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

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

Find θ by solving

$$\frac{\partial}{\partial\theta}\ell(\theta) = \ell'(\theta) = 0$$

The log-likelihood function is defined by

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

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 $\epsilon > 0$

$$P\left(\left|\hat{\theta}_n - \theta\right| > \epsilon\right) \to 0 \quad \text{as } n \to \infty$$

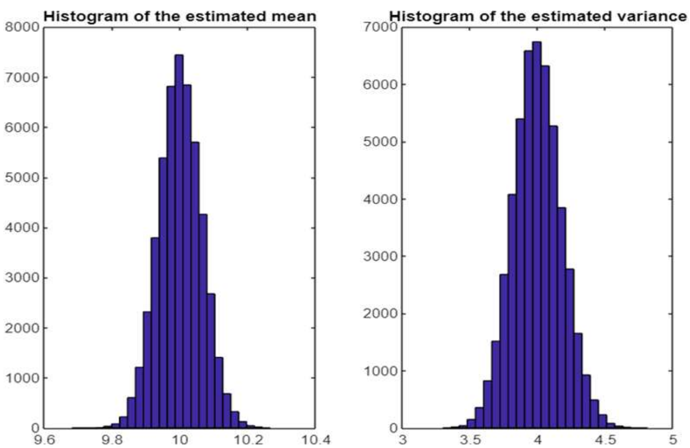
An estimator is asymptotically **normal** if

$$\frac{\hat{\theta}_n - \theta}{\text{se}} \rightsquigarrow N(0,1) \quad \text{as } n \to \infty \qquad \boxed{\hat{\theta}_n \rightsquigarrow N(\theta, \sigma^2_{\hat{\theta}}) \quad \text{as } n \to \infty}$$

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

```
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);
  var_ML(jj)= mean((X-mu_ML(jj)).^2);
end
subplot 121
hist(mu_ML,25)
title('Histogram of the estimated mean')

subplot 122
hist(var_ML,25)
title('Histogram of the estimated variance')
```



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.

$$\frac{\partial}{\partial \theta} \frac{1}{n} \sum_{i=1}^{n} \log[f(x_i \mid \theta)] \to \frac{\partial}{\partial \theta} \mathbb{E} \log f(x \mid \theta) = 0, \text{ when } \hat{\theta}_n \to \theta$$

Proof of Theorem A

Consider maximizing

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

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

$$\frac{1}{n} l(\theta) \to E \log f(X \mid \theta) = \int \log f(x \mid \theta) f(x \mid \theta_0) dx$$

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

$$\frac{\partial}{\partial \theta} \int \log f(x \mid \theta) f(x \mid \theta_0) dx = \int \frac{\frac{\partial}{\partial \theta} f(x \mid \theta)}{f(x \mid \theta)} f(x \mid \theta_0) dx$$

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

$$\int \frac{\partial}{\partial \theta} f(x \mid \theta_0) dx = \frac{\partial}{\partial \theta} \int f(x \mid \theta_0) dx = \frac{\partial}{\partial \theta}(1) = 0$$

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

$$I(\theta) = \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right]^2$$

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

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta)\right]$$

Proof of Lemma A

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

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

Combining this with the identity

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

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

Taking second derivative both sides, we obtain

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

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

Large Sample Theory for Method of Maximum Likelihood

Theorem B: (Rice, 2007)

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

$$\sqrt{nI(\theta_0)}\left(\hat{\theta}-\theta_0\right) \overset{n\to\infty}{\rightsquigarrow} N(0,1)$$

or in other words,

$$\tilde{\theta}=\hat{\theta}-\theta_0 \overset{n\to\infty}{\rightsquigarrow} N\left(0,\frac{1}{nI(\theta_0)}\right)$$

$$\hat{\theta} \overset{n\to\infty}{\rightsquigarrow} N\left(\theta_0,\frac{1}{nI(\theta_0)}\right)$$

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:

$$0=\ell'(\hat{\theta})\approx\ell'(\theta_0)+\left(\hat{\theta}-\theta_0\right)\ell''(\theta_0)$$

$$\left(\hat{\theta}-\theta_0\right)\approx-\frac{\ell'(\theta_0)}{\ell''(\theta_0)}$$

$$n^{1/2}\left(\hat{\theta}-\theta_0\right)\approx-\frac{n^{1/2}\ell'(\theta_0)}{\ell''(\theta_0)}=\frac{-n^{-1/2}\ell'(\theta_0)}{n^{-1}\ell''(\theta_0)}$$

$$\mathrm{var}\left(n^{1/2}\left(\hat{\theta}-\theta_0\right)\right)\approx\mathrm{var}\left(\frac{-n^{-1/2}\ell'(\theta_0)}{n^{-1}\ell''(\theta_0)}\right)$$

First, we consider the denominator:

$$\frac{1}{n}l''(\theta_0)=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta_0^2}\log f(x_i\mid\theta_0)$$

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

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta_0^2}\log f(X\mid\theta_0)\right]=:-I(\theta_0)$$

Proof of Theorem B

$$\mathrm{var}\left(n^{1/2}\left(\hat{\theta}-\theta_0\right)\right) \approx \mathrm{var}\left(\frac{-n^{-1/2}\ell'(\theta_0)}{-I(\theta_0)}\right) = \frac{1}{I(\theta_0)^2}\mathrm{var}\left(-n^{-1/2}\ell'(\theta_0)\right)$$

$$\begin{aligned}\mathrm{var}\left[n^{-1/2}l'(\theta_0)\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\partial}{\partial\theta_0}\log f(X_i \mid \theta_0)\right]^2 - \left(\frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{\partial}{\partial\theta_0}\log f(X_i \mid \theta_0)\right]\right)^2 \\ &= I(\theta_0) - 0 \\ &= I(\theta_0)\end{aligned}$$

$$\mathrm{var}\left(n^{1/2}\left(\hat{\theta}-\theta_0\right)\right) \approx \frac{1}{I(\theta_0)^2}\mathrm{var}\left(-n^{-1/2}\ell'(\theta_0)\right) = \frac{1}{I(\theta_0)^2}I(\theta_0) = \frac{1}{I(\theta_0)}$$

$$\mathrm{var}\left(\widetilde{\theta}\right) = \mathrm{var}\left(\hat{\theta}-\theta_0\right) \approx \frac{1}{nI(\theta_0)}$$