Maximum-likelihood ligningene i en k-parameter eksponensiell familie

Håvard Rue & Bård Skaflestad

17. januar 1999

Dette notatet diskuterer ML-ligningene for k-parameter eksponensielle familier, og er utfyllende til lærebokas bidrag i Appendix side 343-344.

Teorem

Anta

$$x_i \overset{iid}{\sim} f(x \mid \boldsymbol{\theta}), \qquad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_k), \qquad \boldsymbol{x} = (x_1, \ldots, x_n)$$

hvor $f(x \mid \boldsymbol{\theta})$ er en k-parameter eksponensiell familie,

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left\{ \sum_{i=1}^{k} \omega_i(\boldsymbol{\theta}) t_i(x) \right\},$$

og hvor

$$T(\boldsymbol{x}) = \left(\sum_{j=1}^{n} t_1(x_j), \ldots, \sum_{j=1}^{n} t_k(x_j) \right)$$

er minimal suffisient observator[1] for $\boldsymbol{\theta}$. Da er ML-ligningene for $\boldsymbol{\theta}$ gitt ved

$$\frac{1}{n} \sum_{j=1}^{n} t_i(x_j) = E\left(t_i(x)\right) \big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \qquad i = 1, \ldots, k.$$

(Her betyr $E\left(t_1(x)\right) \big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, forventet verdi av $t_1(x)$ som en funksjon av $\boldsymbol{\theta}$, evaluert i $\hat{\boldsymbol{\theta}}$.)

Eksempel

Som kjent er normalfordelingen en 2-parameter eksponensiell familie, og når

$$x_1, \ldots, x_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

[1] Dette krever at variasjonsområdet for $(\omega_1(\boldsymbol{\theta}), \ldots, \omega_k(\boldsymbol{\theta}))$ ikke er en degenerert mengde i $\mathbb{R}^k$.

kjenner vi de (minimale) suffisiente observatorene

$$t_1(\boldsymbol{x}) = \sum_{j=1}^{n} x_j, \qquad \text{og} \qquad t_2(\boldsymbol{x}) = \sum_{j=1}^{n} x_j^2.$$

ML-ligningene blir dermed

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} x_j &= E\left(x\right)|_{(\mu,\sigma^2)=(\hat{\mu},\hat{\sigma}^2)} \\ &= \hat{\mu} \end{aligned}$$

og

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} x_j^2 &= E\left(x^2\right)|_{(\mu,\sigma^2)=(\hat{\mu},\hat{\sigma}^2)} \\ &= \hat{\mu} + \hat{\sigma}^2. \end{aligned}$$

Når vi løser disse ligningene, får vi de kjente ML-estimatene

$$\hat{\mu} = \bar{x}, \qquad \text{og} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2.$$

Bevis for ML-ligningene

Beviset går som følger. Først trenger vi likelihooden

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{x}) &= \prod_{j=1}^{n} f(x_j \mid \boldsymbol{\theta}) \\ &= \left[\prod_{j=1}^{n} h(x_j)\right] c(\boldsymbol{\theta})^n \exp\left\{\sum_{i=1}^{k}\left[\omega_i(\boldsymbol{\theta})\sum_{j=1}^{n} t_i(x_j)\right]\right\} \\ &= h(\boldsymbol{x})c(\boldsymbol{\theta})^n \exp\left\{\sum_{i=1}^{k}\omega_i(\boldsymbol{\theta})t_i(\boldsymbol{x})\right\} \end{aligned}$$

der vi har innført notasjonen $t_i(\boldsymbol{x}) = \sum_{j=1}^{n} t_i(x_j)$. Det er nå gunstig å reparametrisere modellen ved å innføre parametrene $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$, der

$$\eta_1 = \omega_1(\boldsymbol{\theta}), \ \ldots, \eta_k = \omega_k(\boldsymbol{\theta})$$

dvs.

$$\mathcal{L}(\boldsymbol{\eta} \mid \boldsymbol{x}) = h(\boldsymbol{x})\tilde{c}(\boldsymbol{\eta})^n \exp\left\{\sum_{i=1}^{k}\eta_i t_i(\boldsymbol{x})\right\}$$

der $\tilde{c}(\boldsymbol{\eta})$ er ny normaliseringskonstant med hensyn på $\boldsymbol{\eta}$. ML-ligningene finnes da ved

$$\frac{\partial}{\partial \eta_i}\ln\left(\mathcal{L}(\boldsymbol{\eta} \mid \boldsymbol{x})\right) = 0, \qquad i = 1, \ldots, k.$$

Ved å bruke at

$$\ln\left(\mathcal{L}(\boldsymbol{\eta} \mid \boldsymbol{x})\right) = \ln\, h(\boldsymbol{x}) + n \ln\, \tilde{c}(\boldsymbol{\eta}) + \sum_{i=1}^{k} \eta_i\, t_i(\boldsymbol{x})$$

får vi Maximum Likelihood ligningene

$$t_i(\boldsymbol{x}) = -n \frac{\partial}{\partial \eta_i} \ln\, \tilde{c}(\boldsymbol{\eta}), \qquad i = 1, \dots, k$$

eller

$$\frac{1}{n} \sum_{j=1}^{n} t_i(x_j) = -\frac{\partial}{\partial \eta_i} \ln\, \tilde{c}(\boldsymbol{\eta}), \qquad i = 1, \dots, k.$$

Dersom vi kan vise at

$$-n \frac{\partial}{\partial \eta_i} \ln\, \tilde{c}(\boldsymbol{\eta}) = E\, t_i(\boldsymbol{x}) \tag{1}$$

vil derfor teoremet være bevist for parametriseringen $\boldsymbol{\eta}$. For enkelhets skyld viser vi dette kun for $i = 1$. Den momentgenerende funksjonen (mgf) til $g(x)$ er generelt gitt ved

$$M_{g(x)}(s) = E\left(e^{sg(x)}\right)$$

og vi har sammenhengen $E(g(x)) = M'_{g(x)}(0)$. Vi vil derfor finne mgf til $t_1(\boldsymbol{x})$. Fra definisjon av mgf fås

$$\begin{aligned} M_{t_1(\boldsymbol{x})}(s) &= E\, \exp\left(st_1(\boldsymbol{x})\right) \\ &= \int_{\boldsymbol{x}} \exp\left(st_1(\boldsymbol{x})\right) h(\boldsymbol{x})\tilde{c}(\boldsymbol{\eta})^n \exp\left\{\sum_{i=1}^{k} \eta_i t_i(\boldsymbol{x})\right\} d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}} h(\boldsymbol{x})\tilde{c}(\boldsymbol{\eta})^n \exp\left\{(\eta_1 + s)t_1(\boldsymbol{x}) + \sum_{i=2}^{k} \eta_i t_i(\boldsymbol{x})\right\} d\boldsymbol{x} \end{aligned}$$

For å forenkle notasjonen innfører vi $\boldsymbol{\eta}^* = (\eta_1 + s, \eta_2, \dots, \eta_k)$, slik at

$$\begin{aligned} M_{t_1(\boldsymbol{x})}(s) &= \int_{\boldsymbol{x}} h(\boldsymbol{x})\tilde{c}(\boldsymbol{\eta})^n \exp\left\{\sum_{i=1}^{n} \eta_i^* t_i(\boldsymbol{x})\right\} d\boldsymbol{x} \\ &= \frac{\tilde{c}(\boldsymbol{\eta})^n}{\tilde{c}(\boldsymbol{\eta}^*)^n} \underbrace{\int_{\boldsymbol{x}} h(\boldsymbol{x})\tilde{c}(\boldsymbol{\eta}^*)^n \exp\left\{\sum_{i=1}^{k} \eta_i^* t_i(\boldsymbol{x})\right\} d\boldsymbol{x}}_{=1,\ \text{siden } \int_{\boldsymbol{x}} f(\boldsymbol{x} \mid \boldsymbol{\eta}) d\boldsymbol{x} = 1.} \\ &= \left[\frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}(\boldsymbol{\eta}^*)}\right]^n \end{aligned}$$

Dette betyr at

$$
\begin{aligned}
Et_1(\boldsymbol{x}) &= \frac{d}{ds_1} M_{t_1(\boldsymbol{x})}(s) \,|_{s=0} \\
&= \frac{d}{ds_1} \left[\frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}(\boldsymbol{\eta}^*)} \right]^n |_{s=0} \\
&= n \left[\frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}(\boldsymbol{\eta}^*)} \right]^{n-1} \frac{d}{ds_1} \frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}(\boldsymbol{\eta}^*)} |_{s=0} \\
&= n \left[\frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}(\boldsymbol{\eta}^*)} \right]^{n-1} \frac{d}{ds_1} \frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}\left((\eta_1 + s, \eta_2, \ldots, \eta_k)\right)} |_{s=0} \\
&= n \left[\frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}(\boldsymbol{\eta}^*)} \right]^{n-1} \frac{\partial}{\partial \eta_1} \frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}\left((\eta_1 + s, \eta_2, \ldots, \eta_k)\right)} |_{s=0} \\
&= n \left[\frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}(\boldsymbol{\eta}^*)} \right]^{n-1} (-1) \frac{\tilde{c}(\boldsymbol{\eta})}{\tilde{c}(\boldsymbol{\eta}^*)^2} \frac{\partial}{\partial \eta_1} \tilde{c}(\eta_1, \ldots, \eta_k) \,|_{s=0} \\
&= n \cdot 1 \cdot (-1) \frac{1}{\tilde{c}(\boldsymbol{\eta})} \frac{\partial}{\partial \eta_1} \tilde{c}(\boldsymbol{\eta}), \qquad \text{(Bruker at } \boldsymbol{\eta} = \boldsymbol{\eta}^* \text{ når } s = 0.) \\
&= -n \frac{\partial}{\partial \eta_1} \ln \tilde{c}(\boldsymbol{\eta})
\end{aligned}
$$

Dvs, vi har vist ligning (1), og dermed også ML-ligningene med hensyn på $\boldsymbol{\eta}$-parametriseringen. Vi er jo interessert i ML-ligningene for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Da bruker vi egenskapen at ML-estimatoren er invariant for reparametriseringer. Siden

$$\eta_1 = \omega_1(\boldsymbol{\theta}), \; \ldots \;, \eta_k = \omega_k(\boldsymbol{\theta})$$

vil

$$\theta_1 = \xi_1(\boldsymbol{\eta}), \; \ldots \;, \theta_k = \xi_k(\boldsymbol{\eta})$$

der $\xi_j(\cdot)$ er omvendt funksjon av $\omega_j(\cdot)$. Dermed er $\hat{\boldsymbol{\theta}} = (\xi_1(\hat{\boldsymbol{\eta}}), \ldots, \xi_k(\hat{\boldsymbol{\eta}}))$ ML-estimatene for $\boldsymbol{\theta}$, når $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \ldots, \hat{\eta}_k)$ er ML-estimatene for $\boldsymbol{\eta}$. Dette betyr at vi like godt kan løse ligningssystemet

$$\frac{1}{n} \sum_{j=1}^{n} t_i(x_j) = E\left(t_i(x)\right) |_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \qquad i = 1, \ldots, k$$

der forventningene taes med hensyn på $\boldsymbol{\theta}$-parametriseringen for å finne ML-ligningene til $\boldsymbol{\theta}$. ■