

# 표본 분포

# 들어가기

- 통계이론의 주된 활용은 연구대상에 대한 실험이나 관찰 등을 통해 구한 자료로부터 연구대상의 성격을 규명하는 것이다. 연구대상의 총체(totality)를 모두 조사하는 일은 많은 경우 시간 또는 경비 등의 제약으로 인해 불가능하다. 따라서 통계적 방법들은 그 총체의 일부인 표본을 조사함으로써 총체에 대한 성격을 귀납적(inductive)으로 추론하는 길을 택한다.
- 통계적 추론방법들은 불확실성을 계량화하기 위하여 확률을 사용한다.

# 모집단과 표본

- 연구 또는 관찰의 대상이 되는 총체(totality)를 모집단이라 한다.
- 랜덤표본(임의표본)의 수학적 정의는?
  - 확률변수  $X_1, \dots, X_n$ 의 결합확률밀도함수가  $f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f(x_1) \dots f(x_n)$  (단,  $f(\cdot)$ 는 각  $X_i$ 의 공통 확률밀도함수)의 꼴로 나타나면,  $X_1, \dots, X_n$ 을 모확률밀도함수가  $f(\cdot)$ 인 크기가  $n$ 인 랜덤표본 또는 임의표본이라고 한다.  $X_i$ 는  $i$ 번째 관찰변수가 가질 값을 뜻한다. 일단 관찰된 알려진 값은  $x_1, \dots, x_n$ 으로 소문자를 사용하여 표기한다.
  - 랜덤표본을 표현할 때 "확률변수  $X_1, \dots, X_n$ 이 iid(independent and identically distributed)이다" 라고도 한다.

# 모집단과 표본 예제

- 모분포가 "성공"확률이  $p$ 인 베르누이 확률밀도함수는  $f(x) = p^x q^{1-x}$ ,  $x = 0$  또는  $1$ , ( $p + q = 1$ )이다. 이 경우 서로 독립인 베르누이 시행을 10회 반복하였을 때의 표본을  $X_1, X_2, \dots, X_{10}$ 으로 표기하면, 이 표본의 결합 확률밀도함수는?
  - $f_{X_1, X_2, \dots, X_{10}}(x_1, x_2, \dots, x_{10}) = f(x_1)f(x_2) \dots f(x_{10}) = p^{\sum_{i=1}^{10} x_i} q^{10 - \sum_{i=1}^{10} x_i}$  ( $x_i = 0$  또는  $1$ ;  $i = 1, \dots, 10$ )
  - $X_1, \dots, X_{10}$ 을 모확률밀도함수가  $f(x) = p^x q^{1-x}$  ( $x = 0$  또는  $1$ )인 크기 10의 랜덤표본이라고 한다.

# 통계량

- 미지의 모수를 포함하지 않는 랜덤포본의 함수  $T = T(X_1, \dots, X_n)$ 를 통계량(statistic)이라고 한다. 통계량 자체도 확률변수이다.
  - $X_1, \dots, X_n$  이 확률밀도함수  $f(x; \theta)$  ( $\theta$ 는 알려지지 않은 모집단의 모수임)로부터 얻은 랜덤포본이라고 하자. 이 때, 표본평균  $\bar{X}_n = \sum_{i=1}^n \frac{X_i}{n}$  또는 표본최댓값  $\max\{X_1, \dots, X_n\}$ 은 통계량이다. 그러나  $\bar{X}_n - \theta$  또는  $\max\{\frac{X_1}{\theta}, \dots, \frac{X_n}{\theta}\}$  등은 미지의 모수  $\theta$ 에 의존하므로 통계량이 아니다.

# 표본평균과 표본분산

- 평균과 분산이 각각  $\mu$ 와  $\sigma^2$ 인 확률밀도함수  $f(x)$ 로부터 랜덤표본  $X_1, \dots, X_n$ 을 얻었다. 이 때 표본평균은  $\bar{X}_n = \sum_{i=1}^n \frac{X_i}{n}$ , 표본분산은  $S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{(n-1)}$ 이다.
- 대수의 법칙
  - 평균이  $\mu < \infty$ 인 확률밀도함수  $f(x)$ 로부터 랜덤표본  $X_1, \dots, X_n$ 을 얻었다면, 표본의 크기가 커짐에 따라 표본평균이 모평균으로 **확률적으로 수렴**한다.
  - 증명은 체비셰프의 부등식을 사용한다.

# 중심극한정리

- 중심극한정리(central limit theorem)는 확률론과 통계학에 있어서 가장 중요하다고 여겨지는 이론이다.
- 평균과 분산이 각각  $\mu$ 와  $\sigma^2 < \infty$ 인 확률밀도함수  $f(x)$ 로부터의 랜덤표본  $X_1, \dots, X_n$ 
  - 확률변수  $Z_n = \frac{\sum_{i=1}^n X_i - E(\sum_{i=1}^n X_i)}{\sqrt{\text{Var}(\sum_{i=1}^n X_i)}} = \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}}$ 의 분포는 표본의 크기  $n$ 이 무한대에 접근함에 따라 표준정규분포  $N(0,1)$ 로 수렴한다.
- 모분포의 형태와 관계없이 유한한 평균과 분산만 존재하면 확률변수  $Z_n$ 의 분포는 표준정규분포로 수렴한다.

# 중심극한정리 예제 1

균등분포와 정규분포

- ✓  $X_1, \dots, X_n$ 이  $U(0,1)$ 로부터 얻은 랜덤표본이라고 하자. 그러면  $E(X_i) = \frac{1}{2}$ ,  $Var(X_i) = \frac{1}{12}$ 이다. 중심극한정리를 사용하여,  $\sum_{i=1}^n X_i$ 의 분포를  $N(\frac{n}{2}, \frac{n}{12})$ 로 근사할 수 있다.
- ✓ 즉, 중심극한정리에 의해서  $\sum_{i=1}^n X_i$ 를 표준화한  $\frac{\sum_{i=1}^n X_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$ 는 근사적으로 표준정규분포를 따른다.

$$\begin{aligned} \checkmark P[a \leq \sum_{i=1}^n X_i \leq b] &= P\left[\frac{a - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \leq \frac{\sum_{i=1}^n X_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \leq \frac{b - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right] \\ &\approx P\left[\frac{a - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \leq Z \leq \frac{b - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right] = \Phi\left[\frac{b - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right] - \Phi\left[\frac{a - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right] \end{aligned}$$



# 중심극한정리 예제 2

## 이항분포와 정규분포

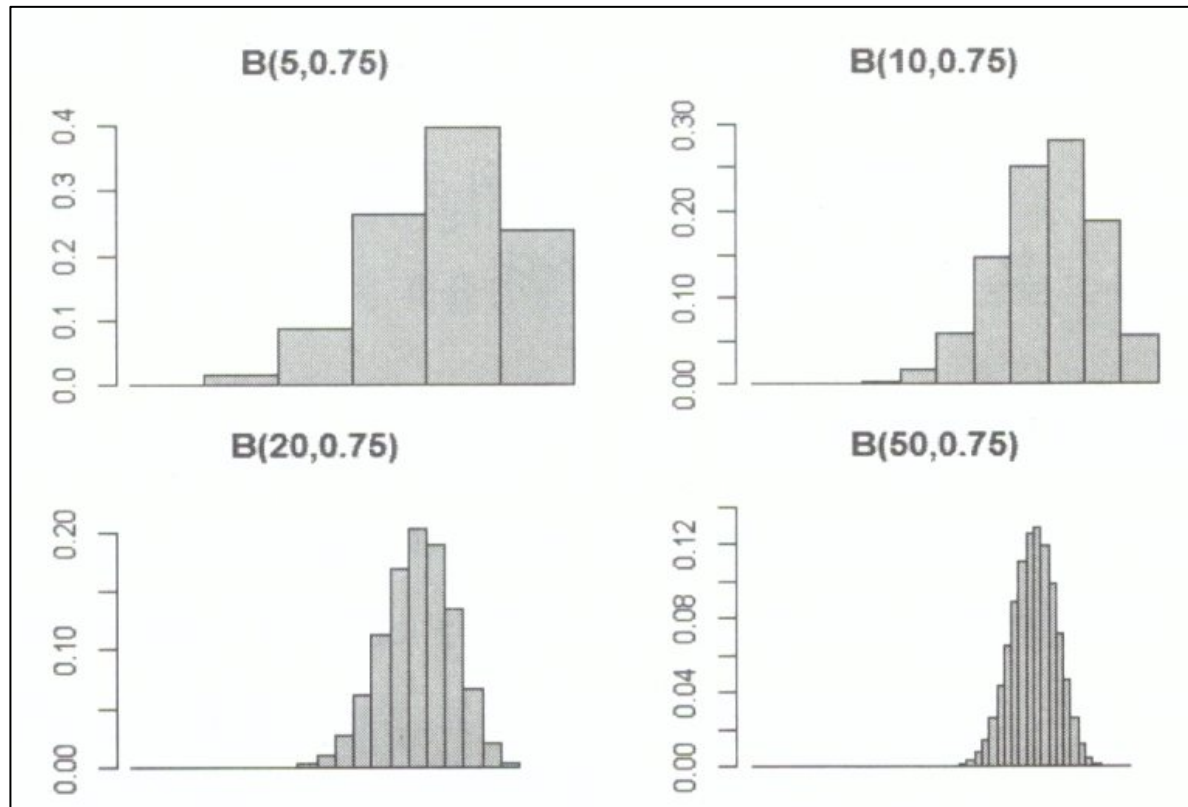
- ✓  $X \sim B(n, p)$ 라고 하면  $X$ 는 서로 독립이며 "성공"확률이  $p$ 인  $n$ 개의 베르누이 확률변수  $X_1, X_2, \dots, X_n$ 들의 합  $\sum_{i=1}^n X_i$ 와 같은 분포를 갖는다.
- ✓  $E(X_i) = p$  ,  $Var(X_i) = pq$  ( $q = 1 - p$ )
- ✓ 중심극한정리에 의해서  $n$ 이 클 때  $\frac{(\sum_{i=1}^n X_i - np)}{\sqrt{npq}}$ 는 근사적으로 표준정규분포를 따른다.
- ✓ 
$$P[a \leq X \leq b] = P[a \leq \sum_{i=1}^n X_i \leq b] = P\left[\frac{a - np}{\sqrt{npq}} \leq \frac{\sum_{i=1}^n X_i - np}{\sqrt{npq}} \leq \frac{b - np}{\sqrt{npq}}\right]$$
$$\approx P\left(\frac{a - np}{\sqrt{npq}} \leq Z \leq \frac{b - np}{\sqrt{npq}}\right)$$
$$= \Phi\left[\frac{b - np}{\sqrt{npq}}\right] - \Phi\left[\frac{a - np}{\sqrt{npq}}\right]$$

## 중심극한정리 예제 2

- ✓ 이항확률변수는 이산형인 반면 정규확률변수는 연속형이므로 이를 반영한 연속성 수정 (continuous correction)을 하여 근사의 정확성을 더 높일 수 있다.
- ✓ 연속성 수정
  - 이항분포는 이산형으로서, 정수값만을 취할 수 있으므로 어떤 정수  $k$ 에 대해  $P(B(n, p) \leq k) = P(B(n, p) < k + 1)$ 이지만, 연속형인 정규분포는 같은 성질을 만족하지 않는다.
  - 정규분포로 근사할 때,  $k$ 나  $k + 1$ 을 쓰지 않고,  $k + \frac{1}{2}$ 을 쓰게 된다.
  - $B(a \leq B(n, p) \leq b)$ 를 근사할 때,  $B\left(a - \frac{1}{2} \leq B(n, p) \leq b + \frac{1}{2}\right)$ 를 이용하여  $\Phi\left[\frac{b-np}{\sqrt{npq}}\right] - \Phi\left[\frac{a-np}{\sqrt{npq}}\right]$ 로 근사한다.

# 중심극한정리 예제 2

✓  $n$ 의 크기에 따른 이항확률변수의 확률밀도함수



# 중심극한정리 예제 3

- 어느 대학교에서 남학생들의 몸무게는 평균 69kg, 표준편차 5kg이라 한다. 100명의 남학생을 랜덤으로 추출할 때, 표본평균  $\bar{X}$ 가 평균  $\mu$ 를 중심으로  $\pm 1\text{kg}$  이내에 있을 확률을 구하여라.

| Z   | 0.00     | 0.01     | 0.02     |
|-----|----------|----------|----------|
| 0.0 | 0.500000 | 0.503989 | 0.507978 |
| 0.1 | 0.539828 | 0.543795 | 0.547758 |
| 0.2 | 0.579260 | 0.583166 | 0.587064 |
| 0.3 | 0.617911 | 0.621720 | 0.625516 |
| 0.4 | 0.655422 | 0.659097 | 0.662757 |
| 0.5 | 0.691462 | 0.694974 | 0.698468 |
| 0.6 | 0.725747 | 0.729069 | 0.732371 |
| 0.7 | 0.758036 | 0.761148 | 0.764238 |
| 0.8 | 0.788145 | 0.791030 | 0.793892 |
| 0.9 | 0.815940 | 0.818589 | 0.821214 |
| 1.0 | 0.841345 | 0.843752 | 0.846136 |
| 1.1 | 0.864334 | 0.866500 | 0.868643 |
| 1.2 | 0.884930 | 0.886861 | 0.888768 |
| 1.3 | 0.903200 | 0.904902 | 0.906582 |
| 1.4 | 0.919243 | 0.920730 | 0.922196 |
| 1.5 | 0.933193 | 0.934478 | 0.935745 |
| 1.6 | 0.945201 | 0.946301 | 0.947384 |
| 1.7 | 0.955435 | 0.956367 | 0.957284 |
| 1.8 | 0.964070 | 0.964852 | 0.965620 |
| 1.9 | 0.971283 | 0.971933 | 0.972571 |
| 2.0 | 0.977250 | 0.977784 | 0.978308 |
| 2.1 | 0.982136 | 0.982571 | 0.982997 |
| 2.2 | 0.986097 | 0.986447 | 0.986791 |
| 2.3 | 0.989276 | 0.989556 | 0.989830 |

표준정규분포표

# 카이제곱분포

- 모수가  $(k, \theta)$ 인 감마분포에서  $k = \frac{n}{2}$ ,  $\theta = 2$ 인 확률변수  $X$ , 즉,  $X \sim GAM(\frac{n}{2}, 2)$ 는 자유도 (degrees of freedom)가  $n$ 인 카이제곱분포(chi-squared distribution)을 따른다고 한다.
- 카이제곱분포는 모분산  $\sigma^2$ 의 추정량인 **표본분산  $s^2$ 의 분포**를 이해하는 데 중요한 역할을 한다.
- 자유도가  $n$ 인 카이제곱분포를 가지는 변수  $X$ 의 확률밀도함수는  $f_X = \frac{1}{\Gamma(\frac{n}{2})\left(\frac{1}{2}\right)^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$ ,  $x > 0$  이다.
  - $X \sim \chi^2(n)$  일 때,  $X$ 의 평균은  $E(X) = n$ ,  $Var(X) = 2n$  이 된다.
  - 서로 독립인 확률변수  $X_i (i = 1, \dots, n)$ 들이 각각 자유도가  $k_i$ 인 카이제곱분포를 따르면 그들의 합인  $Y = \sum_{i=1}^n X_i$ 는 자유도가  $\sum_{i=1}^n k_i$ 인 카이제곱분포를 따른다.

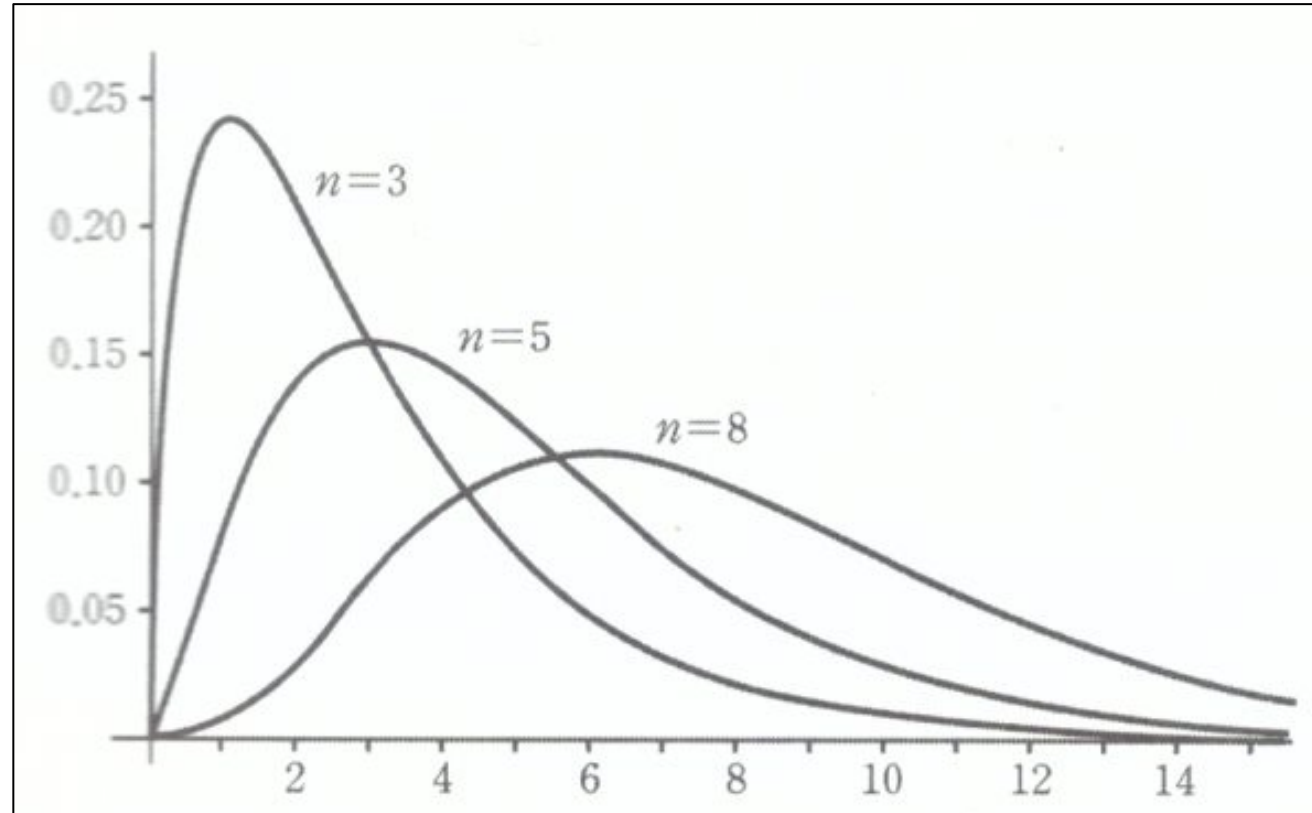
# 자유도

- **자유도**(degree of freedom)란 통계량을 구성하는 변수들 중에서 자유롭게 선택할 수 있는 변수의 개수
- 만약 5개의 값을 모두 더해서 10을 얻었다고 하자. 5개의 값 중 4개의 값이 2, 3, 1, 1임을 알고 있다면 나머지 한 개의 값은 3이 될 것이라는 것을 쉽게 알 수 있다.
- 즉, 5개 값들의 총합을 알고 있을 경우 4개의 값만 알면 나머지 한 개의 값은 자동으로 결정되며, 이와 같이 임의로 결정될 수 있는 값의 개수 4가 바로 자유도인 것이다.

# 자유도

- 통계량  $\sum_{i=1}^n (X_i - \bar{X})^2$  :
  - $n$ 개의  $X_i$  값에 의해서 결정되지만, 여기서  $X_i$ 와  $\bar{X}$ 간에는  $\sum_{i=1}^n (X_i - \bar{X}) = 0$ 이라는 관계가 성립
  - 이러한 성질을 고려할 때  $n$ 개의  $X_i$  중에 자유롭게 택할 수 있는 변수의 개수는  $n - 1$
  - 즉,  $n$ 개 중에서  $n - 1$ 개의  $X_i$  값만 알면 나머지 한개의 값은 자동적으로 결정
  - 이 때의 값  $n - 1$ 이 바로 통계량  $\sum_{i=1}^n (X_i - \bar{X})^2$ 의 자유도
- $\sum_{i=1}^n (X_i - \mu)^2$  :
  - $X_i$ 와  $\mu$ 간에 어떠한 제약 조건이 존재하지 않음
  - 이때의 자유도는 변수  $X_i$ 의 전체 개수인  $n$ 이 됨

# 카이제곱분포의 pdf



자유도가  $n$ 인 카이제곱분포의 확률밀도함수



# 카이제곱분포

- 확률변수  $Z$ 가  $N(0,1)$  분포를 따를 때,  $Y = Z^2$ 은  $\chi^2(1)$  분포를 가진다.
- 서로 독립인 확률변수  $X_i (i = 1, \dots, k)$  들이 각각 정규분포  $N(\mu_i, \sigma_i^2)$ 을 따른다고 하면,
  - $V = \sum_{i=1}^k \left( \frac{X_i - \mu_i}{\sigma_i} \right)^2$ 은 자유도가  $k$ 인 카이제곱분포를 따른다.
- $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ 이고, 크기가  $n$ 인 랜덤표본이라고 하면,
  - $\bar{X}_n$ 와  $S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{(n-1)}$ 은 서로 독립이며,
  - $\frac{(n-1)S^2}{\sigma^2}$ 은 자유도가  $n - 1$ 인 카이제곱분포를 따른다.

# 카이제곱분포 예제 1

- 어떤 종류의 제품을 1개 생산하는 데 걸리는 시간을  $X$ 라 하자.  $X \sim N(6, 2^2)$ 라 하자. 10개의 랜덤표본의 표본분산이 5보다 클 확률은?

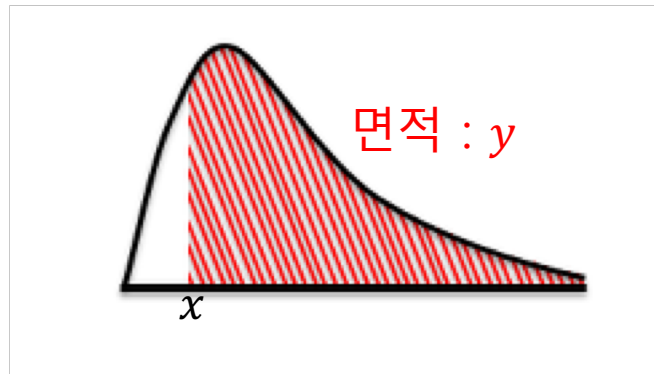
$$\checkmark \frac{9S^2}{4} \sim \chi^2(9)$$

$$\checkmark P(S^2 > 5) = P\left[\frac{9S^2}{4} > \frac{45}{4}\right] = P[\chi^2(9) > 11.25] \approx 0.259$$

## 카이제곱분포 예제 2

- 어떤 양어장에서 1년 동안 양식한 송어의 길이는 분산  $\sigma^2 = (5\text{cm})^2$ 인 정규분포를 따른다고 한다. 이 양식장에서 송어 25마리를 랜덤으로 추출할 때, 표본분산이 16.3cm와 24.3cm 사이일 확률을 구하여라.
- 자유도가 24일 때,  $\chi^2(x) = y$ 의 관계표 (카이제곱분포표 발췌)

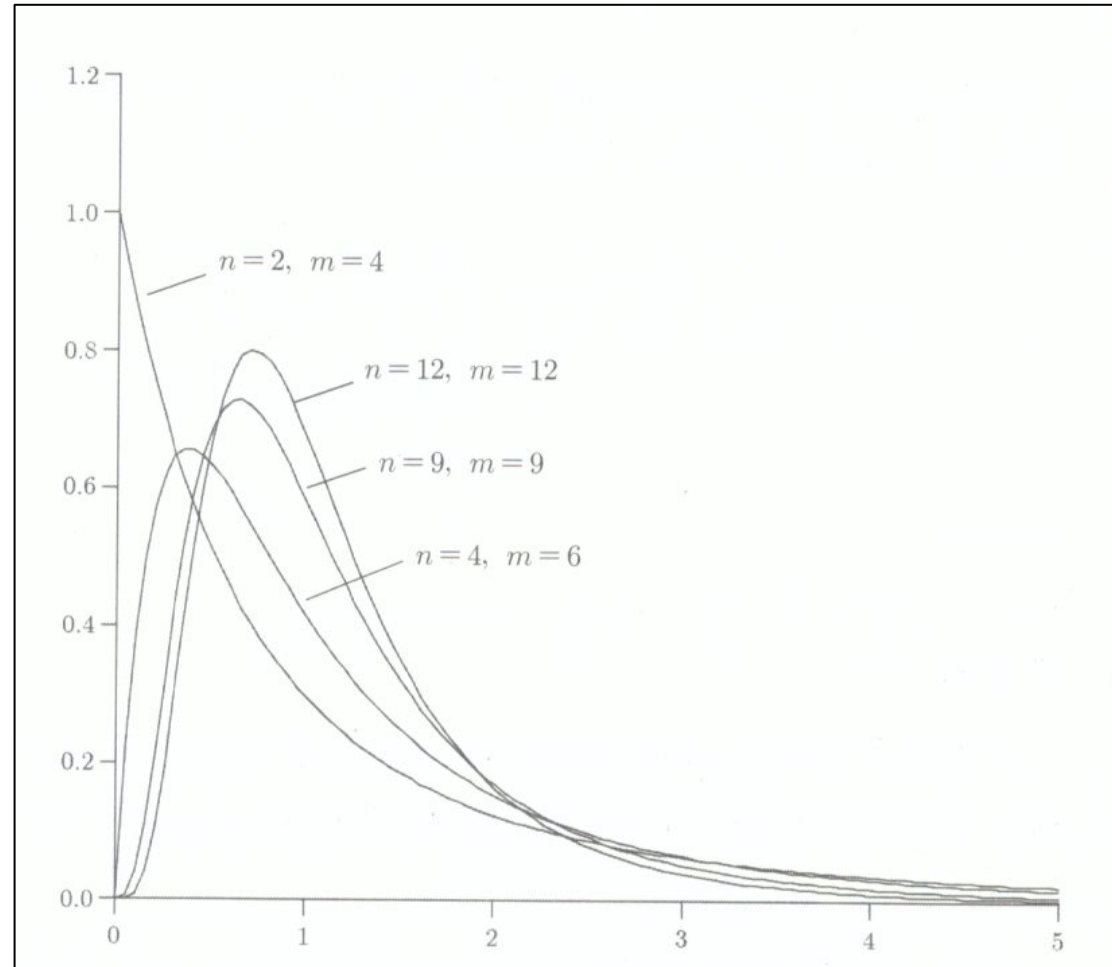
| $x$   | $y$   |
|-------|-------|
| 9.89  | 0.995 |
| 10.86 | 0.99  |
| 12.40 | 0.975 |
| 13.85 | 0.95  |
| 15.66 | 0.9   |
| 23.34 | 0.5   |
| 33.20 | 0.1   |
| 36.42 | 0.05  |



# F분포

- F분포는 정규분포로부터 구한 독립인 **두 표본의 분산비**에 대한 분포를 설명하는 데 사용되며, 분산 분석(ANOVA)에 활용한다.
- 서로 독립인 카이제곱확률변수  $U$ 와  $V$ 의 자유도가 각각  $n, m$ 이라고 할 때,
  - 변수  $X = \frac{U}{n} / \frac{V}{m}$ 은 자유도가  $(n, m)$ 인 F분포를 따른다.
  - $X$ 의 확률밀도함수  $f_X(x) = \frac{\Gamma\left[\frac{n+m}{2}\right]}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} \left(\frac{x^{\frac{n-2}{2}}}{\left[1+\frac{nx}{m}\right]^{\frac{n+m}{2}}}\right), x > 0$  이다.  $X \sim F(n, m)$ 으로 표기한다.
- $X \sim F(n, m)$ 일 때,
  - $X$ 의 평균은  $E(X) = \frac{m}{m-2} (m > 2)$
  - $X$ 의 분산은  $Var(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)} (m > 4)$  이다.

# $F$ 분포의 pdf



$F(n, m)$ 인  $F$ 분포의 확률밀도함수

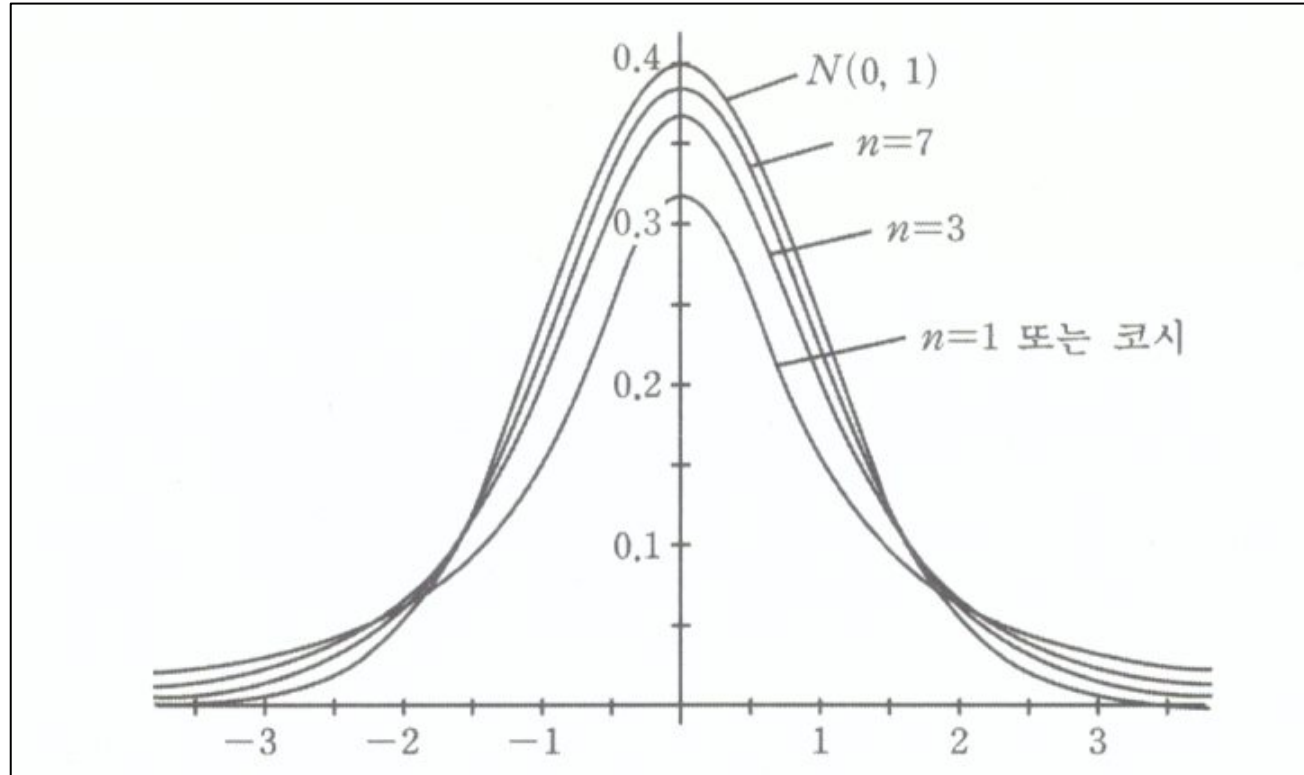
# F분포

- $\frac{1}{F_{\alpha}(n,m)} = F_{1-\alpha}(m,n)$
- $X_1, \dots, X_n$ 은 평균이  $\mu_X$ 이고 분산이  $\sigma_X^2$ 인 정규분포로부터 얻은 크기가  $n$ 인 랜덤표본이고,  $Y_1, \dots, Y_m$ 은 평균이  $\mu_Y$ 이고 분산이  $\sigma_Y^2$ 인 정규분포로부터 얻은 크기가  $m$ 인 랜덤표본이다.
  - 두 표본이 서로 독립이면 확률변수  $F = \frac{S_X^2}{\sigma_X^2} / \frac{S_Y^2}{\sigma_Y^2}$ 은 자유도가  $(n-1, m-1)$ 인  $F$ 분포를 따른다.
  - $\sigma_X^2 = \sigma_Y^2$ 인 경우, 표본분산비  $\frac{S_X^2}{S_Y^2} = \frac{\sum_{i=1}^n \frac{(Y_i - \bar{Y}_n)^2}{n-1}}{\sum_{i=1}^m \frac{(Y_i - \bar{Y}_m)^2}{m-1}}$ 는 자유도가  $(n-1, m-1)$ 인  $F$ 분포를 따른다.

# t분포

- $t$ 분포(Student's  $t$  distribution)는 정규분포의 모평균에 대한 가설검정 등을 포함한 검정론에서 매우 중요한 역할을 한다.
- 자연에서 발생하는 현상 데이터는 실제로 정규분포보다 양 끝단의 비중이 더 크다. 이를 fat tail 현상이라고 하는데, 이러한 데이터 모형에 적합한 것이 스튜던트  $t$ 분포이다.
- 확률변수  $Z$ 가 표준정규분포를 따르며,  $U$ 는 자유도  $k$ 인 카이제곱분포를 따르고,  $Z$ 와  $U$ 가 서로 독립일 때, 확률변수  $X = \frac{Z}{\sqrt{U/k}}$ 는  $t$ 분포를 따른다.
- $t$ 분포를 따르는 확률밀도함수 :
  - $f_X(x) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{\sqrt{k\pi}} \left( \frac{1}{1 + \frac{x^2}{k}} \right)^{\frac{k+1}{2}}, (-\infty < x < \infty)$

# $t$ 분포 pdf



$T(n)$ 인  $t$ 분포의 확률밀도함수



# $t$ 분포

- 평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 정규분포로부터의 랜덤포본  $X_1, \dots, X_n$ 을 고려
  - 확률변수  $T = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}}} = \frac{\bar{X}_n - \mu}{\frac{S}{\sqrt{n}}}$ 은 자유도가  $(n-1)$ 인  $t$ 분포를 따른다.
  - $\mu$ 의 값이 알려져 있을 때  $T$ 를 스튜던트화  $t$ 통계량이라 한다.
  - $\mu$ 의 값이 알려져 있는 않는 경우에도  $T$ 는  $t$ 분포를 따르지만 미지의 모수( $\mu$ )를 포함하므로 통계량은 아니다.
  - 스튜던트화  $t$ 통계량은 주로 모수의 추정량에 포함되는 장애모수(nuisance parameter)인 분산의 추정에서 비롯된다. (장애모수는 추정하고자 하는 모수 이외의 모수이다.)
  - 자유도  $(n-1)$ 이 커짐에 따라 표준정규분포로 가까워진다.

## t분포 예제 1

- 나주지방에서 생산되는 배의 무게는 평균  $\mu = 144\text{g}$ 인 정규분포를 따른다고 한다. 나주에서 생산된 배 9개를 랜덤으로 추출하여 구한 표본표준편차가  $s = 6\text{g}$ 이었다고 할 때 표본평균  $\bar{X}$ 가  $148.612\text{g}$ 보다 클 확률을 구하여라.
  - 모집단의 분포는  $\sigma^2$ 이 알려져 있지 않은 정규분포이므로  $T = \frac{\bar{X}-144}{\frac{6}{\sqrt{9}}}$ 는 자유도 8인  $t$ -분포를 따른다.
  - $P(\bar{X} > 148.612) = P\left(\frac{\bar{X}-144}{\frac{6}{\sqrt{9}}} > \frac{148.612-144}{\frac{6}{\sqrt{9}}}\right) = P(T > 2.306) = 0.025$

| DF | t(0.1) | t(0.05) | t(0.025) | t(0.01) | t(0.005) | t(0.0025) |
|----|--------|---------|----------|---------|----------|-----------|
| 1  | 3.078  | 6.314   | 12.706   | 31.821  | 63.657   | 127.321   |
| 2  | 1.886  | 2.920   | 4.303    | 6.965   | 9.925    | 14.089    |
| 3  | 1.638  | 2.353   | 3.182    | 4.541   | 5.841    | 7.453     |
| 4  | 1.533  | 2.132   | 2.776    | 3.747   | 4.604    | 5.598     |
| 5  | 1.476  | 2.015   | 2.571    | 3.365   | 4.032    | 4.773     |
| 6  | 1.440  | 1.943   | 2.447    | 3.143   | 3.707    | 4.317     |
| 7  | 1.415  | 1.895   | 2.365    | 2.998   | 3.499    | 4.029     |
| 8  | 1.397  | 1.860   | 2.306    | 2.896   | 3.355    | 3.833     |
| 9  | 1.383  | 1.833   | 2.262    | 2.821   | 3.250    | 3.690     |
| 10 | 1.372  | 1.812   | 2.228    | 2.764   | 3.169    | 3.581     |

t분포표

## $t$ 분포 예제 2

- 어느 고등학교에서 학생의 IQ는 평균  $\mu = 120$ 이라고 하자. 이 모집단으로부터 100명의 학생을 랜덤으로 추출하여 표본표준편차  $s = 5$ 를 얻었다. 이 때 표본평균  $\bar{X}$ 가 119보다 크지 않을 확률을 구하여라.
  - 모집단의 분포와 모분산이 알려져 있지 않으나 표본크기가 100으로 크므로, 구하는 확률은 표준정규분포표를 이용하여 근사적으로 구할 수 있다.
  - $$P(\bar{X} \leq 119) = P\left(\frac{\bar{X}-120}{\frac{s}{\sqrt{100}}} \leq \frac{119-120}{\frac{5}{\sqrt{100}}}\right) \cong P(Z \leq -2) = 1 - P(Z \leq 2) = 1 - 0.9773 = 0.0227$$