

# 생생한 사례로 배우는 확률과 통계

## [강의교안 이용 안내]

- 본 강의교안의 저작권은 **이재원**과 **한빛아카데미(주)**에 있습니다.
- 이 자료를 무단으로 전제하거나 배포할 경우 저작권법 136조에 의거하여 벌금에 처할 수 있고 이를 병과(併科)할 수도 있습니다.



## Chapter 07

생생한 사례로 배우는

### 확률과 통계

PROBABILITY & STATISTICS

## 정규분포

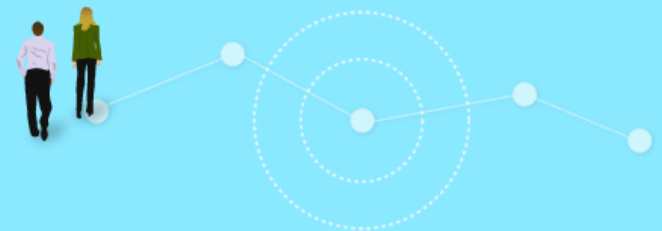
Normal Distribution

# 목 차

## 7.1 정규분포

## 7.2 정규분포와 관련된 분포

# 7.1 정규분포



# 정규분포의 성질

**정규분포** Normal distribution : 확률변수  $X$ 의 상태공간  $S_X = \{x : -\infty < x < \infty\}$ 에서 다음 확률밀도함수를 갖는 확률분포. 이때  $X \sim N(m, s^2)$ 으로 나타낸다. 여기서  $m > 0, -\infty < s < \infty$ .

$$f(x) = \frac{1}{\sqrt{2\pi}s} e^{-(x-m)^2/(2s^2)}, \quad -\infty < x < \infty$$

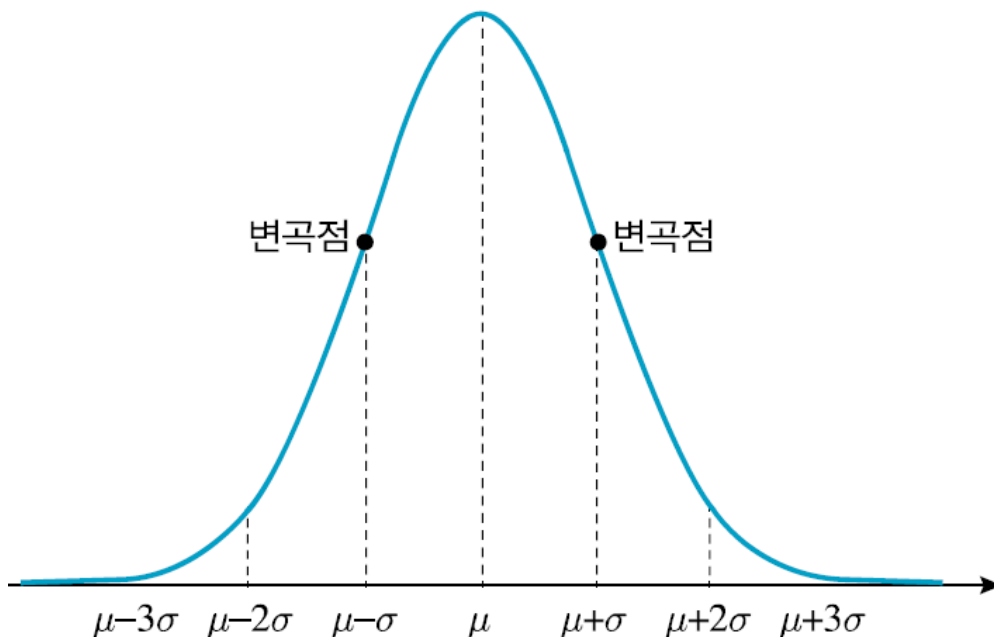
① 평균 :  $\mu = m$

② 분산 :  $\sigma^2 = s^2$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty$$

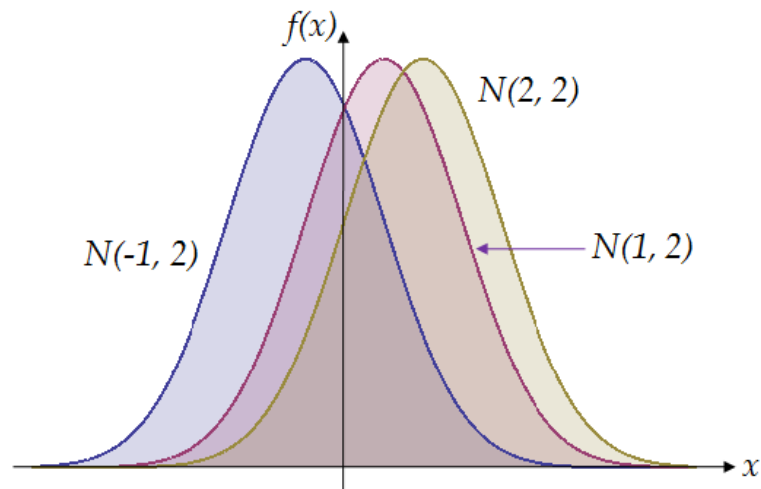
# 정규곡선의 성질

- ①  $f(x)$ 는  $x = \mu$ 에 관하여 좌우대칭이고, 따라서  $X$ 의 중위수는  $Me = \mu$ 이다.
- ②  $f(x)$ 는  $x = \mu$ 에서 최대값을 가지고, 따라서  $X$ 의 최빈값은  $Mo = \mu$ 이다.
- ③  $x = \mu \pm \sigma$ 에서는 변곡점을 갖는다. 즉, 곡선의 모양이 위로 볼록하다가 아래로 볼록하게 바뀐다.
- ④  $x = \mu \pm 3\sigma$ 에서  $x$ -축에 거의 접하는 모양을 가지고  $x \rightarrow \pm\infty$ 이면  $f(x) \rightarrow 0$ 이다.

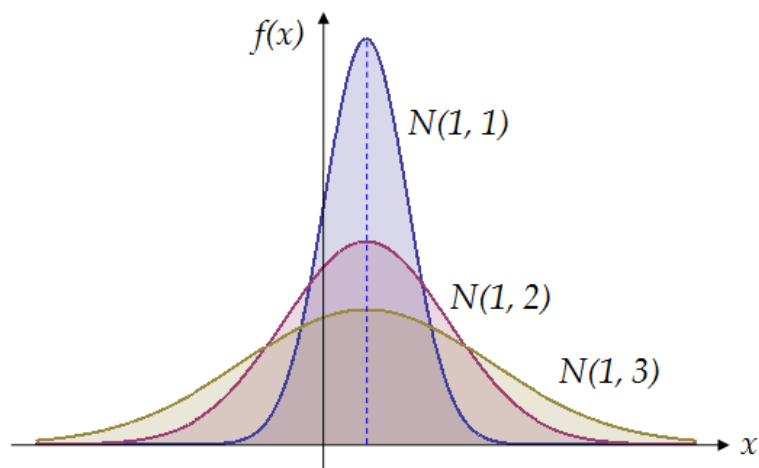


# 정규분포의 성질

❖ 평균이 다르고 분산이 동일한 경우



❖ 평균이 동일하고 분산이 다른 경우



# 표준정규분포

- ❖ 평균  $\mu = 0$ , 분산  $\sigma^2 = 1$ 인 정규분포를 **표준정규분포** standard normal distribution라 하고,  $Z \sim N(0, 1)$ 로 나타낸다.

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

- ①  $\phi(z)$ 는  $z = 0$ 에 관하여 좌우대칭이고, 따라서  $Z$ 의 중위수는  $Me = 0$ 이다.
- ②  $\phi(z)$ 는  $z = 0$ 에서 최대값을 가지고, 따라서  $Z$ 의 최빈값은  $Mo = 0$ 이다.
- ③  $z = \pm 1$ 에서 는 변곡점을 갖는다. 즉, 곡선의 모양이 위로 볼록하다가 아래로 볼록하게 바뀐다.
- ④  $z = \pm 3$ 에서  $z$ -축에 거의 접하는 모양을 가지고  $z \rightarrow \pm\infty$ 이면  $\phi(z) \rightarrow 0$ 이다.



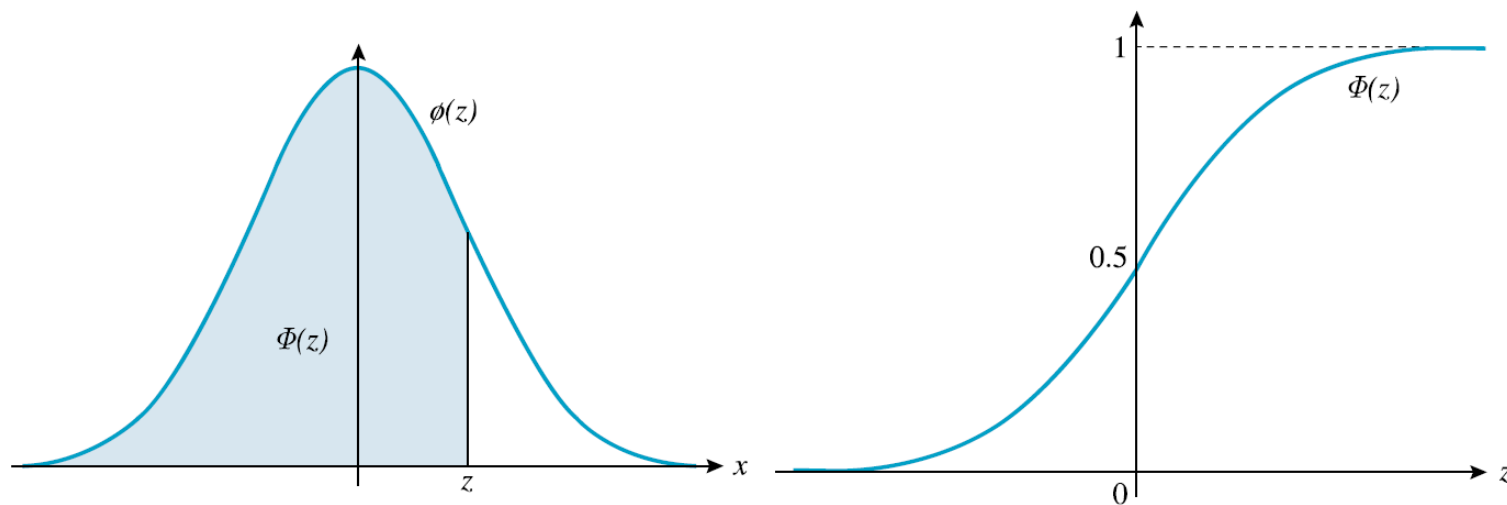
# 표준정규확률변수의 분포함수

❖  $Z \sim N(0, 1)$ 의 분포함수  $\Phi(z) = P(Z \leq z)$ 는 다음과 같다.

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du, \quad -\infty < z < \infty$$

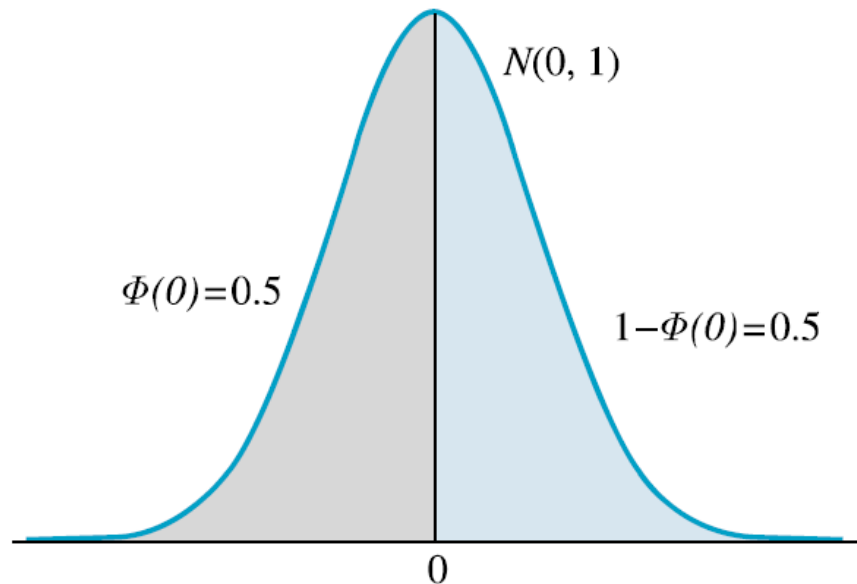
①  $\Phi(z)$ 는  $z$ 와  $\phi(z)$ 아래로 둘러싸인 부분의 넓이다.

②  $\Phi(z)$ 는 그림과 같이 S곡선을 이룬다.



# 표준정규분포의 성질

- ❖ 표준정규분포는  $z = 0$ 에 관하여 좌우대칭이므로 [그림]과 같이  $z = 0$ 의 왼쪽과 오른쪽의 넓이가 동일하다.

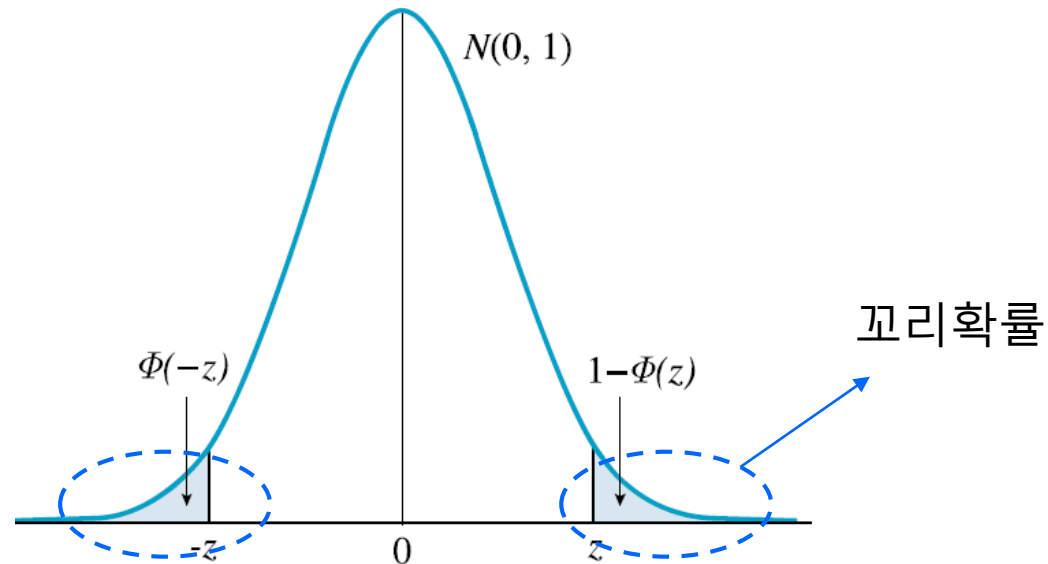


$$\Phi(0) = P(Z \leq 0) = 0.5$$

$$1 - \Phi(0) = P(Z \geq 0) = 0.5$$

# 표준정규분포의 성질

❖ 함수  $\phi(z)$ 의 대칭성으로부터 양쪽 꼬리확률 tail probability은 동일하다.

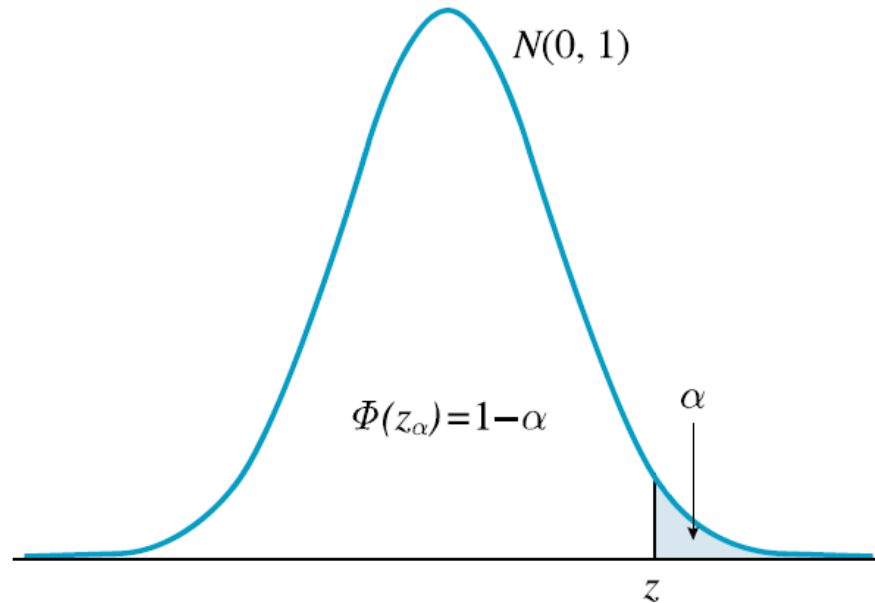


$$\Phi(-z) = P(Z \leq -z) = P(Z \geq z) = 1 - \Phi(z), \quad z > 0$$

$$\Phi(-z) + \Phi(z) = 1$$

# 표준정규분포의 백분위수

- ❖ 표준정규분포의 오른쪽 꼬리확률  $P(Z \geq z) = 1 - \Phi(z)$ 가  $\alpha$ 인  $100(1 - \alpha)\%$  백분위수를  $z_\alpha$ 로 나타낸다.

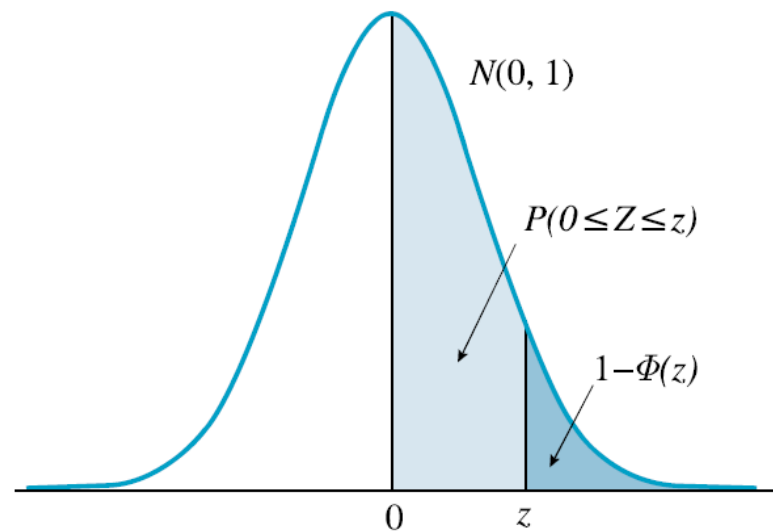
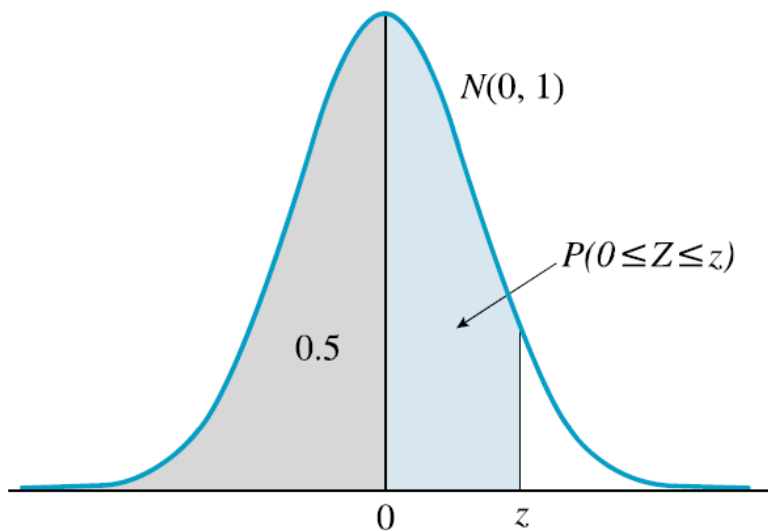


$$P(Z \leq z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$$

$$P(Z \geq z_\alpha) = 1 - \Phi(z_\alpha) = \alpha$$

# 표준정규분포의 백분위수

❖  $P(Z \leq 0) = 0.5$ 이므로 다음을 얻는다.



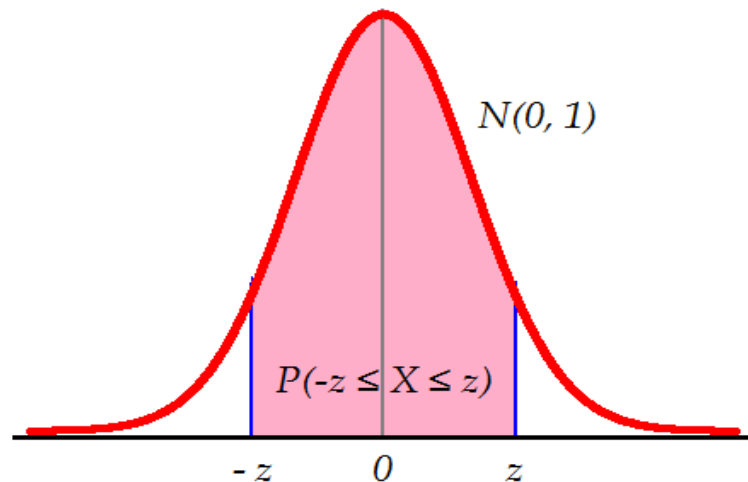
$$\Phi(z) = 0.5 + P(0 \leq Z \leq z), \quad z > 0$$

$$P(0 \leq Z \leq z) = \Phi(z) - 0.5$$

$$P(Z \geq z) = 1 - \Phi(z)$$

# 표준정규분포의 백분위수

❖ 함수  $\phi(z)$ 의 대칭성으로부터 다음을 얻는다.



$$P(-z \leq Z \leq 0) = P(0 \leq Z \leq z), \quad z > 0$$

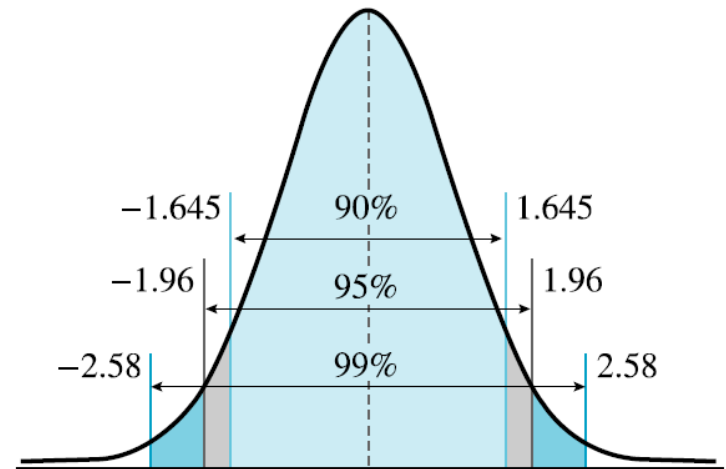
$$P(-z \leq Z \leq z) = 2P(0 \leq Z \leq z) = 2\Phi(z) - 1$$

# 표준정규분포의 백분위수

- ❖ 오른쪽 꼬리확률이  $\alpha = 0.05, 0.025, 0.005$ 인 백분위수  $z_\alpha$

$$P(Z > 1.645) = 0.05, \quad P(Z > 1.96) = 0.025, \quad P(Z > 2.58) = 0.005$$

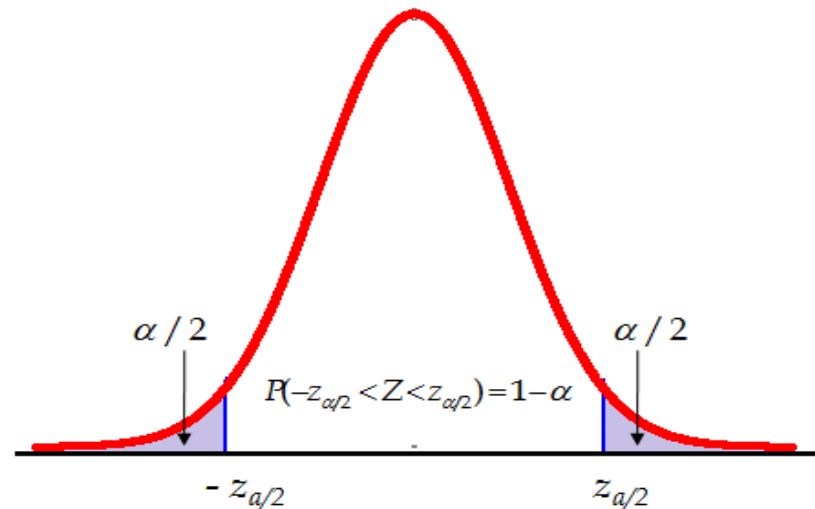
- ❖ 양쪽 꼬리확률이 각각  $\alpha/2 = 0.05, 0.025, 0.005$ 인  $\alpha$ 에 대한 중심확률  $P(|Z| \leq z_{\alpha/2})$



$$P(|Z| \leq 1.645) = 0.9, \quad P(|Z| \leq 1.96) = 0.95, \quad P(|Z| \leq 2.58) = 0.99$$

# 표준정규분포의 백분위수

- ❖ 일반적으로 양쪽 꼬리확률이 각각  $\alpha/2$ 인 두 백분위수-  $z_{\alpha/2}$ ,  $z_{\alpha/2}$ 에 대해 다음이 성립한다.



$$P(|Z| \leq z_{\alpha/2}) = 1 - \alpha$$



# 확률계산

- ❖ 표준정규분포에 대한 확률계산은 <부록 A3>에 제시한 표준정규분포표를 이용한다.
- ❖ 예를 들어,  $P(Z \leq 1.15)$ 를 구한다면, [그림]의 표준정규분포표를 이용하여 다음 순서에 따른다.
  - ① z열에서 소수점 첫 째 자리까지 나타낸 숫자 1.1을 선택한다.
  - ② z행에서 소수점 둘 째 자리를 나타내는 숫자 .05를 선택한다.
  - ③ z열이 1.1인 행과 z행이 .05인 열이 만나는 위치의 수 .8749를 선택한다.
  - ④  $F(1.15) = P(Z \leq 1.15) = 0.8749$ 이다.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8949	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9182	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

# 확률계산

- ❖ 표준정규확률변수  $Z$ 에 대하여,  $P(a \leq Z \leq b)$ 는 분포함수를 이용하여 다음과 같이 구할 수 있다.

$$P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$$

예를 들어,  $P(0.62 \leq Z \leq 1.37)$ 을 구한다면, 표준정규분포표를 이용하여 다음을 구한다.

$$\Phi(1.37) = 0.9147, \Phi(0.62) = 0.7324$$

그러면  $P(0.62 \leq Z \leq 1.37) = \Phi(1.37) - \Phi(0.62) = 0.9147 - 0.7324 = 0.1823$ 이다.

## 예제 7-1

표준정규분포표를 이용하여 다음 확률을 구하라.

(a)  $P(Z \leq 1.74)$

(b)  $P(Z \geq 1.13)$

풀이

(a) 표준정규분포표에서  $z$ 열의 1.7과  $z$ 행의 .04이 만난 위치의 수 0.9591을 선택한다.  
그러면  $P(Z \leq 1.74) = 0.9591$ 이다.

(b) 이 확률을 구하기 위해 먼저  $P(Z \leq 1.13)$ 을 먼저 구한다. 표준정규분포표에서  $z$ 열의 1.1과  $z$ 행의 .03이 만난 위치의 수 0.8708을 선택한다. 그러면  $P(Z \leq 1.13) = 0.8708$ 이다. 그러므로 구하고자 하는 확률은 다음과 같다.

$$P(Z \geq 1.13) = 1 - P(Z \leq 1.13) = 1 - 0.8708 = 0.1292$$

## 예제 7-2

표준정규분포표를 이용하여 다음 확률을 구하라.

(a)  $P(Z \leq -0.57)$

(b)  $P(-1.15 \leq Z \leq 2.12)$

풀이

$$(a) P(Z \leq -0.57) = P(Z \geq 0.57) = 1 - P(Z \leq 0.57) = 1 - 0.7157 = 0.2843$$

$$\begin{aligned}(b) P(-1.15 \leq Z \leq 2.12) &= P(Z \leq 2.12) - P(Z \leq -1.15) \\ &= P(Z \leq 2.12) - P(Z \geq 1.15) \\ &= 0.9830 - 0.1251 = 0.8579\end{aligned}$$

## 예제 7-3

표준정규분포표를 이용하여  $P(-2.13 \leq Z \leq 2.13)$ 을 구하라.

풀이

<부록A3> 에서  $P(Z \leq 2.13) = 0.9834$ 이므로  $P(0 \leq Z \leq 2.13) = 0.4834$ 이고,  
구하고자 하는 확률은 다음과 같다.

$$P(-2.13 \leq Z \leq 2.13) = 2P(0 \leq Z \leq 2.13) = 2 \times 0.4834 = 0.9668$$

# 확률계산

❖  $X \sim N(\mu, \sigma^2)$ 과  $Z \sim N(0, 1)$  사이에 다음이 성립한다.

$$X \sim N(\mu, \sigma^2) \Leftrightarrow Z \sim N(0, 1)$$

❖ 구간  $a \leq X \leq b$ 를 표준화한 구간  $z_l \leq Z \leq z_r$ 로 변경하면 다음과 같다.

$$z_l = \frac{a - \mu}{\sigma}, \quad z_r = \frac{b - \mu}{\sigma}$$

❖ 정규확률  $P(a \leq X \leq b)$ 를 표준정규확률로 변환하면 다음과 같다.

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = \Phi(z_r) - \Phi(z_l)$$

# 확률계산

## 예제 7-4

의료폐수를 하수관로에 흘려보낼 때 발생하는 대장균의 법정 배출허용 기준치는 1ml당 3,000개이다. 어느 의료기관에서 배출되는 폐수로 인한 대장균 수는 평균 3,000개, 표준편차 200개인 정규분포를 따른다고 한다. 이때 다음을 구하라.

- (a) 대장균 수가 2,500개 이하일 확률
- (b) 대장균 수가 3,346개 이상일 확률
- (c) 대장균 수가 2,870개 이상 3,426개 이하일 확률

**풀이** (a) 1ml 표본 안에 들어 있는 박테리아 수를  $X$ 라 하면, 다음 표준화 확률변수는 표준정규분포에 따른다.

$$Z = \frac{X - 3000}{200}$$

$x = 2500$ 을 표준화하면  $z_l = (2500 - 3000)/200 = -2.5$ 이므로 구하고자 하는 확률은 다음과 같다.

$$P(X \leq 2500) = P(Z \leq -2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062$$

- (b)  $x = 3346$ 을 표준화하면  $z_r = (3346 - 3000)/200 = 1.73$ 이므로 구하고자 하는 확률은 다음과 같다.

$$P(X \geq 3346) = P(Z \geq 1.73) = 1 - \Phi(1.73) = 1 - 0.9582 = 0.0418$$

- (c)  $l = 2870$ ,  $r = 3426$ 을 표준화하면 각각 다음과 같다.

$$z_l = \frac{2870 - 3000}{200} = -0.65, \quad z_r = \frac{3426 - 3000}{200} = 2.13$$

따라서 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(2870 \leq X \leq 3426) &= P(-0.65 \leq Z \leq 2.13) \\ &= \Phi(2.13) - \Phi(-0.65) \\ &= \Phi(2.13) - (1 - \Phi(0.65)) \\ &= 0.9821 - (1 - 0.7422) = 0.7243 \end{aligned}$$



# 일반적인 정규분포의 백분위수

- ❖  $X \sim N(\mu, \sigma^2)$ 과  $Z \sim N(0, 1)$  사이의 관계를 이용하여  $P(X \leq x_\alpha) = \alpha$ 를 만족하는  $100(1 - \alpha)\%$  백분위수  $x_\alpha$ 를 다음과 같이 구할 수 있다.

$$z_\alpha = \frac{x_\alpha - \mu}{\sigma}, \quad x_\alpha = \mu + \sigma z_\alpha$$

# 일반적인 정규분포의 백분위수

## 예제 7-5

$X$ 가 평균 15, 분산 2.25인 정규분포를 따른다.

- (a)  $P(X < x_0) = 0.9265$ 를 만족하는  $x_0$ 를 구하라.
- (b)  $P(X > x_0) = 0.0049$ 를 만족하는  $x_0$ 를 구하라.
- (c)  $P(X \leq x_0) = 0.025$ 를 만족하는  $x_0$ 를 구하라.

풀이

- (a)  $x_0$ 의 표준화를  $z_0$ 이라 하면  $z_0 = (x_0 - 15)/1.5$ 이고  $P(X < x_0) = \Phi(z_0) = 0.9265$ 이다. 그러므로 다음과 같이 표준정규확률표에서 임계점  $z_0$ 을 찾으면  $z_0 = 1.45$ 이다.

$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

따라서 구하고자 하는 백분위수는 다음과 같다.

$$x_0 = \mu + \sigma z_0 = 15 + 1.5 \times 1.45 = 17.175$$

# 일반적인 정규분포의 백분위수

- (b)  $x_0$ 의 표준화를  $z_0$ 이라 하면,  $z_0 = (x_0 - 15)/1.5$ 이고  $P(X > x_0) = P(Z > z_0) = 0.0049$ 이므로  $\Phi(z_0) = 0.9951$ 이다. 그러므로 다음과 같이 표준정규확률표에서 임계점  $z_0$ 을 찾으면  $z_0 = 2.58$ 이다.

$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952

따라서 구하고자 하는 백분위수는 다음과 같다.

$$x_0 = \mu + \sigma z_0 = 15 + 1.5 \times 2.58 = 18.87$$

- (c)  $x_0$ 의 표준화를  $z_0$ 이라 하면,  $z_0 = (x_0 - 15)/1.5$ 이고  $\Phi(z_0) = 0.025$ 이므로 표준정규분포의 대칭성으로부터  $P(Z \geq -z_0) = 0.025$ , 즉  $\Phi(-z_0) = 0.9750$ 이다. 그러므로 다음과 같이 표준정규확률표에서 임계점  $-z_0$ 을 찾으면  $-z_0 = 1.96$ , 즉  $z_0 = -1.96$ 이다.

$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

따라서 구하고자 하는 백분위수는 다음과 같다.

$$x_0 = \mu + \sigma z_0 = 15 + 1.5 \times (-1.96) = -12.06$$

# 정규분포의 일차결합

- ❖  $X \sim N(\mu, \sigma^2)$ 이면  $Y = aX + b (a \neq 0)$ 은 평균  $a\mu + b$ , 분산  $a^2\sigma^2$ 인 정규분포에 따른다. 즉, 다음이 성립한다.

$$X \sim N(\mu, \sigma^2) \Leftrightarrow aX + b \sim N(a\mu, a^2\sigma^2), \quad a \neq 0$$

- ❖  $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ 이 독립이면 다음이 성립한다.

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

$$X \pm Y \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$$

- ❖  $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ 이 독립이면  $aX + b$ 와  $X \pm Y$ 의 표준화는 다음과 같다.

$$\textcircled{1} \quad \frac{(aX + b) - (a\mu_1 + b)}{|a|\sigma_1} \sim N(0, 1)$$

$$\textcircled{2} \quad \frac{(X \pm Y) - (\mu_1 \pm \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim N(0, 1)$$

# 정규분포의 일차결합

## 예제 7-6

독립인 두 정규확률변수  $X \sim N(25, 4)$ ,  $Y \sim N(20, 9)$ 에 대해  $U = 2X - Y$ 라 한다.

- (a)  $U$ 의 확률분포를 구하라.
- (b) 확률  $P(28 < U < 39)$ 를 구하라.
- (c)  $P(U \leq u_0) = 0.005$ 를 만족하는  $u_0$ 를 구하라.

풀이

- (a)  $U$ 의 평균과 분산은 각각 다음과 같다.

$$\mu_U = 2\mu_1 - \mu_2 = 2 \times 25 - 20 = 30, \quad \sigma_U^2 = 2^2 \sigma_1^2 + \sigma_2^2 = 4 \times 4 + 9 = 25$$

따라서  $U \sim N(30, 25)$ 이다.

# 정규분포의 일차결합

(b) 28과 39를 표준화하면 다음과 같다.

$$z_l = \frac{28-30}{5} = -0.4, \quad z_r = \frac{39-30}{5} = 1.8$$

따라서 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(28 < X < 39) &= P(-0.4 < Z < 1.8) = \Phi(1.8) - \Phi(-0.4) \\ &= \Phi(1.8) - [1 - \Phi(0.4)] = 0.9641 - (1 - 0.6554) \\ &= 0.6195 \end{aligned}$$

(c)  $u_0$ 을 표준화하면  $z_0 = \frac{u_0 - 30}{5}$  이고  $P(U \leq u_0) = \Phi(z_0) = 0.005$  이다.

따라서  $P(Z \geq -z_0) = 1 - \Phi(-z_0) = 0.005$ ,  $\Phi(-z_0) = 0.99$ 이다. 이제 표준정규확률표에서 임계점  $-z_0$ 을 찾으면 보간법에 의해  $-z_0 = 2.575$ , 즉  $z_0 = -2.575$ 이다. 따라서 구하고자 하는 백분위수는 다음과 같다.

$$u_0 = \mu + \sigma z_0 = 30 + 5 \times (-2.575) = 17.125$$

# 표본평균의 분포

❖  $X_i \sim N(\mu_i, \sigma_i^2)$ ,  $i = 1, 2, \dots, n$  이 독립이면 다음이 성립한다.

$$\textcircled{1} \quad Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n \sim N(\mu, \sigma^2), \quad \mu = \sum_{i=1}^n a_i \mu_i, \quad \sigma^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$$

$$\textcircled{2} \quad Y = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \sim N(\mu, \sigma^2), \quad \mu = \frac{1}{n} \sum_{i=1}^n \mu_i, \quad \sigma^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2$$

❖  $X_i \sim N(\mu, \sigma^2)$ ,  $i = 1, 2, \dots, n$  이 *i.i.d*이면 다음이 성립한다.

**표본평균** sample mean  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  의 확률분포는 다음과 같다.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

# 표본평균의 분포

## 예제 7-7

i.i.d 정규확률변수  $X_i \sim N(12, 4)$ ,  $i = 1, 2, \dots, 16$ 에 대해 표본평균을  $\bar{X}$ 라 한다.

(a)  $\bar{X}$ 의 확률분포를 구하라.

(b) 확률  $P(11.25 < \bar{X} < 13.11)$ 을 구하라.

**풀이** (a)  $n = 16$ 이므로  $\bar{X}$ 의 평균은  $\mu = 12$ , 분산은  $\sigma^2 = \frac{4}{16} = \frac{1}{4}$ 이고  $\bar{X} \sim N\left(12, \frac{1}{2^2}\right)$ 이다.

(b) 11.25와 13.11을 표준화하면 다음과 같다.

$$z_l = \frac{11.25 - 12}{1/2} = -1.5, \quad z_r = \frac{13.11 - 12}{1/2} = 2.22$$

따라서 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(11.25 < \bar{X} < 13.11) &= P(-1.5 < Z < 2.22) \\ &= \Phi(2.22) - \Phi(-1.5) \\ &= \Phi(2.22) - [1 - \Phi(1.5)] \\ &= 0.9868 - (1 - 0.9332) = 0.9200 \end{aligned}$$



# 중심극한정리

❖  $X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$ 이면 표본평균  $\bar{X}$ 의 확률분포는 다음과 같다.

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

❖ 임의의 확률변수  $X_i, i = 1, 2, \dots, n$ 가 *i.i.d*이고 평균  $\mu$ , 분산  $\sigma^2$ 을 갖는다고 하자. 이때  $n$ 이 충분히 커지면 표본평균  $\bar{X}$ 는 다음 정규분포에 근사하며, 이것을 **중심극한정리** central limit theorem이라 한다.

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

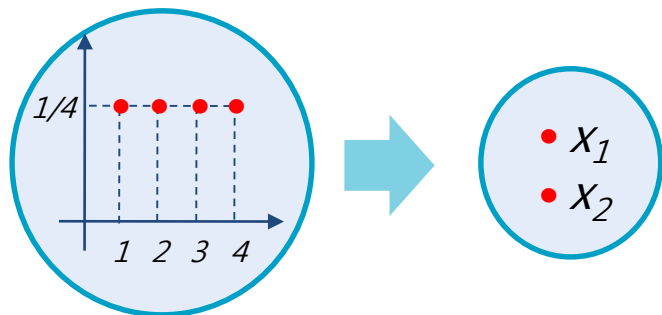
❖ 확률변수  $X_i$ 의 합,  $X_1 + X_2 + \dots + X_n$ 은 다음 정규분포에 근사한다.

$$X_1 + X_2 + \dots + X_n \approx N(n\mu, n\sigma^2)$$

# 중심극한정리

예 모집단 분포가  $p(x) = \frac{1}{4}, x = 1, 2, 3, 4$ 인 모집단에서 크기가 2, 3, 4, 5인 표본을 추출할 때, 표본평균  $\bar{X}$ 의 분포를 살펴보자.

$n = 2$ 인 표본의 표본평균



모평균 :  $\mu = 2.5$ ,      모분산 :  $\sigma^2 = 1.25$

표본평균 :  $\bar{X} = \frac{X_1 + X_2}{2}$ 의 확률분포 ?

$X_1$ 과  $X_2$ 의 결합분포

$X_1$ 과  $X_2$ 가 취하는 값 : 1, 2, 3, 4

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2) &= P(X_1 = x_1) \cdot P(X_2 = x_2) \\ &= \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16} \end{aligned}$$

$(x_1, x_2)$ 의 취하는 값 : (1, 1), (1, 2), ..., (4, 4)

$x_1 \backslash x_2$	1	2	3	4	$f_{X_2}$
1	1/16	1/16	1/16	1/16	4/16
2	1/16	1/16	1/16	1/16	4/16
3	1/16	1/16	1/16	1/16	4/16
4	1/16	1/16	1/16	1/16	4/16
$f_{X_1}$	4/16	4/16	4/16	4/16	1

# 중심극한정리

$\bar{X} = \frac{X_1 + X_2}{2}$  가 취하는 값 : 1, 1.5, 2, 2.5, 3, 3.5, 4

$x_1 \backslash x_2$	1	2	3	4	$p_{x_2}$	
1	1/16	1/16	1/16	1/16	4/16	$\bar{x} = 1$
2	1/16	1/16	1/16	1/16	4/16	$\bar{x} = 1.5$
3	1/16	1/16	1/16	1/16	4/16	$\bar{x} = 2$
4	1/16	1/16	1/16	1/16	4/16	$\bar{x} = 2.5$
$p_{x_1}$	4/16	4/16	4/16	4/16	1	$\bar{x} = 3$

$\bar{x} = 3.5$

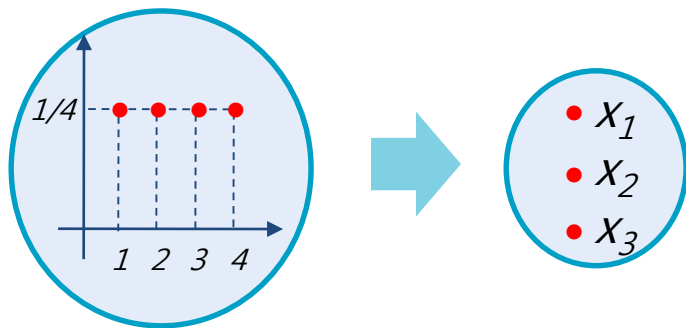
$\bar{x} = 4$

표본평균  $\bar{X} = \frac{X_1 + X_2}{2}$  의 확률분포

	1	1.5	2	2.5	3	3.5	4
$p_{\bar{x}}$	1/16	2/16	3/16	4/16	3/16	2/16	1/16

# 중심극한정리

$n = 3$ 인 표본의 표본평균



표본평균 :  $\bar{X} = \frac{X_1 + X_2 + X_3}{3}$  의 확률분포 ?

$X_1$ 과  $X_2$ 가 취하는 값 : 1, 2, 3, 4

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, X_3 = x_3) &= P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot P(X_3 = x_3) \\ &= \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{64} \end{aligned}$$

$(x_1, x_2, x_3)$ 의 취하는 값 : (1, 1, 1), (1, 1, 2), ..., (4, 4, 4)

# 중심극한정리

$\bar{X} = \frac{X_1 + X_2 + X_3}{3}$  가 취하는 값 : 1, 4/3, 5/3, 2, 7/3, 8/3, 3, 10/3, 11/3, 4

표본평균  $\bar{X} = \frac{X_1 + X_2 + X_3}{3}$  의 확률분포

$\bar{x}$	1	$\frac{4}{3}$	$\frac{5}{3}$	2	$\frac{7}{3}$	$\frac{8}{3}$	3	$\frac{10}{3}$	$\frac{11}{3}$	4
$p_{\bar{x}}$	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{6}{64}$	$\frac{10}{64}$	$\frac{12}{64}$	$\frac{12}{64}$	$\frac{10}{64}$	$\frac{6}{64}$	$\frac{3}{64}$	$\frac{1}{64}$

$$\mu_{\bar{X}} = \mu = 2.5,$$
$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{3} = 0.417$$

# 중심극한정리

$n = 4$ 인 표본의 표본평균

표본평균  $\bar{X} = \frac{X_1 + X_2 + X_3 + X_4}{4}$ 의 확률분포

$\bar{x}$	1	5/4	6/4	7/4	2	9/4	10/4
$p_{\bar{x}}$	$1/4^4$	$4/4^4$	$10/4^4$	$20/4^4$	$31/4^4$	$40/4^4$	$44/4^4$
$\bar{x}$	11/4	3	13/4	14/4	15/4	4	
$p_{\bar{x}}$	$401/4^4$	$31/4^4$	$20/4^4$	$10/4^4$	$4/4^4$	$1/4^4$	

# 중심극한정리

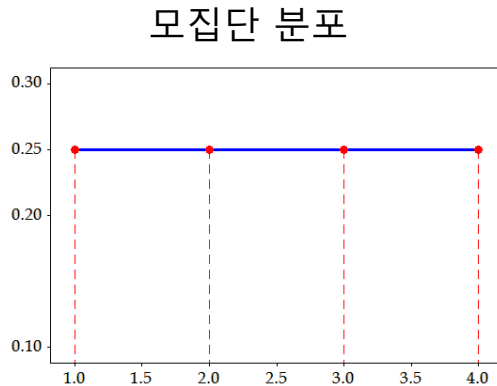
$n = 5$ 인 표본의 표본평균

표본평균  $\bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$  의 확률분포

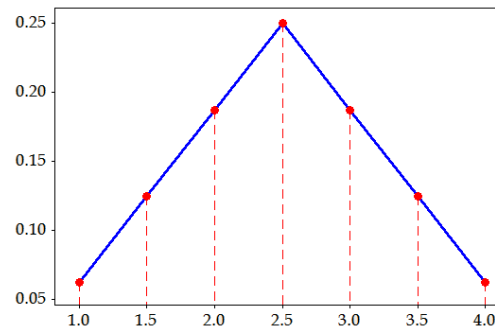
$\bar{x}$	1	6/5	7/5	8/5	9/5	2	11/5
$p_{\bar{x}}$	$1/4^5$	$5/4^5$	$15/4^5$	$35/4^5$	$65/4^5$	$101/4^5$	$135/4^5$
$\bar{x}$	$13/5$	$14/5$	3	$16/5$	$17/5$	$18/5$	$19/5$
$p_{\bar{x}}$	$155/4^5$	$135/4^5$	$101/4^5$	$65/4^5$	$35/4^5$	$15/4^5$	$5/4^5$

# 중심극한정리

$$\text{표본평균의 평균과 분산 : } \mu_{\bar{X}} = \mu, \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

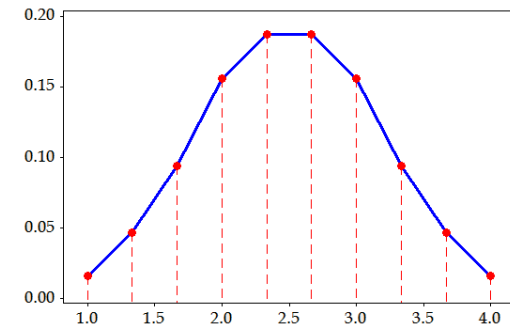


$n = 2$ 인 경우

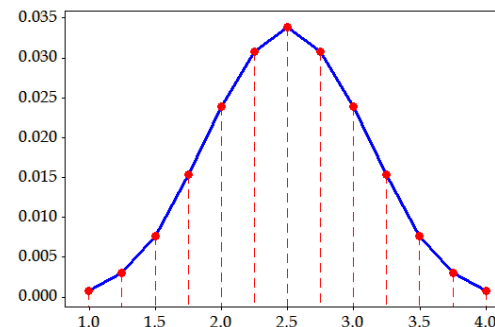


표본평균의 분포

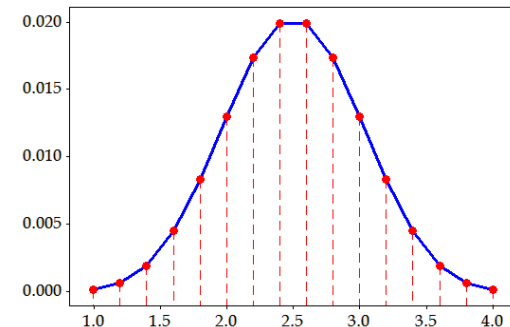
$n = 3$ 인 경우



$n = 4$ 인 경우



$n = 5$ 인 경우



❖  $n$ 이 커질수록 표본평균의 분포는 종모양에 가까워진다.



# 중심극한정리

## 예제 7-8

독립 확률변수  $X_1, X_2, \dots, X_{25}$ 가 각각 평균  $\mu = 120$ , 분산  $\sigma^2 = 36$ 인 항등분포를 따른다고 한다. 표본평균  $\bar{X}$ 에 대해  $P(118 \leq \bar{X} \leq 123)$ 을 구하라.

**풀이** 독립확률변수들  $X_i, i = 1, 2, \dots, 25$ 가 각각 평균  $\mu = 120$ , 분산  $\sigma^2 = 36$ 인 확률분포에 따르므로 표본평균  $\bar{X}$ 의 근사확률분포는 다음과 같다.

$$\bar{X} \approx N\left(120, \frac{36}{25}\right) = N(120, 1.2^2)$$

이제 118과 123을 표준화하면 다음과 같다.

$$z_l = \frac{118 - 120}{1.2} = -1.67, \quad z_r = \frac{123 - 120}{1.2} = 2.5$$

따라서 구하고자 하는 확률은  $P(118 < \bar{X} < 123) = P(-1.67 < Z < 2.5)$  다음과 같다.

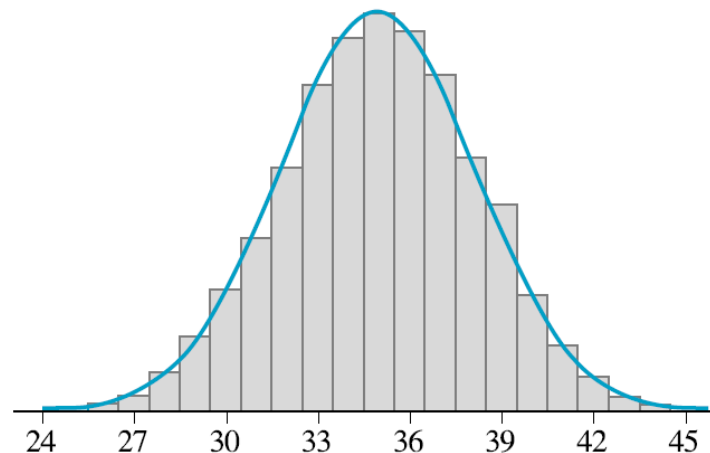
$$\begin{aligned} &\approx \Phi(2.5) - \Phi(-1.67) \\ &= \Phi(2.5) - [1 - \Phi(1.67)] \\ &= 0.9938 - (1 - 0.9525) = 0.9463 \end{aligned}$$

# 이항분포의 정규근사

- ❖ 이항분포  $B(n, p)$ 에 대하여  $np = \mu$ 가 일정하고  $n$ 이 충분히 크면 이항분포는 평균  $\mu$ 인 푸이송분포  $P(\mu)$ 에 근사하는 것을 살펴보았다.  
이때  $n \rightarrow \infty$ 이면  $p \rightarrow 0$ 이다.

$p \rightarrow 0$ 이 아니면 중심극한정리에 의해 이항분포  $B(n, p)$ 는 정규분포  $N(np, npq)$ 에 근사한다.

- ❖ 모수  $n$ 과  $p$ 인 이항분포에 대하여  $np \geq 5, nq \geq 5$ 인 경우,  $n$ 이 커질수록 이항분포  $B(n, p)$ 는 정규분포  $N(np, npq)$ 에 근사하며, 이것을 **이항분포의 정규근사** normal approximation라 한다.



$B(n, p)$ 와  $N(np, npq)$ 의 비교

# 이항분포의 정규근사

## 예제 7-9

확률변수  $X$ 가 모수  $n = 100$ ,  $p = 0.65$ 인 이항분포를 따른다. 그리고  $P(X \leq 62) = 0.29755$ ,  $P(X \leq 73) = 0.96486$ 이다.

- (a) 이항분포에 의해  $P(63 \leq X \leq 73)$ 을 구하라.
- (b) 정규근사에 의해  $P(63 \leq X \leq 73)$ 을 구하라.

**풀이** (a) 이항분포에 의해 주어진 확률을 구하면 다음과 같다.  
$$P(63 \leq X \leq 73) = P(X \leq 73) - P(X \leq 62) = 0.96486 - 0.29755 = 0.66731$$

(b)  $X \sim B(100, 0.65)$ 이므로  $X$ 는 다음 평균과 분산을 갖는 정규분포에 근사한다.

$$\mu = 100 \times 0.65 = 65, \sigma^2 = 100 \times 0.65 \times 0.35 = 22.75$$

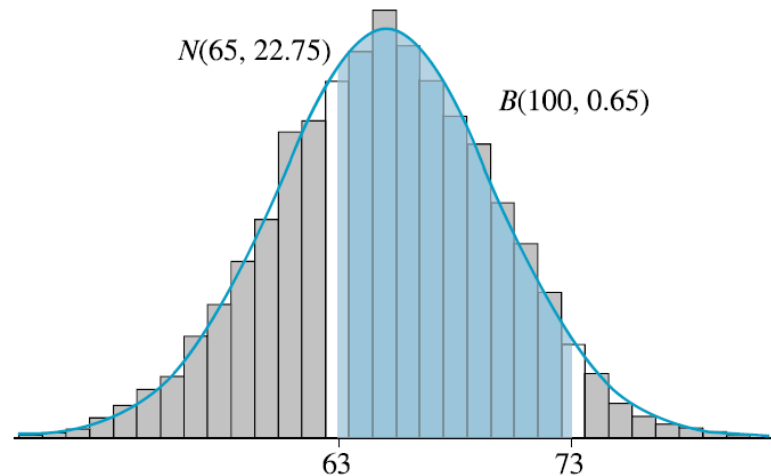
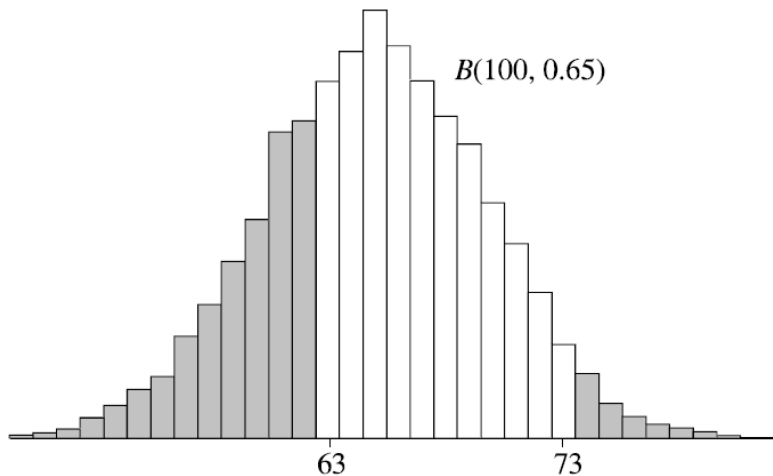
따라서 정규근사에 의한 근사확률은 다음과 같다.

$$\begin{aligned} P(63 \leq X \leq 73) &= P(-0.42 < Z < 1.68) \\ &\approx \Phi(1.68) - \Phi(-0.42) \\ &= \Phi(1.68) - [1 - \Phi(0.42)] \\ &= 0.9535 - (1 - 0.6628) = 0.6163 \end{aligned}$$

# 연속성 수정

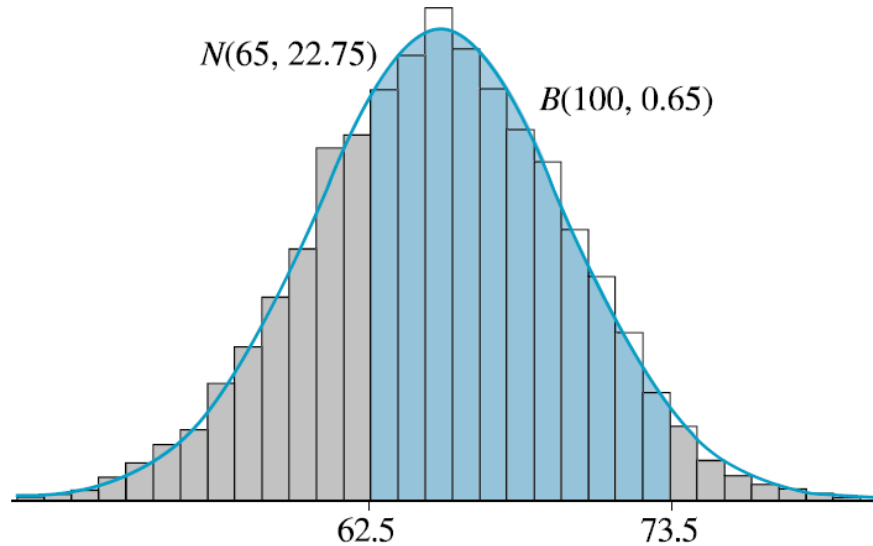
- ❖ [예제 9]에서  $P(63 \leq X \leq 73)$ 의 이항확률과 정규근사 확률 사이에 0.051의 차이가 생긴다.

이때 이항확률은 [그림]과 같이 63과 73을 중심으로 밑변의 길이가 1이고 높이가 확률 값인 사각형의 넓이이다. 따라서 정규근사에 의한 확률은 62.5~63과 73~ 73.5 사이의 확률(사각형의 넓이)이 누락된다.



# 연속성 수정

- ❖ 정규근사에 의한 오차를 줄이기 위하여 그림과 같이  $P(62.5 \leq X \leq 73.5)$ 의 정규근사 확률을 구한다.



- ❖ 이와 같이 보완한 **정규근사를 연속성 수정 정규근사** normal approximation with continuity correction factor라 하며, 보편적으로 정규근사라 함은 연속성 수정 정규근사를 의미한다.

# 연속성 수정

❖ 연속성 수정에 의한 정규근사 확률 계산

$$\textcircled{1} \quad P(X \leq b) \approx \Phi\left(\frac{b+0.5-\mu}{\sigma}\right), \quad \mu = np, \sigma^2 = npq$$

$$\textcircled{2} \quad P(a \leq X) \approx 1 - \Phi\left(\frac{a+0.5-\mu}{\sigma}\right)$$

$$\textcircled{3} \quad P(a \leq X \leq b) \approx \Phi\left(\frac{b+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a-0.5-\mu}{\sigma}\right)$$

$$\textcircled{4} \quad P(X = a) \approx \Phi\left(\frac{a+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a-0.5-\mu}{\sigma}\right)$$

## 예제 7-10

[예제 7-9]에서 연속성을 수정한 정규근사에 의해  $P(63 \leq X \leq 73)$ 을 구하라.

풀이

$P(63 \leq X \leq 73) = P(62.5 \leq X \leq 73.5)$ 이고 62.5와 73.5를 표준화하면 다음과 같다.

$$z_l = \frac{62.5 - 65}{4.77} = -0.57, \quad z_r = \frac{73.5 - 65}{4.77} = 1.78$$

그러므로 연속성을 수정한 정규근사에 의한 근사확률은 다음과 같다.

$$\begin{aligned} P(63 \leq X \leq 73) &= P(-0.57 < Z < 1.78) \\ &\approx \Phi(1.78) - \Phi(-0.57) \\ &= \Phi(1.78) - [1 - \Phi(0.57)] \\ &= 0.9625 - (1 - 0.6985) = 0.6610 \end{aligned}$$

## 7.2 정규분포와 관련된 분포





# 카이제곱분포

카이제곱분포 Chi-squared distribution :

$\alpha = r/2, \beta = 2$ 인 감마분포를 의미하며,  $X \sim \chi^2(r)$ 로 나타낸다.

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1} e^{-x/2}, \quad x > 0$$

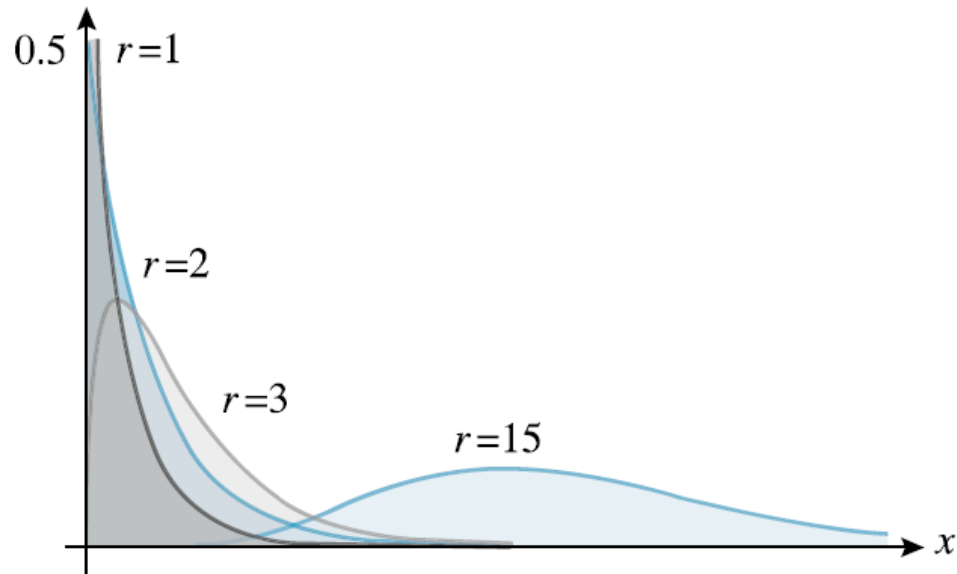
❖  $X \sim \Gamma(\alpha, \beta) \Rightarrow \mu = \alpha\beta, \sigma^2 = \alpha\beta^2$

① 평균 :  $\mu = r$

② 분산 :  $\sigma^2 = 2r$

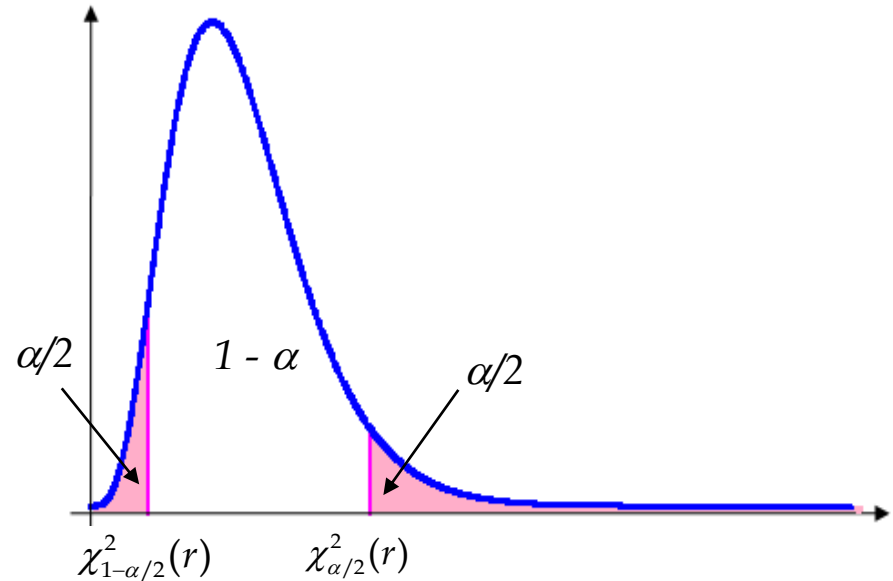
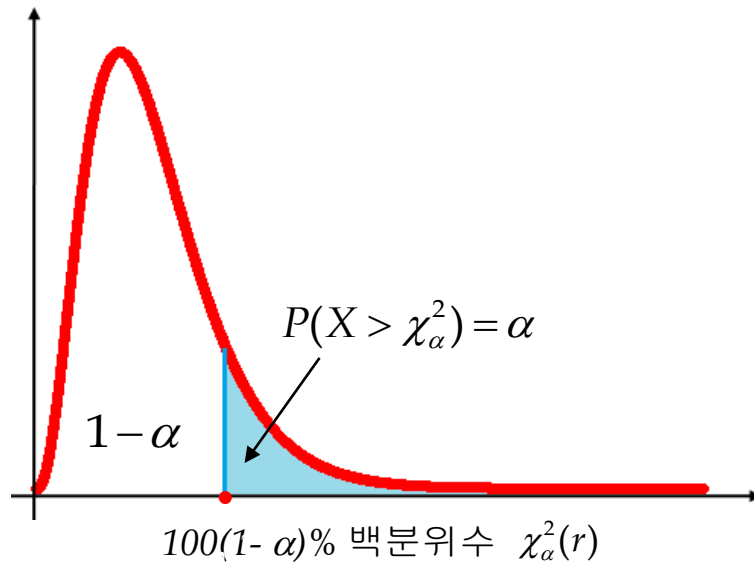
# 카이제곱분포곡선

- ① 왼쪽으로 치우치고 오른쪽으로 긴 꼬리를 갖는 분포, 즉 양의 왜도를 갖는 분포를 이룬다.
- ② 자유도  $r$ 이 커질수록 종모양의 분포에 가까워진다. 즉, 자유도  $r$ 이 커질수록 카이제곱분포는 정규분포에 근사한다.



# 카이제곱분포의 꼬리확률

- ❖ [그림]과 같이  $0 < a < 1$ 에 대하여 오른쪽 꼬리확률  $a = P(X > x_a)$ 를 만족하는  $100(1 - a)\%$  백분위수를  $x_a = \chi_a^2(r)$ 로 나타낸다.
- ❖ 중심확률이  $1 - a$ 인 두 임계점은  $\chi_{1-\alpha/2}^2(r)$ ,  $\chi_{\alpha/2}^2(r)$ 이다.



# 100(1 - $\alpha$ )% 백분위수 구하는 방법

- ❖ 백분위수  $\chi^2_{\alpha}(r)$  은 카이제곱분포표에서 [그림]과 같이 구할 수 있다.
- ❖ 예를 들어,  $X \sim c^2(5)$ 에 대하여  $P(X > \chi^2_{0.05}(5)) = 0.05$  를 만족하는 95% 백분위수  $\chi^2_{0.05}(5)$  는 다음과 같이 구한다.

$\chi^2_{0.05}(5) = 11.07, P(X > 11.07) = 0.05$

$\alpha$		0.250	0.200	0.150	0.100	0.050	0.025	0.020	0.010	0.005	0.0025
자유도 d.f	1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14
	2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98
	3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.31
	4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42
	5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39
	6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25
	7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04
	8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77
	9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46

# 카이제곱분포와 표준정규분포

- ❖ 표준정규 확률변수  $Z_1, Z_2, \dots, Z_n$ 에 대하여  $V = Z_1^2 + Z_2^2 + \dots + Z_n^2$  은 자유도  $n$ 인 카이제곱분포를 이룬다.

$$V = Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \chi^2(n)$$

- ❖ 정규분포  $X \sim N(\mu, \sigma^2)$ 에 대하여 표준화 확률변수의 제곱은 자유도 1인 카이제곱분포를 이룬다.

$$Z^2 = \left( \frac{X - \mu}{\sigma} \right)^2 \sim \chi^2(1)$$

- ❖ 독립인 카이제곱 확률변수  $X \sim \chi^2(r_1)$ ,  $Y \sim \chi^2(r_2)$ 에 대하여 두 확률변수의  $X + Y$ 는 각각의 자유도의 합을 자유도로 가지는 카이제곱분포를 이룬다. 즉, 다음이 성립한다.

$$X + Y \sim \chi^2(r_1 + r_2)$$

# 카이제곱분포와 표준정규분포

## 예제 7-11

$X$ 가 자유도 4인 카이제곱 분포를 따른다고 한다.

(a)  $P(X > \chi_r^2) = 0.05$ 를 만족하는 임계값  $\chi_l^2$ 을 구하라.

(b)  $P(X < \chi_l^2) = 0.05$ 를 만족하는 임계값  $\chi_r^2$ 을 구하라.

풀이

(a)  $P(X > \chi_l^2) = 0.05$  이므로 카이제곱분포표에서  $d.f = 4$ 이고  $\alpha = 0.05$ 가 만나는 위치의 수 9.49를 택하면  $\chi_l^2 = 9.49$  이다.

(b)  $P(X < \chi_r^2) = 0.05$  이므로  $P(X > \chi_l^2) = 0.05$  이고  $d.f = 4$ 이고  $\alpha = 0.95$ 가 만나는 위치의 수 0.71을 택하면  $\chi_r^2 = 0.71$ 이다.

# $t$ -분포

**$t$ -분포**  $t$ -distribution :

독립인  $Z \sim N(0, 1)$ 과  $V \sim \chi^2(r)$ 에 대하여 확률변수  $T = Z / \sqrt{V / r}$ 의 확률분포를 자유도  $r$ 인  $t$ -분포라하고  $T \sim t(r)$ 로 나타낸다.

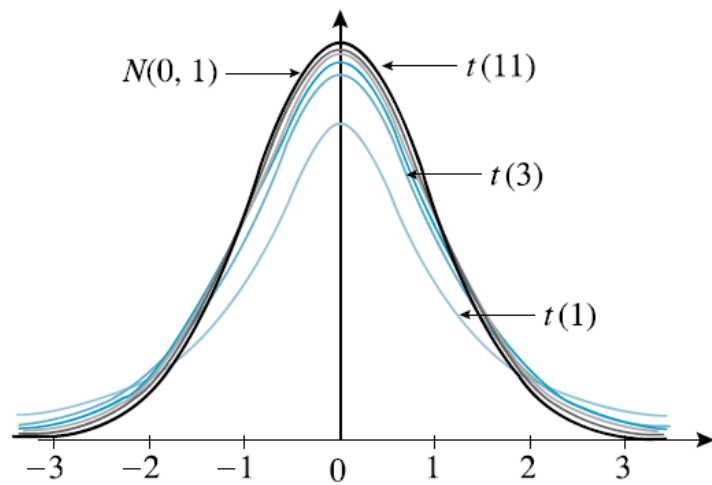
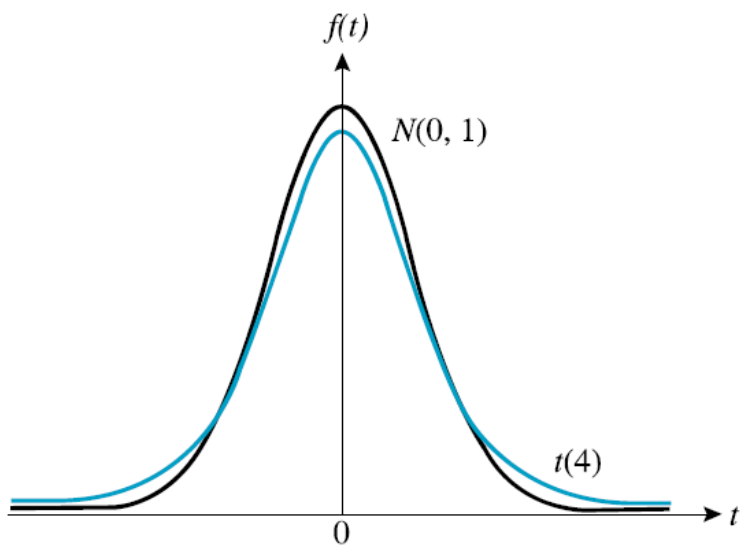
$$f(t) = \frac{\Gamma[(r+1)/2]}{\Gamma(r/2)\sqrt{\pi r}} \left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}}, \quad -\infty < t < \infty$$

① 평균 :  $\mu = 0$

② 분산 :  $\sigma^2 = \frac{r}{r-2}, \quad r > 2$

# $t$ - 분포와 표준정규분포

- ① 분포곡선은  $t = 0$ 에서 최댓값을 갖고 대칭이다.
- ② 분포곡선은 표준정규분포와 같은 종 모양이다.
- ③ [그림] (a)와 같이  $t$  - 분포의 꼬리부분이 표준정규분포보다 약간 두텁다.
- ④ [그림] (b)와 같이 자유도  $r$ 이 증가하면( $r \geq 30$ )  $t$  - 분포는 표준정규분포에 근접한다.

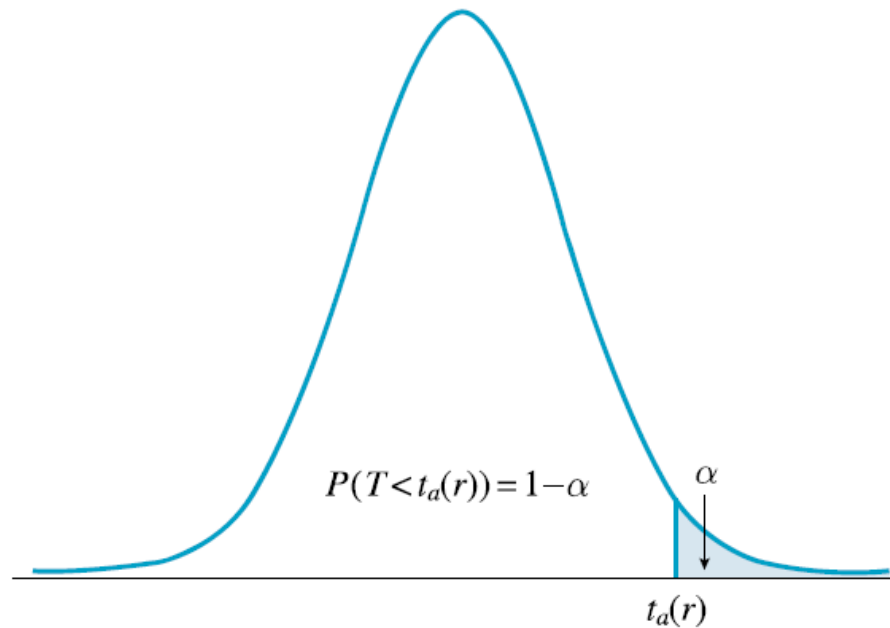




# $t$ - 분포의 백분위수

자유도  $r$ 인  $t$  - 분포에서  $P(T > t_a(r)) = \alpha$ 를 만족하는  $100(1 - \alpha)\%$  백분위수를  $t_a(r)$ 로 나타내며, 다음과 같다.

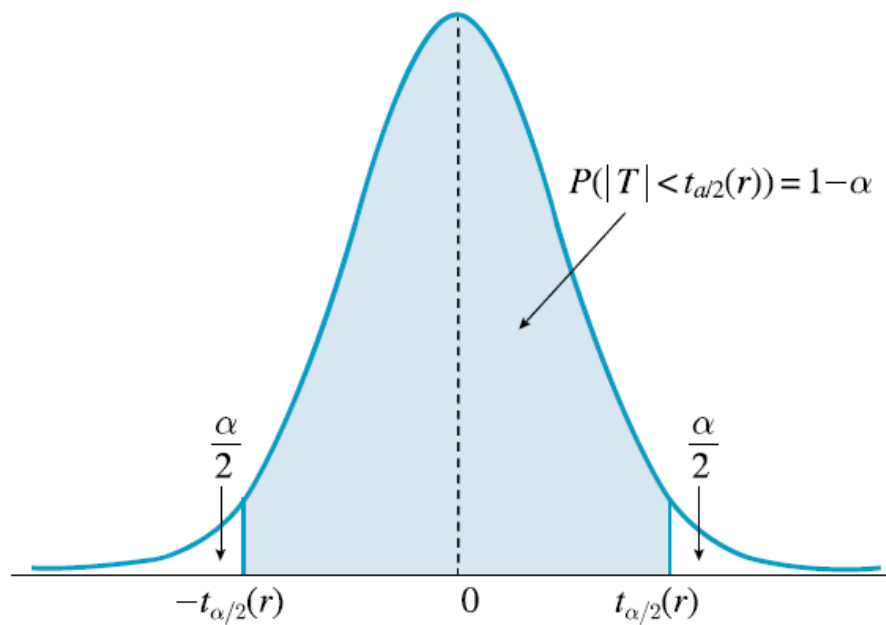
$100\alpha\%$  백분위수 :  $t_{1-\alpha}(r) = -t_\alpha(r)$



# $t$ - 분포의 중심확률

①  $P(T > t_{\alpha}(r)) = P(T < -t_{\alpha}(r)) = \alpha$

②  $P(|T| < t_{\alpha/2}(r)) = 1 - \alpha$



# 100(1 - $\alpha$ )% 백분위수 구하는 방법

❖ 백분위수  $t_\alpha(r)$ 은  $t$ -분포표에서 [그림]과 같이 구할 수 있다.

예를 들어,  $X \sim t(4)$ 에 대하여  $P(X > t_{0.05}(4)) = 0.05$ 를 만족하는 95% 백분위수  $t_{0.05}(4)$ 는 다음과 같이 구한다.

$t_{0.05}(4) = 2.132, P(T > 2.132) = 0.05$

꼬리확률							
$\alpha$		0.25	0.10	0.050	0.025	0.010	0.005
자유도	d.f						
	1	1.000	3.078	6.314	12.706	31.821	63.675
	2	0.816	1.886	2.920	4.303	6.965	9.925
	3	0.765	1.638	2.353	3.182	4.541	5.841
	4	0.741	1.533	2.132	2.776	3.747	4.604
	5	0.727	1.476	2.015	2.571	3.365	4.032
	6	0.718	1.440	1.943	2.447	3.143	3.707

# 100(1 - $\alpha$ )% 백분위수 구하는 방법

## 예제 7-12

확률변수  $T$ 가 자유도 5인  $t$ -분포를 따른다.

- (a)  $P(T > t_0) = 0.01$ 을 만족하는 임계값  $t_0$ 를 구하라.
- (b)  $P(T < t_0) = 0.05$ 를 만족하는 임계값  $t_0$ 를 구하라.
- (c)  $P(-t_0 < T < t_0) = 0.99$ 를 만족하는 임계값  $t_0$ 를 구하라.

풀이

- (a)  $P(T > t_0) = 0.01$ 을 만족하는 임계값은  $t_0 = t_{0.01}(5) = 3.365$ 이다.
- (b)  $P(T < t_0) = P(T > -t_0) = 0.05$ 이므로  $-t_0 = t_{0.05}(5) = 2.015$ 이고  $t_0 = -2.015$ 이다.
- (c)  $P(-t_0 < T < t_0) = 0.99$ 이므로  $t_0 = t_{0.005}(5) = 4.032$ 이고  $t_0 = 4.032$ 이다.

# F - 분포

**F-분포** *F-distribution* :

독립인  $U \sim \chi^2(m)$ 과  $V \sim \chi^2(n)$ 에 대하여 확률변수  $F = (U/m)/(V/n)$ 의 확률분포를 분자와 분모의 자유도가  $m$ 과  $n$ 인  $F$ -분포라 하고,  $F \sim F(m, n)$ 으로 나타낸다.

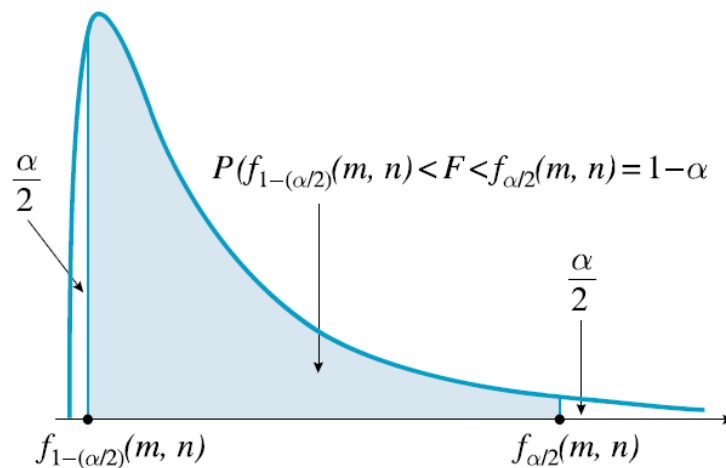
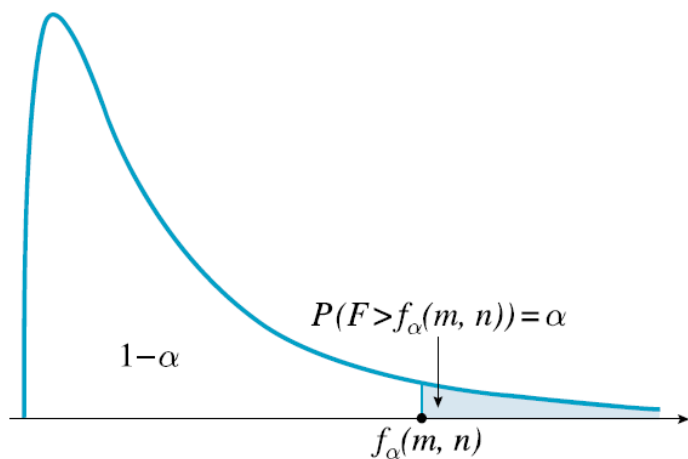
$$f(x) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{(m/2)-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, \quad x > 0$$

① 평균 :  $\mu = \frac{n}{n-2}, \quad n \geq 3$

② 분산 :  $\sigma^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad n \geq 5$

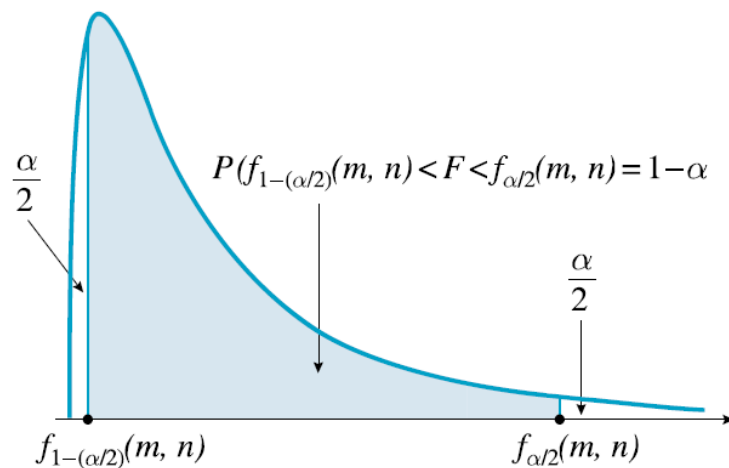
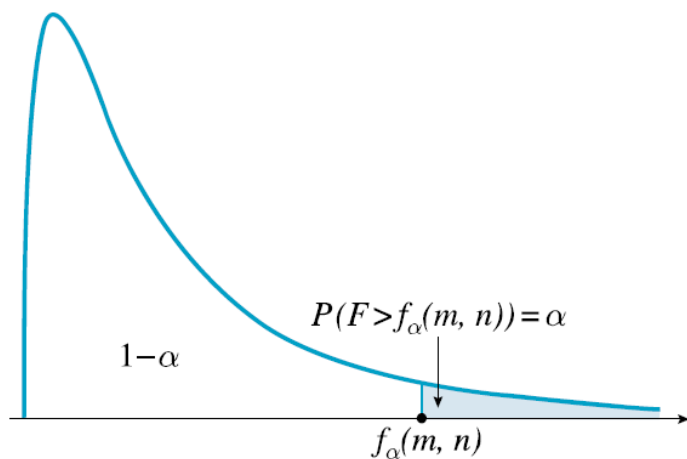
# F- 분포의 특성

- ① [그림]과 같이 왼쪽으로 치우치고 오른쪽으로 긴 꼬리를 갖는 분포, 즉 양의 왜도를 갖는 분포를 이룬다.
- ② 분모의 자유도가 커질수록  $\mu \doteq 1, \sigma^2 \doteq 2/m$ 이다.
- ③ 자유도  $m, n$ 이 커질수록  $\mu = 1$ 을 중심으로 좌유대칭인 정규분포에 근사한다.



# F - 분포의 꼬리확률

- ① 분자, 분모의 자유도가 각각  $m$ 과  $n$ 인  $F$  - 분포에 대하여 오른쪽 꼬리확률  $\alpha = P(X > x_\alpha)$ 를 만족하는  $100(1 - \alpha)\%$  백분위수를  $x_\alpha = f_\alpha(m, n)$ 으로 나타낸다.
- ② 중심확률이  $1 - \alpha$ 인 두 임계점은  $f_{1-\alpha/2}(m, n)$ ,  $f_{\alpha/2}(m, n)$  이다.



# F - 분포의 특성

- ❖  $F \sim F(m, n)$ 이면  $F = (U/m)/(V/n)$ 이므로  $1/F = (V/n)/(U/m)$  이고, 따라서 다음이 성립한다.

$$1/F \sim F(n, m)$$

- ❖ 왼쪽 꼬리확률이  $\alpha$ 인 임계점  $f_{1-\alpha}(m, n)$ 은 다음과 같다.

$$f_{1-\alpha}(m, n) = \frac{1}{f_{\alpha}(n, m)}$$

즉, 왼쪽 꼬리확률이  $\alpha$ 에 대하여 다음이 성립한다.

$$P(F < f_{1-\alpha}(m, n)) = P\left(F < \frac{1}{f_{\alpha}(n, m)}\right) = \alpha$$

- ❖ 왼쪽과 오른쪽 꼬리확률이 각각  $\alpha/2$ 인 중심확률  $1 - \alpha$ 에 대해 다음이 성립한다.

$$P(f_{1-(\alpha/2)}(m, n) < F < f_{\alpha/2}(m, n)) = 1 - \alpha$$



# 100(1 - $\alpha$ )% 백분위수 구하는 방법

❖ 백분위수  $f_\alpha(m, n)$ 은  $F$  - 분포표에서 [그림]과 같이 구할 수 있다.

예를 들어,  $F \sim F(4, 5)$ 에 대하여  $P(F > F_{0.05}(4, 5)) = 0.05$  를 만족하는 95% 백분위수  $f_{0.05}(4, 5)$ 는 다음과 같이 구한다.

꼬리확률  $f_{0.05}(4, 5) = 5.19, P(F > 5.19) = 0.05$

분모의 자유도	$\alpha$	분자의 자유도								
		1	2	3	4	5	6	7	8	9
4	0.010	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	0.025	12.21	10.65	9.98	9.60	9.36	9.20	7.07	8.98	8.90
	0.005	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	0.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
	0.010	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
5	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	0.005	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	0.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24

# 100(1 - $\alpha$ )% 백분위수 구하는 방법

## 예제 7-13

$F$ -분포표를 이용하여 분자와 분모의 자유도가 각각 5와 4인  $F$ -분포에서 중심확률이 0.95인 두 임계점  $f_l$ 과  $f_r$ 을 구하라.

풀이

중심확률이 0.95이므로 왼쪽과 오른쪽 꼬리확률은 동일하게 0.025이다. 왼쪽 꼬리확률이 0.025인 임계점은 다음과 같다.

$$f_l = f_{0.975}(5, 4) = \frac{1}{f_{0.025}(4, 5)} = \frac{1}{9.36} = 0.1068$$

그리고 오른쪽 꼬리확률이 0.025인 임계점은 다음과 같다.

$$f_r = f_{0.025}(5, 4) = 7.39$$

# 로그정규분포

**로그정규분포** log-normal distribution :

정규확률변수  $Y \sim N(m, s^2)$  에 대하여  $X = e^Y$ 의 확률분포이며,  
 $X \sim \text{Log } N(m, s^2)$ 으로 나타낸다.

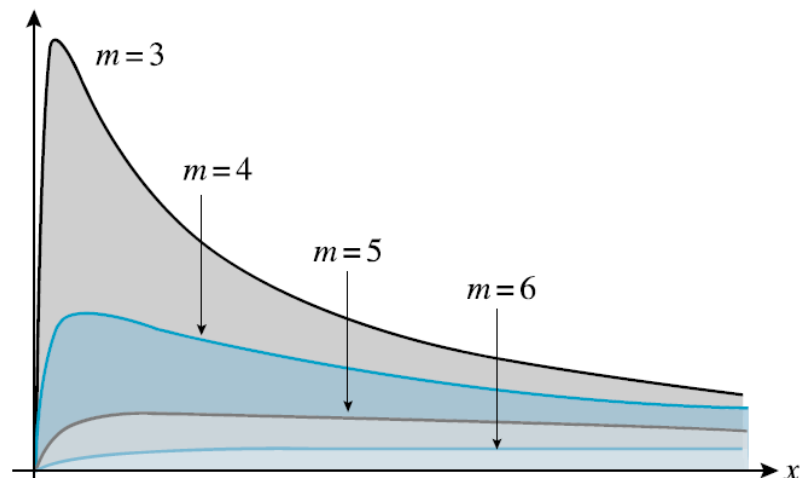
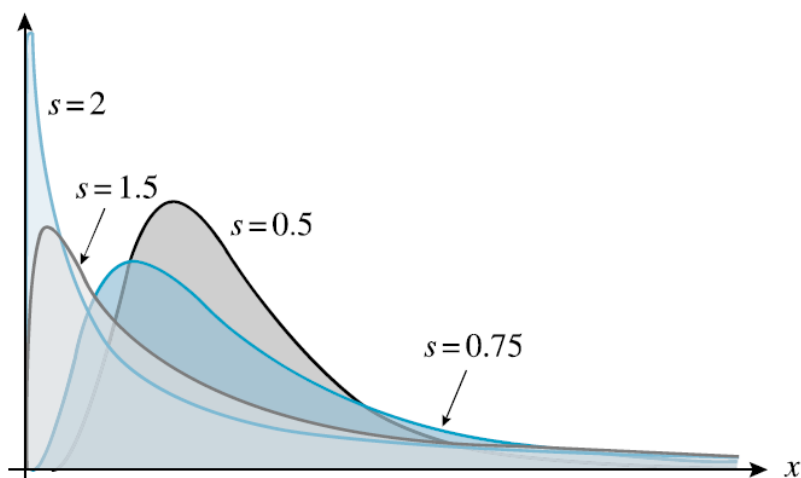
$$f(x) = \frac{1}{\sqrt{2\pi}sx} \exp\left(-\frac{(\ln x - m)^2}{2s^2}\right), \quad x > 0$$

① 평균 :  $\mu = \exp\left(m + \frac{s^2}{2}\right)$

② 분산 :  $\sigma^2 = (e^{s^2} - 1)\exp(2m + s^2)$

# 로그정규분포의 특성

- ❖ [그림]과 같이 왼쪽으로 치우치고 오른쪽으로 긴 꼬리를 갖는 분포, 즉 양의 왜도를 갖는 분포를 이룬다.



# 로그정규분포의 분포함수

$$F(x) = \Phi\left(\frac{\ln x - m}{s}\right) = P\left(Z < \frac{\ln x - m}{s}\right), \quad x > 0$$

# 로그정규분포의 확률계산

$$P(a < X \leq b) = \Phi\left(\frac{\ln b - m}{s}\right) - \Phi\left(\frac{\ln a - m}{s}\right)$$

## 100(1 - a)% 백분위수

$$F(x_\alpha) = P\left(Z < \frac{(\ln x_\alpha) - \mu}{\sigma}\right) = P(Z < z_\alpha) = 1 - \alpha; \quad \frac{(\ln x_\alpha) - \mu}{\sigma} = z_\alpha; \quad x_\alpha = e^{\mu + \sigma z_\alpha}$$

# 로그정규분포의 특성

## 예제 7-14

화학공장에서 배출되는 오염 농도가  $m = 4$ ,  $s = 1.5$ 인 로그정규분포를 따른다고 한다.

- (a) 오염 농도가 10과 20 사이일 확률을 구하라.
- (b) 5% 백분위수  $x_{0.95}$ 를 구하라.

풀이

- (a) 오염농도를  $X$ 라 하면,  $X \sim \text{LogN}(4, 1.5^2)$ 이므로 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(10 < X \leq 20) &= \Phi\left(\frac{\ln 20 - 4}{1.5}\right) - \Phi\left(\frac{\ln 10 - 4}{1.5}\right) = \Phi(-0.67) - \Phi(-1.13) \\ &= \Phi(1.13) - \Phi(0.67) = 0.8708 - 0.7486 \\ &= 0.1222 \end{aligned}$$



# 로그정규분포의 특성

(b)  $P(Z \leq -1.645) = 0.05$ 이므로  $P(X \leq x_{0.05}) = 0.05$ 를 만족하는  $x_{0.05}$ 는 다음과 같이 구한다.

$$P(X \leq x_{0.95}) = \Phi\left(\frac{\ln x_{0.95} - 4}{1.5}\right) = \Phi(-1.645) = 0.05; \quad \frac{\ln x_{0.95} - 4}{1.5} = -1.645$$

$$\ln x_{0.95} = 4 + 1.5 \times (-1.645) = 1.5325$$

$$x_{0.95} = e^{1.5325} = 4.6297$$

# 이변량정규분포

**이변량정규분포** bivariate normal distribution :

양의 상수  $\sigma_X, \sigma_Y$ 와  $-\infty < \mu_X, \mu_Y < \infty, -1 < \rho < 1$ 에 대하여 다음 결합밀도함수를 확률분포. 이때  $(X, Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ 으로 나타낸다.

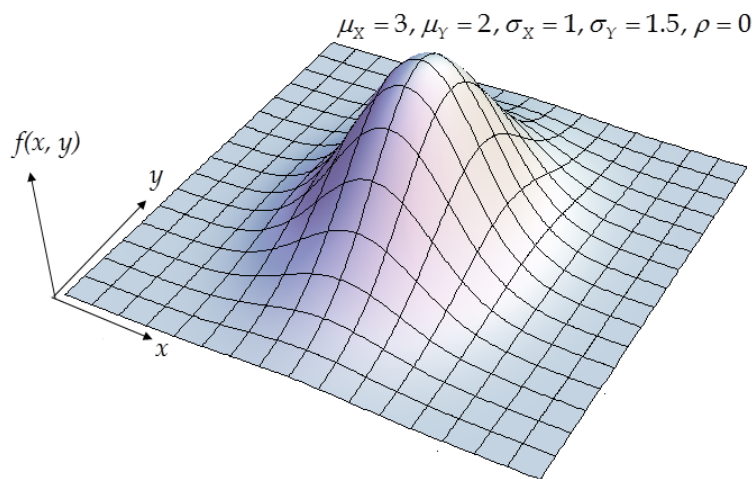
$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{Q}{2}\right), \quad -\infty < x, y < \infty$$

$$, Q = \frac{1}{1-\rho^2} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right]$$

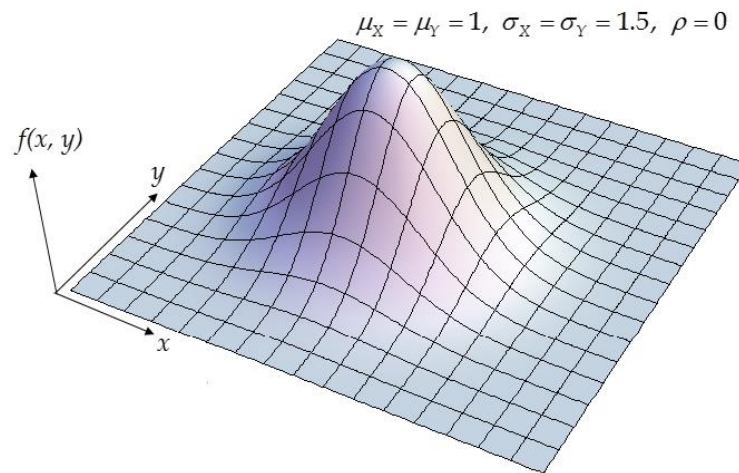
# 이변량 정규분포의 특성

❖ 상수  $\rho$ 는 두 확률변수  $X$ 와  $Y$ 의 상관계수이다.

❖  $\rho = 0$ 인 경우 : 
$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp \left[ -\left( \frac{x - \mu_X}{\sigma_X} \right)^2 - \left( \frac{y - \mu_Y}{\sigma_Y} \right)^2 \right], \quad -\infty < x, y < \infty$$



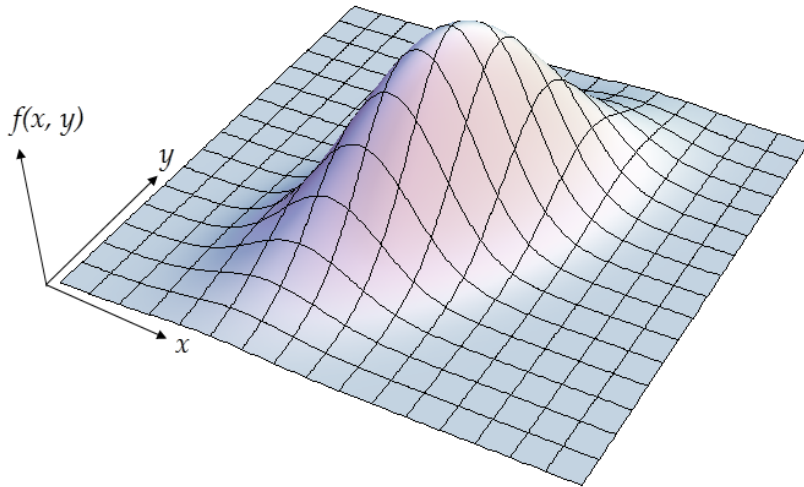
$X$ 와  $Y$ 가 독립인 경우



$X$ 와  $Y$ 가 i.i.d인 경우

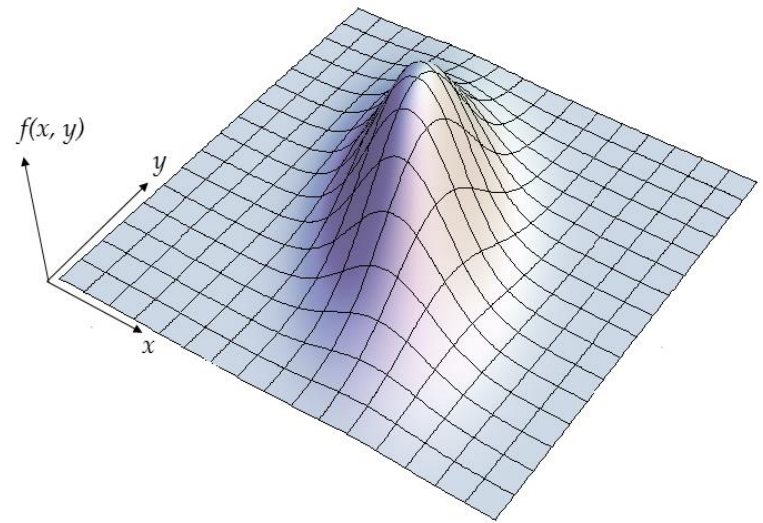
# 이변량 정규분포의 특성

$$\mu_X = 3, \mu_Y = 2, \sigma_X = 1, \sigma_Y = 1.5, \rho = 0.6$$



$\rho > 0$ 인 경우

$$\mu_X = 3, \mu_Y = 2, \sigma_X = 1, \sigma_Y = 1.5, \rho = -0.6$$



$\rho < 0$ 인 경우

# 주변확률밀도함수

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right) \sim N(\mu_X, \sigma_X^2), \quad -\infty < x < \infty$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right) \sim N(\mu_Y, \sigma_Y^2), \quad -\infty < y < \infty$$

# 조건부확률밀도함수

$$f(x|y) = \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} \exp\left(-\frac{(x-b_X)^2}{2\sigma_X^2(1-\rho^2)}\right), \quad -\infty < x < \infty$$

$$f(y|x) = \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{(y-b_Y)^2}{2\sigma_Y^2(1-\rho^2)}\right), \quad -\infty < y < \infty$$

**조건부 평균 :**  $b_X = \mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \quad b_Y = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$

# 이변량정규분포

## 예제 7-15

확률변수  $X$ 와  $Y$ 가  $\mu_X = \mu_Y = 0$ ,  $\sigma_X = \sigma_Y = 1$ ,  $\rho = 0.8$ 인 이변량 정규분포를 따른다고 한다.

- (a)  $X$ 의 주변확률밀도함수를 구하라.
- (b)  $X = 1$ 인 조건 아래서  $Y$ 의 조건부 확률밀도함수를 구하라.

풀이

- (a)  $(X, Y) \sim N(0, 0, 1, 1, 0.8)$ 이므로 구하고자 하는  $X$ 의 주변확률밀도함수는 다음과 같다.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$$

# 이변량정규분포

(b)  $\mu_X = \mu_Y = 0$ ,  $\sigma_X = \sigma_Y = 1$ ,  $\rho = 0.8$  이고  $x = 1$  이 주어졌으므로

$$b_Y = (0.8) \quad x = 0.8$$

이다. 그러므로  $Y$ 의 조건부 확률밀도함수는 다음과 같다.

$$\begin{aligned} f(y | x = 1) &= \frac{1}{\sqrt{2\pi} \sqrt{1 - 0.8^2}} \exp\left(-\frac{(y - 0.8)^2}{2 \times (1 - 0.8^2)}\right) \\ &= \frac{1}{\sqrt{(0.72)\pi}} \exp\left(-\frac{(y - 0.8)^2}{0.72}\right), \quad -\infty < y < \infty \end{aligned}$$



# Q&A

