Topics in Statistical Sciences

Mikkel Findinge

Indhold

Opgave 1

Til forelæsningerne så vi, at $\ell_{\text{obs}}\left(\boldsymbol{\theta}^{(t+1)};\mathbf{y}_{\text{obs}}\right) \geq \ell_{\text{obs}}\left(\boldsymbol{\theta}^{(t)};\mathbf{y}_{\text{obs}}\right)$. Argumentet brugte en eksplicit brug af Jensens ulighed. Ved at bruge Kullbach-Leibler divergens, KL, mellem to sandsynlighedsfordelinger, f og g, er en anden måde at vise, at følgen af log-likelihoods er ikke-aftagende;

$$KL(f:g) = \int f(x)\log\frac{f(x)}{g(x)}\,dx \geq 0.$$

KL-divergensen definerer et asymmetrisk distancemål mellem sandsynlighedsfordelinger, hvor $KL(f:g)=0$, hvis og kun hvis $f=g$ (næsten overalt).

Definér $q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}\right)$ som middelværdien over log-likelihood-ratioen givet $\boldsymbol{\theta}^{(t)}$ og $\mathbf{y}_{\text{obs}}$:

$$q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}\right) = \mathrm{E}\left[\log\frac{f(\mathbf{y};\boldsymbol{\theta})}{f\left(\mathbf{y};\boldsymbol{\theta}^{(t)}\right)}\bigg|\mathbf{y}_{\text{obs}};\boldsymbol{\theta}^{(t)}\right].$$

Derfor er $q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}\right)$ forholdet mellem log-likelihoods, hvor $Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}\right)$ var log-likelihood.

Opgave 1 *Vis, at $q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)})$ kan skrives som*

$$q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) = \ell_{obs}(\boldsymbol{\theta};\mathbf{y}_{obs}) - \ell_{obs}(\boldsymbol{\theta}^{(t)};\mathbf{y}_{obs}) - KL\{f_{mis}(\boldsymbol{\theta}^{(t)}) : f_{mis}(\boldsymbol{\theta})\}, \tag{2.1}$$

hvor $f_{mis}(\boldsymbol{\theta}) = f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs};\boldsymbol{\theta})$.

Bevis
Fra definitionen af $q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)})$ og definitonen af middelværdi, har vi

$$\begin{aligned}
q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) &= \mathrm{E}\left[\log\frac{f(\mathbf{y};\boldsymbol{\theta})}{f\left(\mathbf{y};\boldsymbol{\theta}^{(t)}\right)}\bigg|\mathbf{y}_{\text{obs}};\boldsymbol{\theta}^{(t)}\right] \\
&= \int \log\frac{f(\mathbf{y};\boldsymbol{\theta})}{f\left(\mathbf{y};\boldsymbol{\theta}^{(t)}\right)} f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)\,d\mathbf{y}_{\text{mis}} \\
&= \int \log\frac{f_{\text{obs}}(\mathbf{y}_{\text{obs}};\boldsymbol{\theta})f_{\text{mis}}(\boldsymbol{\theta})}{f_{\text{obs}}\left(\mathbf{y}_{\text{obs}};\boldsymbol{\theta}^{(t)}\right)f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)} f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)\,d\mathbf{y}_{\text{mis}} \\
&= \int \log f_{\text{obs}}\left(\mathbf{y}_{\text{obs}};\boldsymbol{\theta}\right) f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)\,d\mathbf{y}_{\text{mis}} \\
&\quad - \int \log f_{\text{obs}}\left(\mathbf{y}_{\text{obs}};\boldsymbol{\theta}^{(t)}\right) f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)\,d\mathbf{y}_{\text{mis}} \\
&\quad - \int \log\frac{f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)}{f_{\text{mis}}(\boldsymbol{\theta})} f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)\,d\mathbf{y}_{\text{mis}}.
\end{aligned}$$

Vi har efter tredje lighed lavet tæthederne om til betingede tætheder. Efter fjerde lighed har vi brugt logaritme-regneregler samt additivitet af integraler. Bemærk, at det sidste integral har fået skiftet fortegn, og dermed kan vi bytte rundt på tæller og nævner i logaritmen! - Dermed kan sidste integral ses som værende KL-divergensen. Vi ser desuden, at de to første integraler kun afhænger af $\mathbf{y}_{\text{mis}}$. Tæthederne, som ikke afhænger af $\mathbf{y}_{\text{mis}}$, kan altså sættes uden for disse integraler, hvilket betyder, at vi tager integralet af en tæthedsfunktion over alle værdier for denne tæthed. Af generel sandsynlighedsteori giver dette 1. Altså har vi:

$$q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) = \log f_{\text{obs}}(\boldsymbol{\theta};\mathbf{y}_{\text{obs}}) - \log f_{\text{obs}}(\boldsymbol{\theta}^{(t)};\mathbf{y}_{\text{obs}}) - KL\{f_{\text{mis}}(\boldsymbol{\theta}^{(t)}) : f_{\text{mis}}(\boldsymbol{\theta})\}$$

$\blacksquare$

Opgave 2 *Brug udtrykket i* (2.1) *til at vise, at*

$$\frac{\partial}{\partial \boldsymbol{\theta}} q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}} = \frac{\partial}{\partial \boldsymbol{\theta}} \ell_{obs}(\boldsymbol{\theta}; \mathbf{y}_{obs})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}.$$

Bevis
Differentieres begge sider i (2.1) i forhold til $\boldsymbol{\theta}$, så bemærker vi, at $-\ell_{\text{obs}}(\boldsymbol{\theta}^{(t)}; \mathbf{y}_{\text{obs}})$ ikke indeholder $\boldsymbol{\theta}$, altså bliver denne 0, når der differentieres i forhold til denne variabel. Vi behøver altså blot at tjekke, hvad der sker med KL-divergensen. Under tilstrækkelige regularitetsbetingelser, har vi:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} KL\{f_{\text{mis}}(\boldsymbol{\theta}^{(t)}) : f_{\text{mis}}(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \int \log \frac{f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)}{f_{\text{mis}}\left(\boldsymbol{\theta}\right)} f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right) \, d\mathbf{y}_{\text{mis}} \\
&= \int \frac{\partial}{\partial \boldsymbol{\theta}} \log \frac{f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)}{f_{\text{mis}}\left(\boldsymbol{\theta}\right)} f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right) \, d\mathbf{y}_{\text{mis}} \\
&= \int \frac{f_{\text{mis}}(\boldsymbol{\theta})}{f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)} \left(-\frac{f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)}{f_{\text{mis}}^2(\boldsymbol{\theta})}\right) \left(\frac{\partial}{\partial \boldsymbol{\theta}} f_{\text{mis}}(\boldsymbol{\theta})\right) f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right) \, d\mathbf{y}_{\text{mis}} \\
&= -\int \left(\frac{\partial}{\partial \boldsymbol{\theta}} f_{\text{mis}}(\boldsymbol{\theta})\right) \frac{f_{\text{mis}}\left(\boldsymbol{\theta}^{(t)}\right)}{f_{\text{mis}}(\boldsymbol{\theta})} \, d\mathbf{y}_{\text{mis}}
\end{aligned}$$

Evaluerer vi i punktet $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$ kommer brøken i ovenstående til at forsvinde, da nævner og tæller er ens. Byttes der om på integral og differentiation, bliver integralet af tæthedsfunktionen 1. Differentieres denne fås altså 0. Vi har da, at:

$$\frac{\partial}{\partial \boldsymbol{\theta}} KL\{f_{\text{mis}}(\boldsymbol{\theta}^{(t)}) : f_{\text{mis}}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}} = 0.$$

■

Analogt til EM-algoritmen i forelæsningerne er $\boldsymbol{\theta}^{(t+1)}$ fundet i M-trinet ved at evaluere $\arg\max_{\boldsymbol{\theta}} q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)})$.

Opgave 3 *Vis at dette betyder, at* $q(\boldsymbol{\theta}^{(t+1)}; \boldsymbol{\theta}^{(t)}) \geq 0$.

Bevis
Da M-trinet sætter $\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)})$, så er $q\left(\boldsymbol{\theta}^{(t+1)}; \boldsymbol{\theta}^{(t)}\right) \geq q\left(\boldsymbol{\theta}^{(t)}; \boldsymbol{\theta}^{(t)}\right)$. Definitionen af q giver os:

$$q\left(\boldsymbol{\theta}^{(t)}; \boldsymbol{\theta}^{(t)}\right) = \mathrm{E}\left[\log \frac{f(\mathbf{y}; \boldsymbol{\theta}^{(t)})}{f\left(\mathbf{y}; \boldsymbol{\theta}^{(t)}\right)} \Big| \mathbf{y}_{\text{obs}}; \boldsymbol{\theta}^{(t)}\right] = \mathrm{E}\left[\log(1) \Big| \mathbf{y}_{\text{obs}}; \boldsymbol{\theta}^{(t)}\right] = 0.$$

■

Opgave 4 *Baseret på dette resultat, vis at*

$$\ell_{obs}\left(\boldsymbol{\theta}^{(t+1)}; \mathbf{y}_{obs}\right) \geq \ell_{obs}\left(\boldsymbol{\theta}^{(t)}; \mathbf{y}_{obs}\right).$$

Bevis
Fra (2.1) og forrige resultat har vi:

$$q\left(\boldsymbol{\theta}^{(t+1)}, \boldsymbol{\theta}^{(t)}\right) = \ell_{\text{obs}}(\boldsymbol{\theta}^{(t+1)}; \mathbf{y}_{\text{obs}}) - \ell_{\text{obs}}(\boldsymbol{\theta}^{(t)}; \mathbf{y}_{\text{obs}}) - KL\{f_{\text{mis}}(\boldsymbol{\theta}^{(t)}) : f_{\text{mis}}(\boldsymbol{\theta}^{(t+1)})\} \geq 0.$$

Vi kan lægge $\ell_{\text{obs}}(\boldsymbol{\theta}^{(t)}; \mathbf{y}_{\text{obs}})$ til på begge sider. Ydermere er KL-divergensen positiv, hvilket vil sige, at venstre-siden altid vil blive mindre ved, at KL-divergensen bliver trukket fra. Vi kan altså blot fjerne denne, og dermed opnår vi resultatet. ■