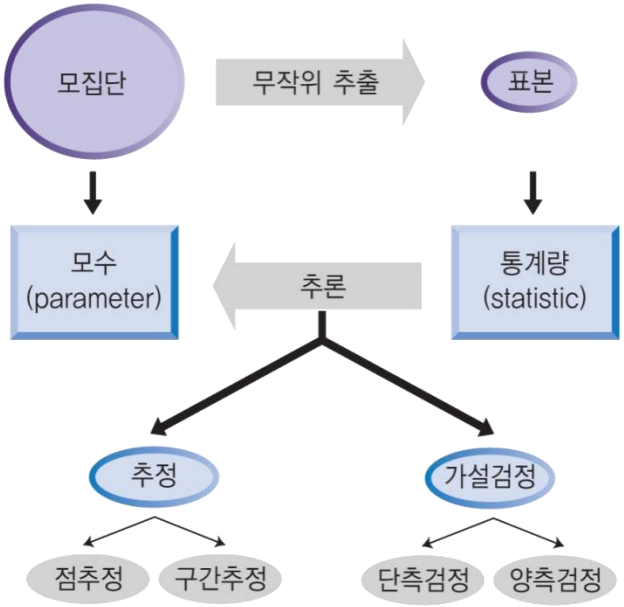
11.1 추정 1

추론(inference)

표본의 정보를 바탕으로 모집단에 관한 의사결정, 추정,예측을 하는 것으로 추정 및 가설검정으로 구분함



추정의 개념

- ✓ 추정량은 표본정보를 이용하여 알지 못하는 모수의 참값을 추정하는 방법이며, 알지 못하는 모수가 θ 라면 추정량은 $\hat{\theta}$ 으로 표기한다.
- ✓ 추정값은 수치로 계산된 $\hat{\theta}$ 의 값이다. (추정량은 여러 가지가 있음)

점추정량	모수를 하나의 값으로 추정하는 방법
구간추정량	모수의 값이 빈번히 포함되는 구간을 추정하는 방법

11-1-2. 좋은 추정량이란?

모집단에서 무수히 많은 표본을 추출하여 각 표본 추정량의 값을 계산했을 때 추정량이 바람직하기 위해서는 추정값들의 확률분포가 모수를 중심으로 밀집되어야 할 것이다.

이 밀집성의 정도는 **평균제곱오차(mean squared error, MSE)**로 측정할 수 있다.

$$MSE(\theta)=E[(\hat{\theta}-\theta)^2]=Var(\hat{\theta})+(E(\hat{\theta})-\theta)^2$$

분산 편의^2

참값=0.5	n=5	n=10	n=500
OLS(평균)	0.507	0.518	0.501
OLS(분산)	1.9353	0.5646	0.0081
LAD(평균)	0.402	0.49	0.504
LAD(분산)	2.6037	0.9765	0.0122
MLE(평균)	0.499	0.454	0.507
MLE(분산)	2.189	0.5377	0.0081
MSE(OLS)	1.935349	0.564924	0.008101
MSE(LAD)	2.613304	0.9766	0.012216
MSE(MLE)	2.189001	0.539816	0.008149

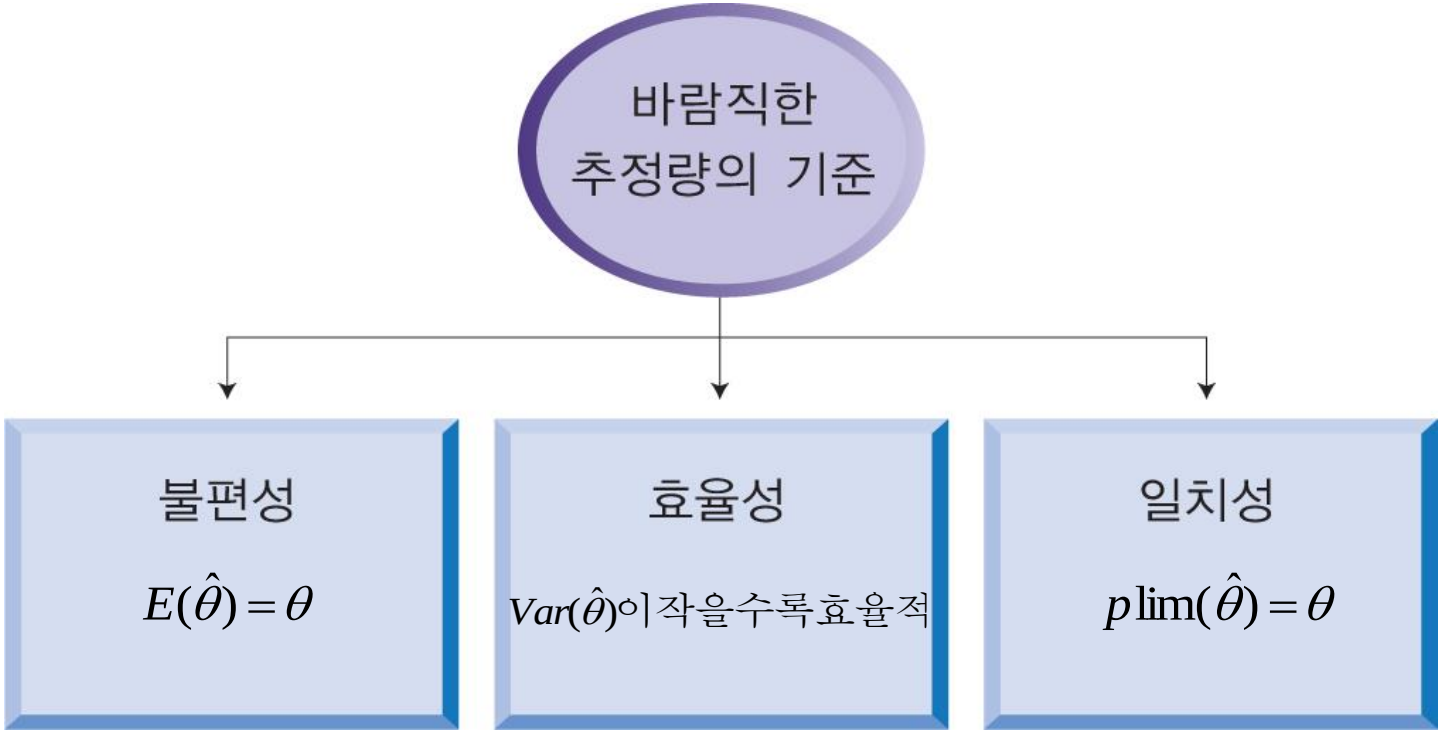
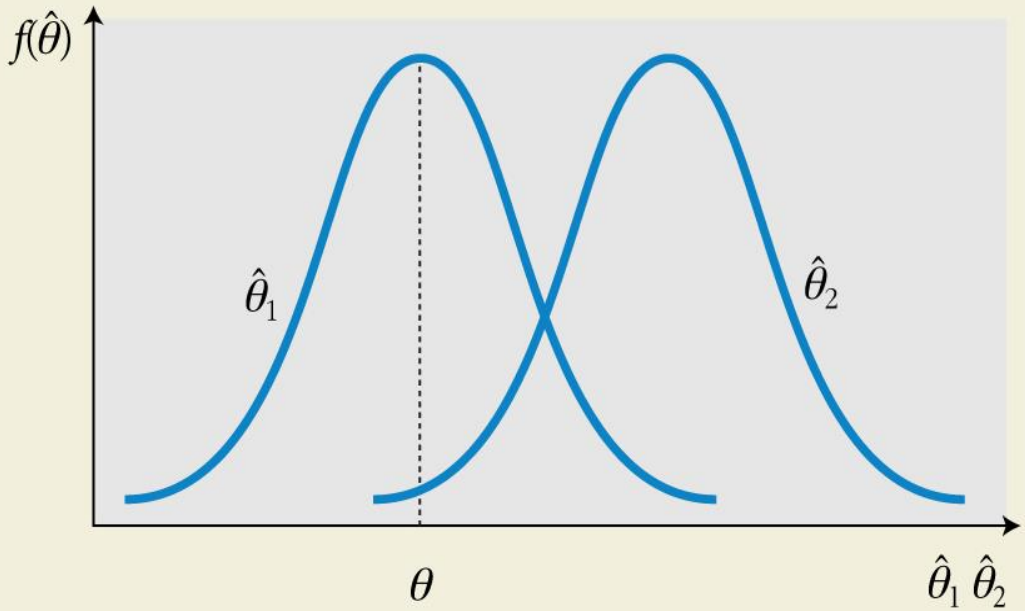

바람직한
추정량의 기준
불편성
$E(\hat{\theta}) = \theta$
효율성
$Var(\hat{\theta})$이 작을수록 효율적
일치성
$p\lim(\hat{\theta}) = \theta$

그림 추정량 $\hat{\theta}_1$과 $\hat{\theta}_2$의 확률밀도함수



불편성

추정량 $\hat{\theta}$ 의 기대값이 모수 θ와 일치하면 $(E(\hat{\theta})=\theta)$, $\hat{\theta}$은 모수 θ의 **불편추정량**이다.

표본평균 $\overline{X}$, 표본분산 S^2과 표본비율 $\hat{p}$ 은 불편추정량이다.

θ와 $(E(\hat{\theta})$의 차이를 편의(bias)라고 한다.

편의 = $(E(\hat{\theta}) - \theta)$

$\overline{X}$는 μ의 불편추정량이다.

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

s^2는 σ^2의 불편추정량이다

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

예

평균 μ와 분산 σ^2을 가지는 모집단으로부터 표본의 크기(n)가 짝수인 확률표본 $x_1, x_1, \cdots, x_n$을 추출하였다. 그리고 다음과 같은 두 추정량이 있다고 하자.

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \hat{\theta}_2 = \frac{1}{n/2}\sum_{i=1}^{n/2} x_{2i-1}$$

두 추정량이 모두 μ의 불편추정량임을 증명해 보자.

풀이

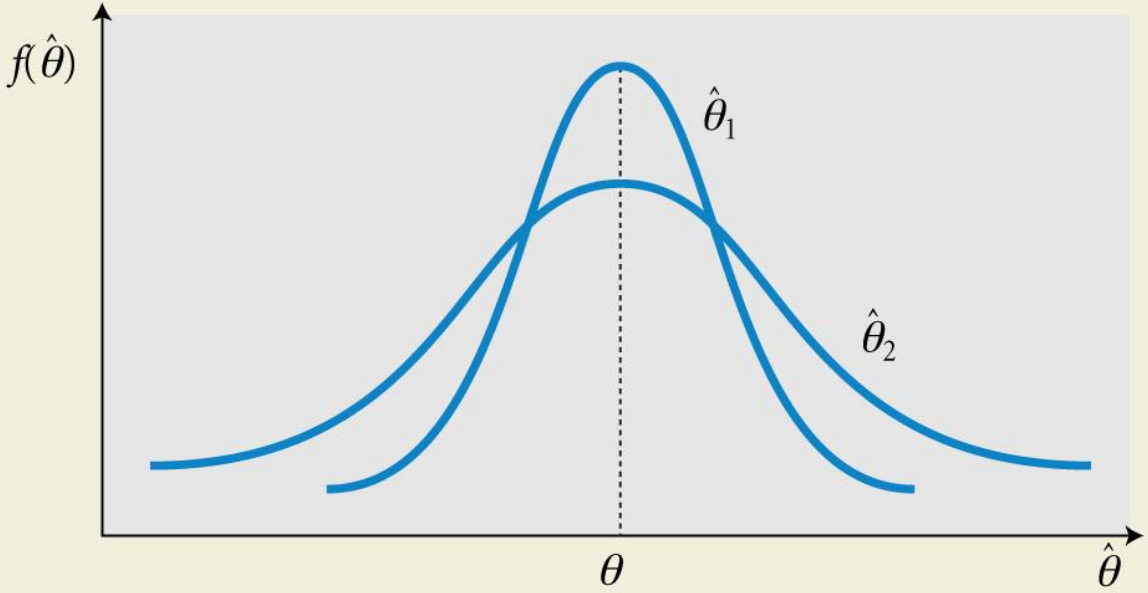
$$E(\hat{\theta}_1)=E(\frac{1}{n}\sum_{i=1}^{n}x_i)=\frac{1}{n}E(\sum_{i=1}^{n}x_i)=\frac{1}{n}\sum_{i=1}^{n}E(x_i)=\frac{1}{n}\times n\times\mu=\mu$$

$E(\hat{\theta}_1)=\mu$이므로 $\hat{\theta}_1$은 μ의 불편추정량이다. 또한

$$E(\hat{\theta}_2)=E(\frac{1}{n/2}\sum_{i=1}^{n/2}x_{2i-1})=\frac{1}{n/2}E(\sum_{i=1}^{n/2}x_{2i-1})=\frac{1}{n/2}\sum_{i=1}^{n/2}E(x_{2i-1})=\frac{1}{n/2}\times\frac{n}{2}\times\mu=\mu$$

$E(\hat{\theta}_2)=\mu$이므로 $\hat{\theta}_2$는 μ의 불편추정량이다.

그림 두 불편추정량의 확률밀도함수



효율성

1. $\hat{\theta}_1$과 $\hat{\theta}_2$가 모두 불편추정량이라고 하자. 이때

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

이면 $\hat{\theta}_1$이 $\hat{\theta}_2$ 보다 더 효율적이라고 한다.

2. $\hat{\theta}_1$이 다른 모든 불편추정량보다 작은 분산을 가지면 $\hat{\theta}_1$은 **가장 효율적인(most efficient)**혹은 **최소분산(minimum variance)** 불편추정량이다.

예 앞 예제의 두 추정량에 대한 분산을 각각 계산하고 어떤 추정량이 더 효율적인지 증명해 보라.

풀이

$$Var(\hat{\theta}_1)=Var(\frac{1}{n}\sum_{i=1}^{n}x_i)=\frac{1}{n^2}Var(\sum_{i=1}^{n}x_i)=\frac{1}{n^2}\sum_{i=1}^{n}Var(x_i)$$

$$=\frac{1}{n^2}(n\cdot\sigma^2)=\frac{\sigma^2}{n}$$

$$Var(\hat{\theta}_2)=Var\left(\frac{1}{n/2}\sum_{i=1}^{n/2}x_{2i-1}\right)=\left(\frac{1}{n/2}\right)^2Var\left(\sum_{i=1}^{n/2}x_{2i-1}\right)=\frac{1}{n^2/4}\sum_{i=1}^{n/2}Var(x_{2i-1})$$

$$=\frac{1}{n^2/4}\left(\frac{n}{2}\cdot\sigma^2\right)=\frac{2}{n}\sigma^2$$

$\hat{\theta}_1$의 분산이 $\hat{\theta}_2$의 분산보다 작으므로 추정량 $\hat{\theta}_1$ 이 추정량 $\hat{\theta}_2$보다 더 효율적이다.

예 평균을 계산하는 다음의 두 추정량 중 어느 추정량이 바람직한 추정량인가?

풀이

$$\overline{X}=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3$$

$$\overline{X}=\frac{1}{4}X_1+\frac{1}{2}X_2+\frac{1}{4}X_3$$

먼저, 불편성을 검토해 보자.

$$E(\overline{X})=\frac{1}{3}3\mu=\mu$$

$$E(\overline{X})=\frac{1}{4}\mu+\frac{1}{2}\mu+\frac{1}{4}\mu=\mu$$

따라서, 둘 다 불편추정량이다. 다음으로 분산을 비교해보자.

$$Var(\overline{X})=\frac{1}{9}\sigma^2+\frac{1}{9}\sigma^2+\frac{1}{9}\sigma^2=\frac{1}{3}\sigma^2$$

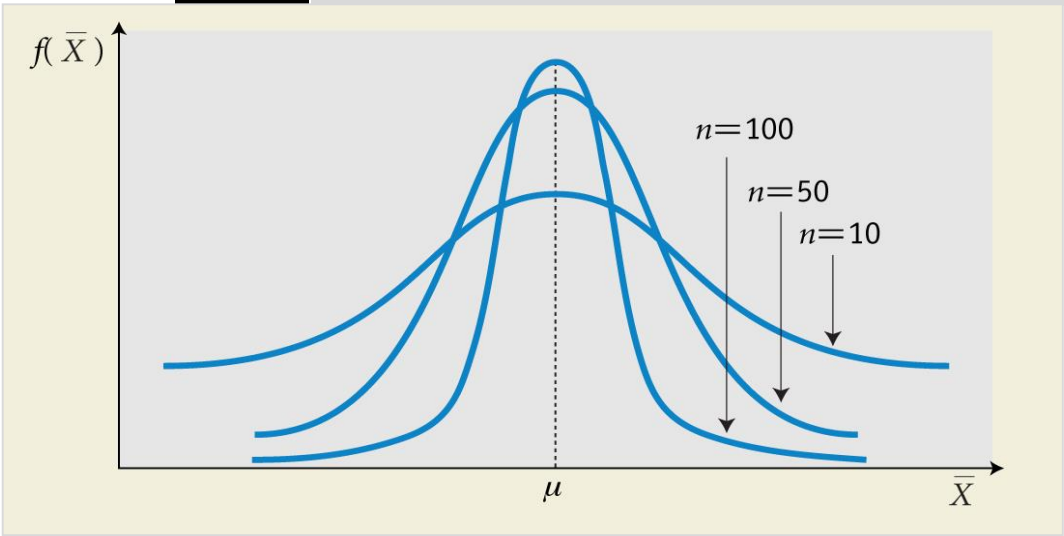
$$Var(\overline{X})=\frac{1}{16}\sigma^2+\frac{1}{4}\sigma^2+\frac{1}{16}\sigma^2=\frac{6}{16}\sigma^2$$

$\therefore$ $\overline{X}$가 $\overline{X}$보다 효율적인 추정량이다.

일치성

바람직한 추정량을 선택하는 또 다른 기준으로는 표본크기가 무한히 증가할 때 추정량 $\hat{\theta}$이 모수 θ에 근접하려는 특성인 **일치성(consistency)**을 들 수 있다.

그림 n의 증가에 따른 $\bar{x}$의 확률밀도함수



$n \to \infty$일 때 임의의 $\varepsilon > 0$에 대해

$$P[\,|\,\hat{\theta} - \theta\,| \le \varepsilon] \to 1$$

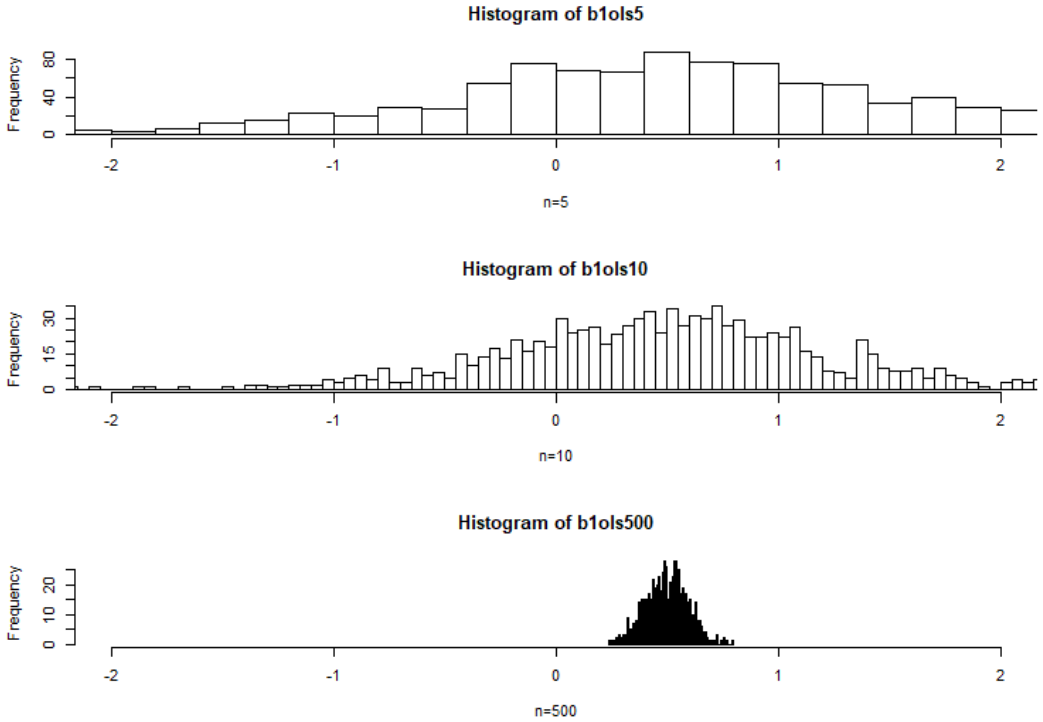
이 성립되면 $\hat{\theta}$은 일치추정량이다. 이를 간단히

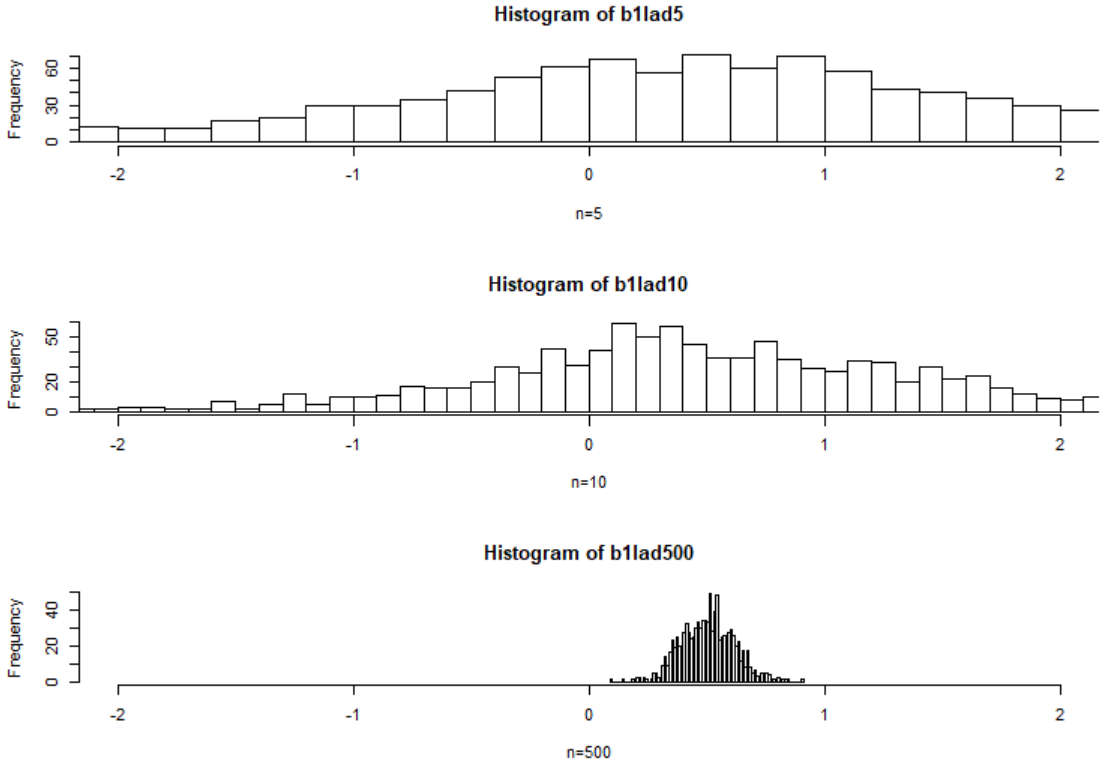
$$\text{plim}\ \hat{\theta} = \theta$$

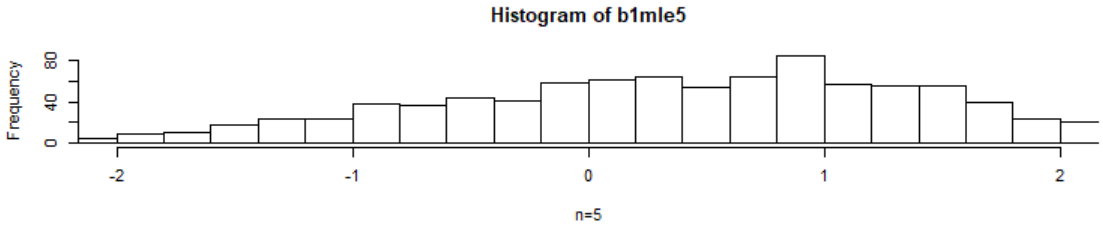
로 표시할 수 있으며, 이때 θ를 $\hat{\theta}$의 **확률극한**(probability limit)이라고 한다.

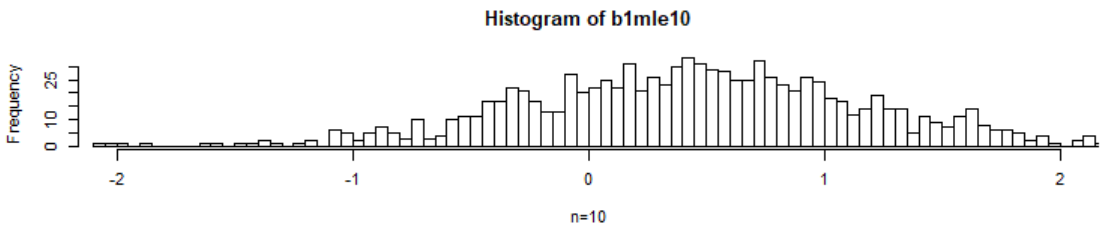
모형	Y = βo + β1 x + u
참값	β1 = 0.5

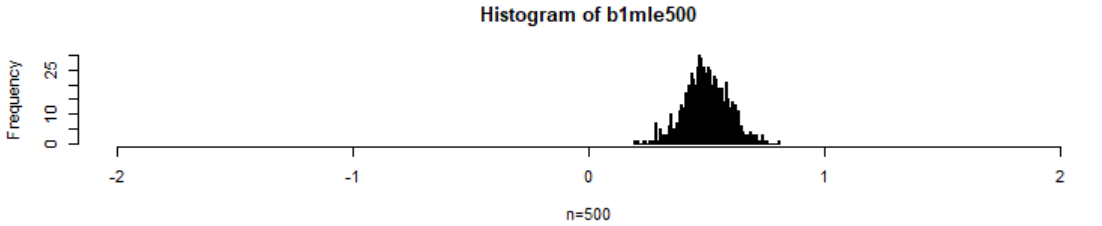
구분	n=5	n=10	n=500
OLS	0.507	0.518	0.501
LAD	0.402	0.49	0.504
MLE	0.499	0.454	0.507




Histogram of b1lad5
Frequency
0
30
60
-2
-1
0
1
2
n=5
Histogram of b1lad10
Frequency
0
20
50
-2
-1
0
1
2
n=10
Histogram of b1lad500
Frequency
0
20
40
-2
-1
0
1
2
n=500


Histogram of b1mle5
Frequency
0
40
80
-2
-1
0
1
2
n=5


Histogram of b1mle10
Frequency
0
10
25
-2
-1
0
1
2
n=10


Histogram of b1mle500
Frequency
0
10
25
-2
-1
0
1
2
n=500