}_\vartheta(H) \quad \forall H \in \mathscr{H},\ \forall \vartheta \in \Theta$

Die Eigenschaften 1. und 2. bestimmen $L(\vartheta\,;\,\mathscr{H})$ eindeutig in dem Sinne, daß für jede $\mathscr{H}$-meßbare Funktion $L_1(\vartheta\,;\,\mathscr{H})$ auf Ω mit der Eigenschaft 2, $L(\vartheta, \cdot) = L_1(\vartheta, \cdot)$ μ-f.ü., für alle $\vartheta \in \Theta$ gilt.

Aussage 4.6. *Ist $\mathscr{H}'$ eine weitere σ-Algebra mit $\mathscr{H} \subseteq \mathscr{H}' \subseteq \mathscr{A}$, und ist $(\mathbb{P}_\vartheta \big|_{\mathscr{H}'},\ \vartheta \in \Theta)$ dominiert durch ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathscr{H}'$, so ist (natürlich) auch $(\mathbb{P}_\vartheta \big|_{\mathscr{H}},\ \vartheta \in \Theta)$ dominiert durch $\mathbb{P}$, und es gilt*

$$L(\vartheta\,;\,\mathscr{H}) = \mathbb{E}_{\mathbb{P}}\big(L(\vartheta\,;\,\mathscr{H}')|\mathscr{H}\big). \tag{4.7}$$

Beweis:
Nach Definition ist $L(\vartheta\,;\,\mathscr{H})$ eine $\mathscr{H}$-meßbare Funktion mit

$$\int_H L(\vartheta\,;\,\mathscr{H})\,d\,\mathbb{P} = \mathbb{P}_\vartheta(H), \quad H \in \mathscr{H},\ \vartheta \in \Theta,$$

und da $\mathscr{H} \subseteq \mathscr{H}'$ gilt, ist auch

$$\mathbb{P}_\vartheta(H) = \int_H L(\vartheta\,;\,\mathscr{H}')\,d\,\mathbb{P}, \quad H \in \mathscr{H},\ \vartheta \in \Theta.$$

Somit gilt

$$\int_H L(\vartheta\,;\,\mathscr{H})\,d\,\mathbb{P} = \int_H L(\vartheta\,;\,\mathscr{H}')\,d\,\mathbb{P}, \quad H \in \mathscr{H},\ \vartheta \in \Theta.$$

Daraus ergibt sich die Behauptung auf Grund der Definition der bedingten Erwartung. □

Folgerung 4.6 (Martingaleigenschaft der Likelihoodfunktion). *Ist $(\mathscr{H}_n,\ n \geq 1)$ eine wachsende Folge von Teil-σ-Algebren von $\mathscr{A}$, d.h. gilt $\mathscr{H}_1 \subseteq \mathscr{H}_2 \subseteq \ldots \subseteq \mathscr{H}_n \subseteq \mathscr{A}$, $(n \geq 1)$, so ist für jedes $\vartheta \in \Theta$ die Folge $\big(L(\vartheta\,;\,\mathscr{H}_n),\ \mathscr{H}_n\big)_{n\geq 1}$ ein nichtnegatives Martingal bezüglich $\mathbb{P}$.*

Auf Grund des Martingalkonvergenzsatzes (siehe z.B. Shiryaev) folgt: Der Grenzwert $\lim\limits_{n\to\infty} L_n(\omega) =: L_\infty$ existiert $\mathbb{P}$-fast sicher und es gilt $\mathbb{E}_{\mathbb{P}}(L_\infty) \leq \overline{\lim}_{n\to\infty}\, \mathbb{E}_{\mathbb{P}}(L_n) = 1$. (Fatou), L_∞ ist $\mathscr{H}_\infty := \bigvee_{n=1}^{\infty} \mathscr{H}_n$ meßbar. Hier wurde $L_n = L(\vartheta\,;\,\mathscr{H}_n)$ gesetzt.

Frage: Sind die Wahrscheinlichkeitsmaße $\mathbb{P}_\vartheta \big|_{\mathscr{H}_\infty}$ auch absolutstetig bezüglich $\mathbb{P} \big|_{\mathscr{H}_\infty}$? Dieser Frage werden wir später (im Abschnitt 4.7) nachgehen.

4.4 Deterministische Likelihoodfunktionen

Im allgemeinen beobachtet man nicht den Ausgang ω des zufälligen Experiments $(\Omega, \mathscr{A}, \mathscr{P})$ als Ganzes, sondern nur Teile davon, z.B. Temperaturmessungen oder Aktienkurse, zu diskreter Zeit (d.h. täglich, wöchentlich, monatlich), im Gegensatz zu kontinuierlicher, also zeitstetiger Beobachtung.

Dieser Sachverhalt wird durch eine Stichprobe $X(\omega)$ modelliert: man beobachtet $X(\omega)$ anstelle von ω.
Folglich hat man nach der Ausführung des Experimentes auch nicht die volle Information, welche der Ereignisse $A \in \mathscr{A}$ eingetreten sind und welche nicht, sondern man weiß es nur von Ereignissen der Form $\{X \in B\} = X^{-1}(B)$ mit $B \in \mathscr{E}$. Da man X beobachtet, kann man entscheiden, ob der beobachtete Wert zu B gehört, d.h., ob $\{X \in B\}$ eingetreten ist.
Die relevante σ-Algebra von Ereignissen ist $\mathscr{A}^X := X^{-1}(\mathscr{E}) \subseteq \mathscr{A}$.
Wir können also auf der Grundlage unserer beobachteten Stichprobe X i.a. die Likelihoodfunktion $L(\vartheta\,;\mathscr{A})(\omega)$ nicht bestimmen. Der Ausweg liegt in der Einführung der Likelihoodfunktion $L(\vartheta\,;\mathscr{A}^X)$. Da $L(\vartheta\,;\mathscr{A}^X)$ eine $\mathscr{A}^X$-meßbare Zufallsgröße ist, muß es für jedes $\vartheta \in \Theta$ eine meßbare Funktion Ψ_ϑ von $(E, \mathscr{E})$ in $(R, \mathscr{B})$ geben mit

$$L(\vartheta\,;\mathscr{A}^X)(\omega) = \Psi_\vartheta\big(X(\omega)\big), \quad \mathbb{P}_\vartheta\text{-f.s.}$$

(Siehe z.B. Bauer, Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie)

Diese Funktion $\Psi_\vartheta(x)$ werden wir hier im Folgenden berechnen.

<u>Vorbereitungen:</u>

a) Die von X erzeugte σ-Algebra $\mathscr{A}^X$
Es seien $(\Omega, \mathscr{A})$ und $(E, \mathscr{E})$ meßbare Räume und X eine meßbare Abbildung von $(\Omega, \mathscr{A})$ in $(E, \mathscr{E})$, es gelte also $\mathscr{A}^X := X^{-1}(\mathscr{E}) \subseteq \mathscr{A}$.
Das Mengensystem $\mathscr{A}^X = X^{-1}(\mathscr{E})$ ist eine σ-Algebra von Teilmengen von Ω, und zwar die kleinste σ-Algebra $\mathscr{H}$, bezüglich der X $(\mathscr{H}\text{-}\mathscr{E})$-meßbar ist. Sie heißt *die von X in Ω erzeugte σ-Algebra* und wird auch mit $\sigma(X)$ bezeichnet.
Die Elemente von $\mathscr{A}^X$ sind genau die Urbilder $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}, \quad B \in \mathscr{E}$, mit anderen Worten,

$$A \in \mathscr{A}^X \Longleftrightarrow \exists\, B \in \mathscr{E} \,:\, A = X^{-1}(B). \tag{4.8}$$

b) Die von X erzeugte Familie $\mathscr{P}^X$
Es sei ν ein σ-finites Maß auf $\mathscr{A}^X$. Dann ist durch $\nu^X(B) := \nu(A)$ mit $A := X^{-1}(B) \in \mathscr{A}^X$ ein σ-finites Maß ν^X auf $\mathscr{E}$ definiert.

Es seien nun $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Modell mit dem Stichprobenraum $(E, \mathscr{E})$

und ν ein σ-finites Maß auf $\mathscr{A}$. Wir setzen $\mathscr{P}\big|_{\mathscr{A}^X} := (\mathbb{P}_\vartheta\big|_{\mathscr{A}^X} : \vartheta \in \Theta)$ und $\mathscr{P}^X := (\mathbb{P}_\vartheta^X : \vartheta \in \Theta)$. Hierbei bezeichnet $Q\big|_{\mathscr{A}^X}$ für jedes Maß Q auf $\mathscr{A}$ die Einschränkung von Q auf $\mathscr{A}^X$.

Lemma 4.7. *Genau dann dominiert $\nu\big|_{\mathscr{A}^X}$ die Familie $\mathscr{P}\big|_{\mathscr{A}^X}$, wenn ν^X die Familie $\mathscr{P}^X$ dominiert.*

Beweis:
Es sei $A \in \mathscr{A}^X$ und $B \in \mathscr{E}$ derart, daß $A = X^{-1}(B)$ gilt. Angenommen, $\nu\big|_{\mathscr{A}^X}$ dominiert $\mathscr{P}\big|_{\mathscr{A}^X}$ und es sei $\nu^X(B) = 0$.
Dann folgt

$$\nu(A) = \nu(X^{-1}(B)) = \nu^X(B) = 0, \text{ folglich } \mathbb{P}_\vartheta(A) = 0 \text{ für alle } \vartheta \in \Theta.$$

Somit gilt $\mathbb{P}_\vartheta^X(B) = 0,\ \vartheta \in \Theta$. Also dominiert ν^X die Familie $\mathscr{P}^X$. Die Umkehrung zeigt man analog. □

Nehmen wir an, daß ν^X die Familie $\mathscr{P}^X$ dominiert, so haben wir insbesondere die Likelihoodfunktion

$$L^X(\vartheta\,;\,x) := \frac{d\,\mathbb{P}_\vartheta^X}{d\nu^X}(x), \quad x \in E,\ \vartheta \in \Theta$$

von $\mathscr{P}^X$ bezüglich ν^X. Sie ist nach Definition $\mathscr{E}$-meßbar. Da die Elemente $x \in E$ als konkrete Stichproben angesehen werden, die nichtzufällig sind, bezeichnen wir $L^X(\vartheta\,;\,x)$ als *deterministische Likelihoodfunktion von $\mathscr{P}^X$ bezüglich ν^X*. Setzt man in $L^X(\vartheta\,;\,x)$ an die Stelle x die mathematische Stichprobe X ein, so erhält man eine Zufallsgröße $L^X\big(\vartheta\,;\,X(\omega)\big)$, für die folgende Aussage gilt:

Aussage 4.7. *Es gilt für jedes $\vartheta \in \Theta$*

$$L\big(\vartheta\,;\,\mathscr{A}^X\big)(\omega) = L^X\big(\vartheta\,;\,X(\omega)\big), \quad \textit{für } \nu\textit{- fast alle } \omega. \tag{4.9}$$

Beweis:
Nach Definition ist für jedes $B \in \mathscr{E}$

$$\nu^X(B) := \nu\left(X^{-1}(B)\right), \qquad \text{also} \quad \int_\Omega \mathbf{1}_B\big(X(\omega)\big)\nu(\,d\omega) = \int_E \mathbf{1}_B(x)\nu^X(\,dx).$$

Damit gilt für alle $\mathscr{E}$-meßbaren nichtnegativen Funktionen g (Beweis mit der Approximationsmethode aus Kapitel 4.1),

$$\int_\Omega g\big(X(\omega)\big)\nu(\,d\omega) = \int_E g(x)\nu^X(\,dx) \text{ (Substitutionsformel).}$$

Wir setzen $g(x) := L^X(\vartheta\,;\,x)\mathbf{1}_B(x)$ und erhalten für jedes $B \in \mathscr{E}$ die Gleichung

$$\int_{X^{-1}(B)} L^X\big(\vartheta, X(\omega)\big)\nu(\,d\omega) = \int_B L^X(\vartheta\,;\,x)\nu^X(\,dx) = \mathbb{P}_\vartheta^X(B) = \mathbb{P}_\vartheta\left(X^{-1}\left(B\right)\right).$$

Da $\mathscr{A}^X = \{X^{-1}(B)\,|\,B \in \mathscr{E}\}$ gilt, ist die Aussage 4.7 bewiesen. Dabei haben wir benutzt, daß $\omega \longrightarrow L^X\big(\vartheta\,;\,X(\omega)\big)$ $\mathscr{A}^X$-meßbar ist. □

Folgerung 4.8. *Ist Y eine Zufallsgröße über $(\Omega, \mathscr{A})$ mit Werten in $(F, \mathscr{F})$, und ist $Y = \Psi(X)$ für eine meßbare Abbildung Ψ von $(E, \mathscr{E})$ in $(F, \mathscr{F})$, so gilt $\mathscr{A}^Y \subseteq \mathscr{A}^X$ (wegen $\mathscr{A}^Y = Y^{-1}(\mathscr{F}) = X^{-1}\big(\Psi^{-1}(\mathscr{F})\big) \subseteq X^{-1}(\mathscr{E}) = \mathscr{A}^X$)*
Folglich gilt (siehe Aussage 4.6):

$$\begin{aligned} L(\vartheta\,;\,\mathscr{A}^Y) &= \mathbb{E}_{\mathbb{P}}\big(L(\vartheta\,;\,\mathscr{A}^X)\mid\mathscr{A}^Y\big), \quad \textit{oder anders ausgedrückt,} \\ L^Y(\vartheta\,;\,Y(\omega)) &= \mathbb{E}_{\mathbb{P}}\big(L^X(\vartheta\,;\,X)\mid Y\big)(\omega) \quad \mathbb{P}\text{-}\textit{f.s.} \end{aligned} \tag{4.10}$$

Diese sehr allgemeinen Folgerungen aus dem Satz von Radon-Nikodym bilden den allgemeinen Rahmen für die Likelihoodtheorie stochastischer Prozesse mit diskreter oder auch mit stetiger Zeit.

4.5 Ausflug in die Welt der Stochastischen Prozesse

Unser statistisches Modell umfaßt insbesondere auch Fälle, bei denen die Stichprobe X aus Trajektorien zeitstetiger stochastischer Prozesse besteht. Der Stichprobenraum E ist also ein Funktionenraum. Die Wahl einer geeigneten σ-Algebra $\mathscr{E}$ von Teilmengen von E erfolgt in Analogie des Falles $E = \mathbb{R}^n$. Wir erinnern uns daran, daß $\mathscr{B}^n$ die kleinste σ-Algebra ist, die alle Rechtecke $B_1 \times \ldots \times B_n$ mit $B_k \in \mathscr{B},\ k = 1, \ldots, n$, enthält. Die Rechtecke werden im Funktionenraum durch die sogenannten Zylindermengen ersetzt. Doch zunächst einige Vorbereitungen.

Es seien $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X := \big(X(t),\, t \in [0,T]\big)$ ein reellwertiger stochastischer Prozeß über $(\Omega, \mathscr{A}, \mathbb{P})$. Für $\mathbb{P}$-fast alle $\omega \in \Omega$ seien die Trajektorien $t \longrightarrow X(t,\,\omega)$ von X an jeder Stelle $t \in [0,T)$ rechtsstetig und für jedes $t \in (0,T]$ existiere der Grenzwert von links und sei endlich. (Praktisch alle gegenwärtig in Anwendungen und in der Theorie vorkommenden stochastischen Prozesse haben diese Eigenschaft.)
Der Vektorraum $E_T := \mathbb{D}\big([0,T]\big)$ aller reellwertigen Funktionen $x = (x_s,\ s \leq T)$ mit den genannten Eigenschaften (cadlag = continues a droite, limites a gauche) ist metrisierbar mit einer Metrik d zu einem vollständigen metrischen Raum (E_T, d). Mit $\mathscr{E}_T$ bezeichnen wir die σ-Algebra der Borelmengen von (E_T, d).

Aussage 4.8. *$\mathscr{E}_T$ ist identisch mit der kleinsten σ-Algebra $\sigma(x_s,\ s \in [0,T])$, bezüglich der alle Abbildungen $x \longrightarrow x_s,\ \ x \in E_T,\ s \in [0,T]$, Borelmeßbar sind.*

Beweis: (Siehe Billigsley, Convergence of Probability Measures)

Für die komplizierte und nicht unmittelbar zugängliche σ-Algebra $\mathscr{E}_T$ gibt es ein Erzeugendensystem $\mathcal{Z}$, das aus speziellen Teilmengen von E_T, den sogenannten Zylindermengen, besteht.

Definition 4.9. Eine Teilmenge Z von E_T heißt eine *Zylindermenge*, falls es ein $m \geq 1$, Zeitpunkte $t_0, t_1, \ldots, t_m$ mit $0 = t_0 < t_1 < \ldots < t_m = T$ und ein $B \in \mathscr{B}^{m+1}$ gibt, mit

$$Z = \{x = (x_s) \in E_T \,|\, (x_{t_0}, x_{t_1}, \ldots, x_{t_m}) \in B\}$$

Mitunter werden wir auch die Notation $Z = Z_{t_1,t_2,\ldots,t_m\,;\,B}$ verwenden; mit $\mathcal{Z}$ werde die Menge aller Zylindermengen aus E_T bezeichnet.

Lemma 4.10. *$\mathcal{Z}$ ist eine Semialgebra von Teilmengen aus E_T.*

Beweis: Übungsaufgabe

Wegen der Aussage 4.8 gilt $\{x_s \in B\} \in \mathscr{E}$ für alle $B \in \mathscr{B}$ und $s \in [0,T]$. Daraus ergibt sich (ohne Beweis) $\mathcal{Z} \subseteq \mathscr{E}_T$ und somit auch $\sigma(\mathcal{Z}) \subseteq \mathscr{E}_T$. (Mit $\sigma(\mathcal{Z})$ bezeichnen wir die kleinste σ-Algebra von Teilmengen von E_T, in der $\mathcal{Z}$ enthalten ist.) Mittels der obigen Aussage 4.8 erhalten wir $\sigma(\mathcal{Z}) = \mathscr{E}_T$.
Um die Wahrscheinlichkeitsverteilung $\mathbb{P}^X$ von X auf $\mathscr{E}_T$ zu bestimmen, genügt es, sie auf $\mathcal{Z}$ anzugeben, da jede σ-additive σ-finite Mengenfunktion Q auf $\mathcal{Z}$ auf eindeutige Weise von $\mathcal{Z}$ auf $\sigma(\mathcal{Z})$ fortgesetzt werden kann.
Für $Z \in \mathcal{Z}$ gilt offenbar

$$\mathbb{P}^X(Z) := \mathbb{P}(X \in Z) = \mathbb{P}\big((X(t_0), \ldots, X(t_m)) \in B\big),$$

wobei Z von der Gestalt

$$Z = \{x = (x_s) \,|\, (x_{t_0}, \ldots, x_{t_m}) \in B\} \text{ für ein } B \in \mathscr{B}^{m+1}$$

sei.
Wir sehen also, daß die Verteilung $\mathbb{P}^X$ von X auf $\mathscr{E}_T$ eindeutig bestimmt ist durch die Wahrscheinlichkeitsverteilungen aller möglichen endlichdimensionalen Vektoren $\big(X(t_0), \ldots, X(t_m)\big)$ mit $\{t_0, t_1, \ldots, t_m\} \subseteq [0,T]$, wobei o.B.d.A. $t_0 = 0$ und $t_m = T$ gesetzt werden und $0 < t_1 < \ldots < t_m$ gefordert werden kann.

Definition 4.11. Für jede Auswahl $\mathscr{T}_n := \{t_0, t_1, \ldots, t_m\} \subseteq [0,T]$ mit den eben angegebenen Eigenschaften, ist durch

$$\Phi_{\mathscr{T}_n}(B) := \mathbb{P}\big((X(t_0), \ldots, X(t_m)) \in B\big), \quad B \in \mathscr{B}^{m+1}$$

eine Wahrscheinlichkeitsverteilung $\Phi_{\mathscr{T}_n}$ auf $\mathscr{B}^{m+1}$ definiert, die zu $\mathscr{T}_n$ gehörende sogenannte *endlichdimensionale Verteilung von* X.

Die endlichdimensionalen Verteilungen stochastischer Prozesse lassen sich in vielen Fällen explizit bestimmen.

Beispiel 4.6. Poissonscher Prozess $\big(N(t),\ t \geq 0\big)$ mit Intensitätsparameter $\vartheta > 0$.

Definition 4.12. Ein stochastischer Prozess $N = \big(N(t),\ t \geq 0\big)$ auf einem Grundraum $(\Omega, \mathscr{F}, \mathbb{P})$ heißt *Poissonscher Prozess mit Intensitätsparameter* $\vartheta > 0$, falls gilt:

1. $N(0) = 0 \quad \mathbb{P}$-f.s.,
2. für $0 = t_0 < t_1 < \ldots < t_m = T$ sind $\big(N(t_k) - N(t_{k-1})\big),\ k = 1, 2, \ldots, m$ unabhängig,
3. für $0 \leq s < t$ ist $N(t) - N(s)$ Poissonverteilt zum Parameter $\vartheta(t-s)$,
4. Jede Funktion $t \longrightarrow N(t, \omega)$ ist stückweise konstant, ihre Sprungpunkte $T_k,\ k \geq 1$, häufen sich nicht im Endlichen und es gilt

   $$N(T_k) = N(T_k + 0) = N(T_k - 0) + 1.$$

   (Aus 1. - 3. folgt bereits, daß es eine Version von N gibt, die die Eigenschaft 4. besitzt.)

Es sei $0 = t_0 < t_1 < t_2 < \ldots < t_n = T$. Dann ist $\Phi^{N,\vartheta}_{n;t_1,t_2,\ldots,t_n}$ eine diskrete Verteilung auf N^n mit $N = \{0, 1, 2, \ldots\}$ und ihre Einzelwahrscheinlichkeiten sind

$$\begin{aligned}
&\Phi^{N,\vartheta}_{n;t_1,t_2,\ldots,t_n}\Big((i_1, i_2, \ldots, i_n)\Big) = \mathbb{P}_\vartheta\big(N(t_1) = i_1, \ldots, N(t_n) = i_n\big) \\
&= \frac{(\vartheta t_1)^{i_1}}{i_1!} e^{-\lambda t_1} \frac{(\vartheta(t_2 - t_1))^{i_2 - i_1}}{(i_2 - i_1)!} e^{-\lambda(t_2 - t_1)} \cdot \ldots \cdot \frac{(\vartheta(t_n - t_{n-1}))^{i_n - i_{n-1}}}{(i_n - i_{n-1})!} e^{-\lambda(t_n - t_{n-1})} \\
&= \vartheta^{i_n} e^{-\vartheta t_n} \nu\big((i_1, \ldots, i_n)\big) \qquad (4.11)
\end{aligned}$$

mit

$$\nu\big((i_1, \ldots, i_n)\big) = \prod_{k=1}^{n} \frac{((t_k - t_{k-1}))^{i_k - i_{k-1}}}{(i_k - i_{k-1})!}.$$

Beispiel 4.7. Wienerprozess $(W(t),\ t \geq 0)$ mit Drift μ und Diffusion $\sigma^2 > 0$.

Definition 4.13. Ein stochastischer Prozess $W = (W(t),\ t \geq 0)$ auf einem Grundraum $(\Omega, \mathscr{F}, \mathbb{P})$ heißt *Wienerprozess mit den Parametern* μ, σ^2, wenn gilt:

1. $W(0) = 0 \quad \mathbb{P}$-f.s.,
2. für $0 = t_0 < t_1 < \ldots < t_m = T$ sind $\big(W(t_k) - W(t_{k-1})\big),\ k = 1, 2, \ldots, m$ unabhängig,
3. für $0 \leq s < t$ ist $W(t) - W(s) \sim \mathcal{N}\big(\mu(t-s),\ \sigma^2(t-s)\big)$-verteilt,
4. $t \longrightarrow W(t, \omega)$ ist stetig für $\mathbb{P}$-fast alle $\omega \in \Omega$.
   (Aus 1. - 3. folgt bereits, daß es eine Version von W gibt, die die Eigenschaft 4. besitzt.)

μ und σ^2 heißen *Drift* bzw. *Diffusion* von W.
Einen Wienerprozess W mit Drift $\mu = 0$ und Diffusion $\sigma^2 = 1$ nennen wir *Standard-Wienerprozess* und bezeichnen ihn mit W^0.

Bemerkung: Ist W^0 ein Standard-Wienerprozess, so bildet

$$W(t) = \mu t + \sigma W^0(t), \quad t \geq 0$$

einen Wienerprozess mit Drift μ und Diffusion σ^2.

Es sei $W = (W(t),\, t \geq 0)$ ein Wiener-Prozess mit dem Parameter $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty) =: \Theta$. Es sei $0 = t_0 < t_1 < \ldots < t_m = T$. Wir setzen

$$X := (W(t_1), \ldots, W(t_m)), \quad E = \mathbb{R}^m,\ \mathscr{E} = \mathscr{B}^m.$$

Für die Wahrscheinlichkeitsdichte von X gilt

Aussage 4.9. *X hat die Dichte bezüglich des Lebesguemaßes*

$$\varphi^\vartheta_{t_1,\ldots,t_m}(x_1, \ldots, x_m) = \prod_{k=1}^m \varphi^\vartheta_{t_k - t_{k-1}}(x_k - x_{k-1}), \tag{4.12}$$

wobei $\quad \varphi^\vartheta_t(x) := (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{ -\dfrac{(x - \mu t)}{2\sigma^2 t} \right\}, \quad x \in \mathbb{R},\ t > 0 \quad$ *gesetzt wurde.*

Beweis:
Es sei $Z := (W(t_1) - W(t_0), \ldots, W(t_m) - W(t_{m-1}))$ mit $t_0 = 0$. Z hat nach Definition des Wiener-Prozesses die Dichte

$$f_Z(z_1, z_2, \ldots, z_m) = \prod_{k=1}^{m} \varphi^{\vartheta}_{t_k - t_{k-1}}(z_k).$$

Es gilt weiter $X = A \cdot Z$ mit der Matrix $A = (e_{ij})_{i,j=1,\ldots,m}$, $e_{ij} = 1$ für $i \geq j$, $e_{ij} = 0$ für $i < j$. Folglich ist für $x = (x_1, \ldots, x_m)$, $\varphi_{t_1,\ldots,t_m}(x_1, \ldots, x_m) = f_Z(A^{-1}x) \cdot |\det A^{-1}|$, woraus sich wegen

$$A^{-1} = (f_{ij})_{i,j=1,\ldots,n}, \quad f_{ij} = \begin{cases} 1 & \text{für } {\scriptstyle i=j} \\ -1 & \text{für } {\scriptstyle i=j+1} \\ 0 & \text{sonst} \end{cases}$$

die Behauptung ergibt. □

4.6 Likelihood am Beispiel des Wienerprozesses

Es seien $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$, $\Theta = \mathbb{R} \times (0, \infty)$, $\vartheta = (\mu, \sigma^2)^T \in \Theta$ und $W = (W(t),\ t \geq 0)$ sei ein reellwertiger stochastischer Prozess auf $(\Omega, \mathscr{A})$, der für jedes $\vartheta = (\mu, \sigma^2)^T \in \Theta$ bezüglich $\mathbb{P}_\vartheta$ ein Wienerprozess mit Drift μ und Diffusion σ^2 ist. Wir nehmen an, μ und σ^2 seien unbekannt.

Ziel ist es, auf der Grundlage der Beobachtung $X = \big(W(t),\ t \in [0, T]\big)$ des Wienerprozesses bis zur Zeit $T < \infty$ Schätzungen für μ und σ^2 vorzunehmen. Wir konzentrieren uns hier auf die Likelihoodmethode.

Beginnen wir mit dem Fall diskreter Beobachtung, d.h., wir kennen W nur zu den Zeitpunkten $t_1 < \ldots < t_m = T$. Die Menge $\mathscr{T}$ sei die Menge dieser Zeitpunkte: $\mathscr{T} = \{t_1, t_2, \ldots, t_m\}$. Als Stichprobe liegt uns dann der Vektor $X = \big(W(t_1), \ldots, W(t_m)\big)$ vor.
Die Verteilung $\mathbb{P}^X_\vartheta$ von X ist eine Normalverteilung mit der Dichte bezüglich des Lebesguemaßes λ_m (vgl. (4.12))

$$\varphi^{\vartheta}_{\mathscr{T}}(x_1, \ldots, x_m) = \prod_{k=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2(t_k - t_{k-1})}} \exp\left\{ -\frac{\big(x_k - x_{k-1} - \mu(t_k - t_{k-1})\big)^2}{2(t_k - t_{k-1})\sigma^2} \right\}, \tag{4.13}$$

$x = (x_1, \ldots, x_m) \in \mathbb{R}^m$.

Da $\varphi_{\mathscr{T}}^{\vartheta}$ eine streng positive Dichte für $\mathbb{P}_{\vartheta}^{X}$ ist, sind $\mathbb{P}_{\vartheta}^{X}$ und das Lebesguemaß λ_m auf $\mathbb{R}^m$ äquivalente Maße, für jedes $\vartheta \in \Theta$.

Somit sind für je zwei $\vartheta, \vartheta_0 \in \Theta$ die Wahrscheinlichkeitsmaße $\mathbb{P}_{\vartheta}^{X}$ und $\mathbb{P}_{\vartheta_0}^{X}$ äquivalent. Daraus ergibt sich (siehe (4.3) und (4.4)) :

$$\frac{d\,\mathbb{P}_{\vartheta}^{X}}{d\,\mathbb{P}_{\vartheta_0}^{X}} = \frac{d\,\mathbb{P}_{\vartheta}^{X}}{d\lambda} \cdot \frac{d\lambda}{d\,\mathbb{P}_{\vartheta_0}^{X}} = \frac{d\,\mathbb{P}_{\vartheta}^{X}}{d\lambda} \cdot \left(\frac{d\,\mathbb{P}_{\vartheta_0}^{X}}{d\lambda}\right)^{-1}.$$

Das bedeutet für die deterministische Likelihoodfunktion

$$L^X(\vartheta\,;\,x) = \frac{d\,\mathbb{P}_{\vartheta}^{X}}{d\,\mathbb{P}_{\vartheta_0}^{X}}(x) = \frac{\varphi_{\mathscr{T}}^{\vartheta}(x)}{\varphi_{\mathscr{T}}^{\vartheta_0}(x)} =$$

$$= \left(\frac{\sigma^2}{\sigma_0^2}\right)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\left(\sigma^{-2} - \sigma_0^{-2}\right)\sum_{k=1}^{m}\frac{(x_k - x_{k-1})^2}{t_k - t_{k-1}} + \left(\frac{\mu}{\sigma^2} - \frac{\mu_0}{\sigma_0^2}\right)x_m - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} - \frac{\mu_0^2}{\sigma_0^2}\right)T\right\}.$$

Da ϑ_0 beliebig aus Θ gewählt werden kann, setzen wir $\mu_0 = 0$.
Unter Beachtung von $t_m = T$ folgt

$$L^X(\vartheta\,;\,x) = \left(\frac{\sigma^2}{\sigma_0^2}\right)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\left(\sigma^{-2} - \sigma_0^{-2}\right)\sum_{k=1}^{m}\frac{(x_k - x_{k-1})^2}{t_k - t_{k-1}} + \frac{\mu}{\sigma^2}x_m - \frac{1}{2}\frac{\mu^2}{\sigma^2}T\right\}$$

mit $x = (x_1, x_2, \ldots, x_m)$.

Für die stochastische Likelihoodfunktion ergibt sich (siehe (4.9)):

$$L(\vartheta\,;\,\mathscr{A}^X)(\omega) = L^X(\vartheta\,;\,X(\omega)) = \frac{d\,\mathbb{P}_{\vartheta}\big|_{\mathscr{A}^X}}{d\,\mathbb{P}_{\vartheta_0}\big|_{\mathscr{A}^X}} =$$

$$= \left(\frac{\sigma^2}{\sigma_0^2}\right)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\left(\sigma^{-2} - \sigma_0^{-2}\right)\sum_{k=1}^{m}\frac{\big(W(t_k) - W(t_{k-1})\big)^2}{t_k - t_{k-1}} + \frac{\mu}{\sigma^2}W(T) - \frac{1}{2}\frac{\mu^2}{\sigma^2}T\right\}$$

Der Maximum-Likelihood-Schätzer für $\vartheta = (\mu, \sigma^2)^T$ ergibt sich aus den Maximum-Likelihood-Gleichungen

$$\frac{\partial}{\partial\mu}\log L(\vartheta\,;\,\mathscr{A}^X)(\omega) = 0 \qquad \text{und}$$

$$\frac{\partial}{\partial\sigma^2}\log L(\vartheta\,;\,\mathscr{A}^X)(\omega) = 0.$$

Wir erhalten $\hat{\vartheta}_{\mathscr{T}} = (\hat{\mu}_{\mathscr{T}}, \hat{\sigma}_{\mathscr{T}}^2)^T$ mit

$$\hat{\mu}_{\mathscr{T}} = \frac{W(T)}{T}, \qquad \text{unabhängig von der konkreten Gestalt von } \mathscr{T}, \text{ und}$$

$$\hat{\sigma}_{\mathscr{T}}^2 = \frac{1}{m}\sum_{k=1}^{m}\frac{\big(W(t_k) - W(t_{k-1})\big)^2}{t_k - t_{k-1}} - \frac{1}{m}\frac{W_T^2}{T}.$$

Wir berechnen den Erwartungswert dieser Schätzer und untersuchen ihre Streuung. Aus der Definition des Prozesses $\big(W(t),\ t \geq 0\big)$ folgt sofort:

$$\mathbb{E}_\vartheta(\hat{\mu}_{\mathscr{T}}) = \mu, \qquad D^2_\vartheta \hat{\mu}_{\mathscr{T}} = \frac{\sigma^2 T}{T^2} = \frac{\sigma^2}{T}, \qquad (\hat{\mu}_{\mathscr{T}} \text{ ist erwartungstreu})$$
$$\mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}}) = \left(1 - \frac{1}{m}\right)\sigma^2 + \frac{\mu^2 T}{m}.$$

$\hat{\sigma}^2_{\mathscr{T}}$ ist ein verfälschter (biased) Schätzer für σ^2 mit dem Bias

$$\mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}}) - \sigma^2 =: b(\vartheta, T, \mathscr{T}) = \frac{\mu^2 T - \sigma^2}{m}.$$

Es gilt

$$D^2_\vartheta \hat{\sigma}^2_{\mathscr{T}} \leq 2 D^2_\vartheta \left(\frac{1}{m}\sum_{k=1}^{m} \frac{\big(W(t_k) - W(t_{k-1})\big)^2}{t_k - t_{k-1}}\right) + 2 D^2_\vartheta \left(\frac{1}{m}\,\frac{W(T)^2}{T}\right). \tag{4.14}$$

(Zum Beweis beachte man $(a+b)^2 \leq 2(a^2+b^2),\ a, b \in \mathbb{R}$.)

Untersuchung des asymptotischen Verhaltens des Schätzers $\hat{\sigma}^2_{\mathscr{T}}$ für $m \to \infty$, $T = \text{const.}$

Es seien für jedes $n \geq 1$

$$\mathscr{T}_n := \{0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_{m_n}^{(n)} = T\} \quad \text{und}$$

$\hat{\mu}_{\mathscr{T}_n}$, $\hat{\sigma}^2_{\mathscr{T}_n}$ die eben konstruierten ML-Schätzer für (μ, σ^2).

Voraussetzung: Die Anzahl der Beobachtungszeitpunkte nimmt unbegrenzt zu: $\lim_{n\to\infty} m_n = \infty$.

Bemerkung: Erwartungswert und Streuung von $\hat{\mu}_{\mathscr{T}_n}$ bleiben konstant bei wachsendem n und hängen nicht von der expliziten Form von $\mathscr{T}_n$ ab. Der Schätzer $\hat{\mu}_{\mathscr{T}_n}$ hängt nur vom letzten Beobachtungswert $W(T)$ ab, seine Streuung läßt sich durch häufigere Beobachtung innerhalb des Zeitraumes $[0, T]$ nicht verkleinern.

Für $\mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n})$ gilt offenbar $\lim_{n\to\infty} \mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n}) = \sigma^2$, d.h., $\hat{\sigma}^2_{\mathscr{T}_n}$ ist asymptotisch erwartungstreu.

Aussage 4.10. *Es gilt für jedes* $\vartheta = (\mu, \sigma^2)^T \in \Theta$

$$\lim_{n\to\infty} D^2_\vartheta \hat{\sigma}^2_{\mathscr{T}_n} = 0 \quad \textit{und insbesondere}$$
$$\mathbb{P}_\vartheta\big(\mid \hat{\sigma}^2_{\mathscr{T}_n} - \sigma^2 \mid > \epsilon\big) \overset{n\to\infty}{\longrightarrow} 0 \quad \textit{für jedes } \epsilon > 0.$$

Falls $\sum_{n=1}^{\infty} \frac{1}{m_n} < \infty$, *so haben wir sogar*

$$\lim_{n\to\infty} \hat{\sigma}^2_{\mathscr{T}_n}(\omega) = \sigma^2, \quad \mathbb{P}_\vartheta \text{ -f.s. für jedes } \vartheta \in \Theta.$$

Beweis:
Aus (4.14) erhalten wir

$$D^2_\vartheta \hat{\sigma}^2_{\mathscr{T}_n} \le 2D^2_\vartheta \left(\frac{1}{m_n} \sum_{k=1}^{m_n} \frac{\left(W(t_k^{(n)}) - W(t_{k-1}^{(n)})\right)^2}{t_k^{(n)} - t_{k-1}^{(n)}} \right) + 2D^2_\vartheta \left(\frac{1}{m_n} \frac{W(T)^2}{T} \right) \tag{4.15}$$

Wir untersuchen beide Summanden auf der rechten Seite. Offenbar ist

$$D^2_\vartheta \left(\frac{1}{m_n} \frac{W(T)^2}{T} \right) = \frac{1}{m_n^2 T^2} D^2_\vartheta \left(W(T)^2\right) \longrightarrow 0 \quad \text{für } m_n \to \infty$$

Wir setzen

$$Z_k^{(n)} := \frac{W(t_k^{(n)}) - W(t_{k-1}^{(n)})}{(t_k^{(n)} - t_{k-1}^{(n)})^{\frac{1}{2}}}.$$

Dann gilt nach Definition des Wienerschen Prozesses $Z_k^{(n)} \sim \mathcal{N}\left(\mu \left(t_k^{(n)} - t_{k-1}^{(n)}\right)^{\frac{1}{2}}, \sigma^2\right)$ und die $Z_1^{(n)}, Z_2^{(n)}, \dots, Z_{m_n}^{(n)}$ sind voneinander unabhängig. Folglich haben wir

$$\frac{1}{m_n^2} D^2_\vartheta \left(\sum_{k=1}^{m_n} \frac{\left(W(t_k^{(n)}) - W(t_{k-1}^{(n)})\right)^2}{t_k^{(n)} - t_{k-1}^{(n)}} \right) = \frac{1}{m_n^2} D^2_\vartheta \big(\sum_{k=1}^{m_n} (Z_k^{(n)})^2 \big).$$

Wir zeigen, daß die rechte Seite dieser Gleichung für $m_n \longrightarrow \infty$ gegen Null konvergiert.

Lemma 4.14. *Ist X eine normalverteilte Zufallsvariable,* $X \sim \mathcal{N}(\mu, \sigma^2)$, *so gilt*

$$D^2 X^2 = \mathbb{E}\big(X^4\big) - \big(\mathbb{E}(X^2)\big)^2 = 4\sigma^2\mu^2 + 2\sigma^4.$$

Beweis:

Für jede $\mathcal{N}(\mu,\sigma^2)$-verteilte Zufallsgröße X haben wir
$\mathbb{E}\left(\frac{(X-\mu)^4}{\sigma^4}\right) = 3$ und $\mathbb{E}\left((X-\mu)^3\right) = 0$.
Damit erhalten wir $\mathbb{E}\left(X^3\right) = \mu^3 + 3(\sigma^2+\mu^2)\mu - 3\mu^3 = \mu^3 + 3\sigma^2\mu$
und $\mathbb{E}(X^4) = 4\mu\,\mathbb{E}(X^3) - 6\mu^2(\mu^2+\sigma^2) + 4\mu^4 - \mu^4 + 3\sigma^4 = \mu^4 + 6\sigma^2\mu^2 + 3\sigma^4$.
Mit $\left(\mathbb{E}(X^2)\right)^2 = (\mu^2+\sigma^2)^2 = \mu^4+\sigma^4+2\mu^2\sigma^2$ gilt
$D^2X^2 = \mathbb{E}\big(X^4\big) - \left(\mathbb{E}(X^2)\right)^2 = 4\sigma^2\mu^2 + 2\sigma^4$.

□

Wendet man Lemma 4.14 auf die Zufallsvariablen $Z_k^{(n)}$ an, so erhält man mit

$$D^2_\vartheta(Z_k^{(n)})^2 = 2\sigma^4 + 4\mu^2\sigma^2(t_k^{(n)} - t_{k-1}^{(n)}) = 2\sigma^2\big(\sigma^2 + 2\mu^2(t_k^{(n)} - t_{k-1}^{(n)})\big)$$

für den ersten Summanden in (4.15):

$$\frac{1}{m_n^2}D^2_\vartheta\left(\sum_{k=1}^{m_n}\frac{\left(W(t_k^{(n)}) - W(t_{k-1}^{(n)})\right)^2}{t_k^{(n)} - t_{k-1}^{(n)}}\right) = \frac{2\sigma^2}{m_n^2}\left(\sum_{k=1}^{m_n}\sigma^2 + 2\mu^2(t_k^{(n)} - t_{k-1}^{(n)})\right) = \frac{2\sigma^4}{m_n} + \frac{4\mu^2\sigma^2}{m_n^2} \le \frac{K}{m_n} \tag{4.16}$$

für eine Konstante $K > 0$.
Folglich gilt $D^2_\vartheta\hat\sigma^2_{\mathscr{T}_n} \stackrel{n\to\infty}{\longrightarrow} 0$. Somit haben wir mittels (4.16)

$$\begin{aligned}
\mathbb{P}_\vartheta\big(\mid \hat\sigma^2_{\mathscr{T}_n} - \sigma^2 \mid > \epsilon\big) &= \mathbb{P}_\vartheta\big(\mid \hat\sigma^2_{\mathscr{T}_n} - \mathbb{E}_\vartheta(\hat\sigma^2_{\mathscr{T}_n}) + \mathbb{E}_\vartheta(\hat\sigma^2_{\mathscr{T}_n}) - \sigma^2 \mid > \epsilon\big)\\
&\le \mathbb{P}_\vartheta\big(\mid \hat\sigma^2_{\mathscr{T}_n} - \mathbb{E}_\vartheta(\hat\sigma^2_{\mathscr{T}_n}) \mid + \mid \mathbb{E}_\vartheta(\hat\sigma^2_{\mathscr{T}_n}) - \sigma^2 \mid > \epsilon\big)\\
&\le \mathbb{P}_\vartheta\big(\mid \hat\sigma^2_{\mathscr{T}_n} - \mathbb{E}_\vartheta(\hat\sigma^2_{\mathscr{T}_n}) \mid > \frac{\epsilon}{2}\big) \le \frac{4K}{m_n\epsilon^2}
\end{aligned}$$

falls n so groß ist, daß $\frac{\mid \mathbb{E}_\vartheta(\hat\sigma^2_{\mathscr{T}_n}) - \sigma^2 \mid}{m_n} = \frac{\mid \mu^2T - \sigma^2 \mid}{m_n} < \frac{\epsilon}{2}$ gilt.
Mit Hilfe der Tschebyschevschen Ungleichung ergibt sich die Behauptung. Ist nun $\sum_{n\ge1}\frac{1}{m_n} < \infty$, so folgt für jedes $\epsilon > 0$

$$\mathbb{P}_\vartheta\big(\limsup_{n\to\infty} A_n^\epsilon\big) = 0 \text{ mit } A_n^\epsilon = \big\{\omega\ :\mid \hat\sigma^2_{\mathscr{T}_n}(\omega) - \sigma^2 \mid > \epsilon\big\}. \qquad \text{(Borel-Cantelli-Lemma)}$$

Das bedeutet $\hat\sigma^2_{\mathscr{T}_n} \longrightarrow \sigma^2 \quad \mathbb{P}_\vartheta$-f.s. für jedes $\vartheta \in \Theta$. □

Folgerungen:

a) Bei unbegrenztem Wachstum der Zahl m_n der Beobachtungspunkte läßt sich σ^2 aus $\mathbb{P}_\vartheta$- fast jeder Trajektorie $\big(W(t),\ t \in [0,T]\big)$ beliebig genau bestimmen ($\vartheta = (\mu, \sigma^2)^T$). Das gilt für jedes $T > 0$ (!).

b) Gilt $\sigma^2 = \sigma_0^2$, so ist

$$L^X(\mu\,;\,X) = \exp\left\{\frac{\mu}{\sigma^2}W(T) - \frac{\mu^2}{\sigma^2}T\right\}$$

für jede Stichprobe $X = \big(W(t_1), \ldots, W(t_m)\big)^T$ mit $t_m = T$.

Es gilt also

$$\mathbb{P}_\vartheta(A) = \int_A \exp\left\{\frac{\mu}{\sigma^2}W(T) - \frac{\mu^2}{\sigma^2}T\right\} d\,\mathbb{P}_{\vartheta_0} \quad \text{mit } \vartheta = (\mu, \sigma^2)^T,\ \vartheta_0 = (0, \sigma^2)^T \tag{4.17}$$

für jedes Ereignis A der Form $A = \{\omega\,:\,(W(t_1), \ldots, W(t_m)) \in B\}$, $\quad B \in \mathscr{B}^m$, mit $t_m = T$, d.h. (4.17) ist gültig für jede Zylindermenge $Z = Z_{t_1,\ldots,t_m\,;\,B}$ mit $0 < t_1 < \ldots < t_m = T$, und somit für jedes A aus $\mathscr{B}\big(\mathbb{C}([0,T])\big)$.
Damit haben wir für die Familie

$$\mathscr{P}_{\sigma^2} := \big(\mathbb{P}_\vartheta, \vartheta = (\mu, \sigma^2)^T,\ \mu \in \mathbb{R}\big)$$

bei fixiertem $\sigma^2 > 0$ erhalten, daß alle $\mathbb{P}_\vartheta^T := \mathbb{P}_\vartheta\,\big|_{\mathscr{B}(\mathbb{C}([0,T]))}$ zueinander äquivalent sind, und daß die Likelihoodfunktion

$$L(\vartheta\,;\,\mathscr{A}_T) = \frac{\mathbb{P}_\vartheta^T}{\mathbb{P}_{\vartheta_0}^T} \quad \text{für } \vartheta = (\mu, \sigma^2)^T,\ \vartheta_0 = (0, \sigma^2)^T$$

die Form

$$L(\vartheta\,;\,\mathscr{A}_T) = \exp\left\{\frac{\mu}{\sigma^2}W(T) - \frac{\mu^2}{\sigma^2}T\right\}$$

besitzt. Hierbei haben wir die Notation

$$\mathbb{P}_\vartheta^T := \mathbb{P}_\vartheta\,\big|_{\mathscr{A}_T}, \quad \text{mit } \mathscr{A}_T := \sigma\big(W(s),\ s \in [0,T]\big)$$

verwendet.