极大似然估计

目录

1 似然函数与极大似然估计

假定$\mathbf{X} \sim P_\theta, \theta \in \Theta$，联合密度为$f(\mathbf{x}|\theta)$。假定我们观察到样本为$\mathbf{X} = \mathbf{x}$。则对应该样本集合的似然函数为：

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

由于样本已知，则该似然函数我们可以看做是关于θ的函数。
如果样本$x_1, x_2, x_3 \ldots\ldots x_n$独立同分布，服从于$f(x|\theta)$，则此时似然函数可以改写为：

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

点估计$\hat{\theta} = \hat{\theta}(x)$是极大似然估计即需要满足：

$$L(\hat{\theta}|x) = \sup \ L(\theta|x)$$

即相当于$\hat{\theta}$最大化该似然函数。一般情况下，似然函数的最大值对应的$\hat{\theta}$唯一，由此可以得到唯一的极大似然估计值：

$$\hat{\theta}(x) = \mathrm{argmax}_\theta L(\theta|x)$$

然而，如果当似然函数在一个平面上同时达到最优值，此时极大似然估计不唯一。

2 以正态分布为例

首先，以正态分布为例，我们通过可视化来理解极大似然估计。假定我们存在以下正态分布：

$$X \sim N(u, \sigma^2)$$

$$f(x|u, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-u)^2}{2\sigma^2}\right)$$

当取$u = 7, \sigma^2 = 1$时，其密度函数图像如下所示：

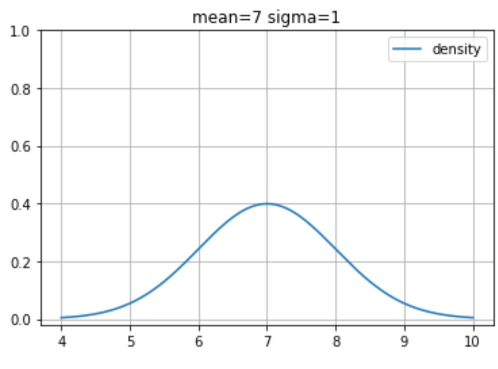

图 1 正态分布密度函数图像

2.1 似然函数

假设我们现在有一个观测值，取值为$x_1 = 6$，则下图中红色点代表其取值，绿色的线代表其概率。此时其似然函数$L(\theta|\mathbf{x}) = L(\theta|x_1 = 3) = f(x_1 = 3|u = 7, \sigma^2 = 1) = 0.24197072451914337$。

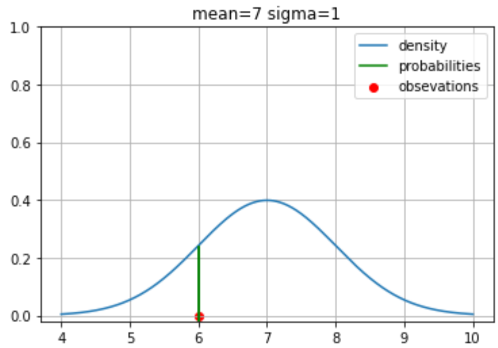

图 2 单个观测值（$u = 7, \sigma^2 = 1$）

如果我们现在有多个观测值（此处为 20 个），那么其在概率密度图上的表现如下。红色的点是每一个观测值，绿色的线为每个观测点出现的概率。

此时其似然函数$L(\theta|\mathbf{x}) = L(\theta|x_1, x_2, \dots, x_{20}) = f(x_1, x_2, \dots, x_{20}|u = 7, \sigma^2 = 1) = 1.734727381395105e - 153$。

似然函数值非常小，我们可以从图中看出，这是由于 0 附近的观测值对应的概率太小所致。

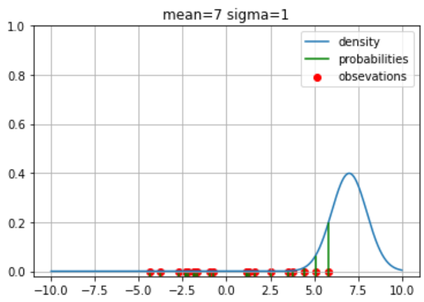

图 3 多个观测值（$u = 7, \sigma^2 = 1$）

所以，接下来，我们采取另一组参数$u = 0, \sigma^2 = 16$生成正态分布密度，重新计算在此组参数下的似然函数值。$L(\theta|\mathbf{x}) = L(\theta|x_1, x_2, \dots, x_{20}) = f(x_1, x_2, \dots, x_{20}|u = 0, \sigma^2 = 16) = 9.426980047109535e - 12$。

可以看到，在该组参数下的似然函数值远远大于上一组。其对应的图像如下。

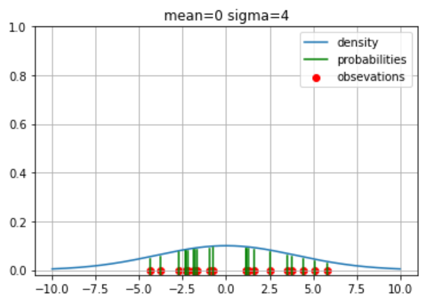

图 4 多个观测值（$u = 0, \sigma^2 = 16$）

从上面两图我们可以得出结论，图 4 的概率分布参数更符合观测到的数据点的概率分布。而极大似然估计的目的就是找到一个最符合当前数据的分布的参数。

2.2 理论结果

从理论的角度，我们可以通过对似然函数$L(\theta|\mathbf{x})$求导，从而得到似然函数的最优值，进一步求得极大似然估计。$(x_1, x_2, \dots, x_n)$的联合对数似然函数为：

$$LnL(\mu,\sigma \mid \boldsymbol{x}) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

两个一阶条件分别为：

$$\frac{LnL}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\frac{LnL}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

可以求出未知参数的估计量分别为：

$$\hat{u} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

接着，我们将通过一个例子来展示该求极大似然估计的过程。

我们从$u = 0, \sigma^2 = 16$的正态分布中随机生成 200 个数据点作为我们的观测数据；

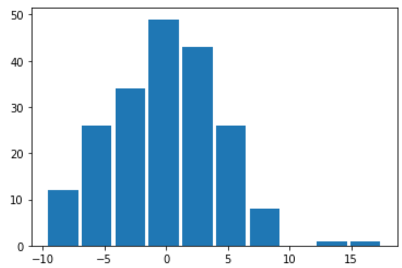

图 5 随机生成的数据

我们首先利用上述公式得到准确的极大似然估计值：

$$\hat{u} = \bar{x} = -0.0719$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = 19.5576$$

2.3 可视化结果

接下来，我们通过可视化来检查我们的极大似然估计结果。我们定义一个待估参数的取值范围，$u \in (-1,1), \sigma^2 \in (9,25)$。我们将该参数组成的网格点一一计算其似然函数值，并 3D 可视化。找似然函数极大值的过程就是极大似然估计的过程。

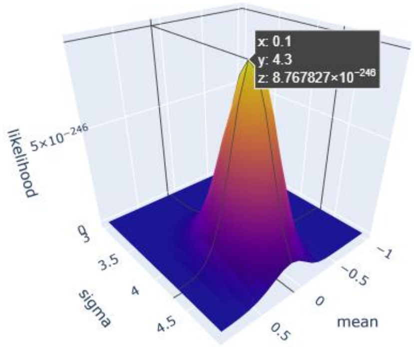

图 6 似然函数可视化

在该图上，我们观察到，似然函数取极大值时，均值和方差大概达到$u=0,\sigma^2=16$。可以看到结果与生成数据几乎一致，也与公式法的结果近似。

上述似然函数未经过对数化，所以我们发现其似然函数值过于小，达到了10^{-253}的量级，这样的数值不适合于数值计算，故必须进行对数化。在进行对数化后，其似然函数的值正常了许多。

以下对对数似然函数进行 3D 可视化。可以看到其最大值与图 6 类似。

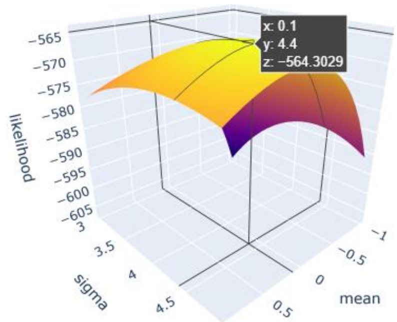

图 7 对数似然函数可视化

3 以指数分布为例

3.1 似然函数

接下来，以指数分布为例。其密度函数为：

$$f(x)=\lambda\exp(-\lambda x),x>0$$

若有 n 个独立同分布的样本，其似然函数为：

$$L(\lambda,x_1,x_2,\dots,x_n)=\lambda^n exp(-\lambda\sum_{j=1}^{n}x_j)$$

对数似然函数为：

$$\ln L(\lambda,x_1,x_2,\dots,x_n)=nIn(\lambda)-\lambda\sum_{j=1}^{n}x_j$$

3.2 理论结果

最大化其极大似然函数得到参数的极大似然估计为：

$$\hat{\lambda}_n = \frac{n}{\sum_{j=1}^{n} x_j}$$

接着，我们将通过一个例子来展示该求极大似然估计的过程。
我们从$\lambda = 1/5$的指数分布中随机生成 200 个数据点作为我们的观测数据；

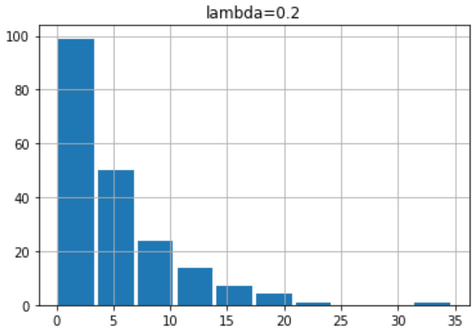

图 8 随机生成的指数分布数据

我们首先利用上述公式得到准确的极大似然估计值：

$$\hat{\lambda}_n = \frac{n}{\sum_{j=1}^{n} x_j} = 0.187$$

可以看出与真实值 0.2 相差不大。

3.3 可视化结果

接下来，我们通过可视化来检查我们的极大似然估计结果。我们定义待估参数的取值范围，$1/\lambda \in (4,6)$。我们每隔 0.1 的步长取一个点，并且一一计算其中各点的似然函数值，之后进行可视化。

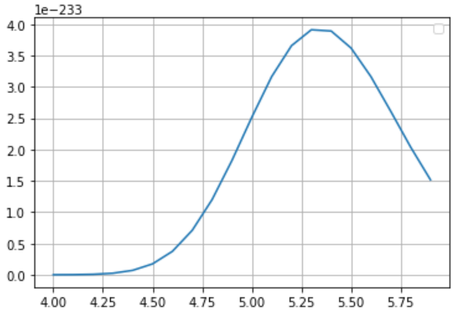

图 9 似然函数可视化

在该图上，我们观察到，似然函数取极大值时，$1/\lambda$大概达到 5.30。可以看到结果与生成数据几乎一致，也与公式法的结果近似。
上述似然函数未经过对数化，所以我们发现其似然函数值过于小，达到了10^{-233}的量级，这样的数值不适合于数值计算，故必须进行对数化。在进行对数化后，其似然函数的值正常了许多。

以下对对数似然函数进行可视化。可以看到其最大值与图 9 类似。

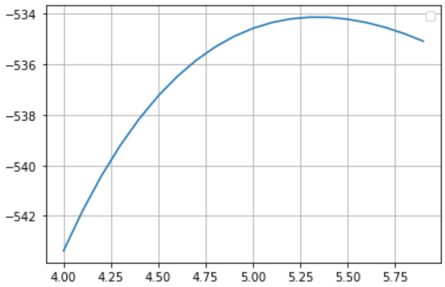

图 10 对数似然函数可视化

4 以伽马分布为例

4.1 似然函数

接下来，我们考虑更加一般化的情形，以伽马分布为例。其密度函数为：

$$f(x;\alpha,\beta)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}e^{-x/\beta}x^{\alpha-1}$$

$$\Theta=\{(\alpha,\beta):\alpha>0,\beta>0\}$$

若有 n 个独立同分布的样本，其对数似然函数为：

$$\begin{aligned}\ell(\alpha,\beta) &= \ln\prod_{i=1}^{n}\frac{1}{\beta^{\alpha}\Gamma(\alpha)}e^{-x_i/\beta}x_i^{\alpha-1}\\ &= \ln\left[\beta^{-n\alpha}\,\Gamma(\alpha)^{-n}\exp(-\frac{1}{\beta}\sum_{i=1}^{n}x_i)\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1}\right]\\ &= -n\alpha\ln\beta-n\ln\Gamma(\alpha)-\frac{1}{\beta}\sum_{i=1}^{n}x_i+(\alpha-1)\sum_{i=1}^{n}\ln x_i\end{aligned}$$

4.2 理论结果

该似然函数含有两个待估参数，我们的目的是求得似然函数的极大值。然而通过求导法，不易求得该最大值（见讲义 55 页），所以这里我们采用优化方法求得两个参数的估计值。

在此我们采用 scipy 库里的最小化算法逐步优化参数 α 和 β，使得负的对数似然函数值小。

$$\ell(\alpha,\beta)=-n\alpha\ln\beta-n\ln\Gamma(\alpha)-\frac{1}{\beta}\sum_{i=1}^{n}x_i+(\alpha-1)\sum_{i=1}^{n}\ln x_i$$

接着，我们将通过一个例子来展示该求极大似然估计的过程。
我们从$\alpha = 0.2, \beta = 2$的指数分布中随机生成 200 个数据点作为我们的观测数据；

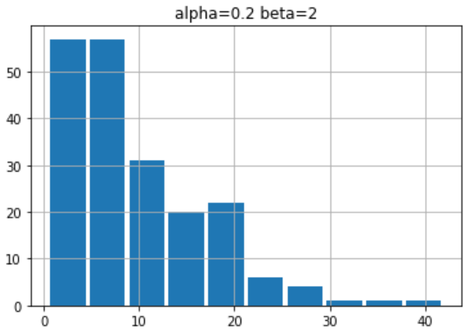

图 11 随机生成的伽马分布数据

在得到 200 个观测值后，我们依据该观测值计算对数似然函数。对其取负号后，我们利用 Nelder-Mead 算法对该二元非线性函数进行优化。
根据算法结果，在$1/\alpha = 5.34, \beta = 1.8$处停止优化，得到局部最优解。

4.2 可视化结果

接下来，我们通过可视化来检查我们的极大似然估计结果。我们定义待估参数的取值范围，$1/\alpha \in (4,6), \beta \in (1,3)$。我们将该参数组成的网格点一一计算其似然函数值，并 3D 可视化。

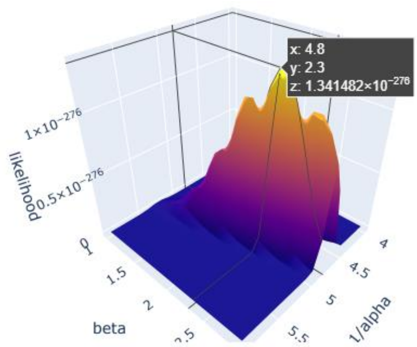

图 12 似然函数可视化

在该图上，我们观察到，似然函数取极大值时，$\frac{1}{\alpha} = 4.8, \beta = 2.3$。可以看到结果与生成数据几乎一致，也与优化求解的结果近似。
上述似然函数未经过对数化，所以我们发现其似然函数值过于小，达到了10^{-276}的量级，这样的数值不适合于数值计算，故必须进行对数化。在进行对数化后，其似然函数的值正常了许多。
以下对对数似然函数进行 3D 可视化。可以看到其最大值与图 12 类似。

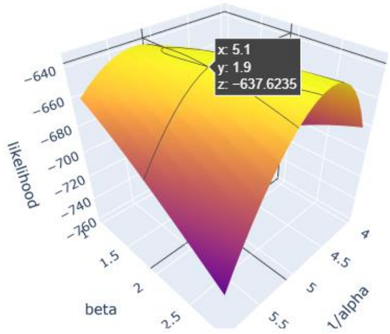

图 13 对数似然函数可视化

5 以泊松分布为例

5.1 似然函数

以上讨论的都是连续型分布，在本节，我们讨论离散型分布——泊松分布。假设有 N 个样本观测值 $(x_1, x_2, \cdots, x_N)$。

每个随机变量的概率密度函数，即似然函数为：

$$L_i = f(x_i \mid \lambda) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

其对数似然函数为

$$LnL_i = Log[f(x_i \mid \lambda)] = x_i \ln(\lambda) - \lambda - \ln(x_i!)$$

由于每个观测值都是独立的，因此这 N 个观测值的对数似然函数为

$$\begin{aligned} LnL &= \sum_{i=1}^{N} LnL_i = \sum_{i=1}^{N} [x_i \ln(\lambda) - \lambda - \ln(x_i!)] \\ &= \ln(\lambda) \sum_{i=1}^{N} x_i - N\lambda - \sum_{i=1}^{N} \ln(x_i!) \end{aligned}$$

5.2 理论结果

对以上似然函数求导，能够得到极大似然估计值。

$$\frac{\partial LnL}{\partial \lambda} = -N + \sum_{i=1}^{N} x_i / \lambda = 0$$

$$\hat{\lambda} = \sum_{i=1}^{N} x_i / N$$

接着，我们将通过一个例子来展示该求极大似然估计的过程。

我们从 $\lambda = 5$ 的泊松分布中随机生成 200 个数据点作为我们的观测数据；

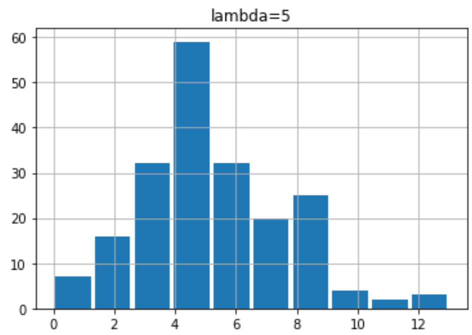

图 14 随机生成的指数分布数据

我们首先利用上述公式得到准确的极大似然估计值：

$$\hat{\lambda}_n = \frac{\sum_{j=1}^{n} x_j}{n} = 5.315$$

可以看出与真实值 5 相差不大。

5.3 可视化结果

接下来，我们通过可视化来检查我们的极大似然估计结果。我们定义待估参数的取值范围，$\lambda \in (4,6)$。我们每隔 0.1 的步长取一个点，并且一一计算其中各点的似然函数值，之后进行可视化。

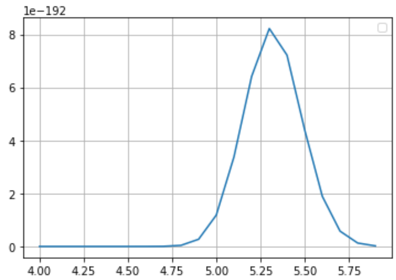

图 15 似然函数可视化

在该图上，我们观察到，似然函数取极大值时，λ大概达到 5.30。可以看到结果与生成数据几乎一致，也与公式法的结果近似。

上述似然函数未经过对数化，所以我们发现其似然函数值过于小，达到了10^{-192}的量级，这样的数值不适合于数值计算，故必须进行对数化。在进行对数化后，其似然函数的值正常了许多。

以下对对数似然函数进行可视化。可以看到其最大值与图 15 类似。

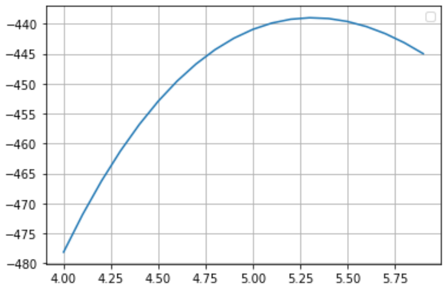

图 16 对数似然函数可视化

源码

本文的代码均可以在我的 Github 上获取：
https://github.com/stxupengyu/maximum-likelihood-estimation/blob/master/MLE.ipynb

参考文献

[1]指数分布极大似然估计,
https://www.statlect.com/fundamentals-of-statistics/exponential-distribution-maximum-likelihood
[2]伽马分布的产生,
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gamma.html
[3]概率论中常见分布总结以及 python 的 scipy 库使用,
https://www.cnblogs.com/pinking/p/7898313.html
[4]Python 产生 gamma 分布,
https://www.geeksforgeeks.org/scipy-stats-gamma-python/
[5]Estimating a Gamma distribution,
https://tminka.github.io/papers/minka-gamma.pdf
[6]Maximum Likelihood,
http://www.utstat.toronto.edu/~brunner/oldclass/appliedf12/lectures/2101f12Likelihood1.pdf
[7]图解极大似然估计,
https://blog.csdn.net/zenglaoshi/article/details/103285033?utm_medium=distribute.pc_relevant.none-task-blog-BlogCommendFromMachineLearnPai2-1.nonecase&depth_1-utm_source=distribute.pc_relevant.none-task-blog-BlogCommendFromMachineLearnPai2-1.nonecase
[8]numpy.random.gamma,
https://numpy.org/doc/stable/reference/random/generated/numpy.random.gamma.html
[9]scipy.optimize.minimize,
https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html#scipy.optimize.minimize
[10]Optimization (scipy.optimize),
https://docs.scipy.org/doc/scipy/reference/tutorial/optimize.html