PARAMETER ESTIMATION

Parameter Estimation

Estimation theory is a branch of statistics that deals with estimating the values of parameters based on measured empirical data that has a random component.

The parameters describe an underlying physical setting in such a way that their value affects the distribution of the measured data.

An estimator attempts to approximate the unknown parameters using the measurements.

The *method of least squares* (*least squares estimation, LSE*) is a standard approach to approximate the solution of overdetermined systems, i.e. sets of equations in which there are more equations than unknowns.

"Least squares" means that the overall solution minimizes the sum of the squares of the ***residuals*** made in the results of every single equation.

Least-squares problems fall into two categories:

- linear or ordinary least squares;
- nonlinear least squares,

depending on whether or not the residuals are linear in all unknowns.

The linear least-squares problem has a closed-form solution. The nonlinear problem is usually solved by iterative refinement; at each iteration the system is approximated by a linear one, and thus the core calculation is similar in both cases.

Parameter Estimation

The objective consists of adjusting the parameters of a model function to best fit a data set.

A simple data set consists of *n* data pairs $(x_i, y_i), i = 1, \dots, n$.

The model function has the form $F(x, \Theta)$, where *m* adjustable parameters are held in the vector Θ.

The goal is to find the parameter values $\hat{\theta}_j$, $j = 1, \dots, m$ for the model that "best" fits the data.

The fit of a model to a data point is measured by its residual, defined as the difference between the actual value of the dependent variable and the value predicted by the model:

$$r(\widehat{\Theta})_i = y_i - F(x_i, \widehat{\Theta})$$

The LSE method finds the optimal parameter values by minimizing the sum of squared residuals:

$$S(\widehat{\Theta}) = \sum_{i=1}^{n} r(\widehat{\Theta})_i^2$$

Ex.: Assume that five identical units are being reliability tested. The units fail during the test after operating the following number of hours: 20, 275, 365, 415, and 1020.

Assuming that the data follow exponential distribution, estimate the value of the parameter λ.

Here, the vector $\widehat{\Theta}$ contains single element - $\hat{\lambda}$, and

$$F(t, \Theta) = 1 - e^{-\Theta t}.$$

The naïve approach is to minimize the sum of squared residuals as previously specified:

$$S(\widehat{\Theta}) = \sum_{i=1}^{5} r(\widehat{\Theta})_i^2 = \sum_{i=1}^{5} \left(MR_i - \left(1 - e^{-\widehat{\Theta} t_i}\right)\right)^2$$

where MR_i are median ranks.

t_i	MR_i
20	0,129
275	0,314
365	0,5
415	0,686
1020	0,871

However, this would lead to a nonlinear equation with respect to $\widehat{\Theta}$.

We can avoid it by linearizing the *cdf:*

$$F(t,\Theta) = 1 - e^{-\Theta t} \qquad 1 - F(t,\Theta) = e^{-\Theta t}$$

$$\ln\big(1 - F(t,\Theta)\big) = -\Theta t \quad \Rightarrow \quad y = kx + b$$

$$y \equiv \ln\big(1 - F(t,\Theta)\big) = \ln(1 - MR)$$

$$k \equiv -\Theta$$

$$x \equiv t$$

$$b \equiv 0$$

Then the sum of squared residuals:

$$S(\hat{k}) = \sum_{i=1}^{5} \left(y_i - \hat{k}x_i\right)^2 = \sum_{i=1}^{5} \left(y_i^2 - 2\hat{k}x_i y_i + \hat{k}^2 x_i^2\right) =$$

$$= \hat{k}^2 \sum_{i=1}^{5} x_i^2 - 2\hat{k} \sum_{i=1}^{5} x_i y_i + \sum_{i=1}^{5} y_i^2$$

To find the minimum of $S(\hat{k})$ we should set $\frac{dS}{d\hat{k}} = 0$.

Parameter Estimation

$$\frac{dS}{d\hat{k}} = 2\hat{k}\sum_{i=1}^{5} x_i^2 - 2\sum_{i=1}^{5} x_i y_i = 0$$

$$\hat{k} = \frac{\sum_{i=1}^{5} x_i y_i}{\sum_{i=1}^{5} x_i^2}$$

Parameter Estimation

	x_i	x_i^2	MR_i	y_i	$x_i y_i$
	20	400	0,129	-0,138	-2,762
	275	75625	0,314	-0,377	-103,641
	365	133225	0,5	-0,693	-252,999
	415	172225	0,686	-1,158	-480,72
	1020	1040400	0,871	-2,048	-2088,902
Σ		1421875			-2929,024

$$\hat{k} = \frac{-2929.024}{1421875} = -2.06 \times 10^{-3}$$

$$\hat{\lambda} = 2.06 \times 10^{-3}$$

The LSE method is quite good for functions that can be linearized.

For these distributions, the calculations are relatively easy and straightforward, having closed-form solutions that can readily yield an answer without having to resort to numerical techniques or tables.

LSE is generally best used with data sets containing complete data, that is, data consisting only of single times-to-failure with no censored or interval data.

Parameter Estimation

In statistics, *maximum likelihood estimation* (*MLE*) is a method of estimating the parameters of a statistical model so the observed data is most probable.

Specifically, this is done by finding the value of the parameter (or parameter vector) $\widehat{\Theta}$ that maximizes the likelihood function $\mathcal{L}(\widehat{\Theta})$, which is the joint probability (or probability density) of the observed data over a parameter space.

The point $\widehat{\Theta}$ that maximizes the likelihood function is called the *maximum likelihood estimate*.

The logic of maximum likelihood is both intuitive and flexible, and as such the method has become a dominant means of inference within much of the quantitative research of the social and medical sciences and engineering.

Parameter Estimation

Consider X – a continuous random variable with pdf:

$$f(x, \Theta) \equiv f(x, \theta_1, \theta_2, \dots, \theta_k)$$

where $\theta_1, \theta_2, \dots, \theta_k$ are k unknown parameters which need to be estimated, with N independent observations, $x_1, x_2, \dots, x_N$, which correspond in the case of life data analysis to failure times.

The *likelihood function* is given by:

$$\mathcal{L}(\widehat{\Theta}) = \prod_{i=1}^{N} f(x_i, \widehat{\Theta})$$

In practice, it is often convenient to work with the natural logarithm of the likelihood function, called the *logarithmic likelihood* (*log-likelihood*) *function*:

$$\Lambda(\widehat{\Theta}) = \ln \mathcal{L}(\widehat{\Theta}) = \sum_{i=1}^{N} \ln f(x_i, \widehat{\Theta})$$

The maximum likelihood estimators (or parameter values) of $\theta_1, \theta_2, \ldots, \theta_k$ are obtained by maximizing $\mathcal{L}(\widehat{\Theta})$ or $\Lambda(\widehat{\Theta})$:

$$\frac{\partial \Lambda}{\partial \hat{\theta}_j} = 0, \qquad j = 1,2, \ldots, k$$

In practice, it is often convenient to work with the natural logarithm of the likelihood function, called the *logarithmic likelihood* (*log-likelihood*) *function*:

$$\Lambda\left(\widehat{\Theta}\right) = \ln \mathcal{L}\left(\widehat{\Theta}\right) = \sum_{i=1}^{N} \ln f\left(x_i, \widehat{\Theta}\right)$$

The maximum likelihood estimators (or parameter values) of $\theta_1, \theta_2, \dots, \theta_k$ are obtained by maximizing $\mathcal{L}\left(\widehat{\Theta}\right)$ or $\Lambda\left(\widehat{\Theta}\right)$:

$$\frac{\partial \Lambda}{\partial \hat{\theta}_j} = 0, \qquad j = 1,2, \dots, k$$

Ex.: Assume that five identical units are being reliability tested. The units fail during the test after operating the following number of hours: 20, 275, 365, 415, and 1020.

Assuming that the data follow exponential distribution, estimate the value of the parameter λ.

Here, the vector $\widehat{\Theta}$ contains single element - $\hat{\lambda}$, and

$$F(t, \Theta) = 1 - e^{-\Theta t}.$$

Parameter Estimation

The log-likelihood function:

$$\Lambda(\widehat{\Theta}) = \sum_{i=1}^{5} \ln\left(\widehat{\Theta} e^{-\widehat{\Theta} t_i}\right) = 5 \ln \widehat{\Theta} - \widehat{\Theta} \sum_{i=1}^{5} t_i$$

Substituting failure times for t_i, we get:

$$\Lambda(\widehat{\Theta}) = 5 \ln \widehat{\Theta} - 2095 \cdot \widehat{\Theta}$$

$$\frac{\partial \Lambda}{\partial \widehat{\Theta}} = \frac{5}{\widehat{\Theta}} - 2095 = 0 \quad \Rightarrow \quad \widehat{\Theta} = \frac{5}{2095} = 2.39 \times 10^{-3}$$

$$\hat{\lambda} = 2.39 \times 10^{-3}$$

Analyzing the results of two previous examples, you should notice that parameter estimates differ from one another, and we can't specify which result is better.

We can evaluate *residual sum of squares* (*RSS*) or *mean squared error* (*MSE*):

$$RSS(\widehat{\Theta}) = \sum_{i=1}^{N} \left(Y_i - F(X_i, \widehat{\Theta})\right)^2 \quad MSE(\widehat{\Theta}) = \frac{RSS(\widehat{\Theta})}{N}$$

We also can determine the likelihood of either result by calculating $\Lambda(\widehat{\Theta})$, or, $-2\Lambda(\widehat{\Theta})$ - the metric used in various statistical quality tests.

Let's compare these metrics obtained with the results of LSE and MLE:

	RSS	-2Λ
$\hat{\lambda}_{LSE} = 2.06 \times 10^{-3}$	0.035	70.482
$\hat{\lambda}_{MLE} = 2.39 \times 10^{-3}$	0.047	70.379

	RSS	-2Λ
$\hat{\lambda}_{LSE} = 2.06 \times 10^{-3}$	0.035	70.482
$\hat{\lambda}_{MLE} = 2.39 \times 10^{-3}$	0.047	70.379

The results we have here are quite obvious: $\hat{\lambda}_{LSE}$ is the value for which *RSS* is minimal, so any other parameter value yields greater *RSS*.

Likewise, $\hat{\lambda}_{MLE}$ is the value that maximizes $\Lambda\left(\widehat{\Theta}\right)$ (and minimizes $-2\Lambda\left(\widehat{\Theta}\right)$).