Q1

我们已经知道，c_{jj}作为方差扩大因子，满足

$$Var(\hat{\beta}_j) = \frac{c_{jj}}{l_{jj}}\sigma^2$$

对于标准化后的样本，事实上

$$Var(\hat{\beta}_j) = c_{jj}\sigma^2$$

另一方面，在线性回归中，我们有

$$Var(\hat{\beta}_j) = \sigma^2(X_s'X_s)^{-1}_{jj}$$

不失一般性，我们假设$X_s = (X_j, x_j)$，即x_j是标准化后的最后一个特征，我们有

$$\begin{aligned}
(X_s'X_s)^{-1} &= \begin{pmatrix} X_j'X_j & X_j'x_j \\ x_j'X_j & x_j'x_j \end{pmatrix}^{-1} & (1)\\
&= \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} & (2)
\end{aligned}$$

这是一个分块矩阵的逆，因为我们关心的是$(X_s'X_s)^{-1}_{jj}$这一对角元，因此我们只关心其中E_{22}的结果，利用分块矩阵求逆公式，我们有

$$\begin{aligned}
E_{22} &= (x_j'x_j - x_j'X_j(X_j'X_j)^{-1}X_j'x_j)^{-1} & (3)\\
Var(\hat{\beta}_j) &= \sigma^2 E_{22} & (4)
\end{aligned}$$

我们接下来考察对于x_j建立的线性回归的模型的性质，我们发现X_j是这一回归模型的设计矩阵，记模型的帽子矩阵为$H_j = X_j(X_j'X_j)^{-1}X_j'$，由于$\sum_i x_{ij} = \bar{x}_j = 0, \sum_i x_{ij}^2 = 1$，我们有

$$\begin{aligned}
SS_{Rj} &= \sum_i (\hat{x}_{ij} - \bar{x}_j)^2 & (5)\\
&= \sum_i \hat{x}_{ij}^2 & (6)\\
&= x_j'H_j'H_jx_j & (7)\\
SS_{Tj} &= \sum_i (x_{ij}^2 - \bar{x}_j)^2 & (8)\\
&= \sum_i x_{ij}^2 & (9)\\
&= 1 & (10)\\
R_j^2 &= \frac{SS_{Rj}}{SS_{Tj}} & (11)\\
&= x_j'H_j'H_jx_j & (12)
\end{aligned}$$

利用上面的讨论，我们重新考察E_{22}的值，注意其中$x_j'x_j = \sum_i x_{ij}^2 = 1$，我们有

$$\begin{aligned}
E_{22} &= (x_j'x_j - x_j'X_j(X_j'X_j)^{-1}X_j'x_j) & (13)\\
&= (x_j'x_j - x_j'H_jx_j) & (14)\\
&= (x_j'x_j - x_j'H_j'H_jx_j) & (15)\\
&= (1 - R_j^2)^{-1} & (16)\\
Var(\hat{\beta}_j) &= \sigma^2(1 - R_j^2)^{-1} & (17)
\end{aligned}$$

同时，我们知道$Var(\hat\beta_j)=c_{jj}\sigma^2$，所以

$$c_{jj}=\frac{1}{1-R_j^2}$$

Q2

我们知道，$E(\hat\beta)=\beta$

$$\begin{aligned}
MSE(\hat\beta) &= E(\hat\beta-\beta)'(\hat\beta-\beta) && (18)\\
&= E(\hat\beta'\hat\beta-\beta'\hat\beta-\hat\beta'\beta+\beta'\beta) && (19)\\
&= E(\hat\beta'\hat\beta)-\beta'\beta-\beta'\beta+\beta'\beta && (20)\\
&= E(\hat\beta'\hat\beta)-\beta'\beta && (21)
\end{aligned}$$

同时，我们知道$E(\hat\beta'\hat\beta)=Tr(Var(\hat\beta))+\hat\beta'\hat\beta=Tr(\sigma^2(X'X)^{-1})+\hat\beta'\hat\beta$，所以

$$\begin{aligned}
MSE(\hat\beta) &= Tr(\sigma^2(X'X)^{-1}) && (22)\\
&= \sigma^2 Tr((X'X)^{-1}) && (23)
\end{aligned}$$

对于满秩实对称阵$X'X$，我们记它的特征值为$\lambda_1,\dots,\lambda_p$，则它的逆矩阵的特征值为$\frac{1}{\lambda_p},\dots\frac{1}{\lambda_1}$，于是我们有

$$\begin{aligned}
Tr((X'X)^{-1}) &= \frac{\sum_i\lambda_i}{\Pi_i\lambda_i} && (24)\\
&= \frac{Tr(X'X)}{|X'X|} && (25)
\end{aligned}$$

因为y经过中心化，X经过标准化，所以我们这里的X为标准化之后的X_s，且β不含截距项，X_s的维度为$n\times p$，且$X_s'X_s$的对角元全为1，所以$Tr(X'X)=p$，则

$$MSE(\hat\beta)=\frac{\sigma^2 p}{|X'X|}$$

上面的结果同样告诉我们，当样本共线性较强时，参数估计的均方误差也会很大。

Q3

我们已经知道，岭回归估计$\hat\beta(k)$是下面这个优化问题的解：

$$\hat\beta(k)=\arg\min_\beta(y-X\beta)'(y-X\beta)+\lambda\beta'\beta$$

我们在这里讨论标准化后的X与中心化后的y，即不考虑截距项β_0，则有

$$\hat\beta(k)=\arg\min_\beta(\sum_{i=1}^n(y_i-x_i\beta)^2+\lambda\sum_{i=1}^p\beta_i^2)$$

上式是我们已有的岭回归的定义，现在我们从贝叶斯统计来解决线性回归的问题，我们采用贝叶斯估计中常见的最大后验估计的方法，即

$$\hat\beta_{bayes}=\arg\max_\beta P(\beta|y)$$

我们接下来开始计算最大化后验估计的结果，我们认为每个参数β_i服从的先验分布为正态分布，且相互独立，即$\beta_i\sim\mathcal{N}(0,\tau^2),i=1,\cdots,p$，同时假定数据分布为$y_i\sim\mathcal{N}(x_i\beta,\sigma^2),i=1,\cdots,n$

$$
\begin{aligned}
\hat{\beta}_{bayes} &= \arg\max_{\beta} P(\beta|y) && (35)\\
&= \arg\max_{\beta} P(\beta)P(y|\beta) && (36)\\
&= \arg\max_{\beta}(\ln P(\beta) + \ln P(y|\beta)) && (37)\\
&= \arg\max_{\beta}(\sum_{i=1}^{p} -\frac{\beta_i^2}{2\tau^2} - \sum_{i=1}^{n}\frac{(y_i - x_i'\beta)^2}{2\sigma^2}) && (38)\\
&= \arg\max_{\beta}(\sum_{i=1}^{p} -\frac{\sigma^2\beta_i^2}{\tau^2} - \sum_{i=1}^{n}(y_i - x_i'\beta)^2) && (39)\\
&= \arg\max_{\beta}(\frac{\sigma^2}{\tau^2}\sum_{i=1}^{p}(-\beta_i^2) - \sum_{i=1}^{n}(y_i - x_i'\beta)^2) && (40)\\
&= \arg\min_{\beta}(\frac{\sigma^2}{\tau^2}\sum_{i=1}^{p}\beta_i^2 + \sum_{i=1}^{n}(y_i - x_i'\beta)^2) && (41)\\
&= \arg\min_{\beta}(\sum_{i=1}^{n}(y_i - x_i'\beta)^2 + \frac{\sigma^2}{\tau^2}\sum_{i=1}^{p}\beta_i^2) && (42)\\
&= \arg\min_{\beta}(\sum_{i=1}^{n}(y_i - x_i\beta)^2 + \lambda\sum_{i=1}^{p}\beta_i^2), \quad let\ \lambda = \frac{\sigma^2}{\tau^2} && (43)
\end{aligned}
$$

注：

- 式(35)到式(36)的推导中，我们忽略了与β无关的分母部分
- 式(37)到式(38)的推导中，我们忽略了与β无关的若干项

这样，我们就证明了岭回归的参数估计$\hat{\beta}(k)$是贝叶斯统计下最大化后验分布的估计$\hat{\beta}_{bayes}$，从而从贝叶斯估计的角度解释了岭回归。