

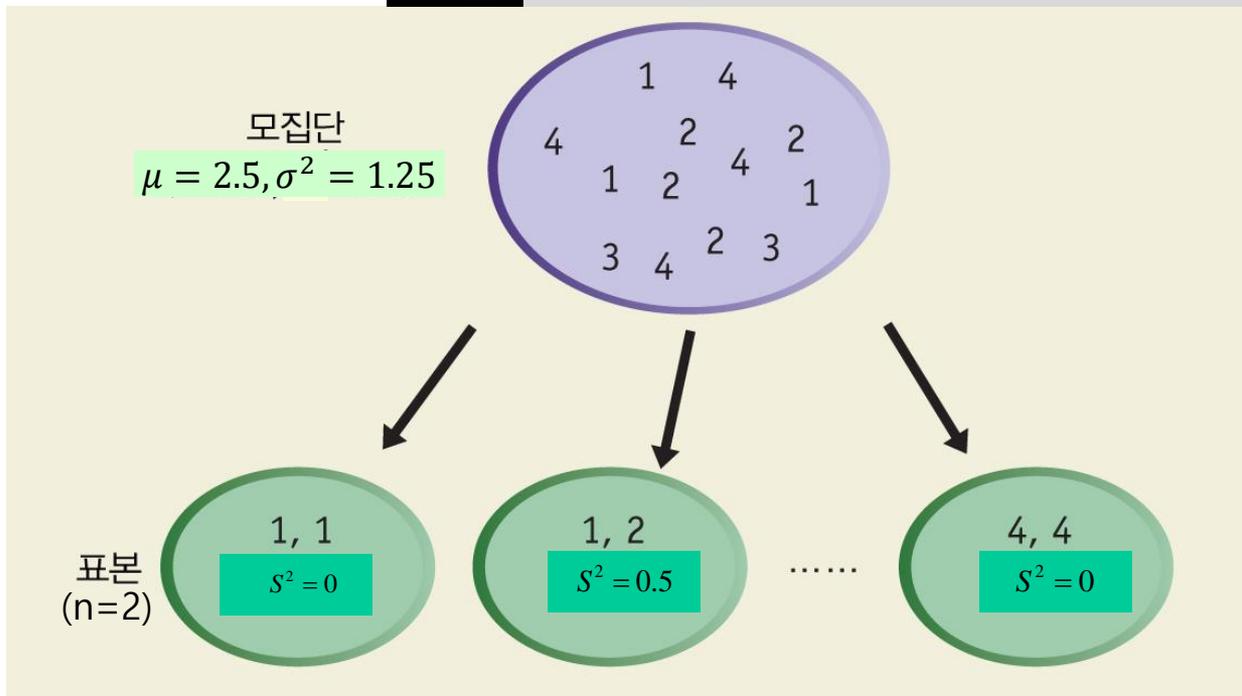
10-2 표본분포 3



10-2-1. 표본분산의 표본분포

그림

모집단으로부터 $n=2$ 인 임의표본 추출



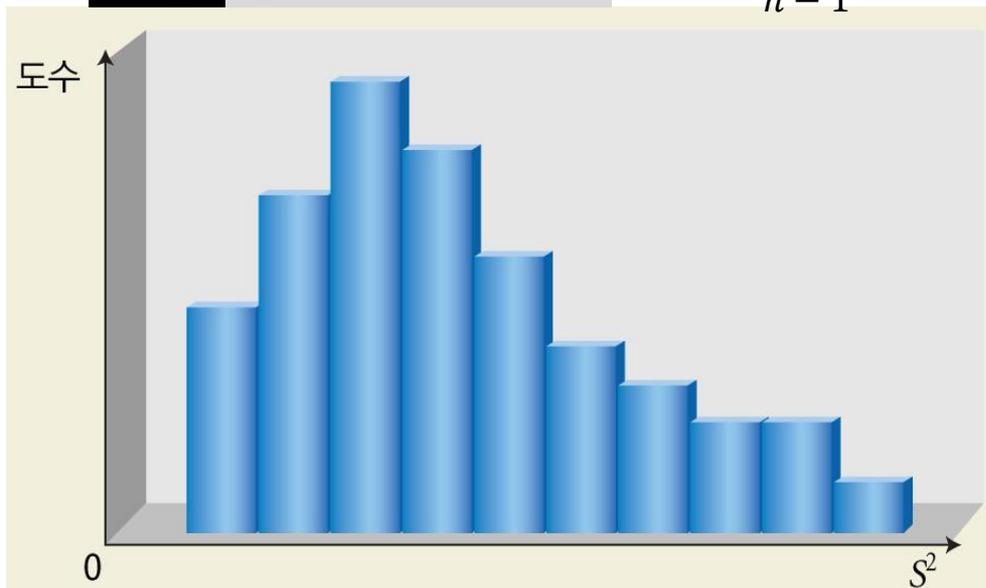
표본평균의 경우 중심극한정리에 의해 표본크기가 증가함에 따라 정규분포를 따르지만, **표본분산의 경우 정규분포를 따르지 않는다.**

표본	표본평균	표본분산	표본	표본평균	표본분산
(1,1)	1	0	(3,1)	2	2
(1,2)	1.5	0.5	(3,2)	2.5	0.5
(1,3)	2	2	(3,3)	3	0
(1,4)	2.5	4.5	(3,4)	3.5	0.5
(2,1)	1.5	0.5	(4,1)	2.5	4.5
(2,2)	2	0	(4,2)	3	2
(2,3)	2.5	0.5	(4,3)	3.5	0.5
(2,4)	3	2	(4,4)	4	0

표본분산의 표본분포

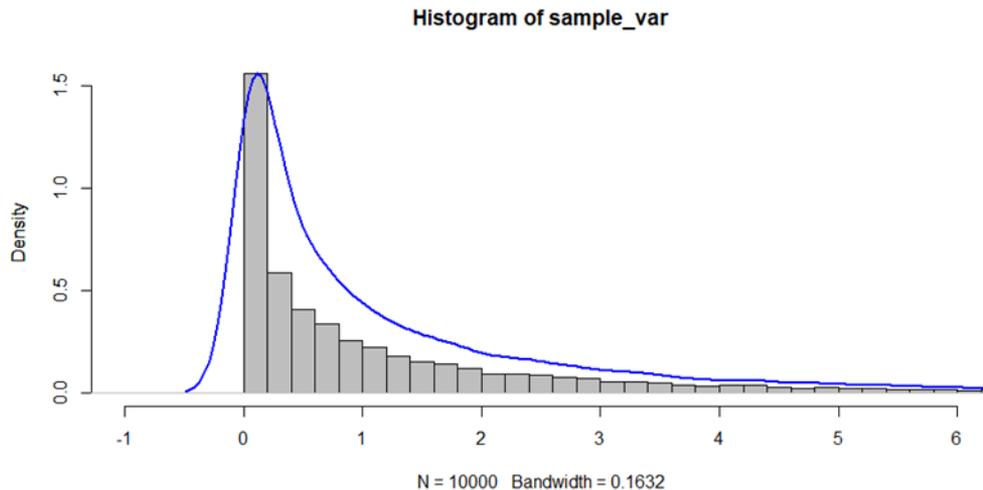
표본분산의 표본분포는 다음 히스토그램과 같은 분포를 이루게 된다.

그림 s^2 의 히스토그램 $s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$



실험

정규분포에 따르는 모집단($\mu=2.5, \sigma=1.118$) 에서 크기 $n=2$ 인 확률표본을 1,000개 추출
 - 표본분산의 분포는 정규분포를 따르지 않고 χ^2 -분포와 근사한 분포를 가진다



정규분포에 따르는 모집단($\mu=10, \sigma=2$)에서 표본크기 $n=11$ 인 확률표본 1,000개 추출.
 - 1,000개 표본 중 처음 5개의 표본이 다음과 같으면 각 표본의 표본분산 $s_1^2, s_2^2, \dots, s_5^2$ 를 구할 수 있듯이, 정규분포로부터 구한 표본분산 $s_1^2, s_2^2, \dots, s_{1000}^2$ 의 분포는 χ^2 -분포와 근사한 분포를 가진다.

• $E(s^2) = \sigma^2 = 4$

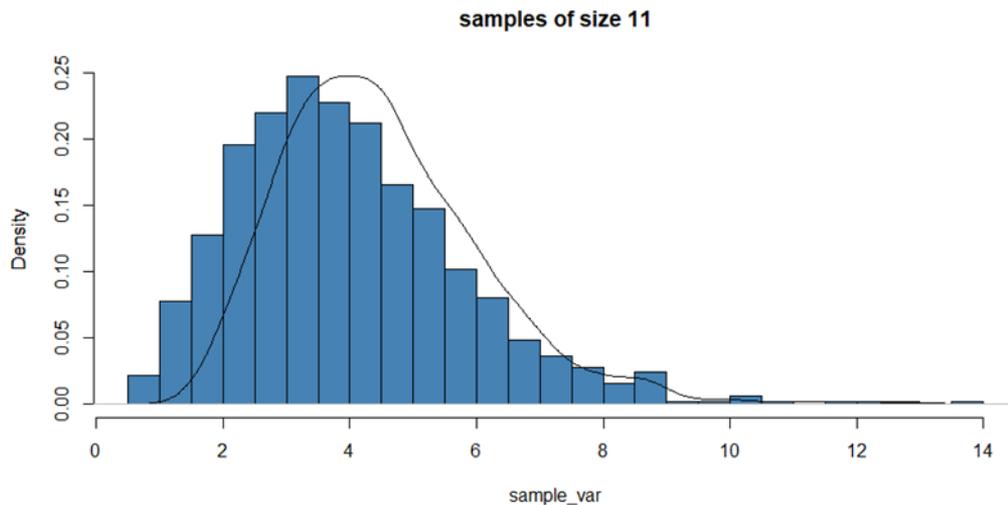
• $Var(s^2) = \frac{2\sigma^4}{n-1} = \frac{2 \times 16}{10} = 3.2$

-즉, 표본분산 s^2 은 근사적으로 $\chi^2(4, 3.2)$ 를 따른다.

표본	관측값											s^2
1	9.040	9.269	10.320	8.434	10.173	11.325	10.566	10.062	11.994	13.417	11.867	2.146
2	8.541	8.526	7.155	11.538	7.843	8.464	7.918	10.553	9.485	9.342	11.417	2.139
3	11.747	8.945	10.636	13.457	13.399	10.499	7.539	7.731	13.615	9.102	9.370	5.048
4	13.943	10.040	9.544	13.875	10.329	10.425	9.672	8.376	8.382	13.057	11.213	4.011
5	11.896	10.571	8.811	6.546	9.141	14.826	12.580	10.986	9.901	7.635	7.754	5.953



- 1,000개 표본의 표본분산의 평균과 분산은 각각 3.986259 및 3.31678로서 모집단의 이론적인 평균과 분산에 근접함을 알 수 있다

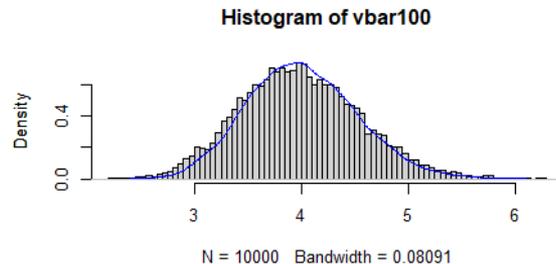
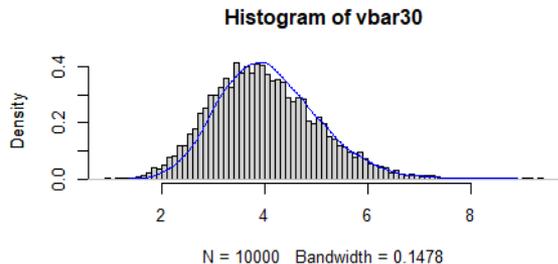
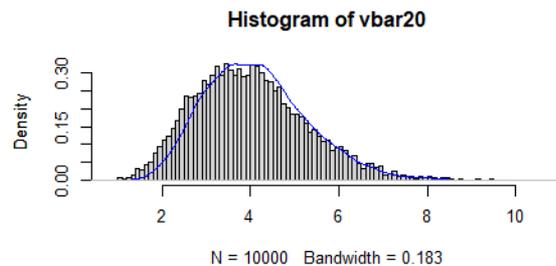
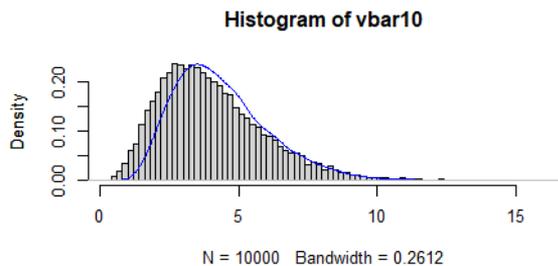


```
> (mean(sample_var))
[1] 3.986259
```

```
> (var(sample_var))
[1] 3.31678
```



즉, 정규분포로부터 구한 표본분산의 분포는 정규분포에 따르지 않고, 표본의 크기에 따라 모집단의 이론적인 평균인 σ^2 과 분산인 $\frac{2\sigma^4}{n-1}$ 에 근접한다.



표본분산의 표본분포 특성

분산 σ^2 을 갖는 모집단으로부터 크기 n 의 임의표본이 추출되었으며 이 표본의 분산이 S^2 이라고 하자.

$$1. E(S^2) = \sigma^2 \quad \leftarrow E\left[\frac{(n-1)s^2}{\sigma^2}\right] = \frac{(n-1)}{\sigma^2} E(s^2) = n-1$$

$$2. \sigma_{S^2}^2 = \frac{2\sigma^4}{n-1} \quad \text{또는} \quad \sigma_{S^2} = \sqrt{\frac{2\sigma^4}{n-1}} \quad \leftarrow \text{Var}\left[\frac{(n-1)s^2}{\sigma^2}\right] = \frac{(n-1)^2}{\sigma^4} \text{Var}(s^2) = 2(n-1)$$

3. 모집단이 정규분포하면 $(n-1)S^2/\sigma^2$ 은 $\chi^2_{(n-1)}$ -분포하게 된다.

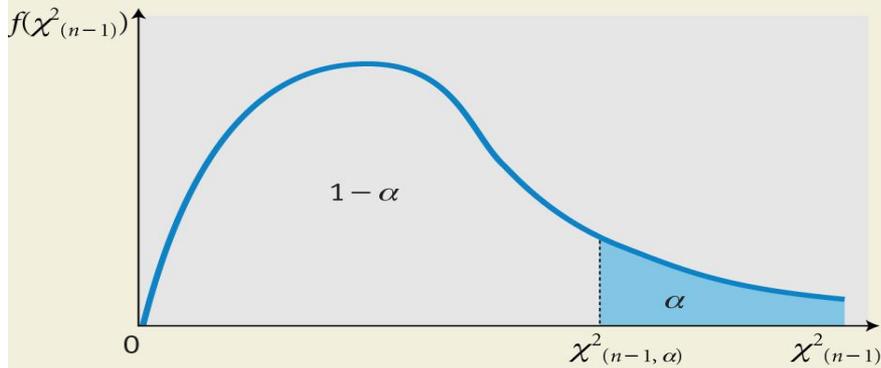


예 어떤 확률변수가 χ^2_5 -분포를 한다면 다음 등식을 만족시켜 주는 K는 무엇인가?

$$P(\chi^2_5 < K) = 0.9 \text{ 또는 } P(\chi^2_5 > K) = 0.1$$

풀이 χ^2 -분포표는 어떤 확률 α 가 주어졌을 때 $\alpha = P(\chi^2_{(n-1)} > \chi^2_{(n-1), \alpha})$ 를 만족하는 임계값 $\chi^2_{(n-1), \alpha}$ 를 나타낸다. 이를 그림으로 보면 다음과 같다.

그림 $\alpha = P(\chi^2_{(n-1)} > \chi^2_{(n-1), \alpha})$ 를 만족하는 임계값 $\chi^2_{(n-1), \alpha}$



따라서 확률변수 χ_5^2 가 K 보다 클 확률이 0.1이 되는 임계값은 $\chi_{5, 0.1}^2=9.24$ 이므로 χ^2 -분포표로부터 구하면 $K=9.24$ 이다.

자유도 <i>d.f.</i>	α									
	.995	.990	.975	.950	.900	.100	.050	.025	.010	.005
1	0.0 ⁴ 393	0.0 ³ 157	0.0 ³ 982	0.0 ² 393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59

10-2-2. 표본비율의 표본분포

모집단 비율 p 는 표본비율 \hat{p} 으로부터 추출될 수 있는데 이 표본비율은 성공의 횟수를 표본크기로 나누어 구할 수 있다.

이 때 성공의 횟수를 나타내는 확률변수 X 는 이항분포를 이루며 표본비율은 $\hat{p} = \frac{\text{성공의 횟수}}{\text{표본의 수}} = \frac{X}{n}$ 이다.



이항분포에 따르는 X 의 평균은 np , 분산은 $np(1-p)$ 이므로 \hat{p} 의 평균과 분산은 다음과 같다

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = p$$

$$\sigma_{\hat{p}}^2 = \text{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2} \text{Var}(X) = \frac{p(1-p)}{n}$$

