1. 증명하시오.

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

i) $\hat{\beta}$ 구하기 (OLS)

$$\min \sum^n e_i^2 = \min (Y - X\beta)^T (Y - X\beta) \quad \text{``} SSR(\beta)$$

$$\rightarrow \frac{d}{d\beta} (Y^T Y - \beta^T X^T Y - Y^T X\beta + \beta^T X^T X\beta)$$

$$= -2Y^T X + \beta^T (X^T X + X^T X) \overset{let}{=} 0$$

$$\rightarrow \beta^T X^T X = Y^T X$$

$$\rightarrow X^T X\beta = X^T Y$$

$$\rightarrow \hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$$

ii) $E(\hat{\beta})$ 구하기

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta + e)$$

$$= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T e$$

$$= \beta + (X^T X)^{-1} X^T e$$

$\therefore E(\hat{\beta}) = \beta \quad \because E(X^T e) = 0$

iii) $Var(\hat{\beta})$ 구하기

$$Var(\hat{\beta}) = Var((X^T X)^{-1} X^T Y)$$

$$= Var(\beta + (X^T X)^{-1} X^T e)$$

$$= Var((X^T X)^{-1} X^T e)$$

$$= E((X^T X)^{-1} X^T e e^T X (X^T X)^{-1})$$

$$= (X^T X)^{-1} X^T E(ee^T) X (X^T X)^{-1}$$

$$= (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1} \quad \because \text{by homoskedasticity, } Var(e_i | X) = \sigma^2$$

$$= \sigma^2 (X^T X)^{-1}$$

iv) $\hat{\beta}$의 분포 형태 추론

$$\begin{cases} \hat{\beta} = (X^T X)^{-1} X^T Y & : \text{linear combination of } Y \\ Y \sim N(\quad) & \end{cases}$$

$$\rightarrow \hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

2. 증명하시오.

Slopes posterior $\beta \mid y, \sigma^2 \sim N(\beta_n, \Sigma_n)$

With $\beta \sim N(\beta_0, \Sigma_0)$ prior and $y \mid \beta, \sigma^2 \sim N(X\beta, \sigma^2)$ likelihood, and **known** σ^2, we have

$$
\begin{aligned}
p(\beta \mid y, \sigma^2) &\propto p(y \mid \beta, \sigma^2) p(\beta) \\
&\propto \exp\Big(-\frac{1}{2\sigma^2}(y^T y - 2\beta^T X^T y + \beta^T X^T X \beta) - \frac{1}{2}(\beta^T \Sigma_0^{-1} \beta - 2\beta^T \Sigma_0^{-1} \beta_0) \Big) \\
&\propto \exp\Big(-\frac{1}{2}\beta^T \underbrace{(X^T X/\sigma^2 + \Sigma_0^{-1})}_{\Sigma_n^{-1}} \beta + \beta^T \underbrace{(X^T y/\sigma^2 + \Sigma_0^{-1}\beta_0)}_{\Sigma_n^{-1}\beta_n} \Big)
\end{aligned}
$$

$$
\therefore \Sigma_n^{-1} = \underbrace{X^T X/\sigma^2}_{\text{data precision}} + \underbrace{\Sigma_0^{-1}}_{\text{prior precision}}, \quad \beta_n = \Sigma_n (X^T y \underbrace{/\sigma^2}_{\text{data weight}} + \underbrace{\Sigma_0^{-1}}_{\text{prior weight}} \beta_0)
$$

for $|\Sigma_0^{-1}| < \epsilon$ (weak prior), $\beta_n \approx \hat\beta_{mle}$. Don't want the bias from β_0? Choose $\beta_0 = 0, \Sigma_0 = g\sigma^2(X^T X)^{-1}$, then $\beta \mid y, \sigma^2 \sim N\Big(\frac{g}{g+1}\hat\beta_{mle}, \frac{g}{g+1}V(\hat\beta_{mle})\Big)$ → **g-prior!** (Higher g means a weaker prior)

HW 2

i) $p(\beta \mid y, X, \sigma^2) \propto p(y \mid \beta, X, \sigma^2) \times p(\beta)$ 이고, $y \mid \beta, X, \sigma^2 \sim MVN(\quad)$ 이므로

β 에 대한 prior를 conjugate한 형태로 맞춰주기 위해서는 똑같이 MVN 형태가 좋다.

이때 $\beta \sim MVN(\beta_0, \Sigma_0)$ 라 하자.

ii) find $E[\beta \mid y, X, \sigma^2]$ in terms of g

$E[\beta \mid y, X, \sigma^2] = (\Sigma_0^{-1} + X^T X/\sigma^2)^{-1}(\Sigma_0^{-1}\beta_0 + X^T y/\sigma^2)$ 이므로 여기에 $\beta_0 = 0$, $\Sigma_0 = g\sigma^2(X^T X)^{-1}$ 대입

$$\rightarrow \left(\frac{1}{g\sigma^2}X^T X + \frac{1}{\sigma^2}X^T X\right)^{-1} \frac{1}{\sigma^2}X^T y$$

$$= \frac{g}{g+1}(X^T X)^{-1}X^T y$$

$$= \frac{g}{g+1}\hat\beta_{MLE}$$ ($\because$ $\hat\beta_{ols}$가 unbiased이므로 $\hat\beta_{MLE}$와 같다.)

iii) find $Var(\beta \mid y, X, \sigma^2)$ in terms of g

$Var(\beta \mid y, X, \sigma^2) = (\Sigma_0^{-1} + X^T X/\sigma^2)^{-1}$

여기에 $\Sigma_0 = g\sigma^2(X^T X)^{-1}$ 대입

$$\rightarrow Var(\beta \mid y, X, \sigma^2) = \left(\frac{1}{g\sigma^2}X^T X + \frac{1}{\sigma^2}X^T X\right)^{-1}$$

$$= \left(\frac{g+1}{g\sigma^2}X^T X\right)^{-1}$$

$$= \frac{g}{g+1}\sigma^2(X^T X)^{-1}$$

$$= \frac{g}{g+1}V(\hat\beta_{MLE})$$

결론: $\beta \mid y, \sigma^2 \sim MVN\left(\frac{g}{g+1}\hat\beta_{MLE}, \frac{g}{g+1}V(\hat\beta_{MLE})\right)$

ㅠ. 괜히 한번 ~

* Posterior for full conjugate prior 계산해보기

$$\cdot\ p(\beta, \sigma^2 | y) \propto \underbrace{p(\beta | y, \sigma^2)}_{\text{posterior for g-prior}}\ p(\sigma^2 | y)$$

$$p(\sigma^2|y) \propto p(\sigma^2)\, p(y|\sigma^2)$$

$$p(y|\sigma^2) = \int p(y|\beta, \sigma^2)\, p(\beta|\sigma^2)\, d\beta$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \frac{1}{|g\sigma^2(X^TX)^{-1}|^{\frac{1}{2}}} \int \exp\left[-\frac{1}{2\sigma^2}(\|y - X\beta\|^2) - \frac{1}{2g\sigma^2}\beta^TX^TX\beta\right] d\beta$$

$$= \quad '' \quad \exp\left(-\frac{1}{2\sigma^2}y^Ty\right) \int \exp\left[\frac{1}{\sigma^2}\beta^TX^Ty - \frac{1}{2\sigma^2}\left(1+\frac{1}{g}\right)\beta^TX^TX\beta\right] d\beta$$

exp 안에서 $\left(1+\frac{1}{g}\right)\beta^TX^TX\beta$를 자세히 보면 다음과 같다.

$\star\ \left(1+\frac{1}{g}\right)\beta^TX^TX\beta$ 구하기

remind $V = \frac{g}{g+1}\sigma^2(X^TX)^{-1}$, $m = \frac{g}{g+1}(X^TX)^{-1}X^Ty$

$\quad \hookrightarrow V^{-1} = \frac{g+1}{g}\cdot\frac{1}{\sigma^2}X^TX$

$$\beta^TX^TX\beta = (\beta-m)^TX^TX(\beta-m) + 2m^TX^TX\beta - m^TX^TXm$$

i) $\frac{g+1}{g}(\beta-m)^TX^TX(\beta-m) = \sigma^2\cdot(\beta-m)^TV^{-1}(\beta-m)$

ii) $\frac{g+1}{g}\,2m^TX^TX\beta = \frac{g+1}{g}\cdot 2\,\beta^TX^TXm$

$$= \cancel{\frac{g+1}{g}}\cdot 2\,\beta^T\cancel{X^TX}\cdot\cancel{\frac{g}{g+1}}\cancel{(X^TX)^{-1}}X^Ty$$

$$= 2\beta^TX^Ty$$

iii) $\frac{g+1}{g}m^TX^TXm = \cancel{\frac{g+1}{g}}m^T\left(\cancel{\frac{g}{g+1}}\cdot\sigma^2\cdot V^{-1}\right)m$

$$= \sigma^2 m^TV^{-1}m$$

$$\rightarrow \left(1+\frac{1}{g}\right)\beta^TX^TX\beta = \sigma^2(\beta-m)^TV^{-1}(\beta-m) + 2\beta^TX^Ty - \sigma^2m^TV^{-1}m$$

$$\rightarrow p(y|\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}}\frac{1}{|g\sigma^2(X^TX)^{-1}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2\sigma^2}y^Ty + \frac{1}{2}m^TV^{-1}m\right)\int \exp\left[-\frac{1}{2}(\beta-m)^TV^{-1}(\beta-m)\right]d\beta$$

$$= \quad '' \quad '' \quad '' \quad \left(\frac{1}{2\pi}\right)^{-\frac{n}{2}}\left|\frac{g}{g+1}\sigma^2(X^TX)^{-1}\right|^{\frac{1}{2}}$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}}(1+g)^{-\frac{p}{2}}\left(\frac{1}{2\pi}\right)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma^2}SSR(g)\right)$$

where $SSR(g) = y^Ty - \sigma^2 m^TV^{-1}m$

$SSR(g) = y^T y - \sigma^2 m^T V^{-1} m$ 에서

$$m^T V^{-1} m = \left[\frac{g}{g+1}(X^TX)^{-1}X^Ty\right]^T \cdot \frac{g+1}{g}\frac{1}{\sigma^2}(X^TX)\left[\frac{g}{g+1}(X^TX)^{-1}X^Ty\right]$$

$$= \frac{1}{\sigma^2}\cdot\frac{g}{g+1}\, y^T X \cancel{(X^TX)^{-1}}\cancel{(X^TX)}(X^TX)^{-1}X^Ty$$

$$= \frac{1}{\sigma^2}\cdot\frac{g}{g+1}\, y^T X(X^TX)^{-1}X^Ty$$

$$\rightarrow SSR(g) = y^Ty - \frac{g}{g+1}y^TX(X^TX)^{-1}X^Ty$$

$$= y^T\left(I - \frac{g}{g+1}X(X^TX)^{-1}X^T\right)y$$

3. 증명하시오

over all possible $\beta \mid \sigma^2, y$, incurred by g. Note that as $g \to \infty$ (weak prior), $SSR(g) \to SSR(\hat\beta_{mle})$.

Hw ③

i) $g \to \infty$

$$SSR(g) \to y^T(I - X(X^TX)^{-1}X^T)y$$

ii) $SSR(\hat\beta_{MLE})$ 구하기

$$SSR(\hat\beta_{MLE}) = (y - X\hat\beta)^T(y - X\hat\beta)$$

$$= (y - X(X^TX)^{-1}X^Ty)^T(y - X(X^TX)^{-1}X^Ty)$$

$$= y^Ty - 2y^TX(X^TX)^{-1}X^Ty + y^TX\cancel{(X^TX)^{-1}}\cancel{X^TX}(X^TX)^{-1}X^Ty$$

$$= y^Ty - 2y^TX(X^TX)^{-1}X^Ty + y^TX(X^TX)^{-1}X^Ty$$

$$= y^Ty - y^TX(X^TX)^{-1}X^Ty$$

★ $g \to \infty$이면 i)과 ii)은 같아진다.
(prior에 대한 믿음수준 ↓)