a note on beyesian linear regression

general form:

$$y_i = \beta^T x_i + \epsilon_i$$

now assume y_i comes from $N(\beta^T x_i, \sigma^2)$

compute the likelihood given assumption that ϵ_i are independent:

$$\begin{aligned} P(y|X,\beta,\sigma^2) &= \prod_{i=1}^{n} P(y_i|x_i,\beta,\sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\left\{\frac{1}{2\sigma^2}(y_i - \beta^T x_i)^2\right\} \\ &= (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta^T x_i)^2\right\} \end{aligned}$$

note that,

$$\begin{aligned} \sum_{i=1}^{n}(y_i - \beta^T x_i)^2 &= (y - X\beta)^T(y - X\beta) \\ &= y^T y - 2\beta^T X^T y + \beta^T X^T X\beta \end{aligned}$$

then we will compute the conditional likelihood $P(\beta|y,x,\sigma^2)$. Suppose we have known the value of σ^2, and prior distribution of β comes from $N_k(\beta_0,\Sigma_0)$. Then $P(\beta) = ((2\pi)^k|\Sigma_0|)^{-1/2} exp\left\{-\frac{1}{2}(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0)\right\}$.

According to bayes throey,

$$
\begin{aligned}
&P(\beta|y,X,\sigma^2)\\
&\propto P(\beta)P(y|X,\beta,\sigma^2)\\
&=((2\pi)^k|\Sigma_0|)^{-1/2}exp\left\{-\frac{1}{2}(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0)\right\}\\
&\qquad\qquad * (2\pi\sigma^2)^{-n/2}exp\left\{-\frac{1}{2\sigma^2}\left(y^Ty-2\beta^TX^Ty+\beta^TX^TX\beta\right)\right\}\\
&\propto exp\left\{-\frac{1}{2}\left(\beta^T\Sigma_0^{-1}\beta_0-\beta^T\Sigma_0^{-1}\beta_0-\beta_0^T\Sigma_0^{-1}\beta+\beta_0^T\Sigma_0^{-1}\beta_0\right)\right\}\\
&\qquad\qquad * exp\left\{-\frac{1}{2\sigma^2}\left(-2\beta^TX^Ty+\beta^TX^TX\beta\right)\right\}\\
&\propto exp\left\{\beta^T\Sigma_0^{-1}\beta_0-\frac{1}{2}\beta^T\Sigma_0^{-1}\beta+\beta^TX^Ty/\sigma^2-\frac{1}{2}\beta^TX^TX\beta/\sigma^2\right\}\\
&=exp\left\{\beta^T\left(\Sigma_0^{-1}\beta_0+X^Ty/\sigma^2\right)-\frac{1}{2}\beta^T\left(\Sigma_0^{-1}+X^TX/\sigma^2\right)\beta\right\}
\end{aligned}
$$

this is proportional to a multivariate normal distribution(to be proved later), that means:

$$\beta|y,X,\sigma^2\sim N(A_{\sigma^2}^{-1}B_{\sigma^2},A_{\sigma^2}^{-1})$$

where $A_{\sigma^2}=\Sigma_0^{-1}+X^TX/\sigma^2$, $B_{\sigma^2}=\Sigma_0^{-1}\beta_0+X^Ty/\sigma^2$

Do something similar, now we have known the value of β. Let $\gamma=\frac{1}{\sigma^2}$ and prior distribution of γ comes from $gamma\left(v_0/2,v_0\sigma^2/2\right)$. Recall that the pdf for a $gamma(x|a,b)$ is

$$f(x|a,b)=\frac{b^ax^{a-1}e^{-bx}}{\Gamma(a)}$$

so the prior distribution of γ is

$$P(\gamma)\propto(v_0\sigma_0^2/2)^{v_0/2}\gamma^{\frac{v_0}{2}-1}exp\left\{-\gamma v_0\sigma^2/2\right\}$$

the likelihood is, as mentioned before,

$$
\begin{aligned}
P(y|X,\beta,\gamma=1/\sigma^2)&=(2\pi\sigma^2)^{-n/2}exp\left\{-\frac{1}{2\sigma^2}\left(y^Ty-2\beta^TX^Ty+\beta^TX^TX\beta\right)\right\}\\
&\propto\gamma^{n/2}exp\left\{-\frac{\gamma}{2}\left(y^Ty-2\beta^TX^Ty+\beta^TX^TX\beta\right)\right\}
\end{aligned}
$$

Again using bayesian theory,

$$
\begin{aligned}
P(\gamma|y,X,\beta) &\propto P(\gamma)P(y|X,\beta,\gamma) \\
&\propto \gamma^{n/2} exp\left\{-\frac{\gamma}{2}\left(y^Ty-2\beta^TX^Ty+\beta^TX^TX\beta\right)\right\} * \gamma^{\frac{v_0}{2}-1} exp\left\{-\gamma v_0\sigma^2/2\right\} \\
&= \gamma^{\frac{1}{2}(v_0+n)-1} exp\left\{-\frac{\gamma}{2}\left(v_0\sigma_0^2+y^Ty-2\beta^TX^Ty+\beta^TX^TX\beta\right)\right\}
\end{aligned}
$$

that is

$$
\gamma|y,X,\beta \sim gamma\left(\frac{v_0+n}{2},\frac{1}{2}\left(v_0\sigma_0^2+y^Ty-2\beta^TX^Ty+\beta^TX^TX\beta\right)\right)
$$

which is equivalent to

$$
\sigma^2|y,X,\beta \sim inverse-gamma\left(\frac{v_0+n}{2},\frac{1}{2}\left(v_0\sigma_0^2+SSR(\beta)\right)\right)
$$

where $SSR(\beta)=y^Ty-2\beta^TX^Ty+\beta^TX^TX\beta$

after we have the posterior distribution, we do the Gibbs sampling.

STEP 1: sample γ_0 fron $P(\gamma)$, $\sigma_0^2=1/\gamma_0$
STEP 2: sample β_t from $N(A_{\sigma_t^2}^{-1}B_{\sigma_t^2},A_{\sigma_t^2}^{-1})$
STEP 3: sample σ_{t+1}^2 from $inverse-gamma(\frac{v_0+n}{2},\frac{1}{2}(v_0\sigma_0^2++SSR(\beta_t))),t=t+1$
STEP 4: return to STEP 2, t=0 to 10000