제3장 확률분포

확률변수

- **확률변수(Random variable)**
  - 일정한 확률을 가지고 발생하는 사건에 대해 값을 부여하는 변수
    - 동전을 던져 앞이면 **1** 뒤면이면 **2**를 부여할 때 두 개를 던져서 둘 다 앞면이 나올 확률: **P(*X*=2) = ¼**
      - ***X***는 앞면의 개수를 나타내는 확률변수

- **확률분포: 확률변수가 가질 수 있는 값과 그에 따른 확률의 집합**
  - 동전을 **2**개 던져 앞면이 나오는 개수의 확률분포

X	$P(X=x)$
0	¼
1	½
2	¼

이산확률분포

- **이산확률분포: 확률변수가 이산적인 값을 가지는 경우의 분포**
- **보통 특정한 확률 값을 함수(확률밀도함수, pdf)로 나타냄**
- **이산확률변수의 확률밀도함수의 성질**
  - **모든 x 값에 대하여 $f(x) \leq 0$**
  - **$\Sigma f(x) = 1$**

- **동전 2개를 던져 앞면이 나오는 갯수의 확률밀도 함수**

  - $$\boldsymbol{f(x) = \begin{cases} \frac{1}{4}, x = 0, 2 \\ \frac{1}{2}, x = 1 \end{cases}}$$

이산확률분포

- **이산확률변수의 평균**
  - $\mu = \sum x \cdot f(x)$

- **이산확률변수의 평균의 기대값**
  - $E(X) = \sum x \cdot f(x) = \mu$

- **동전 2개를 던져 앞면이 나오는 갯수의 기대값(평균)은?**
  - $E(X) = 0 \cdot (1/4) + 1 \cdot (1/2) + 2 \cdot (1/4) = 1 = \mu$

이산확률분포

- **이산확률 변수의 분산과 표준편차**
  - 분산 $\boldsymbol{\sigma^2} = \sum(\boldsymbol{x}-\boldsymbol{\mu})^2 \cdot \boldsymbol{f}(\boldsymbol{x})$
  - 표준편차 $\boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma^2}} = \sqrt{\text{분산}}$

- **이산확률변수의 분산의 기대값**
  - $\boldsymbol{Var}(\boldsymbol{X}) = \boldsymbol{E}(\boldsymbol{x}-\boldsymbol{\mu})^2 = \sum(\boldsymbol{x}-\boldsymbol{\mu})^2 \cdot \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{\sigma^2}$

- **분산의 기대값의 성질**
  - $Var\,(aX) = a^2\, Var\,(X)$
  - $Var\,(X+b) = Var\,(X)$, $Var\,(X-b) = Var\,(X)$
  - $Var\,(aX+b) = a^2\, Var\,(X)$

이항분포 (Binomial Distribution)

- **베르누이 시행: 서로 상반된 두 가지 결과만 나오는 시행**
  - **예) 동전던지기, 신생아의 성별, 생산품의 품질 검사**
- **베르누이 시행의 조건**
  - **확률변수 X와 같은 값이 0 또는 1이다. (1=성공, 0=실패)**
  - **각 시행에서 성공의 결과가 나올 확률($P(X=1)$)은 일정**
  - **여러 번에 걸친 베르누이 시행은 독립적이다.**

- **성공확률이 0.9인 베르누이 시행의 평균과 분산**

x	$f(x)$	$x \cdot f(x)$	$(x-\mu)^2 \cdot f(x)$
0	0.1	0	0.081
1	0.9	0.9	0.009
합계	1.0	0.9	0.09

이항분포 (Binomial Distribution)

- **베르누이 확률변수의 평균, 분산, 표준편차**
  - $\boldsymbol{\mu} = \pi$
  - $\boldsymbol{\sigma^2 = (0-\pi)^2 \cdot (1-\pi) + (1-\pi)^2 \cdot \pi = \pi(1-\pi)}$
  - $\boldsymbol{\sigma = \sqrt{\pi(1-\pi)}}$

- **이항분포의 확률함수**

  $\pi = 0.9$인 경우

X	경우의 수		$f(x)$
0	FF	${}_2C_0$	${}_2C_0(0.9)^0(0.1)^2 = 0.01$
1	FS SF	${}_2C_1$	${}_2C_1(0.9)^1(0.1)^1 = 0.18$
2	SS	${}_2C_2$	${}_2C_2(0.9)^2(0.1)^0 = 0.81$

이항분포 (Binomial Distribution)

- **이항분포 확률함수 (n번 시행 중 x번 성공할 확률)**
  - $f(x) = {}_nC_x(\pi)^x(1-\pi)^{n-x}$
  - **n은 총 시행횟수**
  - **x는 성공횟수**
  - **π는 성공할 확률**
  - **1-π는 실패할 확률**
  - ${}_nC_x = \frac{n!}{x!(n-x)!}$

엑셀 활용

- ❑ $Pr(X = x)$를 구할 때, =BINOMDIST (x, n, π, 0) 이항분포의 확률밀도함수(probability density function, PDF)는 0임에 유의.
- ❑ $Pr(X \leq x)$를 구할 때, =BINOMDIST (x, n, π, 1) 이항분포의 누적밀도함수(cumulative density function, CDF)는 1임에 유의

이항분포 (Binomial Distribution)

- **이항분포의 평균과 분산**

X	$f(x)$	$x \bullet f(x)$	$(x-\mu)^2 \bullet f(x)$
0	0.01	0	0.0324
1	0.18	0.18	0.1152
2	0.81	1.62	0.0324
합계	1.0	1.8	0.18

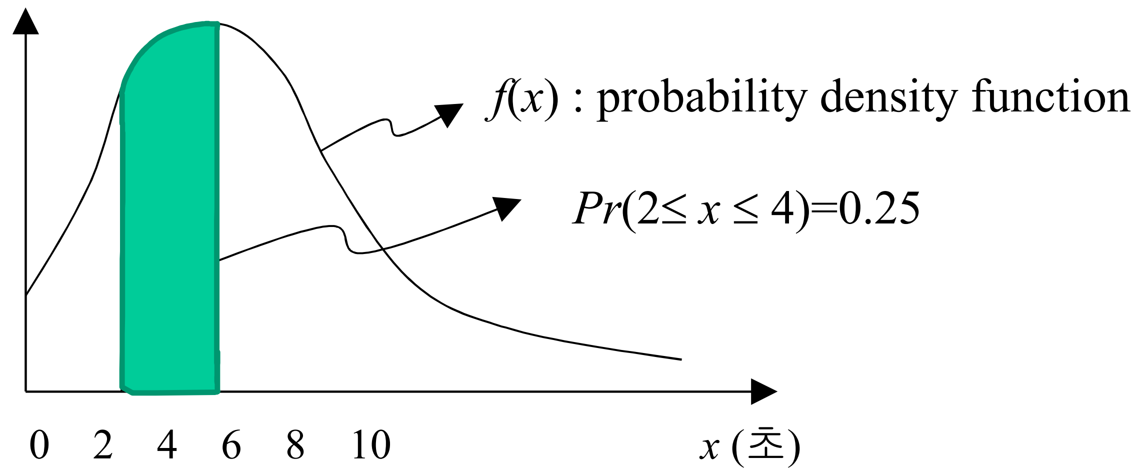
- **평균** $\boldsymbol{\mu = E(X) = n\pi}$
- **분산** $\boldsymbol{\sigma^2 = Var(X) = nE(X) = n\pi(1\text{-}\pi)}$

포아송분포

- 포아송분포는 이산확률분포의 대표
  - 발생 갯수에 관련된 분포를 나타냄
    - 예) 제품의 수요, 전화가 걸려오는 수, 버스터미널의 버스 도착 수, 톨게이트 도착 차량의 수, 타자 한 장 당 오타 수
- 포아송분포의 적용 조건
  - 한 단위 시간 내에서 사건 발생 수는 다른 단위시간에서의 사건 발생 수와 독립적이다.
  - 극히 작은 단위 시간에 둘 이상의 사건이 발생할 확률은 0으로 간주한다.
  - 관심의 대상이 되는 단위시간은 더 작은 단위시간으로 나눌 수 있으며, 작은 단위시간에서 사건이 발생할 확률은 작다. 즉 확률은 구간의 길이에 비례한다.
- 포아송 확률함수
  - $f(x) = \frac{e^{-\lambda} \cdot (\lambda^{x})}{x!}$
    - $e = 2.71828$
    - λ는 일정 단위시간당 평균 발생횟수
    - x는 관심있는 사건의 발생횟수

연속확률분포

- **연속확률분포: 연속확률변수 *X*가 취할 수 있는 특정한 값 *x*에 대응하는 확률을 나타내주는 확률분포**
- **확률변수가 특정한 구간에 있을 확률로 계산**



연속확률분포

- 이산확률변수와 연속확률변수의 확률계산
  - 이산확률변수 $\boldsymbol{Pr(a \leq X \leq b) = \sum_{x=a}^{b} f(x)}$
  - 연속확률변수 $\boldsymbol{Pr(a \leq X \leq b) = \int_{a}^{b} f(x)}$

- 연속확률변수에서의 확률밀도함수의 성질
  - 특정한 값 x 가 발생할 확률, 즉 $Pr(X=x)=0$이다.
  - 모든 x의 값에 대하여 $f(x) \geq 0$ 이다.
  - 확률밀도함수 아래에 있는 전체 면적은 1이다.

    즉, $\boldsymbol{\int_{-\infty}^{+\infty} f(x) = 1}$

연속확률분포의 측정

- 기대값
  - 이산확률변수의 기대값
    - $E(X) = \sum x \cdot f(x) = \mu$
  - 연속확률변수의 기대값
    - $E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) = \mu$
- 분산
  - 이산확률변수의 분산
    - $\mathrm{Var}(X) = \sum(\boldsymbol{x} - \boldsymbol{\mu})^{\mathbf{2}} \cdot f(x) = \sigma^2$
  - 연속확률변수의 분산
    - $\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (\boldsymbol{x} - \boldsymbol{\mu})^{\mathbf{2}} \cdot f(x) = \sigma^2$

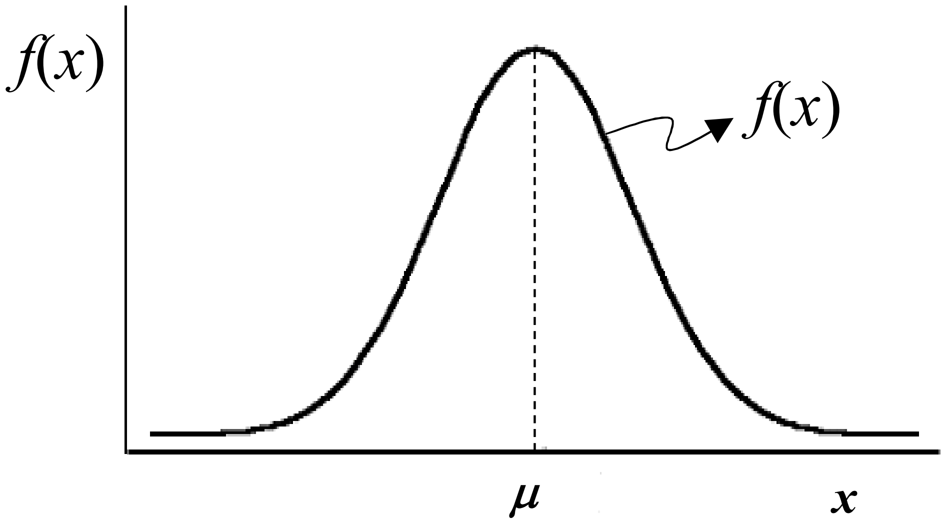
정규분포

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

- 정규확률밀도함수의 평균과 분산
  - $\boldsymbol{E(X) = \mu}$
  - $\boldsymbol{Var(X) = \sigma^2}$

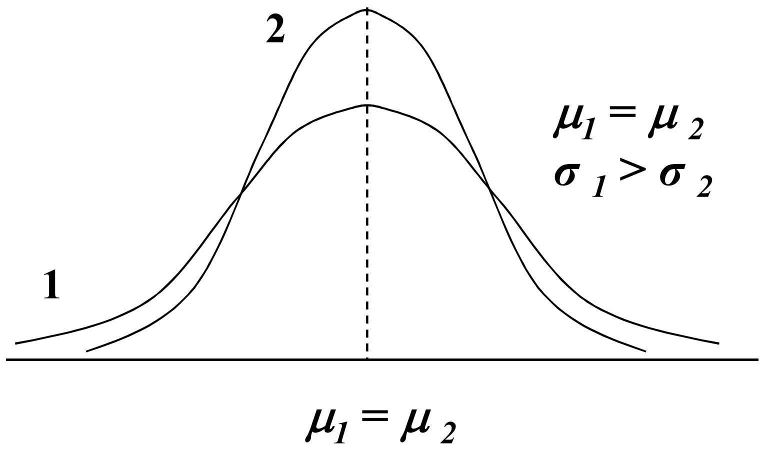
정규분포

- 정규밀도함수

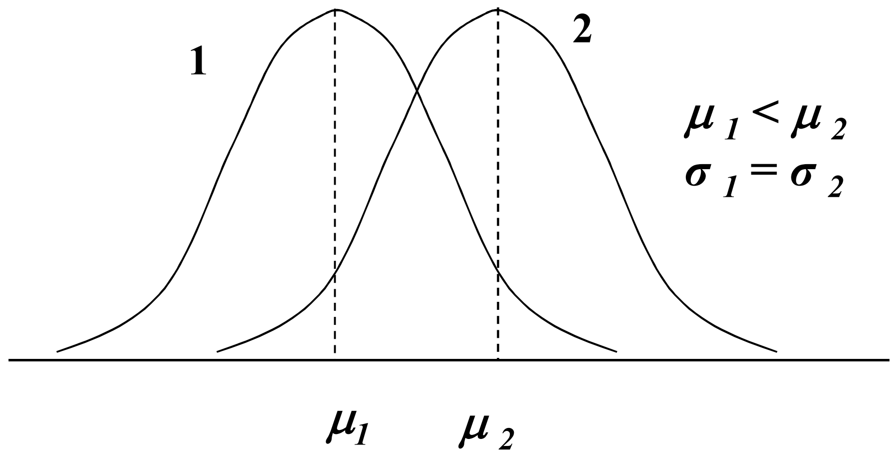


- 정규밀도함수의 특성
  - 평균 μ를 중심으로 좌우 대칭이며 종모양
  - 정규곡선아래와 x축 사이의 면적은 1
  - X가 취할 수 있는 값의 범위는 $-\infty < x < \infty$

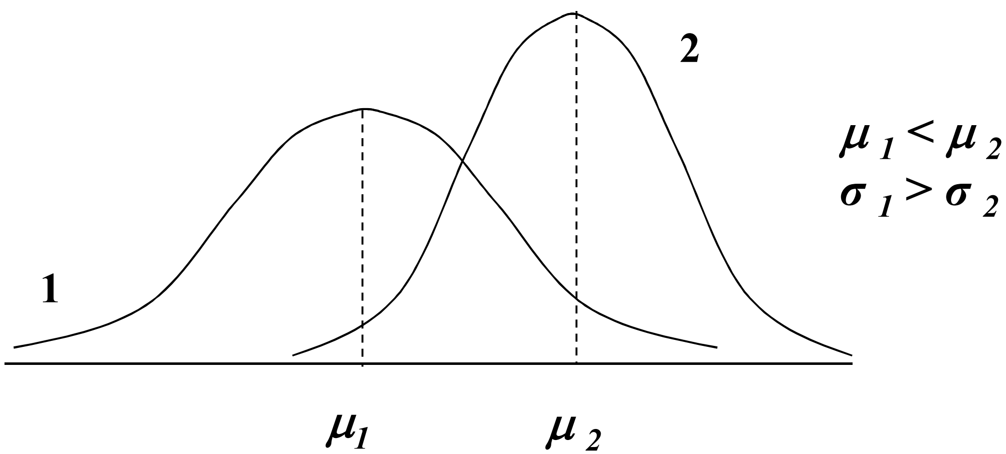
정규분포



평균은 같으나 분산이 다른 경우



평균은 다르나 분산이 같은 경우



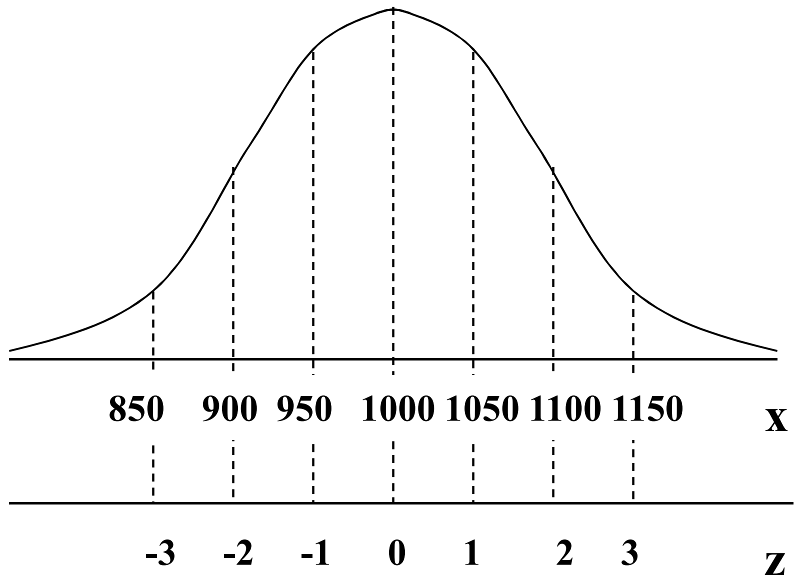
평균과 분산이 모두 다른 경우

정규분포

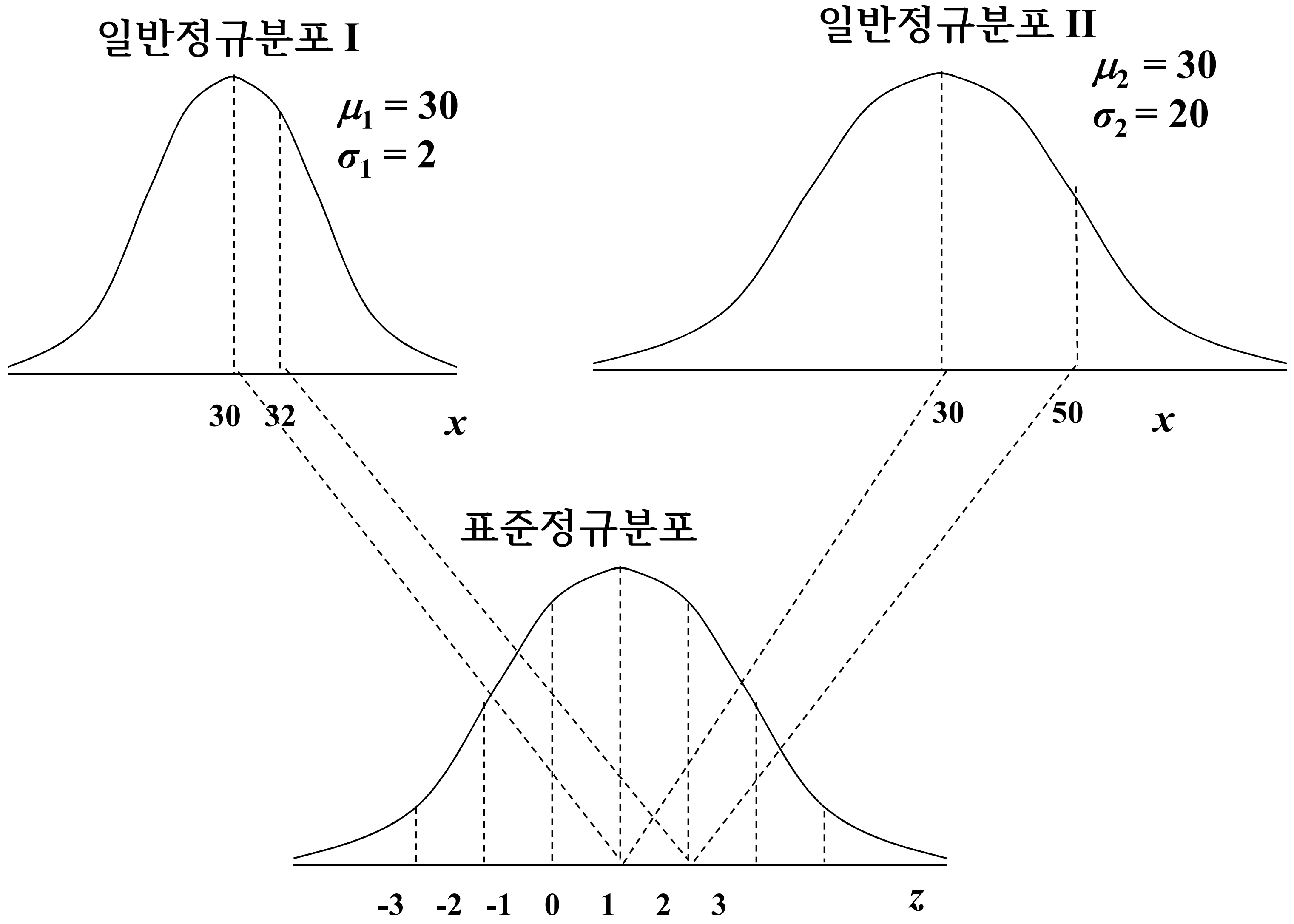
- **정규분포의 표기**
  - **$X \sim N(\mu, \sigma^2)$: 평균이 μ 이고 분산이 σ^2 인 정규분포**
  - **표준정규분포: 평균이 0 이고 분산이 1 인 정규분포**
- **표준화된 정규분포 확률변수 : $Z \sim N(0, 1)$**
  - $Z = \frac{X-\mu}{\sigma}$
    - **X는 관찰치**
    - **μ는 평균**
    - **σ는 표준편차**
- **표준정규확률밀도함수**
  - $f(x) = \frac{1}{\sqrt{2\pi^2}} e^{-\frac{1}{2}Z^2}, \quad -\infty < x < \infty$

정규분포

- ***x*** **값과** ***z*****값의 비교**
- ***x*****~N(100, 50²)**
- ***z*****~N(0, 1²)**



정규분포


일반정규분포 I
$\mu_1 = 30$
$\sigma_1 = 2$
30 32
x
일반정규분포 II
$\mu_2 = 30$
$\sigma_2 = 20$
30 50
x
표준정규분포
-3 -2 -1 0 1 2 3
z