Maximum Likelihood Estimation (MLE) for Gamma Distribution

AE4426-19 Stochastic processes and simulation, 2020-2021

1 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) is a method to estimate the parameters of a probability distribution from the observed data, by maximizing a likelihood function.

Let $\mathbf{y} = [y_1, y_2, \ldots, y_n]$ be the observed data of a random variable Y. We assume that this random variable Y follows a probability distribution. Let $f(y;\theta)$ be the probability density function (pdf) of Y, and its parameters are $\theta = [\theta_1, \theta_2, \ldots, \theta_k]^{\mathrm{T}}$. We observe data $\mathbf{y} = [y_1, y_2, \ldots, y_n]$. The joint probability density of the observed data $\mathbf{y}$ is the product of pdfs at y_i, i.e.,

$$f_n(\mathbf{y};\theta) = \prod_{i=1}^{n} f(y_i;\theta). \tag{1}$$

We define this as a *likelihood function* $L(\theta;\mathbf{y})$, i.e.,

$$L(\theta;\mathbf{y}) = f_n(\mathbf{y};\theta). \tag{2}$$

The goal of MLE is to find the parameters $\hat{\theta}$ that maximizes the likelihood function L, i.e.,

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \ \ L(\theta;\mathbf{y}). \tag{3}$$

In practice, we often maximize *log-likelihood* $l(\theta;\mathbf{y})$, the natural logarithm of the likelihood function, i.e.,

$$l(\theta;\mathbf{y}) = \log L(\theta;\mathbf{y}). \tag{4}$$

Since the logarithm is a monotonic function, the maximum of $l(\theta;\mathbf{y})$ occurs at the same value of θ as does the maximum of $L(\theta;\mathbf{y})$. If $l(\theta;\mathbf{y})$ is differentiable in θ, the necessary conditions for the occurrence of maximum are:

$$\frac{\partial l}{\partial \theta_j} = 0, \quad \forall j \in \{1, \ldots, k\}. \tag{5}$$

Therefore, the parameters $\theta = [\theta_1, \ldots, \theta_k]$ can be found as the solution of the set of equations in eq.(5).

MLE is a useful tool when we know the probability density function f, but not its parameters θ. It can be applied to various distributions such as normal distribution, exponential distribution, Gamma distribution, etc.

2 MLE for the Gamma distribution

Here, we discuss the MLE for the Gamma distribution. The Gamma distribution has two parameters – *shape parameter* α and *scale parameter* β, i.e., $\theta = [\alpha, \beta]$. (Note that some people refer $1/\beta$ as *rate parameter*.) The pdf of the Gamma distribution is:

$$f(y;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} e^{-\frac{y}{\beta}} \tag{6}$$

where $\Gamma(\alpha) = \int_0^\infty s^{\alpha-1}e^{-s}ds$ is the Gamma function.

The likelihood function is obtained as the joint pdf of $\mathbf{y}$. Since y_i are independent, the joint pdf is the product of pdfs as follows:

$$L(\alpha, \beta; \mathbf{y}) = f_n(\mathbf{y}, \alpha, \beta) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}} \tag{7}$$

The log-likelihood is

$$\begin{aligned} l(\alpha, \beta; \mathbf{y}) &= \log L(\alpha, \beta; \mathbf{y}) \\ &= (\alpha - 1)\sum_{i=1}^{n} \log y_i - \frac{1}{\beta}\sum_{i=1}^{n} y_i - n\log\Gamma(\alpha) - n\alpha\log\beta \\ &= n(\alpha - 1)\overline{\log y} - \frac{n}{\beta}\overline{y} - n\log\Gamma(\alpha) - n\alpha\log\beta \end{aligned} \tag{8}$$

where $\overline{\log y} = (\sum_{i=1}^n \log y_i)/n$ and $\overline{y} = (\sum_{i=1}^n y_i)/n$. The necessary condition for the MLE of $\hat{\alpha}$ and $\hat{\beta}$ are two equations in eq.(5), i.e., $\partial l/\partial\alpha = 0$ and $\partial l/\partial\beta = 0$. From the second one, we get the relation between $\hat{\alpha}$ and $\hat{\beta}$ as follows:

$$\frac{\partial l}{\partial \beta} = \frac{n\overline{y}}{\beta^2} - \frac{n\alpha}{\beta} = 0 \tag{9}$$

$$\therefore \beta = \overline{y}/\alpha \tag{10}$$

Substituting this β into eq.(8), we represent the log-likelihood ofthe Gamma distribution as a function of α, i.e.,

$$l(\alpha, \beta; \mathbf{y}) = n(\alpha - 1)\overline{\log y} - n\alpha - n\log\Gamma(\alpha) - n\alpha\log\overline{y} + n\alpha\log\alpha. \tag{11}$$

Now, we can get $\hat{\alpha}$ by finding α that maximizes log-likelihood in eq.(11). The maximization of eq.(11) can be done in several computational ways, but the simplest approach is an exhaustive search, i.e., plot $l(\alpha, \beta; \mathbf{y})$ as a function of α. Figure 1 shows an example of the exhaustive search for MLE of the Gamma distribution. After obtaining $\hat{\alpha}$ that maximizes eq.(11), $\hat{\beta}$ is obtained by substituting $\hat{\alpha}$ into α of eq.(10).

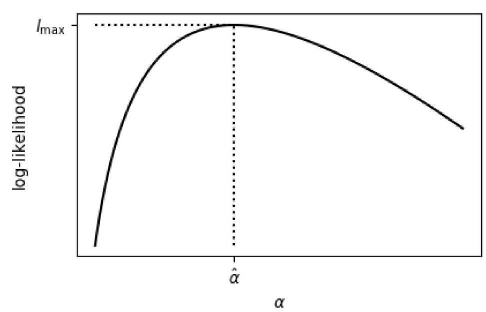

Figure 1: An example of exhaustive search for MLE of Gamma distribution.

Reference

1. Maximum Likelihood Estimation in Wikipedia - https://en.wikipedia.org/wiki/Maximum_likelihood_estimation
2. Gamma distribution in Wikipedia - https://en.wikipedia.org/wiki/Gamma_distribution
3. More efficient method to maximize eq.(11), Thomas P. Minka (2002) - https://tminka.github.io/papers/minka-gamma.pdf