6주

제4장 확률변수(2) - 연속확률변수

4.4 연속확률변수

연속확률변수의 확률분포

- 연속확률변수의 확률모형
  - 확률변수 X가 취할 수 있는 모든 값이 하나씩 셀 수 없는 경우와 같이 확률변수 X가 어떤 구간내의 모든 값을 취할 수 있을 때
  - 즉, 몸무게, 힘, 수명, 온도와 같은 연속적인 척도에 의해 측정된 변수들의 확률분포는 연속적인 곡선으로 나타남
- 연속확률변수의 확률밀도함수
  - 연속확률변수의 확률 밀도 곡선
    - 상대 도수 히스토그램의 극한 형태로써 얻어진 곡선을 연속확률변수의 확률밀도곡선(continuous probability density curve) 이라 함
  - 연속확률변수의 확률밀도함수(continuous probability density function :pdf)
    - 연속확률변수는 실수 또는 실수의 어떤 구간의 값을 가능한 값으로 취하기 때문에 이산확률변수와 같이 그 가능한 값 하나하나에 어떤 확률을 부여하는 것이 가능하지 않음
    - 수학적 함수 $y=f(x)$의 그 구간에서의 면적으로 확률을 표시하도록 정의하여 연속확률변수 X가 취할 수 있는 값이 어떤 구간에 포함된 가능성을 나타낸 것

확률밀도함수의 성질

확률밀도함수 $f(x)$는 연속확률변수 X에 대한 확률의 분포를 나타내며 다음의 성질을 만족한다.

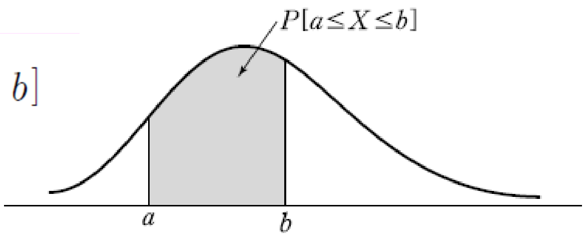
(a) 모든 x에 대해 $f(x) \geq 0$이다.

(b) 확률밀도곡선 아래의 전체 면적은 1이다.

(c) 확률 $P[a \leq X \leq b]$는 a 이상 b 이하의 구간에서 확률밀도곡선 아래의 면적으로 $P[a \leq X \leq b] = \int_a^b f(x)dx$로 계산된다(그림 4.4).

연속확률변수 X는 모든 x에 대해 [X=x]=0일 확률이 항상 0

$$P[a \leq X \leq b] = P[a < X \leq b] = P[a \leq X < b] = P[a < X < b]$$



| 그림 4.5 | 연속확률변수의 확률과 연속확률밀도곡선의 면적

연속확률변수 X의 평균(또는 기대값)

연속확률변수 X의 평균(또는 기대값)은

$$E(X) = \mu = \int x f(x) dx$$

이다. 이 때 적분은 확률변수 X가 정의된 모든 영역의 x에 대하여 적분한다.

연속확률변수 X의 분산과 표준편차

연속확률변수 X의 분산과 표준편차 :

$$\sigma^2 = Var(X) = \int (x-\mu)^2 f(x) dx = \int x^2 f(x) dx - \mu^2$$

$$\sigma = sd(X) = \sqrt{Var(X)}$$

정규분포

- 정규분포(normal distribution)
  - 1733년 프랑스 수학자 드 므와브르(De Moivre, 1667-1754)가 제안

1) 피에르 라플라스(Pierre Laplace, 1749-1827)와 칼 가우스(Carl Gauss 1777-1855)에 의해 이론적인 체계가 이루어짐

2) 정규분포에 관한 이론이 통계학의 근간을 이룸

3) 확률변수의 범위가 실수구간인 연속형 분포

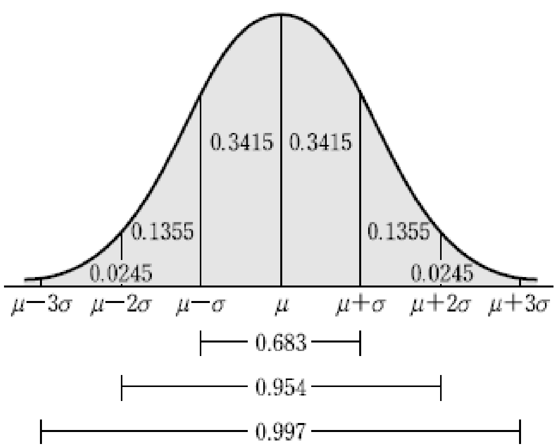
4) 종 모양(bell shape)의 확률밀도함수를 가지며 좌우 대칭의 데이터분포를 나타냄

5) 자연과학 현상은 물론 사회과학 현상을 분석할 때 가장 빈번하게 활용되는 확률분포는 정규분포

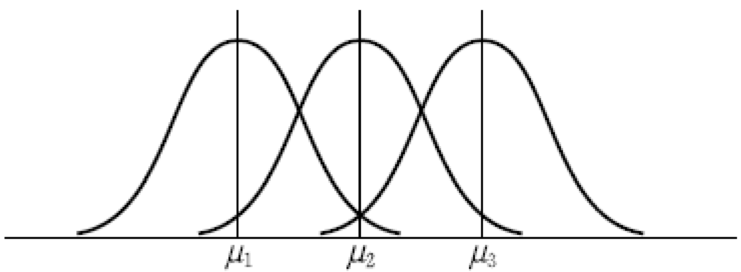
- 확률변수 X 가 평균이 μ 이고 분산이 σ2인 정규분포 할 때
  X ~ N(μ,σ2) 로 표시
- 정규분포 확률밀도함수

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

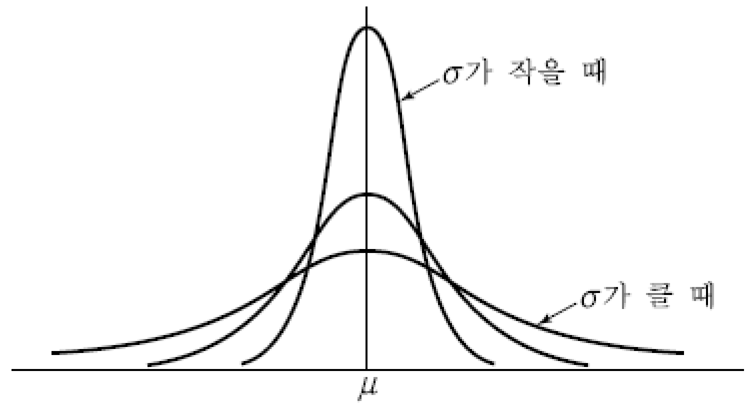
- 이 함수의 그래프에 대한 몇 가지 유의한 성질
  - 정규확률밀도 곡선은 종 모양의 형태로서 최고값을 갖는 평균 μ를 중심으로부터 양 방향으로 각각 표준편차 σ만큼 떨어진 구간
  - 확률변수 X 가 평균이 μ 이고 분산이 σ^2인 정규분포 한다면, μ 로 부터 ±σ, ±2σ, ±3σ 범위 내에 확률변수 값이 포함된 확률은 각각 68.3 %, 95.4 %, 99.7 %



- 정규분포의 위치는 평균 μ에 의해 결정되고, 그 모양은 표준편차 σ의 크기에 의하여 정해짐



| 그림 4.8 | 정규확률분포의 μ(표준편차는 같고 평균이 다를 때)



| 그림 4.9 | 정규확률분포의 σ(평균은 같고 표준편차가 다를 때)

- 정규분포의 특성

첫째, 모든 정규분포는 좌우 대칭이다.

둘째, 정규곡선의 위치는 평균에 의해서 정해지고,

그 모양은 표준편차에 의해 결정된다.

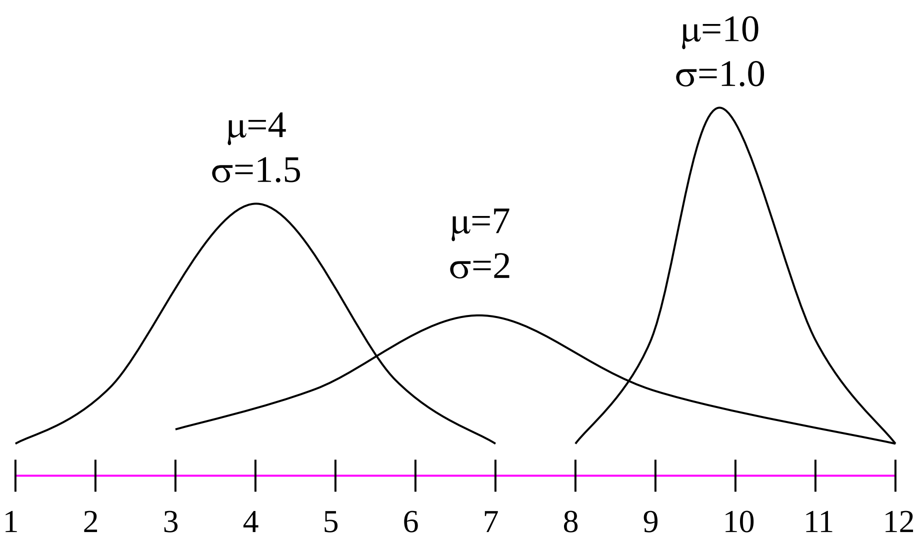
셋째, 자료집단에 따라 평균 μ와 표준편차σ 크기는 다르다.

그러나 평균으로부터 k배 표준편차 이내 범위에서 확률변수 X가 값을 갖게 될 확률은 같다.

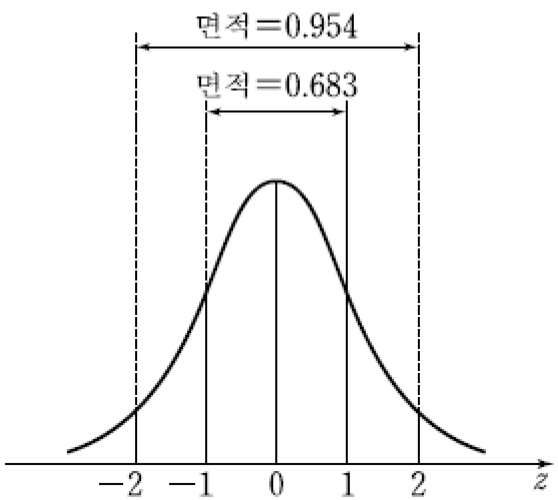
표준정규분포(normal distribution)

- 표준정규분포의 필요성
  - 확률변수 X 가 평균이 μ 이고 분산이 σ^2인 정규분포 할 때

    $X \sim N(\mu, \sigma^2)$
  - 모든 형태에 적용 가능한 표준적인 형태의 분포를 이용하는 것이 편리



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



| 그림 4.10 | 표준정규확률밀도곡선

$X \sim N(0,1)$

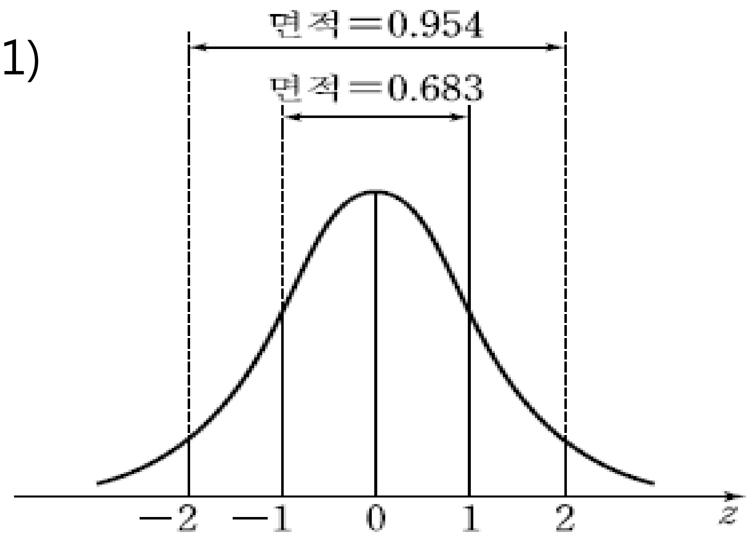
표준정규분포

- 확률변수를 표준화 변수로 변환시킨 것이 표준정규변수이며 Z로 표기

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

- 표준화 변수로 변환된 Z의 확률분포를 표준정규분포

Z~N(0,1)



| 그림 4. 10 | 표준정규확률밀도곡선

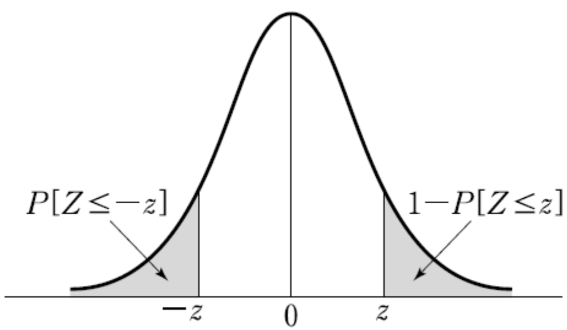
표준정규분포표 (부록B의 표2)

- 부록 B의 표 2에 나와 있는 표준정규분포표는 다음과 같이 지정된 값 z의 왼쪽 부분의 면적을 수록
- 이 면적은 결국 표준정규확률변수가 z값보다 작은 값을 가질 확률

$$P[Z \leqq z] = z\text{의 왼쪽 부분의 곡선 아래의 면적}$$

- 구간 [a,b]에 대한 확률

$$P[a \leqq Z \leqq b] = P[Z \leqq b] - P[Z \leqq a]$$
$$= [b\text{의 왼쪽 부분의 면적}] - [a\text{의 왼쪽 부분의 면적}]$$

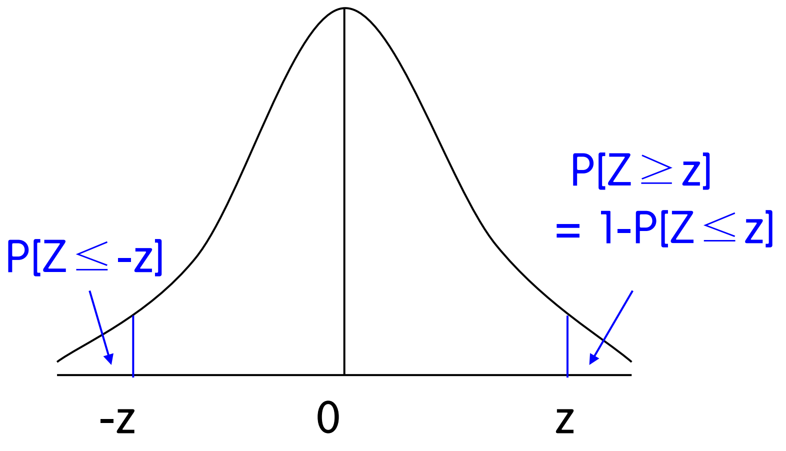


| 그림 4.11 | 표준정규확률변수의 확률

(i) $P[Z \leqq 0] = 0.5$

(ii) 0이상의 값 z에 대해 $P[Z \leqq -z] = 1 - P[Z \leqq z] = P[Z \geqq z]$

- P[Z≤z] = z의 왼쪽 부분의 곡선 아래의 면적
- 구간 [a,b]에 대한 확률
  P[a≤Z≤b] = P[Z≤b] - P[Z≤a]
  = [b의 왼쪽 부분의 면적] - [a의 왼쪽 부분의 면적]
- 표준정규곡선은 0에 관해 대칭
  (i)P[Z≤0]=0.5
  (ii) 0 이상의 값에 대해
  P[Z≤-z] = 1-P[Z ≤z] = P[Z≥z]
  P[Z≥-z] = P[Z ≤z]



- [예제 4.16] 표준확률변수 Z에 대해 다음의 확률을 구하라.

(a) $P[Z \le 1.34] = 0.9099$

(b) $P[Z \ge -0.22] = P[Z \le 0.22] = 0.5871$

(c) $P[0.09 \le Z \le 1.76] = P[Z \le 1.76] - P[Z \le 0.09] = 0.9608 - 0.5359 = 0.4249$

(d) $P[|Z| \ge 0.38] = P[Z \le -0.38]$ 또는 $P[Z \ge 0.38]$

$= P[Z \le -0.38] + P[Z \ge 0.38]$

$= 2 \times P[Z \ge 0.38]$

$= 2 \times (1 - 0.6480) = 0.704$

- 예. 표준정규확률변수 Z 구하기

(a) $P[Z < 1.31]$

(b) $P[Z > 1.20]$

(c) $P[0.67 < Z < 1.98]$

(d) $P[-1.32 < Z < 1.05]$

$$
\begin{aligned}
&(a) P(Z < 1.31) = 0.9049 \\
&(b) P(Z > 1.20) \\
&= 1 - P(Z < 1.20) \\
&= 1 - 0.8849 = 0.1151
\end{aligned}
$$

$$(c)P(0.67 < Z < 1.98)$$
$$= P(Z < 1.98) - P(Z < 0.67)$$
$$= 0.9761 - 0.7486 = 0.2275$$

$$(d)P(-1.32 < Z < 1.05)$$
$$= P(Z < 1.05) - (1 - P(Z < 1.32))$$
$$= 0.8531 - (1 - 0.9066)$$
$$= 0.7597$$

정규분포의 표준화

정규분포의 표준화

- 정규분포를 따르는 확률변수에 대한 확률계산은 새로운 표를 필요로 하지 않고 다음의 관계식을 통해 표준정규분포로 대응시켜서 확률 값을 계산

만일 확률변수 X가 $N(\mu, \sigma^2)$분포를 따르면, 표준화된 확률변수

$$Z = \frac{X - \mu}{\sigma}$$

는 표준정규분포를 따른다.

- X가 주어진 구간에 포함될 확률을 계산하기 위해서는 이 구간을 z 값으로 환산한 다음 표준정규분포표를 이용하여 확률을 구함

만일 확률변수 X가 $N(\mu, \sigma^2)$분포를 따르면,

$$P[a \leq X \leq b] = P\left[\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right]$$

가 성립한다. 여기서 확률변수 Z는 표준정규분포를 따른다.

- [예제 4.19] 어떤 기업체의 입사시험 점수는 평균이 500점이고 표준편차가 100인 정규분포를 따르는 것으로 집계되었다. 임의로 선정된 한 지원자가 다음과 같은 점수를 받을 확률을 구하라.

  (a) 650점보다 높은 점수

  (b) 325점 이상 665점 이하의 점수

기업체의 입사시험 점수를 확률변수 X라고 하면 X는 $N(500, 100^2)$을 따른다.

따라서

$$Z = \frac{X - 500}{100}$$

는 $N(0, 1^2)$을 따른다.

(a) $x = 650$에 대응되는 z 값은

$$z = \frac{650 - 500}{100} = 1.5$$

이므로

$$\begin{aligned} P[X > 650] &= P[Z > 1.5] \\ &= 1 - P[Z \leqq 1.5] \\ &= 1 - 0.9332 \\ &= 0.0668 \end{aligned}$$

이다. 따라서 650점보다 높은 점수를 받을 확률은 0.0668이 된다.

(b) $x = 325$와 $x = 665$에 대응되는 z 값은 각각

$$\frac{325 - 500}{100} = -1.75 \text{와 } \frac{665 - 500}{100} = 1.65$$

이다. 따라서

$$\begin{aligned} P[325 \leqq X \leqq 665] &= P[-1.75 \leqq Z \leqq 1.65] \\ &= 0.9505 - 0.0401 \\ &= 0.9104 \end{aligned}$$

로 얻어져 325점 이상 665점 이하의 점수를 받을 확률은 0.9104가 된다.