# Likelihood Functions

Hao Zhang

January 22, 2015

In this note, I introduce likelihood functions and estimation and statistical tests that are based on likelihood functions.

## 1 History

Likelihood function was introduced by R.A. Fisher (1922) who set up the modern framework of statistical problems as

(i) Specification (of a population model expressed as a family of probability distributions $P_\theta$ with $\theta$ being the parameter).

(ii) Estimation (of $\theta$ from observed data).

(iii) Uncertainty/precision of the estimate.

The observations generally are considered a random sample from some population or distribution. The distribution, if discrete, is specified by its probability mass function (pmf) or if continuous, is specified by its probability density function (pdf). We let $f(x, \theta)$ denote either the pmf or pdf. Denote the observed data by $x_1, \ldots, x_n$. The likelihood is defined as

$$L(\theta) = f(x_1, \ldots, x_n, \theta)$$

where $f(x_1, \ldots, x_n, \theta)$ is joint density (or mass) function. If the data are independent and have identical distribution $f(x, \theta)$. The likelihood function becomes

$$L(\theta) = f(x_1, \theta) \times f(x_1, \theta) \times \ldots \times f(x_n, \theta).$$

The likelihood estimate $\hat{\theta}$ is any value of $\theta$ that maximizes $L(\theta)$, i.e., $L(\hat{\theta}) \geq L(\theta)$.

We often maximize $\log L(\theta)$ instead.

## 2 Examples for Discrete Distributions

### 2.1 Bernoulli Distribution

Data are binary and take values 0 and 1 with $\theta = p(1)$ and $1 - \theta = p(0)$. For $x$ being 0 or 1, we can write the pmf

$$p(x, \theta) = \theta^x (1 - \theta)^{1-x}.$$

Hence the log-likelihood function is

$$\log L(\theta) = \sum_{i=1}^{n} \log(p_i, \theta) = \sum_{i=1}^{n} x_i \log(\theta) + \sum_{i=1}^{n} (1 - x_i) \log(1 - \theta).$$

$$\hat{\theta} = \sum x_i / n.$$

### 2.2 Binomial Distribution

Suppose in the previous example, we aggregate data to get the total count of 1's. Then the total number of 1's has a binomial distribution

$$p(x) = \frac{n!}{x!(n-x)!} \theta^x (1 - \theta)^{n-x}.$$

Hence

$$\log L(\theta) = \log p(x) = c + x \log \theta + (n - x) \log(1 - \theta).$$

where $c$ is a constant that does not depend on the parameter.

$$\hat{\theta} = x/n.$$

### 2.3 Multinomial Distribution

Suppose the population is divided into $k$ categories and the probability that a randomly chosen individual falls into category $i$ is $p_i$, $\sum_{i=1}^{k} p_i = 1$. If $n$ individuals are randomly selected and $x_i$ denotes the number of individuals that are from the $i$th category, then

$$p(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots, x_k!} p_1^{x_1} \cdots p_k^{x_k}.$$

$$L(\theta) = log \left( \frac{n!}{x_1! \cdots, x_k!} \right) + \sum_{i=1}^{k} x_i \log(p_i). \tag{1}$$

$$\text{MLE: } \hat{p}_i = x_i / n. \tag{2}$$

$$\max \log(\theta) = \log \left( \frac{n!}{x_1! \cdots, x_k!} \right) + \sum_{i=1}^{k} x_i \log(x_i / n). \tag{3}$$

### 2.4 Contingency Table

A population or a sample can be divided by two cross categorical variable. For example, a random sample of $n = 200$ people were randomly selected and their preferences (A or B) were recorded. It resulted in the following contingency table

| | A | B |
|---|---|---|
| Female | $n_{11} = 20$ | $n_{12} = 60$ |
| Male | $n_{21} = 40$ | $n_{22} = 80$ |

These counts $(n_{11}, n_{12}, n_{21}, n_{22})$ have a multinomial distribution with probabilities $p_{ij},\ i,j = 1,2$. According to the previous section, the log-likelihood and the estimates of $p_{ij}$ are

$$L(\theta) = \log\left(\frac{n!}{\prod_{ij} n_{ij}!}\right) + \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}\log(p_{ij}). \tag{4}$$

$$\text{MLE: } \hat{p}_{ij} = n_{ij}/n, \text{ for } n = n_{11} + n_{12} + n_{21} + n_{22} \tag{5}$$

$$\max\log(\theta) = \log\left(\frac{n!}{\prod_{ij} n_{ij}!}\right) + \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}\log(n_{ij}/n). \tag{6}$$

## 3 Likelihood Ratio Test

Suppose our data come from a model specified by a vector of parameter $\theta$. Let $\Theta$ be set of all possible values. Suppose we test

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

This is a very general set and includes many special cases. Is there evidence in the data that the null hypothesis should be rejected? The likelihood ratio test examines the ratio

$$\Lambda = \frac{\max\{L(\theta) : \theta \in \Theta_0\}}{\max\{L(\theta) : \theta \in \Theta\}}.$$

This ratio is between 0 and 1. A smaller ratio should result in the rejection of $H_0$. If the null hypothesis is true, $-2\log\Lambda$ has approximately a $\chi^2$ distribution with the number of degrees of freedom equal to dim($\Theta$)-dim($\Theta_0$), where dim($\Theta$) is dimension of $\Theta$ defined as the number of free parameters in $\Theta$.

### 3.1 Goodness-of-Fit Test

Let $(y_1, y_2, \ldots, y_k)$ have a multinomial distribution with probabilities $p_1, p_2, \ldots, p_k$. The test that these probabilities take a particular set of values is called a goodness-of-fit test. Denote by $\pi_1, \pi_2, \ldots, \pi_k$ the hypothesized values of $p_i$. The Pearson $\chi^2$ goodness of fit test is

$$\chi^2 = \sum_{i=1}^{k} \frac{(y_i - E_i)^2}{E_i}$$

where $E_i = n\pi_i$ for $n = \sum_i y_i$. Under the null hypothesis, it has a $\chi^2$ distribution with $k-1$ degrees of freedom. This well-known test is close to the likelihood ratio test.

We already know that

$$\max_{\theta \in \Theta} \log L(\theta) = c + \sum_{i=1}^{k} y_i \log\left(\frac{y_i}{n}\right).$$

Obviously

$$\max_{\theta \in \Theta_0} \log L(\theta) = c + \sum_{i=1}^{k} y_i \log \pi_i.$$

Hence

$$-2\log\Lambda = 2\sum_{i=1}^{k} y_i \log \frac{y_i}{n\pi_i}.$$

This quantity is close to Pearson chi-squared statistics. This can be shown by using the following approximation

$$-\log x \approx \left(\frac{1}{x} - 1\right) - \frac{1}{2}\left(\frac{1}{x} - 1\right)^2$$

for $x$ in a neighborhood of 1.

## 3.2 Test of Independence in a Contingency Table

Let us consider the test of independence for a $2 \times 2$ contingency table. The parameter space is $\Theta = \{\theta = (p_{11}, p_{12}, p_{21}, p_{22}) : \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij} = 1\}$. The null hypothesis is that the row variable and the column variable are independent, which translates into

$$\Theta_0 = \{(p_{11}, p_{12}, p_{21}, p_{22}) : p_{11} = p_r p_c, p_{12} = p_r(1-p_c), \tag{7}$$
$$p_{21} = (1-p_r)p_c, p_{22} = (1-p_r)(1-p_c), p_r \in [0,1], p_c \in [0,1]\}. \tag{8}$$

Based on the MLEs in Sections 2.3 and 2.4, we get

$$-2\log\Lambda = 2\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij} \log \frac{n_{ij}}{E_{ij}}.$$

where

$$E_{ij} = \text{sum of row i } \times \text{sum of column j}/n.$$

It has a degree of freedom as $(r-1)\times(c-1)$ where $r$ is the number of rows and $c$ is the number of columns. Hence the degrees of freedom is 1 for a $2 \times 2$ table.

A more commonly used test statistic is

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

that is close to the likelihood ratio test.