Maximum Likelihood Estimation - Examples*

Ewa Paszek

Abstract

This course is a short series of lectures on Introductory Statistics. Topics covered are listed in the Table of Contents. The notes were prepared by Ewa Paszek and Marek Kimmel. The development of this course has been supported by NSF 0203396 grant.

1 MAXIMUM LIKELIHOOD ESTIMATION - EXAMPLES

1.1 EXPONENTIAL DISTRIBUTION

Let $X_1, X_2, ..., X_n$ be a random sample from the exponential distribution with p.d.f.

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}, 0 < x < \infty, \theta \in \Omega = \{\theta; 0 < \theta < \infty\}.$$

The likelihood function is given by

$$L(\theta) = L(\theta; x_1, x_2, ..., x_n) = \left(\frac{1}{\theta}e^{-x_1/\theta}\right)\left(\frac{1}{\theta}e^{-x_2/\theta}\right)\cdots\left(\frac{1}{\theta}e^{-x_n/\theta}\right) = \frac{1}{\theta^n}\exp\left(\frac{-\sum_{i=1}^n x_i}{\theta}\right), 0 < \theta < \infty.$$

The natural logarithm of $L(\theta)$ is

$$\ln L(\theta) = -(n)\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^n x_i, 0 < \theta < \infty.$$

Thus,

$$\frac{d[\ln L(\theta)]}{d\theta} = \frac{-n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} = 0.$$

The solution of this equation for θ is

$$\theta = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}.$$

*Version 1.3: Oct 8, 2007 3:44 pm -0500

†http://creativecommons.org/licenses/by/2.0/

Note that,

$$\frac{d\left[\ln L\left(\theta\right)\right]}{d\theta}=\frac{1}{\theta}\left(-n+\frac{n\overline{x}}{\theta}\right)>0,\theta<\overline{x},$$

$$\frac{d\left[\ln L\left(\theta\right)\right]}{d\theta}=\frac{1}{\theta}\left(-n+\frac{n\overline{x}}{\theta}\right)=0,\theta=\overline{x},$$

$$\frac{d\left[\ln L\left(\theta\right)\right]}{d\theta}=\frac{1}{\theta}\left(-n+\frac{n\overline{x}}{\theta}\right)<0,\theta>\overline{x},$$

Hence, $\ln L\left(\theta\right)$ does have a maximum at $\overline{x}$, and thus the maximum likelihood estimator for θ is

$$\hat{\theta}=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

This is both an unbiased estimator and the method of moments estimator for θ.

1.2 GEOMETRIC DISTRIBUTION

Let $X_1, X_2, ..., X_n$ be a random sample from the geometric distribution with p.d.f.

$$f\left(x;p\right)=\left(1-p\right)^{x-1}p, x=1,2,3,....$$

The likelihood function is given by

$$L\left(p\right)=\left(1-p\right)^{x_1-1}p(1-p)^{x_2-1}p\cdots\left(1-p\right)^{x_n-1}p=p^n(1-p)^{\sum x_i-n}, 0\le p\le 1.$$

The natural logarithm of $L\left(\theta\right)$ is

$$\ln L\left(p\right)=n\ln p+\left(\sum_{i=1}^{n}x_i-n\right)\ln\left(1-p\right), 0<p<1.$$

Thus restricting **p** to $0<p<1$ so as to be able to take the derivative, we have

$$\frac{d\ln L\left(p\right)}{dp}=\frac{n}{p}-\frac{\sum_{i=1}^{n}x_i-n}{1-p}=0.$$

Solving for **p**, we obtain

$$p=\frac{n}{\sum_{i=1}^{n}x_i}=\frac{1}{\overline{x}}.$$

So the maximum likelihood estimator of **p** is

$$\hat{p}=\frac{n}{\sum_{i=1}^{n}X_i}=\frac{1}{X}$$

Again this estimator is the method of moments estimator, and it agrees with the intuition because, in n observations of a geometric random variable, there are **n** successes in the $\sum_{i=1}^{n}x_i$ trials. Thus the estimate of p is the number of successes divided by the total number of trials.

1.3 NORMAL DISTRIBUTION

Let $X_1, X_2, ..., X_n$ be a random sample from $N(\theta_1, \theta_2)$, where

$$\Omega = ((\theta_1, \theta_2) : -\infty < \theta_1 < \infty, 0 < \theta_2 < \infty).$$

That is, here let $\theta_1 = \mu$ and $\theta_2 = \sigma^2$. Then

$$L(\theta_1, \theta_2) = \prod_{i-1}^{n} \left(\frac{1}{\sqrt{2\pi\theta_2}} \exp\left[-\frac{(x_i - \theta_1)^2}{2\theta_2} \right] \right),$$

or equivalently,

$$L(\theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi\theta_2}} \right)^n \exp\left[-\frac{-\sum_{i=1}^{n} (x_i - \theta_1)^2}{2\theta_2} \right], (\theta_1, \theta_2) \in \Omega.$$

The natural logarithm of the likelihood function is

$$\ln L(\theta_1, \theta_2) = -\frac{n}{2} \ln(2\pi\theta_2) - \frac{-\sum_{i=1}^{n} (x_i - \theta_1)^2}{2\theta_2}.$$

The partial derivatives with respect to θ_1 and θ_2 are

$$\frac{\partial(\ln L)}{\partial\theta_1} = \frac{1}{\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)$$

and

$$\frac{\partial(\ln L)}{\partial\theta_2} = \frac{-n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \theta_1)^2.$$

The equation $\frac{\partial(\ln L)}{\partial\theta_1} = 0$ has the solution $\theta_1 = \overline{x}$. Setting $\frac{\partial(\ln L)}{\partial\theta_2} = 0$ and replacing θ_1 by $\overline{x}$ yields

$$\theta_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

By considering the usual condition on the second partial derivatives, these solutions do provide a maximum. Thus the maximum likelihood estimators

$$\mu = \theta_1$$

and

$$\sigma^2 = \theta_2$$

are

$$\hat{\theta}_1 = \overline{X}$$

and

$$\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2.$$

Where we compare the above example with the introductory one, we see that the method of moments estimators and the maximum likelihood estimators for μ and σ^2 are the same. But this is not always the case. If they are not the same, which is better? Due to the fact that the maximum likelihood estimator of θ has an approximate normal distribution with mean θ and a variance that is equal to a certain lower bound, thus at least approximately, it is unbiased minimum variance estimator. Accordingly, most statisticians prefer the maximum likelihood estimators than estimators found using the method of moments.

1.4 BINOMIAL DISTRIBUTION

Observations:k successes in **n** Bernoulli trials.

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

$$L(p) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \left(\frac{n!}{x_i!(n-x_i)!} p^{x_i} (1-p)^{n-x_i} \right) = \left(\prod_{i=1}^{n} \frac{n!}{x_i!(n-x_i)!} \right) p^{x_i} (1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$\ln L(p) = \sum_{i=1}^{n} x_i \ln p + \left(n - \sum_{i=1}^{n} x_i \right) \ln(1-p)$$

$$\frac{d\ln L(p)}{dp} = \frac{1}{p} \sum_{i=1}^{n} x_i - \left(n - \sum_{i=1}^{n} x_i \right) \frac{1}{1-p} = 0$$

$$\frac{\left(1-\hat{p}\right) \sum_{i=1}^{n} x_i - \left(n - \sum_{i=1}^{n} x_i\right) \hat{p}}{\hat{p}\left(1-\hat{p}\right)} = 0$$

$$\sum_{i=1}^{n} x_i - \hat{p} \sum_{i=1}^{n} x_i - n\hat{p} + \sum_{i=1}^{n} x_i \hat{p} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{k}{n}$$

1.5 POISSON DISTRIBUTION

Observations: $x_1, x_2, ..., x_n,$

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, ...$$

$$L(\lambda) = \prod_{i=1}^{n} \left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \right) = e^{-\lambda n} \frac{\lambda \sum_{i=1}^{n} x_i}{\prod_{i=1}^{n} x_i}$$

$$\ln L(\lambda) = -\lambda n + \sum_{i=1}^{n} x_i \ln \lambda - \ln \left(\prod_{i=1}^{n} x_i \right)$$

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

$$-n + \sum_{i=1}^{n} x_i \frac{1}{\lambda} = 0$$

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