

6강

# 확률분포와 표본분포1

통계·데이터과학과 이기재 교수

## 목차

- 1 이산형 확률분포
- 2 연속형 확률분포

01

# 이산형 확률분포

# 이항분포 적용 사례

- ▶ 동전을 5번 던질 때 앞면의 수
- ▶ 공장에서 생산된 제품 100개를 검사할 때 발견되는 불량품의 수
- ▶ 유권자 50명 중 특정 후보를 지지하는 사람 수

➔ 실험(조사)의 결과가 어떻게 될지 사전에 알 수 없지만 모든 가능한 결과가 두 가지({앞면, 뒷면}, {불량, 양호}, {찬성, 반대} 등)이고 반복됨

# 베르누이시행, 시행, 독립시행

## ■ 베르누이시행

- 각 실험의 결과가 두 가지만 가능한 시행  
(예 : 앞면과 뒷면, 성공과 실패, 합격과 불합격 등)

## ■ 베르누이시행의 예

- 제품 검사로 불량품과 양호품으로 구분하는 경우
- 유권자에게 특정 후보에 대한 지지 여부를 묻는 경우

## ■ 시행

- 같은 실험을 반복할 때 각각의 실험

## ■ 독립시행

- 각 시행의 결과들이 서로 독립인 경우

# 이항분포(binomial distribution)

- ▶ 각 시행에서 성공률이  $p$ 인 베르누이시행을  $n$ 번 독립시행
- ▶  $X = "n\text{번의 베르누이 독립시행에서 얻은 총 성공 횟수}"$

$\Rightarrow$

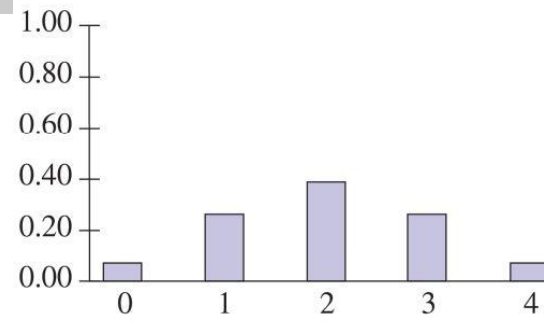
$$X \sim B(n, p)$$

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x=0, 1, 2, \dots, n$$

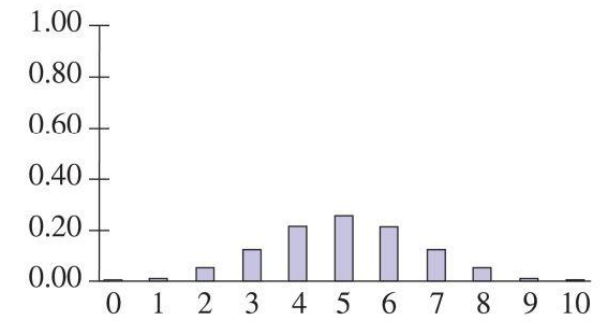
$$E(X) = np, \quad \text{Var}(X) = np(1-p)$$

# 이항분포(binomial distribution)

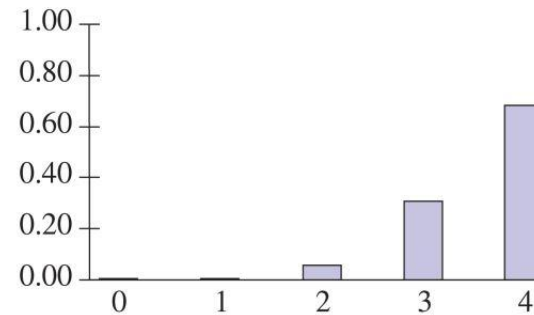
- ▶ 각 시행의 성공률  $p$ 와 시행횟수  $n$ 에 따라 이항확률변수  $X$ 의 확률분포가 결정됨



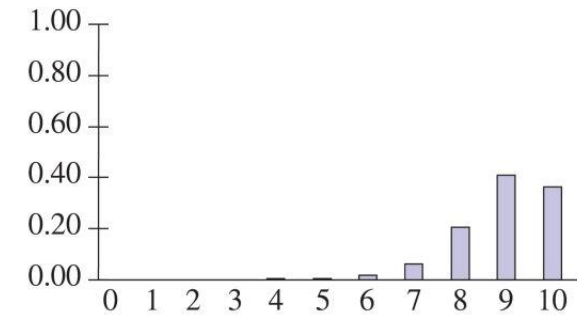
(a)  $n=4, p=0.5$



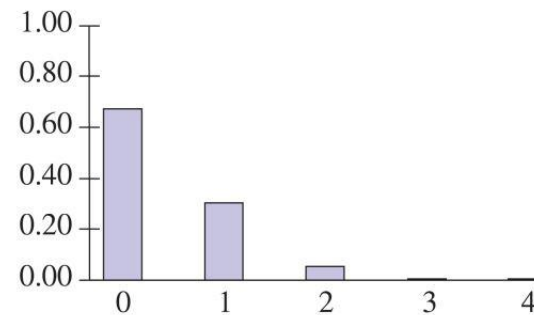
(b)  $n=10, p=0.5$



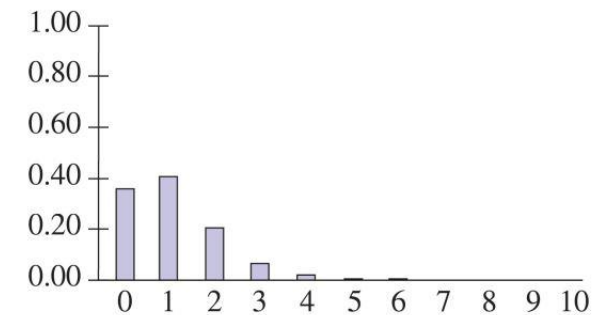
(c)  $n=4, p=0.9$



(d)  $n=10, p=0.9$



(e)  $n=4, p=0.1$



(f)  $n=10, p=0.1$

# 이항분포 사례

- ▶ 공정한 동전을 5번 던지는 실험
  - $X = 5$ 번 던져서 나온 앞면의 수
  - $n = 5, p = 0.5$
  - $X \sim B(5, 0.5)$
  - $P(X = 0) = \binom{5}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = \frac{1}{32}$



# 이항분포 확률계산 1

- 어느 보험회사의 영업사원이 한 고객을 만날 때 그 사람이 보험에 가입하게 될 확률은 경험으로 볼 때 20%이다. 오늘 이 영업사원이 고객 10명을 만날 예정이다.
  - 세 명이 보험에 가입할 확률은?
  - 두 사람 이상이 보험에 가입할 확률은?
  - 10명 중 보험에 가입한 사람의 평균과 분산은?

## 이항분포 확률계산 2

- ▶ 당첨률이 30%인 복권을 8장 구입했다고 한다.
  - 4장이 당첨될 확률은?
  - 3장에서 7장이 당첨될 확률은?
  - 6장 이상이 당첨될 확률?

〈풀이〉  $n = 8, p = 0.3$ 인 이항분포에 대한 누적확률분포표 이용

$x$	0	1	2	3	4	5	6	7	8
$P(X \leq x)$	.058	.255	.552	.806	.942	.989	.999	1.00	1.00

# 초기하분포

- ▶ 전체  $N$ 개인 모집단이 '1' 또는 '0',  
'성공' 또는 '실패' 등으로 두 가지로 분류
- ▶ 확률변수  $X =$  "전체  $N$ 개 중 1이  $D$ 개, 0이  $N - D$ 개로  
구성된 유한모집단에서 크기  $n$ 인 랜덤표본을 뽑을 때  
1이 나오는 수"

## 초기하분포 예제

- ▶ 흰 공 3개와 검은 공 2개가 들어있는 주머니에서 임의로 2개를 꺼냈을 때  $X$ 를 검은 공의 개수인 확률변수로 정의하자.  $P(X = 0)$ 과  $P(X = 2)$ 를 구하시오

풀이

$$P(X = 0) = \frac{\binom{2}{0} \binom{3}{2}}{\binom{5}{2}} = \frac{3}{10} = 0.3$$

$$P(X = 2) = \frac{\binom{2}{2} \binom{3}{0}}{\binom{5}{2}} = \frac{1}{10} = 0.1$$

## 초기하분포의 평균과 분산

- ▶  $X$  = "전체  $N$ 개 중 1이  $D$ 개, 0이  $N - D$ 개로 구성된 유한모집단에서 크기  $n$ 인 랜덤표본을 뽑을 때 1이 나오는 수"
- ▶ 평균 :  $E(X) = np$ , 단  $p = \frac{D}{N}$
- ▶ 분산 :  $Var(X) = np(1 - p) \cdot \frac{N-n}{N-1}$

## 포아송(Poisson)분포 적용 예

- ▶ 고속도로 상에서 하루동안 발생하는 교통사고에 의한 사망자 수
  - ▶ 어느 집에 한 시간 동안에 걸려오는 전화통화 수
  - ▶ 야구 경기 한 게임에서 나오는 홈런 수
  - ▶ 1주일간 어떤 동사무소에서 접수되는 사망신고 수
- ⇒ 발생 확률이 아주 작은 희귀한 사건의 수에 대한 확률분포에 이용

## 포아송분포를 적용하기 위한 가정

- 독립성

- 서로 다른 단위에서 출현하는 횟수는 서로 독립

- 비집락성

- 극히 작은 단위에서 둘 이상이 일어날 확률은 매우 작음

- 비례성

- 단위시간이나 공간에서 성공의 평균 출현횟수는 일정함

# 포아송분포(Poisson distribution)

➤ 단위당 평균 발생률이  $m$ 인 어떤 현상에 대해서

- $X$  = 단위당 발생횟수

- $X \sim \text{Poisson}(m)$

$$P[X = x] = \frac{e^{-m} m^x}{x!}, \quad x = 0, 1, 2, \dots, \quad e = 2.718281 \dots$$

- $E(X) = m, \quad \text{Var}(X) = m$



## 포아송분포 사례

- ▶ 은행에서 하루 평균 6건의 불량수표를 받게 된다고 할 때, 어떤 특정한 날에 불량수표를 4번 받을 확률은?

풀이

$$m = 6, \quad x = 4$$
$$P(X=4) = \frac{6^4 e^{-6}}{4!} = 0.134$$

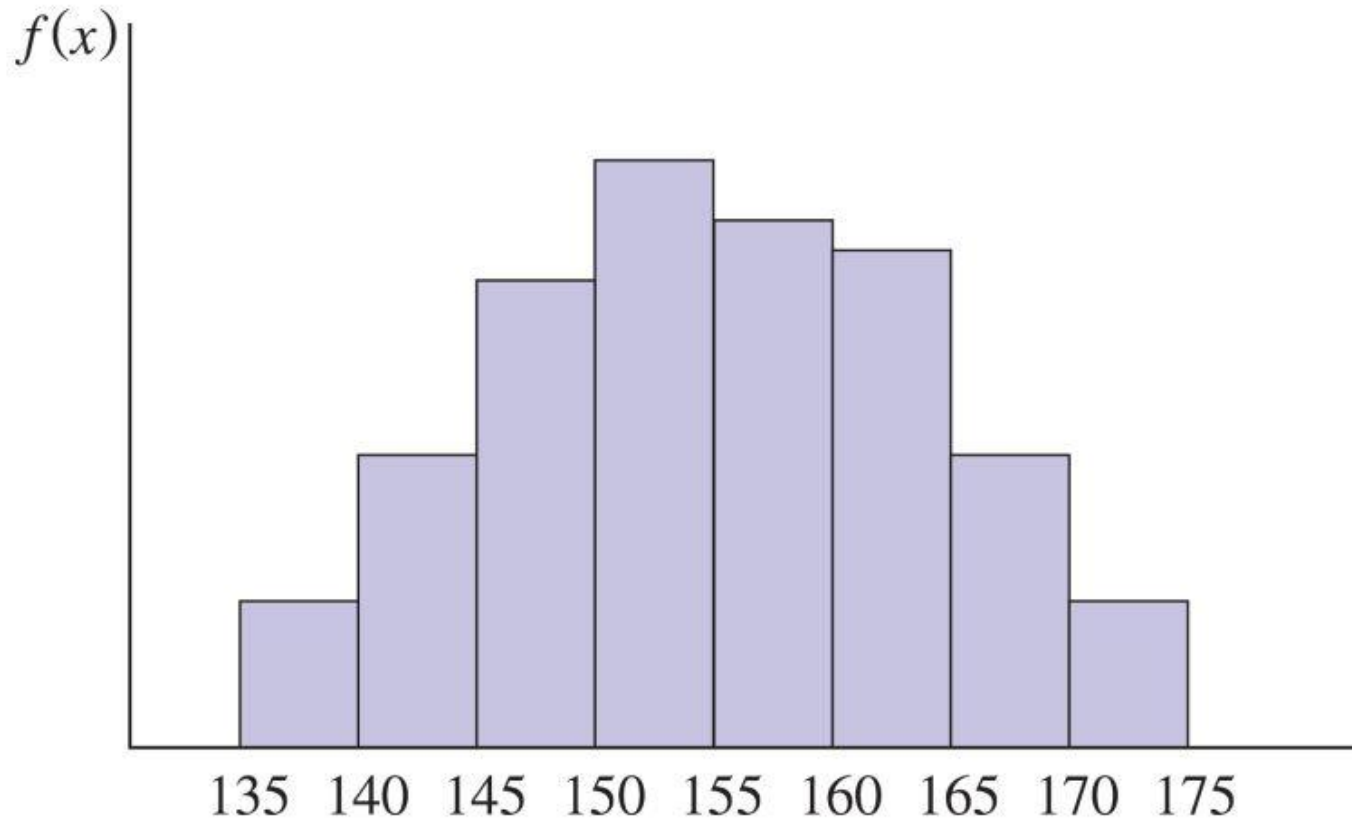
02

# 연속형 확률분포

## 중1 남학생 200명의 신장 도수분포표

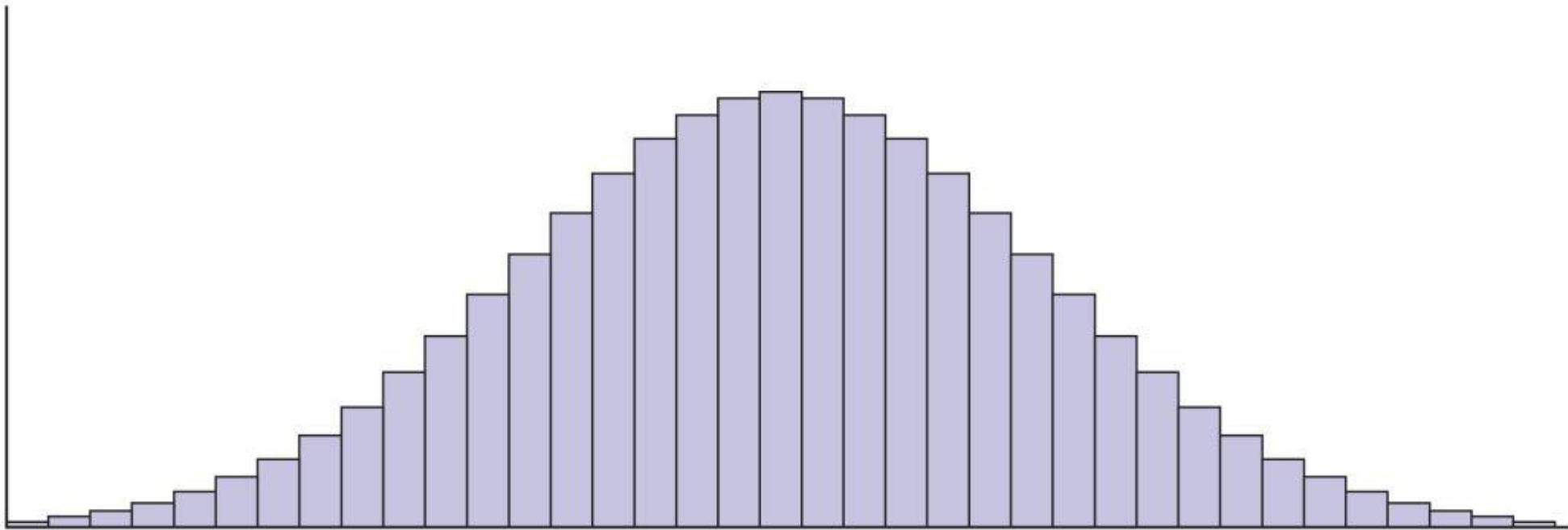
키(cm)	도수	상대도수
135~140	8	0.04
140~145	20	0.10
145~150	32	0.16
150~155	40	0.20
155~160	36	0.18
160~165	34	0.17
165~170	20	0.10
170~175	10	0.05
계	200	1.00

# 중1 남학생 200명의 신장의 분포



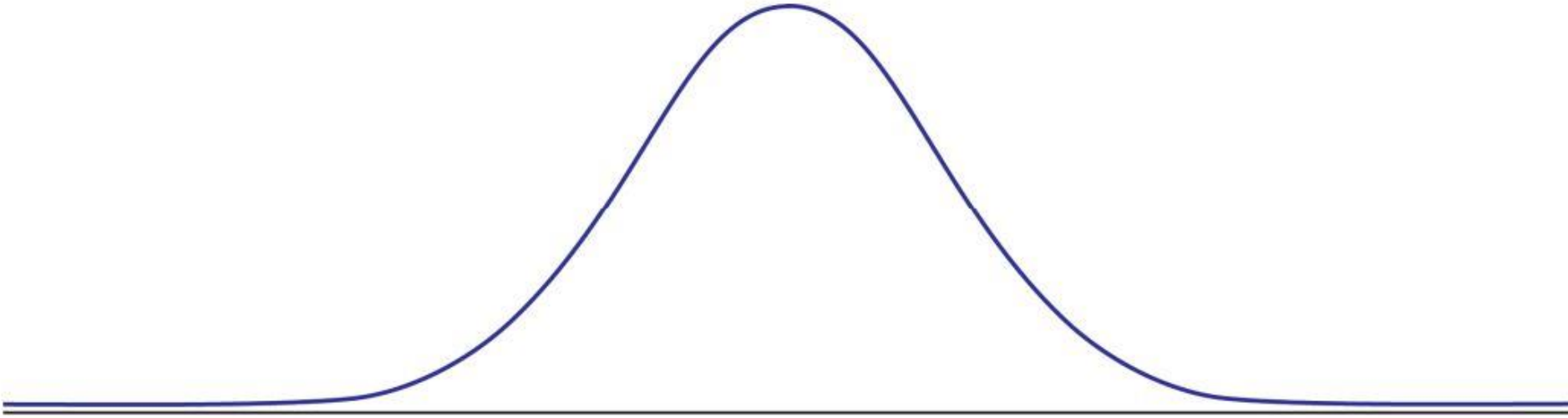
■ 그림 4-2 ■ 중학교 1학년 남학생 200명의 키 히스토그램

# 중1 남학생 3000명의 신장의 분포



■ 그림 4-3 ■ 중학교 1학년 남학생 3,000명의 키 히스토그램

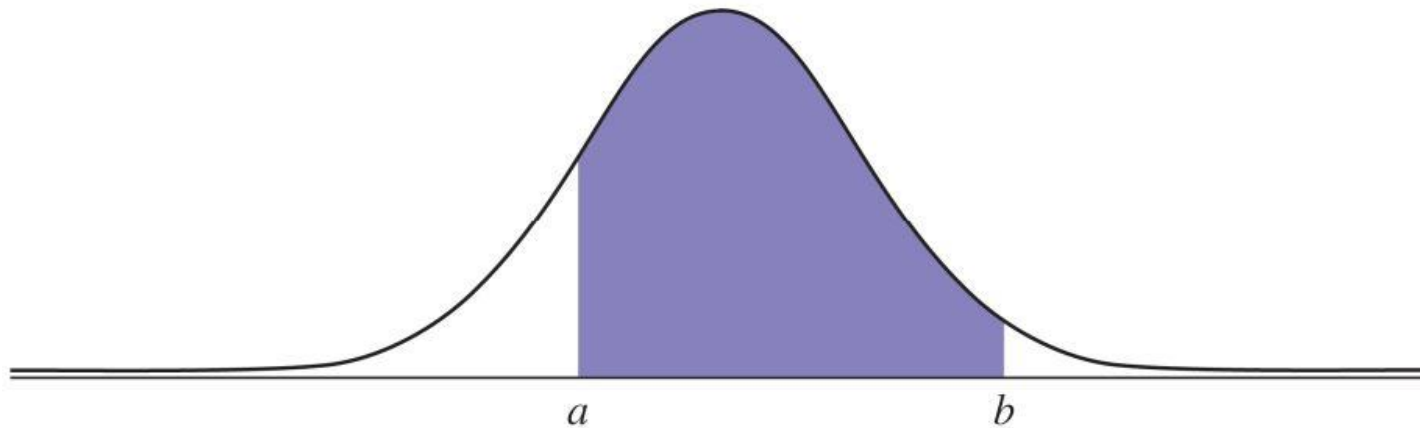
# 아주 많은 중1 남학생의 신장의 분포



■ 그림 4-4 ■ 중학교 1학년 남학생이 아주 많은 경우의 키 히스토그램

# 확률밀도함수 $f(x)$ 의 성질

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) = 1$
- $P(a < X \leq b) = \int_a^b f(x)dx$



# 정규분포(Normal distribution)

▶ "확률변수  $X$ 는 평균  $\mu$ , 표준편차  $\sigma$ 인 정규분포를 따른다."

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

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right], -\infty < x < \infty$$

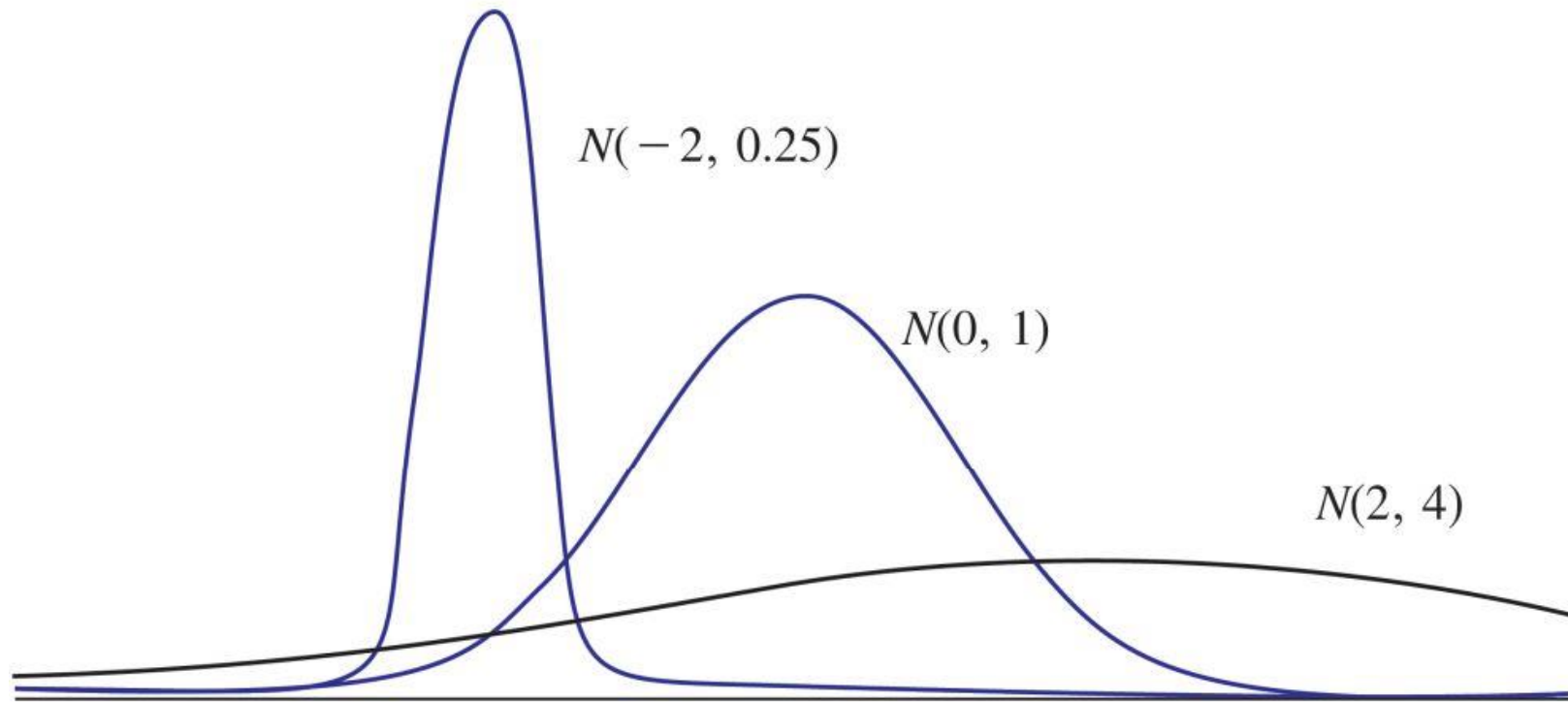
▶ " $\mu$ 와  $\sigma$ 의 값에 의해서 정규분포 결정"

- 종 모양이고, 평균  $\mu$ 에 관해 서로 대칭
- 드 모와브르(A. de Moivre(1667-1754)),  
가우스(C. F. Gauss(1777-1855))



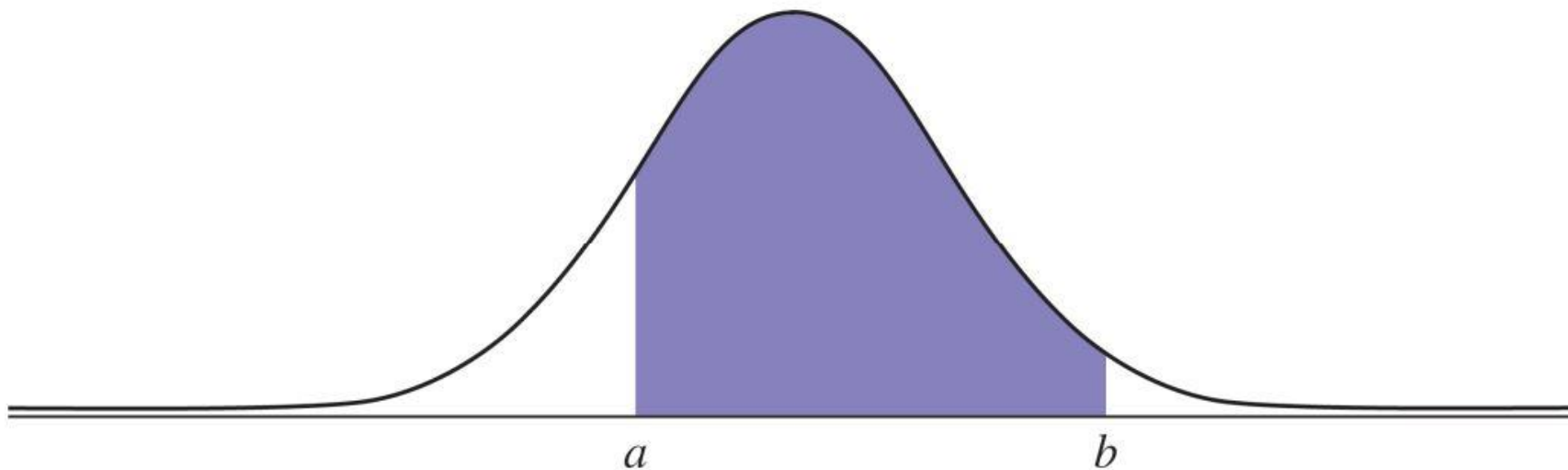
# 세 가지 정규분포

▶ 세 정규분포  $N(-2, 0.5^2)$ ,  $N(0, 1)$ ,  $N(2, 2^2)$ 의 그림



정규분포에서 확률  $P(a \leq X \leq b)$ 

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

| 그림 4-6 | 정규분포에서의 확률  $P(a \leq X \leq b)$

# 정규분포의 표준화

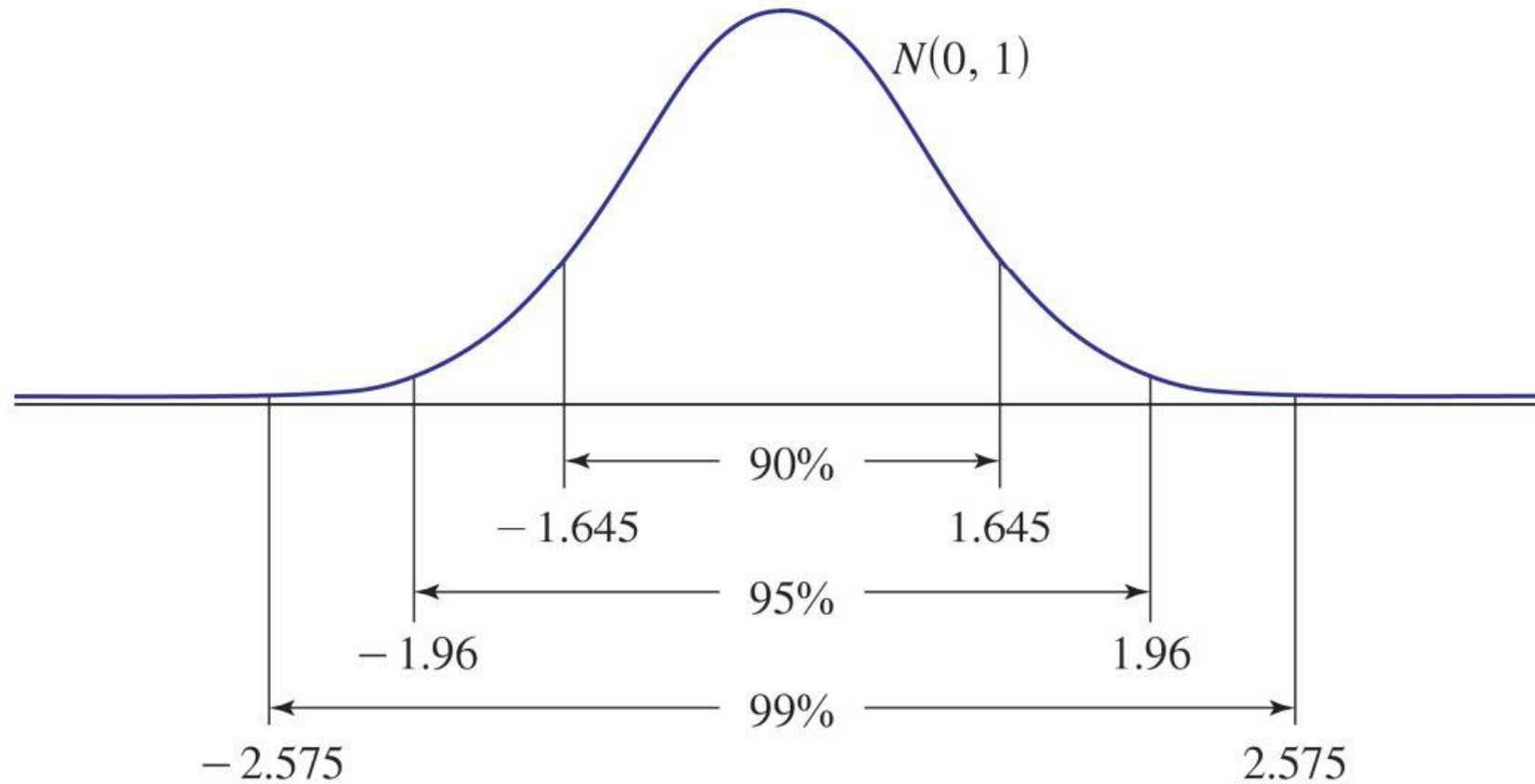
▶  $X \sim N(\mu, \sigma^2)$ 이면  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

▶  $X$ 가 평균이  $\mu$ , 분산이  $\sigma^2$ 인 정규분포를 따를 때

$$P(X < x) = P\left(\frac{X - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right) = P\left(Z < \frac{x - \mu}{\sigma}\right)$$

$$P(a < X < b) = P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right)$$

⇒ 표준정규분포표 이용



【그림 4-7】 표준정규분포에서 양쪽 끝을 제외한 확률이 90%, 95%, 99% 되는 값

# 표준 정규분포표

〈표 4-2〉 표준정규분포표

$z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5358
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8949	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

# 정규분포 확률 계산1

확률변수  $X$ 가 평균이 70, 표준편차가 10인 정규분포를 따를 때

(1)  $P(X < 94.3)$

(2)  $P(X > 57.7)$

풀이

$$(1) P(X < 94.3) = P\left(\frac{X - 70}{10} < \frac{94.3 - 70}{10}\right) = P(Z < 2.43) = 0.9925$$

$$(2) P(X > 57.7) = P\left(\frac{X - 70}{10} > \frac{57.7 - 70}{10}\right) = P(Z > -1.23) = 0.8907$$

## 정규분포 확률 계산2

집에서 회사까지 통근 시간  $X$ (분)은 정규분포  $N(40, 5^2)$ 를 따름.  
통근 시간이 50분 이상 걸릴 확률은?

풀이

$X$ : 집에서 회사까지 통근시간

$$X \sim N(40, 5^2)$$

$$P(X \geq 50) = ?$$



다음시간 안내

## 7강

# 확률분포와 표본분포 2