

Homework 5

Aiden Kenny

STAT GR5205: Linear Regression Models

Columbia University

November 25, 2020

Question 1

Collaborators: None

- (a) Let Y be the number of nurses in the hospital and let X be the available faculty and services.
- (b) The power transformation function is defined as

$$g_\lambda(Y) = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log Y & \text{if } \lambda = 0. \end{cases}$$

Suppose we have our response vector \mathbf{y} and our observed data \mathbf{X} . We are interested in fitting the model $\mathbf{g}_\lambda(\mathbf{y}) = \mathbf{X}\beta + \epsilon$, where $\mathbf{g}_\lambda(\mathbf{y})$ is the transformed response vector (i.e. the i th element is given by $g_\lambda(y_i)$), $\mathbb{E}[\epsilon] = \mathbf{0}$, and $\text{Var}[\epsilon] = \sigma^2 \mathbf{I}$. For notational ease, we will denote $\mathbf{g}_\lambda(\mathbf{y})$ as \mathbf{g}_λ . If we make the further assumption that ϵ is normally distributed, i.e. $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, then our response vector is also normally distributed, where $\mathbf{g}_\lambda \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I})$. It's density function (and thus it's likelihood function) is given by

$$\begin{aligned} f(\mathbf{g}_\lambda | \beta, \sigma^2, \lambda) &= \frac{1}{\sqrt{\det(2\pi\sigma^2 \mathbf{I})}} \cdot \exp\left(-\frac{(\mathbf{g}_\lambda - \mathbf{X}\beta)^T (\sigma^2 \mathbf{I})^{-1} (\mathbf{g}_\lambda - \mathbf{X}\beta)}{2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(-\frac{(\mathbf{g}_\lambda - \mathbf{X}\beta)^T (\mathbf{g}_\lambda - \mathbf{X}\beta)}{2\sigma^2}\right). \end{aligned}$$

Since \mathbf{y} is a transformation of \mathbf{g}_λ , we can derive the density for \mathbf{y} as well. Notationally, this result may be somewhat confusing; even though we are finding the density for \mathbf{y} , we will still express the density (partly) in terms of \mathbf{g}_λ . It is important to remember that \mathbf{g}_λ is a function of \mathbf{y} . Because the i th element of \mathbf{g}_λ only depends on the i th element of \mathbf{y} , the Jacobian will be a diagonal matrix, and so

$$\mathbf{J} = \frac{\partial \mathbf{g}_\lambda}{\partial \mathbf{y}} = \text{diag}\left(\frac{\partial g_\lambda(y_1)}{\partial y_1}, \dots, \frac{\partial g_\lambda(y_n)}{\partial y_n}\right) = \text{diag}(y_1^{\lambda-1}, \dots, y_n^{\lambda-1}),$$

and so the density (and thus the likelihood) of \mathbf{y} is given by

$$g(\mathbf{y} | \beta, \sigma^2, \lambda) = f(\mathbf{g}_\lambda(\mathbf{y})) \cdot |\det(\mathbf{J})| = \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(-\frac{(\mathbf{g}_\lambda - \mathbf{X}\beta)^T (\mathbf{g}_\lambda - \mathbf{X}\beta)}{2\sigma^2}\right) \cdot \prod_{i=1}^n y_i^{\lambda-1}.$$

The log-likelihood $\ell(\mathbf{y}) = \log g(\mathbf{y})$ is given by

$$\ell(\mathbf{y} | \beta, \sigma^2, \lambda) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{(\mathbf{g}_\lambda - \mathbf{X}\beta)^T (\mathbf{g}_\lambda - \mathbf{X}\beta)}{2\sigma^2} + (\lambda - 1) \sum_{i=1}^n \log(y_i).$$

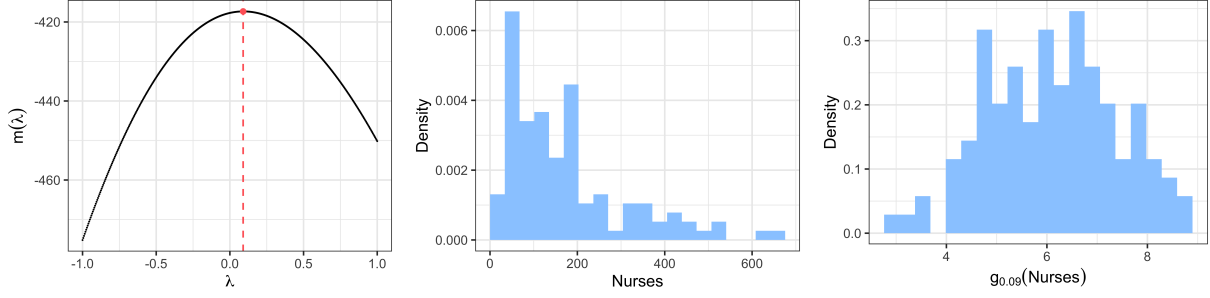


Figure 1: Relevant plots for the Box-Cox transformation of Y .

As is standard with maximum likelihood estimation, we now differentiate ℓ with respect to the unknown parameters, set the derivatives to zero, and solve to get the maximum value of ℓ . For now, we are going to leave λ fixed and differentiate with respect to β and σ^2 . Doing this for both gives us $\hat{\beta}_{\text{MLE}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{g}_\lambda$ and $\hat{\sigma}_{\text{MLE}}^2 = \mathbf{g}_\lambda^T (\mathbf{I} - \mathbf{H}) \mathbf{g}_\lambda$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ is the hat matrix. It is worth noting that both $\hat{\beta}_{\text{MLE}}$ and $\hat{\sigma}_{\text{MLE}}^2$ are functions of λ . Plugging these values back into ℓ will maximize it with respect to β and σ^2 , which means we will only have to maximize it with respect to λ . With some simplification, our new loss function is

$$m(\lambda) := \ell(\mathbf{y} | \hat{\beta}_{\text{MLE}}, \hat{\sigma}_{\text{MLE}}^2, \lambda) = -\frac{n}{2} \log \left(\frac{2\pi e}{n} \right) - \frac{n}{2} \log (\mathbf{g}_\lambda^T (\mathbf{I} - \mathbf{H}) \mathbf{g}_\lambda) + (\lambda - 1) \sum_{i=1}^n \log(y_i).$$

Ideally, we would differentiate m with respect to λ , set $\partial m / \partial \lambda = 0$, and solve for λ . Unfortunately, I was unable to derive a closed form solution for the result. However, it is still possible to use graphical techniques or numerical methods to find the optimal value of λ .

The left panel of Figure 1 shows a plot of $m(\lambda)$ against λ for values $\lambda \in [-1, 1]$. The log-likelihood is maximized when $\hat{\lambda} = 0.09$. The middle panel shows a histogram of the values of Y , which we can see is heavily skewed to the right. The right panel shows a histogram of $g_{0.09}(Y)$, and we can see that after the transformation is applied, the data is much more normally distributed.