

“最大似然估计法”教学设计

刘晓燕, 李沫, 刘波

海军航空大学, 山东 烟台

收稿日期: 2021年10月3日; 录用日期: 2021年11月1日; 发布日期: 2021年11月8日

摘要

课堂教学中不但要注重基本理论的论述及推导, 更应注重理论部分的引入和应用展示。在课堂教学中通过对实际应用案例的讲解, 不但能够有效激发学生的学习热情, 更能帮助他们进一步理解相关知识要点。

关键词

概率统计, 最大似然估计法, 教学设计

Instructional Design on the Maximum Likelihood Estimation

Xiaoyan Liu, Mo Li, Bo Liu

Naval Aviation University, Yantai Shandong

Received: Oct. 3rd, 2021; accepted: Nov. 1st, 2021; published: Nov. 8th, 2021

Abstract

In the classroom teaching, we should not only pay attention to the discussion and deduction of the basic theory, but also pay attention to the introduction and application of the theoretical part and the ideological. In the classroom teaching, the explanation of practical application cases can not only effectively stimulate students' enthusiasm for learning, but also help them to understand the key points of relevant knowledge.

Keywords

Probability and Statistics, Maximum Likelihood Estimation, Instructional Design

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在《概率论与数理统计》中，最大似然估计法是在学生已经学习了点估计中的矩估计法之后学习的知识点。矩估计法的优点非常突出：它简单易行，事先并不需要知道总体服从什么分布，就可以求出未知参数的矩估计。但对应的缺点也很明显：当总体类型已知时，不能充分利用总体分布所提供的信息，因而对某些总体的未知参数，矩估计并不合理[1] [2] [3] [4]。对此，研究其它进行点估计的方法是很有必要的。最大似然估计法是求解参数点估计的一个重要方法，该方法具有很多优良的统计性质，因而是统计学中一种重要且应用广泛的方法，但同时也是学生较难理解的概念之一。

2. 最大似然原理

最大似然估计法源于人们的习惯思维。如果我们遇到这样一个问题：“飞将军李广一日无事，与一副将外出狩猎。忽闻雁叫声声，两人同时弯弓射雁，应声而落。副将纵马视之，雁唯中一箭，惑之：吾中呼？将中呼？”大家觉得是谁射中的呢？分析：我们都知道飞将军李广是西汉赫赫有名的边塞英雄，以箭术高超著称，这一点从唐代诗人卢纶的《塞下曲》：“平明寻白羽，没在石棱中”可窥一斑。因此，我们一般会认为这只雁是飞将军的射中的，这样更符合情理。

又如当机器发生故障，有经验的修理工总先从易损部件、薄弱环节查起；公安人员在侦破案件时，也总先将与受害者来往密切且又有作案可能性的人列为重点怀疑对象等等。这些例子尽管千差万别，但他们具有一个共同的规律，那就是对于不确定性的事件，在一次试验中，人们更愿意相信“概率最大”的事件会发生，这种想法非常自然。

但对有些问题，仅从直观上并不容易做出判断，例如这一战士打靶问题。

战士打靶问题：已知战士甲和战士乙命中靶心的概率分别为 0.9 及 0.4。今有一张靶纸上的弹着点表明 10 枪 6 中，已知这张靶纸肯定是甲或乙之一所射，问最有可能是谁射的呢？

从直观上看，战士甲命中靶心的概率为 0.9，枪法属上乘，不至于打得那么差；而以战士乙的枪法又似乎尚不足以打出这么好的成绩。但这张靶纸的确是甲或乙射出来的，二者中取其一，更像谁射的呢？直观上不好判断，我们就来定量地分析一下。这里的定量分析，其实其底层思维方式与之前的问题实质是一样的，那就是通过计算比较二人分别射出这样成绩的概率。

分析：根据已知条件，这张靶纸为甲所射的概率为

$$P_1 = (0.9)^6 (1-0.9)^4 \approx 0.00005.$$

这张靶纸为乙所射的概率为

$$P_2 = (0.4)^6 (1-0.4)^4 \approx 0.0005.$$

显然 $P_1 < P_2$ ，甲的概率小于乙的概率，乙射出这个成绩的概率为甲的十倍，即认为这张靶纸为乙所射的可能性更大。

不论从直观上判断，还是根据数据进行定量地分析，我们得到这些结论在概率论中都是有理论依据的，那就是人们在长期的实践中总结得到的“实际推断原理”，即：概率很小的事件在一次试验中实际上几乎是不发生的。也就是说，在一次试验中就发生的事件是大概率事件。这也是最大似然原理的基础：

在一次试验中，概率大的事件比概率小的事件易于发生。

基于此就诞生了最大似然估计法。

下面我们就离散型总体的情形和连续型总体的情形分别做一下讨论。

3. 离散型总体的情形

定义 1 若总体 X 属离散型，其分布律 $P\{X = x\} = p(x; \theta)$ ， $\theta \in \Theta$ 的形式为已知，其中 θ 为待估参数， Θ 是 θ 可能取值的范围。设 X_1, X_2, \dots, X_n 是来自总体 X 的样本，则 X_1, X_2, \dots, X_n 的联合分布律为

$$\prod_{i=1}^n p(x_i; \theta)。$$

又设 x_1, x_2, \dots, x_n 是相应于样本 X_1, X_2, \dots, X_n 的一个样本值。易知样本 X_1, X_2, \dots, X_n 取到观察值 x_1, x_2, \dots, x_n 的概率，即事件 $\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 发生的概率为

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta), \theta \in \Theta。$$

这一概率随 θ 的取值而变化，它是 θ 的函数， $L(\theta)$ 称为样本的似然函数。

取 $\hat{\theta}$ ，使得

$$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)，$$

这样得到的 $\hat{\theta}$ 与样本值 x_1, x_2, \dots, x_n 有关，常记为 $\hat{\theta}(x_1, x_2, \dots, x_n)$ ，称为参数 θ 的最大似然估计值，而相应的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 称为参数 θ 的最大似然估计量[1]。

定义 1 说明：

1) 定义中的 X_1, X_2, \dots, X_n 是来自总体 X 的样本，这里的样本全称为“简单随机样本”，其定义告诉我们它具有两条重要的性质。其一：独立性，即 X_1, X_2, \dots, X_n 是相互独立的，故积事件 $\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 发生的概率就等于它们的概率之积

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = P\{X_1 = x_1\} P\{X_2 = x_2\} \cdots P\{X_n = x_n\}。$$

其二：代表性，即样本 X_1, X_2, \dots, X_n 的分布律与总体是一致的，故可用总体分布律的值来表示相应的概率

$$P\{X_i = x_i\} = p(x_i; \theta), i = 1, 2, \dots, n。$$

因此

$$\begin{aligned} P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} &= P\{X_1 = x_1\} P\{X_2 = x_2\} \cdots P\{X_n = x_n\} \\ &= p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta) \\ &= \prod_{i=1}^n p(x_i; \theta)。 \end{aligned}$$

可见，为表示这一积事件的概率，我们充分地使用到了总体的分布律，在利用已知的总体分布信息方面较矩估计有了很大改善。

2) 上述积事件的概率能够表示“样本 X_1, X_2, \dots, X_n 取到观察值 x_1, x_2, \dots, x_n 的概率”，并且这一概率只随 θ 的取值而变化，它是 θ 的函数，我们把这个能够反映试验结果发生概率大小，且可由未知参数表示的函数定义为样本的似然函数

$$L(\theta) = P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = \prod_{i=1}^n p(x_i; \theta), \theta \in \Theta。$$

样本取到其观察值这一事件已经发生了，基于最大似然原理，则必然认为这一事件发生的概率是各种情况中概率最大的。接下来只需在未知参数 θ 的可能取值范围 Θ 内，挑选使得似然函数取得最大值的参数值 θ ，就可作为未知参数的估计值了，相应的统计量则为 θ 的最大似然估计量。

3) 最大似然估计法是由“Maximum likelihood estimation”翻译而来的，有的资料中也翻译成“极大似然估计法”。“似然”是对“likelihood”的一种较为贴近文言文的翻译，似然函数的记法“ L ”就来源于其首字母。“likelihood”用现代的中文翻译就是“可能性”“看起来像”的意思，因此若称之为“最大可能性估计”则更为通俗易懂。

最大似然估计法首先由德国数学家 Gauss 在 1821 年提出，但是这个方法通常归功于英国的统计学家 Fisher。Fisher 是经典数理统计学派的集大成者，他在 1912 年再次提出了这个思想，并且首先探讨了这种方法的一些性质，使得最大似然估计法得到了广泛的应用，“最大似然估计”这一名称也是 Fisher 给出的。当年 Fisher 正是凭借这一方法彻底颠覆了矩估计法提出者皮尔逊在统计学界的统治地位，也正是从这一时期开始，统计学进入了 Fisher 时代，由此可见最大似然估计法的提出在统计学界具有重要的理论意义。

最大似然估计法是在总体分布类型已知的条件下，充分利用这一已知条件，根据最大似然原理，由样本的观测值，对总体分布的未知参数作出估计的方法，是一类完全基于统计数据来确定分布情况的代表方法。因此，要进行未知参数的最大似然估计，前提条件是：总体分布已知；样本观测值已知。

4) 由定义不难总结出使用最大似然估计法的关键两步。第一步写出待估参数的函数——似然函数。这里的待估参数可能是一个，也可能是多个，无论是一个还是多个，构造似然函数的原理和方法都是一样的。第二步就是求似然函数的最大值点。由最大似然原理知，求出的最大值点就是待估参数的最大似然估计值。

下面使用离散型总体的最大似然估计法求解“紧急呼叫问题”。

例 1 某应急部门一天内接到的呼叫次数 $X \sim \pi(\lambda)$ ，现有 42 天的数据如下表 1，试估计参数 λ 的值用于应急方案的制定。

Table 1. Number of calls and their frequency

表 1. 呼叫次数及其频数

呼叫次数	0	1	2	3	4	5	>5
出现的频数	7	10	12	8	3	2	0

分析：在这一问题中，总体的分布形式已知，为 Poisson 分布；样本的观测值也已经在上表 1 中给出；要进行估计的参数只有一个 λ 。由最大似然估计法步骤：首先写出似然函数；然后求解似然函数的最大值点。在求解最值点时，可借助“取对数”的运算技巧来简化计算。

解：由已知，总体 $X \sim \pi(\lambda)$ ， X 的分布律为：

$$P\{X = x\} = \frac{\lambda^x}{x!} e^{-\lambda}, x = 0, 1, \dots$$

设 x_1, x_2, \dots, x_n 为来自总体的样本值，则似然函数

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

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

$$\text{令 } \frac{d[\ln L(\lambda)]}{d\lambda} = \frac{\sum_{i=1}^n x_i}{\lambda} - n = 0，$$

$$\text{解得 } \lambda \text{ 的最大似然估计值为 } \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}。$$

代入数值得 $\hat{\lambda} = 1.905$ 。

到这里研究的问题都是针对离散型总体而言的，若总体的分布是连续型的又该如何呢？

4. 连续型总体的情形

定义 2 若总体 X 属连续型，其概率密度 $f(x; \theta)$ ， $\theta \in \Theta$ 的形式已知，其中 θ 为待估参数， Θ 是 θ 可能取值的范围。设 X_1, X_2, \dots, X_n 是来自 X 的样本， x_1, x_2, \dots, x_n 是相应于样本 X_1, X_2, \dots, X_n 的一个样本值，样本的似然函数

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta), \theta \in \Theta。$$

取 $\hat{\theta}$ ，使得

$$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)，$$

则称 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为参数 θ 的最大似然估计值，称 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为 θ 的最大似然估计量[1]。

定义 2 说明：

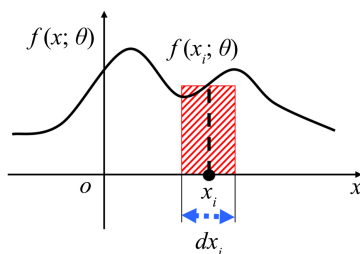


Figure 1. Probability analysis
图 1. 概率分析

1) 当总体 X 属连续型时，谈所谓样本 X_1, X_2, \dots, X_n 取样本值 x_1, x_2, \dots, x_n 的概率是没有意义的，因为连续型随机变量取任意确定值的概率均为 0，即任何具体样本值出现的概率都是 0。因此这里需利用极限的思想，考虑样本在包含样本值的任意小的邻域中出现的概率(图 1)，即用 $f(x_i; \theta) dx_i$ 来反映相应样本值出现的可能性大小，再由样本的独立同总体分布的性质可得：随机点 (X_1, X_2, \dots, X_n) 落在含点 (x_1, x_2, \dots, x_n) 的边长分别为 dx_1, dx_2, \dots, dx_n 的 n 维立方体内的概率近似为

$$\prod_{i=1}^n f(x_i; \theta) dx_i = \prod_{i=1}^n f(x_i; \theta) \prod_{i=1}^n dx_i，$$

其值随 θ 的取值而变化。与离散型的情况类似，只需取 θ 的估计值 $\hat{\theta}$ 使这个概率取到最大值即可。注意到，因子 $\prod_{i=1}^n dx_i$ 不随 θ 而变，因此选取 θ 的估计值，使得 $\prod_{i=1}^n f(x_i; \theta) dx_i$ 最大，就等价于使 $\prod_{i=1}^n f(x_i; \theta)$ 最

大，而后一式子显然更简单一些。

因此在连续型总体的情况下，样本的似然函数可取作这样 n 个概率密度函数值的乘积

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta), \theta \in \Theta.$$

接下来就是在 θ 的可能取值范围内，挑选使得似然函数取得最大值的参数值作为未知参数的估计值，而相应的统计量则为参数的最大似然估计量。这些就与离散型总体一致了。

2) 通过比较不难发现，连续型总体与离散型总体的最大似然估计法的关键步骤是相同的：首先写出似然函数；然后求解似然函数的最大值点。不同点就在于似然函数的形式：离散型对应的是分布律的乘积；连续型是概率密度的乘积。但它们都充分利用了已知的总体分布信息。

3) 最大似然估计的性质(不变性)：设 θ 的函数 $u = u(\theta)$ ， $\theta \in \Theta$ 具有单值反函数 $\theta = \theta(u)$ ， $u \in U$ 。又设 $\hat{\theta}$ 是 X 的概率分布中参数 θ 的最大似然估计，则 $u = u(\hat{\theta})$ ， $\theta \in \Theta$ 是 $u(\theta)$ 的最大似然估计。当总体分布中含有多个未知参数时，也具有上述性质[1]。

最大似然估计法是目前应用最为广泛的一种点估计法，基于最大似然估计的算法在雷达探测领域、数字信号处理方面，以及遥控水下航行器的目标跟踪等技术中都起着无可取代的作用，现以较简单的水雷对目标的判断识别为例进行讨论。

水雷是布设在水中的爆炸性武器，现代水雷多采用利用目标自身产生的辐射噪声工作的被动声引信，这种引信不需暴露雷位，更符合水雷隐蔽打击的特点。水雷上通常装有声接收器，相当于水雷的耳朵，当接收器接收到目标的噪声后，会将声波变成电信号，通过识别系统判断有没有超出预先设置的电信号阈值，若超过就会引爆水雷攻击目标。因此，设置电信号阈值非常关键。而在这一过程中，首先得对测得的噪声进行分析和识别。

舰船航行时，由于其机械的运转和航行时与水流的接触，不可避免的会产生噪声。在水下环境中，噪声的产生和组成是及其复杂的，工程上一般认为，在一段时间内测量得到的舰船噪声服从正态分布，只不过不同型号的舰船对应的参数是有区别的。因此，就需要根据已知样本的统计特征，采用适当有效的估计目标参数的方法，来对未知类别的样本进行分类。

例 2 工程上一般认为，一段时间内测量得到的舰船噪声 $X \sim N(\mu, \sigma^2)$ ，不同型号的舰船所对应的参数是不同的。现水下测量系统测得某舰船的噪声值(dB)如下表 2，试估计其参数。

Table 2. Noise value of a ship (dB)

表 2. 某舰船的噪声值(dB)

152	151	152	155	151	155	151	158	157	157
150	156	156	157	153	151	152	150	155	150

设置水雷声引信动作参数(水雷声引信动作所需的最小噪声值) L 时，需考虑目标漏检(目标噪声小于 L)的情况，若要求漏检的概率不超过 0.005，试利用估计出的参数确定 L 的最大似然估计值。

解： $X \sim N(\mu, \sigma^2)$ ， μ ， σ^2 均未知，设 x_1, x_2, \dots, x_n 是测量得到的一个样本值，则似然函数

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2},$$

$$\text{又 } \ln L(\mu, \sigma^2) = -\frac{n}{2} (\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

$$\text{令} \begin{cases} \frac{\partial}{\partial \mu}(\ln L) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \hat{=} 0 \\ \frac{\partial}{\partial \sigma^2}(\ln L) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \hat{=} 0 \end{cases}$$

解得 $\mu = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$, $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 。

带入数值得 μ , σ^2 的最大似然估计值分别为:

$$\hat{\mu} = 153.45(\text{dB}); \hat{\sigma}^2 = 7.25(\text{dB}^2)。$$

进一步, 根据设计要求, 应有 $P\{X < L\} \leq 0.005$, 即

$$P\left\{\frac{X - \mu}{\sigma} < \frac{L - \mu}{\sigma}\right\} \leq 0.005, \text{ 其中 } \frac{X - \mu}{\sigma} \sim N(0, 1)。$$

由分位点定义有 $\frac{L - \mu}{\sigma} \leq -u_{0.005}$, 故 $L \leq \mu - u_{0.005}\sigma$ 。

取 $L = \mu - u_{0.005}\sigma$, 根据最大似然估计的不变性 $\hat{L} = \hat{\mu} - u_{0.005}\sqrt{\hat{\sigma}^2}$ 。

代入数据计算得 $\hat{L} = 146.5(\text{dB})$ 。

5. 小结与思考

在本次课的教学环节设计中, 针对最大似然估计法计算复杂、学生理解较为困难的问题, 根据人们思维习惯由直观到抽象的特点, 首先给出两个简单的例子让学生从直观上去使用最大似然估计原理做出判断, 这样对于接下来较为抽象的理论有较好的引导作用, 不仅可以激发学生的学习兴趣, 而且还避免因直接给出抽象复杂的理论给学生带来困惑。接下来分别从“离散型总体”和“连续型总体”两方面介绍最大似然估计法的基本思想、原则和解题步骤, 并在其中融入基本的统计思想, 引导学生建立统计模型, 从而归纳出最大似然估计法的基本步骤。最后, 联系军事实际问题, 并对其详细求解。例题的设计非常关键, 规范解题步骤的同时融入基本思想, 通过两个例题让学生能够进一步理解最大似然估计法这一概念。在案例选择的时候要能说明问题, 而不是越复杂越好。由对统计模型的求解过程完成从理论到案例的过渡, 使学生明确学有所用, 培养学生运用数学方法分析和解决实际问题的能力。

参考文献

- [1] 盛骤, 等. 概率论与数理统计[M]. 第五版. 北京: 高等教育出版社, 2019.
- [2] 同济大学应用数学系. 概率统计简明教程[M]. 北京: 高等教育出版社, 2003.
- [3] Sheldon M. Ross. 概率论基础教程[M]. 北京: 人民邮电出版社, 2010.
- [4] 赵鲁涛, 等. 概率论与数理统计教学设计[M]. 北京: 机械工业出版社, 2015.