Kapitel 7

Exponentialfamilien

Exponentialfamilien sind dominierte statistische Räume, deren Likelihoodfunktion eine besonders einfache Struktur besitzt, ihr Logarithmus ist von affiner Gestalt. Neben der daraus resultierenden sehr guten analytischen Handhabbarkeit zeichnen sie sich durch zahlreiche statistische Eigenschaften aus. Hinzu kommt, daß viele der gängigen klassischen statistischen Räume Exponentialfamilien bilden.
In diesem Kapitel untersuchen wir Exponentialfamilien von Verteilungen und von Lévy-Prozessen. Ausführliche und tiefer gehende Ergebnisse findet man in Küchler, Sørensen, (1997).

7.1 Wahrscheinlichkeitsverteilungen

Es sei μ ein σ-finites Maß auf $(\mathbb{R}^k, \mathscr{B}^k)$. Wir definieren

$$I := \left\{ u \in \mathbb{R}^k \ : \ \int_{\mathbb{R}^k} \exp\left(< u, x >\right) \mu\left(dx\right) < \infty \right\}$$

und setzen voraus, daß I einen nichtleeren offenen Kern enthält: $\overset{\circ}{I} \neq \emptyset$. Für jedes $u \in I$ definieren wir ein Wahrscheinlichkeitsmaß $\mathbb{P}^{(u)}$ auf $\mathscr{B}^k$ durch

$$\mathbb{P}^{(u)}(A) = \int_A \exp\big\{< u, x > -\psi(u)\big\}\mu(\, dx), \quad A \in \mathscr{B}^k$$

mit

$$\psi(u) := \ln \int_{\mathbb{R}^k} \exp\big\{< u, x >\big\}\mu\left(\, dx\right), \quad u \in I.$$

Definition 7.1. Für jede Teilmenge J von I mit mehr als einem Element nennt man $(\mathbb{P}^{(u)},\ u \in J)$ eine *Exponentialfamilie von Wahrscheinlichkeitsmaßen auf* $\mathbb{R}^k$*, erzeugt von* μ. Der Parameter u heißt *kanonischer Parameter* der Exponentialfamilie.

Exponentialfamilien bilden statistische Räume und jedes $\mathbb{P}^{(u_0)}$ aus $\mathscr{P} = (\mathbb{P}^{(u)},\ u \in I)$ mit $u_0 \in I$ ist ein dominierendes privilegiertes Wahrscheinlichkeitsmaß für $\mathscr{P}$.

Beispiel 7.1.

$$\mu = \mathcal{N}(0,1), \quad d\,\mathbb{P}^{(u)} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2} + ux - \frac{u^2}{2}\right\} dx = \exp\left\{ux - \frac{u^2}{2}\right\} d\mu,$$

$\psi(u) = \frac{u^2}{2}, \quad u \in I = \mathbb{R}\,.$ $\quad \mathbb{P}^{(u)}$ ist eine Normalverteilung mit dem Erwartungswert u.

Beispiel 7.2.

$$\mu(\{k\}) = \frac{\lambda_0^k}{k!} e^{-\lambda_0}, \qquad \mathbb{P}^{(u)}\left(\{k\}\right) = \exp\left\{uk - \psi(u)\right\} \mu\left(\{k\}\right),\ u \in I = \mathbb{R},$$

mit $\psi(u) = \lambda_0(e^u - 1)$. $\quad \mathbb{P}^{(u)}$ ist eine Poissonverteilung mit dem Parameter $\lambda_0 e^u$.

Beispiel 7.3.

$$\mu = \Gamma(\alpha_0, \lambda_0), \quad \alpha_0, \lambda_0 > 0, \ \text{ d.h. } \mu(\,dx) = \frac{1}{\Gamma(\alpha_0)} x^{\alpha_0 - 1} \lambda_0^{\alpha_0} e^{-\lambda_0 x} \mathbf{1}_{(0,\infty)}(x)\,dx, \quad x \in \mathbb{R}\,.$$

$$\mathbb{P}^{(u)}(\,dx) = \frac{1}{\Gamma(\alpha_0)} x^{\alpha_0 - 1} \lambda_0^{\alpha_0} e^{-(\lambda_0 - u)x} e^{-\psi(u)} \mathbf{1}_{(0,\infty)}(x)\,dx$$

mit $\psi(u) = \alpha_0 \ln\left(\frac{\lambda_0}{\lambda_0 - u}\right)$ und $u \in I = (-\infty, \lambda_0)$.

$\mathbb{P}^{(u)}$ ist eine $\Gamma(\alpha_0, \lambda_0 - u)$-Verteilung.

Ist γ eine Abbildung von $\Theta \subseteq \mathbb{R}^m$ in I, so nennt man auch $\left(\mathbb{P}_\vartheta := \mathbb{P}^{(\gamma(\vartheta))},\ \vartheta \in \Theta\right)$ eine Exponentialfamilie (sofern $\gamma(\Theta)$ mehrelementig ist).

Die folgende Aussage ist eine mehrdimensionale Verallgemeinerung der Aussagen, die wir in der ersten Übung kennengelernt haben.

Aussage 7.1. *Die Menge I ist konvex. Wenn $u \in \mathring{I}$ gilt, so ist $\psi(\cdot)$ in u unendlich oft differenzierbar, und es gilt für alle $i_j \geq 0,\ j = 1, 2, \ldots, k$:*

$$\int |\, x_1^{i_1} x_2^{i_2} \cdots x_k^{i_k} \,|\, e^{<u,x>} \mu(\,dx) < \infty$$

$$\frac{\partial^{i_1 + \ldots + i_k}}{\partial u_1^{i_1} \ldots \partial u_k^{i_k}} \exp\left\{\psi(u)\right\} = \int x_1^{i_1} x_2^{i_2} \cdots x_k^{i_k} e^{<u,x>} \mu(\,dx).$$

Für einen Beweis siehe Dacunha-Castelle, Duflo, Band I.

Es seien $H = (H_1, H_2, \dots, H_k)$ eine $\mathbb{R}^k$-wertige Zufallsgröße über $(\Omega, \mathscr{A})$ und ν ein σ-finites Maß auf $(\Omega, \mathscr{A})$.
Dann ist durch

$$\mathbb{P}^{(u)}(A) = \int_A \exp\{< u, H(\omega) > -\psi(u)\}\, \nu(\,d\omega)$$

mit

$$\psi(u) = \ln \int_\Omega \exp\{< u, H(\omega) >\}\, \nu(\,d\omega), \quad u \in I$$

und

$$I := \Big\{ v \in \mathbb{R}^k \ : \ \int_\Omega \exp\big\{< u, H(\omega) >\big\}\nu(\,d\omega) < \infty \Big\}$$

auf $(\Omega, \mathscr{A})$ eine Familie $\mathscr{P}_H$ von Wahrscheinlichkeitsmaßen $\mathbb{P}^{(u)}$, $u \in I$ gegeben, die wir *die von H und ν erzeugte Exponentialfamilie* nennen. Die Zufallsgröße H heißt *kanonische Zufallsgröße für $\mathscr{P}_H$.*
Die Anwendung der vorangegangenen Aussage auf $\nu^H := \nu \circ H^{-1}$ liefert:

Aussage 7.2. 1. *I ist konvex und $\psi(\cdot)$ ist im Inneren von I unendlich oft differenzierbar,*

2. *Wenn $u \in \overset{\circ}{I}$, so sind alle Momente $\int_\Omega H_1^{i_1} H_2^{i_2} \cdot \ldots \cdot H_k^{i_k} \exp\big\{< u, H > -\psi(u)\big\}\mu(\,d\omega)$ endlich, und es gilt*

$$\operatorname{grad} \psi(u) \cdot \exp\{\psi(u)\} = \int H \exp\{< u, H >\}\, d\mu, \quad \textit{also} \quad \mathbb{E}^{(u)}(H) = \operatorname{grad}\psi(u),$$

3. *Für $u \in \overset{\circ}{I}$ haben wir $Kov(H_i, H_j) = \Big(\dfrac{\partial^2 \psi}{\partial u_i \partial u_j}\Big)(u)$.*

Aussage 7.3. *Die Stichprobenfunktion H ist eine minimal suffiziente Statistik für $\mathscr{P}_H$ bezüglich $\mathscr{A}$.*

Beweis:
Wir erinnern zunächst an die Bezeichnung $\mathscr{M}_0 := \sigma\big(L(\vartheta\,;\,\cdot) \mid \vartheta \in \Theta\big)$. Es gilt $\mathscr{M}_0 \subseteq \sigma(H)$, da $e^{<u,H>}$ bezüglich $\sigma(H)$ meßbar für alle $u \in \mathbb{R}^k$ ist.

Andererseits, ist $u \in \overset{\circ}{I}$ und v irgend ein Element von $\mathbb{R}^k$, so gilt

$$< v, H >= \lim_{\substack{h \to 0 \\ h \in \mathbb{R}}} \frac{\exp\{< u + vh, H >\} \exp\{< u, H >\}}{h \exp\{< u, H >\}}.$$

Also ist für jedes $v \in \mathbb{R}^k$ die Zufallsgröße $< v, H >$ bezüglich $\mathscr{M}_0$ meßbar, und somit auch H selbst, d.h., es gilt $\sigma(H) \subseteq \mathscr{M}_0$. $\square$

$(\Omega, \mathscr{A}, \mathscr{P})$ bilde eine Exponentialfamilie mit der kanonischen Zufallsgröße H.
Das zufällige Experiment werde n-mal mit demselben wahren Parameter $\vartheta \in \Theta$ unabhängig voneinander durchgeführt. $(\Omega_k, \mathscr{A}_k, \mathbb{P}_\vartheta)$ sei der Wahrscheinlichkeitsraum für das k-te Experiment, ω_k sei das Ergebnis des k-ten Experiments, A_k Ereignisse, die mit dem k-ten Experiment zusammenhängen. Dann wird das Gesamtexperiment, das aus den n Einzelexperimenten besteht, beschrieben durch:

$$\tilde{\Omega} := \prod_{j=1}^{n} \Omega_j, \quad \tilde{\mathscr{A}} := \bigotimes_{j=1}^{n} \mathscr{A}_j, \quad \mathscr{P}^{\otimes n} := \left(\mathbb{P}_\vartheta^{\otimes n} \ : \ \vartheta \in \Theta\right)$$

mit

$$\mathbb{P}_\vartheta^{\otimes n} := \mathbb{P}_\vartheta \otimes \cdots \otimes \mathbb{P}_\vartheta, \quad \omega = (\omega_1, \ldots, \omega_n) \in \tilde{\Omega}.$$

Für die diesem Gesamtexperiment entsprechende Wahrscheinlichkeitsverteilung gilt

$$\begin{aligned}
&\mathbb{P}_\vartheta \otimes \cdots \otimes \mathbb{P}_\vartheta(A_1 \times \ldots \times A_n) = \\
&= \prod_{j=1}^{n} \mathbb{P}_\vartheta(A_j) = \prod_{j=1}^{n} \int_{A_j} \exp\Big\{< \gamma(\vartheta), H(\omega_j) > -\psi(\vartheta)\Big\}\, d\mu(\omega_j) \\
&\qquad = \int_{A_1 \times \ldots \times A_n} \!\!\cdots \int \exp\left\{< \gamma(\vartheta), \sum_{j=1}^{n} H(\omega_j) > -n\psi(\vartheta)\right\} d\mu^{\otimes n}(\omega).
\end{aligned}$$

$$\mathbb{P}^{\otimes n}(A) = \int_A \exp\left\{< \gamma(\vartheta), \sum_{j=1}^{n} H(\omega_j) > -n\psi(\vartheta)\right\} d\mu^{\otimes n}(\omega), \quad A \in \mathscr{A}.$$

Folgerung 7.2. Ist $X = X = (X_1, \ldots, X_n)$ eine mathematische Stichprobe und bildet $(\mathbb{P}_\vartheta^{X_1})$ eine Exponentialfamilie mit der kanonischen Variablen $H(x)$, so ist $H(X_1) + \ldots + H(X_n)$ eine minimal suffiziente Statistik für $\mathscr{P}^X$.

Beweis:
Es gilt

$$L^X(\vartheta; x) = \prod_{k=1}^{n} \exp\Big\{< \gamma(\vartheta), H(x_k) > -\psi(\vartheta)\Big\} = \frac{d\,\mathbb{P}_\vartheta^X}{d\mu^{\otimes n}}(x), \quad (L(\vartheta; \omega) = L^X(\vartheta; X(\omega)))$$

d.h., $\sum_{j=1}^{n} H(X_j)$ ist die kanonische Zufallsgröße der Produkt-Exponentialfamilie $(\mathbb{P}_\vartheta^X,\ \vartheta \in \Theta)$. □

7.2 Lévy-Prozesse

Es seien $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(X(t)\ t \geq 0)$ ein Lévy-Prozess über $(\Omega, \mathscr{A}, \mathbb{P})$. Darunter verstehen wir einen stochastisch stetigen reellwertigen zufälligen Prozess mit unabhängigen stationären Zuwächsen, der $\mathbb{P}$-fast sicher bei Null startet und dessen Trajektorien die càdlàg-Eigenschaft haben. Beispiele sind der Wienersche, der Poissonsche, der Gamma-, und jeder zusammengesetzte Poissonsche Prozess.

Für jedes $t > 0$ sei F_t die Verteilungsfunktion von $X(t)$ bezüglich $\mathbb{P}$:

$$F_t(x) = \mathbb{P}(X(t) \leq x), \quad x \in \mathbb{R},\ t > 0.$$

Lemma 7.3.

$$\textit{Ist}\ \ \mathbb{E}_{\mathbb{P}}\Big(\exp\{uX(t)\}\Big) < \infty \ \ \textit{für ein}\ \ u \in \mathbb{R} \ \ \textit{und ein}\ t > 0,$$
$$\textit{so gilt}\ \ \mathbb{E}_{\mathbb{P}}\Big(\exp\{uX(s)\}\Big) < \infty \ \ \textit{für alle}\ \ s > 0.$$

Beweis:
Es sei zunächst $s \in (0, t)$. Damit ist $X(t)$ nach Voraussetzung die Summe der zwei unabhängigen Zufallsgrößen $X(s)$ und $X(t) - X(s)$, die die Verteilungsfunktionen F_s bzw. F_{t-s} besitzen. Folglich gilt

$$F_t = F_s \star F_{t-s}. \tag{a}$$

Diese Gleichung schreiben wir in der Form

$$F_t(z) = \int_{\mathbb{R}} F_s(z-y) F_{t-s}(\,dy) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{(-\infty, z-y]}(x) F_s(\,dx) \right) F_{t-s}(\,dy) =$$
$$\int_{\mathbb{R}^2} \mathbf{1}_{(-\infty, z]}(x+y) F_s(\,dx) \otimes F_{t-s}(\,dy), \quad z \in \mathbb{R}\,.$$

(Wir haben die Folgerung von Tonelli-Hobson zum Satz von Fubini benutzt.)
Daraus ergibt sich mit der üblichen Approximationsmethode für alle nichtnegativen meßbaren Funktionen f auf $\mathbb{R}$ die Gleichung

$$\int_{\mathbb{R}} f(z) F_t(\,dz) = \int_{\mathbb{R}^2} f(x+y) F_s(\,dx) \otimes F_{t-s}(\,dy), \tag{b}$$

wobei beide Seiten gleichzeitig endlich oder beide unendlich sind. Setzt man $f(z) = \exp\{uz\}$, so ist nach Voraussetzung

$$\int_{\mathbb{R}} \exp\{uz\} F_t(\,dz) < \infty$$

und aus (b) und dem Satz von Fubini folgt

$$\int_{\mathbb{R}} \exp\{ux\} F_s(\,dx) < \infty.$$

Ist dagegen $s > t$, so wählt man $m \geq 1$ derart, daß $mt > s$ gilt. Nun ergibt sich mit analogen Mitteln und aus der genannten Folgerung des Satzes von Fubini

$$\left(\int_{\mathbb{R}} e^{ux} F_t(\,dx) \right)^m = \int_{\mathbb{R}} e^{u(x_1+\ldots+x_m)} \prod_{j=1}^{m} F_t(\,dx_j) = \int_{\mathbb{R}} e^{uz} F_{mt}(\,dz) < \infty.$$

Daraus folgt, wie bereits gezeigt,

$$\int_{\mathbb{R}} e^{sx} F_s(\,dx) < \infty.$$

$\square$

Das Lemma erlaubt folgende

Definition 7.4. Das Intervall I werde definiert durch

$$I = \left\{ u \in \mathbb{R} \,:\, \int_{\mathbb{R}} \exp\{uX(t)\}\, d\,\mathbb{P} = \int_{\mathbb{R}} \exp\{ux\} F_t(\,dx) < \infty \right\}.$$

Dabei ist $t > 0$ beliebig gewählt, I ist unabhängig von t. Es sei $u \in I$. Dann ist die Funktion $\psi_t(u)$, definiert durch

$$\psi_t(u) = \ln \int_{\mathbb{R}} \exp\{ux\} F_t(\,dx), \quad t > 0$$

endlich. Außerdem gilt

$$\begin{aligned} \psi_{s+t}(u) = \ln \int_{\mathbb{R}} \exp\{uz\} F_{s+t}(\,dz) = \ln \int_{\mathbb{R}} \exp\left\{u(x+y)\right\} F_s(\,dx) \otimes F_t(\,dy) = \\ \ln \left(\int_{\mathbb{R}} \exp\{ux\} F_s(\,dx) \cdot \int_{\mathbb{R}} \exp\{uy\} F_t(\,dy) \right) = \psi_s(u) + \psi_t(u), \quad s,t > 0. \end{aligned}$$

Wegen der stochastischen Stetigkeit von $X(\cdot)$ ist auch $t \longrightarrow \psi_t(u)$ stetig, und es folgt

$$\psi_t(u) = t\psi_1(u), \quad \text{wir setzen } \ \psi(u) := \psi_1(u).$$

Lemma 7.5. *Für jedes $u \in I$ ist der Prozess $(M_t^{(u)},\ t \geq 0)$, definiert durch*

$$M_t^{(n)} := \exp\bigl\{uX(t) - t\psi(u)\bigr\}, \quad t \geq 0$$

ein Martingal bezüglich der Filtration $\bigl(\mathscr{A}_t^X,\ t \geq 0\bigr)$ mit $\mathscr{A}_t^X := \sigma\bigl(X(s),\ s \leq t\bigr), \quad t \geq 0.$

Beweis:
Es gilt

$$\begin{aligned} \mathbb{E}\left(M_{t+s}^{(u)} \mid \mathscr{A}_s^X\right) &= \mathbb{E}\Bigl(\exp\bigl\{u(X_{t+s} - X_s) + uX_s\bigr\}\big|\mathscr{A}_s^X\Bigr) \cdot e^{-(t+s)\psi(u)} \\ &= \exp\bigl\{uX_s - s\psi(u)\bigr\}\, \mathbb{E}\Bigl(\exp\bigl\{u(X_{t+s} - X_s) - t\psi(u)\bigr\}\big|\mathscr{A}_s^X\Bigr) \\ &= M_s^{(u)} \cdot \mathbb{E}\bigl(M_t^{(u)}\bigr) = M_s^{(u)}. \end{aligned}$$

□

Wir definieren durch

$$\mathbb{P}_t^{(u)}(A) := \int_A M_t^{(u)}\, d\,\mathbb{P}, \quad A \in \mathscr{A}_t^X$$

ein Wahrscheinlichkeitsmaß $\mathbb{P}_t^{(u)}$ auf $\mathscr{A}_t^X$. Die Familie $\bigl(\mathbb{P}_t^{(u)},\ t > 0\bigr)$ ist verträglich in dem Sinne, daß gilt:

$$\text{Ist } A \in \mathscr{A}_s^X \text{ und } s < t, \text{ so ist } A \in \mathscr{A}_t^X \text{ mit } \mathbb{P}_s^{(u)}(A) = \mathbb{P}_t^{(u)}(A).$$

Das ergibt sich aus der Martingaleigenschaft von $\bigl(M_t^{(u)}\bigr)$:

$$\mathbb{P}_s^{(u)}(A) = \int_A M_s^{(u)}\, d\,\mathbb{P} = \int_A M_t^{(u)}\, d\,\mathbb{P} = \mathbb{P}_t^{(u)}(A).$$

Somit ist eine Mengenfunktion $\mathbb{P}^{(u)}$ auf $\bigcup_{t>0} \mathscr{A}_t^X$ definiert:

$$\mathbb{P}^{(u)}(A) = \mathbb{P}_t^{(u)}(A), \quad \text{falls } A \in \mathscr{A}_t^X.$$

Sie ist σ-additiv und somit eindeutig erweiterbar zu einer Wahrscheinlichkeitsverteilung $\mathbb{P}^{(u)}$ auf

$$\mathscr{A}^X := \bigvee_{t>0} \mathscr{A}_t^X = \sigma\bigl(X(t),\ t \geq 0\bigr).$$

Die eindeutige Fortsetzung von $\mathbb{P}^{(u)}$ auf $\mathscr{A}^X$ werde ebenfalls mit $\mathbb{P}^{(u)}$ bezeichnet.

Aussage 7.4. *Die Wahrscheinlichkeitsmaße* $\mathbb{P}^{(u)}$, $u \in I$, *sind lokal absolutstetig, d.h.* $\mathbb{P}_t^{(u)} \ll \mathbb{P}_t := \mathbb{P}\big|_{\mathscr{A}_t^X}$, $\forall t > 0$, *und es gilt nach Definition*

$$\frac{d\,\mathbb{P}_t^{(u)}}{d\,\mathbb{P}_t} = \exp\{uX(t) - \psi(u)t\}, \quad t > 0,\ u \in I.$$

Die Familie $(\mathbb{P}^{(u)},\ u \in I)$ heißt *die zu* $\mathbb{P}$ *und* $(X(t),\ t \geq 0)$ *gehörende Exponentialfamilie (von Wahrscheinlichkeitsmaßen auf* $\mathscr{A}^X$*).*

Aussage 7.5. *Für jedes* $u \in I$ *ist* $(X(t), t \geq 0)$ *bezüglich* $\mathbb{P}^{(u)}$ *ein Lévy-Prozess. Gilt* $u \in \overset{\circ}{I}$, *so haben wir*

$$\mathbb{E}^{(u)}\big(X(t)\big) = \psi'(u) \cdot t, \qquad Var^{(u)}\big(X(t)\big) = \psi''(u) \cdot t.$$

Beweis:
Sind X und Y zwei n-dimensionale zufällige Vektoren über $(\Omega, \mathscr{A}, \mathbb{P})$ und ist A eine $n \times n$-Matrix mit $X = A \cdot Y$, so gilt offenbar

$$\mathbb{E}\big(f(X)\big) = \mathbb{E}\big(f(A \cdot Y)\big), \quad \text{also}$$
$$\int_{\mathbb{R}^n} f(x)\,\mathbb{P}^X(\,dx) = \int_{\mathbb{R}^n} f(Ay)\,\mathbb{P}^Y(\,dy).$$

Ist nun $X = (X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ und $Y = (X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$, so haben wir mit

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

für jede beschränkte meßbare Funktion f

$$\mathbb{E}^{(u)}\big(f(X)\big) = \mathbb{E}^{(u)}\big(f(A \cdot Y)\big),$$

also

$$\int_\Omega \exp\Big\{uX(t_n) - t_n\psi(u)\Big\} f(X)\,d\,\mathbb{P} = \int_\Omega \exp\Big\{uX(t_n) - t_n\psi(u)\Big\} f(A \cdot Y)\,d\,\mathbb{P}$$

und somit wegen

$$X_{t_n} = X_{t_1} + (X_{t_2} - X_{t_1}) + \ldots + (X_{t_n} - X_{t_{n-1}})$$
$$\int_{\mathbb{R}^n} e^{ux_n - t_n\psi(u)} f(x_1, x_2, \ldots, x_n)\,\mathbb{P}^X(\,dx_1, \ldots, \,dx_n)$$
$$= \int_{\mathbb{R}^n} f(y_1, y_1 + y_2, \ldots, y_1 + y_2 + \cdots + y_n) e^{u(y_1+y_2+\cdots+y_n) - \psi(u)t_n} \prod_{k=1}^{n}{}^{\otimes} F_{t_k - t_{k-1}}(\,dy_k)$$
$$= \int_{\mathbb{R}^n} f(y_1, y_1 + y_2, \ldots, y_1 + y_2 + \cdots + y_n) \prod_{k=1}^{n}{}^{\otimes} F^{(u)}_{t_k - t_{k-1}}(\,dy_k)$$

mit

$$F_t^{(u)}(\,dx) = \exp\big\{ux - t\psi(u)\big\}F_t(\,dx).$$

Da f beliebig ist, folgt daraus, daß für jedes $u \in I$ der Prozess $\big(X(t),\ t \geq 0\big)$ bezüglich $\mathbb{P}^{(u)}$ unabhängige und stationäre Zuwächse besitzt. Das bedeutet, die eindimensionalen Verteilungen $\big(F_t^{(u)} \mid u \in I\big)$ bilden selbst wieder Exponentialfamilien. □

Folgerung 7.6. X_t ist eine minimal suffiziente Statistik für $\mathscr{P} = (\mathbb{P}^{(u)},\ u \in I)$ hinsichtlich $\mathscr{A}_t^X$. Außerdem bildet $\dfrac{X_t}{t}$ eine erwartungstreue Schätzung für $\psi'(u)$. Für ihre Streuung gilt

$$\operatorname{Var}\left(\frac{X_t}{t}\right) = \frac{\psi''(u)}{t}.$$

7.3 Vollständige Statistiken

Es seien $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und T eine Statistik über $(\Omega, \mathscr{A})$ mit Werten in einem meßbaren Raum $(E, \mathscr{E})$.

Definition 7.7. Die Statistik T heißt *vollständig (b-vollständig)*, falls jede reellwertige Borelmeßbare Funktion Φ auf $(E, \mathscr{E})$, die $\mathbb{P}_\vartheta^T$-integrierbar ist (bzw. die beschränkt ist), und für die gilt

$$\mathbb{E}_\vartheta\left(\Phi(T)\right) = 0, \quad \text{für alle } \ \vartheta \in \Theta,$$

notwendig Null ist ($\mathbb{P}_\vartheta^T$-f.s., für alle $\vartheta \in \Theta$).

Aussage 7.6. *Es sei U ein erwartungstreuer Schätzer für $\gamma(\vartheta)$ mit $\mathbb{E}_\vartheta(U^2) < \infty$, $\vartheta \in \Theta$. Weiterhin sei T eine vollständige und suffiziente Statistik für ϑ bezüglich $\mathscr{A}$. Dann ist $\mathbb{E}\,(U \mid T)$ ein erwartungstreuer Schätzer für $\gamma(\vartheta)$ mit kleinerer Varianz:*

$$Var\big(\mathbb{E}\,(U \mid T)\big) \leq Var(U). \tag{7.1}$$

$\mathbb{E}\,(U \mid T)$ ist eine Funktion von T und zwar die einzige Funktion $h(T)$, die einen erwartungstreuen Schätzer für $\gamma(\vartheta)$ darstellt. Folglich ist $\mathbb{E}\,(U \mid T)$ der Schätzer mit minimaler Varianz unter allen erwartungstreuen Schätzern für $\gamma(\vartheta)$: ein MVUE, minimum variance unbiased estimator.

Beweis: Die Ungleichung (7.1) ist eine Eigenschaft bedingter Erwartungen. Ist S ein erwartungstreuer Schätzer für $\gamma(\vartheta)$ und ist S eine Funktion von T, d.h. gilt $S = h(T)$, so haben wir

$$\mathbb{E}_\vartheta\big(h(T)\big) = \mathbb{E}_\vartheta\big(\mathbb{E}(U \mid T)\big) \quad \text{für alle } \vartheta \in \Theta.$$

Aus der Vollständigkeit von T folgt $h(T) = \mathbb{E}(U \mid T)$. □

Beispiel 7.4. Ist $\mathscr{P}$ eine Exponentialfamilie mit der kanonischen Zufallsgröße H, so ist H minimal suffizient und vollständig.
Beweis:
Es gilt

$$0 = \mathbb{E}_\vartheta\left(f(H)\right) = \int_{\mathbb{R}} f(x)\exp\big\{< x, u > -\varphi(u)\big\}\mu^H(\,dx) =$$
$$\exp\big\{-\psi(u)\big\}\int_{\mathbb{R}} f(x)\exp\big\{< x, u >\big\}\mu^H(\,dx).$$

Das bedeutet, die Laplace-Transformierten von f_+ und f_- sind gleich auf Θ. Daraus ergibt sich $f_+ = f_- = 0 \quad \mu^H$-fast überall und somit $f(H) = 0 \quad \mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$. (Wir haben hier benutzt, daß $\gamma(\Theta)$ in I ein nichtleeres Inneres besitzt.) □

Beispiel 7.5. Es seien $(\Omega, \mathscr{A}, \mathscr{P})$ ein dominiertes statistisches Modell, T eine Statistik über $(\Omega, \mathscr{A})$. Ist T suffizient und b-vollständig, so ist T minimal suffizient.
Beweis:
Die σ-Algebra $\sigma(T)$, vervollständigt bezüglich $\mathbb{P}$ ($\mathbb{P}$ ist ein privilegiertes dominierendes Maß), enthält die minimal suffiziente σ-Algebra $\mathscr{M}_0$. Umgekehrt:

$$\text{Ist } C \in \sigma(T), \quad \text{so ist} \quad \mathbb{E}_\vartheta\left(I_C - \mathbb{E}\left(\mathbf{1}_C \mid \mathscr{M}_0\right)\right) = 0 \quad \text{für alle } \vartheta \in \Theta.$$

Damit gilt $\mathbf{1}_C = \mathbb{E}\left(\mathbf{1}_C \mid \mathscr{M}_0\right) \quad \mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$ wegen der Vollständigkeit von T. Somit hat $\mathbf{1}_C$ eine $\mathbb{P}$-Version, die $\mathscr{M}_0$-meßbar ist, d.h. $C \in \mathscr{M}_0^{\mathbb{P}}$. □

Aus der Minimalsuffizienz folgt nicht die b-Vollständigkeit (siehe 6. Übung).

7.4 Die Cramer-Rao-Ungleichung und Exponentialfamilien

Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$, $\Theta \subseteq \mathbb{R}^k$ und einem dominierenden σ-finiten Maß μ. Wir können o.B.d.A. annehmen, daß μ ein Wahrscheinlichkeitsmaß ist und bezeichnen es deshalb der Gewohnheit gemäß mit $\mathbb{P}$: $\quad \mu = \mathbb{P}$.
Die Likelihoodfunktion $L(\vartheta\,;\,\omega)$ $(\vartheta \in \Theta, \omega \in \Omega)$ ist definiert durch

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

Wegen $\mathbb{P}_\vartheta\big(L(\vartheta\,;\,\cdot) > 0\big) = 1$ ist für jedes $\vartheta \in \Theta$

$$l(\vartheta\,;\,\cdot) := \ln L(\vartheta\,;\,\cdot)$$

eine $\mathbb{P}_\vartheta$-fast sicher definierte Zufallsgröße über $(\Omega, \mathscr{A})$.

Voraussetzung: Die Funktion $\vartheta \longrightarrow L(\vartheta\,;\,\omega)$ sei differenzierbar für $\mathbb{P}$-fast alle $\omega \in \Omega$ und es gelte

$$\mathbb{E}_\vartheta\left(\|\operatorname{grad}_\vartheta l(\vartheta\,;\,\cdot)\|^2\right) < \infty, \quad \text{für alle } \vartheta \in \Theta. \ (\|\cdot\| \text{ ist die Euklidische Norm in } \mathbb{R}^k)$$

Definition 7.8. Die Matrix $I(\vartheta)$, definiert durch

$$I(\vartheta) := \mathbb{E}_\vartheta\Big(\operatorname{grad}_\vartheta l(\vartheta\,;\,\cdot)\operatorname{grad}_\vartheta^T l(\vartheta\,;\,\cdot)\Big)$$

heißt *Fisher-Informationsmatrix* (des statistischen Modells $(\Omega, \mathscr{A}, \mathscr{P})$).

Aussage 7.7 (Cramer-Rao-Ungleichung). *Es gelte*

a) Für $\mathbb{P}_\vartheta$-fast alle ω sei $\vartheta \longrightarrow L(\vartheta\,;\,\omega)$ stetig differenzierbar bezüglich ϑ,

b) $\mathbb{E}_\vartheta\left(\|\operatorname{grad}_\vartheta l(\vartheta\,;\,\cdot)\|^2\right) < \infty, \quad \vartheta \in \Theta$

c) Für jede Zufallsgröße Y mit $\mathbb{E}_\vartheta\big(Y^2\big) < \infty$ gelte

$$\operatorname{grad}_\vartheta \mathbb{E}_\vartheta\big(L(\vartheta\,;\,\cdot)Y\big) = \mathbb{E}_\vartheta\big((\operatorname{grad}_\vartheta L(\vartheta\,;\,\cdot)Y\big), \quad \vartheta \in \Theta$$

d) $I(\vartheta)$ sei regulär, $\quad \vartheta \in \Theta$.

Dann gilt für jedes $\vartheta \in \Theta$ und für jede Zufallsgröße Y mit $\mathbb{E}_\vartheta\big(|Y|\big) < \infty$ die Ungleichung

$$\mathbb{E}_\vartheta\left((Y - \mathbb{E}_\vartheta(Y))^2\right) \geq \big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T I^{-1}(\vartheta)\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big).$$

Spezialisierungen:
Wenn Y eine erwartungstreue Schätzung für $\gamma(\vartheta)$ ist, so gilt:

$$D^2_\vartheta Y \geq \left(\mathrm{grad}_\vartheta\, \gamma\,(\vartheta)\right)^T I^{-1}(\vartheta)\left(\mathrm{grad}_\vartheta\, \gamma\,(\vartheta)\right),$$

und ist Θ eindimensional, $\Theta \subseteq \mathbb{R}$, so haben wir

$$D^2_\vartheta Y \geq \frac{\left(\dot\gamma(\vartheta)\right)^2}{I(\vartheta)}.$$

Beweis der Cramer-Rao-Ungleichung:

1. Schritt: Aus c) folgt

$$\mathbb{E}\big(\mathrm{grad}_\vartheta\, L(\vartheta\,;\,\cdot)\big) = 0, \quad \text{d.h.} \quad \mathbb{E}_\vartheta\big(\mathrm{grad}_\vartheta \ln L(\vartheta\,;\,\cdot)\big) = 0 \;\; (\text{Wegen } \mathbb{E}\big(L(\vartheta\,;\,\cdot)\big) = 1)$$

Es gilt folglich:

$$\mathrm{grad}_\vartheta\, \mathbb{E}_\vartheta(Y) = \mathrm{grad}_\vartheta\, \mathbb{E}\big(L(\vartheta\,;\,\cdot)Y\big) = \mathbb{E}\big((\mathrm{grad}_\vartheta\, L(\vartheta\,;\,\cdot))\,Y\big) = \\ \mathbb{E}_\vartheta\Big((\mathrm{grad}_\vartheta \ln L(\vartheta\,;\,\cdot))\,Y\Big) = \mathbb{E}_\vartheta\Big((\mathrm{grad}_\vartheta \ln L(\vartheta\,;\,\cdot))\,(Y - \mathbb{E}_\vartheta(Y))\Big) \quad (7.2)$$

2. Schritt: Für alle $u \in \mathbb{R}^k$ gilt mit $< u, v >= \sum_{l=1}^k u_k v_k = u^T v$

$$< u, \mathrm{grad}_\vartheta\, \mathbb{E}_\vartheta(Y) > \overset{(7.2)}{=} \mathbb{E}_\vartheta\Big(< u, \big(\mathrm{grad}_\vartheta \ln L(\vartheta\,;\,\cdot)\big) > \cdot \big(Y - \mathbb{E}_\vartheta(Y)\big)\Big),$$

und somit wegen der Cauchy-Schwarzschen Ungleichung

$$< u, \mathrm{grad}_\vartheta\, \mathbb{E}_\vartheta(Y) >^2 \leq \mathbb{E}_\vartheta\Big(\big(< u, \big(\mathrm{grad}_\vartheta \ln L(\vartheta\,;\,\cdot)\big) >\big)^2\Big) \cdot \mathbb{E}_\vartheta\Big(\big(Y - \mathbb{E}_\vartheta(Y)\big)^2\Big).$$

Das heißt,

$$\mathbb{E}_\vartheta\Big(\big(Y - \mathbb{E}_\vartheta(Y)\big)^2\Big) \geq \frac{< u, \mathrm{grad}_\vartheta\, \mathbb{E}_\vartheta(Y) >^2}{\mathbb{E}_\vartheta\big(< u, \mathrm{grad}_\vartheta \ln L(\vartheta\,;\,\cdot) >^2\big)}, \quad \forall u \in \mathbb{R}^k,\ \forall \vartheta \in \Theta. \quad (7.3)$$

Wir fixieren $\vartheta \in \Theta$ und wählen $u \in \mathbb{R}^k$ so, daß

$$< u, \mathrm{grad}_\vartheta\, \mathbb{E}_\vartheta(Y) >= 1 \quad (7.4)$$

erfüllt ist. Die rechte Seite von (7.3) lautet dann

$$\frac{1}{u^T I(\vartheta) u}$$

3. Schritt: Bestimmung des Maximums der rechten Seite von (7.3) unter der Nebenbedingung (7.4).

$$u^T I(\vartheta)u - \lambda\big(< u, \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) > -1\big) \quad \text{soll minimiert werden.}$$

Wir erhalten als notwendige Bedingung

$$\operatorname{grad}_\vartheta : \quad u^T I(\vartheta) - \lambda\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T = 0 \quad \text{und}$$
$$\frac{d}{d\lambda} : \quad < u, \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) >= 1.$$

Auflösung nach u und λ: Erste Gleichung nach u^T auflösen und in die zweite einsetzen liefert:

$$1 = u^T \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) = \lambda\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T I^{-1}(\vartheta)\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big).$$

Also:

$$\lambda = \frac{1}{\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T I^{-1}(\vartheta)\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)},$$

$$u^T = \lambda \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) I^{-1}(\vartheta).$$

Damit folgt

$$\frac{1}{u^T I(\vartheta)u} = \frac{1}{\lambda^2\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T I^{-1}(\vartheta)\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)}.$$

□

Bemerkung 7.9. Die Fishersche Informationsmatrix ist definiert durch

$$I(\vartheta) := \mathbb{E}_\vartheta\Big(\operatorname{grad}_\vartheta \ln L(\vartheta\,;\,\cdot)\operatorname{grad}_\vartheta^T \ln L(\vartheta\,;\,\cdot)\Big).$$

Unter der Bedingung, daß $\vartheta \to L(\vartheta\,;\,\cdot)$ zweimal differenzierbar ist und die Vertauschungsgleichung

$$\mathbb{E}\Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j} L(\vartheta\,;\,\cdot)\Big) = \frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\mathbb{E}\Big(L(\vartheta\,;\,\cdot)\Big) = 0$$

gilt ($\mathbb{E}$ ist der Erwartungswert bezüglich des dominierenden privilegierten Maßes, $L(\vartheta;\cdot) = \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$), haben wir wegen

$$\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\ln L(\vartheta\,;\,\cdot) = \Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}L(\vartheta\,;\,\cdot)\Big)\frac{1}{L(\vartheta\,;\,\cdot)} - \Big(\frac{\partial}{\partial\vartheta_i}L(\vartheta\,;\,\cdot)\Big)\Big(\frac{\partial}{\partial\vartheta_j}L(\vartheta\,;\,\cdot)\Big)\frac{1}{\big(L(\vartheta\,;\,\cdot)\big)^2}$$

und

$$\mathbb{E}_\vartheta\Big(\frac{1}{L(\vartheta\,;\,\cdot)}\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}L(\vartheta\,;\,\cdot)\Big) = \mathbb{E}\Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}L(\vartheta\,;\,\cdot)\Big) = 0$$

die Beziehung

$$I(\vartheta) = -\,\mathbb{E}_\vartheta\Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\ln L(\vartheta\,;\,\cdot)\Big).$$

Beispiel 7.6. $\mathscr{P}$ bilde eine Exponentialfamilie. Dann ist

$$\begin{aligned}\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} &= \exp\Big\{<\vartheta, H> -\psi(\vartheta)\Big\}, \quad \vartheta\in\Theta,\ \Theta \text{ sei offen,}\\ I(\vartheta) &= \mathrm{grad}_\vartheta\big(\mathrm{grad}_\vartheta\,\psi(\vartheta)\big) = \Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\psi(\vartheta)\Big) \quad \text{sei regulär.}\end{aligned}$$

Weiterhin sei

$$\begin{aligned} l(\vartheta) :&= \ln L(\vartheta\,;\,\cdot) =<\vartheta, H> -\psi(\vartheta) \quad \text{und es gilt}\\ \mathrm{grad}_\vartheta\, l(\vartheta) &= H - \mathrm{grad}_\vartheta\,\psi(\vartheta).\end{aligned}$$

Es folgt

$$\mathbb{E}_\vartheta(H) = \mathrm{grad}_\vartheta\,\psi(\vartheta), \quad \text{da } \mathbb{E}_\vartheta\big(\mathrm{grad}_\vartheta\, l(\vartheta)\big) = 0 \text{ ist.}$$

Als Zufallsgröße Y nehmen wir $< u, H >$ für ein $u \in \mathbb{R}^k$. Es ergibt sich

$$\begin{aligned}\mathbb{E}_\vartheta(Y) &=< u, \mathrm{grad}_\vartheta\,\psi(u) > \quad \text{und}\\ \mathrm{grad}_\vartheta\,\mathbb{E}_\vartheta(Y) &=< u, I(\vartheta) > .\end{aligned}$$

Folglich gilt:

$$\begin{aligned} D^2_\vartheta Y &= \mathbb{E}_\vartheta\Big(< H - \mathrm{grad}_\vartheta\,\psi(\vartheta), u >^2\Big)\\ &= \mathbb{E}_\vartheta\Big(u^T\big(H - \mathrm{grad}_\vartheta\,\psi(\vartheta)\big)\big(H - \mathrm{grad}_\vartheta\,\psi(\vartheta)\big)^T u\Big)\\ &= u^T I(\vartheta) u.\end{aligned}$$

Somit ist:

$$\begin{aligned} D^2_\vartheta Y &= u^T I(\vartheta) u = u^T I(\vartheta) I^{-1}(\vartheta) I(\vartheta) u\\ &= \Big(\mathrm{grad}_\vartheta\,\mathbb{E}_\vartheta(Y)\Big)^T I^{-1}(\vartheta)\Big(\mathrm{grad}_\vartheta\,\mathbb{E}_\vartheta(Y)\Big).\end{aligned}$$

Also wird im Fall der Exponentialfamilie die untere Schranke in der Cramer-Rao-Ungleichung erreicht.

In der Ungleichung von Cramer-Rao muß es keine erwartungstreue Schätzung für $\gamma(\vartheta)$ geben, deren Streuung die untere Schranke erreicht.

Beispiel 7.7. Es sei X eine zum Parameter $\lambda > 0$ Poissonverteilte Zufallsgröße. X ist eine suffiziente und vollständige Statistik für λ, da $L(\lambda\,;\,x) = \lambda e^{-\lambda x}$. Es sei $T(X) := \mathbf{1}_{\{0\}}(X)$, dann gilt $\mathbb{E}_\lambda(T) = e^{-\lambda}$. D.h., T ist eine erwartungstreue Schätzung mit minimaler Varianz für $g(\lambda) = e^{-\lambda}$. Es gilt

$$D^2T = \mathbb{E}_\lambda\Big(\mathbf{1}_{\{0\}}(X) - e^{-\lambda}\Big)^2 = e^{-\lambda} - 2e^{-2\lambda} + e^{-2\lambda} = e^{-\lambda}(1-e^{-\lambda}).$$

Andererseits ist

$$I(\lambda) = \mathbb{E}_\lambda\Big(\big(\frac{\partial}{\partial\lambda}\ln L(\lambda\,;\,\cdot)\big)^2\Big) = \frac{1}{\lambda^2},\ \text{außerdem, mit } g'(\lambda) = -e^{-\lambda} \text{ gilt}$$

$$\frac{\big(g'(\lambda)\big)^2}{I(\lambda)} = \lambda^2 e^{-2\lambda} \;<\; e^{-\lambda}(1-e^{-\lambda}), \quad \text{wegen } 1+\lambda^2 < e^\lambda.$$

Beispiel 7.8. Es seien $X = (X_1,\dots,X_n)$ eine klassische mathematische Stichprobe unter $\mathbb{P}_\vartheta$, $\vartheta\in\Theta$, $Q_\vartheta := \mathbb{P}_\vartheta^{X_1}$, und $\frac{dQ_\vartheta}{d\,\mathbb{P}}(x) =: f_\vartheta(x)$. Es gilt:

$$\begin{aligned}
L(\vartheta\,;\,\omega) &= \prod_{k=1}^n \frac{dQ_\vartheta}{d\,\mathbb{P}}(X_k) = \prod_{k=1}^n f_\vartheta(X_k),\\
l(\vartheta\,;\,\omega) &= \sum_{k=1}^n \ln f_\vartheta(X_k),\\
\operatorname{grad}_\vartheta l(\vartheta\,;\,\omega) &= \sum_{k=1}^n \operatorname{grad}_\vartheta \ln f_\vartheta(X_k) = \sum_{k=1}^n \frac{\operatorname{grad}_\vartheta f_\vartheta(X_k)}{f_\vartheta(X_k)},\\
I(\vartheta) &= \mathbb{E}_\vartheta\big(\operatorname{grad}_\vartheta l(\vartheta\,;\,\omega)\operatorname{grad}_\vartheta l(\vartheta\,;\,\omega)^T\big)\\
&= \sum_{k=1}^n\sum_{l=1}^n \mathbb{E}_\vartheta\Big(\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_k)\big)\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_l)\big)^T\Big)
\end{aligned}$$

$\Big[$für $k \neq l$ sind X_k und X_l unabhängig, also gilt

$$\begin{aligned}
\mathbb{E}_\vartheta\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_k)\big) &= \mathbb{E}_\vartheta\Big(\frac{\operatorname{grad}_\vartheta f_\vartheta(X_k)}{f_\vartheta(X_k)}\Big) = \mathbb{E}_{\mathbb{P}}\Big(\operatorname{grad}_\vartheta \frac{dQ_\vartheta}{d\,\mathbb{P}}(X_k)\Big)\\
&= \operatorname{grad}_\vartheta \mathbb{E}\Big(\frac{dQ_\vartheta}{d\,\mathbb{P}}(X_1)\Big) = \operatorname{grad}_\vartheta 1 = 0.\Big]\\
&= \sum_{k=1}^n \underbrace{\mathbb{E}_\vartheta\Big(\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_k)\big)\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_k)\big)^T\Big)}_{=:\, I_1(\vartheta)}\\
&= n\cdot I_1(\vartheta).
\end{aligned}$$