Lecture 09 Large Sample Theory of MLE

Dr. Shafayat Abrar

The Method of Maximum Likelihood for Point Estimation

The most common method for estimating parameters in a parametric model is the maximum likelihood method. Let X_1, \ldots, X_n be IID with PDF $f(x; \theta)$.

Definition: The likelihood function is defined by

$$ext{lik}(heta) = \mathcal{L}_n(heta) = \prod_{i=1}^n f(X_i; heta)$$

The log-likelihood function is defined by

$$\ell(heta) = \log ext{lik}(heta) = \sum_{i=1}^n \log [f(X_i| heta)]$$

Find heta by solving

$$rac{\partial}{\partial heta}\ell(heta)=\ell'(heta)=0$$

Large Sample Theory

Let $\hat{\theta}_n$ be an estimate of a parameter θ based on a sample of size n. Then $\hat{\theta}_n$ is said to be consistent in probability if $\hat{\theta}_n$ converges in probability to θ as n approaches infinity; that is, for any $\epsilon > 0$

$$P\Big(\left| \hat{ heta}_n - heta
ight| > \epsilon \Big) o 0 \quad ext{ as } n o \infty$$

An estimator is asymptotically **normal** if

$$rac{\hat{ heta}_n - heta}{ ext{se}} \leadsto N(0,1) \ ext{ as } n o \infty \qquad \qquad \hat{ heta}_n \leadsto N(heta, \sigma^2_{\hat{ heta}}) \ ext{ as } n o \infty$$

$$\hat{ heta}_n \leadsto N(heta, \sigma^2_{\hat{ heta}}) \;\; ext{ as } \; n o \infty$$

where se $=\sigma_{\hat{\theta}_n}$.

```
clear
n=1000;
mu true = 10;
sigmga square true = 4;
for jj=1:50000
  X=randn(1,n)*sqrt(sigmga_square_true)+mu_true;
  mu_ML(jj) = mean(X);
                                                      Histogram of the estimated mean
  var_ML(jj)= mean((X-mu_ML(jj)).^2);
                                                  8000
end
                                                  7000
subplot 121
                                                                                       6000
hist(mu ML,25)
                                                  6000
title('Histogram of the estimated mean')
                                                                                       5000
                                                  5000
subplot 122
                                                                                       4000
hist(var_ML,25)
                                                  4000
title('Histogram of the estimated variance')
                                                                                       3000
                                                  3000
                                                                                       2000
                                                  2000
                                                                                       1000
                                                  1000
```

Large Sample Theory for Method of Maximum Likelihood

Theorem A (Rice, 2007)

Under appropriate smoothness conditions on f, the mle from an i.i.d. sample is consistent.

This theorem simply means that the derivative of log likelihood is really zero when the optimizing parameter θ is close to its true but unknown value θ_0 . Or in other words, optimization of likelihood does yield a consistent estimator.

$$rac{\partial}{\partial heta} rac{1}{n} \sum_{i=1}^n \log [f(x_i \mid heta)]
ightarrow rac{\partial}{\partial heta} \mathbb{E} \log f(x \mid heta) = 0, \ ext{ when } \hat{ heta}_n
ightarrow heta$$

Proof of Theorem A

Consider maximizing

$$rac{1}{n}l(heta) = rac{1}{n}\sum_{i=1}^n \log[f(X_i\mid heta)]$$

As n tends to infinity, the law of large numbers implies that

$$rac{1}{n}l(heta)
ightarrow E \log f(X \mid heta) = \int \log f(x \mid heta) f(x \mid heta_0) dx$$

To maximize $E \log f(X \mid \theta)$, we consider its derivative:

$$rac{\partial}{\partial heta} \int \log f(x \mid heta) f(x \mid heta_0) dx = \int rac{rac{\partial}{\partial heta} f(x \mid heta)}{f(x \mid heta)} f(x \mid heta_0) dx$$

If $\theta = \theta_0$, this equation becomes

$$\int rac{\partial}{\partial heta} f(x \mid heta_0) dx = rac{\partial}{\partial heta} \int f(x \mid heta_0) dx = rac{\partial}{\partial heta} (1) = 0$$

which shows that θ_0 is a stationary point and hopefully a maximum.

Large Sample Theory for Method of Maximum Likelihood

Lemma A (Rice, 2007)

Define $I(\theta)$ by

$$I(heta) = \mathbb{E}igg[rac{\partial}{\partial heta} \! \log f(X \mid heta)igg]^2$$

Under appropriate smoothness conditions on $f, I(\theta)$ may also be expressed as

$$I(heta) = -\mathbb{E}iggl[rac{\partial^2}{\partial heta^2} \!\log f(X \mid heta)iggr]$$

Proof of Lemma A

First, we observe that since $\int f(x \mid \theta) dx = 1$,

$$\frac{\partial}{\partial heta} \int f(x \mid heta) dx = 0$$

Combining this with the identity

$$\frac{\partial}{\partial \theta} f(x \mid \theta) = \left[\frac{\partial}{\partial \theta} \log f(x \mid \theta) \right] f(x \mid \theta)$$

$$0 = rac{\partial}{\partial heta} \int f(x \mid heta) dx = \int \left[rac{\partial}{\partial heta} \log f(x \mid heta)
ight] f(x \mid heta) dx$$

Taking second derivative both sides, we obtain

$$0 = \int \left[\frac{\partial^2}{\partial \theta^2} \log f(x \mid \theta) \right] f(x \mid \theta) dx + \int \left[\frac{\partial}{\partial \theta} \log f(x \mid \theta) \right]^2 f(x \mid \theta) dx$$

$$Eiggl[rac{\partial^2}{\partial heta^2} {
m log}\, f(x\mid heta)iggr] + Eiggl[rac{\partial}{\partial heta} {
m log}\, f(x\mid heta)iggr]^2 = 0$$

Large Sample Theory for Method of Maximum Likelihood Theorem B: (Rice, 2007)

Under smoothness conditions on f, the probability distribution of $\sqrt{nI(\theta_0)} (\hat{\theta} - \theta_0)$ tends to a standard normal distribution. i.e.,

$$\sqrt{nI(heta_0)} \Big(\hat{ heta} - heta_0 \Big) \overset{n o \infty}{\leadsto} N(0,1)$$

or in other words.

 $ilde{ heta} = \hat{ heta} - heta_0 \overset{n o \infty}{\leadsto} Nigg(0, rac{1}{nI(heta_0)}igg) \ \hat{ heta} \overset{n o \infty}{\leadsto} Nigg(heta_0, rac{1}{nI(heta_0)}igg)$

To prove this theorem, we need to evaluate the variance of estimation error $\tilde{\theta}$.

Proof of Theorem B

From a Taylor series expansion, we may expand $\ell'(\hat{\theta}) = 0$ around the true value θ_0 as follows:

$$egin{aligned} 0 &= \ell'(\hat{ heta}) pprox \ell'(heta_0) + \left(\hat{ heta} - heta_0
ight) \ell''(heta_0) \ \left(\hat{ heta} - heta_0
ight) pprox - rac{\ell'(heta_0)}{\ell''(heta_0)} \ n^{1/2} \Big(\hat{ heta} - heta_0\Big) pprox - rac{n^{1/2}\ell'(heta_0)}{\ell''(heta_0)} = rac{-n^{-1/2}\ell'(heta_0)}{n^{-1}\ell''(heta_0)} \end{aligned}$$

 $ext{var}\Big(n^{1/2}\Big(\hat{ heta}- heta_0\Big)\Big)pprox ext{var}igg(rac{-n^{-1/2}\ell'(heta_0)}{n^{-1}\ell''(heta_0)}igg)$

First, we consider the denominator:

$$rac{1}{n}l''(heta_0) = rac{1}{n}\sum_{i=1}^nrac{\partial^2}{\partial heta_0^2}{
m log}\,f(x_i\mid heta_0)$$

By the law of large numbers, the latter expression converges to

$$\mathbb{E} \left[rac{\partial^2}{\partial heta_0^2} {
m log} \, f(X \mid heta_0)
ight] =: -I(heta_0)$$

Proof of Theorem B

$$egin{aligned} ext{var}\Big(n^{1/2}\Big(\hat{ heta}- heta_0\Big)\Big) &pprox ext{var}igg(rac{-n^{-1/2}\ell'(heta_0)}{-I(heta_0)}igg) = rac{1}{I(heta_0)^2} ext{var}\Big(-n^{-1/2}\ell'(heta_0)\Big) \ & ext{var}\Big[n^{-1/2}l'(heta_0)\Big] = rac{1}{n}\sum_{i=1}^n \mathbb{E}igg[rac{\partial}{\partial heta_0}\log f(X_i\mid heta_0)\Big]^2 - \Big(rac{1}{\sqrt{n}}\mathbb{E}igg[rac{\partial}{\partial heta_0}\log f(X_i\mid heta_0)\Big]\Big)^2 \ &= I(heta_0) - 0 \ &= I(heta_0) \ & ext{var}\Big(n^{1/2}\Big(\hat{ heta}- heta_0\Big)\Big) pprox rac{1}{I(heta_0)^2} ext{var}\Big(-n^{-1/2}\ell'(heta_0)\Big) = rac{1}{I(heta_0)^2}I(heta_0) = rac{1}{I(heta_0)} \ & ext{var}\Big(\hat{ heta}- heta_0\Big) pprox rac{1}{nI(heta_0)} \end{aligned}$$