

제 6 장 연속확률분포와 정규분포

평균이 100, 표준편차가 10인 정규분포에서 추출한 10만개의 데이터

순서	관찰값
1	93.3975771054312
2	95.5682204017059
3	117.6887216376710
4	94.4098050601060
5	93.8422855869881
6	105.2421637066770
7	102.4420392282740
8	102.5331174473210
...	...
99,999	104.4473561954800
100,000	82.2465827008973

A1				
A	B	C	D	E
119.6448				

도수분포표로 정리하면

구간	빈도	상대빈도	구간	빈도	상대빈도	구간	빈도	상대빈도
50	0	-	99.990	1	0.0000	99.9990	-	-
55	1	0.0000	99.991	1	0.0000	99.9991	-	-
60	4	0.0000	99.992	4	0.0000	99.9992	2	0.0000
65	24	0.0002	99.993	4	0.0000	99.9993	-	-
70	105	0.0011	99.994	5	0.0001	99.9994	1	0.0000
75	494	0.0049	99.995	0	-	99.9995	-	-
80	1,726	0.0173	99.996	6	0.0001	99.9996	-	-
85	4,280	0.0428	99.997	6	0.0001	99.9997	-	-
90	9,282	0.0928	99.998	2	0.0000	99.9998	-	-
95	15,000	0.1500	99.999	2	0.0000	99.9999	1	0.0000
100	19,221	0.1922	100.000	4	0.0000	100.0000	-	-
105	19,121	0.1912	100.001	6	0.0001	100.0001	-	-
110	14,940	0.1494	100.002	4	0.0000	100.0002	3	0.0000
115	9,175	0.0918	100.003	4	0.0000	100.0003	-	-
120	4,347	0.0435	100.004	3	0.0000	100.0004	-	-
125	1,682	0.0168	100.005	4	0.0000	100.0005	1	0.0000
130	476	0.0048	100.006	6	0.0001	100.0006	1	0.0000
135	106	0.0011	100.007	2	0.0000	100.0007	-	-
140	12	0.0001	100.008	2	0.0000	100.0008	-	-
145	4	0.0000	100.009	6	0.0001	100.0009	1	0.0000
150	0	-	100.010	3	0.0000	100.0010	-	-

자료는 연속 자료로서, 값의 경우는 무한대이다.

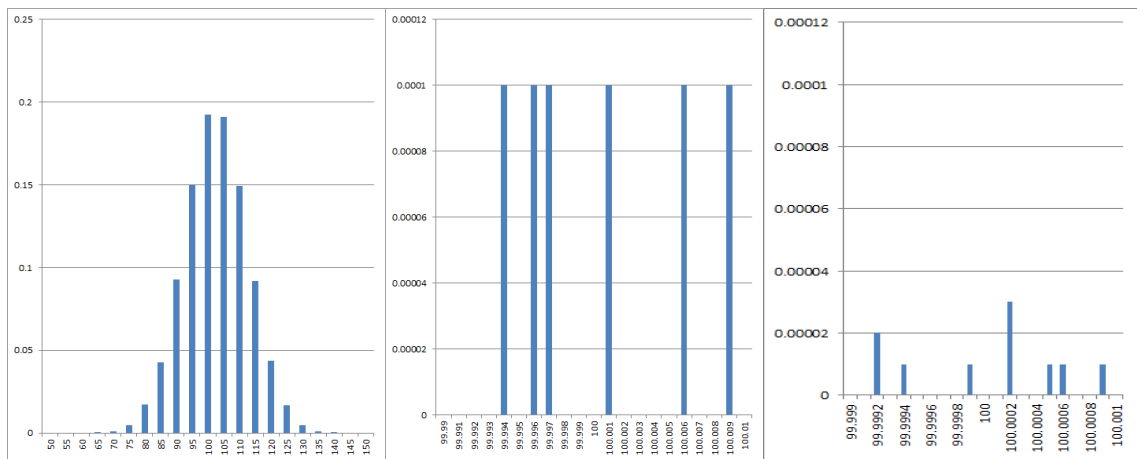
99와 100 사이에도 무한대의 경우의 수가 존재한다.

이항분포의 경우, 자료값은 $0, 1, 2, \dots, n$ 으로 경우의 수는 $n+1$ 이다. (이산적)

포아송의 경우, 자료값은 $0, 1, 2, \dots, \infty$ 이다. (이산적)

도수분포표에서 구간을 세분화할수록 대부분의 구간에서 상대빈도(확률)은 0이고, 0이 아니라도 0에 가까운 값이 된다.

예) 3번째 표에서 21개의 구간 중 빈도가 0인 구간이 14개이며, 빈도가 0이 아닌 경우에도 그 값이 3 이하이다.



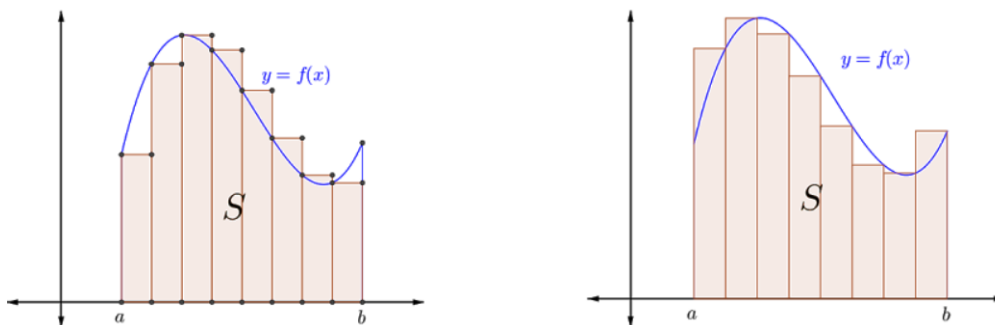
1. 연속확률분포

1.1 [기초정리] 적분(Integral)

1.1.1 적분 기본 개념

적분이란 넓이, 부피 등을 구할 때 매우 유용

적분의 개념은 주어진 함수를 직사각형으로 아주 작게 잘라 나누어 그 직사각형의 합을 구하는 작업 → 적분 계산은 미분의 역연산



1.1.2 적분 표기법

$$\int_a^b f(x)dx \leftarrow \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k)\Delta x \quad \text{where } x_k = a + k\Delta x, \quad \Delta x = \frac{b-a}{n}$$

1.1.3 미적분의 기본공식

$$\int_a^b f(x)dx = F(b) - F(a) \text{ where } f \text{ is continuous on } [a, b] \text{ and } (F') = f$$

1.1.4 다항함수(polynomials) 적분 공식

$$\int k dx = kx + C$$

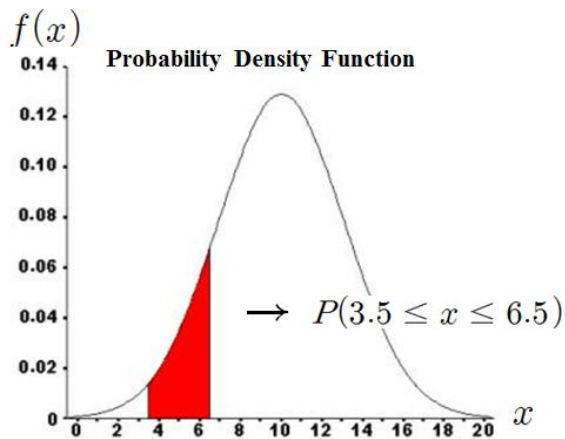
$$\int x^n dx = \frac{x^{n+1}}{n+1} + C \text{ for } n \neq -1$$

$$\int (ax+b)^n dx = \frac{(ax+b)^{n+1}}{a(n+1)} + C \text{ for } n \neq -1$$

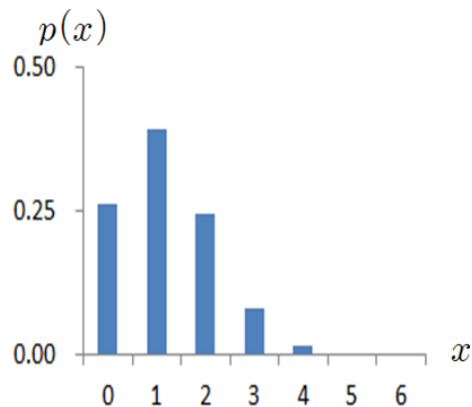
$$\int \frac{1}{x} dx = \ln|x| + C$$

1.2 연속확률변수의 확률밀도함수(probability density function, pdf)

연속확률분포의 양상을 나타내는 곡선을 식으로 표현한 것. 보통 $f(x)$ 로 표기.



연속확률변수



이산확률변수

이산확률변수 $P(a \leq X \leq b) = \sum_{x=a}^b p(x)$

높이의 합

$\sum_x p(x) = 1$

전체 높이의 합은 1

연속확률변수 $P(a \leq X \leq b) = \int_a^b f(x)dx$

구간 면적

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

전체 구간 면적의 합은 1

1.3 연속확률변수의 확률밀도함수 특징

1. $P(X = x^*) = 0$

$$\rightarrow \int_{x^*}^{x^*} f(X)dX = F(X^*) - F(X^*) = 0$$

구간만 양의 확률을 가짐

\rightarrow 구간의 면적 ≥ 0

구간이 아닌 선(특정한 값)은 항상 확률 0 \rightarrow 선의 면적 = 0

2. 모든 x 에 대하여 $f(x) \geq 0$

$f(x)$ 는 0보다 작을 수 없다.

3. $P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$

확률밀도함수의 면적의 합은 항상 1이다.

1.3.1 Example 아래 $f(x)$ 의 확률밀도함수 여부

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x, & 0 \leq x \leq 2 \\ 0, & x > 2 \end{cases}$$

(1) 특정한 값에 대해서는 확률이 항상 0 성립

(2) 모든 x 에 대하여 $f(x) \geq 0$ 이 성립

$$(3) P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = \int_0^2 \frac{1}{2}x dx = \left[\frac{x^2}{4} \right]_0^2 = 1 - 0 = 1$$

연습문제 1. 확률변수 x 의 확률밀도함수는 아래와 같이 정의된다고 하자.

$$f(x) = \begin{cases} 0.1, & 0 \leq x \leq 10 \\ 0, & \text{otherwise} \end{cases}$$

문제 1. $P(X = 1)$

문제 2. $P(X = 10)$

문제 3. $f(x = -3,000)$ x 가 -3,000 일 때의 확률밀도함수의 값(높이)

문제 4. $f(x = -3)$ x 가 -3 일 때의 확률밀도함수의 값(높이)

문제 5. $f(x = 3)$ x 가 3 일 때의 확률밀도함수의 값(높이)

문제 6. $f(x = 15)$ x 가 15 일 때의 확률밀도함수의 값(높이)

문제 7. $f(x = 15,000)$ x 가 15,000 일 때의 확률밀도함수의 값(높이)

문제 8. $f(x) < 0$ 인 구간이 존재한다. True, False

연습문제 2. 확률변수 x 의 확률밀도함수는 아래와 같이 정의된다고 하자.

$$f(x) = \begin{cases} 0.5, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

문제 1. $P(x = 1)$

문제 2. $P(x = 10)$

문제 3. $P(0 \leq x \leq 1)$

문제 4. $P(-1 \leq x \leq 1)$

문제 5. $P(0 \leq x \leq 2)$

문제 6. $P(-\infty \leq x \leq \infty)$

문제 7. $f(x = -1)$ x 가 -1 일 때의 확률밀도함수의 값(높이)

문제 8. $f(x = 1)$ x 가 1 일 때의 확률밀도함수의 값(높이)

2. 연속확률분포의 측정

2.1 평균[기대값]

이산확률변수 $E(X) = \sum_x x \cdot p(x)$

연속확률변수 $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$ $\int_{-\infty}^{\infty} () \cdot f(x) dx$ 는 ()의 평균을 의미

..... 여기서는 x 의 평균

2.2 분산과 표준편차

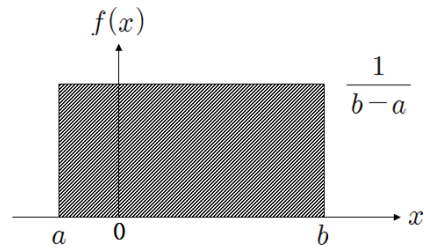
이산확률변수 $V(X) = \sum_x (x - \mu)^2 \cdot p(x)$

연속확률변수 $V(X) = \int_{-\infty}^{\infty} [x - E(x)]^2 \cdot f(x) dx$ 여기서는 x 편차 제곱의 평균

3. 일양분포(uniform distribution)

3.1 일양분포의 확률밀도함수

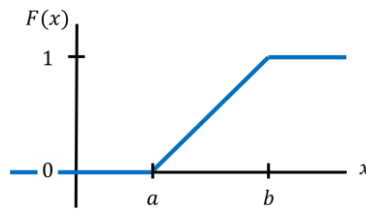
$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \quad b > a \\ 0, & \text{elsewhere} \end{cases}$$



a 와 b 는 상수. Uniform Distribution의 형태는 두 상수 a 와 b 에 의해 결정된다. 이처럼 확률밀도함수의 형태를 결정짓는 상수를 그 함수의 parameter라고 한다.

3.2 일양분포의 누적분포함수(cumulative distribution function)

$$f(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$



3.3 평균과 분산

$$E(x) = \mu = \frac{a+b}{2} \quad ^1, \quad V(x) = \sigma^2 = \frac{(b-a)^2}{12} \quad ^2$$

$$\begin{aligned} ^1 E(x) &= \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x \frac{1}{b-a} dx = \left[\frac{1}{b-a} \frac{1}{2} x^2 \right]_a^b \\ &= \frac{1}{2(b-a)} (b^2 - a^2) = \frac{a+b}{2} \end{aligned}$$

$^2 V(x) = E(x^2) - [E(x)]^2$ 이 성립한다.

$$\begin{aligned} E(x^2) &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left[\frac{1}{3} x^3 \right]_a^b \\ &= \frac{1}{3(b-a)} [b^3 - a^3] = \frac{(b-a)(b^2 + ba + a^2)}{3(b-a)} = \frac{(b^2 + ba + a^2)}{3} \end{aligned}$$

$$[E(x)]^2 = \left[\frac{a+b}{2} \right]^2 = \frac{a^2 + 2ab + b^2}{4}$$

3.4 각 확률분포의 평균과 분산

분포명	parameter	평균	분산
이항분포	n, p	np	$np(1-p)$
포아송분포	λ	λ	λ
일양분포	a, b	$\mu = \frac{a+b}{2}$	$\sigma^2 = \frac{(b-a)^2}{12}$

→ μ 와 σ 값이 이렇게 정의되는 것이 아니라 μ 와 σ 값은 이 값들과 일치

3.5 모의실험

확률변수 $x \sim U[5, 15]$

3.5.1 컴퓨터로 데이터 10,000개 생성[a=5, b=15, 일양분포]

번호	값 x_i	편차 제곱 $(x_i - \mu)^2$
1	7.8825	4.23515
2	5.9465	15.95118
3	8.5088	2.049637
...
9,999	8.2486	2.862464
10,000	10.2604	0.102375
합계	$99,404.42 = \sum_{i=1}^{10,000} x_i$	$83,041.67 = \sum_{i=1}^{10,000} (x_i - \mu)^2$
평균	$9.9404 = \frac{\sum_{i=1}^{10,000} x_i}{10,000}$ $= E(x)$	$8.3042 = \frac{\sum_{i=1}^{10,000} (x_i - \mu)^2}{10,000}$ $= V(x)$

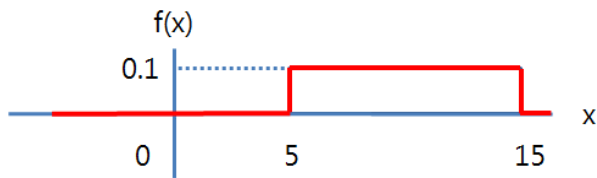
3.5.2 공식을 이용한 평균과 분산 추정

$$a = 5, b = 15 \text{이므로 } \mu = \frac{a+b}{2} = 10, \quad \sigma^2 = \frac{(b-a)^2}{12} = 8.3333$$

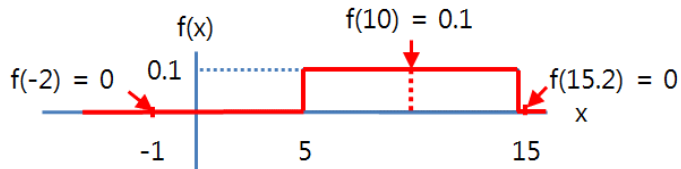
3.5.3 예제 [1]

(1) $f(x)$ 를 붉은 색으로 표기하시오.

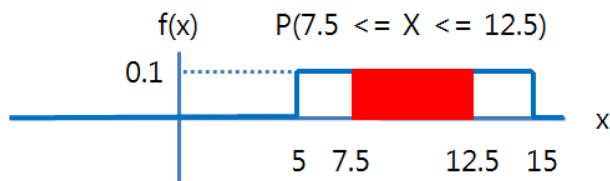
$$V(x) = E(x^2) - [E(x)]^2 = \frac{(b^2+ba+a^2)}{3} - \frac{a^2+2ab+b^2}{4} = \frac{(b-a)^2}{12}$$



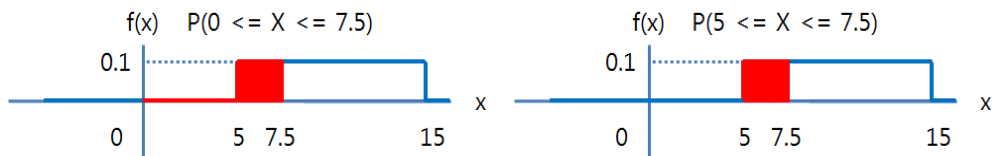
(2) $f(X=-2)$, $f(X=10)$, $f(X=15.2)$ 를 그래프 상에서 표기하시오.



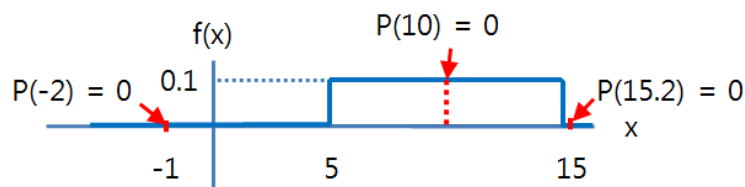
(3) $P(7.5 \leq X \leq 12.5)$ 를 그래프 상에서 표기하시오.



(4) $P(0 \leq X \leq 7.5)$ 와 $P(5 \leq X \leq 7.5)$ 를 그래프 상에서 표기하시오.



(5) $P(X=-2)$, $P(X=10)$, $P(X=15.2)$ 를 그래프 상에서 표기하시오.



연습문제 3. 확률변수 x 는 최소값이 -5, 최대값이 5 인 일양분포를 따른다.

문제 1. μ 와 σ^2 값

연습문제 4. Excel 에서 난수(random number)는 최소값이 0, 최대값이 1 인 일양분포를 따른다.

문제 1. μ

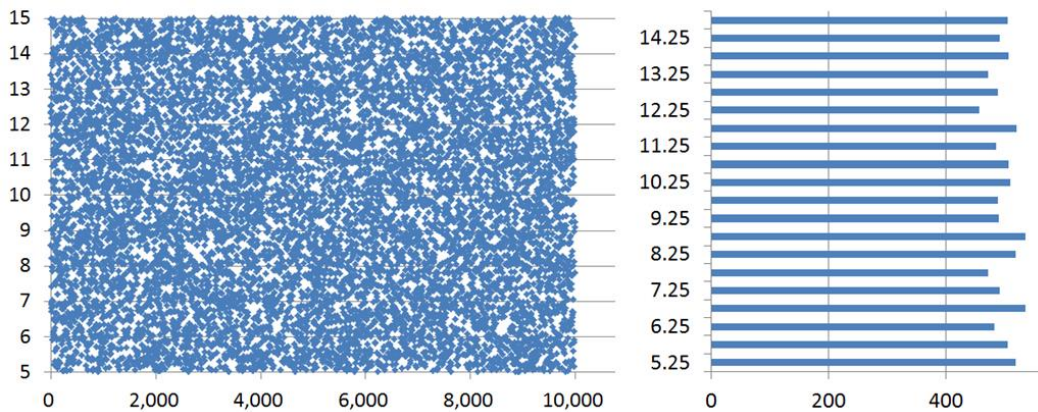
문제 2. σ^2

문제 3. $P(0.4 < x < 0.5)$

문제 4. $P(0.4 \leq x \leq 0.5)$

3.5.4 예제 [2].

데이터 10,000개의 분산도와 도수분포그래프를 그린 결과이다.



데이터 값은 다음과 같다.

번호	데이터
1	13.3973
2	8.5712
3	12.5207
4	12.3271
5	6.7949
...	...
10,000	12.1666

문제 1. 데이터값을 X 라 하자. X 는 어떤 분포를 따른다고 할 수 있는가?

Ans. 최소값이 5이고, 최대값이 15인 일양분포

문제 2. X 의 평균은 얼마로 추정되는가?

Ans. 10 (실제 실험에 사용된 10,000개의 평균은 10.0396이다.)

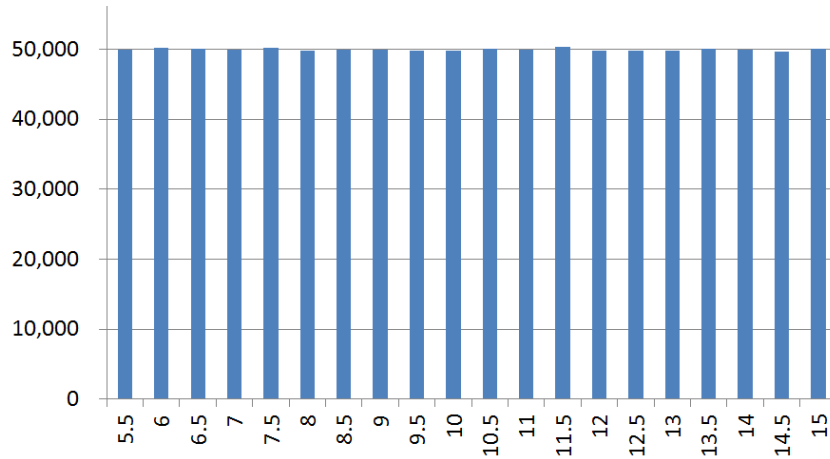
문제 3. $(13.3973 - \text{평균}), (8.5712 - \text{평균}), (12.5207 - \text{평균}), \dots, (12.1666 - \text{평균})$ 들의 평균은 얼마인가?

Ans. 편차의 합은 항상 0이므로, 편차의 평균도 0이다.

문제 4. $(13.3973 - \text{평균})^2, (8.5712 - \text{평균})^2, (12.5207 - \text{평균})^2, \dots, (12.1666 - \text{평균})^2$ 들의 평균은 얼마로 추정되는가?

Ans. 편차 제곱의 평균은 분산이다. 일양분포의 경우, 분산은 $\sigma^2 = \frac{(b-a)^2}{12} = 8.3333$ 이다. (실제 실험에 사용된 10,000개의 편차 제곱의 평균은 8.3173이다.)

데이터 수가 백만일 때의 데이터 평균, 분산, 도수분포그래프
 - 평균: 9.9987, 분산: 8.3352



4. 정규분포(normal distribution)

n 과 x 가 커지면 이항분포 ${}_nC_x p^x q^{n-x}$ 계산이 어려워진다.

De Moivre: 이항분포의 극한분포로 정규분포 발견 (1733년)

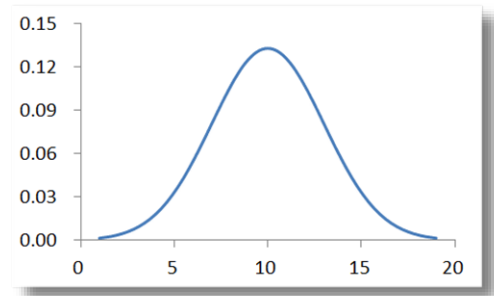
$${}_nC_x p^x q^{n-x} \simeq \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(x-np)^2}{2npq}} \quad (\text{이항분포 } \mu = np, \sigma^2 = npq)$$

4.1 정규분포의 확률밀도함수(probability density function)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{where } \sigma > 0. \quad \text{위 식의 } npq = \sigma^2 \text{로, } np \text{는 } \mu \text{로 표현한 것}$$

4.1.1 $\mu = 10, \sigma = 3$ 인 정규분포의 확률밀도함수 pdf

x	pdf	x	pdf
1.0	0.00148	10.6	0.13035
1.6	0.00264	11.2	0.12276
2.2	0.00453	11.8	0.11107
2.8	0.00746	12.4	0.09656
3.4	0.01182	13.0	0.08066
4.0	0.01800	13.6	0.06473
4.6	0.02632	14.2	0.04991
5.2	0.03697	14.8	0.03697
5.8	0.04991	15.4	0.02632
6.4	0.06473	16.0	0.01800
7.0	0.08066	16.6	0.01182
7.6	0.09656	17.2	0.00746
8.2	0.11107	17.8	0.00453
8.8	0.12276	18.4	0.00264
9.4	0.13035	19.0	0.00148
10.0	0.13298	19.6	0.00079



A1		f _x	=NORM.DIST(1, 10, 3, 0)
A	B	C	D
0.00148			

4.2 정규분포의 특성

- (1) parameter는 μ 와 σ^2 이다.

$X \sim N(\mu, \sigma^2)$ 확률변수 X 는 평균이 μ 이고 분산이 σ^2 인 정규분포를 따른다.

- (2) $f(x)$ 는 μ 를 중심으로 대칭인 bell shape이다.

- (3) $E(x) = \mu, V(x) = \sigma^2$.

- (4) $\int_{-\infty}^{\infty} f(x)dx = 1$.

- (5) $-\infty < x < \infty$

4.3 정규분포의 평균과 분산

$$E(x) = \mu$$

$$V(x) = \sigma^2$$

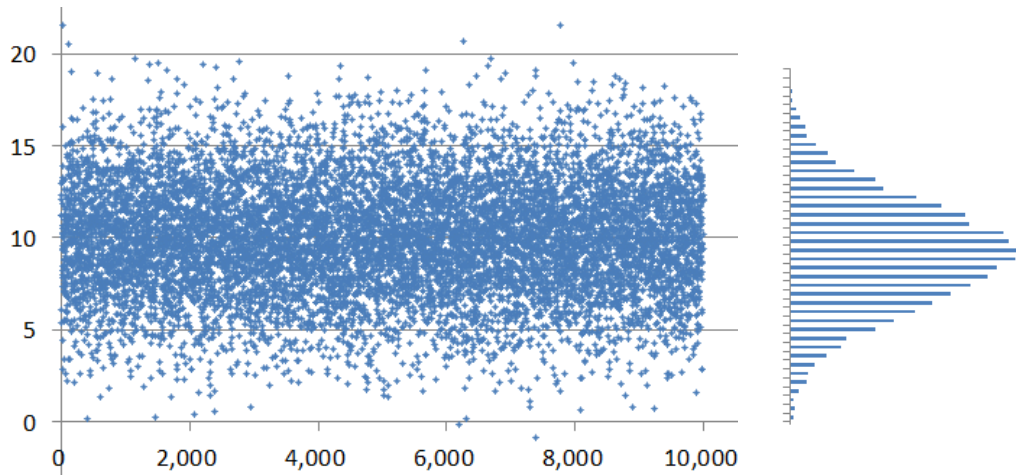
4.3.1 모의실험

데이터를 이용한 평균과 분산 추정 (x 는 $\mu = 10, \sigma = 3$ 인 정규분포 따름)

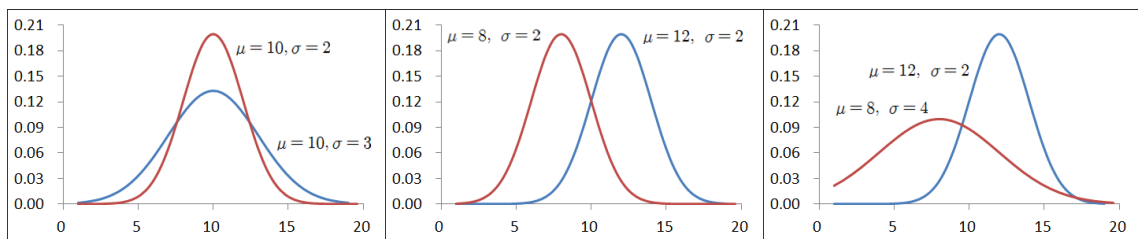
번호	값 x_i	편차 $(x_i - \mu)$	편차 제곱 $(x_i - \mu)^2$
1	1.33995	-8.66140	75.01991
2	11.39961	1.39826	1.95512
3	9.74808	-0.25328	0.06415
4	8.24693	-1.75442	3.07799
5	10.64904	0.64769	0.41951
6	2.68680	-7.31455	53.50260
7	10.80508	0.80372	0.64597
8	10.76176	0.76041	0.57823
9	9.66882	-0.33253	0.11058
10	11.69247	1.69112	2.85989
...
10,000	7.25732	-2.74403	7.52972
합계	100,013.51455	0.00000	90,149.13633
평균	10.00135 $\rightarrow E(X)$	0.00000	9.01491 $\rightarrow V(X)$

10,000개의 데이터 분석 결과, 평균(기대값)은 정규분포의 parameter인 μ 와 일치한다.

10,000개의 데이터 분석 결과, 분산은 정규분포의 parameter인 σ 의 제곱과 일치한다.



4.4 μ 와 σ 에 따른 정규분포의 모양



4.5 정규분포의 특징

μ 와 σ 값과는 무관하게 $P(x < \mu + k \cdot \sigma)$ 는 항상 일정하다.

예 1. (평균+1표준편차) 보다 작을 확률

평균 100, 표준편차 10인 정규분포 $P(x < 110) = P(x < 100 + 1 \cdot 10) = .8413$

평균 200, 표준편차 10인 정규분포 $P(x < 210) = P(x < 200 + 1 \cdot 10) = .8413$

평균 0, 표준편차 1인 정규분포 $P(x < 1) = P(x < 0 + 1 \cdot 1) = .8413$

예 2. (평균+2표준편차) 보다 작을 확률

평균 100, 표준편차 10인 정규분포 $P(x < 120) = P(x < 100 + 2 \cdot 10) = .9772$

평균 200, 표준편차 10인 정규분포 $P(x < 220) = P(x < 200 + 2 \cdot 10) = .9772$

평균 0, 표준편차 1인 정규분포 $P(x < 2) = P(x < 0 + 2 \cdot 1) = .9772$

정리하면, μ 와 σ 값이 무엇이든 정규분포를 따르는 확률변수 x 가

(평균 + 1·표준편차)보다 작을 확률은 항상 .8413이며,

(평균 + 2·표준편차)보다 작을 확률은 항상 .9772이다.

정규분포의 parameter는 μ 와 σ^2 이다. x 의 확률분포는 x 의 구체적인 값에 의해서만 결정되는 것이 아니라 x 값이 자신의 평균보다 몇 표준편차 큰가에 따라 결정된다.

4.6 표준정규분포(standard normal distribution)

4.6.1 z

표준화된 정규분포 확률변수(standardized normal random variable), $z \sim N(0,1)$

일반 정규분포 확률변수 x 를 z 로 환산하는 방법

$$z = \frac{x - \mu}{\sigma}$$

Note: ① $x - \mu$ 는 편차이다.

② $z = \frac{x - \mu}{\sigma}$ 는 x 가 평균보다 몇 표준편차 큰가를 의미한다.

$z = 1$ 이면 $x = \mu + 1 \cdot \sigma$ 이며, $z = -0.51$ 이면 $x = \mu - 0.5 \cdot \sigma$ 이다.

4.6.1.1 Example 1. $\mu = 1,000, \sigma = 50$

x	z	해석
-----	-----	----

850	$z = \frac{850-1,000}{50} = -3$	평균보다 3 표준편차 작은 값
900	$z = \frac{900-1,000}{50} = -2$	평균보다 2 표준편차 작은 값
950	$z = \frac{950-1,000}{50} = -1$	평균보다 1 표준편차 작은 값
1,000	$z = \frac{1,000-1,000}{50} = 0$	평균과 같은 값
1,050	$z = \frac{1,050-1,000}{50} = 1$	평균보다 1 표준편차 큰 값
1,100	$z = \frac{1,100-1,000}{50} = 2$	평균보다 2 표준편차 큰 값
1,150	$z = \frac{1,150-1,000}{50} = 3$	평균보다 3 표준편차 큰 값

4.6.1.2 Example 2.

$$(1) \quad \mu=100, \sigma=10 \text{인 정규분포 } P(x < 110) = P\left(\frac{x-\mu}{\sigma} < \frac{110-100}{10}\right) = P(z < 1)$$

$$\mu=200, \sigma=10 \text{인 정규분포 } P(x < 210) = P\left(\frac{x-\mu}{\sigma} < \frac{210-200}{10}\right) = P(z < 1)$$

$$\mu=0, \sigma=1 \text{인 정규분포 } P(x < 1) = P\left(\frac{x-\mu}{\sigma} < \frac{1-0}{1}\right) = P(z < 1)$$

$$(2) \quad \mu=100, \sigma=10 \text{인 정규분포 } P(x < 120) = P\left(\frac{x-\mu}{\sigma} < \frac{120-100}{10}\right) = P(z < 2)$$

$$\mu=200, \sigma=10 \text{인 정규분포 } P(x < 220) = P\left(\frac{x-\mu}{\sigma} < \frac{220-200}{10}\right) = P(z < 2)$$

$$\mu=0, \sigma=1 \text{인 정규분포 } P(x < 2) = P\left(\frac{x-\mu}{\sigma} < \frac{2-0}{1}\right) = P(z < 2)$$

4.6.2 표준정규분포의 평균과 분산

$x \sim N(\mu, \sigma^2)$ 라 가정: [확률변수 x 는 평균이 μ 이고 표준편차가 σ 인 정규분포를 따른다.]

$z = \frac{x-\mu}{\sigma}$ 의 평균과 분산은 다음과 같다.

$$E(z) = E\left(\frac{x-\mu}{\sigma}\right) = E\left(\frac{x}{\sigma} - \frac{\mu}{\sigma}\right) = E\left(\frac{x}{\sigma}\right) - \frac{\mu}{\sigma} = \frac{E(x)}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$V(z) = V\left(\frac{x-\mu}{\sigma}\right) = V\left(\frac{x}{\sigma} - \frac{\mu}{\sigma}\right) = V\left(\frac{x}{\sigma}\right) = \frac{1}{\sigma^2}V(x) = \frac{\sigma^2}{\sigma^2} = 1$$

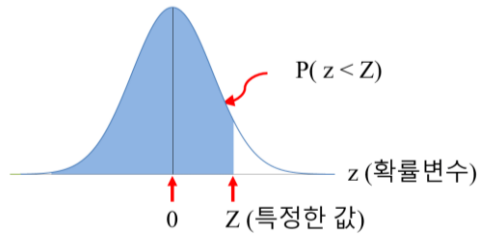
결론: $x \sim N(\mu, \sigma^2)$ 이 성립하면, $z \sim N(0, 1^2)$ 이다. $\rightarrow \mu_z = 0, \sigma_z^2 = 1$

4.6.3 x 와 z 와의 관계

Example $x \sim N(10, 2^2)$ 와 z 와의 관계

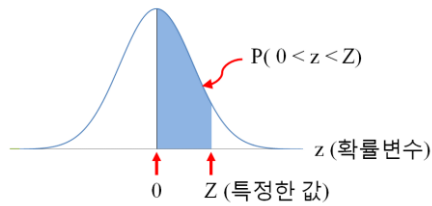
	x	z
1	10	0
2	11	0.5
3	13	1.5
4	8	-1
5	7	-1.5
6	9	-0.5
7	12	1

1. x 값이 크면, z 값도 크다. \rightarrow 크기 순서는 바뀌지 않는다.
2. 상대적인 거리도 동일하다.
 - (1) x : (13과 10의 차이)는 (11과 10의 차이)의 3배이다.
 z : (1.5와 0의 차이)는 (0.5와 0의 차이)의 3배이다.
 - (2) x : (12와 7의 차이)는 (9와 7의 차이)의 2.5배이다.
 z : (1과 -1.5의 차이)는 (-0.5와 -1.5의 차이)의 3배이다.
3. x 에 대응하는 오직 하나의 z 가 존재하고, z 에 대응하는 오직 하나의 x 가 존재한다.
4. z 는 x 를 평균만큼 이동하고 단위를 σ 로 조정한 값이다.

4.6.4 표준정규분포의 누적확률표 $P(z < Z)$ 

A1		f_x	=NORM.S.DIST(1.2, 1)
A	B	C	D
0.88493			

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
4.0	1.0000									

4.6.5 표준정규분포의 확률표 $P(0 \leq z \leq Z)$ 

A1		f_x	=NORM.S.DIST(1.2, 1)-0.5	
A	B	C	D	
0.38493				

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
4.0	0.5000									

4.6.5.1 표 읽는 법

$P(0 \leq z \leq Z) \rightarrow$ 여기서 확률변수는 z

		Z_2		
Z_1		$P(0 < z < [Z_1 + Z_2])$		

여기서 Z_1 은 Z 값 중 소숫점 1자리까지만 표기한 값, Z_2 는 Z 값 중 소숫점 2자리값만 표기한 값이다.

예) $P(0 \leq z \leq 1.28)$ 값을 구한다면,

- ① Z_1 은 1.2, Z_2 는 0.08이 된다.
- ② $P(0 \leq z \leq 1.28)$ 은 Z_1 이 1.2행과 Z_2 가 0.08인 열의 교차점에 있는 0.3997이다.

f_x	=NORM.S.DIST(1.28, 1)-0.5	
	C	D
	0.399727	

Excel에서 찾는다면,

예) $P(0 \leq z \leq Z) = 0.4750$ 을 성립시키는 Z 를 찾으려면,

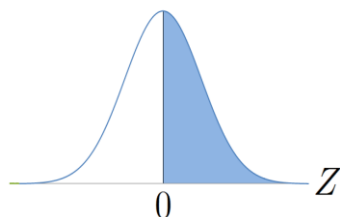
- ① 확률값 중 0.4750을 찾는다. 이 확률값은 $Z_1 = 1.9$ 인 행과 $Z_2 = 0.06$ 인 열의 교차점에 위치한다.
- ② 찾는 Z 값은 $Z_1 + Z_2 = 1.96$ 이다.

f_x	=NORM.S.INV(0.475+0.5)	
	C	D
	1.959964	

Excel에서 찾는다면,

4.7 정규분포의 응용

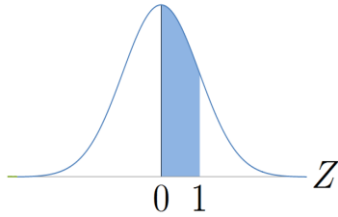
(1) $P(z \geq 0) = 0.5$... $z = 0$ 을 중심으로 좌우대칭형 $\rightarrow x$ 가 평균보다 클 확률



$Z = 0$ 상위 50%에 해당하는 값 $P(z > Z) = 0.5$

하위 50%에 해당하는 값 $P(z < Z) = 0.5$

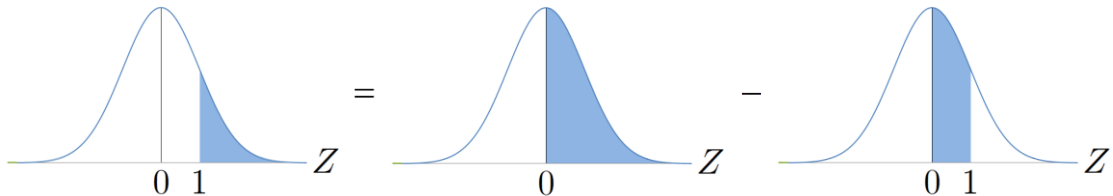
(2) $P(0 \leq z \leq 1) = 0.3413 \rightarrow x$ 가 평균과 평균+1·표준편차 사이에 값일 확률



$Z = 1$	상위 15.87%에 해당하는 값	$P(z > 1) = 0.1587$
	하위 84.13%에 해당하는 값	$P(z < 1) = 0.8413$
$Z = -1$	상위 84.13%에 해당하는 값	$P(z > -1) = 0.8413$
	하위 15.87%에 해당하는 값	$P(z < -1) = 0.1587$

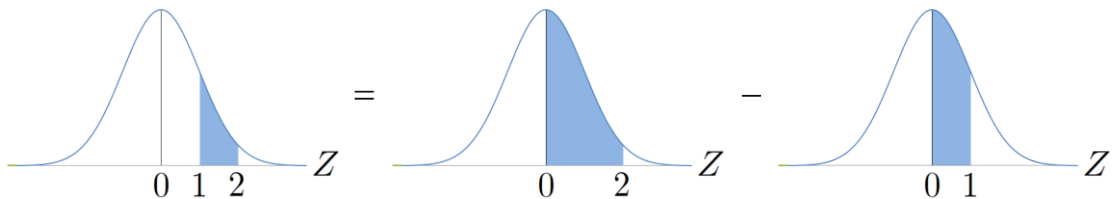
(3) $P(z \geq 1) = P(z \geq 0) - P(0 \leq z \leq 1) = 0.5 - 0.3413 = 0.1587$

$\rightarrow x$ 가 평균+1·표준편차보다 클 확률



(4) $P(1 \leq z \leq 2) = P(0 \leq z \leq 2) - P(0 \leq z \leq 1) = 0.4772 - 0.3413 = 0.1359$

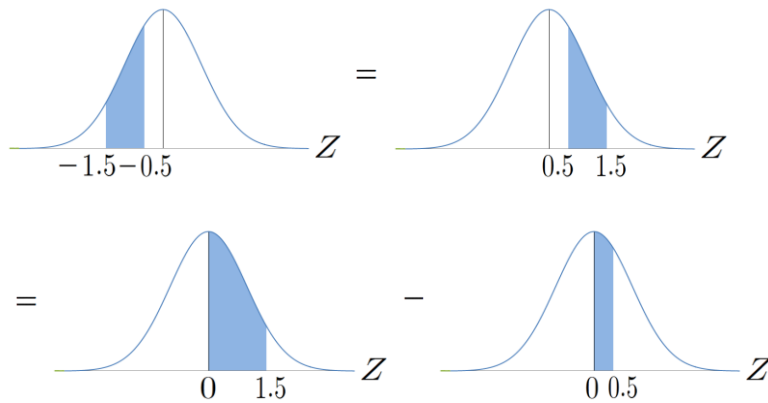
$\rightarrow x$ 가 평균+1·표준편차와 평균+2·표준편차 사이에 있을 확률



$Z = 2$	상위 2.28%에 해당하는 값	$P(z > 2) = 0.0228$
	하위 97.72%에 해당하는 값	$P(z < 2) = 0.9772$
$Z = -2$	상위 97.72%에 해당하는 값	$P(z > -2) = 0.9772$
	하위 2.287%에 해당하는 값	$P(z < -2) = 0.0228$

(5) $P(-1.5 \leq z \leq -0.5) = P(0.5 \leq z \leq 1.5) = P(0 \leq z \leq 1.5) - P(0 \leq z \leq 0.5)$

$\rightarrow x$ 가 평균보다 1.5 표준편차 작은 값과 평균보다 0.5 표준편차 작은 값 사이에 있을 확률



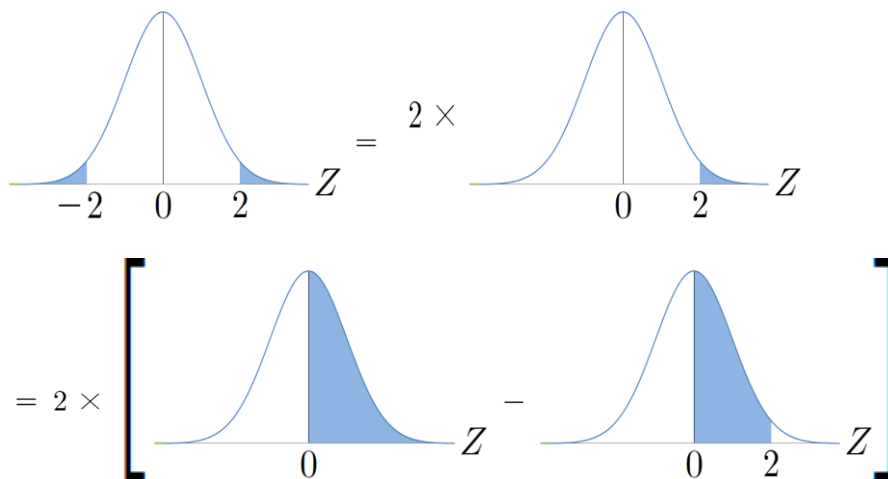
$Z = 1.5$	상위 6.68%에 해당하는 값 하위 93.32%에 해당하는 값	$P(z > 1.5) = 0.0668$ $P(z < 1.5) = 0.9332$
$Z = -1.5$	상위 93.32%에 해당하는 값 하위 6.68%에 해당하는 값	$P(z > -1.5) = 0.9332$ $P(z < -1.5) = 0.0668$
$Z = 0.5$	상위 30.82%에 해당하는 값 하위 69.15%에 해당하는 값	$P(z > 0.5) = 0.0228$ $P(z < 0.5) = 0.6915$
$Z = -0.5$	상위 69.15%에 해당하는 값 하위 30.85%에 해당하는 값	$P(z > -0.5) = 0.6915$ $P(z < -0.5) = 0.0228$

(6) z 가 -2보다 작거나 2보다 클 확률

$$= P(z \leq -2) + P(z \geq 2) = 2 \times P(z \geq 2)$$

$$= 2 \times \{0.5 - P(0 \leq z \leq 2)\}$$

→ x 가 평균보다 2 표준편차 작은 값과 평균보다 2표준편차 큰 값 사이에 있을 확률



(7) 다음을 만족시키는 Z 의 값은?

$$P(0 \leq z \leq Z) = 0.45 \rightarrow \text{표로부터 } Z = 1.645$$

상위 5%, 하위 95%에 해당하는 z 값은? : 1.645 (평균보다 1.645 표준편차 큰 값)

(8) 통계학 점수는 평균 80, 분산 100인 정규분포를 따르며, 수강인원은 100명이다.

8.1 어떤 학생이 80점에서 85점 사이의 점수를 받았을 확률은?

x : 정규분포확률변수, μ : 평균, σ : 표준편차

$$\begin{aligned} P(80 \leq x \leq 85) &= P\left(\frac{80-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{85-\mu}{\sigma}\right) = P\left(\frac{80-80}{10} \leq \frac{x-\mu}{\sigma} \leq \frac{85-80}{10}\right) \\ &= P(0 \leq z \leq 0.5) = 0.1915 \end{aligned}$$

① $\mu = 80$, $\sigma^2 = 100$ 일 때, $X = 80$ 과 85에 해당하는 z 값을 구한다.

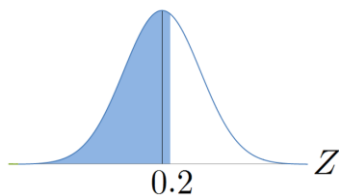
② 학생이 80점에서 85점 사이의 점수를 받았을 확률을 Z version으로 변환한다.
 $\rightarrow P(0 \leq z \leq 0.5)$

8.2 82점 이하를 받은 학생의 수

$$\begin{aligned} P(x \leq 82) &= P\left(\frac{x-\mu}{\sigma} \leq \frac{82-\mu}{\sigma}\right) = P(z \leq 0.2) \\ &= 0.5 + P(0 \leq z \leq 0.2) = 0.5793 \end{aligned}$$

① $\mu = 80$, $\sigma^2 = 100$, $x = 82$ 이면 $z = 0.2$ 에 해당한다.

② $z = 0.2$ 는 하위 57.93%에 해당한다.



$$\text{예상 학생수} = \text{총 학생수} \times 82\text{점 이하 받을 확률} = 100 \times 0.5793$$

4.7.1 암기 필요한 z 값

z	상위
1.00	0.1587
1.28	0.1000
1.645	0.0500
1.96	0.0250
2.00	0.0228
3.00	0.0013

4.7.2 예제

문제 1. $x \sim N(50, 3^2)$, x 를 z 로 변환하고 확률값을 제시하시오.

- (1) $P(x > 53)$ Ans. $P(z > 1) = .1587$
 (2) $P(x < 53)$ Ans. $P(z < 1) = .8413$
 (3) $P(50 < x < 53)$ Ans. $P(0 < z < 1) = .3413$
 (4) $P(x > 55.88)$ Ans. $P(z > 1.96) = .025$
 (5) $P(x < 45.065)$ Ans. $P(z < -1.645) = .050$

문제 2. 전세계적으로 동시에 실시되는 시험이 있는데, 시험 점수는 평균이 200, 표준편차는 30인 정규분포를 따른다고 한다. 백만명의 점수는 다음과 같다.

번호	점수
1	186
2	203
3	205
4	197
...	...
1,000,000	197

- (1) i 번째 응시자의 점수를 X_i 라 하자. X_1 부터 $X_{1,000,000}$ 의 (예상) 평균은?
 Ans. 200 (모의실험 자료의 평균은 199.4164)
 (2) $(X_1\text{-평균}), (X_2\text{-평균}), \dots, (X_{1,000,000}\text{-평균})$ 들의 (예상) 평균은?
 Ans. 0 (편차의 합은 항상 0, 그러므로 편차의 평균도 항상 0)
 (3) $(X_1\text{-평균})^2, (X_2\text{-평균})^2, \dots, (X_{1,000,000}\text{-평균})^2$ 들의 (예상) 평균은?
 Ans. 편차 제곱의 평균은 곧 분산이다. 분산 예상값은 (표준편차의 제곱)이므로 900 (모의실험 자료의 분산은 900.0505)
 (4) 해당 자료가 200과 230 사이에 있으면 1, 200 미만이거나 230보다 크면 0이라는 값을 기입하시오.

번호	점수	$200 \leq X_i \leq 230$
1	186	
2	203	
3	205	
4	197	
...
1,000,000	197	

점수	$200 \leq X_i \leq 230$
186	0
203	1
205	1
197	0
...	...
197	0

Ans.

- (5) (4)의 3번째 열의 값들의 (예상)합계는?

$$\text{Ans. } P(200 \leq X \leq 230) = P\left(\frac{200-200}{30} \leq \frac{X-\mu}{\sigma} \leq \frac{230-200}{30}\right) = P(0 \leq z \leq 1) = .3413$$

각각의 값들이 1일 확률이 .3413이므로, 백만개 중 1인 예상 갯수는 341,300 (실제 백만 개의 데이터 중에 해당되는 데이터는 348,985)

(6) (4)의 3번째 열의 값들의 평균은?

$$\text{Ans. } \frac{341,300}{1,000,000} = .3413$$

실험의 경우, 발생 건수는 348,985, 총 관찰횟수는 1,000,000, 상대빈도는 .34895 → 실험의 횟수가 증가하면 이 값은 .3413에 수렴한다. (이 값이 바로 상대빈도정의에 의한 확률값이다.)

연습문제 5.

문제 1. $P(z \geq 0)$

문제 2. $P(z \leq 0)$

문제 3. $P(0 \leq z \leq 1)$

문제 4. $P(z \geq 1)$

문제 5. $P(0 \leq z \leq 1.645)$

문제 6. $P(z \geq 1.645)$

연습문제 6. 학생 성적은 평균이 100, 표준편차가 10인 정규분포를 따른다고 하자. 다음 확률값을 구하시오.

문제 1. 어떤 학생 점수를 보았을 때 그 점수가 100 점 이상일 확률

문제 2. 어떤 학생 점수를 보았을 때 그 점수가 100 점 이하일 확률

문제 3. 어떤 학생 점수를 보았을 때 그 점수가 100 점에서 110 점 사이에 있을 확률

문제 4. 어떤 학생 점수를 보았을 때 그 점수가 110 점 이상일 확률

5. 지수분포(exponential distribution)

지수분포의 예: 매장에서 고객 도착시간 간격, 환자 치료 시간 등 한 사건 발생 후 다음 사건이 발생할 때까지의 시간 분포

지수분포는 수명 또는 대기시간 등에 관한 확률모형에 사용된다.

일반적으로 $x \sim \text{Exp}(\lambda)$ 로 표현

5.1 포아송분포와 지수분포

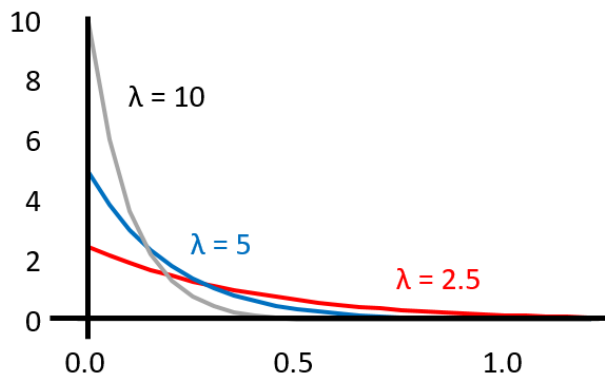
포아송 분포에서 단위 시간당 평균 발생 수가 λ 이면, 한 사건 발생 후 다음 사건이 일어날 때까지의 시간(대기시간)은 지수분포를 따르며 평균은 $\frac{1}{\lambda}$ 가 된다.

5.2 지수분포의 확률밀도함수

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

여기서, x = 다음 사건이 발생할 때까지의 (대기)시간, λ = 단위시간당 사건의 발생 수

5.3 지수분포 확률밀도함수 특징



- (1) $0 \leq x \leq \infty$
- (2) $f(x) = 0$ 에서 최대가 된다. $f(x=0) = \lambda$
 x 가 증가할수록 감소한다. (오른쪽 꼬리)
- (3) λ 가 증가할수록 초기 감소폭이 크다.

5.4 지수분포의 평균과 분산

$$E(x) = \frac{1}{\lambda}$$

3

$$\begin{aligned} E(x) &= \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx \\ \lambda x &= t \text{라고 하면 } \lambda dx = dt \\ \int_0^{\infty} \frac{t}{\lambda} \lambda e^{-t} \frac{1}{\lambda} dt &= \int_0^{\infty} t e^{-t} \frac{1}{\lambda} dt = \frac{1}{\lambda} \int_0^{\infty} t e^{-t} dt \\ &= \frac{1}{\lambda} [-e^{-t}(t+1)]_0^{\infty} = \frac{1}{\lambda} (0 - (-1)) = \frac{1}{\lambda} \end{aligned}$$

$$V(x) = \frac{1}{\lambda^2}$$

5.5 지수분포 누적분포함수

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

여기서, x = 다음 사건이 발생할 때까지의 (대기)시간, λ = 단위시간당 사건의 발생 수

5.6 Example.

어떤 매장의 시간당 도착하는 고객의 수는 포아송분포를 따르고, 평균 4.0이다.

(1) 고객 도착 간의 시간이 30분 이내일 확률을 구하시오.

$$\text{Ans. } P\left(x < \frac{30}{60}\right) = F(x = 0.5) = 1 - e^{-4.0 \times 0.5} = 1 - e^{-2.0} = 0.8647$$

(2) 최소 30분 이후에 다음 고객이 도착할 확률은?

$$\text{Ans. } P\left(x > \frac{30}{60}\right) = 1 - P(x < 0.5) = 1 - F(x = 0.5) = 1 - [1 - e^{-4.0 \times 0.5}] = 0.1353$$

연습문제 7. 어떤 창구의 고객 응대 시간은 지수분포를 따른다고 한다. 평균 응대 시간은 15분이라 알려져 있다.

문제 1. 고객 응대시간이 10 분 이내일 확률을 구하시오.

문제 2. 고객 응대시간이 20 분 이내일 확률을 구하시오.

$$\begin{aligned} E(x^2) &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2} \int_0^{\infty} t^2 e^{-t} dt \\ &= \frac{1}{\lambda^2} [-e^{-t}(t^2 + 2t + 2)]_0^{\infty} = \frac{1}{\lambda^2} (0 - (-2)) = \frac{2}{\lambda^2} \\ V(x) &= E(x^2) - [E(x)]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \end{aligned}$$

문제 3. 고객 응대시간이 10 분에서 20 분 사이일 확률을 구하시오.

문제 4. 고객 응대시간이 20 분을 초과할 확률을 구하시오.

5.7 포아송분포에서 지수분포 유도

5.7.0 [사전 학습] 포아송 분포, 기간 t 동안 발생 건수

5.7.0.1 (포아송분포) 단위 시간당 평균 발생 건수가 λ 일 때, (단위 시간 동안) 발생 건수가 y 일 확률

$$P(y | \lambda) = \frac{e^{-\lambda} \cdot \lambda^y}{y!}, \quad y = 0, 1, 2, \dots$$

5.7.0.2 단위 시간당 평균 발생 건수가 λ 일 때, 기간 t 동안의 평균 발생 건수 $= \lambda t$

5.7.0.3 기간 t 동안 발생 건수가 y 일 확률

$$P(y | \lambda t) = \frac{e^{-\lambda t} (\lambda t)^y}{y!}$$

5.7.1 지수분포 유도

5.7.1.1 단위시간당 평균 발생횟수가 λ 인 포아송분포에서 어떤 사건이 처음 발생하기 까지 걸린 시간을 t 라 하자.

5.7.1.2 t 시간 만에 처음 사건이 발생했으므로, 시점 0부터 시점 t 동안 (시점 0은 포함, t 는 미포함; 총 기간은 t) 발생횟수는 0이 된다. 발생횟수를 y 로 표현하면, 이 확률은 아래와 같다.

$$P(y = 0 | \lambda t) = \frac{e^{-\lambda t} (\lambda t)^0}{0!} = e^{-\lambda t}$$

5.7.1.3 어떤 사건이 처음 발생할 때까지 걸린 시간을 확률변수 x 라 하자. 이 확률변수 x 가 시간 t 를 초과할 확률은 다음과 같다.

$$P(x > t) = e^{-\lambda t}$$

5.7.1.4 확률변수 x 에 대한 누적분포함수는 다음과 같다

$$P(0 \leq x \leq t) = F(t) = 1 - e^{-\lambda t}$$

5.7.1.5 누적분포함수를 미분하면 확률밀도함수가 되므로 확률밀도함수는 다음과 같다.

$$f(x = t) = \frac{dF(t)}{dt} = \frac{d}{dt}(1 - e^{-\lambda t}) = \lambda e^{-\lambda t}$$

연습문제 정답

1. (1) 0 (2) 0 (3) 0 (4) 0 (5) 0.1 (6) 0 (7) 0 (8) False
2. (1) 0 (2) 0 (3) 0.5 (4) 0.5 (5) 1.0 (6) 1.0 (7) 0 (8) 0.5
3. (1) $\mu = 0, \sigma^2 = 8.33$
4. (1) 0.5 (2) 0.0833 (3) 0.1
 (4) 0.1 연속확률변수에서 구간의 확률값을 측정할 때 = 사인은 무의미하다.
 연속확률분포에서 어느 특정한 값일 확률은 항상 0이다.
 그러므로 $0.4 < x < 0.5$ 구간의 확률이나 $0.4 \leq x \leq 0.5$ 구간의 확률이나 동일하다.
5. (1) 0.5 (2) 0.5 (3) 0.3413 (4) 0.1587 (5) 0.450 (6) 0.050
6. (1) 연습문제 5의 문제 1 (2) 연습문제 5의 문제 2 (3) 연습문제 5의 문제 3
 (4) 연습문제 5의 문제 4와 동일
7. (1) 0.4866 (2) 0.7364 (3) 0.2498 (4) 0.2636