

# 10-1 표본분포 2



# 10-1-1. 중심극한정리

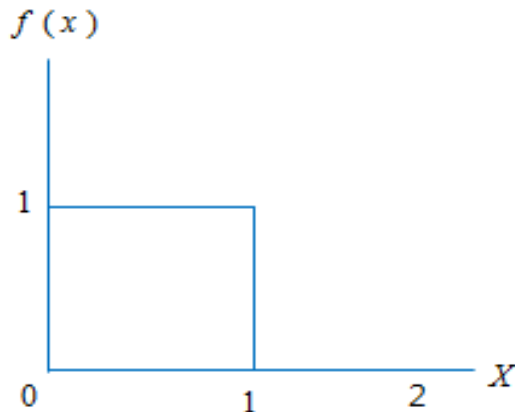
## 배경

관찰된 자료의 모집단이 정규분포에 따를 경우 관찰자료 역시 정규분포에 따르고, 관찰된 자료의 모집단이 실제로 정규분포가 아니면 관찰 자료 역시 정규분포에 따르지 않는다.

그러나 관찰된 자료의 모집단이 실제로 정규분포가 아닌 경우에도 중심극한정리에 의해 정규확률분포를 이용한 추정량의 근사확률을 구할 수 있다.

## 예

확률변수  $X$ 가 균등분포  $U(0,1)$ 을 따른다고 할 때  $X \sim U(0,1)$ 이며,  $X$ 는 0과 1사이에서 균등한 분포를 갖는 연속형 확률변수이다.

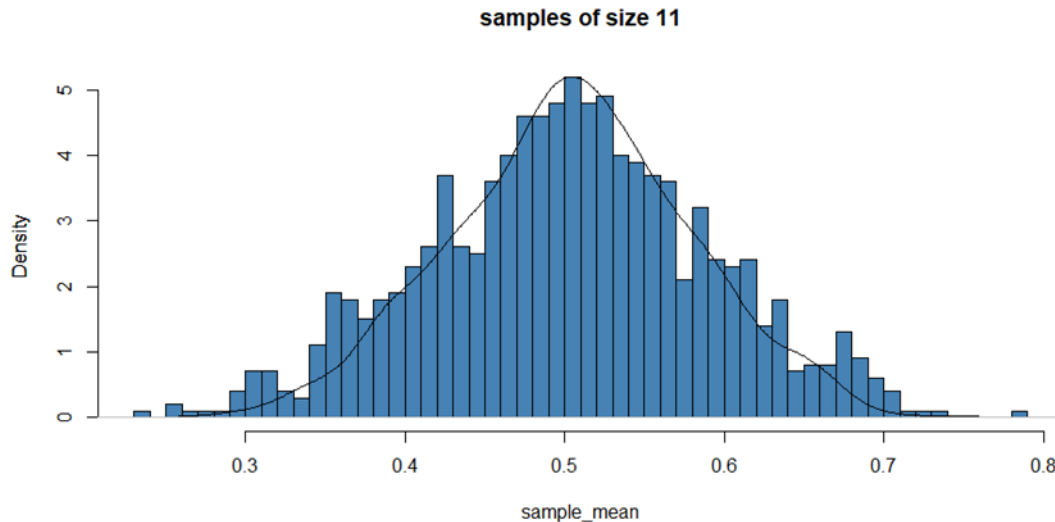


이러한 확률분포로부터 크기  $n=11$ 인 확률표본을 1,000개 추출하는 실험을 해 보자.  
1,000개 표본 중 처음 5개의 표본이 다음과 같다고 하면,  
각 표본의 평균  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_5$ 을 구할 수 있다.

| 표본 | 관측값   |       |       |       |       |       |       |       |       |       |       | $\bar{x}$ |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------|
| 1  | 0.316 | 0.377 | 0.357 | 0.548 | 0.564 | 0.312 | 0.217 | 0.373 | 0.535 | 0.963 | 0.746 | 0.482     |
| 2  | 0.481 | 0.611 | 0.478 | 0.512 | 0.611 | 0.841 | 0.152 | 0.956 | 0.604 | 0.825 | 0.402 | 0.589     |
| 3  | 0.233 | 0.724 | 0.231 | 0.779 | 0.077 | 0.839 | 0.779 | 0.487 | 0.140 | 0.498 | 0.221 | 0.455     |
| 4  | 0.907 | 0.149 | 0.624 | 0.609 | 0.291 | 0.398 | 0.596 | 0.371 | 0.929 | 0.761 | 0.949 | 0.599     |
| 5  | 0.809 | 0.055 | 0.299 | 0.466 | 0.625 | 0.088 | 0.958 | 0.742 | 0.955 | 0.885 | 0.598 | 0.589     |

1,000개 표본으로부터 구한 표본평균  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{1000}$ 을 이용하여 막대그림표를 그리면 다음과 같다.

즉, 균등분포로부터 구한 표본평균의 분포가 정규분포와 근사한 분포를 갖는다



```
> (mean(sample_avgs))  
[1] 0.5019778  
> (sd(sample_avgs))  
[1] 0.08852728  
> (var(sample_avgs))  
[1] 0.007837079
```

$X \sim U(0,1)$ 에서  $E(X) = \frac{1}{2}$ ,  $Var(X) = \frac{1}{12}$ 이므로  $n=11$ 인 경우 표본평균  $\bar{X}$ 의 평균과 분산은 다음과 같다.

$$E(\bar{X}) = \frac{1}{n}nE(X) = E(X) = \frac{1}{2}$$

$$Var(\bar{X}) = \frac{1}{n^2}nVar(X) = \frac{1}{n}Var(X) = \frac{1}{11} \times \frac{1}{12} = \frac{1}{132} = 0.008$$

$$\sqrt{Var(\bar{X})} = \sqrt{0.008} = 0.087$$

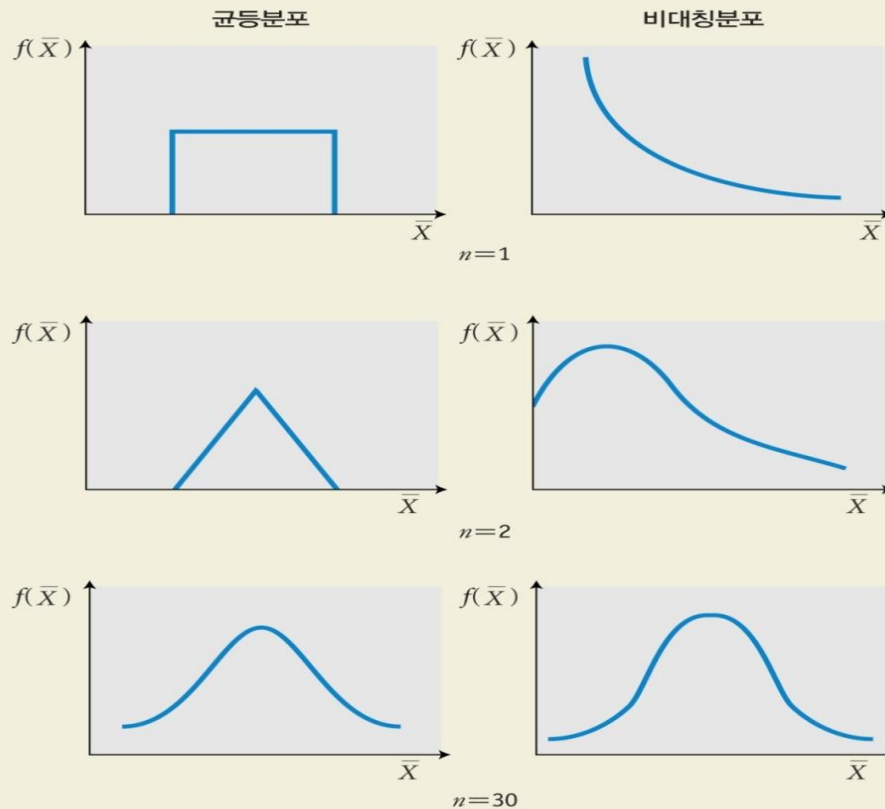
즉, 표본평균  $\bar{X}$ 는 근사적으로  $N(\frac{1}{2}, \frac{1}{132})$  을 따른다.

1,000개 표본의 표본평균과 표준편차는 각각 0.501와 0.088로서 모집단의 이론적인 평균과 표준편차에 근접함을 알 수 있다.

| Variable | Obs  | Mean     | Std. Dev. |
|----------|------|----------|-----------|
| xbar     | 1000 | .5019778 | .08852728 |

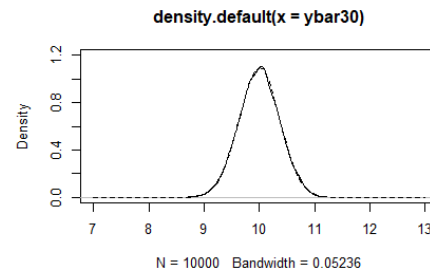
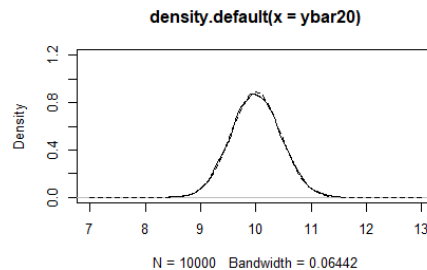
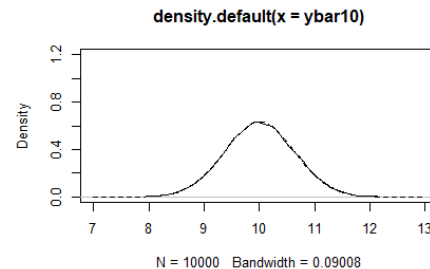
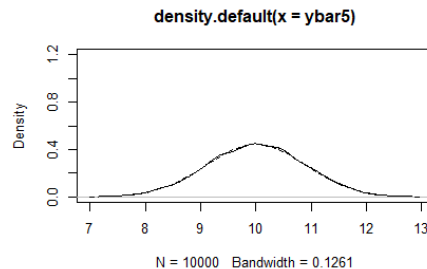
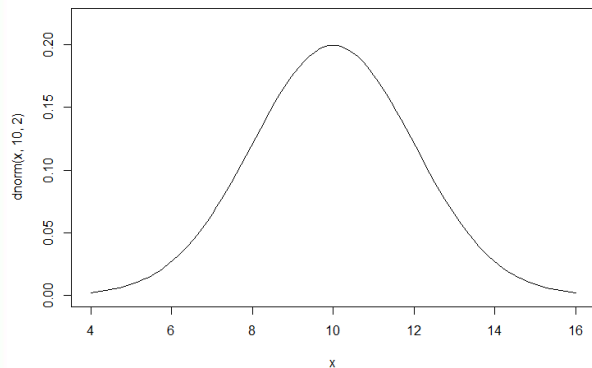
## 그림

$n$ 의 변화에 따른  $\bar{X}$ 의  
확률밀도함수의 모양



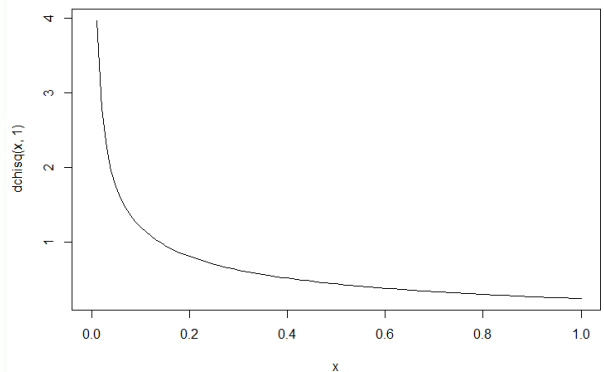
## 그림

n의 변화에 따른  $\bar{X}$ 의 확률밀도함수의 모양

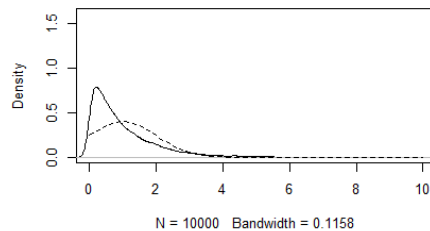


## 그림

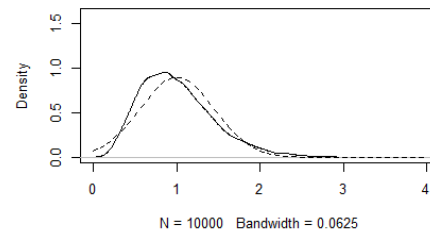
n의 변화에 따른  $\bar{X}$ 의 확률밀도함수의 모양



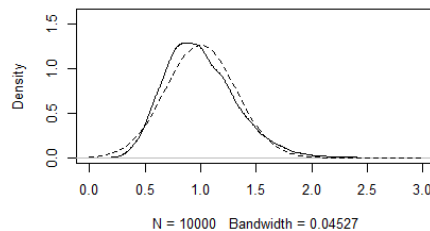
density.default(x = ybar2)



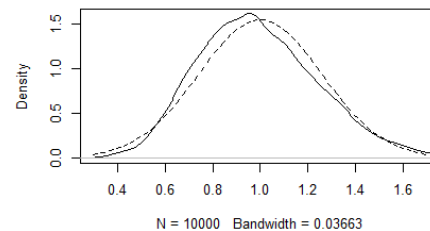
density.default(x = ybar10)



density.default(x = ybar20)

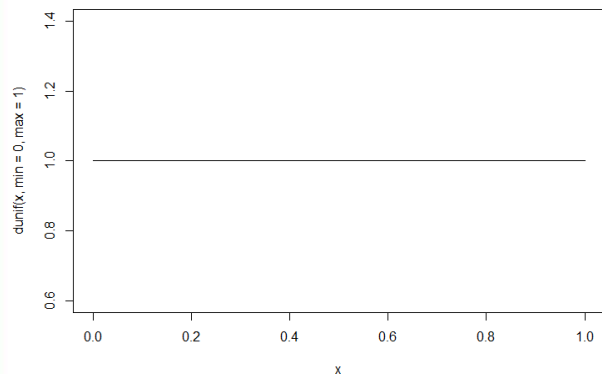


density.default(x = ybar30)

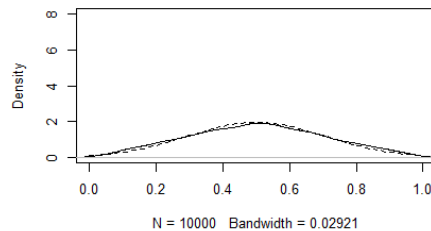


## 그림

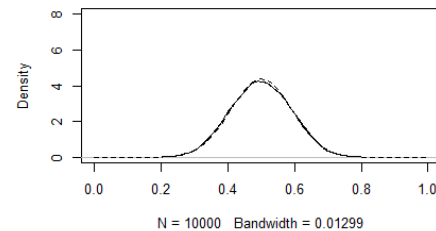
n의 변화에 따른  $\bar{X}$ 의 확률밀도함수의 모양



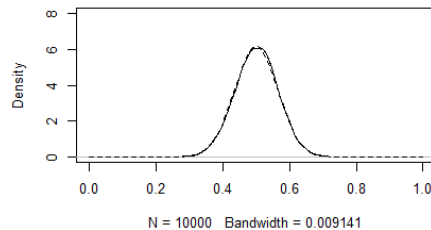
density.default(x = ybar2)



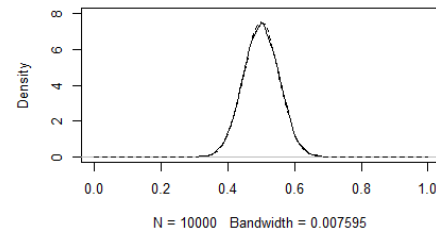
density.default(x = ybar10)



density.default(x = ybar20)



density.default(x = ybar30)



## 중심극한정리

평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 확률분포로부터 크기가  $n$ 인 확률표본  $(X_1, X_2, \dots, X_n)$ 을 추출할 때, 표본평균  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 는  $n$ 이 클수록 평균이  $\mu$ 이고 분산이  $\frac{\sigma^2}{n}$ 인 정규분포와 근사한 분포를 갖는다.

즉,  $\bar{X}$ 의 분포는

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{과 같이 표현한다.}$$

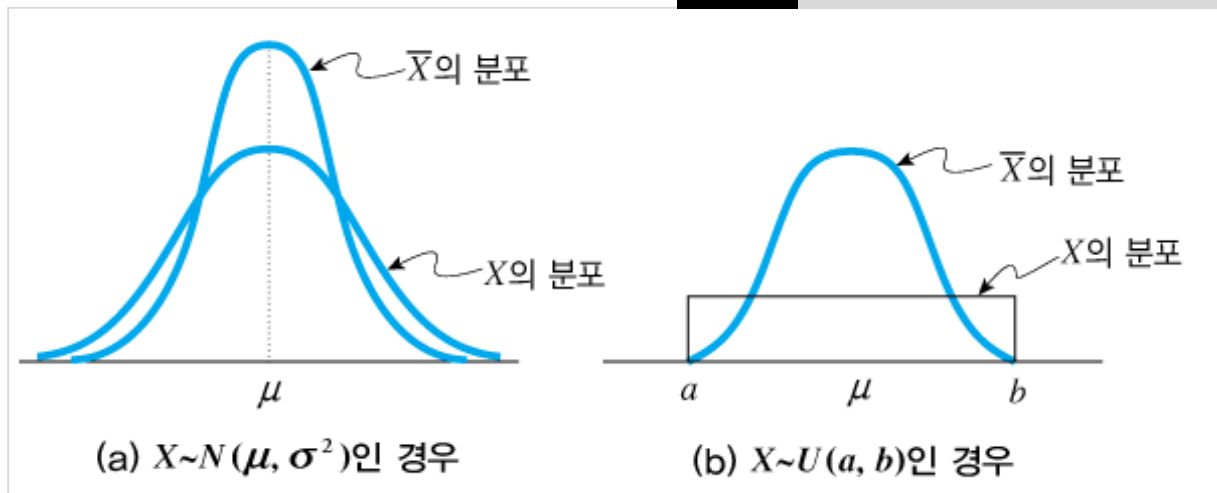
만약에 확률표본  $(X_1, X_2, \dots, X_n)$  이 평균  $\mu$ 와 분산  $\sigma^2$ 을 갖는 정규분포에서 추출되었다면, 표본평균  $\bar{X}$ 의 분포는  $n$ 의 크기에 관계없이 평균  $\mu$ 와

분산  $\frac{\sigma^2}{n}$ 을 갖는 정규분포를 따른다.

즉,  $\bar{X}$ 의 분포는

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{이다.}$$

## 그림 $X$ 와 $\bar{X}$ 의 분포



$\bar{X}$ 의 분포는  $X$ 의 분포와 중심은 같으나 분산은 작아지며, 분포의 모양은  $X$ 의 분포와 관계없이 표본의 수  $n$ 이 커질수록 정규분포 모양을 갖는다.  
위 그림은  $X$ 의 분포가 정규분포인 경우와 균등분포인 경우에 대해 그림으로 표현한 것이다.

## 10-1-2. 표본평균의 표준화

평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 확률변수의  $X$ 를 표준화하는 방법은 평균을 빼주고 표준편차로 나눈다. 따라서 평균이  $\mu$ 이고 분산이  $\frac{\sigma^2}{n}$ 인 표본평균  $\bar{X}$ 의 표준화된

확률변수는  $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$  이다.

