Kapitel 6

Suffiziente Statistiken

In diesem Kapitel untersuchen wir einen weiteren statistischen Begriff, der eng mit Likelihoodfunktionen zusammenhängt und mit der Frage nach eventuell möglicher Datenreduktion zu tun hat. Wir gehen der Frage nach, in welchen Fällen man an Stelle der gesamten Stichprobe X lediglich Stichprobenfunktionen $H(X)$ zu verwenden braucht, ohne Informationen über den zugrunde liegenden Parameter ϑ zu verlieren.

6.1 Vorbetrachtungen

Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und X als Stichprobe mit Werten in einem meßbaren Raum $(E, \mathscr{E})$. Die Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta^X$ von X auf $(E, \mathscr{E})$ hängt von ϑ ab. Hat man eine konkrete Stichprobe x, also eine Realisierung von X erhalten, so kann man i.a. daraus Schlüsse über die dem Experiment zugrunde liegende Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta$, also den wahren Parameter $\vartheta \in \Theta$ ziehen.

Beispiel 6.1. Es sei $X = (X_1, X_2, \ldots, X_n)$ eine mathematische Stichprobe aus einer auf $[\vartheta, \vartheta + 1]$ gleichmäßig verteilten Grundgesamtheit, $\vartheta \in \mathbb{R}$ sei unbekannt. Hat man eine Realisierung $x = (x_1, \ldots, x_n)$ von X erhalten, so kann man schließen, daß

$$(\max_{k=1,\ldots,n} x_k) - 1 \leq \vartheta \leq \min_{k=1,\ldots,n} x_k$$

gilt. Man erhält also aus x Informationen über ϑ.

Beispiel 6.2. Ist $X = (X_1, X_2, \ldots, X_n)$ eine mathematische Stichprobe aus einer exponentiell mit dem Parameter λ verteilten Grundgesamtheit, so kann man erwarten, daß

$$\bar{X}_n := \frac{1}{n} \sum_{k=1}^{n} X_k$$

in der Nähe von $\mathbb{E}_\lambda(X_1) = \dfrac{1}{\lambda}$ liegt (Gesetz der großen Zahlen).

Beispiel 6.3. Aus der Beobachtung einer Trajektorie $X = \big(W(s),\ s \in [0,T]\big)$ des Wienerschen Prozesses mit den Parametern μ und σ^2 auf dem Intervall $[0,T]$ $(T > 0)$ läßt sich σ^2 exakt berechnen (siehe Kapitel 4).

In jedem der genannten Fälle enthält also die Stichprobe X Informationen über den wahren Parameter ϑ.

Wir kehren zum eingangs formulierten allgemeinen Fall zurück.
Ist $H(\cdot)$ eine Stichprobenfunktion auf $(E, \mathscr{E})$ (d.h. eine Statistik) mit Werten in $(F, \mathscr{F})$, so hängt auch ihre Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta^H$ auf $(F, \mathscr{F})$ im allgemeinen von ϑ ab. Wir haben das im Beispiel 6.2 mit $H(X) = \bar{X}_n$ gesehen.

Es entsteht die Frage, ob in $H(X)$ noch genau soviel Information über ϑ enthalten ist, wie in der Stichprobe X selbst. Ist dies der Fall, so nennt man $H(X)$ eine *suffiziente* (oder auch *erschöpfende*) *Statistik für $\vartheta \in \Theta$ bzw. für $\mathscr{P}$* $= (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$. Wir gehen darauf im Folgenden näher ein.

6.2 Suffiziente Statistiken und suffiziente σ-Algebren

Wir verwenden die Bezeichnungen des Abschnittes 6.1. Ausgangspunkt der folgenden Überlegungen ist die Formel

$$\mathbb{P}_\vartheta^X(B) = \int_F \mathbb{P}_\vartheta^X(B \mid H = h)\, \mathbb{P}_\vartheta^H(\,dh), \quad B \in \mathscr{E}, \tag{6.1}$$

die man als Verallgemeinerung der Formel der totalen Wahrscheinlichkeit ansehen kann. Die Abhängigkeit der Verteilung $\mathbb{P}_\vartheta^X$ von ϑ wird also aufgespaltet in die Abhängigkeit von ϑ der Verteilung $\mathbb{P}_\vartheta^H$ der Statistik H und die Abhängigkeit von ϑ der bedingten Verteilung $\mathbb{P}_\vartheta^X(\cdot \mid H = h)$ von X unter $H = h$. Beschränkt man sich bei der Datenerhebung auf die Statistik $H(X)$ anstelle X, so geht also ein Teil der Abhängigkeit (und damit ein Teil der Information über ϑ, die in X steckt) verloren.

Beispiel 6.4. Es sei $X = (X_1, X_2, \dots, X_n)$ eine mathematische Stichprobe aus einer zweipunktverteilten Grundgesamtheit mit dem Parameter $p \in (0,1)$, genauer, $X_1, X_2, \dots, X_n$ seien unabhängig und identisch verteilt, und es gelte

$$\mathbb{P}_p(X_k = 1) = p, \quad \mathbb{P}_p(X_k = 0) = 1 - p. \qquad \text{(Bernoullischema mit dem Parameter } p)$$

Wir definieren $H(X) := \max_{k=1,\dots,n} X_k$.
Die Statistik H hat die möglichen Werte 0 und 1 und besitzt die Verteilung

$$\mathbb{P}(H=0) = (1-p)^n, \quad \mathbb{P}(H=1) = 1-(1-p)^n.$$

Für $\mathbb{P}_p(X = x \mid H = i), \quad x = (i_1, i_2, \dots, i_n) \in \{0,1\}^n,\ i \in \{0,1\}$ gilt

$$\begin{aligned}\mathbb{P}_p(X = x \mid H = 0) =& 1, \quad \text{falls } x = (0,0,\dots,0), \quad \mathbb{P}_p(X = x \mid H = 0) = 0 \ \text{ sonst},\\ \mathbb{P}_p(X = x \mid H = 1) =& \frac{p^{\sum_{j=1}^n i_j}(1-p)^{n-\sum_{j=1}^n i_j}}{1-(1-p)^n}, \quad \text{falls}\\ & x = (i_1, i_2, \dots, i_n) \in \{0,1\}^n \text{ mit } H(x) = 1.\end{aligned}$$

Offensichtlich enthält $H(X) = \max_{k=1,\dots,n} X_k$ weit weniger Information über den wahren Parameter p als die Stichprobe X selbst, aus der man zum Beispiel durch $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ eine bessere Schätzung für p gewinnt als es auf der Grundlage von $H(X)$ möglich wäre.

Wir behandeln im Folgenden die am Ende des Abschnittes 6.1 aufgeworfene Frage in einem allgemeineren Rahmen.

Es seien $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und $\mathscr{H}$ eine Teil-σ-Algebra von $\mathscr{A}$.

Definition 6.1. Die σ-Algebra $\mathscr{H}$ heißt *suffizient für $\mathscr{P}$ (bzw. für $\vartheta \in \Theta$) hinsichtlich $\mathscr{A}$*, falls es für alle $A \in \mathscr{A}$ eine nicht von ϑ abhängende $\mathscr{H}$-meßbare Zufallsgröße Z_A gibt mit

$$Z_A = \mathbb{E}_\vartheta(\mathbf{1}_A \mid \mathscr{H}) \quad \mathbb{P}_\vartheta\text{-fast sicher}, \quad \forall \vartheta \in \Theta. \tag{6.2}$$

Anstelle Z_A schreiben wir symbolisch $\mathbb{P}(A \mid \mathscr{H})$ um anzudeuten, daß Z_A für jedes ϑ eine Version der bedingten Wahrscheinlichkeit $\mathbb{P}_\vartheta(A \mid \mathscr{H})$ von A unter $\mathscr{H}$ ist, allerdings aber nicht von ϑ abhängt.
Ist $\mathscr{H}$ suffizient, so gilt also

$$\mathbb{P}_\vartheta(A) = \int_\Omega \mathbb{P}(A \mid \mathscr{H})\, d\,\mathbb{P}_\vartheta\big|_{\mathscr{H}} \quad \text{für alle } A \in \mathscr{A},\ \vartheta \in \Theta. \tag{6.3}$$

Hier bestimmt bereits die Abhängigkeit der Einschränkung $\mathbb{P}_\vartheta\big|_{\mathscr{H}}$ von ϑ diejenige der $\mathbb{P}_\vartheta$ auf $\mathscr{A}$.

Definition 6.2. Eine Zufallsgröße T auf $(\Omega, \mathscr{A})$ mit Werten in $(E, \mathscr{E})$ nennt man *suffizient für $\mathscr{P}$ hinsichtlich $\mathscr{A}$*, falls $\mathscr{A}^T := T^{-1}(\mathscr{E})$ suffizient im genannten Sinne ist.

In diesem Fall gilt für alle $A \in \mathscr{A}$ und $\vartheta \in \Theta$:

$$\mathbb{P}_\vartheta(A) = \int_\Omega \mathbb{P}(A \mid \sigma(T))\, d\,\mathbb{P}_\vartheta = \int_E \mathbb{P}(A \mid T = t)\, \mathbb{P}_\vartheta^T(\,dt). \tag{6.4}$$

Hier bestimmt die Abhängigkeit der Verteilung $\mathbb{P}_\vartheta^T$ von ϑ bereits die Abhängigkeit der Maße $\mathbb{P}_\vartheta$ von ϑ.

Bemerkung 6.3. Anstelle der σ-Algebra $\mathscr{A}$ setzt man häufig auch $\mathscr{A}^X$ für eine Stichprobe X. In diesem Fall fordert man in der Definition der Suffizienz von $\mathscr{H}$ die Eigenschaft $\mathscr{H} \subseteq \mathscr{A}^X$, was im Fall $\mathscr{H} = \mathscr{A}^T$ für eine Zufallsgröße T bedeutet, daß T sich als meßbare Funktion H von X darstellen läßt:

$$T(\omega) = H(X(\omega)).$$

Damit haben wir den im Abschnitt 6.1 betrachteten Fall eingeordnet.

Beispiel 6.5. $(X_1, X_2, \ldots, X_n)$ sei ein Bernoullischema mit dem Parameter $p \in (0,1)$. (Die $X_k,\ k = 1, \ldots, n$ sind unabhängig und es gilt $\mathbb{P}_p(X_k = i) = p^i(1-p)^{1-i},\ i = 0, 1$.) Dann ist $S_n := \sum_{k=1}^n X_k$ eine suffiziente Statistik für $(\mathbb{P}_p,\, p \in (0,1)\,)$ hinsichtlich $\mathscr{A}^X$.
Beweis:
Die σ-Algebra $\mathscr{A}^X$ wird erzeugt durch die Zerlegung

$$\{\omega \mid X(\omega) = (i_1, i_2, \ldots, i_n)\}\,, \quad (i_1, i_2, \ldots, i_n) \in \{0,1\}^n.$$

Es gilt

$$\mathbb{P}_p\big(X = (i_1, i_2, \ldots, i_n) \mid S_n = m\big) = \mathbf{1}_{\{m\}}(i_1 + \ldots + i_n)\frac{p^{\sum_{j=1}^n i_j}(1-p)^{n-\sum_{j=1}^n i_j}}{\mathbb{P}_p(S_n = m)}$$
$$= \frac{\mathbf{1}_{\{m\}}(i_1 + \ldots + i_n)}{\binom{n}{m}}$$

und diese bedingte Wahrscheinlichkeit ist unabhängig von p. $\square$

Beispiel 6.6. Es sei $\big(W(t),\, t \in [0,T]\big)$ ein Wienerscher Prozess mit den Parametern μ und σ^2. Der Parameter σ^2 sei bekannt, μ sei unbekannt. Dann ist $W(T)$ eine suffiziente Statistik für μ bezüglich $\mathscr{A}^T = \sigma\big(W(t),\, t \in [0,T]\big)$.

Beweis:
Ist Z eine Zylindermenge der Form

$$Z = \{\omega \in \Omega \mid \big(W(t_1), \ldots, W(t_n)\big) \in B\}$$

für gewisse $t_1 < t_2 < \ldots < t_n = T$ und ein $B \in \mathscr{B}^n$, so gilt

$$\mathbb{P}_\mu\big(Z \mid W(T) = w\big) = \int_B \cdots \int \varphi_{t_1,\ldots,t_n}(x_1, \ldots, x_{n-1}, w)\, dx_1 \ldots dx_{n-1}$$

$$\text{mit} \quad \varphi_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi(t_k - t_{k-1})\sigma^2}} \frac{\exp\left\{-\frac{1}{2}\sum_{k=1}^n \frac{(x_k - x_{k-1} - \mu(t_k - t_{k-1}))^2}{\sigma^2(t_k - t_{k-1})}\right\}}{\exp\left\{-\frac{1}{2}\frac{(w-\mu T)^2}{\sigma^2 T}\frac{1}{\sqrt{2\pi T\sigma^2}}\right\}}.$$

Eine kurze Rechnung zeigt, daß dieser Ausdruck nicht vom Parameter μ abhängt. Wenn $\mathbb{P}_\mu(A \mid W(T) = w)$ für jede Zylindermenge A aus $\mathscr{A}^T$ nicht von μ abhängt, so gilt dasselbe für alle $A \in \mathscr{A}^T$.

6.3 Suffiziente σ-Algebren und Statistiken in dominierten Modellen

Es sei $\mathscr{P} = (\mathbb{P}_\vartheta, \vartheta \in \Theta)$ eine durch ein σ-finites Maß μ auf dominierte Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathscr{A})$. Die Likelihoodfunktion $L(\vartheta\,; \omega)$ ist definiert durch

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

und ist für jedes $\vartheta \in \Theta$ μ-fast überall eindeutig bestimmt.

Die Suffizienz einer σ-Algebra $\mathscr{H}$ bzw. einer Statistik T (mit Werten in einem meßbaren Raum $(E, \mathscr{E})$) läßt sich an Hand von L feststellen. Das ist der Inhalt des folgenden Satzes. Mit seiner Hilfe werden wir weitere Beispiele suffizienter σ-Algebren und Statistiken gewinnen.
Wir beginnen mit einem Lemma technischer Natur.

Lemma 6.4. *Die Familie $\mathscr{P} = (\mathbb{P}_\vartheta, \vartheta \in \Theta)$ sei dominiert durch ein σ-finites Maß μ. Dann existiert eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $(\Omega, \mathscr{A})$, so daß gilt:*

a) $\mathbb{P}(A) = 0 \iff (\mathbb{P}_\vartheta(A) = 0, \ \forall \vartheta \in \Theta)$ *für alle* $A \in \mathscr{A}$.

b) Die Verteilung $\mathbb{P}$ aus a) kann als konvexe Linearkombination $\mathbb{P} = \sum_{k\geq 1} c_k\, \mathbb{P}_{\vartheta_k}$ einer höchstens abzählbar unendlichen Parametermenge $\{\vartheta_k, k \geq 1\} \subseteq \Theta$ gewählt werden.

(Für einen Beweis siehe z.B. Dacunha-Castelle, Duflo, Band I.)

Eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ mit den Eigenschaften a) und b) nennt man *ein bezüglich* $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ *privilegiertes Maß*. Offenbar gilt $\mathbb{P} \ll \mu$ und

$$L(\vartheta\,;\,\omega) = \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \cdot \frac{d\,\mathbb{P}}{d\mu}. \tag{6.5}$$

Zur Untersuchung der Abhängigkeit der Likelihoodfunktion L von ϑ genügt es also, sich auf privilegierte Maße zu beschränken.
Sind $\mathbb{P}$ und $\mathbb{P}'$ zwei privilegierte dominierende Wahrscheinlichkeitsmaße bezüglich $\mathscr{P}$, so sind sie äquivalent: $\mathbb{P} \equiv \mathbb{P}'$.

Folgerung 6.5. Ist $\mathbb{P}$ ein privilegiertes Wahrscheinlichkeitsmaß bezüglich $\mathscr{P}$, so gilt für jede nichtnegative Zufallsgröße Z auf $(\Omega, \mathscr{A})$ die Gleichung

$$\mathbb{E}_{\mathbb{P}}(Z) = \sum_{k \geq 1} c_k\, \mathbb{E}_{\vartheta_k}(Z) \tag{6.6}$$

mit der Bezeichnung aus Punkt b) des obigen Lemmas.

Beweis:
Die Gleichung (6.6) ist richtig für $Z = \mathbf{1}_A,\ A \in \mathscr{A}$, somit auch für $Z = \sum_{k=1}^m \alpha_k \mathbf{1}_{A_k}$. Nun wendet man die Approximationsmethode an. □

Aussage 6.1 (Charakterisierungssatz). *Folgende Eigenschaften sind äquivalent:*

i) $\mathscr{H}$ *(bzw.* T*) ist suffizient für* $\mathscr{P}$ *hinsichtlich* $\mathscr{A}$

ii) Für jedes bezüglich $\mathscr{P}$ *privilegiertes Wahrscheinlichkeitsmaß* $\mathbb{P}$ *und jedes* $\vartheta \in \Theta$ *gibt es eine* $\mathbb{P}$*-Version von* $\dfrac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$*, die bezüglich* $\mathscr{H}$ *(bzw.* $\sigma(T)$*) meßbar ist*[1]

iii) für jedes σ*-finite dominierende Maß* μ *gibt es eine* $\mathscr{A}$*-meßbare Zufallsgröße* h *(unabhängig von* ϑ*) und eine* $\mathscr{H}$*-meßbare Zufallsgröße* G_ϑ *(bzw. eine Borelmeßbare Funktion* g_ϑ *von* $(E, \mathscr{E})$ *in* $(\mathbb{R}, \mathscr{B})$*), so daß gilt:*

$$L(\vartheta\,;\,\omega) = h(\omega) G_\vartheta(\omega) \quad \mu\text{-fast überall}\ , \vartheta \in \Theta.$$
$$\text{(bzw. } L(\vartheta\,;\,\omega) = h(\omega) g_\vartheta(T(\omega)) \quad \mu\text{-fast überall}\ , \vartheta \in \Theta)^1$$

[1]Man beachte: Eine reellwertige Zufallsgröße Y ist $\sigma(T)$-meßbar genau dann, wenn es eine $(\mathscr{E}$-$\mathscr{B})$-meßbare Funktion ψ von E in $\mathbb{R}$ gibt mit $Y(\omega) = \psi(T(\omega)),\ \omega \in \Omega$.

Beweis:
(i)$\Longrightarrow$(ii):
$\mathscr{H}$ sei suffizient für $\mathscr{P}$ hinsichtlich $\mathscr{A}$ und $\mathbb{P}$ sei ein privilegiertes Maß für $\mathscr{P}$. Wir verwenden die Bezeichnungen des Abschnittes 6.2. Nach Definition gibt es für jedes A eine nicht von ϑ abhängende Zufallsgröße Z_A, die $\mathscr{H}$-meßbar ist und für die gilt

$$Z_A = \mathbb{P}_\vartheta(A \mid \mathscr{H}) \quad \mathbb{P}_\vartheta\text{-f.s.}, \quad \forall \vartheta \in \Theta.$$

Das bedeutet

$$\int_H Z_A \, d\,\mathbb{P}_\vartheta = \mathbb{P}_\vartheta(A \cap H) \quad \text{für alle } H \in \mathscr{H} \text{ und alle } \vartheta \in \Theta.$$

Daraus ergibt sich (siehe (6.6))

$$\int_H Z_A \, d\,\mathbb{P} = \mathbb{P}(A \cap H)$$

und somit $\mathbb{E}_{\mathbb{P}}(\mathbf{1}_A \mid \mathscr{H}) = Z_A \quad \mathbb{P}$-fast sicher.
Sind $\vartheta \in \Theta$ und $A \in \mathscr{A}$, so gilt folglich

$$\begin{aligned}
&\mathbb{P}_\vartheta(A) = \int_\Omega \mathbb{E}_{\mathbb{P}}(\mathbf{1}_A \mid \mathscr{H}) \, d\,\mathbb{P}_\vartheta = \int_\Omega \mathbb{E}_{\mathbb{P}}(\mathbf{1}_A \mid \mathscr{H}) \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \cdot d\,\mathbb{P} = \\
&\int_\Omega \mathbb{E}_{\mathbb{P}}\Big(\mathbb{E}_{\mathbb{P}}(\mathbf{1}_A \mid \mathscr{H}) \cdot \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}\Big) \, d\,\mathbb{P} = \int_\Omega \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_A \mid \mathscr{H}\right) \cdot \mathbb{E}_{\mathbb{P}}\left(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}\right) d\,\mathbb{P} = \\
&\int_\Omega \mathbb{E}_{\mathbb{P}}\Big(\mathbf{1}_A \, \mathbb{E}_{\mathbb{P}}(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}) \mid \mathscr{H}\Big) \, d\,\mathbb{P} = \int_A \mathbb{E}_{\mathbb{P}}\Big(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}\Big) \, d\,\mathbb{P}\,. \qquad (6.7)
\end{aligned}$$

Andererseits gilt

$$\mathbb{P}_\vartheta(A) = \int_A \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \, d\,\mathbb{P}, \quad \forall\, A \in \mathscr{A}. \qquad (6.8)$$

Somit folgt aus (6.7) und (6.8)

$$\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} = \mathbb{E}_{\mathbb{P}}\left(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}\middle|\mathscr{H}\right) \quad \mathbb{P}\text{-fast überall},$$

d.h., es gibt eine $\mathscr{H}$-meßbare Version von $\dfrac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$.
(ii)$\Longrightarrow$(iii)
Dieser Schritt ergibt sich aus Formel (6.5).
(ii)$\Longrightarrow$(i)

Mit der Formel für die Umrechnung bedingter Erwartungen aus der 4. Übung ergibt sich für jede nichtnegative Zufallsgröße Y

$$\mathbb{E}_\vartheta(Y \mid \mathscr{H}) = \frac{\mathbb{E}_{\mathbb{P}}\left(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} Y \mid \mathscr{H}\right)}{\mathbb{E}_{\mathbb{P}}\left(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}\right)} = \mathbb{E}_{\mathbb{P}}(Y \mid \mathscr{H}) \quad \mathbb{P}\text{-fast sicher für alle } \vartheta \in \Theta. \tag{6.9}$$

(iii)$\Longrightarrow$(ii)
Es sei $\mathbb{P}$ privilegiert, folglich gilt $\mathbb{P} \ll \mu$, und

$$\frac{d\,\mathbb{P}}{d\mu} = h \sum c_k G_{\vartheta_k} = h(\omega) \cdot G, \tag{6.10}$$

G ist nach Konstruktion $\mathscr{H}$-meßbar. Nach Voraussetzung ist

$$\frac{d\,\mathbb{P}_\vartheta}{d\mu} = h(\omega) G_\vartheta(\omega). \tag{6.11}$$

Aus den beiden Gleichungen (6.10) und (6.11) erhalten wir

$$h(\omega)G_\vartheta(\omega) = \frac{d\,\mathbb{P}_\vartheta}{d\mu} = \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \cdot \frac{d\,\mathbb{P}}{d\mu} = \frac{d\,\mathbb{P}_\vartheta}{d\mu} \cdot h(\omega)G(\omega) \quad \mu\text{-fast sicher.} \tag{6.12}$$

Wegen

$$\mathbb{P}_\vartheta(h=0) = \int_{\{h=0\}} \frac{d\,\mathbb{P}_\vartheta}{d\mu}\, d\mu = \int_{\{h=0\}} h \cdot G_\vartheta \, d\mu = 0, \quad \vartheta \in \Theta,$$

gilt $\mathbb{P}(h=0) = 0$. Analog ergibt sich $\mathbb{P}(G=0) = 0$ aus (6.10). Somit folgt aus (6.12)

$$\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} = \frac{G_\vartheta(\omega)h(\omega)}{G(\omega)h(\omega)} = \frac{G_\vartheta(\omega)}{G(\omega)} \quad \mathbb{P}\text{-fast sicher,}$$

und $\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$ hat somit eine $\mathscr{H}$-meßbare Version. □

6.4 Minimal suffiziente Statistiken und σ-Algebren

Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistischer Raum mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$. $\mathscr{P}$ werde durch ein σ-finites Maß μ dominiert und $\mathscr{H}$ sei eine suffiziente σ-Algebra für $\mathscr{P}$ hinsichtlich $\mathscr{A}$.
Ist $\mathscr{H}'$ eine Teil-σ-Algebra von $\mathscr{A}$ mit $\mathscr{H} \subseteq \mathscr{H}'$, so ist auch $\mathscr{H}'$ suffizient. Das ergibt sich unmittelbar aus dem Faktorisierungssatz. Da andererseits nicht jede Teil-σ-Algebra $\mathscr{H}''$ von $\mathscr{H}$ suffizient ist (man prüfe das für die triviale σ-Algebra $\mathscr{H}'' = \{\emptyset, \Omega\}$), ergibt sich die Frage, ob es eine minimale suffiziente σ-Algebra gibt und ob diese gegebenenfalls

eindeutig ist.

Es seien $\mathbb{P}$ eine privilegierte dominierende Wahrscheinlichkeitsverteilung für $\mathscr{P}$ und

$$L(\vartheta\,;\,\omega) := \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}(\omega) \ \text{ die Likelihoodfunktion.}$$

Die σ-Algebra $\mathscr{H}_0$, definiert durch

$$\mathscr{H}_0 := \sigma\big(L\,(\vartheta\,;\,\cdot),\ \vartheta \in \Theta\big)$$

(kleinste σ-Algebra von Teilmengen von Ω, bezüglich der alle $L\,(\vartheta\,;\,\cdot)\,,\ \vartheta \in \Theta$, meßbar sind)

ist eine suffiziente σ-Algebra für $\mathscr{P}$ bezüglich $\mathscr{A}$. Davon überzeugt man sich leicht durch Anwendung von Formel (6.9). Für Z_A ergibt sich dabei $\mathbb{P}(A \mid \mathscr{H}_0)$.
Die σ-Algebra $\mathscr{H}_0$ hängt davon ab, welche Versionen der Zufallsgrößen $L\,(\vartheta\,;\,\cdot)$ zu ihrer Bildung eingesetzt werden. ($L\,(\vartheta\,;\,\cdot)$ ist ja nach Definition nur $\mathbb{P}$-fast sicher bestimmt.)
Um von diesem Sachverhalt unabhängig zu werden, gehen wir zu Vervollständigung $\mathscr{H}_0^{\mathbb{P}}$ von $\mathscr{H}_0$ bezüglich $\mathbb{P}$ über.
(Bezeichnet man mit $\mathscr{M}$ die Menge aller $A \in \mathscr{A}$ mit $\mathbb{P}(A) = 0$, so ist $\mathscr{M}_0^{\mathbb{P}}$ definiert als $\sigma\big(\mathscr{M}_0 \cup \mathscr{M}\big)$. Es gilt: X ist $\mathscr{M}_0^{\mathbb{P}}$-meßbar genau dann, wenn es eine $\mathscr{M}_0$-meßbare Zufallsgröße $\tilde{X}$ gibt mit $\mathbb{P}(X \neq \tilde{X}) = 0$.)

Lemma 6.6. *Sind $\mathscr{H}_0$ und $\tilde{\mathscr{H}_0}$ zwei σ-Algebren , die durch verschiedene Versionen von $L\,(\vartheta\,;\,\cdot)\,,\ \vartheta \in \Theta$, erzeugt werden, so gilt*

$$\mathscr{H}_0^{\mathbb{P}} = \tilde{\mathscr{H}}_0^{\mathbb{P}}.$$

Beweis:
Jede Version von $L\,(\vartheta\,;\,\cdot)$ ist meßbar bezüglich $\mathscr{H}_0^{\mathbb{P}}$ und auch bezüglich $\tilde{\mathscr{H}}_0^{\mathbb{P}}$. Daraus folgt

$$\mathscr{H}_0 \subseteq \tilde{\mathscr{H}}_0^{\mathbb{P}} \text{ und } \tilde{\mathscr{H}_0} \subseteq \mathscr{H}_0^{\mathbb{P}} \text{ und somit } \mathscr{H}_0^{\mathbb{P}} \subseteq \tilde{\mathscr{H}}_0^{\mathbb{P}},\ \text{ sowie } \tilde{\mathscr{H}}_0^{\mathbb{P}} \subseteq \mathscr{H}_0^{\mathbb{P}}.$$

□

Die σ-Algebra $\mathscr{H}_0^{\mathbb{P}}$ ist suffizient (siehe die Bemerkung eingangs dieses Abschnittes) und minimal im Sinne der folgenden Aussage

Aussage 6.2. *Ist $\mathscr{H}$ eine suffiziente σ-Algebra für $\mathscr{P}$ bezüglich $\mathscr{A}$, so gilt $\mathscr{H}_0^{\mathbb{P}} \subseteq \mathscr{H}^{\mathbb{P}}$.*

Beweis:
Ist $\mathscr{H}$ suffizient, so besitzt $\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$ auf Grund des Faktorisierungssatzes eine $\mathscr{H}$-meßbare Version, folglich ist $\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$ eine $\mathscr{H}^{\mathbb{P}}$-meßbare Zufallsgröße für alle $\vartheta \in \Theta$. Somit gilt $\mathscr{H}_0 \subseteq \mathscr{H}^{\mathbb{P}}$ und damit $\mathscr{H}_0^{\mathbb{P}} \subseteq \mathscr{H}^{\mathbb{P}}$. □

Diese Aussage führt zu folgender

Definition 6.7. Eine suffiziente σ-Algebra $\mathscr{H}$ heißt *minimal suffizient*, falls $\mathscr{H}^{\mathbb{P}} \subseteq \tilde{\mathscr{H}}^{\mathbb{P}}$ für alle suffizienten σ-Algebren $\tilde{\mathscr{H}}$ gilt.

Die obige Aussage erlaubt nun die

Folgerung 6.8. $\mathscr{H}$ ist minimal suffizient genau dann, wenn $\mathscr{H}^{\mathbb{P}} = \mathscr{H}_0^{\mathbb{P}}$ gilt.

Definition 6.9. Eine Zufallsgröße H auf $(\Omega, \mathscr{A})$ heißt *minimal suffizient*, wenn $\sigma(H)$ minimal suffizient ist.

Das bedeutet, daß es zu jedem $\vartheta \in \Theta$ eine $\mathbb{P}$-Version von $L(\vartheta\,;\,\cdot)$ gibt mit

$$\sigma(H) = \sigma\left(L(\vartheta\,;\,\cdot) \mid \vartheta \in \Theta\right).$$

Kurz ausgedrückt heißt das: Ist die Likelihoodfunktion für jedes $\vartheta \in \Theta$ in Abhängigkeit von ω die Funktion einer Statistik $H = H(\omega)$, so ist H minimal suffizient.

Beispiel 6.7. $X = (X_1, \ldots, X_n)$ sei eine mathematische Stichprobe aus einer gleichmäßig auf $[a, b]$ verteilten Grundgesamtheit, a und b seien unbekannt. Wir setzen

$$\Theta = \{\vartheta = (a, b)\ :\ -\infty < a < b < \infty\}.$$

Als dominierendes Maß für $\mathscr{P}^X$ verwenden wir das Lebesguemaß λ_n auf $(\mathbb{R}^n, \mathscr{B}^n)$. Dann gilt

$$L^X(\vartheta\,;\,X(\omega)) = (b-a)^{-n} \prod_{k=1}^n \mathbf{1}_{[a,b]}(X_k) = (b-a)^{-n} \mathbf{1}_{[a,b]}(\min_{k=1,\ldots,n} X_k) \cdot \mathbf{1}_{[a,b]}(\max_{k=1,\ldots,n} X_k).$$

Folglich ist die Stichprobenfunktion $T(X) = \left(\min_{k=1,\ldots,n} X_k, \max_{k=1,\ldots,n} X_k\right)$ eine suffiziente Statistik für $\mathscr{P}\big|_{\mathscr{A}^X}$ bezüglich $\mathscr{A}^X$, sie ist sogar minimal suffizient.

Beispiel 6.8. Es sei $X = (X_0, X_1, \ldots, X_n)$ eine Markovsche Kette mit dem Zustandsraum $E = \{e_1, e_2, \ldots, e_m\}$, der Anfangsverteilung ρ und den Übergangswahrscheinlichkeiten $\pi_{ij},\ i, j \in E$. Als unbekannte Parameter werden die π_{ij} angesehen: $\Theta := \left\{\vartheta = (\pi_{ij})_{i,j\in E}\right\}$. Die Anfangsverteilung sei bekannt.

Dann gilt

$$L_n(\vartheta\,;\,X) = \rho(X_0) \cdot \pi_{X_0X_1} \cdot \pi_{X_1X_2} \cdot \ldots \cdot \pi_{X_{n-1}X_n} = \rho(X_0) \prod_{i,j\in E} \pi_{ij}^{N_n^{ij}}$$

mit

$$N_n^{ij} = \sum_{l=0}^{n-1} \mathbf{1}_{\{X_l = i, X_{l+1} = j\}}.$$

Folglich ist $(N_n^{ij},\, i, j \in E)$ eine minimal suffiziente Statistik für $\big(\mathbb{P}_\vartheta^X,\, \vartheta \in \Theta\big)$ bezüglich $\mathscr{A}^X$.

Beispiel 6.9. Es sei $(X(t),\, t \geq 0)$ ein Ornstein-Uhlenbeck-Prozess, d.h. eine Lösung der Differentialgleichung

$$\begin{aligned} &dX(t) = \rho X(t)\, dt + \sigma\, dW(t), \quad t \geq 0 \\ &X(0) = X_0. && \text{(Anfangsbedingung)} \end{aligned}$$

Dann gilt

$$\frac{d\,\mathbb{P}_\rho^T}{d\,\mathbb{P}_0^T}(\omega) = \exp\left\{\rho \int_0^T X(t)\, dX(t) - \frac{\rho^2}{2} \int_0^T X^2(t)\, dt\right\}.$$

Folglich ist

$$H\big((X(t),\ 0 \leq t \leq T)\big) = \left(\int_0^T X(t)\, dX(t),\ \int_0^T X^2(t)\, dt\right)$$

eine minimal suffiziente Statistik für ρ.