Lecture 07

Lecture on Estimating Distribution Parameters Using the Method of Maximum Likelihood

Parameter Estimation

in order to fit a probability law to data, one typically has to estimate parameters associated with the probability law from the data.

The Method of Maximum Likelihood

Suppose that random variables $X_1, \ldots, X_n$ have a joint density or frequency function $f(x_1, x_2, \ldots, x_n|\theta)$. Given observed values $X_i = x_i$, where $i = 1, \ldots, n$, the likelihood of θ as a function of $x_1, x_2, \ldots, x_n$ is defined as

$$\text{lik}(\theta) = f(x_1, x_2, \ldots, x_n|\theta)$$

Note that we consider the joint density as a function of θ rather than as a function of the x_i. If the distribution is discrete, so that f is a frequency function, the likelihood function gives the probability of observing the given data as a function of the parameter θ. The **maximum likelihood estimate (mle)** of θ is that value of θ that maximizes the likelihood—that is, makes the observed data "most probable" or "most likely."

If the X_i are assumed to be i.i.d., their joint density is the product of the marginal densities, and the likelihood is

$$\text{lik}(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

Rather than maximizing the likelihood itself, it is usually easier to maximize its natural logarithm (which is equivalent since the logarithm is a monotonic function). For an i.i.d. sample, the **log likelihood** is

$$l(\theta) = \sum_{i=1}^{n} \log[f(X_i|\theta)]$$

(In this text, "log" will always mean the natural logarithm.)

Setting the first derivative of the log likelihood equal to zero, we find the estimator

$$l'(\theta) = 0$$

An intuitive explanation

If $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, their joint density is the product of their marginal densities:

$$f(x_1, x_2, \ldots, x_n \mid \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x_i - \mu}{\sigma}\right]^2\right)$$

Let us assume that μ is unknown and σ^2 is known. Let μ_{true} be the true value of μ that to be estimated.

Let X_i be written as $X_i = \epsilon_i + \mu_{\text{true}} \sim N(\mu_{\text{true}}, \sigma^2)$, where $\epsilon_i \sim N(0, \sigma^2)$.
The joint density may be written as

$$f(x_1, x_2, \ldots, x_n \mid \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{\epsilon_i + \mu_{\text{true}} - \mu}{\sigma}\right]^2\right)$$

Clearly the joint density is maximized when $\mu = \mu_{\text{true}}$.

Normal Distribution

If $X_1, X_2, \ldots, X_n$ are i.i.d. $N\left(\mu, \sigma^2\right)$, their joint density is the product of their marginal densities:

$$f\left(x_1, x_2, \ldots, x_n \mid \mu, \sigma^2\right) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x_i - \mu}{\sigma}\right]^2\right)$$

Regarded as a function of μ and σ^2, this is the likelihood function. The log likelihood is thus

$$l(\mu, \sigma^2) = -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

$$l(\mu, \sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

The partials with respect to μ and σ are

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (X_i - \mu)^2$$

Setting the first partial equal to zero and solving for the mle, we obtain

$$\widehat{\mu} = \widehat{\mu_1} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Setting the second partial equal to zero and substituting the mle for μ, we find that the mle for σ^2 is

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \widehat{\mu_1}\right)^2$$

Again, these estimates and their sampling distributions are the same as those obtained by the method of moments.

Gamma Distribution

Since the density function of a gamma distribution is

$$f(x|\alpha, \lambda) = \frac{1}{\Gamma(\alpha)}\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}, \qquad 0 \le x < \infty$$

the log likelihood of an i.i.d. sample, $X_i, \ldots, X_n$, is

$$\begin{aligned} l(\alpha, \lambda) &= \sum_{i=1}^{n}[\alpha \log \lambda + (\alpha - 1) \log X_i - \lambda X_i - \log \Gamma(\alpha)] \\ &= n\alpha \log \lambda + (\alpha - 1) \sum_{i=1}^{n} \log X_i - \lambda \sum_{i=1}^{n} X_i - n \log \Gamma(\alpha) \end{aligned}$$

The partial derivatives are

$$\frac{\partial l}{\partial \alpha} = n \log \lambda + \sum_{i=1}^{n} \log X_i - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$\frac{\partial l}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i$$

Setting the second partial equal to zero, we find

$$\widehat{\lambda} = \frac{n\widehat{\alpha}}{\sum_{i=1}^{n} X_i} = \frac{\widehat{\alpha}}{\widehat{\mu_1}}$$

Exercise 1

Code a computer program to first generate Gamma distributed data samples for the given values of λ and α.

Then using the data, estimate the true values of λ and α by solving these nonlinear equations in an iterative manner.

But when this solution is substituted into the equation for the first partial, we obtain a nonlinear equation for the mle of α :

$$n \log \hat{\alpha} - n \log \widehat{\mu_1} + \sum_{i=1}^{n} \log X_i - n \frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})} = 0$$

This equation cannot be solved in closed form; an iterative method for finding the roots has to be employed. To start the iterative procedure, we could use the initial value obtained by the method of moments.

An Angular Distribution

The angle θ at which electrons are emitted in muon decay has a distribution with the density

$$f(x|\alpha) = \frac{1 + \alpha x}{2}, \qquad -1 \leq x \leq 1 \qquad \text{and} \qquad -1 \leq \alpha \leq 1$$

where $x = \cos\theta$. The parameter α is related to polarization. Physical considerations dictate that $|\alpha| \leq \frac{1}{3}$, but we note that $f(x|\alpha)$ is a probability density for $|\alpha| \leq 1$.

The log likelihood is

$$l(\alpha) = \sum_{i=1}^{n} \log(1 + \alpha X_i) - n \log 2$$

Setting the derivative equal to zero, we see that the mle of α satisfies the following nonlinear equation:

$$\sum_{i=1}^{n} \frac{X_i}{1+\widehat{\alpha} X_i} = 0$$

Again, we would have to use an iterative technique to solve for $\widehat{\alpha}$.
The method of moments estimate could be used as a starting value.

Exercise 2

Code a computer program to first generate angular distributed data samples for the given value of α.

Then using the data, estimate the true value of α by solving the nonlinear equation in an iterative manner.

Example

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. with density function

$$f(x \mid \theta) = e^{-(x-\theta)}, \quad x \geq \theta$$

and $f(x \mid \theta) = 0$ otherwise.

Find the maximum likelihood estimate of θ.

Hint: This is a difficult problem, and needs order statistics.
Refer to the next slide for its solution.

Let n be the sample size, and let $X_1, \ldots, X_n$ be independent identically distributed random variables with the same density function (the one described in the exercise).
To find the MLE of θ, we first define the likelihood function:

$$\operatorname{lik}(\theta) = f(x_1, \ldots, x_n \mid \theta) = f(x_1 \mid \theta) \cdots f(x_n \mid \theta).$$

Substituting the definition of the density function of X yields

$$\operatorname{lik}(\theta) = e^{-(x_1-\theta)} \cdot e^{-(x_2-\theta)} \cdots e^{-(x_n-\theta)} = e^{n\theta-(x_1+\cdots x_n)}.$$

It's easier to work with the natural logarithm of the given expression, so we define

$$l(\theta) = \ln(\operatorname{lik}(\theta)) = n \cdot \theta - \sum_{i=1}^{n} x_i$$

Notice that this log-likelihood is defined this way only when $\theta \leq x_1, \ldots, x_n$, because the density f is non-zero only when $\theta \leq x$ (and is 0 otherwise).

If θ is greater than some x_i, then the log-likelihood is not defined, because the likelihood is 0 .

So, we can notice that the log-likelihood is an affine function (i.e. its graph is a straight line), 4which is strictly increasing, since the coefficient next to θ is n, which is always positive.

Therefore, the log-likelihood is the greatest when θ is the greatest possible, but with the constraint that

$$\theta \leq \{x_1, \ldots, x_n\}$$

So, if we set that θ is the minimum of those nx_i 's, then it should be clear that at that point the log-likelihood is the greatest.
In other words, we have that

$$\widehat{\theta} = X_{(1)}$$

is the MLE for θ, where $X_{(1)}$ denotes the minimum of the sample, i.e. $X_{(1)} = \min_{1 \leq i \leq n} X_i$.

Exercise 3

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables on the interval $[0, 1]$ with the density function

$$f(x|\alpha) = \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} x^{\alpha-1}(1-x)^{2\alpha-1}$$

where $\alpha > 0$ is a parameter to be estimated from the sample. It can be shown

$$E(X) = \frac{1}{3}$$

$$\mathrm{Var}(X) = \frac{2}{9(3\alpha+1)}$$

Show that the maximum likelihood estimate of α is given by

$$\frac{\Gamma'(2\alpha)\cdot\Gamma(\alpha) - \Gamma(2\alpha)\cdot\Gamma'(\alpha)}{\Gamma(\alpha)\cdot\Gamma(2\alpha)} = -\frac{1}{2n}\cdot\sum_{i=1}^{n}\ln(x_i\cdot(1-x_i))$$