3. 연속확률변수와 확률분포

3-1. 연속확률변수

3-1-1. 연속확률변수

; 어떤 구간의 모든 실숫값을 취하는 확률변수 X

〚예〛

3-1-2. 확률밀도함수

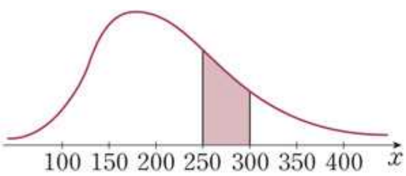
구간 $[\alpha,\ \beta]$의 임의의 값을 취하는 연속확률변수 X에 대하여 다음의 성질을 가지는 함수 $f(x)$

① $f(x) \geq 0$

② $\int_{\alpha}^{\beta} f(x)dx = 1$

③ $\mathrm{P}(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$ (단, $\alpha \leq x_1 \leq X \leq x_2 \leq \beta$)

〚예〛 연속확률변수 X의 확률밀도함수의 그래프가 오른쪽 그림과 같을 때 확률 $\mathrm{P}(250 \leq X \leq 300)$은 색칠한 부분의 넓이와 같다.



3-1-3. 연속확률변수 X의 평균, 분산, 표준편차

① $\mathrm{E}(X) = \int_{\alpha}^{\beta} xf(x)dx = m$

② $\mathrm{V}(X) = \mathrm{E}\{(X-m)^2\} = \int_{\alpha}^{\beta} (x-m)^2 f(x)dx$

③ $\sigma(X) = \sqrt{\mathrm{V}(X)}$

〚예〛 연속확률변수 X의 확률밀도함수가 $f(x) = 2x\,(0 \leq x \leq 1)$일 때, X의 평균과 표준편차를 각각 구하여라.

[풀이]

㉠ $\mathrm{E}(X) = \int_{0}^{1} x \cdot\ 2x\,dx = \left[\frac{2}{3}x^3\right]_0^1 = \frac{2}{3}$

㉡ $\mathrm{V}(X) = \int_{0}^{1} \left(x - \frac{2}{3}\right)^2 \cdot\ 2x\,dx = 2\int_{0}^{1} \left(x^3 - \frac{4}{3}x^2 + \frac{4}{9}x\right) dx = \frac{1}{18}$

㉢ $\sigma(X) = \sqrt{\frac{1}{18}} = \frac{\sqrt{2}}{6}$

3-2. 정규분포(Normal Distribution)

<잠시>

키, 몸무게, IQ, 제품의 수명, 측정 오차 등 자연 현상이나 사회 현상에서 나타나는 여러 가지 통계 자료, 자연과학이나 공학에서의 실험자료 또는 사회경제적인 조사 자료

$\Downarrow\leftarrow$ '정리하여 히스토그램 나타내기'

자료의 수가 커짐에 따라 좌우 대칭인 종 모양의 곡선에 가까워짐
: 정규분포 혹은 정규분포의 형태에 가까운 분포

(단) 자료가 정규분포의 형태를 많이 벗어나는 경우에도 중심극한 정리(central limit theorem)에 의해 정규분포로 근사할 수 있는 경우가 많음

3-2-1. 정규분포의 확률밀도함수

① 연속확률변수 X가 모든 실숫값을 취하고, 그 확률밀도함수가

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}\quad(-\infty<x<\infty,\ e=2.718281\cdots\cdots)$$

일 때, X의 확률분포를 정규분포라고 한다.

$$X\backsim \mathrm{N}(m,\ \sigma^2)$$

과 같이 나타낸다.

3-2-2. 정규분포의 그래프 성질

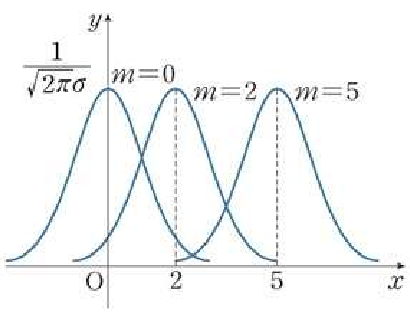
① 직선 $x=m$에 대하여 대칭이다.

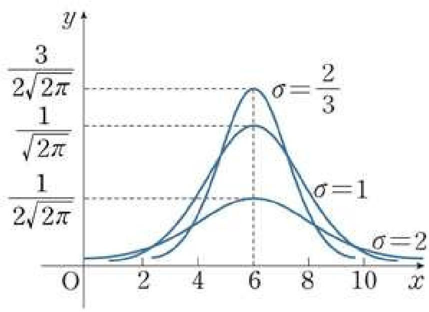
② $x=m$일 때 최댓값 $\dfrac{1}{\sqrt{2\pi}\,\sigma}$을 갖는다.

③ x축을 점근선으로 한다.

④ 곡선과 x축 사이의 넓이는 1이다.

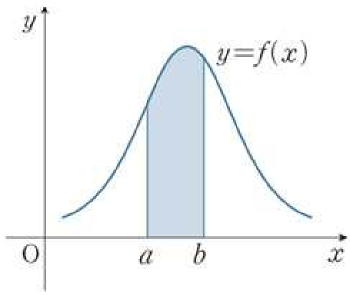
⑤ m과 σ의 값에 따라 다음과 같은 모양이 된다.





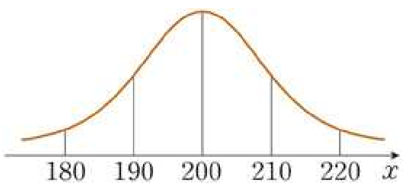
3-2-3. 확률 구하기

① 확률변수 X가 정규분포를 따를 때, X가 a 이상 b 이하의 값을 취할 확률 $\mathrm{P}(a \le X \le b) = \int_a^b f(x)dx$



〖예〗 어느 회사의 초콜릿 한 봉지의 무게는 평균 200 g, 표준편차 10 g인 정규분포를 따른다고 한다. 초콜릿 한 봉지의 무게가 210 g 이하일 확률은?

[풀이]

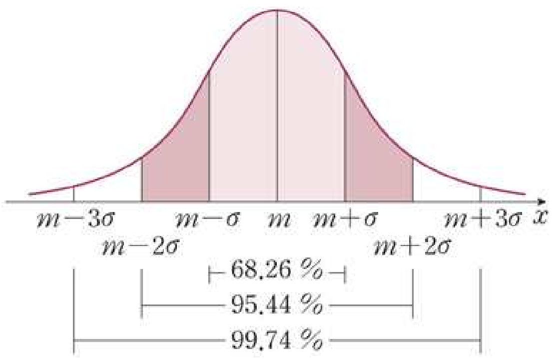


3-2-4. 중요한 확률계산

① $P(m-\sigma < X < m+\sigma) = 0.6826$

② $P(m-2\sigma < X < m+2\sigma) = 0.9544$

③ $P(m-3\sigma < X < m+3\sigma) = 0.9974$



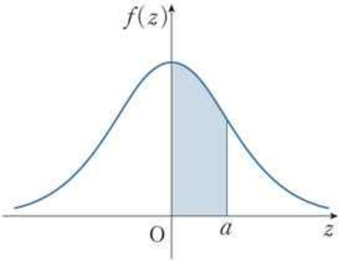
3-3. 표준정규분포

3-3-1. $Z \backsim \mathrm{N}(0,\ 1)$

① 표준정규분포를 따르는 확률변수 Z의 확률밀도함수

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

② Z가 닫힌 구간 $[0,\ a]$에 속할 확률 $\mathrm{P}(0 \le Z \le a)$



㉠ $z = 0$에 대하여 좌우 대칭

㉡ $\mathrm{P}(Z \ge 0) = \mathrm{P}(Z \le 0) = 0.5$

㉢ $\mathrm{P}(0 \le Z \le a) = \mathrm{P}(-a \le Z \le 0)$

〖예〗 $Z \sim \mathrm{N}(0,\ 1)$

㉠ $\mathrm{P}(Z \le 1.32)$

㉡ $\mathrm{P}(-0.56 \le Z \le 1.95)$

[풀이]

① $\mathrm{P}(Z \le 1.32)$

$= \mathrm{P}(Z \le 0) + \mathrm{P}(0 \le Z \le 1.32)$

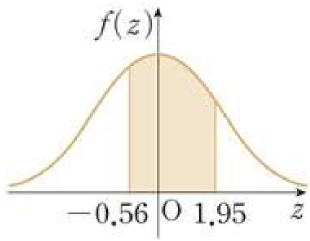
$= 0.5 + 0.4066 = 0.9066$

② $\mathrm{P}(-0.56 \le Z \le 1.95)$

$= \mathrm{P}(-0.56 \le Z \le 0) + \mathrm{P}(0 \le Z \le 1.95)$

$= \mathrm{P}(0 \le Z \le 0.56) + \mathrm{P}(0 \le Z \le 1.95)$

$= 0.2123 + 0.4744 = 0.6867$



3-3-2. 정규분포의 표준화(Standarizing)

①

$X \sim \mathrm{N}(m,\ \sigma^2)$		$Z \sim \mathrm{N}(0,\ 1)$
X		$Z=\dfrac{X-m}{\sigma}$
㉠ $\mathrm{E}(X)=m$ ㉡ $\mathrm{V}(X)=\sigma^2$	⇒	㉠ $\mathrm{E}(Z)=\mathrm{E}\left(\dfrac{X-m}{\sigma}\right)=\mathrm{E}\left(\dfrac{1}{\sigma}X-\dfrac{m}{\sigma}\right)$ $=\dfrac{1}{\sigma}\mathrm{E}(X)-\dfrac{m}{\sigma}=\dfrac{m}{\sigma}-\dfrac{m}{\sigma}=0$ ㉡ $\mathrm{V}(Z)=\mathrm{V}\left(\dfrac{X-m}{\sigma}\right)$ $=\left(\dfrac{1}{\sigma}\right)^2\mathrm{V}(X)=\dfrac{1}{\sigma^2}\cdot\ \sigma^2=1$

② 확률변수 X가 정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따를 때

$$\mathrm{P}(x_1 \le X \le x_2)=\mathrm{P}\left(\frac{x_1-m}{\sigma}\le \frac{X-m}{\sigma}\le \frac{x_2-m}{\sigma}\right)$$
$$=\mathrm{P}\left(\frac{x_1-m}{\sigma}\le Z\le \frac{x_2-m}{\sigma}\right)$$

〚예〛 $X \sim \mathrm{N}(30,\ 4^2)$일 때, $\mathrm{P}(34 \le X \le 38)$는?

[풀이]

$$\mathrm{P}(34 \le X \le 38)=\mathrm{P}\left(\frac{34-30}{4}\le \frac{X-30}{4}\le \frac{38-30}{4}\right)$$
$$=\mathrm{P}(1\le Z\le 2)$$
$$=\mathrm{P}(0\le Z\le 2)-\mathrm{P}(0\le Z\le 1)$$
$$=0.4772-0.3413$$
$$=0.1359$$

〚예〛 어느 도시의 고등학교 학생들의 통학 거리는 평균이 4 km이고, 표준편차가 0.6 km인 정규분포를 따른다고 한다. 통학 거리가 4.6 km 이상인 학생들은 전체의 몇 %인가?

[풀이]

통학 거리 X는 정규분포 $\mathrm{N}(4,\ 0.6^2)$를 따름

확률변수 $Z=\dfrac{X-4}{0.6}$는 표준정규분포 $\mathrm{N}(0,\ 1)$를 따름

$$\mathrm{P}(X\ge 4.6)=\mathrm{P}\left(Z\ge \frac{4.6-4}{0.6}\right)=\mathrm{P}(Z\ge 1)=0.5-0.3413=0.1587$$

따라서 통학 거리가 4.6 km 이상인 학생들은 전체의 약 15.9 %이다.

〖예〗 어떤 합성세제 생산 공장에서 자동포장기를 사용하여 생산된 합성세제를 봉지에 주입하여 포장하고 있다. 봉지당 세제 주입량 규격은 $(97,\ 103)gr$으로, 각 봉지에 주입된 세제의 양이 규격의 범위 내에 들어야 품질검사에서 합격될 수 있다고 한다. 자동포장기에 의한 포장 봉지당 세제 주입량을 조사한 결과, 평균 $102gr$, 표준편차 $2gr$의 정규분포를 따르는 것으로 밝혀졌다. 자동포장기에서 포장된 세제 봉지 중 몇 %가 검사에서 합격할 수 있나?

[풀이]

정규분포 $\mathrm{N}(102,\ 2^2)$에 대하여

$$\begin{aligned} P(97 < X < 103) &= P(X < 103) - P(X < 97) \\ &= P\left(Z < \frac{103-102}{2}\right) - P\left(Z < \frac{97-102}{2}\right) \\ &= P(Z < 0.5) - P(Z < -2.5) \\ &= 0.6915 - 0.0062 = 0.6853 \end{aligned}$$

즉, 68.53%가 검사에서 합격될 수 있다.

3-3-3. 정규분포의 선형결합

① n개의 확률변수 $X_1,\ X_2,\ \cdots,X_n$이 서로 독립으로 각각 $\mathrm{N}(m_1,\ {\sigma_1}^2)$, $\mathrm{N}(m_2,\ {\sigma_2}^2)$, $\cdots$, $\mathrm{N}(m_n,\ {\sigma_n}^2)$의 정규분포를 따를 때, 임의의 상수 $a_1,\ a_2,\ \cdots,a_n$에 의해 선형결합이 된다.

② $X = \sum_{i=1}^{n} a_i X_i \sim \mathrm{N}(m_X,\ {\sigma_X}^2)$

$$\Rightarrow\ m_X = \sum_{i=1}^{n} a_i m_i,\ \ {\sigma_X}^2 = \sum_{i=1}^{n} {a_i}^2 {\sigma_i}^2$$

〖예〗 세제 봉지 100개씩을 한 상자에 포장한다고 하자. 상자 내의 i번째 봉지에 주입된 세제의 중량을 X_i, 봉지 자체의 무게를 B_i 그리고 상자 자체의 무게를 C라고 할 때, 제품 한 상자 전체의 무게 W는

$$W = \sum_{i=1}^{100} (X_i + B_i) + C$$

여기서, $X_i \sim N(102,\ 2^2)$, $B_i \sim N(5,\ 0.5^2)$ 그리고 $C \sim N(106,\ 8^2)$이며, 이들 X_i, B_i, C는 서로 독립이라고 하면, W는 평균과 분산은?

[풀이]

평균 $\mu_W = \sum_{i=1}^{100} (102+5) + 160 = 10{,}860$

분산 ${\sigma_W}^2 = \sum_{i=1}^{100} (2^2 + 0.5^2) + 8^2 = 489$

을 갖는 정규분포에 따른다.