Now Q_1/σ^2 $(= \sum_i \varepsilon_i^2/\sigma^2)$ is χ_n^2, and $Q_2/\sigma^2 \sim \chi_p^2$ [by (ii)]. Also, Q_2 is a continuous function of $\hat{\beta}$, so that by Example 1.11 and (iii), Q is independent of Q_2. Hence $Q/\sigma^2 \sim \chi_{n-p}^2$ (Example 1.10, Section 1.6). □

EXERCISES 3d

1. Given $Y_1, Y_2, \ldots, Y_n$ independently distributed as $N(\theta, \sigma^2)$, use Theorem 3.5 to prove that:

   (a) $\overline{Y}$ is statistically independent of $Q = \sum_i (Y_i - \overline{Y})^2$. → from $\hat{\beta} \perp\!\!\!\perp S^2$ for $X = 1_n$

   (b) $Q/\sigma^2 \sim \chi_{n-1}^2$. → Thm 3.5 (iv).

2. Use Theorem 2.5 to prove that for the full-rank regression model, RSS is independent of $\underbrace{(\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)}_{\|X(\hat\beta - \beta)\|_2^2}$.

$RSS = \|y - X\hat\beta\|_2^2$

NTS indep w/ Thm 2.5. $X(\hat\beta - \beta)$ normal b/ $\hat\beta$ is. $y - X\hat\beta = \varepsilon - (X(\hat\beta - \beta))$ normal. Clear they're joint Gaussian too.

3.5 MAXIMUM LIKELIHOOD ESTIMATION

Assuming normality, as in Section 3.4, the likelihood function, $L(\beta, \sigma^2)$ say, for the full-rank regression model is the probability density function of $\mathbf{Y}$, namely,

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\beta)||^2\right\}.$$

Let $l(\beta, v) = \log L(\beta, \sigma^2)$, where $v = \sigma^2$. Then, ignoring constants, we have

$$l(\beta, v) = -\frac{n}{2}\log v - \frac{1}{2v}||\mathbf{y} - \mathbf{X}\beta||^2,$$

and from (3.6) it follows that

$$\frac{\partial l}{\partial \beta} = -\frac{1}{2v}(-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta)$$

Then, $y - X\hat\beta = (I - P)y = (I-P)(y - X\beta) = (I - P)\varepsilon$

and

$$\frac{\partial l}{\partial v} = -\frac{n}{2v} + \frac{1}{2v^2}||\mathbf{y} - \mathbf{X}\beta||^2.$$

Setting $\partial l/\partial \beta = \mathbf{0}$, we get the least squares estimate of β, which clearly maximizes $l(\beta, v)$ for any $v > 0$. Hence

$$L(\beta, v) \leq L(\hat{\beta}, v) \quad \text{for all } v > 0$$

with equality if and only if $\beta = \hat{\beta}$.

We now wish to maximize $L(\hat{\beta}, v)$, or equivalently $l(\hat{\beta}, v)$, with respect to v. Setting $\partial l/\partial v = 0$, we get a stationary value of $\hat{v} = ||\mathbf{y} - \mathbf{X}\hat{\beta})||^2/n$. Then

$$\begin{aligned} l(\hat{\beta}, \hat{v}) - l(\hat{\beta}, v) &= -\frac{n}{2}\left[\log\left(\frac{\hat{v}}{v}\right) + 1 - \frac{\hat{v}}{v}\right] \\ &\geq 0, \end{aligned}$$

Since $X(\hat\beta - \beta) \in C(X)$,

$X(\hat\beta - \beta) = P(X(\hat\beta - \beta))$
$= X\hat\beta - X(X^TX)^{-1}X^TX\beta$
$= X(X^TX)^{-1}X^Ty - \ldots$
$= P(y - X\beta)$
$= P\varepsilon$

$(I-P)P = 0$