SS3858B Tutorial

Jan 27, 2010

1 Uniform

Sometimes Method of Moments and MLE provide the same estimates, but not always. Let $X_1, \cdots, X_n$ be iid from uniform$[-\theta, \theta]$ with $\theta > 0$.

Method of Moments

We have to use the second moment since $\mathrm{E}(X) = 0$.

$$\mathrm{E}(X^2) = \int_{-\theta}^{\theta} x^2 \frac{1}{2\theta} dx = \frac{\theta^2}{3}.$$

Then

$$\tilde{\theta} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2}.$$

MLE

$$\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} L(\theta) \\
&= \arg\max_{\theta} \prod_{i=1}^{n} \frac{1}{2\theta} \mathbb{I}_{\{-\theta \le X_i \le \theta\}} \\
&= \arg\max_{\theta} \frac{1}{(2\theta)^n} \prod_{i=1}^{n} \mathbb{I}_{\{-\theta \le X_i \le \theta\}} \\
&= \max\{|X_{(1)}|, |X_{(n)}|\}.
\end{aligned}$$

Numeric Example in R

```
> r=runif(10000,-3,3) #unif[-3,3]
> sqrt(3/10000*sum(r^2)) #Method of Moments estimate
[1] 2.98994
> max(abs(min(r)),abs(max(r))) #MLE
[1] 2.999256
```

2 Uniform # 2

Suppose $X_1, \cdots, X_n$ are iid with uniform$[0, \theta]$.

$$\begin{aligned}\tilde{\theta} &= 2\bar{X}. \quad \text{(done in last week's tutorial)} \\ \hat{\theta} &= \arg\max_{\theta} \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{I}_{\{0 \le X_i \le \theta\}} = X_{(n)}.\end{aligned}$$

$$\begin{aligned}\mathrm{E}(\tilde{\theta}) &= 2\mathrm{E}(\bar{X}) = 2 \cdot \frac{\theta}{2} = \theta; \\ \mathrm{Var}(\tilde{\theta}) &= \frac{4}{n^2} n \mathrm{Var}(X) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}.\end{aligned}$$

Thus, $\tilde{\theta}$ is unbiased.

Since

$$\mathrm{P}(X_{(n)} \le x) = \mathrm{P}(X \le x)^n = \left(\frac{x}{\theta}\right)^n,$$

the density of $X_{(n)}$ is

$$f_{X_{(n)}}(x) = \frac{n x^{n-1}}{\theta^n}.$$

Then

$$\begin{aligned}\mathrm{E}(\hat{\theta}) &= \int_0^{\theta} \frac{n x^n}{\theta^n} = \frac{n}{n+1}\theta. \\ \mathrm{Var}(\hat{\theta}) &= \mathrm{E}(\hat{\theta}^2) - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+2)(n+1)^2}\theta^2 = \frac{1}{n^2 + 4n + 5 + \frac{2}{n}}\theta^2 < \mathrm{Var}(\tilde{\theta}).\end{aligned}$$

$\hat{\theta}$ is biased, but with smaller variance. If n is small, the bias is quite large. But we are more preferable to $\hat{\theta}$ when n is large.

3 Discrete Case

Sometimes the derivative of the likelihood (or log-likelihood) may not exist. The above is one example. Here is another one. Only one observation is taken on a discrete random variable X with the following pmf, where $\theta \in \{1, 2, 3\}$.

x	$f(x\|1)$	$f(x\|2)$	$f(x\|3)$
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

The MLE's are the followings given different values of the observation.

x	0	1	2	3	4
$\hat{\theta}$	1	1	2 or 3	3	3

This example also shows that the MLE may not be unique and sometimes the sample size may be extremely small.

4 Regression

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where $\epsilon_1, \cdots, \epsilon_n$ are iid $N(0, \sigma^2)$ and x_i's are fixed. Find the MLE of β and σ^2.

We know that $Y_i \sim N(\beta x_i, \sigma^2)$. The likelihood function (for the y_i's)

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \beta x_i)^2}{2\sigma^2}\right\}.$$

The negative log-likelihood function

$$l(\beta, \sigma^2) = -\log L(\beta, \sigma^2) = \frac{n}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{2\sigma^2},$$

and

$$(\hat{\beta}, \hat{\sigma}^2) = \arg\min_{(\beta, \sigma^2)} l(\beta, \sigma^2).$$

Consider the two partial derivatives of l,

$$\begin{aligned}
\frac{\partial l(\beta, \sigma^2)}{\partial \beta} &= \sum_{i=1}^{n} \frac{y_i - \beta x_i}{2\sigma^2}(-x_i) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i y_i - \beta x_i^2) \overset{\text{set}}{=} 0 \\
&\Rightarrow \quad \hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}; \\
\frac{\partial l(\hat{\beta}, \sigma^2)}{\partial \sigma^2} &= \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \hat{\beta} x_i)^2 \overset{\text{set}}{=} 0 \\
&\Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta} x_i)^2.
\end{aligned}$$

5 Time Series

Likelihood function is not always the product of individual densities. This is true when the observations are independent. Consider the simple AR(1) model

$$x_t = \alpha x_{t-1} + \epsilon_t,$$

where ϵ_t is iid white noise with mean 0 and variance σ^2. Notice that ϵ does not have to be normally distributed. In order to find MLE for α and σ^2, quasi-MLE is usually used in this case, by pretending the disturbance term follows a normal distribution. Suppose the observed data is $x_1, \cdots, x_n$. Then the likelihood function (quasi-likelihood) is given by

$$\begin{aligned}
L(\alpha, \sigma^2) &= f_{1:n}(x_1, \cdots, x_n; \alpha, \sigma^2) \\
&= f_{1:(n-1)}(x_1, \cdots, x_{n-1}; \alpha, \sigma^2) f_n(x_n | x_1, \cdots, x_{n-1}; \alpha, \sigma^2) \\
&= \cdots \\
&= f_1(x_1; \alpha, \sigma^2) f_2(x_2 | x_1; \alpha, \sigma^2) \cdots f_n(x_n | x_1, \cdots, x_{n-1}; \alpha, \sigma^2)
\end{aligned}$$

where f_i denotes the conditional density of x_i given the past and $f_{i:j}, i < j$ denotes the joint density of $x_i, \cdots, x_j$.

All the conditional distributions are normal. We could think that the initial value x_1 is drawn from some distribution with density f_1. But in this case, there is no closed form solution and some numeric methods such as Newton-Raphson have to be used to maximize the likelihood. To simplify our problem, we treat the first observation as a deterministic number. Then

$$\begin{aligned}
L(\alpha, \sigma^2) &= \prod_{i=2}^{n} f_i(x_i | x_{i-1}; \alpha, \sigma^2) \\
&= \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_i - \alpha x_{i-1})^2}{2\sigma^2} \right\}.
\end{aligned}$$

This is similar with result we attained in the regression problem. Therefore,

$$\begin{aligned}
\hat{\alpha} &= \frac{\sum_{i=2}^{n} x_{i-1} x_i}{\sum_{i=2}^{n} x_{i-1}^2}; \\
\hat{\sigma}^2 &= \frac{1}{n-1} \sum_{i=2}^{n} (x_i - \hat{\alpha} x_{i-1})^2.
\end{aligned}$$