

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

$$x_i \stackrel{iid}{\sim} f(x | \theta), \quad \theta = (\theta_1, \dots, \theta_k), \quad \mathbf{x} = (x_1, \dots, x_n)$$

hvor $f(x | \theta)$ er en k -parameter eksponensiell familie,

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

og hvor

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

er minimal suffisient observator¹ for θ . Da er ML-ligningene for θ gitt ved

$$\frac{1}{n} \sum_{j=1}^n t_i(x_j) = E(t_i(x)) |_{\theta=\hat{\theta}} \quad i = 1, \dots, k.$$

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

Eksempel

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

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

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

kjenner vi de (minimale) suffisiente observatorene

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

ML-ligningene blir dermed

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

og

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

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

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

Bevis for ML-ligningene

Beviset går som følger. Først trenger vi likelihooden

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

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

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

dvs.

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

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

$$\frac{\partial}{\partial \eta_i} \ln(\mathcal{L}(\boldsymbol{\eta} \mid \mathbf{x})) = 0, \quad i = 1, \dots, k.$$

Ved å bruke at

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

får vi Maximum Likelihood ligningene

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

eller

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

Dersom vi kan vise at

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

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

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

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

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

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

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

Dette betyr at

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

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 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Da bruker vi egenskapen at ML-estimatoren er invariant for reparametriseringer. Siden

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

vil

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

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

$$\frac{1}{n} \sum_{j=1}^n t_i(x_j) = E(t_i(x)) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad i = 1, \dots, k$$

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