6 Exponentialfamilien

Es sei $(\mathfrak{X}, \mathcal{B})$ Messraum, $\mathcal{M}^1(\mathfrak{X}, \mathcal{B})$ Menge aller Wahrscheinlichkeitsmaße auf \mathcal{B} .

6.1 Definition

Eine Verteilungsklasse $\wp = \{P_{\vartheta} : \vartheta \in \Theta\} \subset \mathcal{M}^1(\mathfrak{X}, \mathcal{B})$ heißt **Exponentialfamilie** : \Leftrightarrow es existiert ein σ -endliches dominierendes Maß μ auf \mathcal{B} , für ein $k \in \mathbb{N}$ existieren $q_1, \ldots, q_k, c : \Theta \to \mathbb{R}$ und messbare Funktionen $T_1, \ldots, T_k : \mathfrak{X} \to \mathbb{R}, h : \mathfrak{X} \to \mathbb{R}_{>0}$ mit

$$f(x,\vartheta) := \frac{dP_\vartheta}{d\mu}(x) = c(\vartheta) \cdot e^{\sum_{j=1}^k q_j(\vartheta) T_j(x)} \cdot h(x) \quad \text{μ-f.\"{\it u}$.}$$

6.2 Bemerkungen

- a) Mit $q(\vartheta) := (q_1(\vartheta), \dots, q_k(\vartheta))^T$ und $T(x) := (T_1(x), \dots, T_k(x))^T$ ist $f(x,\vartheta) = c(\vartheta) e^{q(\vartheta)^T T(x)} h(x)$
- b) c ist Normierungskonstante:

$$c(\vartheta) = \left[\int e^{q(\vartheta)^T T(x)} h(x) \mu(dx) \right]^{-1} > 0$$

c) Der Träger $\{x: f(x,\vartheta) > 0\}$ hängt nicht von ϑ ab, insbesondere gilt

$$\forall N \in \mathcal{B}: P_{\vartheta_1}(N) = 0 \Leftrightarrow P_{\vartheta_2}(N) = 0 \qquad (\vartheta_1, \vartheta_2 \in \Theta)$$

(d.h. es gilt $P_{\vartheta_1} \ll P_{\vartheta_2}, P_{\vartheta_2} \ll P_{\vartheta_1}$).

- d) Im Folgenden gelte immer:
 - (i) Die Funktionen $1, q_1, \ldots, q_k$ sind linear unabhängig
 - (ii) Die Funktionen $1, T_1, \ldots, T_k$ sind linear unabhängig auf dem Komplement jeder μ -Nullmenge

(sogenannte (strikt) k-parametrige Exponentialfamilie).

Dann ist k kleinstmöglich gewählt, und q sowie T sind bis auf nicht ausgeartete affine Transformationen $q\mapsto Aq+a,\,T\mapsto BT+b$ (μ -f.ü.) eindeutig bestimmt.

6.3 Beispiele

a) $P_{\vartheta} := \mathcal{N}(\mu, \sigma^2), \ \vartheta := (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} =: \Theta.$ Die Lebesguedichte ist

$$f(x,\vartheta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)}_{=:c(\vartheta)} \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2\right) \cdot \underbrace{1}_{=:h(x)}$$

Mit $q(\vartheta) := (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}), \ T(x) := (x, x^2)$ folgt, dass hier eine (strikt) zweiparametrige Exponentialfamilie vorliegt.

b) $P_{\vartheta} := \mathcal{N}(\vartheta, \vartheta^2), \ \vartheta \in \mathbb{R}_{>0} =: \Theta.$ Die Lebesguedichte ist

$$f(x,\vartheta) = \frac{1}{\sqrt{2\pi\vartheta^2}} \exp\left(-\frac{(x-\vartheta)^2}{2\vartheta^2}\right)$$
$$= \underbrace{\frac{1}{\sqrt{2\pi\vartheta^2}}}_{=:c(\vartheta)} \exp\left(\frac{1}{\vartheta}x - \frac{1}{2\vartheta^2}x^2\right) \cdot \underbrace{1}_{=:h(x)}$$

Mit $q(\vartheta) := (\frac{1}{\vartheta}, -\frac{1}{2\vartheta^2}), \ T(x) := (x, x^2)$ folgt wieder, dass eine (strikt) zweiparametrige Exponentialfamilie vorliegt (obwohl der Parameterraum Θ eindimensional ist!)

c) $P_{\vartheta} := \text{Bin}(n, \vartheta), \ \vartheta \in (0, 1) =: \Theta.$ Die Zähldichte ist

$$f(x,\vartheta) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} = (1-\vartheta)^n \exp\left(x \log \frac{\vartheta}{1-\vartheta}\right) \binom{n}{x}.$$

Mit $c(\vartheta) := (1-\vartheta)^n$, $q(\vartheta) := \log \frac{\vartheta}{1-\vartheta}$, T(x) := x und $h(x) := \binom{n}{x}$ folgt, dass $\wp := \{ \operatorname{Bin}(n,\vartheta) : \vartheta \in \Theta \}$ eine einparametrige Exponentialfamilie ist.

d) Die Menge aller Gleichverteilungen $\{U(0,\vartheta), \vartheta \in \mathbb{R}_{>0}\}$ ist nach 6.2(c) keine Exponentialfamilie.

6.4 Satz 45

6.4 Satz

Es seien $X_1, \ldots, X_n \stackrel{uiv}{\sim} P_{\vartheta}$, wobei P_{ϑ} Element einer k-parametrigen Exponentialfamilie $\{P_{\vartheta}: \vartheta \in \Theta\}$ ist. Dann gehöhrt auch die Verteilung von $X := (X_1, \ldots, X_n)$ zu einer k-parametrigen Exponentialfamilie mit

$$q(\vartheta)$$
 und $T_{(n)}(x) := \sum_{j=1}^{n} T(x_j)$.

Beweis:

Sei $\mu^n := \mu \otimes \cdots \otimes \mu$ das n-fache Produktmaß auf $\mathcal{B}^n := \mathcal{B} \otimes \cdots \otimes \mathcal{B}$ und

$$P_{\vartheta}^n := P_{\vartheta} \otimes \cdots \otimes P_{\vartheta}$$

die Verteilung von X unter P_{ϑ} . Wir erhalten mit $x := (x_1, \dots, x_n)$:

$$\frac{dP_{\vartheta}^{n}}{d\mu^{n}}(x) = \prod_{j=1}^{n} \frac{dP_{\vartheta}}{d\mu}(x_{j}) \quad \mu\text{-f.\"{u}}.$$

$$= \prod_{j=1}^{n} \left[c(\vartheta) \exp\left(q^{T}(\vartheta)T(x_{j})\right) h(x_{j}) \right] \quad \mu\text{-f.\"{u}}.$$

$$= c(\vartheta)^{n} \exp\left(q^{T}(\vartheta) \sum_{j=1}^{n} T(x_{j})\right) \prod_{J=1}^{n} h(x_{j}) \quad \mu\text{-f.\"{u}}.$$

Bemerkung:

In der Situation von Satz 6.4 ist der ML-Schätzer $\hat{\vartheta}_n$ für ϑ eine Funktion von $\sum_{j=1}^n T(X_j)$.

In der Darstellung

$$f(x,\vartheta) = c(\vartheta) \exp(q^T(\vartheta)T(x))h(x)$$

hängt $c(\cdot)$ von ϑ nur über $q:=q(\vartheta)\in Q:=q(\Theta)\subset\mathbb{R}^k$ ab, das heißt es gilt

$$c(\vartheta) = C\left(q(\vartheta)\right)$$

für ein geeignetes $C: Q \to \mathbb{R}$.

q heißt natürlicher Parameter. Somit lässt sich f ausdrücken als

$$f(x,q) = \frac{dP_q}{d\mu}(x) = C(q)e^{q^T \cdot T(x)}h(x)$$

Die Menge

$$Q_* := \{ q \in \mathbb{R}^k : 0 < \int e^{q^T T(x)} h(x) \mu(dx) < \infty \}$$

heißt natürlicher Parameterraum der Exponentialfamilie. Es gilt

$$Q = q(\Theta) \subset Q_*$$
.

6.5 Satz

 Q_* ist konvex und enthält ein nicht-ausgeartetes k-dimensionales Intervall.

Beweis:

Für $q, r \in Q_*$ und $\lambda \in [0, 1]$ gilt

$$0 < \int e^{(\lambda q^T + (1-\lambda)r^T)T} h d\mu$$

$$= \int \left(e^{q^T T}\right)^{\lambda} \left(e^{r^T T}\right)^{1-\lambda} h d\mu$$

$$\leq \int \max\left(e^{q^T T}, e^{r^T T}\right) h d\mu$$

$$= \int \left(e^{q^T T} + e^{r^T T}\right) h d\mu < \infty$$

Die zweite Aussage folgt dann aus der linearen Unabhängigkeit von $1, q_1, \dots, q_k$.

Bemerkung:

Im Folgenden setzen wir $\vartheta := q$, betrachten also Exponentialfamilien

$$f(x,\vartheta) = \frac{dP_{\vartheta}}{d\mu}(x) = C(\vartheta)e^{\vartheta^T T(x)}h(x) \tag{1}$$

mit $\vartheta \in \Theta := \left\{ \vartheta \in \mathbb{R}^k : \ 0 < \int e^{\vartheta^T T(x)} h(x) \mu(dx) < \infty \right\}.$ Weiter sei

$$b(\vartheta) := -\log C(\vartheta).$$

6.6 Lemma

Es sei $\varphi: \mathfrak{X} \to \mathbb{R}$ eine messbare Abbildung mit

$$E_{\vartheta}|\varphi| = \int |\varphi(x)|f(x,\vartheta)\mu(dx) < \infty$$

6.7 Satz 47

Sei

$$A_{\varphi}(\vartheta) := \int \varphi(X) e^{\vartheta^T T(x)} h(x) \mu(dx), \quad \vartheta \in \Theta^0$$
 (2)

Dann ist $A_{\varphi}:\Theta^0\to\mathbb{R}$ beliebig oft differenzierbar und die Differentiation in (2) kann unter dem Integralzeichen vorgenommen werden beziehungsweise Integration und Differentiation können vertauscht werden.

Beweis:

Witting, 1985, S. 151f.

6.7 Satz

- a) Die Funktion $b(\vartheta)$, $\vartheta \in \Theta^0$, ist beliebig oft differenzierbar.
- b) Besitzt X die Dichte $f(x, \vartheta)$ aus (1), so gilt:

$$E_{\vartheta}T(X) = \frac{d}{d\vartheta}b(\vartheta)$$

$$\operatorname{Var}_{\vartheta} T(X) = \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

Beweis:

a)
$$\varphi \equiv 1$$
 in $6.6 \Rightarrow A_{\varphi}(\vartheta) = C(\vartheta)^{-1} = e^{b(\vartheta)}$
 $6.6 \Rightarrow \text{Behauptung}$

b)

$$E_{\vartheta}T(X) = e^{-b(\vartheta)} \int T(x)e^{\vartheta^{T}T(x)}h(x)\mu(dx)$$

$$= e^{-b(\vartheta)} \int \frac{d}{d\vartheta}e^{\vartheta^{T}T(x)}h(x)\mu(dx)$$

$$\stackrel{6.6}{=} e^{-b(\vartheta)} \frac{d}{d\vartheta} \underbrace{\int e^{\vartheta^{T}T(x)}h(x)\mu(dx)}_{=e^{b(\vartheta)}}$$

$$= \frac{d}{d\vartheta}b(\vartheta)$$

$$E_{\vartheta}[T(X) \cdot T(X)^{T}] = e^{-b(\vartheta)} \int T(x) e^{\vartheta^{T}T(x)} T(x)^{T} h(x) \mu(dx)$$

$$= e^{-b(\vartheta)} \int \frac{d^{2}}{d\vartheta d\vartheta^{T}} e^{\vartheta^{T}T(x)} h(x) \mu(dx)$$

$$= e^{-b(\vartheta)} \frac{d^{2}}{d\vartheta d\vartheta^{T}} e^{b(\vartheta)}$$

$$= \frac{d^{2}}{d\vartheta d\vartheta^{T}} b(\vartheta) + \underbrace{\left(\frac{d}{d\vartheta} b(\vartheta)\right) \left(\frac{d}{d\vartheta} b(\vartheta)\right)^{T}}_{=E_{\vartheta}T(X) \cdot (E_{\vartheta}T(X))^{T}}$$

$$\Rightarrow \operatorname{Var}_{\vartheta} T(X) = \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

6.8 CR-Effizienz in Exponentialfamilien

Seien $X_1, \ldots, X_n \stackrel{uiv}{\sim} f_1(\xi, \vartheta) = e^{-b(\vartheta)} e^{\vartheta^T T(\xi)} h(\xi)$ wie in (1). $\Rightarrow X = (X_1, \ldots, X_n)$ besitzt die Dichte

$$f(x, \vartheta) = e^{-nb(\vartheta)} \cdot \exp(\vartheta^T \sum_{i=1}^n T(x_i)) \prod_{j=1}^n h(x_j)$$

Sei
$$S(X) = \frac{1}{n} \sum_{j=1}^{n} T(X_j)$$
.

$$\Rightarrow E_{\vartheta}S(X) = E_{\vartheta}T(X_1) \stackrel{6.7}{=} \frac{d}{d\vartheta}b(\vartheta), \ \vartheta \in \Theta$$

 \Rightarrow S erwartungstreu für $\frac{d}{d\vartheta}b(\vartheta)$. Behauptung: S(X) ist CR-effizient.

Beweis:

$$\operatorname{Var}_{\vartheta} S(X) = \frac{1}{n} \operatorname{Var}_{\vartheta} T(X_1) \stackrel{6.7}{=} \frac{1}{n} \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

CR-Ungleichung:

$$\operatorname{Var}_{\vartheta} S(X) \ge C_n(\vartheta)^T I_n(\vartheta)^{-1} C_n(\vartheta)$$

wobei

$$C_n(\vartheta) = \frac{d}{d\vartheta} E_{\vartheta}[S(X)^T] = \frac{d}{d\vartheta} \left[\frac{d}{d\vartheta} b(\vartheta) \right]^T = \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

$$I_n(\vartheta) = n \cdot I_1(\vartheta) = n \cdot E_{\vartheta} \left[\frac{d}{d\vartheta} \log f_1(X_1, \vartheta) \cdot \frac{d}{d\vartheta} \log f_1(X_1, \vartheta)^T \right]$$

$$\log f_1(X_1, \vartheta) = -b(\vartheta) + \vartheta^T T(X_1) + \log h(X_1)$$

$$\frac{d}{d\vartheta} \log f_1(X_1, \vartheta) = -\frac{d}{d\vartheta} b(\vartheta) + T(X_1) = T(X_1) - E_{\vartheta} T(X_1)$$

6.9 Beispiel 49

$$\Rightarrow I_n(\vartheta) = n \cdot \operatorname{Var}_{\vartheta} T(X_1) = n \cdot \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$
$$\Rightarrow C_n(\vartheta)^T I_n(\vartheta)^{-1} C_n(\vartheta) = \frac{1}{n} \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

6.9 Beispiel

$$f_1(\xi, \vartheta) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\mu^2}{2\sigma^2})}_{=C(\vartheta)} \exp(\frac{\mu}{\sigma^2} \cdot \xi - \frac{1}{2\sigma^2} \xi^2)$$

$$\vartheta = (\vartheta_1, \vartheta_2) := (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$$

$$b(\vartheta) = -\log C(\vartheta) = \frac{\mu^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2) = -\frac{1}{4}\frac{\vartheta_1^2}{\vartheta_2} + \frac{1}{2}\log(\frac{-\pi}{\vartheta_2})$$

$$\frac{d}{d\vartheta}b(\vartheta) = (-\frac{1}{2}\frac{\vartheta_1}{\vartheta_2}, \frac{1}{4}\frac{\vartheta_1^2}{\vartheta_2^2} - \frac{1}{2\vartheta_2})^T = (\mu, \sigma^2 + \mu^2)^T$$

Fazit:

$$S(X) = (\frac{1}{n} \sum_{j=1}^{n} X_j, \frac{1}{n} \sum_{j=1}^{n} X_j^2)$$

ist erwartungstreu und CR-effizient für $(E_{\vartheta}X_1, E_{\vartheta}X_1^2)$.

$$\frac{\text{Frage:}}{\text{Ist } S_n^2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \text{ CR-effizient für } \sigma^2?$$