

Stat 8003, Homework 4

Group G: `sample(c("David" , "Andrew", "Salam"))`

September 25, 2014

Question 4.1. Suppose X is a discrete random variables with $P(X = 1) = \theta$ and $P(X = 2) = 1 - \theta$. Three independent observations of X are made: $x_1 = 1, x_2 = 2, x_3 = 2$.

- (a) Find the method of moments estimate of θ ;
- (b) What is the likelihood function?
- (c) What is the MLE of θ ?:

Answer: X is Bernoulli with parameter θ :

$$X = \begin{cases} 1 & \text{with probability } \theta \\ 2 & \text{with probability } 1 - \theta \end{cases}$$

We write its pdf compactly as:

$$f(x | \theta) = \theta^{2-x}(1 - \theta)^{x-1}$$

(This function evaluates to θ when $X = 1$ and to $1 - \theta$ when $X = 2$. The likelihood function is the product of pdfs viewed as a function of θ for the observed data: $L(\theta; \tilde{x}) = \prod_i f(x_i | \theta)$. Here the data consist of three independent observations: $x_1 = 1, x_2 = 2, x_3 = 2$.)

- (a) The method of moments (MOM) estimate of θ :

$$m_1 = E(X) = (1)(\theta) + (2)(1 - \theta) = \theta + 2 - 2\theta = 2 - \theta$$

So,

$$\hat{\theta} = 2 - m_1$$

From our data,

$$m_1 = \bar{x} = \frac{1 + 2 + 2}{3} = \frac{5}{3}$$

so our method of moment estimate for $\hat{\theta}$ is

$$\hat{\theta} = 2 - \frac{5}{3} = \frac{1}{3}$$

(b) The likelihood function for this data is:

$$\begin{aligned} L(\theta; x) &= \theta^{2-1}(1-\theta)^{1-1}\theta^{2-2}(1-\theta)^{x-2}\theta^{2-2}(1-\theta)^{x-2} \\ &= \theta(1-\theta)^2 \end{aligned}$$

(c) To get the maximum likelihood estimate (MLE) for θ , we can differentiate the likelihood function with respect to θ , set it to zero, and solve for θ . Alternatively, we can take the log of the likelihood function and set that to zero to solve for θ :

$$\frac{d}{d\theta} \log(\theta(1-\theta)^2) = \frac{d}{d\theta} (\log \theta + 2 \log(1-\theta)) = 0$$

Or

$$\frac{1}{\theta} = \frac{2}{1-\theta}$$

$$1 - \theta = 2\theta$$

From where we get:

$$\boxed{\hat{\theta} = \frac{1}{3}}$$

We now check whether this extremum is indeed a *maximum*. We take the second derivative of $L(\theta; x)$, evaluate it at $\frac{1}{3}$ and see if it's negative. Indeed:

$$\begin{aligned} \frac{d^2}{d\theta^2} L(\theta; x) &= \frac{d}{d\theta} \left(\frac{1}{\theta} + \frac{2}{1-\theta} \right) \\ &= -\frac{1}{\theta^2} + \frac{2}{(1-\theta)^2} \\ &= -\frac{1}{(\frac{1}{3})^2} + \frac{2}{(1-\frac{1}{3})^2} \\ &= -9 + \frac{9}{2} \\ &< 0 \end{aligned}$$

So we indeed have a *maximum* at $\hat{\theta} = \frac{1}{3}$.

Question 4.2. Consider an i.i.d. sample of random variables with density function

$$f(x | \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right)$$

- (a) Find the MOM estimate of σ ;
- (b) Find the MLE estimate of σ ;

Answer:

- (a)

$$m_1 = E(X) = \int_{-\infty}^{+\infty} \frac{x}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx$$

This is an integral of an *odd* function, which should evaluate to zero:

$$\begin{aligned} &= \int_{-\infty}^0 \frac{x}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx + \int_0^{+\infty} \frac{x}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx \\ &= - \int_0^{+\infty} \frac{x}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx + \int_0^{+\infty} \frac{x}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx \\ &= 0 \end{aligned}$$

This isn't any help to us. So let us consider m_2 .

$$m_2 = E(X^2) = \int_{-\infty}^{+\infty} \frac{x^2}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx$$

$$= 2 \int_0^{+\infty} \frac{x^2}{2\sigma} \exp\left(-\frac{x}{\sigma}\right) dx$$

$$= \frac{1}{\sigma} \int_0^{+\infty} x^2 \exp\left(-\frac{x}{\sigma}\right) dx$$

Substitute ($y = \frac{x}{\sigma}$)

$$= \sigma^2 \int_0^{+\infty} y^2 \exp(-y) dy$$

We could recognize this as the Gamma function and write the result.

Or continue:

$$\begin{aligned}
&= \sigma^2 \left(-y^2 e^{-y} \Big|_0^{+\infty} + 2 \int_0^{+\infty} y \exp(-y) dy \right) \\
&= \sigma^2 \left(0 + (-2y e^{-y}) \Big|_0^{+\infty} + 2 \int_0^{+\infty} \exp(-y) dy \right) \\
&= \sigma^2 \left(0 + 0 - 2 \exp(-y) \Big|_0^{+\infty} \right) = 2\sigma^2
\end{aligned}$$

So $m_2 = 2\sigma^2$.

Therefore

$$\hat{\sigma} = \sqrt{\frac{m_2}{2}}$$

(b) The likelihood function is given by

$$L(\sigma; X) = \prod_{i=1}^n \frac{1}{2\sigma} \exp\left(-\frac{|x_i|}{\sigma}\right)$$

Taking the *log* of this gives

$$l(\sigma; X) = -n \log(2\sigma) + \sum_{i=1}^n \frac{-|x_i|}{\sigma}$$

Then taking the derivative with respect to σ gives

$$\frac{dl}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n |x_i|$$

Setting this equal to zero we get

$$\frac{n}{\sigma} = \frac{1}{\sigma^2} \sum_{i=1}^n |x_i|$$

i.e.,

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n |x_i|$$

Question 4.3. In the shuttle example, let X_i denote the number of damaged o-rings and t_i the temperature, where $i = 1, 2, \dots, n$. Assume the model as

$$\begin{cases} X_i | p_i \sim \text{Binom}(2, p_i) \\ p_i = e^{(\beta_0 + \beta_1 t_i)} / (1 + e^{(\beta_0 + \beta_1 t_i)}) \end{cases}$$

(a) Derive the log-likelihood function;

Since

$$\begin{aligned} f(x_i | p_i) &= \binom{2}{x_i} \left(\frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right)^{x_i} \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 t_i)} \right)^{2-x_i} \\ &= \binom{2}{x_i} \frac{\exp(\beta_0 + \beta_1 t_i)^{x_i}}{(1 + \exp(\beta_0 + \beta_1 t_i))^2} \end{aligned}$$

The likelihood is

$$L(\beta_0, \beta_1; \tilde{x}, \tilde{t}) = \prod_{i=1}^n \binom{2}{x_i} \frac{\exp(\beta_0 + \beta_1 t_i)^{x_i}}{(1 + \exp(\beta_0 + \beta_1 t_i))^2}$$

And thus the log likelihood is:

$$\begin{aligned} l(\beta_0, \beta_1; \tilde{x}, \tilde{t}) &= \log \prod_{i=1}^n \binom{2}{x_i} + \sum_{i=1}^n x_i (\beta_0 + \beta_1 t_i) - 2 \sum_{i=1}^n \log(1 + \exp(\beta_0 + \beta_1 t_i)) \\ &= \text{Some Constant} + \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i t_i - 2 \sum_{i=1}^n \log(1 + \exp(\beta_0 + \beta_1 t_i)) \end{aligned}$$

(b) Set the equations for the maximum likelihood estimator of β_0, β_1 Since

$$\begin{cases} \frac{\partial l}{\partial \beta_0} = \sum_{i=1}^n x_i - 2 \sum_{i=1}^n \frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \\ \frac{\partial l}{\partial \beta_1} = \sum_{i=1}^n x_i t_i - 2 \sum_{i=1}^n \frac{t_i \exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \end{cases}$$

We set these partial derivatives to 0 and solve them for β_0 and β_1 :

$$\begin{cases} \sum_{i=1}^n x_i - 2 \sum_{i=1}^n \frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} = 0 \\ \sum_{i=1}^n x_i t_i - 2 \sum_{i=1}^n \frac{t_i \exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} = 0 \end{cases}$$

(c) Derive the steps for the Newton-Raphson algorithm...

In previous equations, we denote the first one by $f_1(\beta_0, \beta_1)$, the second with $f_2(\beta_0, \beta_1)$ and then apply Newton-Raphson to following two-variable problem:

$$\begin{cases} f_1(\beta_0, \beta_1) = 0 \\ f_2(\beta_0, \beta_1) = 0 \end{cases}$$

The Jacobian matrix is:

$$\begin{aligned}
J &= \begin{pmatrix} \frac{\partial f_1(\beta_0, \beta_1)}{\partial \beta_0} & \frac{\partial f_1(\beta_0, \beta_1)}{\partial \beta_1} \\ \frac{\partial f_2(\beta_0, \beta_1)}{\partial \beta_0} & \frac{\partial f_2(\beta_0, \beta_1)}{\partial \beta_1} \end{pmatrix} = \begin{pmatrix} -2 \frac{\partial}{\partial \beta_0} \left(\sum_{i=1}^n \frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right) & -2 \frac{\partial}{\partial \beta_1} \left(\sum_{i=1}^n \frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right) \\ -2 \frac{\partial}{\partial \beta_0} \left(\sum_{i=1}^n \frac{t_i \exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right) & -2 \frac{\partial}{\partial \beta_1} \left(\sum_{i=1}^n \frac{t_i \exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right) \end{pmatrix} \\
&= \\
&= -2 \sum_i \begin{pmatrix} \frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} - \left(\frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right)^2 & t_i \left(\frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} - \left(\frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right)^2 \right) \\ t_i \left(\frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} - \left(\frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right)^2 \right) & t_i^2 \left(\frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} - \left(\frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)} \right)^2 \right) \end{pmatrix} \\
&= -2 \sum_i \begin{pmatrix} \frac{\exp(\beta_0 + \beta_1 t_i)}{(1 + \exp(\beta_0 + \beta_1 t_i))^2} & t_i \frac{\exp(\beta_0 + \beta_1 t_i)}{(1 + \exp(\beta_0 + \beta_1 t_i))^2} \\ t_i \frac{\exp(\beta_0 + \beta_1 t_i)}{(1 + \exp(\beta_0 + \beta_1 t_i))^2} & t_i^2 \frac{\exp(\beta_0 + \beta_1 t_i)}{(1 + \exp(\beta_0 + \beta_1 t_i))^2} \end{pmatrix}
\end{aligned}$$

We could invert this Jacobian and use the following Newton-Raphson update rule to find β_0 and β_1 :

$$\begin{pmatrix} \beta_0^{i+1} \\ \beta_1^{i+1} \end{pmatrix} = \begin{pmatrix} \beta_0^i \\ \beta_1^i \end{pmatrix} - J^{-1} \begin{pmatrix} f_1(\beta_0^i, \beta_1^i) \\ f_2(\beta_0^i, \beta_1^i) \end{pmatrix}$$

Before implementing this in R, and since these expressions look complicated, we try to simplify life for ourselves by making the following observations.

We can simplify these formulas, and their R implementation, if we note that the model given in this problem is equivalent to the following:

$$\log \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 t_i$$

We will denote $\log \left(\frac{p_i}{1 - p_i} \right) = y_i$ and write

$$y_i = \beta_0 + \beta_1 t_i$$

Once we have estimates for β_0 and β_1 , we can then recover p_i from y_i .

(We note that we could use R's linear regression model to get the least squares estimates of β_0 and β_1 . However, we will estimate these using Newton-Raphson instead.)

First,

$$L(\beta_0, \beta_1; \tilde{t}) = \prod_{i=1}^n (\beta_0 + \beta_1 t_i)$$

and

$$l(\beta_0, \beta_1; \tilde{t}) = \sum_{i=1}^n \log(\beta_0 + \beta_1 t_i)$$

And we set the partial derivatives with respect to β_0 and β_1 to zero:

$$\begin{cases} \frac{\partial l}{\partial \beta_0} = \sum_{i=1}^n \frac{1}{\beta_0 + \beta_1 t_i} = 0 \\ \frac{\partial l}{\partial \beta_1} = \sum_{i=1}^n \frac{t_i}{\beta_0 + \beta_1 t_i} = 0 \end{cases}$$

We then apply Newton-Raphson to these guys.

- (d) Use the Newton-Raphson algorithm to calculate the maximum likelihood estimator of β_0 and β_1

The estimates for the parameters are:

$$\begin{aligned} \beta_0 &= 15.0429 \\ \beta_1 &= -0.2322 \end{aligned}$$

- (e) On January 28, 1986, the outside temperature was 31 degrees. Based on your estimated β_0 and β_1 , what is the probability of an o-ring failure?

$$y = 15.0429 - 0.2322 * 31 = 7.8447$$

$$p = \frac{e^y}{1 + e^y}$$

$$\boxed{p = 0.9996083}$$

- (f) Based on your estimator, plot the probability p against the temperature by letting temperature go from 30 degrees to 90 degrees.

The plot of the probability of failure for our estimates of β s looks like this:

