# 제8장 연속형 확률분포

## 1. 연속 확률분포

◆ 연속 확률변수 $X$의 확률분포는 다음과 같은 2가지 성질을 가지고 있다.

**성질 1.** $X$가 어떤 구간에 속할 확률은 0과 1사이이다.

**성질 2.** $X$가 값을 가질 수 있는 서로 배반인 모든 구간의 확률을 합하면 1이다. 전체 면적은 1임을 나타낸다.

◆ **Note :**

· 성질 1은 확률변수 $X$의 값이 구간 $[a, b]$에 속할 확률은 0과 1사이의 값임을 나타내고 기호로는 $P(a \le X \le b)$로 나타내며,

· 성질 2는 확률분포곡선 아래의 전체면적은 1 임을 나타낸다.

· 연속 확률변수의 확률은 확률분포곡선에서 구간에 대한 면적으로 계산된다. 따라서 한 점에 대한 확률은 항상 0이다.

· 따라서 임의의 점 $a$와 $b$에 대하여,

$$P(a \le X \le b) = P(a < X < b) = P(a \le X < b) = P(a < X \le b)$$

가 성립한다.

## 2. 정규분포

◆ **정규분포**(normal distribution)는 연속 확률변수가 취할 수 있는 많은 확률분포 중 하나로서 가장 중요하고 가장 많이 사용되는 확률분포이다. 현실 속에서 일어나는 많은 현상들의 분포가 정확하게 혹은 근사적으로 정규분포를 따른다.

· 사람들의 키 혹은 체중을 나타내는 변수, 수능시험의 점수, 과자 혹은 시리얼 한 통의 무게, 1리터 팩에 들어 있는 우유의 양, 어떤 전구나 전자 부품의 수명, 어떤 일을 완성하는데 걸리는 시간 등을 나타내는 변수는 근사적으로 정규분포를 따르는 것으로

관찰될 수 있다.

· 정규분포의 확률분포곡선 혹은 정규곡선(normal curve)은 종을 엎어놓은 모양으로 대칭적(symmetric)인 곡선이다.

· $\mu$ 와 $\sigma^2$ 은 각각 평균과 분산을 나타내며 정규곡선을 확률분포로 갖는 확률변수를 정규확률변수라 한다.

◆ 정규분포는 다음과 같은 성질을 가지고 있다.

**성질 1.** 정규분포곡선 아래의 전체 면적은 1 혹은 100%이다.

**성질 2.** 정규분포는 평균 $\mu$ 에 대하여 대칭이기 때문에 평균의 왼쪽부분의 면적이나 오른쪽 부분의 면적은 같으며 각각 $\frac{1}{2}$이다.

**성질 3.** 정규분포의 꼬리부분은 무한대까지 뻗어있으며, 평균에서 $3\sigma$ 를 벗어난 부분의 면적은 거의 0이다.

◆ 평균 $\mu$ 와 표준편차 $\sigma$ 는 정규분포의 모수이며 이들 모수의 값이 주어졌을 때 어떤 구간에 대한 정규분포의 면적, 즉 확률을 계산할 수 있다. 또한 $\mu\ (-\infty < \mu < \infty)$와 $\sigma\ (\geq 0)$ 의 값이 변함에 따라 다른 정규분포가 된다. 즉, $\mu$ 의 값은 분포의 중앙을 결정하고 $\sigma$ 는 분포의 퍼짐정도를 나타낸다.

# 3 표준정규분포

◆ **표준정규분포**(standard normal distribution)는 정규분포의 특수한 경우로 평균이 0이고 표준편차가 1인 정규분포를 말한다. 그리고 표준정규분포를 따르는 확률변수를 $z$ 로 나타내며, 표준정규분포곡선의 단위를 $z$ **값**(표준단위; standard unit) 혹은 $z$ **점수**(표준점수; standard score)라 부른다.

· 부록의 표7에서 표준정규분포표를 제시하고 있다. 이 표를 이용하여 표준정규분포곡선에서 두 점 사이의 면적의 근사값을 계산할 수 있는데 이는 두 점 사이의 확률을 나타낸다.

## 예제 6-1

표준정규곡선에서 두 점 $z=0$ 과 $z=1.95$ 사이의 면적을 구하여라.

(풀이)

아래 표 6.4를 보자. 부록의 표7에서 정규분포표를 사용하는 방법을 소개하고 있다.

표 6.4 표준정규곡선에서 두 점 $z=0$ 과 $z=1.95$ 사이의 면적

| $z$ | .00 | .01 | ⋯ | .05 | ⋯ | .09 |
|---|---|---|---|---|---|---|
| 0.0 | .0000 | .0040 | ⋯ | .0199 | ⋯ | .0359 |
| 0.1 | .0398 | .0438 | ⋯ | .0596 | ⋯ | .0753 |
| 0.2 | .0793 | .0832 | ⋯ | .0987 | ⋯ | .1141 |
| . | . | . | ⋯ | . | ⋯ | . |
| . | . | . | ⋯ | . | ⋯ | . |
| . | . | . | ⋯ | . | ⋯ | . |
| 1.9 | .4713 | .4719 | ⋯ | .4744 | ⋯ | .4767 |
| . | . | . | ⋯ | . | ⋯ | . |
| . | . | . | ⋯ | . | ⋯ | . |
| . | . | . | ⋯ | . | ⋯ | . |
| 3.0 | .4987 | .4987 | ⋯ | .4989 | ⋯ | .4990 |

따라서 두 점 0과 1.95사이의 면적 = $P(0 < Z < 1.95) = 0.4744$ 가 된다. 6.1절에서 언급한 바와 같이 연속 확률변수가 어떤 특정한 값을 가질 확률은 0이므로

$$P(0 < Z < 1.95) = P(0 \le Z \le 1.95) = P(0 < Z \le 1.95) = P(0 \le Z < 1.95) = 0.4744$$

가 된다.

## 예제 6-2

표준정규곡선에서 두 점 $z=-2.17$ 과 $z=0$ 사이의 면적을 구하여라.

(풀이)

정규분포는 평균에 대하여 대칭이므로 두 점 $z=-2.17$ 과 $z=0$ 사이의 면적은 두 점 $z=0$ 과 $z=2.17$ 사이의 면적과 같다. 따라서 표 7에서 $z$ 값이 2.17에 해당되는 값을 읽으면 0.4850이 됨을 알 수 있다. 즉,
두점 $z=-2.17$ 과 $z=0$ 사이의 면적 = $P(-2.17 < Z < 0) = 0.4850$
가 된다.

## 예제 6-3

표준정규분포에서 다음 부분의 면적을 구하여라.

(1) 점 $z=2.32$의 오른쪽 부분

(2) 점 $z=-1.54$의 왼쪽 부분

(풀이)

(1) $z=0$ 에서 대칭이므로 $z=0$ 의 오른쪽 부분의 면적은 0.5이다. 따라서 표 7에서 두 점 $z=0$ 와 $z=2.32$ 사이의 면적이 0.4898이므로 구하려고 하는 면적은 $0.5-0.4898=0.0102$가 된다. 즉, $z=2.32$ 의 오른쪽 부분의 면적 = $P(Z>2.32)=0.5-0.4898 = 0.0102$ 이다.

(2) 두 점 $z=-1.54$ 와 $z=0$ 사이의 면적은 두 점 $z=0$ 과 $z=1.54$ 사이의 면적과 같으므로 표 7에서 0.4382를 얻는다. 따라서 구하려고 하는 면적은 0.5-0.4382 = 0.0618이 된다. 즉,
$z=-1.54$ 의 왼쪽 부분의 면적 = $P(Z<-1.54)=0.5-0.4382=0.0618$ 이다.

## 예제 6-4

표준정규분포에서 다음의 확률을 구하여라.

(1) $P(1.19<Z<2.12)$

(2) $P(-1.56<Z<2.31)$

(3) $P(Z>-0.75)$

(풀이)

(1) 확률 $P(1.19<Z<2.12)$ 은 표준정규분포에서 두 점 $z=1.19$ 와 $z=2.12$ 사이의 면적과 같다. 따라서 두 점 $z=0$ 과 $z=1.19$사이의 면적이 0.3830이고 두 점 $z=0$ 와 $z=2.12$ 사이의 면적이 0.4830이므로,

$$P(1.19<Z<2.12)=0.4830-0.3830=0.1000$$

가 된다.

(2) 확률 $P(-1.56<Z<2.31)$ 은 표 7에서 두 점 $z=-1.56$ 과 $z=0$ 사이의 면적은 0.4406이고 $z=0$ 과 $z=2.31$ 사이의 면적은 0.4896이므로 $P(-1.56<Z<2.31)=0.4406+0.4896=0.9302$ 가 된다.

(3) 표 7에서 두 점 $z=-0.75$ 와 $z=0$ 사이의 면적은 0.2734이므로 (2)와

같은 방법으로 확률 $P(Z > -0.75)$ 은

$$P(Z > -0.75) = 0.2734 + 0.5 = 0.7734$$

가 된다.

◆ 3.4절에서 언급한 경험법칙은 표준정규분포에 근거한 것으로 <부록>의 표4에서 다음과 같은 결과를 얻는다.

$$P(-1.0 < Z < 1.0) = 2\times 0.3413 = 0.6826$$

$$P(-2.0 < Z < 2.0) = 2\times 0.4772 = 0.9544$$

$$P(-3.0 < Z < 3.0) = 2\times 0.4987 = 0.9974$$

즉, 평균에서 표준편차이내에 있을 확률은 68.26%, 2배의 표준편차이내에 있을 확률은 95.44%, 그리고 3배의 표준편차 이내에 있을 확률은 99.74%임을 알 수 있다.

## 예제 6-5

표준정규분포에서 다음 확률을 구하여라.

(1) $P(0 < Z < 5.67)$

(2) $P(Z < -5.35)$

(풀이)

(1) 표 7에는 $z = 5.67$에 대한 값이 없다. 그러나 5.67은 3.09보다 크므로 확률 $P(0 < Z < 5.67)$ 은 적어도 0.4990보다 큼을 알 수 있다. 따라서 위의 확률은 근사적으로 0.5임을 알 수 있다.

(2) 같은 방법으로 확률 $P(Z < -5.35)$ 은 근사적으로 0임을 알 수 있다.

## 예제 6-6

표준정규분포의 오른쪽 꼬리부분의 면적이 0.0050되는 점를 계산하여라.

(풀이)

두 점 0과 $z$ 사이의 면적이 $0.5 - 0.0050 = 0.495$가 되는 $z$를 <부록>의 표4에서 찾으면 된다. <부록>의 표4에서 $z = 2.57$ 일때는 0.4949이고 $z = 2.58$ 일때는 0.4951이 된다. 따라서 우리가 찾는 $z$는 2.57과 2.58사이의

값이다. 보간법을 사용하면 $z = 2.575$ 가 된다.

# 4 정규분포의 표준화

◆ <부록>의 표4에서는 평균이 0이고 표준편차가 1인 정규분포만을 다룬다. 따라서 <부록>의 표4을 이용하여 일반적인 정규분포의 확률을 계산할 수 없다. 그러나 정규분포를 표준화하면 이를 해결할 수 있다.

확률변수 $X$ 가 평균이 $\mu$ 이고 표준편차가 $\sigma$ 인 정규분포를 따른다고 하자. 그러면 $X$ 의 표준화변수(standardized variable) $Z$ 는 다음과 같다.

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

그러면 $Z$ 는 항상 평균이 0이고 표준편차가 1이 되어 표준정규분포를 따른다.

## 예제 6-7

확률변수 $X$ 가 평균이 25이고 표준편차가 4인 정규분포를 따른다고 하자. 다음을 계산하여라.

(1) 두 점 $x = 25$와 $x = 32$사이의 면적

(2) 두 점 $x = 18$와 $x = 34$사이의 면적

(풀이)

(1) $Z = \frac{X-25}{4}$ 라 두면 $z$ 는 표준정규분포를 따른다. 두 점 $x = 25$와 $x = 32$사이의 면적은 두 점 $Z = \frac{25-25}{4} = 0$과 $Z = \frac{32-25}{4} = 1.75$ 사이의 면적과 같다. 따라서 표 7에서 두 점 $x = 25$ 와 $x = 32$ 사이의 면적은 다음 확률로 계산된다. 즉,

$$P(25 < X < 32) = P(0 < Z < 1.75) = 0.4599$$

이다.

(2) 위와 같은 방법으로 두 점 $x = 18$와 $x = 34$사이의 면적은 두 점 $Z = \frac{18-25}{4} = -1.75$과 $Z = \frac{34-25}{4} = 2.25$사이의 면적과 같다. 따라서

<부록>의 표4을 이용하면

$$P(18 < X < 34) = P(-1.75 < Z < 2.25) = 0.4599 + 4878 = 0.9477$$

이 된다.

## 예제 6-8

확률변수 $X$가 평균이 40이고 표준편차가 5인 정규분포를 따른다고 하자. 다음 확률을 계산하여라.

(1) $P(X > 55)$

(2) $P(X < 49)$

(3) $P(32 < X < 42)$

(4) $P(X > 60)$

(풀이)

(1) $x = 55$의 표준단위는 $Z = \dfrac{55-40}{5} = 3.00$이므로

$$P(X > 55) = P(Z > 3.00) = 0.5 - 0.4987 = 0.0013$$

가 된다.

(2) $x = 49$의 표준단위는 $Z = \dfrac{49-40}{5} = 1.80$ 이므로

$$P(x < 49) = P(z < 1.80) = 0.5 + 0.4641 = 0.9641$$

가 된다.

(3) 같은 방법으로

$$\begin{aligned} P(32 < X < 42) &= P\left(\frac{32-40}{5} < Z < \frac{42-40}{5}\right) = P(-1.60 < Z < 0.40) \\ &= 0.4452 + 0.1554 = 0.6006 \end{aligned}$$

가 된다.

(4) <부록>의 표4로부터 $P(X > 60) = P(Z > 4.00) \approx 0$ 가 된다.

## 예제 6-9

평균이 30이고 표준편차가 10인 정규분포에서
$P(30 - c < X < 30 + c) = 0.95$ 가 되는 $c$ 를 구하여라.

(풀이)

$x$ 의 확률분포가 30에 대하여 대칭이므로
$P(30 < X < 30 + c) = P\left(0 < Z < \frac{c}{10}\right) = 0.4750$ 가 되는 $c$ 를 찾으면 된다. 따라서 <부록>의 표4에서 $\frac{c}{10} = 1.960$ 이다. 즉, $c = 19.6$가 된다.