STA 6351, Report.1.4

Carson Slater *Baylor University*

1.4

GLM standard error. Prove the following:

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\beta})^\top (\mathbf{y} - \mathbf{X}\hat{\beta}).$$

Consider the normal-theory linear model

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n),$$

where

- $\mathbf{Y} \in \mathbb{R}^n$ is the response vector and $\mathbf{X}$ is a known $n \times p$ design matrix of full rank,
- $\beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$ are regression coefficients,
- For $\sigma \in \mathbb{R}^+$, σ^2 is the error variance,
- I_n is the $n \times n$ identity matrix.

The distribution of $\mathbf{Y}$ is

$$\mathbf{Y} \mid \beta, \sigma^2 \sim \mathcal{N}_n(\mathbf{X}\beta, \sigma^2 I_n),$$

with joint density

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

The parameter space is $\Omega \equiv \mathbb{R}^p \times \mathbb{R}^+$.

The likelihood is, of course, $f(y \mid \beta, \sigma^2)$ treated as a function of its parameters given an observed response, $\mathbf{Y} = y$: $L(\beta, \sigma^2 \mid y)$. The log-likelihood is

$$\log L(\beta, \sigma^2 \mid y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - \mathbf{X}\beta)^\top (y - \mathbf{X}\beta).$$

Differentiating with respect to β gives

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2}\mathbf{X}^\top (y - \mathbf{X}\beta).$$

Setting this equal to zero, and noting that $\mathbf{X}$ has full rank, we obtain

$$\mathbf{X}^\top \mathbf{X}\hat{\beta} = \mathbf{X}^\top y \quad \Rightarrow \quad \hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top y.$$

Differentiating with respect to σ^2 yields

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - \mathbf{X}\beta)^\top (y - \mathbf{X}\beta).$$

Setting this equal to zero gives

$$-n\sigma^2 + (y - \mathbf{X}\beta)^\top (y - \mathbf{X}\beta) = 0.$$

Thus, the maximum likelihood estimator of the error variance is[1]

$$\hat{\sigma}^2 = \frac{1}{n}(y - \mathbf{X}\hat{\beta})^\top (y - \mathbf{X}\hat{\beta}).$$

Note: the MLE divides by n, whereas the usual unbiased estimator divides by $n - p$, since

$$\mathbb{E}[(y - \mathbf{X}\hat{\beta})^\top (y - \mathbf{X}\hat{\beta})] = (n - p)\sigma^2.$$

[1]See the next page for concavity.

Lemma 1. *Let $\ell(\beta, \sigma^2)$ be the log-likelihood of the normal linear model*

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n),$$

with full-rank design matrix $\mathbf{X}$. Then the Hessian of ℓ evaluated at the MLE $(\hat{\beta}, \hat{\sigma}^2)$ is negative definite, so $(\hat{\beta}, \hat{\sigma}^2)$ is indeed a maximum.

Proof. For fixed σ^2, the log-likelihood as a function of β is

$$\ell(\beta \mid \sigma^2) = -\frac{1}{2\sigma^2}(y - \mathbf{X}\beta)^\top (y - \mathbf{X}\beta) + \text{const.}$$

Its Hessian is

$$\frac{\partial^2 \ell}{\partial \beta \, \partial \beta^\top} = -\frac{1}{\sigma^2}\mathbf{X}^\top \mathbf{X}.$$

Since $\mathbf{X}$ has full column rank, $\mathbf{X}^\top \mathbf{X}$ is positive definite, and hence the Hessian is negative definite in β. This shows that $\ell(\beta \mid \sigma^2)$ is concave in β.

For $\sigma^2 = u > 0$, the log-likelihood profile is

$$\ell(u \mid \hat{\beta}) = -\frac{n}{2}\log u - \frac{1}{2u}(y - \mathbf{X}\hat{\beta})^\top (y - \mathbf{X}\hat{\beta}) + \text{const.}$$

Differentiating twice with respect to u gives

$$\frac{\partial^2 \ell}{\partial u^2} = \frac{n}{2u^2} - \frac{(y - \mathbf{X}\hat{\beta})^\top (y - \mathbf{X}\hat{\beta})}{u^3}.$$

At the critical point $u = \hat{\sigma}^2 = \frac{1}{n}(y - \mathbf{X}\hat{\beta})^\top (y - \mathbf{X}\hat{\beta})$, this reduces to

$$\left.\frac{\partial^2 \ell}{\partial u^2}\right|_{u=\hat{\sigma}^2} = -\frac{n}{2(\hat{\sigma}^2)^2} < 0.$$

Thus, the log-likelihood is concave in both β and σ^2 at the MLE $(\hat{\beta}, \hat{\sigma}^2)$, confirming it is indeed a maximum. □