

Maximum Likelihood Estimation: Numerical Solution for Bernoulli Distribution¹

The Bernoulli distribution has the probability mass function:

$$f(y|\theta) = \theta^y(1-\theta)^{1-y} \quad (1)$$

where: $y = \{0, 1\}$ and $\theta \in [0, 1]$.

Maximum Likelihood Estimation

Let y_1, \dots, y_n , denote the data. Assume, $\forall i : y_i$ independent random variables, share the same parameter from a Bernoulli distribution described in (1).

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

then, their join probability distribution is:

$$\begin{aligned} f(y_1, \dots, y_n|\theta) &= \prod_{i=1}^n f(y_i|\theta) \\ &= \prod_{i=1}^n \theta^{y_i}(1-\theta)^{1-y_i} \end{aligned} \quad (3)$$

The likelihood function is:

$$\mathcal{L}(\theta|y_1, \dots, y_n) = f(y_1, \dots, y_n|\theta) \quad (4)$$

The log-likelihood function is:

$$\begin{aligned} \ell(\theta|y_1, \dots, y_n) &= \log(\mathcal{L}(\theta|y_1, \dots, y_n)) \\ &= \log \prod_{i=1}^n \theta^{y_i}(1-\theta)^{1-y_i} \\ &= \sum_{i=1}^n \log(\theta^{y_i}(1-\theta)^{1-y_i}) \\ &= \sum_{i=1}^n (y_i \log \theta + (1-y_i) \log(1-\theta)) \end{aligned} \quad (5)$$

The maximum likelihood estimator (MLE), denoted by $\hat{\theta}$, is such that:

$$\hat{\theta} = \operatorname{argmax} \{ \ell(\theta|y_1, \dots, y_n) \} \quad (6)$$

¹To refer this document and the implemented code, please cite as: Alvarado, M. (2020, August 12). Maximum Likelihood Estimation: Numerical Solution for Bernoulli Distribution (Version v1.0.0). Zenodo. <http://doi.org/10.5281/zenodo.3981462>. Also at GitHub: https://github.com/miguel-alvarado-stats/MLE_Bernoulli.

To maximize the log-likelihood function (5), requires the derivative with respect to θ . The resulted function is called the Score function, denoted by $U(\theta|y_1, \dots, y_n)$.

$$\begin{aligned} U(\theta|y_1, \dots, y_n) &= \frac{\partial \ell(\theta|y_1, \dots, y_n)}{\partial \theta} \\ &= \sum_{i=1}^n \left(y_i \frac{1}{\theta} + (1 - y_i) \frac{1}{1 - \theta} (-1) \right) \\ &= \frac{1}{\theta(1 - \theta)} \sum_{i=1}^n y_i - \frac{n}{1 - \theta} \end{aligned} \quad (7)$$

Then, the MLE $\hat{\theta}$ is the solution of:

$$U(\theta = \hat{\theta}|y_1, \dots, y_n, \lambda) = 0 \quad (8)$$

Maximum Likelihood Estimation: Newton-Raphson Method

Just for notation, let write equation (8) as:

$$U(\theta^*) = 0 \quad (9)$$

The equation (9), generally, is a nonlinear equation, that can be aproximate by Taylor Series:

$$U(\theta^*) \approx U(\theta^{(t)}) + U'(\theta^{(t)})(\theta^* - \theta^{(t)}) \quad (10)$$

Then, using (10) into (9), and solving for θ^* :

$$\begin{aligned} U(\theta^{(t)}) + U'(\theta^{(t)})(\theta^* - \theta^{(t)}) &= 0 \\ \theta^* &= \theta^{(t)} - \frac{U(\theta^{(t)})}{U'(\theta^{(t)})} \end{aligned} \quad (11)$$

where U' is the derivative of the Score function (7) respect of θ .

$$\begin{aligned} U'(\theta|y_1, \dots, y_n, \lambda) &= \frac{\partial U(\theta|y_1, \dots, y_n, \lambda)}{\partial \theta} \\ &= \frac{2\theta - 1}{\theta^2(1 - \theta)^2} \sum_{i=1}^n y_i - \frac{n}{(1 - \theta)^2} \end{aligned} \quad (12)$$

Then, with the Newton-Raphson method: starting with an initial guess $\theta^{(1)}$ successive approximations are obtained using (13), until the iterative process converges.

$$\theta^{(t+1)} = \theta^{(t)} - \frac{U(\theta^{(t)})}{U'(\theta^{(t)})} \quad (13)$$

In order to example the use of Newton-Raphson method, we use the data of total July rainfall (in millimeters) at Quilpie, Australia stored into the data set “quilpie”, where the variable y is a dicotomic variable².

²This data was taken from package: GLMsData.

```
# load library
library(GLMsData)

# load data
data("quilpie")
Y <- quilpie$y
```

We load the code developed into our R function MLE_NR_Bernoulli, stored in the R object with the same name.

```
# load the function to solve by Newton-Raphson
load("MLE_NR_Bernoulli.RData")
```

The function MLE_NR_Bernoulli takes $\theta = 0.5$ as a first guess for the iterative process and, besides some other default parameters that can be modified, only needs the data vector Y.

```
# MLE by Newton-Raphson (NR) for Bernoulli distribution
MLE_NR_Bernoulli(Y)

##      ML Estimator Likelihood      Log-Likelihood
## [1,] "0.5000000" "3.3881317890172e-21" "-47.1340082780763"
## [2,] "0.5147059" "3.48927745358427e-21" "-47.1045922714519"
## [3,] "0.5147059" "3.48927745358427e-21" "-47.1045922714519"
```

Then, the MLE by Newton-Raphson method: $\hat{\theta} = 0.5147059$.

Maximum Likelihood Estimation: Fisher-Scoring Method

A distribution belongs to the exponential family if it can be written in the form:

$$f(y|\theta) = \exp \left\{ \frac{a(y)b(\theta) - c(\theta)}{\phi} + d(y, \phi) \right\} \quad (14)$$

Since (1) can be written as a member of exponential family as in (14):

$$\begin{aligned} f(y|\theta, \lambda) &= \exp \{ \log(\theta^y (1-\theta)^{1-y}) \} \\ &= \exp \left\{ y \log \left(\frac{\theta}{1-\theta} \right) - (-\log(1-\theta)) \right\} \end{aligned} \quad (15)$$

where, $a(y) = y$, $b(\theta) = \log \left(\frac{\theta}{1-\theta} \right)$, $c(\theta) = -\log(1-\theta)$, $\phi = 1$, and $d(y, \phi) = 0$.

Then, since the Bernoulli distribution belongs to the exponential family, it can be show that the variance of U , denoted by \mathcal{J} , is:

$$\mathcal{J} = \text{Var} \{U\} = -E \{U'\} \quad (16)$$

where:

$$E \{U'\} = -\frac{1}{\phi} \left(b''(\theta) \frac{c'(\theta)}{b'(\theta)} - c''(\theta) \right) \quad (17)$$

For MLE, it is common to approximate U' by its expected value $E\{U'\}$. In this case:

$$\begin{aligned}
\mathcal{J} &= -E\{U'\} \\
&= E\{-U'\} \\
&= E\left\{-\sum_{i=1}^n U'_i\right\} \\
&= \sum_{i=1}^n -E\{U'_i\} \\
&= \sum_{i=1}^n -\frac{1}{\phi} \left(b''(\theta) \frac{c'(\theta)}{b'(\theta)} - c''(\theta) \right)
\end{aligned} \tag{18}$$

where, using (1), the previous derivaties:

$$\begin{aligned}
b'(\theta) &= \frac{1}{\theta(1-\theta)} \\
b''(\theta) &= \frac{2\theta-1}{\theta^2(1-\theta)^2} \\
c'(\theta) &= \frac{1}{1-\theta} \\
c''(\theta) &= \frac{1}{(1-\theta)^2} \\
\frac{1}{\phi} &= 1
\end{aligned}$$

Then, replacing them into (18):

$$\begin{aligned}
\mathcal{J} &= \sum_{i=1}^n -\left\{ \frac{2\theta-1}{\theta^2(1-\theta)^2} \frac{\frac{1}{1-\theta}}{\frac{1}{\theta(1-\theta)}} - \frac{1}{(1-\theta)^2} \right\} \\
&= \sum_{i=1}^n \frac{1}{\theta(1-\theta)} \\
&= \frac{n}{\theta(1-\theta)}
\end{aligned} \tag{19}$$

Finally:

$$\begin{aligned}
\mathcal{J} &= -E\{U'\} = \frac{n}{\theta(1-\theta)} \\
-\mathcal{J} &= E\{U'\} = \frac{n}{\theta(1-\theta)}
\end{aligned} \tag{20}$$

Then, approximating U' by its expected value $E\{U'\}$, the equation (13) results into:

$$\theta^{(t+1)} = \theta^{(t)} + \frac{U(\theta^{(t)})}{\mathcal{J}(\theta^{(t)})} \tag{21}$$

In order to example the use of Fisher-Scoring method, we use the same data used in the Newton-Rapshon method. We load the code developed into our R function `MLE_FS_Bernoulli`, stored in the R object with the same name.

```
# load the function to solve by Fisher-Scoring
load("MLE_FS_Bernoulli.RData")
```

The function `MLE_FS_Bernoulli` takes $\theta = 0.5$ as a first guess for the iterative process and, besides some other default parameters that can be modified, only needs the data vector `Y`.

```
# MLE by Fisher-Scoring (FS) for Bernoulli distribution
MLE_FS_Bernoulli(Y)
```

##	ML Estimator	Likelihood	Log-Likelihood
## [1,]	"0.5000000"	"3.3881317890172e-21"	"-47.1340082780763"
## [2,]	"0.5147059"	"3.48927745358427e-21"	"-47.1045922714519"
## [3,]	"0.5147059"	"3.48927745358427e-21"	"-47.1045922714519"

Then, the MLE by Fisher-Scoring method: $\hat{\theta} = 0.5147059$.