最大似然估计法

最大似然估计就是利用已知的样本结果信息，反推最具有可能（最大概率）导致这些样本结果出现的模型参数值。

举例：给定一组数据$x_1, x_2 ... x_n$，已知他们都服从于同一个高斯分布$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$，但不知道这个高斯分布的两个参数$\sigma,\mu$分别是多少，这时候就需要通过最大似然估计法求出各个参数。

步骤

写出最大似然函数，以上面的例子为例，此时的最大似然函数为

$$L(x;\mu,\sigma)=f(\mu,\sigma;x_1)\cdot f(\mu,\sigma;x_2)\cdots f(\mu,\sigma;x_n)=\prod f(x_i;\ \mu,\sigma)$$

取对数，得到$\ln L(x;\mu,\sigma)$

分别对参数求偏导（$\frac{d\ln L(x;\mu,\sigma)}{d\mu}$和$\frac{d\ln L(x;\mu,\sigma)}{d\sigma}$）并令其为零，求出当似然函数取极大值时参数的值。

对正态分布的假设检验

1. 可以把残差画成直方图，并与正态分布的密度函数比较。但直方图是不连续的，为了得到对密度函数的光滑估计，可以使用“核密度估计法”（kernel density estimation）并与正态密度相比较。
2. 将正态分布的分位数与残差的分位数画成散点图。如果残差来自正态分布，则该图上的散点应该集中在45°线附近。称这种图为“分位数-分位数图”(Quantile-Quantile plot，简记QQ plot)。
3. JB检验。JB检验是对数据服从正态分布进行的检验，检验基于数据的样本偏度和峰度构造的统计量。在正态分布的假设下，JB统计量渐进地服从自由度为2的卡方分布，

   $$\mathrm{JB}\rightarrow\chi^2(2)$$

   。如果JB统计量值较大，比如为11，则可以计算出卡方值大于11的概率为0.004，这个概率过小，因此不能认为样本来自正态分布。反之，成立。

Stata 代码

```
hist mpg, normal
```

hist 绘制直方图，变量为 mpg，并使用 normal 叠加一个正态密度曲线进行比较

```
kdensity mpg, normal lpattern("-")
kdensity mpg, lpattern("-") normal
```

kdensity 绘制核密度图，并使用 normal 叠加一个正态密度曲线，`lpattern("-")` 跟在 normal 后面表示 normal 曲线用虚线表示，在 normal 前面则是 mpg 用虚线表示。

```
qnorm mpg
```

绘制 QQ 图。

```
su mpg, detail
```

显示 mpg 变量的偏度和峰度，detail 是必须的，因为 su（summarize）只能显示描述性统计内容，detail 才能显示到偏度和峰度。

手搓 JB 统计量

```
di (r(N)/6)*((r(skewness)^2)+[(1/4)*(r(kurtosis)-3)^2])
```

这条命令依赖于前面 `su mpg, detail` 生成的偏度和峰度，因此不能单独运行。

1. `di` 是 Stata 中 "display" 的缩写，用于显示结果。
2. `r()` 是 Stata 中用于存储和访问最近执行的命令结果的函数。这些结果被存储在 Stata 的内存中，可以在后续的操作中被调用。

```
di chi2tail(2, 14.031924)
```

计算自由度为2的卡方分布观察到大于等于 14.03（上一步算出来的值）的概率。

调包

```
ssc install jb6
```

`ssc` 是 "Statistical Software Components" 的缩写，它是 Stata 用户贡献的命令和程序的在线存储库，类似于 Python 的 pip。

```
jb6 mpg
```

直接对 mpg 进行 JB 检验，输出 JB 统计量和检验结果。

```
sktest mpg
```

其他检验

```
sktest mpg // D'Agostino 检验
swilk mpg // 非参数 Shapiro-Wilk 检验
sfrancia mpg // 非参数 Shapiro-Francia 检验
```