Homework 1

Ki Hyun
FINM 33210: Bayesian Statistical Inference and Machine Learning

Due: 23:59 (CT) Mar 31st 2023

1: Property C5

Let U be the probability space.

The five given axioms are:

$$\mathbb{P}\{\phi\} = 0 \tag{C0}$$

$$\text{If } A_1 \cap A_2 = \phi, \text{ then } \mathbb{P}\{A_1 \cup A_2\} = \mathbb{P}\{A_1\} + \mathbb{P}\{A_2\} \tag{C1}$$

$$\mathbb{P}\{A^c\} = 1 - \mathbb{P}\{A\} \tag{C2}$$

$$0 \leq \mathbb{P}\{A\} \leq 1 \tag{C3}$$

$$\text{If } A \subset B, \text{ then } \mathbb{P}\{B\} = \mathbb{P}\{A\} + \mathbb{P}\{B \backslash A\} \geq \mathbb{P}\{A\} \tag{C4}$$

We want to show:

$$\mathbb{P}\{A \cup B\} = \mathbb{P}\{A\} + \mathbb{P}\{B\} - \mathbb{P}\{A \cap B\} \tag{C5}$$

First we know from the definition of \:

$$A \cap (B \backslash A) = \phi \tag{1}$$

However, we also know that:

$$A \cup (B \backslash A) = A \cup B \tag{2}$$

Therefore, from (1), (2) and (C1):

$$\begin{aligned}\mathbb{P}\{A \cup B\} &= \mathbb{P}\{A \cup (B \backslash A)\} \\ &= \mathbb{P}\{A\} + \mathbb{P}\{B \backslash A\}\end{aligned}$$

Here, we know by definition that $(A \cap B) \subset A$ and therefore from (1):

$$(B \backslash A) \cap (A \cap B) = \phi \tag{3}$$

Also, by definition,

$$(B \backslash A) \cup (A \cap B) = B \tag{4}$$

Therefore, from (3), (4) and (C1):

$$\mathbb{P}\{B\} = \mathbb{P}\{(B \backslash A) \cup (A \cap B)\} = \mathbb{P}\{B \backslash A\} + \mathbb{P}\{A \cap B\} \tag{5}$$

Moreover,

$$\begin{aligned}\mathbb{P}\{A \cup B\} &= \mathbb{P}\{A \cup (B \backslash A)\} \\ &= \mathbb{P}\{A\} + \mathbb{P}\{B \backslash A\} \\ &= \mathbb{P}\{A\} + \mathbb{P}\{B\} - \mathbb{P}\{A \cap B\} \\ &(\because (5))\end{aligned}$$

$$Q.E.D.$$

2: Univariate Linear Regression

(a)

Let $\mathbf{X}$ be the $n \times 2$ matrix representing each data (including the constant) such that

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

Similarly, let $\vec{y}$ be the vector of the observed endogenous variable such that

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Since it was given that

$$y_i - \theta_0 - \theta_1 x_i = \epsilon \sim N(0, s^2)$$

From the notation above, assuming that the data are i.i.d.,

$$\vec{y} \sim MVN(X\theta, s^2\mathbf{I})$$

Since this is a special case of the multivariate normal distribution where the correlation between the entries in $\vec{y}$ are independent the probability density function can be written as:

$$p(\vec{y} \mid \theta) = \frac{e^{-\frac{1}{2}\frac{\|\vec{y}-\mathbf{X}\theta\|^2}{s^2}}}{(2\pi s^2)^{\frac{n}{2}}}$$

(b)

$$\begin{aligned}
p(\theta \mid \vec{y}) &\propto p(\vec{y} \mid \theta)p(\theta) \\
&\propto \frac{\exp(-\frac{1}{2}||\vec{y} - X\theta||^2/s^2)}{(2\pi s^2)^{\frac{n}{2}}} \cdot \frac{\exp\left(-\frac{1}{2}\left[\frac{(\theta_0-\mu_0)^2}{\sigma_0^2} + \frac{(\theta_1-\mu_1)^2}{\sigma_1^2}\right]\right)}{2\pi\sigma_0\sigma_1} \\
&\propto \exp(-\frac{1}{2}||\vec{y} - X\theta||^2/s^2) \cdot \exp\left(-\frac{1}{2}\left[\frac{(\theta_0-\mu_0)^2}{\sigma_0^2} + \frac{(\theta_1-\mu_1)^2}{\sigma_1^2}\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^n (y_i - \theta_o - \theta_1 x_i)^2}{s^2}\right) \cdot \exp\left(-\frac{1}{2}\left[\frac{(\theta_0-\mu_0)^2}{\sigma_0^2} + \frac{(\theta_1-\mu_1)^2}{\sigma_1^2}\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&\propto \exp\left(-\frac{1}{2}\left\{\frac{\sum_{i=1}^{n} -2\theta_0 y_i - 2\theta_1 x_i y_i + \theta_0^2 + 2\theta_0\theta_1 x_i + \theta_1^2 x_i^2}{s^2} + \left[\frac{\theta_0^2 - 2\mu_0\theta_0}{\sigma_0^2} + \frac{\theta_1^2 - 2\mu_1\theta_1}{\sigma_1^2}\right]\right\}\right) \\
&\propto \exp(-\frac{1}{2}\{\theta_0^2\left[\frac{1}{\sigma_0^2} + \frac{n}{s^2}\right] - 2\theta_0\left[\frac{\sum_{i=1}^{n} y_i}{s^2} + \frac{\mu_0}{\sigma_0^2}\right] + \theta_1^2\left[\frac{\sum_{i=1}^{n} x_i}{s^2} + \frac{1}{\sigma_1^2}\right] \\
&\quad - 2\theta_1\left[\frac{\sum_{i=1}^{n} y_i x_i}{s^2} + \frac{\mu_i}{\sigma_1^2}\right] + 2\theta_0\theta_1\frac{\sum_{i=1}^{n} x_i}{s^2}\}) \\
&\propto \exp\left(-\frac{1}{2}\left\{\left[\frac{1}{\sigma_0^2} + \frac{n}{s^2}\right]\left[\theta_0 - \frac{\frac{\sum_{i=1}^{n} y_i}{s^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{n}{s^2}}\right]^2 + \left[\frac{\sum_{i=1}^{n} x_i^2}{s^2} + \frac{1}{\sigma_1^2}\right]\left[\theta_0 - \frac{\frac{\sum_{i=1}^{n} x_i y_i}{s^2} + \frac{\mu_1}{\sigma_1^2}}{\frac{x_i^2}{s^2} + \frac{1}{\sigma_1^2}}\right]^2 + C\right\}\right)
\end{aligned}
$$

Therefore, in proportional form:

$$
p(\theta \mid \vec{y}) \propto exp\left(-\frac{1}{2}\left\{\frac{(\theta_0 - \tilde{\mu}_0)^2}{\tilde{\sigma}_0^2} + \frac{(\theta_1 - \tilde{\mu}_1)^2}{\tilde{\sigma}_1^2}\right\} + C\right)
$$

Here, the posterior parameters are:

$$
\begin{aligned}
\tilde{\mu}_0 &= \frac{\frac{\sum_{i=1}^{n} y_i}{s^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{n}{s^2}} \\
\tilde{\sigma}_0^2 &= \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{s^2}} \\
\tilde{\mu}_1 &= \frac{\frac{\sum_{i=1}^{n} x_i y_i}{s^2} + \frac{\mu_1}{\sigma_1^2}}{\frac{x_i^2}{s^2} + \frac{1}{\sigma_1^2}} \\
\tilde{\sigma}_1^2 &= \frac{1}{\frac{x_i^2}{s^2} + \frac{1}{\sigma_1^2}} \\
\tilde{\rho} &= \frac{\frac{\sum_{i=1}^{n} x_i y_i}{s^2} - \frac{\frac{\sum_{i=1}^{n} y_i}{s^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{n}{s^2}} \cdot \frac{\frac{\sum_{i=1}^{n} x_i^2}{s^2} + \frac{1}{\sigma_1^2}}{\frac{\sum_{i=1}^{n} x_i^2}{s^2} + \frac{1}{\sigma_1^2}}}{\sqrt{\frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{s^2}} \cdot \frac{1}{\frac{\sum_{i=1}^{n} x_i^2}{s^2} + \frac{1}{\sigma_1^2}}}}
\end{aligned}
$$

(c)

The posterior distribution also follows a normal distribution.

Since the normal distribution was a prior, this prior would be a conjugate prior for this likelihood.