Coursework 2: STAT 570

Philip Pham

October 10, 2018

1. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, \ldots, n,$$

where the error terms ϵ_i are such that $\mathbb{E}\left[\epsilon_i\right] = 0$, $\operatorname{Var}\left(\epsilon_i\right) = \sigma^2$, and $\operatorname{Cov}\left(\epsilon_i, \epsilon_j\right) = 0$ for $i \neq j$.

In the following you will consider $x_i \sim_{\text{iid}} \mathcal{N}\left(20, 3^2\right)$, with $\beta_0 = 2$ and $\beta_1 = -2.5$ and $n = 15, 30$.

Consider the model in Equation 1 with the error terms ϵ_i, independent and identically distributed, from the distributions:

- The normal distribution with mean 0 and variance 2^2.
- The uniform distribution on the range $(-r, r)$ for $r = 2$.
- A skew normal distribution with $\alpha = 5$, $\omega = 1$, and ξ chosen to given mean 0.

(a) What is the theoretical bias for $\hat{\beta}$ if the errors are of the form specified?

Solution: The theoretical bias for $\hat{\beta}$ is 0. Let

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \tag{1}$$

If we use the least squares estimate, we have

$$\begin{aligned} \hat{\beta} &= \left(X^\intercal X\right)^{-1} X^\intercal y \\ &= \left(X^\intercal X\right)^{-1} X^\intercal \left(X\beta + \epsilon\right) \\ &= \beta + \left(X^\intercal X\right)^{-1} X^\intercal \epsilon, \end{aligned} \tag{2}$$

Thus, using Equation 2 and linearity of expectations, we have

$$\boxed{\operatorname{bias}\left(\hat{\beta}\right) = \mathbb{E}\left[\hat{\beta}\right] - \beta = \beta + \left(X^\intercal X\right)^{-1} X^\intercal \mathbb{E}\left[\epsilon\right] - \beta = 0.} \tag{3}$$

(b) Compare the variance of the estimator as reported by least squares, with that which follows from the sampling distribution of the estimator.

Solution:

(c) Examine the distribution of the resultant estimators (across simulations) of β_0 and β_1, in particular with respect to normality. For each parameter find the coverage probability of a 95% confidence interval, that is the proportion of times that the confidence intervals contain the true value.

Solution:

(d) **Bonus:** Can you "break" least squares? i.e., find a distribution of the errors (with mean zero) that produces poor confidence interval coverage?

Solution:

2. Consider the exponential regression problem with independent responses

$$p\left(y_i \mid \lambda_i\right) = \lambda_i \exp\left(-\lambda_i y_i\right), y_i > 0, \tag{4}$$

and $\log \lambda_i = \beta_0 + \beta_1 x_i$ for given covariates x_i, $i = 1, \ldots, n$. We wish to estimate the 2×1 regression parameter $\beta = \begin{pmatrix}\beta_0 & \beta_1\end{pmatrix}^\intercal$ using maximum likelihood estimation (MLE).

(a) Find expressions for the likelihood function $L\left(\beta\right)$, log-likelihood function $l\left(\beta\right)$, score function $S\left(\beta\right)$, and Fisher's information matrix $I\left(\beta\right)$.

Solution: We can rewrite Equation 4 in terms of β, which gives us

$$\begin{aligned} p\left(y_i \mid \beta_0, \beta_1\right) &= \exp\left(\beta_0 + \beta_1 x_i\right) \exp\left(-y_i \exp\left(\beta_0 + \beta_1 x_i\right)\right) \\ &= \exp\left(\beta_0 + \beta_1 x_i - y_i \exp\left(\beta_0 + \beta_1 x_i\right)\right). \end{aligned} \tag{5}$$

Using Equation 5, we can write the likelihood function

$$\begin{aligned} L\left(\beta\right) &= \prod_{i=1}^{n} p\left(y_i \mid \beta_0, \beta_1\right) \\ &= \exp\left(n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \exp\left(\beta_0 + \beta_1 x_i\right)\right). \end{aligned} \tag{6}$$

Taking the log of Equation 6, we have the log-likelihood function as

$$\begin{aligned} l\left(\beta\right) &= \log L\left(\beta\right) \\ &= n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \exp\left(\beta_0 + \beta_1 x_i\right). \end{aligned} \tag{7}$$

Taking the gradient of Equation 7, we have the score function

$$\begin{aligned} S\left(\beta\right) &= \nabla l\left(\beta\right) \\ &= \begin{pmatrix} n - \sum_{i=1}^{n} y_i \exp\left(\beta_0 + \beta_1 x_i\right) \\ \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i \exp\left(\beta_0 + \beta_1 x_i\right) \end{pmatrix}. \end{aligned} \tag{8}$$

One definition of the Fisher information is the expected value of the observed information which is the negative of the second derivative of the log-likelihood function. For a single observation,

$$\begin{aligned}\mathcal{I}_1\left(\beta\right) &= \mathbb{E}\left[\begin{pmatrix} Y\exp\left(\beta_0+\beta_1 x_i\right) & x_i Y\exp\left(\beta_0+\beta_1 x_i\right) \\ x_i Y\exp\left(\beta_0+\beta_1 x_i\right) & x_i^2 Y\exp\left(\beta_0+\beta_1 x_i\right)\end{pmatrix} \mid X = x_i\right] \\ &= \frac{1}{\exp\left(\beta_0+\beta_1 x_i\right)}\begin{pmatrix} \exp\left(\beta_0+\beta_1 x_i\right) & x_i\exp\left(\beta_0+\beta_1 x_i\right) \\ x_i\exp\left(\beta_0+\beta_1 x_i\right) & x_i^2\exp\left(\beta_0+\beta_1 x_i\right)\end{pmatrix} \\ &= \begin{pmatrix} 1 & x_i \\ x_i & x_i^2\end{pmatrix} \qquad (9)\end{aligned}$$

by properties of the exponential distribution. Thus, Fisher information is

$$\mathcal{I}_n\left(\beta\right) = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2\end{pmatrix}. \qquad (10)$$

(b) Find expressions for the maximum likelihood estimate $\hat{\beta}$. If no closed form solution exists, then instead provide a functional form that could be simply implemented for solution.

Solution: We can solve for $\hat{\beta}_0$ in terms of $\hat{\beta}_1$. We know that $S\left(\hat{\beta}\right) = \mathbf{0}$.

From Equation 8, we can solve for $\hat{\beta}_0$,

$$\hat{\beta}_0 = \log n - \log\sum_{i=1}^n y_i\exp\left(\hat{\beta}_1 x_i\right). \qquad (11)$$

Substituing Equation 11 into the second entry of Equation 8, we have

$$\begin{aligned} 0 &= \sum_{i=1}^n x_i - \exp\left(\hat{\beta}_0\right)\sum_{i=1}^n x_i y_i\exp\left(\hat{\beta}_1 x_i\right) \\ &= \sum_{i=1}^n x_i - \frac{n}{\sum_{i=1}^n y_i\exp\left(\hat{\beta}_1 x_i\right)}\sum_{i=1}^n x_i y_i\exp\left(\hat{\beta}_1 x_i\right), \qquad (12)\end{aligned}$$

which we can solve numerically with a root-finding algorithm.

(c) For the data in Table 1, numerically maximize the likelihood function to obtain estimates of β. These data consist of the survival times (y) of rats as function of concentration of a contaminant (x). Find the asymptotic covariance matrix for your estimate using the information $\mathcal{I}\left(\beta\right)$. Provide a 95% confidence interval for each element of β_0 and β_1.

Solution: Numerically solving Equations 11 and 12, we have that

$$\hat{\beta} = \begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{pmatrix} = \begin{pmatrix} -2.821150253077923 \\ 0.30133576292327585\end{pmatrix}. \qquad (13)$$

The Fisher information gives a lower bound on the variance according to the Cramér-Rao bound. Asymptotic normality of the MLE tells us that

$$\hat{\beta}_n - \beta \to \mathcal{N}\left(0, \mathcal{I}_n^{-1}\left(\beta\right)\right)$$

in distribution.

i	x_i	y_i
1	6.1	0.8
2	4.2	3.5
3	0.5	12.4
4	8.8	1.1
5	1.5	8.9
6	9.2	2.4
7	8.5	0.1
8	8.7	0.4
9	6.7	3.5
10	6.5	8.3
11	6.3	2.6
12	6.7	1.5
13	0.2	16.6
14	8.7	0.1
15	7.5	1.3

Table 1: Each observation is a rat. x_i are the concentrations of the contaminant, and y_i are the survival times.

Thus, we have the covariance matrix

$$\operatorname{Var}\left(\hat{\beta}\right) \approx \begin{pmatrix} 15 & 90.1 \\ 90.1 & 671.07 \end{pmatrix}^{-1} = \begin{pmatrix} 0.34448471 & -0.04625162 \\ -0.04625162 & 0.00770005 \end{pmatrix}. \tag{14}$$

Using this we can approximate 95% confidence intervals as $\hat{\beta}_j \pm z_{0.975}\sqrt{\operatorname{Var}\left(\hat{\beta}_j\right)}$, where $z_p = \Phi^{-1}(p)$ and Φ is the cumulative distribution function of the normal distribution.

We have the confidence intervals

$$\begin{aligned} \left(\hat{\beta}_0 - 1.150358, \hat{\beta}_0 + 1.150358\right) &= (-3.97150839, -1.67079212) \\ \left(\hat{\beta}_1 - 0.171986669, \hat{\beta}_1 + 0.171986669\right) &= (0.12934909, 0.47332243) \end{aligned} \tag{15}$$

for $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively.

All calcuations can be found in `exponential_regression.ipynb`.

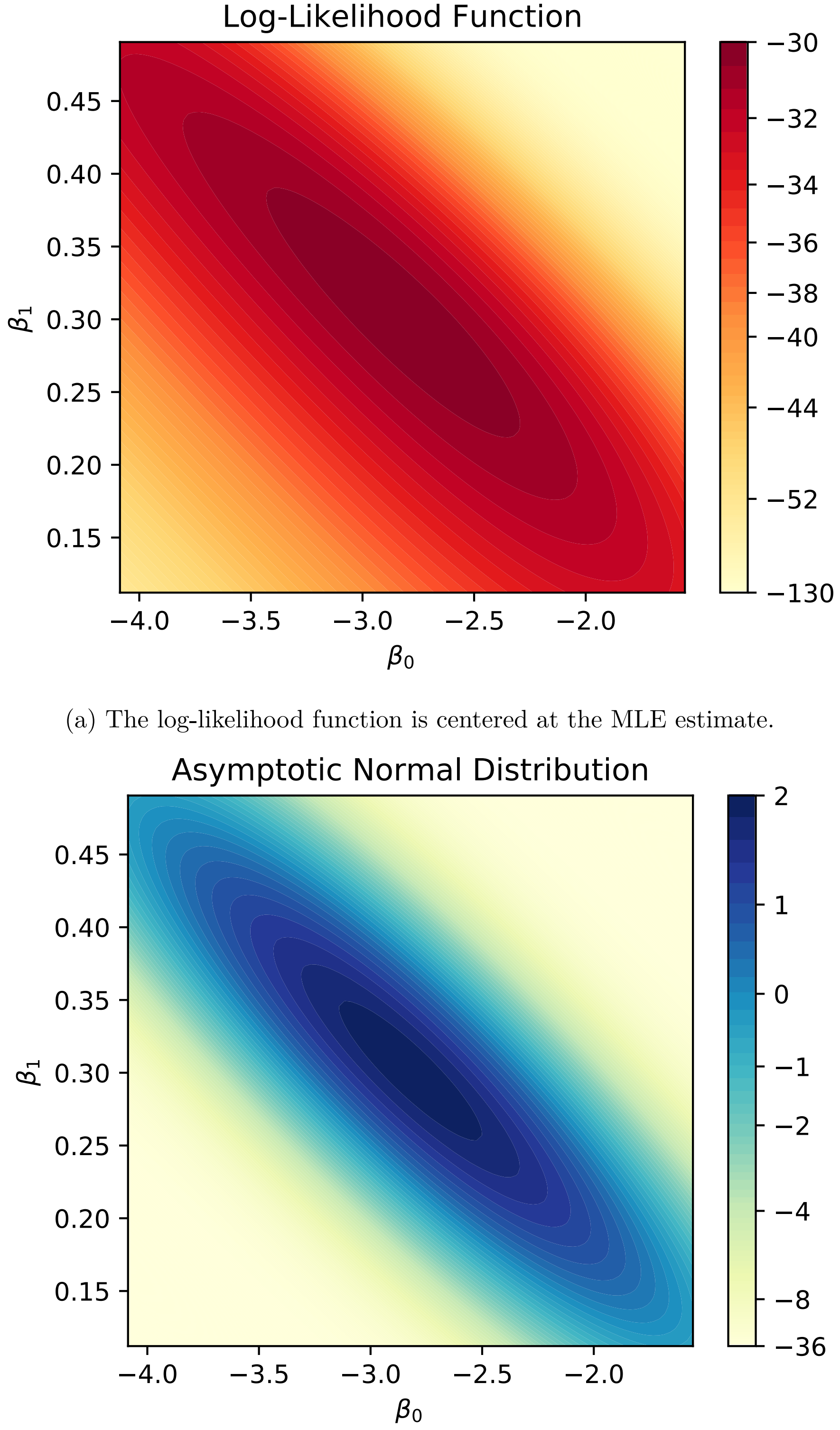
(d) Plot the log-likelihood function $l(\beta_0, \beta_1)$ and compare with the log of the asymptotic normal approximation to the sampling distribution of the MLE.

Solution: The two plots can be found in Figure 1.

The log-likelihood function can be found in Figure 1a, and the asymptotic normal approximation is plotted in Figure 1b.

The distributions are similar, and the asymptotic normal approximation accurately captures the covariance structure of $\hat{\beta}$. The asymptotic normal approximation, however, does underestimate the variance. This is unsurprising since the inverse of the Fisher information matrix is a lower bound on variance according to the Cramér-Rao bound.

Code for the plots can be found in `exponential_regression.ipynb`.



(a) The log-likelihood function is centered at the MLE estimate.

(b) The asymptotic normal approximation mirrors the log-likelihood with slightly less variance.

Figure 1: Plots of the distributions of $\hat{\beta}$ for Part 2d.

(e) Summarize the results of the estimation presented above in a manner that would address the question of whether increasing concentrations of the contaminant had an effect on a rat's life expectancy.

Solution: λ_i can be viewed as the rate of death, so the mean survival time is λ_i^{-1}. $\log \lambda_i$ has a linear relationship with concentrations of the contaminant described by β_1. $\beta_1 > 0$ indicates that increasing concentrations of the contaminant increase death rates, and therefore, decrease life expectancy. Indeed from Equation 13, our estimate of β_1, $\hat{\beta}_1 = 0.30133576292327585 > 0$. From Equation 15, we see that the 95% confidence interval lies entirely on the positive half-line, so the result is statistically significant at level 0.05. Thus, it is quite likely that increasing concentrations of the contaminant decrease life expectancy.