6.1 最大似然估计与 EM 算法

2. 设总体概率函数如下，$x_1, x_2, \cdots, x_n$ 是样本，试求未知参数的最大似然估计。

(1) $p(x;\theta) = c\theta^c x^{-(c+1)}, x > \theta, \theta > 0, c > 0$ 已知；

解：

对数似然函数

$$
\begin{aligned}
\ln L(\theta) &= \ln \prod_{i=1}^{n} p(x_i|\theta) = \ln \prod_{i=1}^{n} c\theta^c x_i^{-(c+1)} \\
&= \sum_{i=1}^{n} \ln(c\theta^c x_i^{-(c+1)}) = \sum_{i=1}^{n} (\ln c + c\ln\theta - (c+1)\ln x_i) \\
&= n\ln c + nc\ln\theta - (c+1)\sum_{i=1}^{n} \ln x_i
\end{aligned}
$$

只需要让 θ 尽量大即可使似然函数取到最大值，又因为 $\theta < x$，所以 θ 的最大似然估计为 $\widehat{\theta} = x_{(1)}$。 □

(2) $p(x;\theta,\mu) = \frac{1}{\theta} e^{-\frac{x-\mu}{\theta}}, x > \mu, \theta > 0$；

解：

对数似然函数

$$
\begin{aligned}
\ln L(\theta,\mu) &= \ln \prod_{i=1}^{n} p(x;\theta,\mu) = \ln \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x-\mu}{\theta}} \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\theta} e^{-\frac{x-\mu}{\theta}}\right) = \sum_{i=1}^{n} (-\ln\theta - \frac{x-\mu}{\theta}) \\
&= -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i + \frac{n\mu}{\theta}
\end{aligned}
$$

对于 μ，由于 $\ln L(\theta,\mu)$ 关于 μ 是线性关系，所以只需要 μ 尽量大即可使似然函数取到最大值，而 $\mu < x$，所以 $\widehat{\mu} = x_{(1)}$。

对于 θ，则需要求偏导，令

$$
\frac{\partial \ln L(\theta,\mu)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i - \frac{n\mu}{\theta^2} = 0
$$

则可解得 $\theta = \frac{1}{n}\sum_{i=1}^{n} x_i - \mu = \bar{x} - \mu$。此时 $\ln L(\theta,\mu)$ 关于 θ 最大。

所以 $\widehat{\mu} = x_{(1)}$, $\widehat{\theta} = \bar{x} - x_{(1)}$。 □

(3) $p(x;\theta)=(k\theta)^{-1},\theta<x<(k+1)\theta,\theta>0,k>0$ 已知。

解：

对数似然函数

$$\ln L(\theta)=\ln\prod_{i=1}^{n}(k\theta)^{-1}=\sum_{i=1}^{n}\ln(k\theta)^{-1}=\sum_{i=1}^{n}(-k\theta)=-nk\theta$$

只要 θ 尽量小即可使似然函数取得最大值。由于 $\theta<x<(k+1)\theta$ 且 $k>0$，所以 $\frac{\theta}{k+1}<\frac{x}{k+1}<\theta$，所以 θ 的最大似然估计为 $\widehat{\theta}=\dfrac{x_{(n)}}{k+1}$。 □

4. 一地质学家为研究密歇根湖的湖滩地区的岩石成分，随机地自该地区取 100 个样品，每个样品有 10 块石子，记录了每个样品中属石灰石的石子数。假设这 100 次观察柜互独立，求这地区石子中石灰石的比例 p 的最大似然估计。该地质学家所得的数据如下：

样本中的石子数	0	1	2	3	4	5	6	7	8	9	10
样品个数	0	1	6	7	23	26	21	12	3	1	0

解：

当已知石灰石的比例为 p 时，并且如果每次抽样都是随机抽样，那么每个石子是石灰石的概率就是 p，由于每个样品有 10 块石子，所以一次抽样服从二项分布 $b(10,p)$，则概率函数为

$$p(k;p)=\mathrm{C}_{10}^{k}p^{k}(1-p)^{10-k}$$

设表格中的第一行为 $x_i(i=0,1,\cdots,10)$，第二行为 $a_i(i=0,1,\cdots,10)$，则对数似然函数为

$$\begin{aligned}\ln L(p)&=\ln\prod_{i=1}^{n}\left(\mathrm{C}_{10}^{x_i}p^{x_i}(1-p)^{10-x_i}\right)^{a_i}\\&=\sum_{i=1}^{n}a_i\left(\ln\mathrm{C}_{10}^{x_i}+x_i\ln p+(10-x_i)\ln(1-p)\right)\\&=\sum_{i=1}^{n}a_i\ln\mathrm{C}_{10}^{x_i}+\ln p\sum_{i=1}^{n}a_ix_i+\ln(1-p)\sum_{i=1}^{n}a_i(10-x_i)\end{aligned}$$

令

$$\frac{\mathrm{d}\ln L(p)}{\mathrm{d}p}=\frac{\sum_{i=1}^{n}a_ix_i}{p}-\frac{\sum_{i=1}^{n}a_i(10-x_i)}{1-p}=0$$

解得

$$p=\frac{\sum_{i=1}^{n}a_ix_i}{10\sum_{i=1}^{n}a_i}=\frac{\sum_{i=1}^{n}a_i\frac{x_i}{10}}{\sum_{i=1}^{n}a_i}$$

即以样品个数为权重，样品中石灰石比例的加权平均值。

所以

$$\widehat{p}=\frac{\sum_{i=1}^{n}a_ix_i}{10\sum_{i=1}^{n}a_i}=\frac{\begin{array}{c}0\times0+1\times1+6\times2+7\times3+23\times4+26\times5\\+21\times6+12\times7+3\times8+1\times9+0\times10\end{array}}{10\times100}=0.499$$

□

5. 在遗传学研究中经常要从截尾二项分布中抽样，其总体概率函数为

$$p(X=k;p)=\frac{\binom{m}{k}p^k(1-p)^{m-k}}{1-(1-p)^m},\quad k=1,2,\cdots,m$$

若已知 $m=2,x_1,x_2,\cdots,x_n$ 是样本，试求 p 的最大似然估计。

解：

对数似然函数为

$$\begin{aligned}\ln L(p)&=\ln\prod_{i=1}^{n}\frac{\binom{m}{x_i}p^{x_i}(1-p)^{m-x_i}}{1-(1-p)^m}\\&=\sum_{i=1}^{n}\left[\ln\binom{m}{x_i}+x_i\ln p+(m-x_i)\ln(1-p)-\ln(1-(1-p)^m)\right]\\&=\sum_{i=1}^{n}\ln\binom{m}{x_i}+\ln p\sum_{i=1}^{n}x_i+\ln(1-p)\sum_{i=1}^{n}(m-x_i)-n\ln(1-(1-p)^m)\end{aligned}$$

令

$$\frac{\mathrm{d}\ln L(p)}{\mathrm{d}p}=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{\sum_{i=1}^{n}(m-x_i)}{1-p}-n\frac{-m(1-p)^{m-1}}{1-(1-p)^m}=0$$

由于 $m=2$，所以

$$\frac{\sum_{i=1}^{n}x_i}{p}-\frac{\sum_{i=1}^{n}(2-x_i)}{1-p}+\frac{2n(1-p)}{1-(1-p)^2}=0$$

即

$$\frac{n\bar{x}}{p}-\frac{2-n\bar{x}}{1-p}+\frac{2n(1-p)}{1-(1-p)^2}=0$$

解得 p 的最大似然估计为

$$\widehat{p}=\frac{\bar{x}n+4n+4}{4(n+1)}\pm\frac{\sqrt{\bar{x}^2n^2-8\bar{x}n^2-8\bar{x}n+16n+16}}{4(n+1)}$$

□

6. 已知在文学家萧伯纳的"The Intelligent Woman's Guide to Socialism and Capitalism" 一书中，一个句子的单词数 X 近似地服从对数正态分布，即 $Z=\ln X\sim N(\mu,\sigma^2)$。今从该书中随机地取 20 个句子，这些句子中的单词数分别为

52 24 15 67 15 22 63 26 16 32 7 33 28 14 7 29 10 6 59 30

求该书中一个句子单词数均值 $E(X)=e^{\mu+\frac{\sigma^2}{2}}$ 的最大似然估计。

解：

□

根据题意，由于 $Z=\ln X\sim N(\mu,\sigma^2)$，可以将一个句子的单词数先取自然对数，此时即可使用正态分布的最大似然估计来估计 μ 和 σ^2。

```
import numpy as np

a = np.array([52,24,15,67,15,22,63,26,16,32,7,33,28,14,7,29,10,6,59,30])
print(np.log(a))
# [3.95124372 3.17805383 2.7080502  4.20469262 2.7080502  3.09104245
#  4.14313473 3.25809654 2.77258872 3.4657359  1.94591015 3.49650756
#  3.33220451 2.63905733 1.94591015 3.36729583 2.30258509 1.79175947
#  4.07753744 3.40119738]

print(np.mean(np.log(a)))
# 3.0890326915239807
print(np.var(np.log(a)))
# 0.5081312851436304
```

所以 $\widehat{\mu} \approx 3.0890326915239807, \widehat{(\sigma^2)} \approx 0.5081312851436304$。

再根据最大似然估计的不变性，直接计算 $e^{\widehat{\mu}+\frac{\widehat{(\sigma^2)}}{2}}$。

```
np.exp(np.mean(np.log(a)) + np.var(np.log(a)) / 2)
# 28.306694575039742
```

则该书中一个句子单词数均值 $E(X) = e^{\mu+\frac{\sigma^2}{2}}$ 的最大似然估计约为 28.306694575039742。

7. 设总体 $X \sim U(\theta, 2\theta)$，其中 $\theta > 0$ 是未知参数，$x_1, x_2, \cdots, x_n$ 为取自该总体的样本，$\bar{x}$ 为样本均值。

(1) 证明 $\widehat{\theta} = \frac{2}{3}\bar{x}$ 是参数 θ 的无偏估计和相合估计；

证明：

$$E\widehat{\theta} = E\frac{2}{3}\bar{x} = E\frac{2}{3}\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{2}{3}\frac{1}{n}\sum_{i=1}^{n} EX = \frac{2}{3}\frac{1}{n}n\frac{\theta+2\theta}{2} = \theta$$

$$\operatorname{Var}\widehat{\theta} = \operatorname{Var}\frac{2}{3}\bar{x} = \frac{2}{3}\frac{n\operatorname{Var}X}{n^2} = \frac{2\operatorname{Var}X}{3n} \xrightarrow{n\to\infty} 0$$

所以 $\widehat{\theta} = \frac{2}{3}\bar{x}$ 是参数 θ 的无偏估计和相合估计。 □

(2) 求 θ 的最大似然估计，它是无偏估计吗？是相合估计吗？

解：

$$\ln L(\theta) = \ln\prod_{i=1}^{n} 1_{[\theta,2\theta]}(x_i)\cdot\frac{1}{\theta} = \frac{1}{\theta}\sum_{i=1}^{n}\ln 1_{[\theta,2\theta]}(x_i)$$

要使似然函数最大，则需要 θ 尽量小，同时要满足 $\theta \leqslant x_i \leqslant 2\theta$，即 $\frac{\theta}{2} \leqslant \frac{x_i}{2} \leqslant \theta$，所以 θ 的最大似然估计为 $\widehat{\theta} = \dfrac{x_{(n)}}{2}$。

下面验证无偏性。

$$E\widehat{\theta} = \frac{1}{2}\int_{\theta}^{2\theta} x\frac{n}{\theta}\left(\frac{x-\theta}{\theta}\right)^{n-1}\mathrm{d}x = \frac{\theta(2n+1)}{2(n+1)} \xrightarrow{n\to\infty} \theta$$

所以 $\widehat{\theta}$ 不是无偏估计，但是是渐近无偏估计。

下面验证相合性。

$$E\widehat{\theta}^2 = \frac{1}{4}\int_{\theta}^{2\theta} x^2\frac{n}{\theta}\left(\frac{x-\theta}{\theta}\right)^{n-1}\mathrm{d}x = \frac{\theta^2(n^2+2n+\frac{1}{2})}{n^2+3n+2}$$

$$\operatorname{Var}\widehat{\theta} = E\widehat{\theta}^2 - (E\widehat{\theta})^2 = \frac{n\theta^2}{4(n^3+4n^2+5n+2)} \xrightarrow{n\to\infty} 0$$

所以 $\widehat{\theta}$ 是相合估计。 □

8. 设 $x_1, x_2, \cdots, x_n$ 是来自密度函数为 $p(x;\theta) = e^{-(x-\theta)}, x > \theta$ 的总体的样本。

(1) 求 θ 的最大似然估计 $\widehat{\theta}_1$，它是否是相合估计？是否是无偏估计？

解：

$$\ln L(\theta) = \ln\prod_{i=1}^{n} e^{-(x-\theta)} = \sum_{i=1}^{n}(-(x_i-\theta)) = -\sum_{i=1}^{n} x_i + n\theta$$

要让似然函数最大，θ 要尽量大，同时 $\theta < x$，所以 θ 的最大似然估计为 $\widehat{\theta} = x_{(1)}$。

$\widehat{\theta} = x_{(1)}$ 的概率函数为

$$p(x) = n\left[1 - \int_{\theta}^{x} e^{-(t-\theta)}\mathrm{d}t\right]^{n-1} e^{-(x-\theta)} = n(e^{\theta-x})^n$$

则可以验证无偏性

$$E\widehat{\theta}_1 = \int_{\theta}^{+\infty} xn(e^{\theta-x})^n\mathrm{d}x = \frac{1}{n} + \theta \xrightarrow{n\to\infty} \theta$$

所以 $\widehat{\theta}_1$ 不是无偏估计，但是是渐近无偏估计。

下面验证相合性。

$$E\widehat{\theta}_1^2 = \int_{\theta}^{+\infty} x^2 n(e^{\theta-x})^n\mathrm{d}x = \frac{2}{n^2} + \frac{2}{n}\theta + \theta^2$$

$$\operatorname{Var}\widehat{\theta}_1 = E\widehat{\theta}_1^2 - (E\widehat{\theta})^2 = \frac{2}{n^2} + \frac{2}{n}\theta + \theta^2 - \left(\frac{1}{n}+\theta\right)^2 = \frac{1}{n^2} \xrightarrow{n\to\infty} 0$$

所以 $\widehat{\theta}_1$ 是相合估计。 □

(2) 求 θ 的矩估计 $\widehat{\theta}_2$，它是否是相合估计？是否是无偏估计？

解：

$$EX = \int_{\theta}^{+\infty} x e^{-(x-\theta)} \mathrm{d}x = \theta + 1$$

所以 $\widehat{\theta}_2 = 1 - \bar{x}$。

$$E\widehat{\theta}_2 = E(1-\bar{x}) = 1 - EX = \theta$$

所以 $\widehat{\theta}_2$ 是无偏估计。

$$\operatorname{Var}\widehat{\theta}_2 = \operatorname{Var}(1-\bar{x}) = \frac{\operatorname{Var}X}{n} \xrightarrow{n\to\infty} 0$$

所以 $\widehat{\theta}_2$ 是相合估计。 □

10. 证明：对正态分布 $N(\mu,\sigma^2)$，若只有一个观测值，则 μ,σ^2 的最大似然估计不存在。

证明：

设此观测值为 x，则似然函数为

$$L(\mu,\theta) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$$

要使似然函数最大，则 $\left(\frac{x-\mu}{\sigma}\right)^2$ 应尽量小，则 $\frac{(x-\mu)^2}{\sigma^2} \to 0$，所以 $\mu = x, \sigma^2 = \infty$，由于 $\infty \notin \mathbb{R}$，所以 μ,σ^2 的最大似然估计不存在。 □