Stat 4620: Day 15

(3/19)

1 Chapter 6: Maximum Likelihood Estimation

Recall: The likelihood function is

$$L(\theta) = L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$$

The *log-likelihood* function is

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i; \theta)$$

Definition 6.1.1
Given a random sample $\mathbf{x} = (x_1, \ldots, x_n)$ from a pdf $f(x; \theta)$ where $\theta \in \Omega$. The *maximum likelihood estimator* (mle) of θ is $\hat{\theta}$ such that

$$L(\hat{\theta}; \mathbf{x}) \geq L(\theta; \mathbf{x}) \text{ for all } \theta \in \Omega$$

In practice, we often, but not always, maximize $l(\theta) = \ln L(\theta)$.

1.1 Properties of the mle

Theorem 1 (Chebyshev's inequality, Sec. 1.10)**.** *Let the random variable X have density $f(\cdot)$ with mean μ and variance σ^2. Then for every $k > 0$,*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

or equivalently

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Proof.

$$\begin{aligned}
\sigma^2 = E(X-\mu)^2 &= \int (x-\mu)^2 f(x)dx \\
&= \int_{|x-\mu|\geq k\sigma} (x-\mu)^2 f(x)dx + \int_{|x-\mu|<k\sigma} (x-\mu)^2 f(x)dx \\
&\geq \int_{|x-\mu|\geq k\sigma} (x-\mu)^2 f(x)dx \qquad \text{(since integral is nonnegative)} \\
&\geq (k\sigma)^2 \int_{|x-\mu|\geq k\sigma} f(x)dx \\
&= k^2\sigma^2 P(|x-\mu| \geq k\sigma)
\end{aligned}$$

□

Definition (Convergence in probability). A sequence of random variables $\{X_n\}$ *converges in probability* to a constant a if for all $\epsilon > 0$,

$$\lim_{n\to\infty} P\left[|X_n - a| < \epsilon\right] = 1$$

Theorem 2 (Weak law of Large Numbers, Sec. 5.1)**.** *Let $X_1, \ldots, X_n$ be a random sample from a density $f(\cdot)$ with mean μ and variance σ^2. Let $\overline{X}_n = (1/n)\sum_{i=1}^n$. Then*

$$\overline{X}_n \xrightarrow{P} \mu$$

Proof. For every $\epsilon > 0$,

$$P\left[|\overline{X}_n - \mu| < \epsilon\right] = P\left[|\overline{X}_n - \mu| < \frac{\epsilon\sqrt{n}}{\sigma}\frac{\sigma}{\sqrt{n}}\right] \le 1 - \frac{\sigma^2}{\epsilon^2 n} \to 1$$

using Chebyshev's inequality with $k = \frac{\epsilon\sqrt{n}}{\sigma}$. □

Theorem 3 (Jensen's Inequality, Sec. 1.10)**.** *Let ϕ be a convex function on an open interval I and let X be a random variable whose support is contained in I and has finite mean μ. Then*

$$\phi\left[E(X)\right] \le E\left[\phi(X)\right]$$

Proof. For this proof, assume that ϕ has a second derivative, but in general only convexity is required. Expand $\phi(x)$ into a Taylor series about μ

$$\phi(x) = \phi(\mu) + \phi'(\mu)(x-\mu) + \phi''(\xi)(x-\mu)^2$$

where ξ is between x and μ. By convexity, the last term is nonnegative, so

$$\phi(x) \ge \phi(\mu) + \phi'(\mu)(x-\mu)$$

Taking expectations of both sides leads to the result. □

Theorem 4. *Let θ_0 denote the true parameter value. Under the assumptions*

(R0) $f(x_i;\theta) \ne f(x_i;\theta')$ if $\theta \ne \theta'$

(R1) The pdf have common support for all θ

Then for all $\theta \ne \theta_0$,

$$\lim_{n\to\infty} P_{\theta_0}\left[L(\theta_0;\mathbf{x}) > L(\theta;\mathbf{x})\right] = 1$$

Proof.

$$L(\theta_0;\mathbf{x}) > L(\theta;\mathbf{x}) \Leftrightarrow \frac{L(\theta;\mathbf{x})}{L(\theta_0;\mathbf{x})} < 1 \Leftrightarrow \frac{\prod f(x_i;\theta)}{\prod f(x_i;\theta_0)} < 1$$
$$\Leftrightarrow \prod \frac{f(x_i;\theta)}{f(x_i;\theta_0)} < 1 \Leftrightarrow \frac{1}{n}\sum \ln\frac{f(x_i;\theta)}{f(x_i;\theta_0)} < 0$$

Using the law of large numbers and Jensen's inequality,

$$\frac{1}{n}\sum \ln \frac{f(x_i;\theta)}{f(x_i;\theta_0)} \xrightarrow{P} E_{\theta_0}\left[\ln \frac{f(X;\theta)}{f(X;\theta_0)}\right] < \ln E_{\theta_0}\left[\frac{f(X;\theta)}{f(X;\theta_0)}\right] = \ln(1) = 0$$

since

$$E_{\theta_0}\left[\frac{f(X;\theta)}{f(X;\theta_0)}\right] = \int \left[\frac{f(X;\theta)}{f(X;\theta_0)}\right] f(X;\theta_0)\, dx = \int f(X;\theta)\, dx = 1$$

□

Comment:

- The theorem says: the likelihood function is maximized at the true value θ_0, with probability approaching 1 as sample size increases.

1.2 Examples

1.2.1 Normal distribution

Let $x_1, \dots, x_n$ be a random sample from a distribution with continuous pdf

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2},\ -\infty < x < \infty$$

We will find the mle of θ. The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} = \left(\frac{1}{2\pi}\right)^{n/2} e^{-\frac{1}{2}\sum(x_i-\theta)^2}$$

Taking ln of the likelihood function and calculating its derivative

$$l(\theta) = \ln L(\theta) = -(n/2)\ln(2\pi) - \frac{1}{2}\sum(x_i-\theta)^2$$

$$l'(\theta) = \frac{d}{d\theta} l(\theta) = \sum(x_i - \theta)$$

Equating to 0,

$$0 = l'(\hat{\theta}) = \sum(x_i - \hat{\theta})$$

implies $\hat{\theta} = \frac{1}{n}\sum x_i = \overline{x}$.

1.2.2 Example 6.1: Laplace or double exponential

Let $x_1, \dots, x_n$ be a random sample from a distribution with continuous pdf

$$f(x;\theta) = \frac{1}{2} e^{-|x-\theta|},\ -\infty < x < \infty$$

The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{2} e^{-|x_i-\theta|} = \left(\frac{1}{2}\right)^{n} e^{-\sum|x_i-\theta|}$$

Taking ln of the likelihood function and calculating its derivative

$$l(\theta) = -n\ln(2) - \sum |x_i - \theta|$$

$$l'(\theta) = \frac{d}{d\theta} l(\theta) = -\sum \frac{d}{d\theta} |x_i - \theta|$$

But

$$|x_i - \theta| = \begin{cases} (x-\theta), & \text{if } x > \theta \\ 0, & \text{if } x = \theta \\ (\theta - x), & \text{if } x < \theta \end{cases}$$

$$\begin{aligned} \frac{d}{d\theta}|x_i - \theta| &= \begin{cases} -1, & \text{if } x > \theta \\ 0, & \text{if } x = \theta \\ 1, & \text{if } x < \theta \end{cases} \\ &= -\text{sgn}(x-\theta) \end{aligned}$$

so that $0 = l'(\hat{\theta}) = \sum \text{sgn}(x_i - \hat{\theta})$ implies that $\hat{\theta} = \text{med}\{x_i\}$.

1.2.3 Example 6.1.2: Logistic (no closed form mle)

Let $x_1, \ldots, x_n$ be a random sample from a distribution with continuous pdf

$$f(x;\theta) = \frac{e^{-(x-\theta)}}{\left[1 + e^{-(x-\theta)}\right]^2}, \ -\infty < x < \infty$$

Then

$$L(\theta) = \prod_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{\left[1 + e^{-(x-\theta)}\right]^2}$$

$$\begin{aligned} l(\theta) = \sum \ln \frac{e^{-(x_i-\theta)}}{\left[1 + e^{-(x-\theta)}\right]^2} &= \sum_{i=1}^{n} \left[-(x_i - \theta) - 2\ln\left(1 + e^{-(x-\theta)}\right)\right] \\ &= -n\overline{x} + n\theta - 2\sum \ln\left(1 + e^{-(x-\theta)}\right) \end{aligned}$$

$$l'(\theta) = n - 2\sum \frac{1}{1 + e^{-(x-\theta)}} e^{-(x-\theta)}$$

Equating to zero, the mle is $\hat{\theta}$ that solves the *estimating equation*

$$\sum_{i=1}^{n} \frac{e^{-(x-\hat{\theta})}}{1 + e^{-(x-\hat{\theta})}} = \frac{n}{2} \tag{1}$$

1.3 Maximizing likelihood calculations in R using optimize()

:

```
> xvec<-c(1,4,5,7,8,21,24)                                    # Data
> mean(xvec)
[1] 10
> median(xvec)
[1] 7
> fn1<-function(theta){ prod(exp(-.5*(xvec-theta)^2))}       # Normal likelihood
> optimize(fn1, c(0,15), maximum=T)                           # Max using optimize
$maximum
[1] 9.999982

$objective
[1] 3.20998e-103

> fn2<-function(theta){ prod(.5*exp(-1*abs(xvec-theta)))}    # Laplace likelihood
> optimize(fn2, c(0,15), maximum=T)
$maximum
[1] 6.999993

$objective
[1] 1.652435e-21

> fn3<-function(theta){ prod((exp(-1*(xvec-theta)))/( 1+exp(-1*(xvec-theta)) )^2 )} # Lo
> optimize(fn3, c(0,15), maximum=T)
$maximum
[1] 6.834033

$objective
[1] 2.012726e-20

> sto1<-rep(0,length(thetavec1))
> thetavec1<-seq(6,8,by=.10)
> for(i in 1:length(thetavec1)){sto1[i]<-fn3(thetavec[i])}    # Confirm mle for logistic
> cbind(thetavec1, sto1)
      thetavec1         sto1
 [1,]       6.0 1.319842e-20
 [2,]       6.1 1.451727e-20
 [3,]       6.2 1.577451e-20
 [4,]       6.3 1.693319e-20
 [5,]       6.4 1.795716e-20
 [6,]       6.5 1.881289e-20
 [7,]       6.6 1.947142e-20
 [8,]       6.7 1.990992e-20
 [9,]       6.8 2.011319e-20
```

```
[10,]       6.9 2.007451e-20
[11,]       7.0 1.979611e-20
[12,]       7.1 1.928902e-20
[13,]       7.2 1.857242e-20
[14,]       7.3 1.767239e-20
[15,]       7.4 1.662036e-20
[16,]       7.5 1.545127e-20
[17,]       7.6 1.420160e-20
[18,]       7.7 1.290749e-20
[19,]       7.8 1.160300e-20
[20,]       7.9 1.031874e-20
[21,]       8.0 9.080843e-21

> thetavec2<-seq(6.7, 6.9, by=.01)
> for(i in 1:length(thetavec2)){sto2[i]<-fn3(thetavec[i])}
> cbind(thetavec2, sto2)
      thetavec2         sto2
 [1,]      6.70 1.990992e-20
 [2,]      6.71 1.994102e-20
 [3,]      6.72 1.996974e-20
 [4,]      6.73 1.999608e-20
 [5,]      6.74 2.002003e-20
 [6,]      6.75 2.004159e-20
 [7,]      6.76 2.006074e-20
 [8,]      6.77 2.007748e-20
 [9,]      6.78 2.009181e-20
[10,]      6.79 2.010371e-20
[11,]      6.80 2.011319e-20
[12,]      6.81 2.012025e-20
[13,]      6.82 2.012487e-20
[14,]      6.83 2.012707e-20
[15,]      6.84 2.012683e-20
[16,]      6.85 2.012417e-20
[17,]      6.86 2.011908e-20
[18,]      6.87 2.011156e-20
[19,]      6.88 2.010163e-20
[20,]      6.89 2.008927e-20
[21,]      6.90 2.007451e-20
```