Stats Topic

Exercises on Maximum Likelihood Estimation

Updated on January 21, 2024

© 2024 Nicky老师
nickycanhelp@qq.com

The maximum likelihood estimation (MLE) is one of the core topics in statistics. This note reviews several classical and representative exercises in this area.

I. Preliminaries

Assume that a parametric statistical model $\mathcal{F}$ admits a distribution μ_θ[1], whose parameter θ is unknown (but fixed) and is supposed to lie in a given set Θ, namely,

$$\mathcal{F} = \{\mu_\theta : \theta \in \Theta\}. \tag{1}$$

We also have a set of independent and identically-distributed (i.i.d.) samples generated from the true distribution μ_θ with a certain θ. We denote $\mathcal{X}_n = \{X_1, \dots, X_n\}$, where $X_i \overset{i.i.d.}{\sim} \mu_\theta$, $i = 1, \dots, n$.

The MLE is one of the methods to estimate the unknown θ from the dataset $\mathcal{X}_n$ of n i.i.d. samples. The idea of MLE says that such estimator $\hat{\theta}^{MLE}$ should make the current samples occur with the largest possibility. This is achieved by maximizing the so-called likelihood function $\mathcal{L}(\theta)$, i.e.,

$$\hat{\theta}^{MLE} \in \arg\max_{\theta \in \Theta} \mathcal{L}(\theta), \tag{2}$$

where

$$\mathcal{L}(\theta) := \prod_{i=1}^{n} \mu_\theta(X_i). \tag{3}$$

Sometimes it is much easier to consider the log-likelihood function defined as $\ell(\theta) := \ln(\mathcal{L}(\theta))$. Since the logarithm function is increasing, the solution sets of maximizing $\mathcal{L}(\theta)$ and $\ell(\theta)$ are equivalent.

Forming the likelihood function depends on the concrete problems. In the simplest case where the samples come from a continuous distribution with probability density function (pdf) f_θ, the likelihood function is usually (but not always) the product of pdf's evaluated at each sample, namely,

$$\mathcal{L}(\theta|\mathcal{X}_n) = \prod_{i=1}^{n} f_\theta(X_i). \tag{4}$$

In other cases, however, determining the likelihood function is more tricky. In the following, we consider several exercises on solving MLE.

[1] Or, in general, a Lebesgue measure

II. MLE Exercises

1 Normal distribution

Question: Suppose $\mathcal{X}_n = \{X_1, \dots, X_n\}$ are i.i.d. samples from a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with unknown mean μ and unknown standard deviation $\sigma > 0$. Find their maximum likelihood estimators $\hat{\mu}^{MLE}$ and $\hat{\sigma}^{MLE}$.
Solution: The parametric statistical model of this problem is

$$\mathcal{F} = \{\mathcal{N}(\mu, \theta) : \mu \in \mathbb{R}, \sigma > 0\}. \tag{5}$$

We know that the pdf of $\mathcal{N}(\mu, \sigma^2)$ is

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{6}$$

Therefore, the likelihood and log-likelihood are given by

$$\begin{aligned}
\mathcal{L}(\mu, \sigma|\mathcal{X}_n) &= \prod_{i=1}^{n} f(X_i|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{\sum_i (X_i-\mu)^2}{2\sigma^2}}, && (7) \\
\ell(\mu, \sigma|\mathcal{X}_n) &= \ln(\mathcal{L}(\mu, \sigma|\mathcal{X}_n)) \\
&= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{\sum_i (X_i - \mu)^2}{2\sigma^2}. && (8)
\end{aligned}$$

The maximum likelihood estimators $\hat{\mu}^{MLE}$ and $\hat{\sigma}^{MLE}$ are solving from

$$\max_{\mu \in \mathbb{R}, \sigma > 0} \ell(\mu, \sigma|\mathcal{X}_n). \tag{9}$$

The first-order condition (FOC) gives

$$\begin{cases} \frac{\partial \ell}{\partial \mu} = -\frac{\sum_i (X_i - \mu)}{\sigma^2} \overset{\text{set}}{=} 0 \\ \frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_i (X_i-\mu)^2}{\sigma^3} \overset{\text{set}}{=} 0 \end{cases}. \tag{10}$$

Solving the above system of equations leads to

$$\hat{\mu}^{MLE} = \bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{11}$$

$$\hat{\sigma}^{MLE} = \sqrt{\frac{1}{n}\sum_i (X_i - \hat{\mu}^{MLE})^2}. \tag{12}$$

2 Exponential distribution

Question: Suppose the lifetime of a bulb, denoted by T, follows an exponential distribution with unknown parameter λ. Now, we have n bulbs of this type and we turn on all these bulbs in a room at $t = 0$ and come back until $t = \tau$ and find that m out of them are out of work while the rest is still lighting. Find the maximum likelihood estimators $\hat{\lambda}^{MLE}$.
Solution: The parametric statistical model of this problem is

$$\mathcal{F} = \{Exp(\lambda) : \lambda > 0\}. \tag{13}$$

We know that the pdf of $Exp(\lambda)$ is

$$f(t|\lambda) = \lambda e^{-\lambda t},\ t > 0. \tag{14}$$

Therefore, its cumulative density function (cdf) is given by, for $t > 0$,

$$F(t|\lambda) = \mathbb{P}(T \leq t) = \int_0^t f(s|\lambda)ds = 1 - e^{-\lambda t}. \tag{15}$$

In this problem, we have n i.i.d. bulbs with random lifetime T_i, $i = 1, \ldots, n$. When we examine the bulbs at $t = \tau$, we find that m out of n do not work anymore. Without loss of generality, let us number the broken bulbs from 1 to m and the rest from $m + 1$ to n. Therefore, the likelihood function in this case is given by

$$\begin{aligned}\mathcal{L}(\lambda) &= \prod_{i=1}^{m} \mathbb{P}(T_i \leq \tau) \prod_{i=m+1}^{n} \mathbb{P}(T_i > \tau) \\ &= (1 - e^{-\lambda\tau})^m (e^{-\lambda\tau})^{n-m},\end{aligned} \tag{16}$$

and the log-likelihood function is then given by

$$\ell(\lambda) = \ln(\mathcal{L}(\lambda)) = m\ln(1 - e^{-\lambda\tau}) - (n - m)\lambda\tau. \tag{17}$$

The maximum likelihood estimators $\hat{\lambda}^{MLE}$ is solved from $\max_{\lambda>0} \ell(\lambda)$. FOC leads to

$$\frac{\partial \ell}{\partial \lambda} = \frac{\tau m e^{-\lambda\tau}}{1 - e^{-\lambda\tau}} - (n - m)\tau \overset{\text{set}}{=} 0. \tag{18}$$

We finally obtain

$$\hat{\lambda}^{MLE} = -\frac{1}{\lambda} \ln \frac{n - m}{n}. \tag{19}$$

◊ 这道题虽然也是连续分布，但很明显不是pdf的连乘来构造似然函数这么简单的情形了，是有点trick在身上的.

3 Multinomial distribution - Example 1

<u>Question:</u> Suppose $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ are i.i.d. samples from a trinomial distribution whose probability mass function (pmf) is given by

$$\mathbb{P}(X = a_k) = p_k,\ k = 1, 2, 3, \tag{20}$$

where a_k are given constants and $p_k > 0$ are unknown probabilities satisfying $\sum_k p_k = 1$. Find the maximum likelihood estimators $\hat{p}_k^{MLE}$ for all k.
<u>Solution:</u> The parametric statistical model of this problem is

$$\mathcal{F} = \{Tri(p_1, p_2, p_3) : p_1 + p_2 + p_3 = 1, p_k > 0\}. \tag{21}$$

The likelihood function of this problem is defined as

$$\begin{aligned}\mathcal{L}(p_1, p_2, p_3) &= \prod_{i=1}^{n} \mathbb{P}(X_i) \\ &= \prod_{i=1}^{n} (p_1 \mathbb{1}_{\{X_i=a_1\}} + p_2 \mathbb{1}_{\{X_i=a_2\}} + p_3 \mathbb{1}_{\{X_i=a_3\}}) \\ &= p_1^{\sharp_1} p_2^{\sharp_2} p_3^{\sharp_3},\end{aligned} \tag{22}$$

where

$$\mathbb{1}_{\{x \in A\}} = \begin{cases} 1 & x \in A \\ 0 & o/w \end{cases} \tag{23}$$

is known as the indicator function and

$$\sharp_k = \sum_{i=1}^{n} \mathbb{1}_{\{X_i = a_k\}} \tag{24}$$

just counts occurrences of a_k out of n samples, and obviously $\sharp_1 + \sharp_2 + \sharp_3 = n$.

The log-likelihood function is then given by

$$\ell(p_1, p_2, p_3) = \ln(\mathcal{L}(p_1, p_2, p_3)) = \sharp_1 \ln(p_1) + \sharp_2 \ln(p_2) + \sharp_3 \ln(p_3). \tag{25}$$

The maximum likelihood estimators $\hat{p}_k^{MLE}$ are solved from

$$\begin{aligned}\max_{p_k, \forall k} \quad & \ell(p_1, p_2, p_3) \\ \text{s.t.} \quad & p_1 + p_2 + p_3 = 1 \\ & p_k > 0.\end{aligned}$$

The above programming can be easily solved by the Lagrangian method. We first define the Lagrangian as

$$L(p_1, p_2, p_3, \lambda) = \ell(p_1, p_2, p_3) - \lambda(p_1 + p_2 + p_3 - 1). \tag{26}$$

The FOC leads to

$$\begin{cases} \frac{\partial L}{\partial p_k} = \frac{\sharp_k}{p_k} - \lambda \overset{\text{set}}{=} 0,\ k = 1, 2, 3 \\ \frac{\partial L}{\partial \lambda} = -(p_1 + p_2 + p_3 - 1) \overset{\text{set}}{=} 0 \end{cases}. \tag{27}$$

After solving the above system of equations for all p_k and λ, we obtain

$$\hat{p}_k^{MLE} = \frac{\sharp_k}{n},\ k = 1, 2, 3. \tag{28}$$

◊ 这道题的求解过程看似复杂和抽象, 但结论非常地intuitive: p_k的MLE正是a_k出现的次数($\sharp_k$)在总次数(n)中所占的比例.

4 Multinomial distribution - Example 2

Question: Suppose $\mathcal{X}_n = \{X_1, \ldots, X_n\}$, $X_i \in \mathbb{R}^m$, are i.i.d. samples from a m-dimensional multinomial distribution whose probability mass function (pmf) is given by

$$\mathbb{P}(X = e_j) = p_j,\ j = 1, \ldots, m, \tag{29}$$

where $e_j = (0, \ldots, 1, \ldots, 0)^T \in \mathbb{R}^m$ is the vector whose jth coordinate is 1 while others all zeros, and $p_j > 0$ are unknown probabilities satisfying $\sum_j p_j = 1$. Find the maximum likelihood estimators $\hat{p}_j^{MLE}$ for all j.

Solution: The parametric statistical model of this problem is

$$\mathcal{F} = \{Multi(p_1, \ldots, p_m) : \sum_j p_j = 1, p_j > 0\}. \tag{30}$$

The likelihood function of this problem is defined as

$$\begin{aligned} \mathcal{L}(p_1, \ldots, p_m) &= \prod_{i=1}^n \mathbb{P}(X_i) \\ &= \prod_{i=1}^n \prod_{j=1}^m p_j^{X_i^j}, \end{aligned} \tag{31}$$

where X_i^j represents jth element of X_i. Note that in this problem $X_i^j = 1$ with probability p_j and $X_i^j = 0$ with probability $1 - p_j$, and obviously $\sum_{j=1}^m X_i^j = 1$.

The log-likelihood function is then given by

$$\ell(p_1, \ldots, p_m) = \ln(\mathcal{L}(p_1 \ldots, p_m)) = \sum_{i=1}^n \sum_{j=1}^m X_i^j \ln p_j. \tag{32}$$

The maximum likelihood estimators $\hat{p}_j^{MLE}$ are solved from

$$\begin{aligned} \max_{p_1, \ldots, p_m} \quad & \ell(p_1 \ldots, p_m) \\ \text{s.t.} \quad & \sum_j p_j = 1 \\ & p_j > 0. \end{aligned}$$

The above programming can be easily solved by the Lagrangian method. We first define the Lagrangian as

$$L(p_1,\ldots,p_m,\lambda)=\ell(p_1,\ldots,p_m)-\lambda(\sum_j p_j-1). \tag{33}$$

The FOC leads to

$$\begin{cases} \frac{\partial L}{\partial p_j}=\frac{\sum_{i=1}^n X_i^j}{p_j}-\lambda\overset{\text{set}}{=}0,\ j=1,\ldots,m \\ \frac{\partial L}{\partial\lambda}=-(\sum_j p_j-1)\overset{\text{set}}{=}0 \end{cases}. \tag{34}$$

After solving the above system of equations for all p_j and λ and also noting that $\sum_{j=1}^m\sum_{i=1}^n X_i^j=\sum_{i=1}^n\sum_{j=1}^m X_i^j=n$, we obtain

$$\hat{p}_j^{MLE}=\frac{\sum_{i=1}^n X_i^j}{n},\ j=1,\ldots,m. \tag{35}$$

◊ 以上两道涉及discrete distribution的MLE题型, 最重要的是在likelihood function定义式$\mathcal{L}=\prod_{i=1}^n\mathbb{P}(X_i)$中, 如何通过未知参数(如$p_j$)与样本实现$X_i$之间的联系来正确表达$\mathbb{P}(X_i)$, 这是需要花点心思具体问题具体分析的, 如 in Exercise 3 $\mathbb{P}(X_i)=\sum_k p_k\mathbb{1}_{\{X_i=a_k\}}$ while in Exercise 4 就变成了$\mathbb{P}(X_i)=\prod_j p_j^{X_i^j}$. 另外注意log-likelihood的应用.

End.