Assignment3

Benjamin Moran

30 August 2016

Given the random vector $\boldsymbol{Y}$, the **Kullback-Leibler Divergence** is the **distance** between two joint pdfs in the same family, indexed by a **parameter** $\boldsymbol{\theta}$, say $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_1)$ and $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_2)$, defined as

$$I(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2) = \frac{1}{n} E_1 \left[log \left(\frac{f_{\boldsymbol{Y}}(\boldsymbol{Y};\boldsymbol{\theta}_1)}{f_{\boldsymbol{Y}}(\boldsymbol{Y};\boldsymbol{\theta}_2)} \right) \right]$$

where E_1 denotes **expectation** with respect to the **joint pdf** determined by $\boldsymbol{\theta}_1$. That is,

$$I(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2) = \frac{1}{n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} log \left(\frac{f_{\boldsymbol{Y}}(\boldsymbol{Y};\boldsymbol{\theta}_1)}{f_{\boldsymbol{Y}}(\boldsymbol{Y};\boldsymbol{\theta}_2)} \right) f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_1) d\boldsymbol{y}$$

Question 1

Let us consider the $n \times 1$ **dependent vector X** and the $n \times k$ **observed matrix of independent variables** Z as defined in Slide 10. We construct two **regression models** between **X** and Z as follows:

$$\begin{cases} (i) & \mathbf{X} = Z\beta_1 + \boldsymbol{\mathcal{W}}_1 \\ (ii) & \mathbf{X} = Z\beta_2 + \boldsymbol{\mathcal{W}}_2 \end{cases}$$

where $\boldsymbol{\mathcal{W}}_1 = (\mathcal{W}_{11}, \cdots, \mathcal{W}_{n1})'$ and $\boldsymbol{\mathcal{W}}_2 = (\mathcal{W}_{12}, \cdots, \mathcal{W}_{n2})'$, $\mathcal{W}_{11}, \cdots, \mathcal{W}_{n1} \overset{iid}{\sim} N[0, \sigma_1^2]$ and $\mathcal{W}_{12}, \cdots, \mathcal{W}_{n2} \overset{iid}{\sim} N[0, \sigma_2^2]$ are two **independent** series. If we define $\boldsymbol{\theta}_1 := (\beta_1', \sigma_1^2)'$ and $\boldsymbol{\theta}_2 := (\beta_2', \sigma_2^2)'$, show that the **Kullback-Leibler Divergence** between the joint pdf of **X** based on models *(i)* and *(ii)* is given as:

$$I(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2) = \frac{1}{2} \left(\frac{\sigma_1^2}{\sigma_2^2} - log \left(\frac{\sigma_1^2}{\sigma_2^2} \right) - 1 \right) + \frac{(\beta_1 - \beta_2)' Z' Z (\beta_1 - \beta_2)}{2n\sigma_2^2}$$

Answer: We are given (in the lecture notes) that the random observation vector $\boldsymbol{X}$ - corresponding to the Regression Model defined in the question - has a **Multivariate Normal** distribution with **mean** vector $Z\beta$ and the **Covariance** matrix $\sigma_{\mathcal{W}}^2 I = \Gamma$, where I is an $n \times n$ identity matrix (Chapter 2: slide 14). The pdf for a Mutivariate Normal Distribution is given by:

$$f_{\boldsymbol{X}}(x;\boldsymbol{\theta}) = (2\pi)^{-\frac{n}{2}} |\Gamma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(x - Z\beta)'\Gamma^{-1}(x - Z\beta) \right\}$$

We can rewrite the expression for the **Kullback-Leibler Divergence** above as:

$$
\begin{aligned}
I(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2) &= \frac{1}{n}\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1) log\left(\frac{f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)}{f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_2)}\right) d\boldsymbol{X} \\
&= \frac{1}{n}\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\, log\left(\frac{(2\pi)^{-\frac{n}{2}}|\Gamma_1|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x-Z\beta_1)'\Gamma_1^{-1}(x-Z\beta_1)\right\}}{(2\pi)^{-\frac{n}{2}}|\Gamma_2|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x-Z\beta_2)'\Gamma_2^{-1}(x-Z\beta_2)\right\}}\right) d\boldsymbol{X} \\
&= -\frac{1}{2n}\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\, log\left(\frac{|\Gamma_1|}{|\Gamma_2|}\right) d\boldsymbol{X} - \frac{1}{2n}\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\,(x-Z\beta_1)'\Gamma_1^{-1}(x-Z\beta_1)\, d\boldsymbol{X} \\
&\qquad + \frac{1}{2n}\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\,(x-Z\beta_2)'\Gamma_2^{-1}(x-Z\beta_2)\, d\boldsymbol{X} \\
&= -\frac{n}{2n} log\left(\frac{\sigma_1^2}{\sigma_2^2}\right)\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\, d\boldsymbol{X} - \frac{(x-Z\beta_1)'(x-Z\beta_1)}{2n\,\Gamma_1}\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\, d\boldsymbol{X} \\
&\qquad + \frac{1}{2n\,\Gamma_2}\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\,(x-Z\beta_2)'(x-Z\beta_2)\, d\boldsymbol{X}
\end{aligned}
$$

Here we need to recall a few things. Firstly, $f_{\boldsymbol{X}}(x;\boldsymbol{\theta})$ is a pdf and therefore integrates to 1. Secondly, we have that:

$$
\begin{aligned}
&(x-Z\beta_1)'(x-Z\beta_1) = \Gamma_1 \\
\implies &\frac{(x-Z\beta_1)'(x-Z\beta_1)}{2n\,\Gamma_1} = \frac{tr(\Gamma_1\Gamma_1^{-1})}{2n} = \frac{tr(I_n)}{2n} = \frac{n}{2n} = \frac{1}{2}
\end{aligned}
$$

Substituting these results into the equation leaves:

$$
I(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2) = -\frac{1}{2} log\left(\frac{\sigma_1^2}{\sigma_2^2}\right) - \frac{1}{2} + \frac{1}{2n\,\Gamma_2}\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\,(x-Z\beta_2)'(x-Z\beta_2)\, d\boldsymbol{X}
$$

Now let's simplify the remaining integral:

$$
\begin{aligned}
\int\limits_{\mathbb{R}^n} f_{\boldsymbol{X}}(x;\boldsymbol{\theta}_1)\,(x-Z\beta_2)'(x-Z\beta_2)\, d\boldsymbol{X} &= E_1\left[(x-Z\beta_2)'(x-Z\beta_2)\right] \\
&= \Gamma_1 + (\beta_1-\beta_2)'Z'Z(\beta_1-\beta_2)
\end{aligned}
$$

Substituting this back into the original, we get:

$$
\begin{aligned}
I(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2) &= -\frac{1}{2} log\left(\frac{\sigma_1^2}{\sigma_2^2}\right) - \frac{1}{2} + \frac{1}{2n\,\Gamma_2}\left(\Gamma_1 + (\beta_1-\beta_2)'Z'Z(\beta_1-\beta_2)\right) \\
&= -\frac{1}{2} log\left(\frac{\sigma_1^2}{\sigma_2^2}\right) - \frac{1}{2} + \frac{tr(I_n)\sigma_1^2}{2n\,\sigma_2^2} + \frac{(\beta_1-\beta_2)'Z'Z(\beta_1-\beta_2)}{2n\,\sigma_2^2} \\
&= -\frac{1}{2} log\left(\frac{\sigma_1^2}{\sigma_2^2}\right) - \frac{1}{2} + \frac{\sigma_1^2}{2\,\sigma_2^2} + \frac{(\beta_1-\beta_2)'Z'Z(\beta_1-\beta_2)}{2n\,\sigma_2^2} \\
&= \frac{1}{2}\left(\frac{\sigma_1^2}{\sigma_2^2} - log\left(\frac{\sigma_1^2}{\sigma_2^2}\right) - 1\right) + \frac{(\beta_1-\beta_2)'Z'Z(\beta_1-\beta_2)}{2n\,\sigma_2^2}
\end{aligned}
$$

as required.

Bonus Question

If the **true** value of the parameter vector is $\boldsymbol{\theta} = (\beta', \sigma^2)'$ and the **estimated** value based on the **sample** $\widehat{\boldsymbol{\theta}} = (\widehat{\beta}', \widehat{\sigma^2})'$, one may argue that the **best** model would be one that **minimizes** the **Kullback-Leibler distance** between the joint-pdfs of **theoretical** value and the **sample** estimation, say $I(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}})$. Because $\boldsymbol{\theta}$ will not be known, Hurvich and Tsai (1989) considered finding an **unbiased estimator** for $E_{\boldsymbol{\theta}}[I(\beta, \sigma^2; \widehat{\beta}, \widehat{\sigma}^2)]$, where

$$I(\beta, \sigma^2; \widehat{\beta}, \widehat{\sigma}^2) = \frac{1}{2}\left(\frac{\sigma^2}{\widehat{\sigma}^2} - log\left(\frac{\sigma^2}{\widehat{\sigma}^2}\right) - 1\right) + \frac{(\beta - \widehat{\beta})' Z' Z (\beta - \widehat{\beta})}{2n\widehat{\sigma}^2}$$

and β is a $k \times 1$ regression parameter vector. Show that

$$E_{\boldsymbol{\theta}}[I(\beta_1, \sigma_1^2; \widehat{\beta}, \widehat{\sigma}^2)] = \frac{1}{2}\left(-log(\sigma^2) + E_{\boldsymbol{\theta}}[log(\widehat{\sigma}^2)] + \frac{n+k}{n-k-2} - 1\right).$$

Answer: Expectation is a linear function, so we can rewrite the above as

$$\begin{aligned}
E_{\boldsymbol{\theta}} I(\beta, \sigma^2; \widehat{\beta}, \widehat{\sigma}^2) &= \frac{1}{2} E\left[\frac{\sigma^2}{\widehat{\sigma}^2} - log\left(\frac{\sigma^2}{\widehat{\sigma}^2}\right) - 1 + \frac{(\beta - \widehat{\beta})' Z' Z (\beta - \widehat{\beta})}{n\widehat{\sigma}^2}\right] \\
&= \frac{1}{2}\left(E\left[\frac{\sigma^2}{\widehat{\sigma}^2}\right] - E\left[log\left(\frac{\sigma^2}{\widehat{\sigma}^2}\right)\right] - E\left[1\right] + E\left[\frac{(\beta - \widehat{\beta})' Z' Z (\beta - \widehat{\beta})}{n\widehat{\sigma}^2}\right]\right)
\end{aligned}$$

From the reference text, we are given that:

$$\begin{aligned}
\frac{n\widehat{\sigma}^2}{\sigma^2} &\sim \chi^2_{n-k} \\
\frac{(\widehat{\beta} - \beta)' Z' Z (\widehat{\beta} - \beta)}{n\widehat{\sigma}^2} &\sim \chi^2_k
\end{aligned}$$

We are also given that if $x \sim \chi^2_n \implies E[(\frac{1}{x})] = \frac{1}{n-2}$. So:

$$\begin{aligned}
E\left[\left(\frac{\sigma^2}{\widehat{\sigma}^2}\right)\right] &= n * \left(\frac{1}{(n-k)-2}\right) \\
&= \frac{n}{n-k-2}
\end{aligned}$$

In class it was shown that:

$$\frac{k}{n-k} \frac{(\widehat{\beta} - \beta)' Z' Z (\widehat{\beta} - \beta)}{n\widehat{\sigma}^2} \sim F_{k, n-k}$$

Given that $E\left[F_{k,n-k}\right] = (n-k)/(n-k-2)$, we get:

$$\begin{aligned}
E\left[\frac{(\widehat{\beta} - \beta)' Z' Z (\widehat{\beta} - \beta)}{n\widehat{\sigma}^2}\right] &= \frac{k}{n-k} \frac{n-k}{n-k-2} \\
&= \frac{k}{n-k-2}
\end{aligned}$$

Taking the expectation of a scalar returns the scalar. So we can simplify the original equation as:

$$\begin{aligned}
E_{\boldsymbol{\theta}} I(\beta, \sigma^2; \widehat{\beta}, \widehat{\sigma}^2) &= \frac{1}{2}\left[\frac{n}{n-k-2} - E_{\boldsymbol{\theta}}\left[log\left(\frac{\sigma^2}{\widehat{\sigma}^2}\right)\right] - 1 + \frac{k}{n-k-2}\right] \\
&= \frac{1}{2}\left[-E_{\boldsymbol{\theta}}\left[log\left(\frac{\sigma^2}{\widehat{\sigma}^2}\right)\right] + \frac{n+k}{n-k-2} - 1\right]
\end{aligned}$$

We can rewrite the remaining log term using the fact that $log(a/b) = log(a) - log(b)$.

$$\begin{aligned}
E_{\boldsymbol{\theta}} I(\beta, \sigma^2; \widehat{\beta}, \widehat{\sigma}^2) &= \frac{1}{2}\left[-\left(E_{\boldsymbol{\theta}}\left[log\left(\sigma^2\right) - log\left(\widehat{\sigma}^2\right)\right]\right) + \frac{n+k}{n-k-2} - 1\right] \\
&= \frac{1}{2}\left[-E_{\boldsymbol{\theta}}\left[log\left(\sigma^2\right)\right] + E_{\boldsymbol{\theta}}\left[log\left(\widehat{\sigma}^2\right)\right] + \frac{n+k}{n-k-2} - 1\right]
\end{aligned}$$

We don't know how to evaluate $E_{\boldsymbol{\theta}}\left[log\left(\widehat{\sigma}^2\right)\right]$, but $log\left(\sigma^2\right)$ is a constant - not a R.V - so we can rewrite the equation as:

$$E_{\boldsymbol{\theta}} I(\beta, \sigma^2; \widehat{\beta}, \widehat{\sigma}^2) = \frac{1}{2}\left[-log\left(\sigma^2\right) + E_{\boldsymbol{\theta}}\left[log\left(\widehat{\sigma}^2\right)\right] + \frac{n+k}{n-k-2} - 1\right]$$

as required.
