Lecture 8
Methods of Point Estimation

Chao Song

College of Ecology

Lanzhou University

October 29, 2024

Essential terminologies

In statistics, we often consider random variables for which the functional form of the PMF or PDF is known, but the distribution depends on unknown parameters.. We take a random sample $X_1, X_2, \ldots, X_n$ from the distribution to elicit some information about the unknown parameters, say θ.

- All potential values of the parameters is called the **parameter space**;
- The statistic $u(X_1, X_2, \ldots, X_n)$ used to estimate θ is called an **estimator**;
- The computed value of the estimator is called an **estimate**.
- Since we are estimating one single value for each parameter, this is also referred to as **point estimation**

While intuition can lead us to good estimators, we need a more methodical way of estimating parameters for complex problems we encounter in practice.

Method of moments

Empirical distribution of a sample should "converge" to the probability distribution. Hence, the corresponding moments should be about equal.

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with PMF or PDF $f(x|\theta_1, \theta_2, \ldots, \theta_n)$. **Method of moments** estimator can be found by equating the moments of the distribution to the moments of the sample as

$$E(X^k) = \frac{1}{n}\sum_{i=1}^{n} X^k$$

until all parameters can be solved. Here, $E(X^k)$ is the moments of distribution and $\sum_{i=1}^{n} X_i^k / n$ is the moments of the sample.

Method of moments

Example: Suppose $X_1, X_2, \ldots, X_n$ are a random sample from $N(\mu, \sigma^2)$. Find the method of moment estimators of μ and σ^2.

The moment generating function of X is $m(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. Thus, the moments of the distribution is

$$m'(0) = (\mu + \sigma^2 t) e^{\mu t + \frac{1}{2}\sigma^2 t^2}\Big|_{t=0} = \mu$$
$$m''(0) = \left(\sigma^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2} + (\mu + \sigma^2 t)^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right)\Big|_{t=0} = \mu^2 + \sigma^2$$

Thus, by equating moments of the distribution and moments of the sample

$$\mu = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

Method of moments

We obtain the method of moments estimator for parameters of a normal distribution

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\end{aligned}$$

The method of moments estimator coincides with our intuition and perhaps gives some credence to both. The method is somewhat more helpful when no obvious estimator suggests itself.

Methods of moments

Example: Let $X_1, X_2, \ldots, X_n$ be a random sample from a binomial distribution with unknown parameters k and p. Find the method of moment estimators.

Equating the first two sample moments to those of the distribution yields the system of equations

$$
\begin{aligned}
kp &= \overline{X} \\
kp(1-p) + k^2p^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2
\end{aligned}
$$

Solving the equations yields the estimates

$$
\begin{aligned}
\hat{k} &= \frac{\overline{X}^2}{\overline{X} - (1/n)\sum_{i=1}^{n}(X_i - \overline{X}^2)} \\
\hat{p} &= \frac{\overline{X}}{\hat{k}}
\end{aligned}
$$

Maximum likelihood estimators

Suppose we flipped a coin 3 times and observed heads, heads, and tail. What is the probability of observing such a result if $p = 0.5$ or $p = 0.6$?

The result of a coin flipping follows a Bernoulli distribution. Thus, the probability of observing heads, heads, and tail is

$$P(HHT) = p \times p \times (1 - p)$$

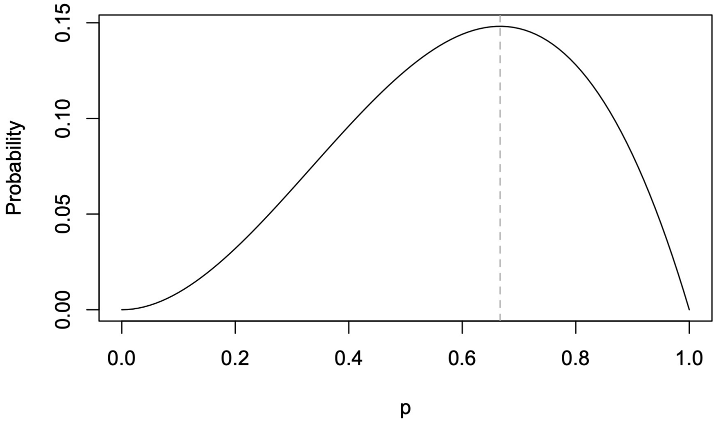
Thus, we have

$$\begin{aligned} P(HHT) &= 0.5 \times 0.5 \times (1 - 0.5) = 0.125 \quad \text{if } p = 0.5 \\ P(HHT) &= 0.6 \times 0.6 \times (1 - 0.6) = 0.144 \quad \text{if } p = 0.6 \end{aligned}$$

In this case, if we do not know the probability of success and want to estimate it from observation, what would be the best estimates?

Maximum likelihood estimates

The probability of getting heads, heads, and tail depends on what the success probability is. Thus, the value of p that makes the observation most likely seems to be a reasonable estimator.



Maximum likelihood estimators

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with PDF or PMF $f(x \big| \theta_1, \theta_2, \ldots, \theta_k)$, the joint PDF or PMF regarded as a function the $\theta_1, \theta_2, \ldots, \theta_k$ is called the **likelihood** function

$$\begin{aligned} L(\theta|X) &= L(\theta_1, \theta_2, \ldots, \theta_k \big| X_1, X_2, \ldots, X_n) \\ &= f(X_1, X_2, \ldots, X_n | \theta_1, \theta_2, \ldots, \theta_k) \end{aligned}$$

Most commonly, the random samples we draw are independent and are from the same distribution. We refer to as **independent and identically distributed (iid)** data. In this case, the likelihood function is

$$L(\theta|X) = \prod_{i=1}^{n} f(x_i|\theta_1, \theta_2, \ldots, \theta_k)$$

Maximum likelihood estimators

Definition: For a partcular sample, let $\hat{\theta}$ be the parameter value at which $L(\theta|X)$ attains its maximum as a function of θ, with X held fixed. $\hat{\theta}$ is called the **maximum likelihood estimator (MLE)** of the parameter θ based on the sample X.

Invariance property of MLEs: If $\hat{\theta}$ is the maximum likelihood estimator of θ, then for any function $\tau(\theta)$, the maximum likelihood estimator for $\tau(\theta)$ is $\tau(\hat{\theta})$.

Maximum likelihood estimators

Let $X_1, X_2, \ldots, X_n$ be the results a n independent Bernoulli trials. We know the PDF of X is $f(x) = p^x(1-p)^{1-x}$. Find the maximum likelihood estimate of p.

The likelihood function is

$$L(p|X) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

The derivative of $L(p|X)$ is

$$L'(p|X) = (\sum x_i)p^{\sum x_i - 1}(1-p)^{n-\sum x_i} + (n - \sum x_i)p^{\sum x_i}(1-p)^{n-\sum x_i - 1}$$

Setting $L'(p|X) = 0$, we have

$$p^{\sum x_i}(1-p)^{n-\sum x_i}\left(\frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p}\right) = 0$$

We solve for p and get

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Maximum likelihood estimators

We finding maximum likelihood estimators, it is often easier to deal with the natural logarithm of the likelihood function. Because logarithm is a monotonically increasing function, parameter values that maximize likelihood function also maximize the logarithm of likelihood.

In the example above, the log likelihood is

$$\log L(p) = (\sum_{i=1}^{n} x_i) \log p + (n - \sum_{i=1}^{n} x_i) \log(1 - p)$$

Thus, the derivative is

$$\frac{d \log L(p)}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1 - p}$$

Setting the derivative to zero and we obtain

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Maximum Likelihood estimators

Let Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Find the maximum likelihood estimator of μ and σ^2.

The likelihood function is

$$\begin{aligned} L(\mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left[-\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2} \right] \\ &= \left(2\pi\sigma^2 \right)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2} \right] \end{aligned}$$

The log likelihood function is

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

Maximum Likelihood estimators

Taking derivative of the log likelihood function with respect to μ and σ^2

$$\frac{\partial \log L}{\partial \mu} = -\frac{\sum_{i=1}^{n}(x_i - \mu)}{\sigma^2}$$

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^4}$$

Setting the derivative with respect to μ and σ to 0, we get

$$\hat{\mu} = \overline{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Evaluating estimators

We want an estimator to be “close” to the true value of the parameter it tries to estimate. The “closeness” can be evaluated in multiple ways:

- **Bias**: how far away is the estimator from the true value on average?
- **Efficiency**: how much uncertainty do we have in the estimator?
- **Consistency**: does the estimator become closer to the true value as sample size increases?
- **Sufficiency**: has the estimator used all available information from the sample to estimator the parameter of interest?

Bias

Definition: An estimator $u(X_1, X_2, \ldots, X_n)$ is an **unbiased estimator** of a parameter θ if $E[u(X_1, X_2, \ldots, X_n)] = \theta$. Otherwise, it is a biased estimator.

Example: Let X be a random variable with mean μ and variance σ^2 and $X_1, X_2, \ldots, X_n$ be a random sample. The sample mean and variance

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

are unbiased estimators of μ and σ^2

Proof:

$$\begin{aligned} E(\overline{X}) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n} n\mu = \mu \end{aligned}$$

Bias

$$
\begin{aligned}
E(s^2) &= E\Big[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\Big] = \frac{1}{n-1}E\Big[\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2)\Big] \\
&= \frac{1}{n-1}\Big[E(\sum_{i=1}^{n}X_i^2)-2E(\sum_{i=1}^{n}X_i\overline{X})+E(\sum_{i=1}^{n}\overline{X}^2)\Big] \\
&= \frac{1}{n-1}\Big[nE(X_i^2)-2nE(\overline{X}^2)+nE(\overline{X}^2)\Big] = \frac{1}{n-1}\Big[nE(X_i^2)-nE(\overline{X}^2)\Big] \\
&= \frac{1}{n-1}\Big[n(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\Big] = \sigma^2
\end{aligned}
$$

using the fact that $Var(X) = E(X^2) - [E(x)]^2$. Hence, sample mean and variance are unbiased estimator of μ and σ^2.

Consistency

A sequence $X_1, X_2, \ldots, X_n$ of random variables **converges in probability** towards the random variable X if for $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

A sequence $X_1, X_2, \ldots, X_n$ of random variables **converges almost surely** or **converge with probability 1** to a random variable X if

$$P(\lim_{n\to\infty} X_n = X) = 1$$

As sample size $n \to \infty$, if an estimator $\hat{\theta}$ converges to θ in probability, $\hat{\theta}$ is a consistent estimator. If $\hat{\theta}$ converges to θ in probability 1, then $\hat{\theta}$ is a strongly consistent estimator.

Bias and consistency

Bias and consistency describe different aspects of an estimator and are not equivalent concepts. An estimator can be unbiased but inconsistent, or biased but consistent. However, if an estimator is unbiased and it converges to a value as $n \to \infty$, then it must be an consistent estimator and converges to the true value of the parameter.

Unbiased but inconsistent: For a random sample $X_1, X_2, \ldots, X_n$, X_1 is an unbiased estimator of the population mean μ because $E(X_1) = \mu$, but it is clear that the property of X_1 as an estimator does not change with sample size n and it is not a consistent estimator.

Biased but consistent: The maximum likelihood estimate of population variance σ^2, $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2$ is an biased estimator of σ^2, but it is a consistent estimator.

Efficiency

Definition: The **mean square error (MSE)** of an estimator $u(X)$ of a parameter θ is the function of θ defined as $E(u(X) - \theta)^2$.

Mean square error can be decomposed into two components as

$$E(u(X) - \theta)^2 = Var[u(X)] + (E[u(X)] - \theta)^2$$

For an unbiased estimator, i.e., $E[u(X)] = \theta$, the mean square error is equal to the variance of the estimator.

Sufficiency

Definition: A statistic $u(X)$ is a sufficient statistic for θ if the conditional distribution of the sample $X_1, X_2, \ldots, X_n$ given the value of $u(X)$ does not depend on θ.

Fisher-Neyman Factorization Theorem: Let $X_1, X_2, \ldots, X_n$ denote random variables with joint PDF or PMF $f(X_1, X_2, \ldots, X_n|\theta)$, which depends on the parameter θ. The statistic $Y = u(X_1, X_2, \ldots, X_n)$ is a sufficient statistic for θ if and only if

$$f(X_1, X_2, \ldots, X_n|\theta) = \phi(u(X_1, X_2, \ldots, X_n)|\theta)h(X_1, X_2, \ldots, X_n)$$

where ϕ depends on $X_1, X_2, \ldots, X_n$ only through $u(X_1, X_2, \ldots, X_n)$ and $h(X_1, X_2, \ldots, X_n)$ does not depend θ.

Sufficiency

Rao-Blackwell Theorem: Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with PDF or PMF $f(x|\theta)$. Let $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ be a sufficient statistic for θ, and let $Y_2 = u_2(X_1, X_2, \ldots, X_n)$ be an unbiased estimator of θ, where Y_2 is not a function of Y_1 alone. Then $E(Y_2|Y_1)$ defines a statistic $u(Y_1)$, a function of the sufficient statistic Y_1, which is an unbiased estimator of θ, and its variance is less than that of Y_2.

The important implication is that for every other unbiased estimator of θ, we can always find an unbiased estimator based on the sufficient statistic that has a variance at least as small as the first unbiased estimator. Hence, we might as well search for an unbiased estimator by considering only those unbiased estimators based on sufficient statistics.

Asymptotic properties of maximum likelihood estimator

Maximum likelihood estimators is widely used because it has many desirable properties. Under regular conditions, maximum likelihood estimator is,

- asymptotically consistent;
- approximately normally distributed;
- asymptotically efficient;
- is sufficient for a large family of distributions

In the following slides, rough sketch proof of these properties of the maximum likelihood estimator is provided. These contents are much more advanced than necessary for an introductory course like this. Therefore, these materials are optional.

Consistency of maximum likelihood estimator

Law of large numbers: If the distribution of the i.i.d. sample $X_1, X_2, \ldots, X_n$ is such that X_i has finite expectation, i.e., $|E(X)| < \infty$, then the sample average

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \to E(X)$$

converges to its expectation in probability.

Using law of large numbers, the likelihood function normalized by $1/n$

$$\frac{1}{n} \ln \left[L(\theta)\right] = \frac{1}{n} \sum_{i=1}^{n} \ln \left[f(X_i|\theta)\right] \to E\left[\ln f(X_i|\theta)\right]$$

in probability. We also know that the maximum likelihood estimator $\hat{\theta}$ maximizes $\ln\left[L(\theta)\right]$ and thus $\frac{1}{n}\ln\left[L(\theta)\right]$. Since $\frac{1}{n}\ln\left[L(\theta)\right]$ converges to $E\left[\ln f(X_i|\theta)\right]$, $\hat{\theta}$ should also converge to the value that maximizes $E\left[\ln f(X_i|\theta)\right]$.

Consistency of maximum likelihood estimator

What value maximizes $E\left[\ln f(X_i|\theta)\right]$?

Let θ_0 be the true value of θ and θ_1 be any other estimates of θ.

$$E\left[\ln f(X_i|\theta_1)\right] - E\left[\ln f(X_i|\theta_0)\right] = E\left[\ln \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right]$$

Because $\ln(x)$ is a concave function,

$$E\left[\ln \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right] < \ln\left[E\left(\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right)\right] = \ln\left(\int_{-\infty}^{\infty} \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} f(X_i|\theta_0)dx\right)$$
$$= \ln\left(\int_{-\infty}^{\infty} f(X_i|\theta_1)dx\right) = \ln(1) = 0.$$

Thus, θ_0 maximizes $E\left[\ln f(X_i|\theta)\right]$. Therefore, we conclude that $\hat{\theta}$ converges to θ_0 and is thus a consistent estimator.

Normality of maximum likelihood estimator

The maximum likelihood estimator $\hat{\theta}$ for a parameter θ has asymptotic normal distribution. To prove this, we first note that we obtain $\hat{\theta}$ by setting the derivatives of the log likelihood function to 0, i.e.,

$$\frac{\partial[\ln L(\hat{\theta})]}{\partial\theta} = 0$$

Approximating the left hand side of the equation using the first two terms in the Taylor expansion, we have

$$\frac{\partial[\ln L(\theta)]}{\partial\theta} + (\theta - \hat{\theta})\frac{\partial^2[\ln L(\theta)]}{\partial\theta^2} \approx 0$$

Rearranging the equation, we obtain

$$\hat{\theta} - \theta = \frac{\frac{\partial[\ln L(\theta)]}{\partial\theta}}{-\frac{\partial^2[\ln L(\theta)]}{\partial\theta^2}}$$

Normality of maximum likelihood estimator

We first consider the numerator. Recall that

$$\ln L(\theta) = \ln f(X_1|\theta) + \ln f(X_2|\theta) + \cdots + \ln f(X_n|\theta)$$

and thus

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial[\ln f(X_i|\theta)]}{\partial \theta}$$

This is the sum of n independent and identically distributed random variables and thus, by central limit theorem, has an approximate normal distribution. The mean of the distribution is

$$n\int_{-\infty}^{\infty} \frac{\partial[\ln f(X_i|\theta)]}{\partial \theta} f(X_i|\theta)dx = n\int_{-\infty}^{\infty} \frac{\partial[f(X_i|\theta)]}{\partial \theta} \frac{f(X_i|\theta)}{f(X_i|\theta)}dx$$

$$= n\int_{-\infty}^{\infty} \frac{\partial[f(X_i|\theta)]}{\partial \theta}dx = n\frac{\partial}{\partial \theta}\left[\int_{-\infty}^{\infty} f(X_i|\theta)\right] = n\frac{\partial(1)}{\partial \theta} = 0.$$

Normality of maximum likelihood estimator

We can consider the variance of the distribution. We just show that

$$\int_{-\infty}^{\infty} \frac{\partial[\ln f(X_i|\theta)]}{\partial\theta} f(X_i|\theta)dx = 0$$

Take derivative with respect to θ, we have

$$\int_{-\infty}^{\infty} \left\{ \frac{\partial^2[\ln f(X_i|\theta)]}{\partial\theta^2} f(X_i|\theta) + \frac{\partial[\ln f(X_i|\theta)]}{\partial\theta} \frac{\partial[f(X_i|\theta)]}{\partial\theta} \right\} dx = 0$$

Note that

$$\frac{\partial[f(X|\theta)]}{\partial\theta} = \frac{\partial[\ln f(X|\theta)]}{\partial\theta} f(X|\theta)$$

Using this in the equation above, we have

$$\int_{-\infty}^{\infty} \left\{ \frac{\partial[\ln f(X|\theta)]}{\partial\theta} \right\}^2 f(X|\theta)dx = -\int_{-\infty}^{\infty} \frac{\partial^2[\ln f(X|\theta)]}{\partial\theta^2} f(X|\theta)dx$$

Normality of maximum likelihood estimator

Because $E(\frac{\partial \ln f(X|\theta)}{\partial \theta}) = 0$, this last expression provides the variance of $\frac{\partial \ln f(X|\theta)}{\partial \theta}$. Thus,

$$\begin{aligned} Var(\frac{\partial \ln L(\theta)}{\partial \theta}) &= Var(\sum_{i=1}^{n} \frac{\partial \ln f(X|\theta)}{\partial \theta}) = nVar(\frac{\partial \ln f(X|\theta)}{\partial \theta}) \\ &= -nE\left[\frac{\partial^2 \ln f(X|\theta)}{\partial \theta^2}\right] \end{aligned}$$

This is commonly denoted as $I(\theta)$ and is referred to as the Fisher information for the sample. Thus far, we have shown that the numerator in the expression of $\hat{\theta} - \theta$ has a normal distribution with mean 0 and variance $I(\theta)$. We now consider the denominator.

Normality of maximum likelihood estimator

The denominator is also the sum of independent and identically distributed random variables

$$-\frac{\partial^2 \ln L(\theta)}{\partial\theta^2} = \sum_{i=1}^{n} -\frac{\partial^2 \ln f(X|\theta)}{\partial\theta^2}$$

Based on law of large numbers, as sample size n increases,

$$-\frac{\partial^2 \ln L(\theta)}{\partial\theta^2} \to nE\left(\frac{\partial^2 f(X|\theta)}{\partial\theta^2}\right) = I(\theta).$$

Thus, we have shown that asymptotically,

$$\frac{\partial \ln L(\theta)}{\partial\theta} \sim N(0, I(\theta))$$

$$-\frac{\partial^2 \ln L(\theta)}{\partial\theta^2} \to I(\theta)$$

Normality of maximum likelihood estimator

We therefore have, asymptotically,

$$Var(\hat{\theta} - \theta) = Var\left(\frac{\frac{\partial \ln L(\theta)}{\partial \theta}}{-\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}}\right) = Var\left(\frac{\frac{\partial \ln L(\theta)}{\partial \theta}}{I(\theta)}\right)$$

$$= \frac{1}{I(\theta)^2} Var\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right) = \frac{1}{I(\theta)}$$

We therefore can see, at least roughly, that $(\hat{\theta} - \theta) \sim N(0, I(\theta)^{-1})$. That is, for any maximum likelihood estimator $\hat{\theta}$, we have asymptotically

$$\hat{\theta} \sim N(\theta, I(\theta)^{-1})$$

Efficiency of maximum likelihood estimator

Cramer–Rao lower bound: Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a population with density function $f(x|\theta)$. Let $Y = u(X_1, X_2, \ldots, X_n)$ be an unbiased estimator of θ. Then

$$Var(Y) \geqslant \frac{1}{I(\theta)}$$

where $I(\theta)$ is the **Fisher information** of the sample and is defined as

$$I(\theta) = E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right] = -E\left(\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right) = -nE\left(\frac{\partial^2 \ln f(x|\theta)}{\partial \theta^2}\right)$$

Because the maximum likelihood estimator has asymptotic variance of $I(\theta)^{-1}$. The Cramer–Rao lower bound suggests that maximum likelihood estimator has the minimum variance that an unbiased estimator can achieve. Thus, maximum likelihood estimator is asymptotically efficient.