7강

확률분포와 표본분포2

통계·데이터과학과 이기재 교수

한국방송통신대학교

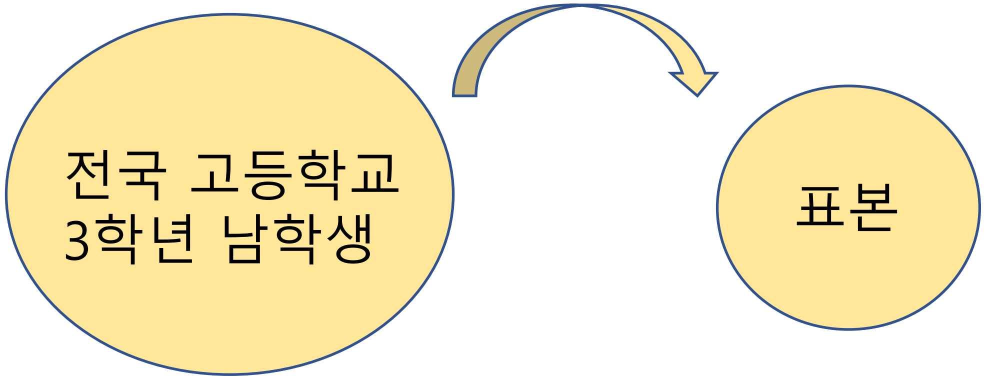
목차

01

표본분포

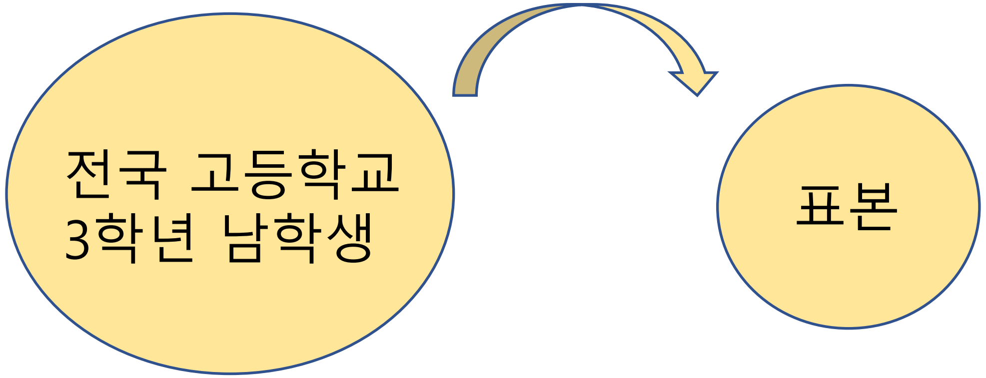
기본용어 1

- 통계적 추론(statistical inference)
  - 모집단에서 추출한 표본을 이용하여 모집단에 관한 추측이나 결론을 이끌어내는 과정
- 모수(parameter)
  - 모집단의 특성값(예 : 평균, 비율, 분산 등)



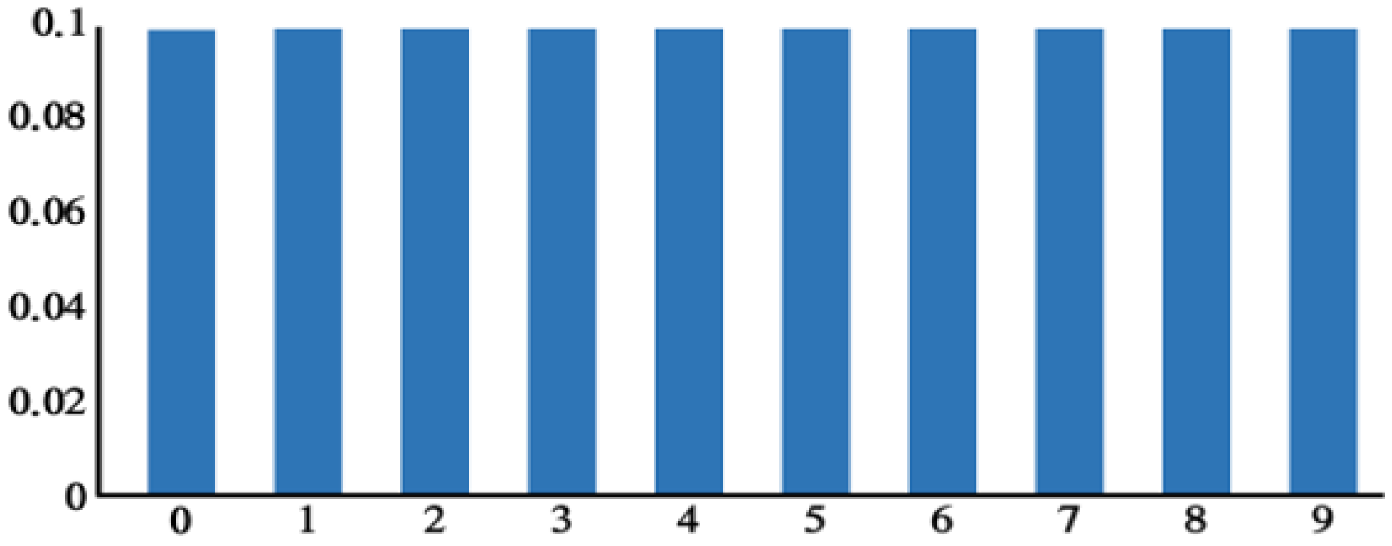
기본용어 2

- 랜덤표본(random sample)
  - 모집단에서 랜덤하게 추출된 일부로 서로 독립이며 동일한 분포를 따름
- 표본추출변동
  - 통계량 값이 표본에 따라 달라지는 것
- 표본분포(표집분포, sampling distribution)
  - 표본 통계량의 분포



표본평균의 표본추출변동 사례

- 0, 1, 2, ⋯, 9의 정수값이 될 확률이 각각 0.1인 이산형 균등분포에서 랜덤추출하는 경우



- $\mu = E(X) = 4.5,\ \ \sigma^2 = Var(X) = 8.25$

표본평균의 표본추출변동 사례

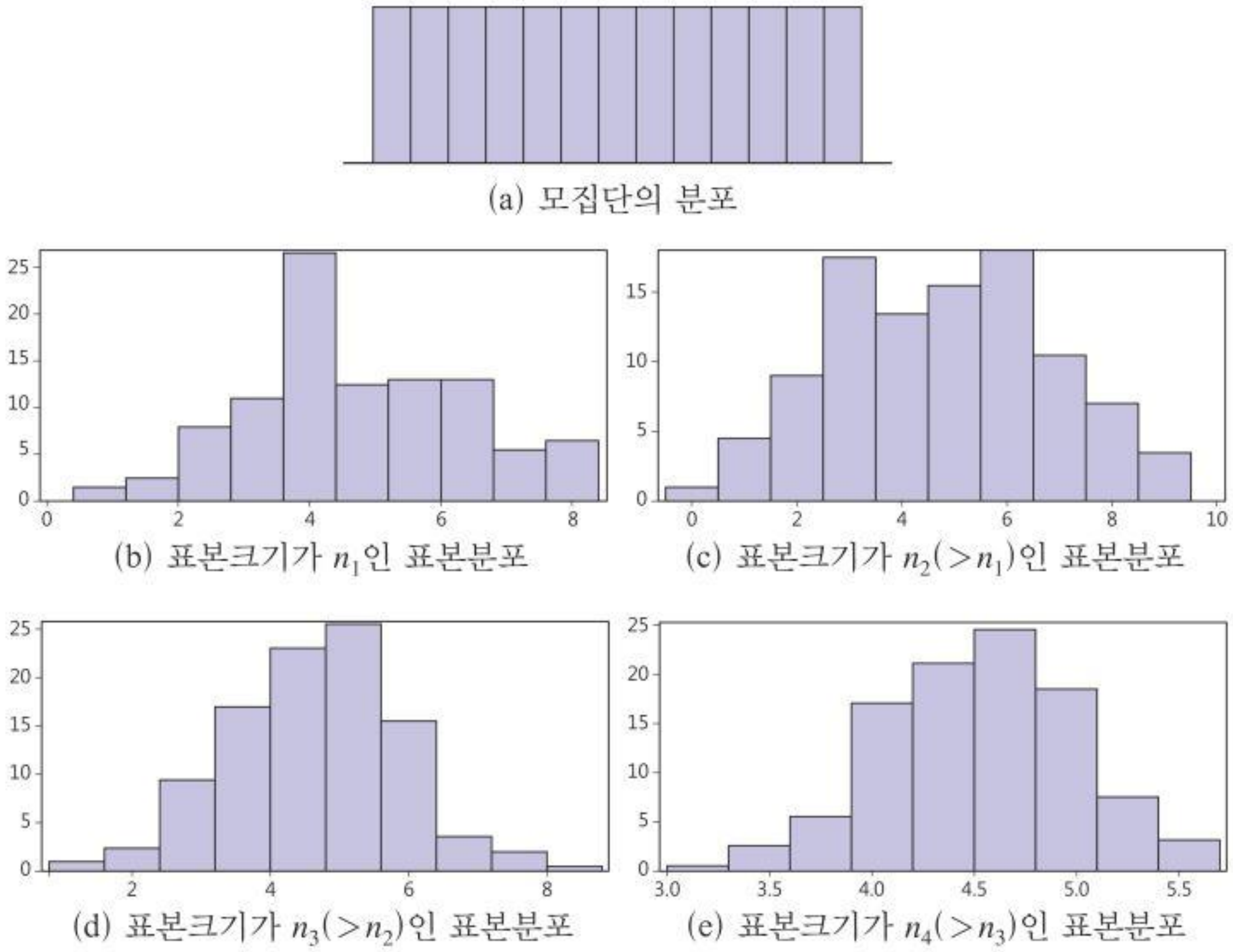


그림 4-9 표본평균의 분포: 모집단이 정규분포가 아닌 경우

표본평균의 기댓값과 분산

- 평균이 μ이고, 분산이 σ^2인 무한모집단에서 표본의 크기 n인 랜덤표본의 표본평균 $\bar{X}$에 대하여
  - $E(\bar{X}) = \mu$
  - $Var(\bar{X}) = \frac{\sigma^2}{n}$

표본평균의 분포(정규모집단의 경우)

- 모집단의 분포가 정규분포 $N(\mu, \sigma^2)$일 때 표본 크기 n인 랜덤표본의 표본평균 $\bar{X}$는 정규분포 $N(\mu, \frac{\sigma^2}{n})$을 따른다.

중심극한정리(Central Limit Theorem)

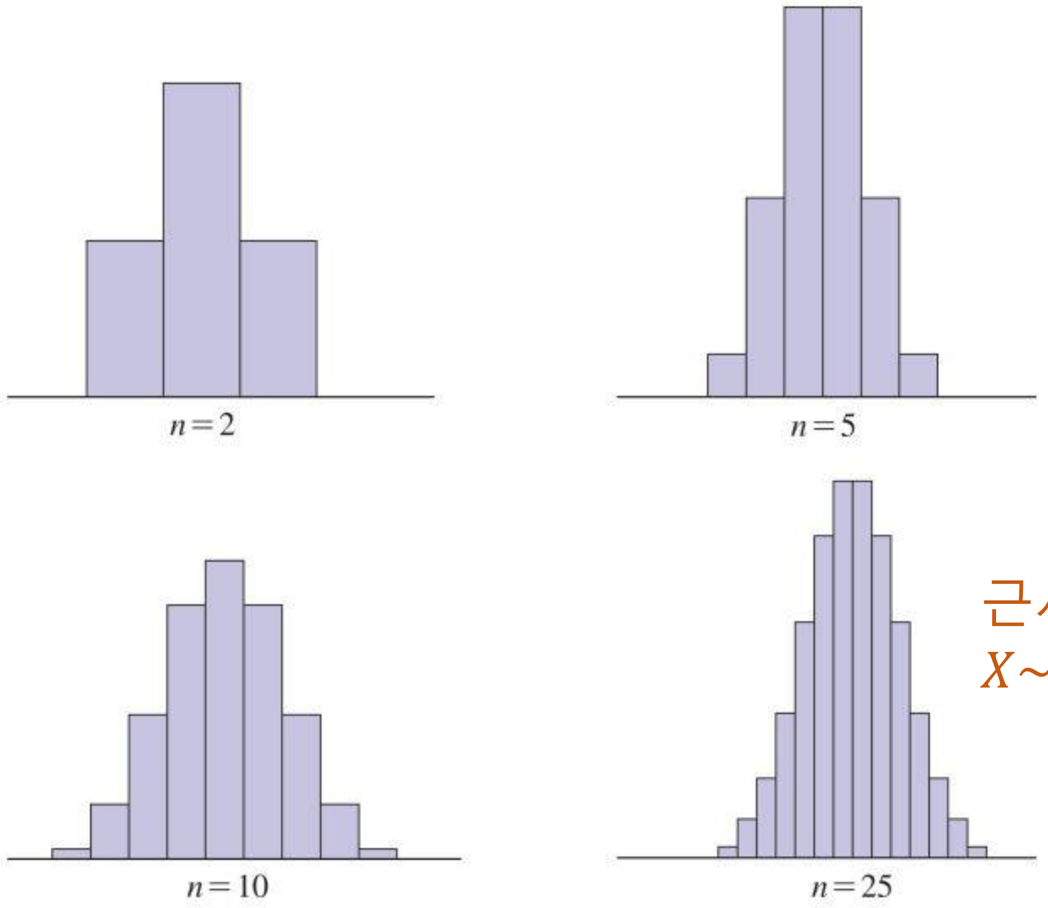
- 평균이 μ이고, 분산이 σ^2인 임의의 모집단에서 표본의 크기 n이 충분히 크면,
  - $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
  - $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

이항분포의 정규근사

- 이항분포 $B(n,p)$를 따르는 확률변수 X는 n이 클 때 근사적으로 정규분포 $N(np, np(1-np))$를 따른다.

$$\frac{X-np}{\sqrt{np(1-p)}} \sim N(0,\, 1)$$

이항분포의 정규근사 예시



근사적으로
$X \sim N(np, np(1-p))$을 따름

성공률 $p = \frac{1}{2}$일 때의 이항분포

이항분포의 정규근사 예제

- 한 공장에서 생산되는 제품의 불량률이 5%라고 함. 어느 날 제품 100개를 랜덤추출 하였을 때, 이 중에 불량품이 3개에서 7개일 확률은?

풀이

X = 100개 중에 포함된 불량품의 개수

$$X \sim B(100,\ 0.05)$$
$$E(X) = np = 100 \times 0.05 = 5$$
$$Var(X) = np(1-p) = 100 \times 0.05 \times 0.95 = 4.75$$
$$P(3 \le X \le 7) = P\left(\frac{3-5}{\sqrt{4.75}} \le Z \le \frac{7-5}{\sqrt{4.75}}\right)$$
$$= P(-0.918 \le Z \le 0.918) = 0.642$$

$t-$분포

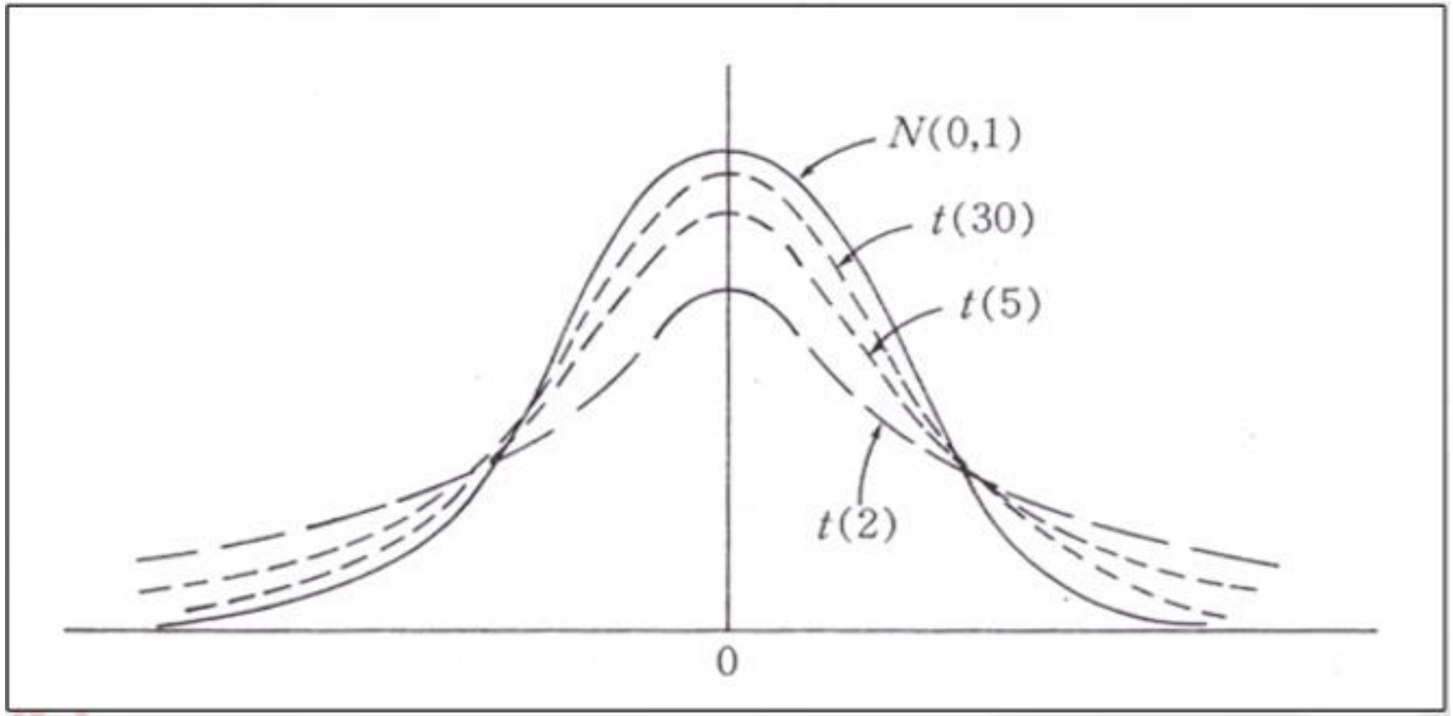
- $X_1, \ldots, X_n$이 $N(\mu, \sigma^2)$으로부터의 확률표본일 때,

  $$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)\,,\ \text{여기서 } S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

- 0에 대해서 좌우대칭이며, 자유도 n이 커지면 표준정규분포에 가까워짐
  - 통계학자 Gosset이 스튜던트(student)라는 필명으로 발표(1907년)

    ⇒ *스튜던트 t-분포(student's t-distribution)*

t –분포와 표준정규분포



<$N(0,1)$과 t-분포곡선>

모평균 μ의 $100(1-\alpha)\%$ 구간추정

- 모집단이 정규분포이고, σ^2을 알 수 있는 경우

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- 모집단이 정규분포이고, σ^2을 모르는 경우

$$\left[\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\ ,\ \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]$$

모분산 추정이 중요한 사례

- 어느 거리측정기 생산업체의 제품 정밀도 평가

  ⇒ 측정 거리의 편차가 크면 불량품으로 간주함

  ⇒ 측정 거리의 모분산이 중요함

- 어느 플라스틱판 생산 공장의 공정관리

  ⇒ 판 두께의 표준편차가 1.5mm보다 크면 공정 이상으로 판단

모분산, 모표준편차의 점추정

- 모분산(σ^2)의 점추정량
  - 표본분산 $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$

- 모표준편차(σ)의 점추정
  - 표본표준편차

$$S = \sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)}$$

표본분산의 분포

- 모분산 σ^2인 정규분포에서 뽑은 랜덤표본일 때

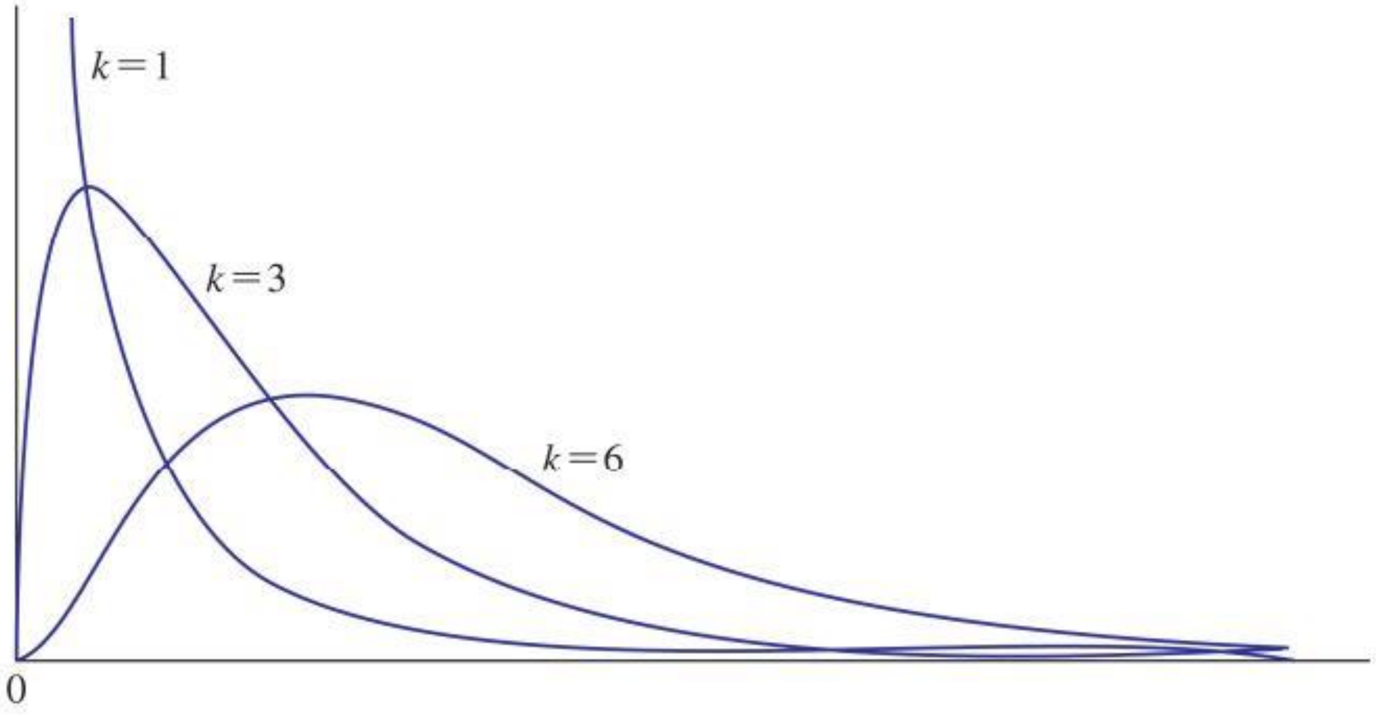
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

여기서, $S^2 = \sum(X_i - \overline{X})^2/(n-1)$

" 통계량 $(n-1)S^2/\sigma^2$은 여기서, 자유도 $n-1$인 카이제곱분표(χ^2 distribution)를 따른다."

카이제곱분포(χ^2 distribution)의 특징

- 자유도에 따라서 모양이 결정됨
- 비대칭분포



<여러 자유도에 대한 카이제곱분포>

F 분포

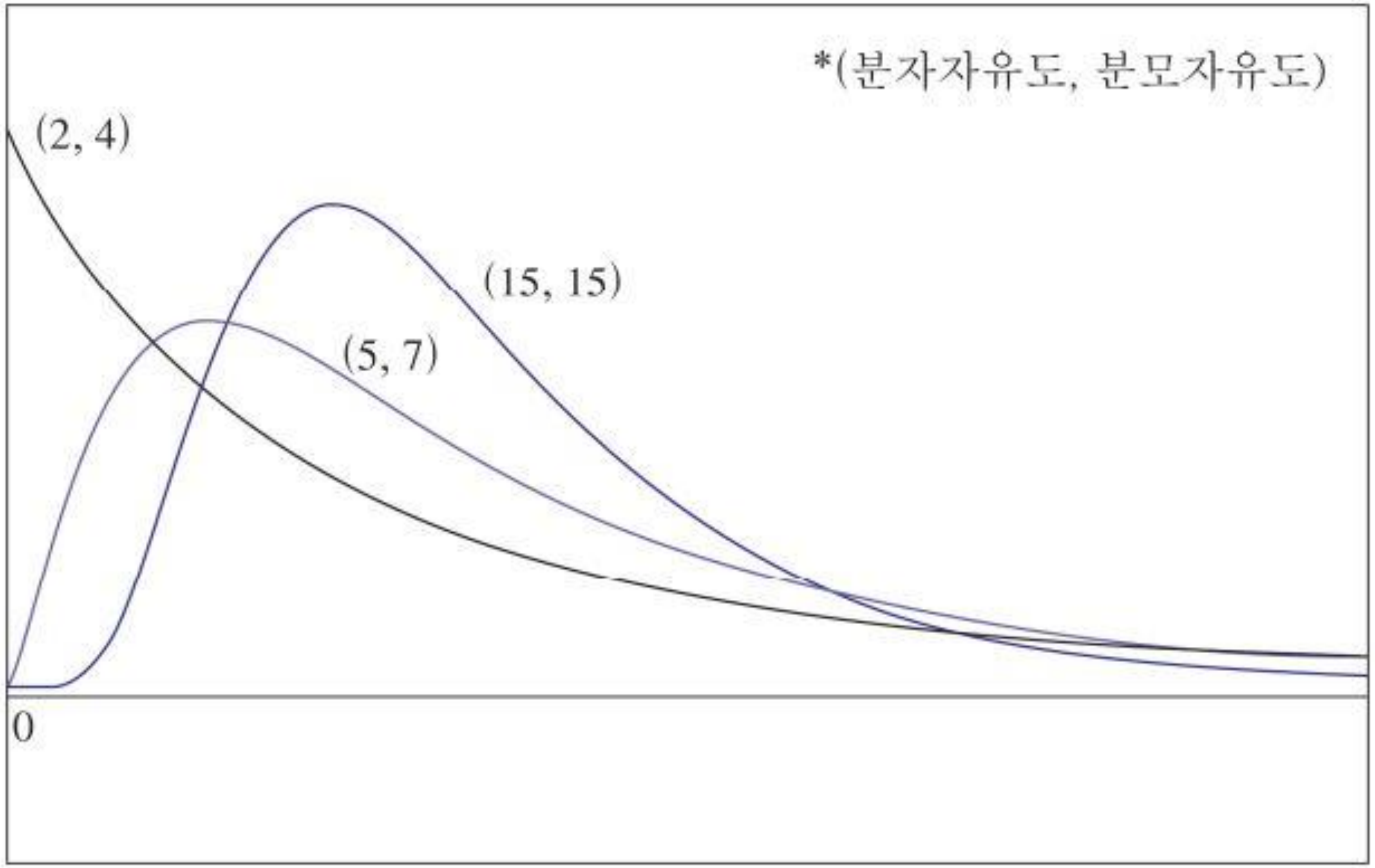
표본분산 비의 분포

두 모집단이 각각 모분산 σ_1^2, σ_2^2인 정규분포를 따를 때 이 두 모집단에서 독립적으로 추출한 크기 n_1, n_2인 표본에서 구한 표본분산을 각각 s_1^2, s_2^2이라고 하면,

$$F = \frac{\left(\dfrac{S_1^2}{\sigma_1^2}\right)}{\left(\dfrac{S_2^2}{\sigma_2^2}\right)} \sim F(n_1 - 1,\ n_2 - 1)$$

" 분자자유도 $n_1 - 1$, 분모자유도 $n_2 - 1$ 인 F를 따른다."

F 분포 그림



❙그림 4-14❙ 여러 가지 자유도에 따른 F분포의 그림

다음시간 안내

8강

추정