Maximum Likelihood

Max Turgeon

STAT 3150–Statistical Computing

Maximum Likelihood i

- **Maximum Likelihood Estimation** is a general strategy for finding “good” estimators that was first proposed by R.A. Fisher.
- I will give the general definition, but a more thorough discussion is beyond the scope of today’s lecture.
- Notation:
  - $X_1, \ldots, X_n$ is a random sample.
  - $\mathbf{X} = (X_1, \ldots, X_n)$.
  - θ is a (population) parameter of interest.
  - S_θ is the parameter space, i.e. the set of possible values for θ.
  - $f(x; \theta)$ will denote the density function (or PMF) of the data.

Maximum Likelihood ii

Definition
The **likelihood function** $L(\theta \mid \mathbf{X})$ is the joint distribution of the observations considered as a function of θ:

$$L(\theta \mid \mathbf{X}) = \prod_{i=1}^{n} f(X_i; \theta).$$

A value $\hat{\theta}$ that maximises $L(\theta \mid \mathbf{X})$, in other words

$$L(\hat{\theta} \mid \mathbf{X}) = \max_{\theta \in S_\theta} L(\theta \mid \mathbf{X}),$$

is a **Maximum Likelihood Estimate** of θ.

Remarks

- In general, the MLE may not be unique. We need to make some assumptions (called *identifiability* assumptions) to ensure uniqueness.
- Since $\log$ is a monotone increasing function, maximising $L(\theta \mid \mathbf{X})$ is equivalent to maximising

$$\ell(\theta \mid \mathbf{X}) = \log L(\theta \mid \mathbf{X}).$$

  - Why would this be helpful?

Example i

- Suppose $X_1, \ldots, X_n$ is a random sample from a normal distribution $N(\mu, \sigma^2)$.
  - So $\theta = (\mu, \sigma^2)$.
- We have

$$
\begin{aligned}
L(\theta \mid \mathbf{X}) &= \prod_{i=1}^{n} f(X_i; \theta) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_i - \mu)^2\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right)
\end{aligned}
$$

Example ii

- By taking the log, we get the log-likelihood:

$$\ell(\theta \mid \mathbf{X}) = -\frac{n}{2}\log(2\pi\sigma^2) - -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2.$$

- We need to maximise this expression.
    - Take derivative.
    - Numerical methods?

Exercise 1

- Suppose we have a random sample $X_1, \ldots, X_n$ from an exponential $Exp(\lambda)$, where

$$f(x; \lambda) = \lambda \exp(-\lambda x), \quad x > 0.$$

- Write down the log-likelihood and find its derivative.
- Use numerical methods to find the MLE using the `aircondit` dataset in the `boot` package.

Solution 1 i

- The likelihood is given by

$$L(\lambda \mid \mathbf{X}) = \prod_{i=1}^{n} \lambda \exp(-\lambda X_i) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} X_i).$$

- The log-likelihood is therefore

$$\ell(\lambda \mid \mathbf{X}) = n \log \lambda - \lambda \sum_{i=1}^{n} X_i.$$

- The derivative with respect to λ is

$$\frac{d}{d\lambda}\ell(\lambda \mid \mathbf{X}) = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i.$$

Solution 1 ii

```r
library(boot)

log_lik_der <- function(lambda) {
    n <- nrow(aircondit)
    n/lambda - sum(aircondit$hours)
}

# We will look for a solution on [0.001, 1]
# We found the bounds by trial and error
uniroot(log_lik_der,
        c(0.001, 1))
```

Solution 1 iii

```
## $root
## [1] 0.009250825
##
## $f.root
## [1] 0.1816193
##
## $iter
## [1] 11
##
## $init.it
## [1] NA
##
```

Solution 1 iv

```
## $estim.prec
## [1] 0.0001107722
```

```r
# Check whether we get the same value
# as analytical solution
1/mean(aircondit$hours)
```

```
## [1] 0.00925212
```

Exercise 2

- This is an example of **grouped** or **binned** data:

Interval	Count
[0, 2)	2
[2, 3)	3
[3, 4)	1
[4, 5)	2
[5, 6)	1
[6, Inf)	1

- If the data follows $Exp(\lambda)$, what is the probability that X_i falls in a given bin?

Solution 2 i

- The probability an observation X_i falls in the interval $[a, b)$ can be computed using the CDF:

$$P(X_i \in [a, b)) = F(b) - F(a) = \exp(-\lambda a) - \exp(-\lambda b).$$

- If our data is in k bin $[a_j, b_j)$, and n_j is the number of elements in bin j, then our likelihood function is

$$L(\lambda \mid \mathbf{X}) = \prod_{j=1}^{k} \left(\exp(-\lambda a_j) - \exp(-\lambda b_j)\right)^{n_j}.$$

Solution 2 ii

- The log-likelihood is

$$\ell(\lambda \mid \mathbf{X}) = \sum_{j=1}^{k} n_j \log\left(\exp(-\lambda a_j) - \exp(-\lambda b_j)\right).$$

- The derivative with respect to λ is

$$\frac{d}{d\lambda}\ell(\lambda \mid \mathbf{X}) = \sum_{j=1}^{k} \frac{n_j\left(-a_j \exp(-\lambda a_j) + b_j \exp(-\lambda b_j)\right)}{\exp(-\lambda a_j) - \exp(-\lambda b_j)}.$$

Solution 2 iii

```r
# Create three vectors:
# 1. Lower bounds of bins
# 2. Upper bounds of bins
# 3. Number of values in bins
a_vec <- c(0, 2, 3, 4, 5, 6)
b_vec <- c(2, 3, 4, 5, 6, Inf)
n_vec <- c(2, 3, 1, 2, 1, 1)
```

Solution 2 iv

```r
log_lik_der_binned <- function(lambda) {
    num <- n_vec*(-a_vec*exp(-lambda*a_vec) +
                  b_vec*exp(-lambda*b_vec))
    # Need to fix last value manually
    # to avoid NaN value
    num[6] <- -n_vec[6]*a_vec[6]*exp(-lambda*a_vec[6])
    denom <- exp(-lambda*a_vec) - exp(-lambda*b_vec)
    sum(num/denom)
}
```

Solution 2 v

```r
# We will look for a solution on [0.1, 1]
uniroot(log_lik_der_binned,
        c(0.1, 1))
```

```
## $root
## [1] 0.271396
##
## $f.root
## [1] 3.952816e-05
##
## $iter
```

Solution 2 vi

```
## [1] 7
##
## $init.it
## [1] NA
##
## $estim.prec
## [1] 6.103516e-05
```