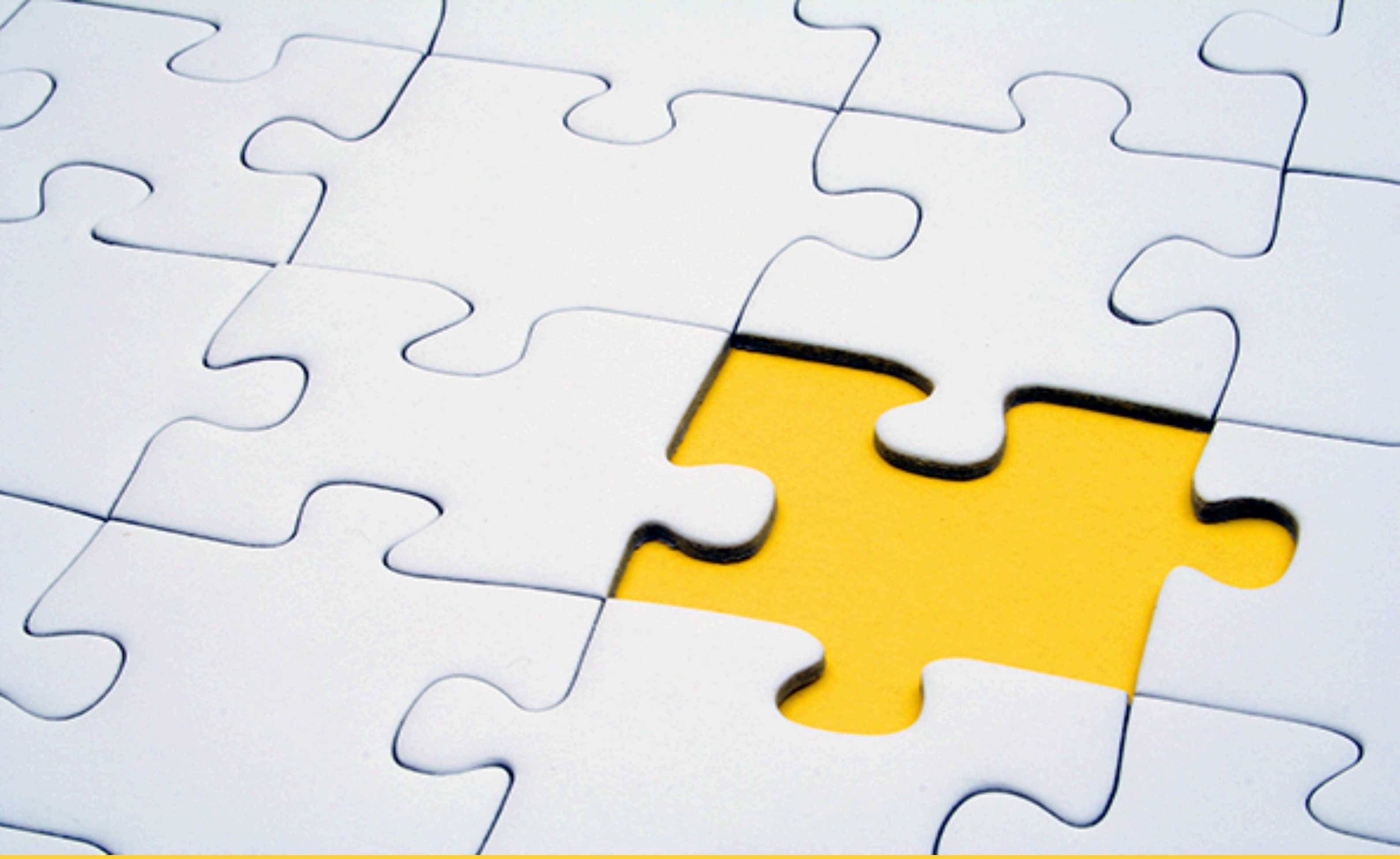


고등학교

확률과 통계

수악중독

고등학교 확률과 통계

경우의 수

1

1. 순열과 조합
2. 이항정리

1 원순열

원순열

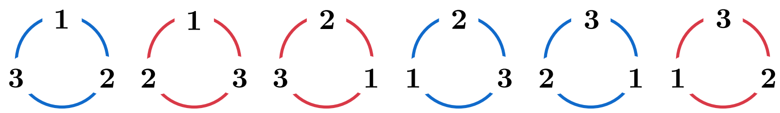
n개의 서로 다른 것을 원형으로 배열하는 경우의 수는 다음과 같다.

$$n! \times \frac{1}{n} = (n-1)!$$

➤ 1, 2, 3을 일렬로 배열하는 것은 다음과 같이 6가지가 가능하다.

$$(1,\ 2,\ 3),\ (1,\ 3,\ 2),\ (2,\ 1,\ 3),\ (2,\ 3,\ 1),\ (3,\ 1,\ 2),\ (3,\ 2,\ 1)$$

위의 서로 다른 6가지를 원형으로 배열하면 아래 그림과 같다.



이 중에서 첫 번째, 네 번째, 다섯 번째와 두 번째, 세 번째, 여섯 번째는 서로 상대적인 위치가 같다. 이와 같이 회전하여 일치하는 경우는 모두 같은 것으로 보는 순열을 원순열이라고 한다. 위 여섯 가지 경우 중 세 가지 씩 겹치므로 원순열의 수는 $\frac{3!}{3} = 2!$이 된다.

➤ 일반적으로 서로 다른 n개를 원형으로 배열할 때, 회전하여 일치하는 것이 n개 씩 생기므로 원순열의 수는 순열의 수의 $\frac{1}{n}$이 된다.

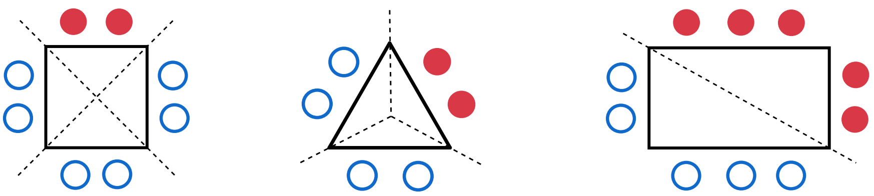
$$(\text{원순열의 수}) = n! \times \frac{1}{n} = (n-1)!$$

➤ 원형이 아닌 다각형인 경우(대칭인 경우)의 순열의 수

(1) 정사각형의 식탁에 n명을 앉히는 방법의 수 $= (n-1)! \times \frac{n}{4}$ (단, $n = 4k$, k는 자연수)

(2) 정삼각형의 식탁에 n명을 앉히는 방법의 수 $= (n-1)! \times \frac{n}{3}$ (단, $n = 3k$, k는 자연수)

(3) 직사각형의 식탁에 n명을 앉히는 방법의 수 $= (n-1)! \times \frac{n}{2}$ (단, $n = 2k$, k는 자연수)



예제1

여섯 명의 학생 중 네 명의 학생을 뽑아 원탁에 앉히는 경우의 수를 구하시오.

$${}_6\mathrm{P}_4 \times \frac{1}{4} = 360 \times \frac{1}{4} = 90$$

예제2

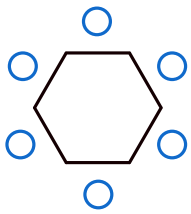
여섯 명의 학생 A, B, C, D, E, F가 원형의 탁자에 둘러 앉을 때, A, B, C가 이웃하여 앉는 경우의 수를 구하시오.

세 명의 학생 A, B, C를 한 묶음으로 생각하면, 모두 네 명을 원형으로 배열하는 원순열이 되므로 그 경우의 수는 $(4-1)! = 6$이다. 이때, 이 6가지 각각의 경우에 대하여 A, B, C가 자리를 바꾸어 앉는 경우의 수가 $3! = 6$이다. 따라서 구하는 경우의 수는 $6 \times 6 = 36$이 된다.

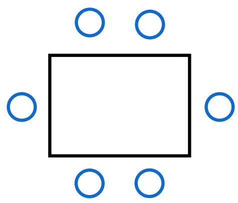
예제3

다음 그림과 같은 모양의 탁자에 여섯 명이 둘러 앉는 경우의 수를 구하시오.

(1)



(2)



(1) $(6-1)! = 5! = 120$

(2) $(6-1)! \times \dfrac{6}{2} = 5! \times 3 = 360$

2 중복순열

중복순열

서로 다른 n개에서 중복을 허락하여 r개를 택하는 중복순열의 수는 다음과 같다.

$$ {}_n\Pi_r = n^r $$

➤ 서로 다른 n개에서 중복을 허락하여 r개를 뽑아 일렬로 나열할 때, 첫 번째 자리에 올수 있는 것은 n개이고, 두 번째 자리에 올 수 있는 것도 n개다. 이렇게 생각하면 각 자리에 올 수 있는 것의 개수가 모두 n개이므로 다음이 성립한다.

$$ {}_n\Pi_r = \underbrace{n \times n \times \cdots \times n}_{r\text{개}} = n^r $$

예제4

$1,\ 2,\ 3,\ 4$의 네 숫자를 한 번씩만 사용하여 만들 수 있는 세 자리 정수의 개수를 α, 중복 사용하여 만들 수 있는 세 자리 정수의 개수를 β라고 할 때, $\beta - \alpha$의 값을 구하시오.

$\alpha = {}_4\mathrm{P}_3 = 4 \times 3 \times 2 = 24$

$\beta = {}_4\Pi_3 = 4 \times 4 \times 4 = 64$

$\therefore \beta - \alpha = 40$

예제5

$0,\ 1,\ 2,\ 3,\ 4$의 다섯 숫자를 중복 사용하여 만들 수 있는 네 자리 자연수의 개수를 구하시오.

천의 자리에는 0을 제외한 $1,\ 2,\ 3,\ 4$만 올 수 있으므로 경우의 수는 4가지다.
백의 자리, 십의 자리, 일의 자리에는 $0,\ 1,\ 2,\ 3,\ 4$가 모두 올 수 있으므로 경우의 수는 ${}_5\Pi_3 = 5^3 = 125$가지다.

따라서 만들 수 있는 네 자리 자연수의 개수는 $4 \times 125 = 500$개다.

3 같은 것이 있는 순열

같은 것이 있는 순열

n개 중 p개, q개, r개, $\cdots$, s개가 각각 같은 것일 때, 이들을 일렬로 나열하는 순열의 수는 다음과 같다.

$$\frac{n!}{p! \times q! \times r! \times \cdots \times s!} \quad (\text{단}, p+q+r+\cdots+s=n)$$

➤ 네 개의 알파벳 $a,\ a,\ b,\ c$를 일렬로 나열하는 경우의 수

$a,\ b,\ c,\ d$ 네 개의 알파벳을 일렬로 나열하는 경우의 수는 $4!$가지인데, 만약 d가 a로 바뀌게 되면 $a,\ d$가 서로 자리를 바꾸는 경우의 수만큼 중복이 된다. 예를 들어, $c,\ a,\ b,\ d$와 $c,\ d,\ b,\ a$는 서로 다르지만, d가 a로 바뀌게 되면 둘 다 $c,\ a,\ b,\ a$가 되어 중복이 된다. 즉, $a,\ b,\ c,\ d$를 일렬로 나열하는 경우의 수 $4!$을 $a,\ d$가 서로 자리를 바꾸는 경우의 수인 $2!$로 나누어야 한다. 따라서 구하는 경우의 수는 $\dfrac{4!}{2!}=12$이다.

➤ 네 개의 알파벳 $a,\ a,\ b,\ b$를 일렬로 나열하는 경우의 수

$a,\ b,\ c,\ d$를 일렬로 나열한 후, c는 a로 바꾸고 d는 b로 바꾸었다고 생각하면, $a,\ c$가 서로 자리를 바꾸는 경우의 수 $2!$ 만큼 중복되고, 동시에 $b,\ d$가 서로 자리를 바꾸는 경우의 $2!$ 만큼 중복된다. 따라서 구하는 경우의 수는 $\dfrac{4!}{2! \times 2!}=6$이다.

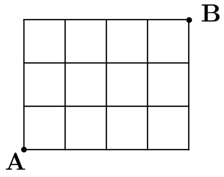
예제6

tomorrow의 8개 알파벳을 일렬로 나열하는 경우의 수를 구하시오.

8개의 알파벳 중에 o가 세 개, r이 두 개 있으므로 구하는 경우의 수는 $\dfrac{8!}{2! \times 3!}=3360$이다.

예제7

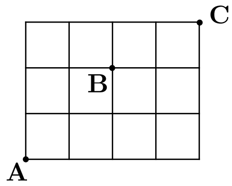
아래 그림과 같은 도로망이 있다. A지점에서 B지점으로 가는 최단 경로의 수를 구하시오.



A에서 B까지 최단 경로로 가려면 오른쪽이나 위쪽 방향으로만 이동해야 한다. 결과적으로는 오른쪽으로 4번, 위쪽으로 3번 이동해야 하는데 오른쪽으로 이동하는 것을 r, 위쪽으로 이동하는 것을 u로 나타내면, 최단 경로의 수는 $r,\ r,\ r,\ r,\ u,\ u,\ u$를 나열하는 경우의 수와 같게 된다. 따라서 구하는 최단 경로의 수는 $\dfrac{7!}{4! \times 3!} = 35$이다.

예제8

아래 그림과 같이 바둑판 모양의 도로망이 있다. A지점에서 B지점을 거쳐 C지점까지 가는 최단 경로의 수를 구하시오.



A $\rightarrow$ B로 가는 최단 경로의 수 : $\dfrac{4!}{2! \times 2!} = 6$

B $\rightarrow$ C로 가는 최단 경로의 수 : $\dfrac{3!}{2!} = 3$

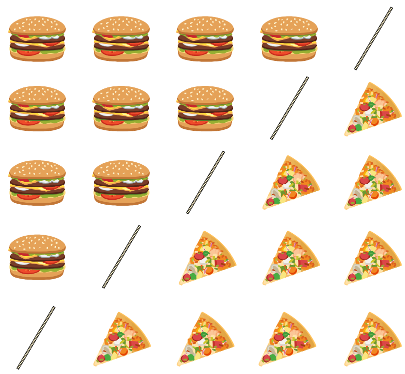
따라서 구하는 최단 경로의 수는 $6 \times 3 = 18$이다.

4 중복조합

중복조합

서로 다른 n개에서 r개를 택하는 중복조합의 수는 다음과 같다.

$$_{n}\mathrm{H}_{r} = {}_{n+r-1}\mathrm{C}_{r}$$



➤ 패스트푸드점에 네 명의 손님이 와서 햄버거와 피자를 주문하는 경우, 주방 알바의 입장에서 조리해야 하는 음식의 조합이 몇 가지가 가능하지를 생각보면 이해하기 쉽다. 예를 들어, 4개의 빈 쟁반과 빨대 하나를 나열한 후, 빨대 왼쪽의 쟁반에는 햄버거를 조리해 담고, 빨대 오른쪽의 쟁반에는 피자를 조리해 담는다고 생각하면 가능한 경우의 수는 5가지임을 알 수 있다. 따라서 구하는 경우의 수는 5자리 중에서 빨대가 놓일 자리 한 개를 선택하는 경우의 수 $_{5}\mathrm{C}_{1}$과 같다. 혹은 5자리 중에서 쟁반이 들어갈 4자리를 선택하는 경우의 수 $_{5}\mathrm{C}_{4}$와 같다고 생각해도 된다. 즉, 서로 다른 n개에서 r개를 택하는 중복조합의 수는 r개의 빈 쟁반과 $(n-1)$개의 빨대를 일렬로 나열하는 경우의 수와 같다. 결국 전체 $(n+r-1)$자리 중에서 빈 쟁반을 놓을 r자리를 택하는 조합의 수와 같아진다. 따라서 $_{n}\mathrm{H}_{r} = {}_{n+r-1}\mathrm{C}_{r}$이 된다.

예제9

다항식 $(x+y+z)^3$을 전개해서 생기는 서로 다른 항의 개수를 구하시오.

$(x+y+z)^3$을 전개해서 생기는 각 항들은 다음과 같은 형태를 갖는다.

$$x^3,\ x^2y,\ x^2z,\ \cdots,\ yz^2,\ z^3$$

위에 나열된 항들은 문자의 차수의 합이 3이라는 공통점이 있다. 또한 곱셈에서는 교환법칙이 성립하기 때문에 곱해지는 순서는 고려하지 않는다. 결국 구하는 항의 개수는 $x,\ y,\ z$ 중 순서를 고려하지 않고, 중복을 허락하여 3개를 택하는 중복조합의 수와 같다. 따라서 다항식 $(x+y+z)^3$을 전개해서 생기는 서로 다른 항의 개수는 $_{3}\mathrm{H}_{3} = {}_{3+3-1}\mathrm{C}_{3} = {}_{5}\mathrm{C}_{3} = 10$개다.

예제10

방정식 $x+y+z=10$의 음이 아닌 정수해의 개수를 구하시오.

$x=2,\ y=3,\ z=5$는 등식 $x+y+z=10$을 만족한다. 이것을

$$\underbrace{xx}_{2\text{개}}\overbrace{yyy}^{3\text{개}}\underbrace{zzzzz}_{5\text{개}}$$

와 같이 나타내면, 주어진 방정식을 만족시키는 음이 아닌 정수해의 개수는 서로 다른 세 개의 문자 $x,\ y,\ z$ 중에서 10개를 택하는 중복조합의 수와 같다고 볼 수 있다. 따라서 구하는 해의 개수는 ${}_3\mathrm{H}_{10}={}_{3+10-1}\mathrm{C}_{10}={}_{12}\mathrm{C}_{10}=66$개다.

예제11

집합 $X=\{1,\ 2,\ 3\}$, $Y=\{1,\ 2,\ 3,\ 4,\ 5\}$에 대하여 X에서 Y로의 함수 f에 대하여 다음 물음에 답하시오.

(1) X의 임의의 원소 $x_1,\ x_2$에 대하여 $x_1<x_2$이면 $f(x_1)<f(x_2)$를 만족하는 함수 f의 개수를 구하시오.

(2) X의 임의의 원소 $x_1,\ x_2$에 대하여 $x_1<x_2$이면 $f(x_1)\leq f(x_2)$를 만족하는 함수 f의 개수를 구하시오.

(1) 공역의 원소 1, 2, 3, 4, 5 중 순서를 고려하지 않고 3개를 선택하는 조합의 수와 같으므로 조건을 만족하는 함수 f의 개수는 ${}_5\mathrm{C}_3=10$개다.

(2) 공역의 원소 $1,\ 2,\ 3,\ 4,\ 5$ 중 순서를 고려하지 않고, 중복을 허락하여 3개를 선택하는 중복조합의 수와 같으므로 조건을 만족하는 함수 f의 개수는 ${}_5\mathrm{H}_3={}_{5+3-1}\mathrm{C}_3={}_7\mathrm{C}_3=35$개다.

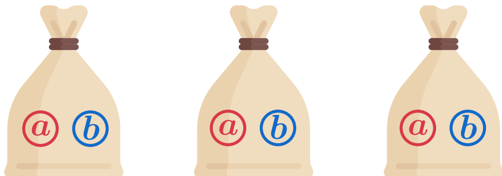
1 이항정리

이항정리

n이 자연수일 때, 다음이 성립한다.

$$(a+b)^n = {}_n\mathrm{C}_0 a^n + {}_n\mathrm{C}_1 a^{n-1}b + \cdots + {}_n\mathrm{C}_n b^n = \sum_{r=0}^{n} {}_n\mathrm{C}_r a^{n-r} b^r$$

➤ 다항식 $(a+b)^3 = (a+b)(a+b)(a+b)$의 전개식에서 a^2b항은 3개의 인수 $a+b$, $a+b$, $a+b$ 중 어느 하나의 인수에서 b를 택하고, 나머지 두 인수에서 각각 a를 택하여 곱한 것과 같다. 이런 상황을 다음과 같이 생각할 수 있다.
세 개의 주머니 A, B, C에 a가 표시된 공과 b가 표시된 공이 각각 하나씩 들어 있다고 한다.



주머니 **A** 주머니 **B** 주머니 **C**

각 주머니에서 공을 하나씩 꺼내면 aaa, aab, abb, bbb의 네 가지 결과가 있다는 것을 알 수 있다. 이때, 각 주머니에서 뽑힌 공에 적힌 문자들을 곱하면 4가지의 항 a^3, a^2b, ab^2, b^3을 만들 수 있고, 이 과정이 $(a+b)^3$을 전개할 때 각각의 항이 생성되는 원리와 같다는 것을 알 수 있다. 이 과정을 표로 정리하면 다음과 같다.

A	a	a	a	b	a	b	b	b
B	a	a	b	a	b	a	b	b
C	a	b	a	a	b	b	a	b
문자의 곱	a^3	a^2b	a^2b	a^2b	ab^2	ab^2	ab^2	b^3

위 표를 보면 a^2b가 나오는 경우는 세 가지가 있고, 이것은 $(a+b)^3$를 전개식할 때 a^2b항이 세 번 등장한다는 것을 의미한다. 즉, $(a+b)^3$의 전개식에서 a^2b항의 계수가 3이 되는 것이다. 결국 $(a+b)^3$의 전개식에서 a^2b의 계수는 세 개의 주머니에서 b가 적힌 공이 나오는 주머니 한 개를 선택하는 경우의 수 ${}_3\mathrm{C}_1 = 3$로 생각할 수 있다. 마찬가지로 b^3의 계수는 세 개의 주머니에서 b가 적힌 공이 나오는 주머니 세 개를 선택하는 경우의 수 ${}_3\mathrm{C}_3 = 1$로 생각할 수 있다. 이런 식으로 생각하면 다음 식이 성립함을 알 수 있다.

$$(a+b)^3 = {}_3\mathrm{C}_0 a^3 + {}_3\mathrm{C}_1 a^2 b + {}_3\mathrm{C}_2 ab^2 + {}_3\mathrm{C}_3 b^3$$

➤ 이것을 일반적으로 확장시키면 $(a+b)^n$ 에서 $a^{n-r}b^r$ 의 계수는

$$(a+b)^n = \underbrace{(a+b)(a+b)\cdots(a+b)}_{n\text{개의 }(a+b)}$$

에서 우변 $(a+b)(a+b)\cdots(a+b)$ 의 n 개의 인수 중 b 가 선택되는 인수 r 개를 택하는 경우의 수 ${}_n\mathrm{C}_r$ 와 같다는 것을 알 수 있다. 따라서 다음과 같은 전개식을 얻을 수 있다.

$$(a+b)^n = {}_n\mathrm{C}_0 a^n + {}_n\mathrm{C}_1 a^{n-1}b + \cdots + {}_n\mathrm{C}_n b^n = \sum_{r=0}^{n} {}_n\mathrm{C}_r a^{n-r}b^r$$

예제12

이항정리를 이용하여 $(3x-2y)^4$ 을 전개하시오.

$$\begin{aligned}(3x-2y)^4 &= {}_4\mathrm{C}_0(3x)^4 + {}_4\mathrm{C}_1(3x)^3(-2y)^1 + {}_4\mathrm{C}_2(3x)^2(-2y)^2 + {}_4\mathrm{C}_3(3x)^1(-2y)^3 + {}_4\mathrm{C}_4(-2y)^4 \\ &= 81x^4 - 216x^3y + 216x^2y^2 - 96xy^3 + 16y^4\end{aligned}$$

예제13

$\left(2x-\frac{1}{x}\right)^7$ 의 전개식에서 x^3 의 계수를 구하시오.

$\left(2x-\frac{1}{x}\right)^7 = \sum_{r=0}^{7} {}_7\mathrm{C}_r(2x)^{7-r}\left(-\frac{1}{x}\right)^r$ 이다.

이때, $(2x)^{7-r}\left(-\frac{1}{x}\right)^r = 2^{7-r}\times(-1)^r\times x^{7-2r}$ 이고, x 의 지수 $7-2r$ 이 3과 같아야 하므로,

$7-2r=3$ 에서 $r=2$ 임을 알 수 있다.

$r=2$ 이면 ${}_7\mathrm{C}_r(2x)^{7-r}\left(-\frac{1}{x}\right)^r = {}_7\mathrm{C}_2(2x)^{7-2}\left(-\frac{1}{x}\right)^2 = 21\times32x^5\times x^{-2} = 672x^3$ 이 된다.

따라서 x^3 의 계수는 672 이다.

이항계수의 성질

(1) ${}_{n-1}\mathrm{C}_{r-1} + {}_{n-1}\mathrm{C}_r = {}_n\mathrm{C}_r$

(2) ${}_n\mathrm{C}_0 + {}_n\mathrm{C}_1 + {}_n\mathrm{C}_2 + {}_n\mathrm{C}_3 + \cdots + {}_n\mathrm{C}_n = 2^n$

(3) ${}_n\mathrm{C}_0 - {}_n\mathrm{C}_1 + {}_n\mathrm{C}_2 - {}_n\mathrm{C}_3 + \cdots + {}_n\mathrm{C}_n(-1)^n = 0$

(4) $n = 2k$ (k는 자연수)인 경우

① ${}_n\mathrm{C}_0 + {}_n\mathrm{C}_2 + {}_n\mathrm{C}_4 + \cdots + {}_n\mathrm{C}_n = 2^{n-1}$

② ${}_n\mathrm{C}_1 + {}_n\mathrm{C}_3 + {}_n\mathrm{C}_5 + \cdots + {}_n\mathrm{C}_{n-1} = 2^{n-1}$

(5) $n = 2k - 1$ (k는 자연수)인 경우

① ${}_n\mathrm{C}_0 + {}_n\mathrm{C}_2 + {}_n\mathrm{C}_4 + \cdots + {}_n\mathrm{C}_{n-1} = 2^{n-1}$

② ${}_n\mathrm{C}_1 + {}_n\mathrm{C}_3 + {}_n\mathrm{C}_5 + \cdots + {}_n\mathrm{C}_n = 2^{n-1}$

➤ 다음과 같이 (1)이 성립함을 보일 수 있다.

$$\begin{aligned}
{}_{n-1}\mathrm{C}_{r-1} + {}_{n-1}\mathrm{C}_r &= \frac{(n-1)!}{(n-r)! \times (r-1)!} + \frac{(n-1)!}{(n-r-1)! \times r!} \\
&= \frac{r(n-1)!}{(n-r)! \times r!} + \frac{(n-r)(n-1)!}{(n-r)! \times r!} \\
&= (n-1)! \times \left\{ \frac{r + (n-r)}{(n-r)! \times r!} \right\} \\
&= \frac{n!}{(n-r)! \times r!} \\
&= {}_n\mathrm{C}_r
\end{aligned}$$

➤ $(1+x)^n = {}_n\mathrm{C}_0 + {}_n\mathrm{C}_1 x + {}_n\mathrm{C}_2 x^2 + \cdots + {}_n\mathrm{C}_r x^r + \cdots + {}_n\mathrm{C}_n x^n \quad \cdots\cdots$ ㉠은 x에 어떤 값을 대입해도 성립하는 항등식이다.

(2) ㉠의 양변에 $x = 1$을 대입하면

$${}_n\mathrm{C}_0 + {}_n\mathrm{C}_1 + {}_n\mathrm{C}_2 + {}_n\mathrm{C}_3 + \cdots + {}_n\mathrm{C}_n = 2^n \quad \cdots\cdots ㉡$$

(3) ㉠의 양변에 $x = -1$을 대입하면

$${}_n\mathrm{C}_0 - {}_n\mathrm{C}_1 + {}_n\mathrm{C}_2 - {}_n\mathrm{C}_3 + \cdots + {}_n\mathrm{C}_n(-1)^n = 0 \quad \cdots\cdots ㉢$$

(4) ㉡+㉢을 한 후, 양변을 2로 나누어 보일 수 있다.

(5) ㉡−㉢을 한 후, 양변을 2로 나누어 보일 수 있다.

예제14

식 ${}_2\mathrm{C}_0 + {}_3\mathrm{C}_1 + {}_4\mathrm{C}_2 + {}_5\mathrm{C}_3 + \cdots + {}_{10}\mathrm{C}_8$의 값을 구하시오.

${}_{n-1}\mathrm{C}_{r-1} + {}_{n-1}\mathrm{C}_r = {}_n\mathrm{C}_r$을 이용하면

$$\begin{aligned}
{}_2\mathrm{C}_0 + {}_3\mathrm{C}_1 &= {}_3\mathrm{C}_0 + {}_3\mathrm{C} + 1 = {}_4\mathrm{C}_1 \\
{}_4\mathrm{C}_1 + {}_4\mathrm{C}_2 &= {}_5\mathrm{C}_2 \\
{}_5\mathrm{C}_2 + {}_5\mathrm{C}_3 &= {}_6\mathrm{C}_3 \\
&\vdots \\
{}_9\mathrm{C}_6 + {}_9\mathrm{C}_7 &= {}_{10}\mathrm{C}_7 \\
{}_{10}\mathrm{C}_7 + {}_{10}\mathrm{C}_8 &= {}_{11}\mathrm{C}_8
\end{aligned}$$

$\therefore {}_2\mathrm{C}_0 + {}_3\mathrm{C}_1 + {}_4\mathrm{C}_2 + {}_5\mathrm{C}_3 + \cdots + {}_{10}\mathrm{C}_8 = {}_{11}\mathrm{C}_8 = 165$

예제15

다음 식의 값을 구하시오.

(1) ${}_8\mathrm{C}_0 + {}_8\mathrm{C}_1 + {}_8\mathrm{C}_2 + {}_8\mathrm{C}_3 + \cdots + {}_8\mathrm{C}_8$

(2) ${}_9\mathrm{C}_0 - {}_9\mathrm{C}_1 + {}_9\mathrm{C}_2 - {}_9\mathrm{C}_3 + \cdots - {}_9\mathrm{C}_9$

(1) ${}_8\mathrm{C}_0 + {}_8\mathrm{C}_1 + {}_8\mathrm{C}_2 + {}_8\mathrm{C}_3 + \cdots + {}_8\mathrm{C}_8 = 2^8 = 256$

(2) ${}_9\mathrm{C}_0 - {}_9\mathrm{C}_1 + {}_9\mathrm{C}_2 - {}_9\mathrm{C}_3 + \cdots - {}_9\mathrm{C}_9 = 0$

예제16

자연수 n에 대하여 $f(n) = \sum_{k=1}^{n} ({}_{2k}\mathrm{C}_1 + {}_{2k}\mathrm{C}_3 + {}_{2k}\mathrm{C}_5 + \cdots + {}_{2k}\mathrm{C}_{2k-1})$일 때, $f(5)$의 값을 구하시오.

$$\begin{aligned}
f(n) &= \sum_{k=1}^{n} ({}_{2k}\mathrm{C}_1 + {}_{2k}\mathrm{C}_3 + {}_{2k}\mathrm{C}_5 + \cdots + {}_{2k}\mathrm{C}_{2k-1}) \\
&= \sum_{k=1}^{n} 2^{2k-1} = \sum_{k=1}^{n} (2 \times 4^{k-1}) = \frac{2(4^n - 1)}{4-1} = \frac{2}{3}(4^n - 1)
\end{aligned}$$

$\therefore f(5) = \frac{2}{3}(4^5 - 1) = \frac{2}{3} \times 1023 = 682$

파스칼의 삼각형

n =0, 1, 2, 3, 4, 5일 때, $(a+b)^n$의 전개식은 다음과 같다.

$$
\begin{aligned}
(a+b)^0 &= 1 \\
(a+b)^1 &= {}_1\mathrm{C}_0 a + {}_1\mathrm{C}_1 b \\
&= 1a + 1b \\
(a+b)^2 &= {}_2\mathrm{C}_0 a^2 + {}_2\mathrm{C}_1 ab + {}_2\mathrm{C}_2 b^2 \\
&= 1a^2 + 2ab + 1b^2 \\
(a+b)^3 &= {}_3\mathrm{C}_0 a^3 + {}_3\mathrm{C}_1 a^2 b + {}_3\mathrm{C}_2 ab^2 + {}_3\mathrm{C}_3 b^3 \\
&= 1a^3 + 3a^2 b + 3ab^2 + 1b^3 \\
(a+b)^4 &= {}_4\mathrm{C}_0 a^4 + {}_4\mathrm{C}_1 a^3 b + {}_4\mathrm{C}_2 a^2 b^2 + {}_4\mathrm{C}_3 ab^3 + {}_4\mathrm{C}_4 b^4 \\
&= 1a^4 + 4a^3 b + 6a^2 b^2 + 4ab^3 + 1b^4 \\
(a+b)^5 &= {}_5\mathrm{C}_0 a^5 + {}_5\mathrm{C}_1 a^4 b + {}_5\mathrm{C}_2 a^3 b^2 + {}_5\mathrm{C}_3 a^2 b^3 + {}_5\mathrm{C}_4 ab^4 + {}_5\mathrm{C}_5 b^5 \\
&= 1a^5 + 5a^4 b + 10a^3 b^2 + 10a^2 b^3 + 5ab^4 + 1b^5
\end{aligned}
$$

이때, 각 항의 계수를 다음과 같이 삼각형 모양으로 나타낼 수 있다.

$$
\begin{array}{c}
n=0 \\ n=1 \\ n=2 \\ n=3 \\ n=4 \\ n=5 \\ \vdots
\end{array}
\qquad
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & {}_1\mathrm{C}_0 & & {}_1\mathrm{C}_1 & & & & \\
 & & & {}_2\mathrm{C}_0 & & {}_2\mathrm{C}_1 & & {}_2\mathrm{C}_2 & & & \\
 & & {}_3\mathrm{C}_0 & & {}_3\mathrm{C}_1 & & {}_3\mathrm{C}_2 & & {}_3\mathrm{C}_3 & & \\
 & {}_4\mathrm{C}_0 & & {}_4\mathrm{C}_1 & & {}_4\mathrm{C}_2 & & {}_4\mathrm{C}_3 & & {}_4\mathrm{C}_4 & \\
{}_5\mathrm{C}_0 & & {}_5\mathrm{C}_1 & & {}_5\mathrm{C}_2 & & {}_5\mathrm{C}_3 & & {}_5\mathrm{C}_4 & & {}_5\mathrm{C}_5 \\
 & & & & & \vdots & & & & &
\end{array}
\qquad
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & & & & \vdots & & & & &
\end{array}
$$

이 수의 배열에서 각 수는 그 수의 왼쪽 위와 오른쪽 위에 있는 두 수의 합과 같다. 예를 들어, 왼쪽 삼각형에서 ${}_3\mathrm{C}_1 + {}_3\mathrm{C}_2 = {}_4\mathrm{C}_2$가 됨을 오른쪽 삼각형을 통해 알 수 있다.
이와 같이 이항계수를 삼각형 모양으로 나열한 것을 파스칼의 삼각형이라고 한다.

고등학교 확률과 통계

확률

2

1. 확률의 뜻과 활용
2. 조건부 확률

1 확률의 뜻

시행과 사건

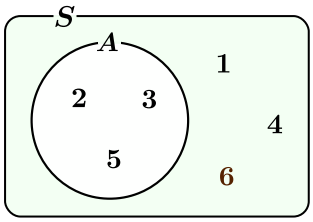
(1) 시행 : 동일한 조건에서 반복이 가능하고, 그 결과가 우연에 의해 지배되는 실험이나 관찰

(2) 표본공간 : 어떤 시행에서 일어날 수 있는 모든 결과들의 집합

(3) 사건 : 표본공간의 부분집합

(4) 근원사건 : 표본공간의 부분집합 중 원소의 개수가 한 개인 사건

➤ 한 개의 주사위를 던지는 것은 동일한 조건에서 반복이 가능하고, 그 결과가 우연에 의해 정해지므로 시행이다. 이때, 표본공간을 S라고 하면



$$S = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$$

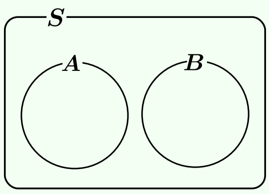
이고, 이 시행의 근원사건은

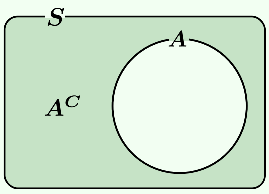
$$\{1\},\ \{2\},\ \{3\},\ \{4\},\ \{5\},\ \{6\}$$

이다. 또, 소수의 눈이 나오는 사건을 A라고 하면 $A = \{2,\ 3,\ 5\}$이다.

배반사건과 여사건

표본공간 S의 두 사건 $A,\ B$에 대하여





(1) 사건 A와 사건 B가 동시에 일어나지 않을 때, 즉

$$A \cap B = \varnothing$$

일 때, A와 B를 서로 배반사건이라고 한다.

(2) 사건 A에 대하여 A가 일어나지 않는 사건을 사건 A의 여사건이라고 하고 A^C로 나타낸다.

➤ A와 B의 합사건 : A 또는 B가 일어나는 사건으로 $A \cup B$로 나타낸다.

➤ A와 B의 곱사건 : A와 B가 동시에 일어나는 사건으로 $A \cap B$로 나타낸다.

➤ $A \cap A^C = \varnothing$이므로 A와 A^C는 서로 배반사건이다.

예제1

1부터 12까지의 자연수가 각각 하나씩 적힌 12개의 공이 들어 있는 상자에서 임의로 한 개의 공을 꺼낼 때, 공에 적힌 수가 8의 약수인 사건을 A, 12의 약수인 사건을 B라고 하자. 다음을 구하시오.

(1) $A \cap B$

(2) $A \cup B$

(3) A^{C}

(4) $A^{\mathrm{C}} \cup B^{\mathrm{C}}$

$A = \{1,\ 2,\ 4,\ 8\}$, $B = \{1,\ 2,\ 3,\ 4,\ 6,\ 12\}$이므로

(1) $A \cap B = \{1,\ 2,\ 4\}$

(2) $A \cup B = \{1,\ 2,\ 3,\ 4,\ 6,\ 8,\ 12\}$

(3) $A^{\mathrm{C}} = \{3,\ 5,\ 6,\ 7,\ 9,\ 10,\ 11,\ 12\}$

(4) $A^{\mathrm{C}} \cup B^{\mathrm{C}} = (A \cap B)^{\mathrm{C}} = \{3,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12\}$

예제2

한 개의 주사위를 던지는 시행에서 2의 배수의 눈이 나오는 사건을 A, 5의 약수의 눈이 나오는 사건을 B, 소수의 눈이 나오는 사건을 C라고 할 때, 다음을 구하시오.

(1) 사건 C의 여사건

(2) 사건 $A,\ B,\ C$ 중에서 서로 배반인 두 사건

$A = \{2,\ 4,\ 6\}$, $B = \{1,\ 5\}$, $C = \{2,\ 3,\ 5\}$이므로

(1) $C^{\mathrm{C}} = \{1,\ 4,\ 6\}$

(2) $A \cap B = \varnothing$이므로 사건 A와 사건 B가 서로 배반사건이다.

수학적 확률

어떤 시행에서 표본공간 S에 속하는 각각의 근원사건이 일어날 가능성이 같은 정도로 기대된다고 할 때, 사건 A가 일어날 확률 $\mathrm{P}(A)$를

$$\mathrm{P}(A) = \frac{n(A)}{n(S)} = \frac{(\text{사건 } A\text{가 일어날 모든 경우의 수})}{(\text{일어날 수 있는 모든 경우의 수})}$$

와 같이 정의하고, 이것을 수학적 확률이라고 한다.

예제3

서로 다른 두 개의 주사위를 동시에 던질 때, 눈의 수의 합이 7이 될 확률을 구하시오.

일어날 수 있는 모든 결과들의 집합인 표본공간 S는

$$S = \{(1,1),\ (1,2), \cdots,\ (6,5),\ (6,6)\}$$

이므로 $n(S) = 36$이다. 혹은 곱의 법칙을 이용하여 $n(S) = 6 \times 6 = 36$임을 알아낼 수도 있다.
이때, 눈의 수의 합이 7인 사건을 A라고 하면

$$A = \{(1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1)\}$$

이고 $n(A) = 6$이다.
따라서 구하는 확률은 $\mathrm{P}(A) = \dfrac{n(A)}{n(S)} = \dfrac{6}{36} = \dfrac{1}{6}$이다.

예제4

빨간 공 4개와 파란 공 3개가 들어 있는 주머니에서 임의로 3개의 공을 꺼낼 때, 빨간 공 1개와 파란 공 2개가 나올 확률을 구하시오.

표본공간을 S라고 하면 $n(S) = {}_7\mathrm{C}_3 = 35$이고, 빨간 공 1개와 파란 공 2개가 나오는 사건을 A라고 하면 $n(A) = {}_4\mathrm{C}_1 \times {}_3\mathrm{C}_2 = 12$이다.

따라서 구하는 확률은 $\mathrm{P}(A) = \dfrac{n(A)}{n(S)} = \dfrac{12}{35}$이다.

통계적 확률

어떤 시행을 n번 반복할 때, 사건 A가 일어난 횟수를 r_n이라 하자. 일반적으로 n이 매우 커질수록 사건 A가 일어나는 상대도수 $\frac{r_n}{n}$은 일정한 값 p에 가까워진다. 이때, p를 사건 A가 일어날 통계적 확률이라고 한다.

$$\mathrm{P}(A) = \lim_{n\to\infty} \frac{r_n}{n} = p$$

- 표본공간의 근원사건들이 일어날 가능성이 모두 같다고 기대할 수 없을 때, 각 근원사건이 일어날 확률을 통계적 확률로 구하게 된다. 예를 들어, 윷의 등과 배가 나올 확률은 각각 $\frac{1}{2}$라고 할 수 없다. 즉, 각 근원사건이 일어날 가능성이 같다고 볼 수 없다. 이런 경우에는 윷가락을 n번 던져서 그 중 배가 나온 횟수 r_n을 사용하여 다음과 같이 배가 나올 사건 A와 등이 나올 사건 B에 대한 확률을 구할 수 있다.

$$\mathrm{P}(A) = \lim_{n\to\infty} \frac{r_n}{n}, \quad \mathrm{P}(B) = 1 - \mathrm{P}(A)$$

  보통의 경우 시행 횟수 n이 충분히 크면 $\frac{r_n}{n}$을 통계적 확률로 생각해도 무방하다.

- 어떤 사건 A가 일어날 수학적 확률이 p일 때, 시행 횟수 n을 충분히 크게 하면 사건 A가 일어나는 상대도수 $\frac{r_n}{n}$은 수학적 확률 p에 가까워 진다는 것이 알려져 있다.

예제5

어떤 야구 선수가 1000번의 타석에서 270개의 안타를 쳤을 때, 이 타자가 안타를 칠 확률을 구하시오.

이 타자가 안타를 칠 통계적 확률은 $\frac{270}{1000} = \frac{27}{100}$이다.

확률의 기본 성질

표본공간 S의 임의의 사건 A에 대하여

(1) $0 \leq \mathrm{P}(A) \leq 1$

(2) $\mathrm{P}(S) = 1$

(3) $\mathrm{P}(\varnothing) = 0$

➤ 사건 A는 표본공간 S의 부분집합이므로 $0 \leq n(A) \leq n(S)$가 성립하고, 이 부등식의 양변을 $n(S)$로 나누면 $0 \leq \dfrac{n(A)}{n(S)} \leq 1$, 즉 $0 \leq \mathrm{P}(A) \leq 1$이 성립한다.

➤ $\mathrm{P}(S) = \dfrac{n(S)}{n(S)} = 1$ ➡ 반드시 일어나는 사건(전사건)

➤ $\mathrm{P}(\varnothing) = \dfrac{n(\varnothing)}{n(S)} = 0$ ➡ 절대로 일어나지 않는 사건(공사건)

예제6

한 개의 주사위를 두 번 던질 때, 다음을 구하시오.

(1) 눈의 수의 합이 1이 될 확률

(2) 눈의 수의 곱이 40 이하일 확률

(1) 눈의 수의 합이 1이 되는 사건은 공사건이므로 확률은 0이다.

(2) 눈의 수의 곱이 40 이하가 될 사건은 전사건이므로 확률은 1이다.

2 확률의 덧셈정리

확률의 덧셈정리

표본공간의 두 사건 A, B에 대하여 다음이 성립한다.

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$$

➤ 두 사건 A, B에 대하여

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

이므로, 사건 A 또는 B가 일어날 확률은 다음과 같이 구할 수 있다.

$$\begin{aligned}\mathrm{P}(A \cup B) &= \frac{n(A \cup B)}{n(S)} \\ &= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)} \\ &= \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)\end{aligned}$$

➤ 두 사건 A, B가 서로 배반사건이면 $\mathrm{P}(A \cap B) = 0$이므로 $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$이다.

예제7

한 개의 주사위를 던질 때, 짝수의 눈이 나오는 사건을 A, 6의 약수의 눈이 나오는 사건을 B라고 하자. A 또는 B가 일어날 확률을 구하시오.

$A = \{2,\ 4,\ 6\}$, $B = \{1,\ 2,\ 3,\ 6\}$, $A \cap B = \{2,\ 6\}$이므로 구하는 확률과 같다.

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$$

예제8

두 개의 주사위를 동시에 던질 때, 눈의 합이 5가 되는 사건을 A, 눈의 합이 7이 되는 사건을 B라고 하자. A 또는 B가 일어날 확률을 구하시오.

$A = \{(1,4),\ (2,3),\ (3,2),\ (4,1)\}$, $B = \{(1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1)\}$,
$A \cap B = \varnothing$이므로 구하는 확률은 다음과 같다.

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) = \frac{4}{36} + \frac{6}{36} = \frac{10}{36} = \frac{5}{18}$$

여사건의 확률

사건 A의 여사건 A^{C}가 일어날 확률은 다음과 같다.

$$\mathrm{P}\left(A^{\mathrm{C}}\right)=1-\mathrm{P}(A)$$

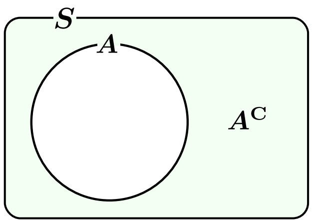
➤ $A \cap A^{\mathrm{C}}=\varnothing$이므로 두 사건 $A,\ A^{\mathrm{C}}$은 서로 배반사건이다. 따라서 확률의 덧셈정리에 의하여

$$\mathrm{P}\left(A \cup A^{\mathrm{C}}\right)=\mathrm{P}(A)+\mathrm{P}\left(A^{\mathrm{C}}\right)$$

이다. 그런데 $A \cup A^{\mathrm{C}}$는 시행에서 일어날 수 있는 모든 결과들의 집합인 표본공간 S와 같기 때문에 $\mathrm{P}\left(A \cup A^{\mathrm{C}}\right)=1$이 된다. 따라서

$$\mathrm{P}\left(A^{\mathrm{C}}\right)=1-\mathrm{P}(A)$$

가 성립한다.



예제9

두 개의 주사위를 동시에 던질 때, 서로 다른 눈이 나올 확률을 구하시오.

서로 같은 눈이 나오는 사건을 A라고 하면, 구하는 확률은 $\mathrm{P}\left(A^{\mathrm{C}}\right)$와 같다.
$A=\{(1,1),\ (2,2),\ (3,3),\ (4,4),\ (5,5),\ (6,6)\}$이므로 $\mathrm{P}\left(A^{\mathrm{C}}\right)$는 다음과 같이 구할 수 있다.

$$\mathrm{P}\left(A^{\mathrm{C}}\right)=1-\mathrm{P}(A)=1-\frac{6}{36}=1-\frac{1}{6}=\frac{5}{6}$$

예제10

3개의 당첨 제비가 포함된 10개의 제비 중에서 임의로 두 개의 제비를 뽑을 때, 적어도 한 개가 당첨 제비일 확률을 구하시오.

적어도 한 개의 당첨 제비를 뽑는 사건을 A라고 하면, 여사건 A^{C}는 당첨 제비를 한 개도 뽑지 못하는 사건이므로 $\mathrm{P}\left(A^{\mathrm{C}}\right)=\frac{{}_{7}\mathrm{C}_{2}}{{}_{10}\mathrm{C}_{2}}=\frac{21}{45}=\frac{7}{15}$가 된다.

$\therefore \mathrm{P}(A)=1-\mathrm{P}\left(A^{\mathrm{C}}\right)=1-\frac{7}{15}=\frac{8}{15}$

1 조건부확률과 확률의 곱셈정리

조건부확률

사건 A, B가 표본공간 S의 두 부분집합이라 할 때, 사건 A가 일어났다는 조건 하에 사건 B가 일어날 확률을 사건 A가 일어났을 때의 사건 B의 조건부확률이라고 하고, 이것을 기호로 $\mathrm{P}(B|A)$로 나타낸다. 이때, 조건부확률 $\mathrm{P}(B|A)$는 다음과 같이 구할 수 있다.

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)} \quad (\text{단, } \mathrm{P}(A) > 0)$$

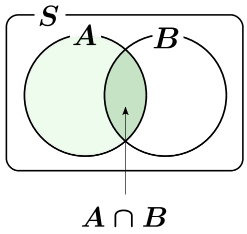
➤ 표본공간 S에서 사건 A가 일어났을 때의 사건 B의 조건부확률은

$$\mathrm{P}(B|A) = \frac{n(A\cap B)}{n(A)}$$

이다. 이때, $\mathrm{P}(A) = \dfrac{n(A)}{n(S)}$, $\mathrm{P}(A\cap B) = \dfrac{n(A\cap B)}{n(S)}$ 이므로

$$\mathrm{P}(B|A) = \frac{n(A\cap B)}{n(A)} = \frac{\frac{n(A\cap B)}{n(S)}}{\frac{n(A)}{n(S)}} = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)}$$

가 성립한다.



예제11

한 학급 30명의 학생 중 여학생이 20명, 안경을 착용한 학생이 17명, 안경을 착용하지 않은 남학생이 5명이라고 한다. 이 중 한 명을 선택했더니 남학생이었다. 이 학생이 안경을 착용하고 있을 확률을 구하시오.

오른쪽 표로부터 조건부확률을 다음과 같이 구할 수 있다.

$$\mathrm{P}(\text{안경}|\text{남학생}) = \frac{\mathrm{P}(\text{안경}\cap\text{남학생})}{\mathrm{P}(\text{남학생})} = \frac{\frac{5}{30}}{\frac{10}{30}} = \frac{5}{10} = \frac{1}{2}$$

	안경	눈	계
남자	5	5	10
여자	12	8	20
계	17	13	30

확률의 곱셈정리

$\mathrm{P}(A)>0$, $\mathrm{P}(B)>0$일 때, 사건 A와 B가 동시에 일어날 확률은

$$\mathrm{P}(A\cap B)=\mathrm{P}(A)\times\mathrm{P}(B|A)=\mathrm{P}(B)\times\mathrm{P}\times\mathrm{P}(A|B)$$

➤ 조건부확률 $\mathrm{P}(B|A)=\dfrac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)}$의 양변에 $\mathrm{P}(A)$를 곱하면

$$\mathrm{P}(A\cap B)=\mathrm{P}(A)\times\mathrm{P}(B|A)$$

가 된다. 마찬가지로 조건부확률 $\mathrm{P}(A|B)=\dfrac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}$의 양변에 $\mathrm{P}(B)$를 곱하면

$$\mathrm{P}(A\cap B)=\mathrm{P}(B)\times\mathrm{P}(A|B)$$

가 된다.

예제12

주머니 안에 빨간 공이 6개, 파란 공이 4개 들어 있다. 이 주머니에서 차례로 2개의 공을 꺼낼 때, 2개 모두 빨간 공일 확률을 구하시오. (단, 꺼낸 공은 다시 넣지 않는다.)

첫 번째 꺼낸 공이 빨간 공인 사건을 A, 두 번째 꺼낸 공이 빨간 공인 사건을 B라고 하자.

첫 번째에 빨간 공을 꺼낼 확률은 $\mathrm{P}(A)=\dfrac{6}{10}=\dfrac{3}{5}$이고,

첫 번째 꺼낸 공이 빨간색일 때, 두 번째 꺼낸 공도 빨간색일 확률은 $\mathrm{P}(B|A)=\dfrac{5}{9}$이다.

따라서 구하는 확률은 $\mathrm{P}(A\cap B)=\mathrm{P}(A)\times\mathrm{P}(B|A)=\dfrac{3}{5}\times\dfrac{5}{9}=\dfrac{1}{3}$이다.

예제13

빨간 공 4개와 파란 공 2개가 들어 있는 주머니에서 정국이가 임의로 하나의 공을 꺼낸 후, 지민이가 남아 있는 공 중에서 임의로 하나의 공을 꺼냈다. 지민이가 꺼낸 공이 빨간색이었을 때, 정국이가 꺼낸 공도 빨간색이었을 확률을 구하시오.

정국이가 꺼낸 공이 빨간색인 사건을 A, 지민이가 꺼낸 공이 빨간색인 사건을 B라고 하면

$$\begin{aligned}\mathrm{P}(B) &= \mathrm{P}(A\cap B)+\mathrm{P}\left(A^{\mathrm{C}}\cap B\right)\\ &= \{\mathrm{P}(A)\times\mathrm{P}(B|A)\}+\left\{\mathrm{P}\left(A^{\mathrm{C}}\right)\times\mathrm{P}\left(B|A^{\mathrm{C}}\right)\right\}\\ &= \left(\frac{4}{6}\times\frac{3}{5}\right)+\left(\frac{2}{6}\times\frac{4}{5}\right)=\frac{2}{3}\end{aligned}$$

따라서 구하는 확률은 $\mathrm{P}(A|B)=\dfrac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}=\dfrac{\frac{2}{5}}{\frac{2}{3}}=\dfrac{3}{5}$이다.

사건의 독립과 종속

두 사건 $A,\ B$에 대하여 어느 한 사건이 다른 사건이 일어날 확률에 영향을 미치지 않을 때, 즉

$$\mathrm{P}(B|A)=\mathrm{P}\left(B|A^{\mathrm{C}}\right)=\mathrm{P}(B) \text{ 또는 } \mathrm{P}(A|B)=\mathrm{P}\left(A|B^{\mathrm{C}}\right)=\mathrm{P}(A)$$

일 때, 두 사건 $A,\ B$를 서로 독립이라고 한다.
두 사건이 독립이 아닐 때, 두 사건은 서로 종속이라고 한다.

두 사건이 서로 독립일 조건

두 사건 $A,\ B$가 서로 독립이기 위한 필요충분조건은 다음과 같다.

$$\mathrm{P}(A\cap B)=\mathrm{P}(A)\times\mathrm{P}(B)\quad(\text{단, }\mathrm{P}(A)>0,\ \mathrm{P}(B)>0)$$

➤ 두 사건 $A,\ B$가 서로 독립이면 확률의 곱셈정리에 의하여 다음이 성립한다.

$$\mathrm{P}(A\cap B)=\mathrm{P}(A)\times\mathrm{P}(B|A)=\mathrm{P}(A)\times\mathrm{P}(B)$$

➤ 역으로, $\mathrm{P}(A\cap B)=\mathrm{P}(A)\times\mathrm{P}(B)$이고 $\mathrm{P}(A)>0$이면

$$\mathrm{P}(B|A)=\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)}=\frac{\mathrm{P}(A)\times\mathrm{P}(B)}{\mathrm{P}(A)}=\mathrm{P}(B)$$

이 성립하므로 두 사건 $A,\ B$는 서로 독립이다.

예제14

한 개의 주사위를 던져 짝수의 눈이 나오는 사건을 A, 4이상의 눈이 나오는 사건을 B, 6의 약수의 눈이 나오는 사건을 C라 할 때, 다음 두 사건이 서로 독립인지 아니면 종속이지 조사하시오.

(1) 사건 $A,\ B$

(2) 사건 $A,\ C$

$A=\{2,\ 4,\ 6\}$, $B=\{4,\ 5,\ 6\}$, $C=\{1,\ 2,\ 3,\ 6\}$이므로 $\mathrm{P}(A)=\dfrac{1}{2}$, $\mathrm{P}(B)=\dfrac{1}{2}$, $\mathrm{P}(C)=\dfrac{2}{3}$이다.

(1) $A\cap B=\{4,\ 6\}$이므로 $\mathrm{P}(A\cap B)=\dfrac{1}{3}$이고, $\mathrm{P}(A)\times\mathrm{P}(B)=\dfrac{1}{2}\times\dfrac{1}{2}=\dfrac{1}{4}$이다.
따라서 $\mathrm{P}(A\cap B)\neq\mathrm{P}(A)\times\mathrm{P}(B)$이므로 두 사건 $A,\ B$는 서로 종속이다.

(2) $A\cap C=\{2,\ 6\}$이므로 $\mathrm{P}(A\cap C)=\dfrac{1}{3}$이고, $\mathrm{P}(A)\times\mathrm{P}(C)=\dfrac{1}{2}\times\dfrac{2}{3}=\dfrac{1}{3}$이다.
따라서 $\mathrm{P}(A\cap C)=\mathrm{P}(A)\times\mathrm{P}(C)$이므로 두 사건 $A,\ C$는 서로 독립이다.

예제15

두 사건 $A,\ B$가 서로 독립일 때, 다음 두 사건도 서로 독립임을 보이시오.
(단, $0<\mathrm{P}(A)<1,\ 0<\mathrm{P}(B)<1$)

(1) 사건 $A^{\mathrm{C}},\ B$

(2) 사건 $A^{\mathrm{C}},\ B^{\mathrm{C}}$

두 사건이 서로 독립이므로 $\mathrm{P}(A\cap B)=\mathrm{P}(A)\times\mathrm{P}(B)$이다.

(1) $\mathrm{P}\left(A^{\mathrm{C}}\cap B\right)$
$$\begin{aligned}&=\mathrm{P}(B-A)\\&=\mathrm{P}(B)-\mathrm{P}(A\cap B)\\&=\mathrm{P}(B)-\mathrm{P}(A)\times\mathrm{P}(B)\\&=\mathrm{P}(B)(1-\mathrm{P}(A))\\&=\mathrm{P}(B)\times\mathrm{P}\left(A^{\mathrm{C}}\right)\end{aligned}$$

(2) $\mathrm{P}\left(A^{\mathrm{C}}\cap B^{\mathrm{C}}\right)$
$$\begin{aligned}&=\mathrm{P}\left((A\cup B)^{\mathrm{C}}\right)\\&=1-\mathrm{P}(A\cup B)\\&=1-\mathrm{P}(A)-\mathrm{P}(B)+\mathrm{P}(A\cap B)\\&=1-\mathrm{P}(A)-\mathrm{P}(B)-\mathrm{P}(A)\times\mathrm{P}(B)\\&=(1-\mathrm{P}(A))\times(1-\mathrm{P}(B))\\&=\mathrm{P}\left(A^{\mathrm{C}}\right)\times\mathrm{P}\left(B^{\mathrm{C}}\right)\end{aligned}$$

독립시행

어떤 시행을 여러 번 반복하는 경우, 매 시행마다 일어나는 사건이 서로 독립일 때, 이러한 시행을 독립시행이라고 한다.

독립시행의 확률

매회 사건 A가 일어날 확률이 p, 사건 A가 일어나지 않을 확률이 $(1-p)$로 일정할 때, n회의 독립시행 중 사건 A가 r번 일어날 확률 P_r은 다음과 같다.

$$\mathrm{P}_r = {}_n\mathrm{C}_r p^r (1-p)^{n-r} \quad (r = 0,\ 1,\ 2,\ \cdots,\ n)$$

➤ 독립시행의 확률은 i번 째 시행에서 사건 A가 일어나는 경우와 그렇지 않은 경우로 나누어 다음과 같이 구할 수 있다.

(1) i번 째 시행에서 사건 A가 일어나는 것을 ◯, 일어나지 않는 것을 ✕로 나타낼 때, n번의 시행 중 사건 A가 r번 일어나는 경우의 수는

r개의 ◯와 $n-r$개의 ✕를 나열하는 경우의 수 $\dfrac{n!}{r! \times (n-r)!} = {}_n\mathrm{C}_r$

와 같다.

(2) 각 시행이 모두 독립이므로 사건 A가 r번 일어나고, 사건 A^{C}가 $n-r$번 일어났다면 확률은 $p^r \times (1-p)^{n-r}$이 된다.

(3) (1)과 (2)로부터 n회의 시행 중 사건 A가 r번 일어날 확률은

$${}_n\mathrm{C}_r p^r (1-p)^{n-r} \quad (r = 0,\ 1,\ 2,\ \cdots,\ n)$$

과 같이 구할 수 있다.

예제16

주사위를 6번 던졌을 때, 1 또는 2의 눈이 2번 나올 확률을 구하시오.

주사위를 던지는 시행은 독립시행이다. 주사위를 던져서 1 또는 2의 눈이 나오는 사건을 A 라고 하면, $\mathrm{P}(A) = \dfrac{1}{3}$ 이다. 결국 6번의 독립시행에서 $p = \dfrac{1}{3}$ 인 사건이 2번 일어날 독립시행의 확률을 구하면 된다.

$$\therefore\ {}_6\mathrm{C}_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^4 = 15 \times \frac{1}{9} \times \frac{16}{81} = \frac{80}{243}$$

예제17

어떤 사격 선수가 매회의 시행에서 10점 과녁을 맞히는 사건이 서로 독립이고, 각 시행에서 10점 과녁을 맞힐 확률은 $\dfrac{3}{4}$ 이라고 한다. 이 선수가 4발을 쏘았을 때, 3발 이상 10점 과녁에 맞힐 확률을 구하시오.

(1) 3발을 10점 과녁에 맞힐 확률은 ${}_4\mathrm{C}_3 \left(\dfrac{3}{4}\right)^3 \left(\dfrac{1}{4}\right)^1 = \dfrac{27}{64}$

(2) 4발을 10점 과녁에 맞힐 확률은 ${}_4\mathrm{C}_4 \left(\dfrac{3}{4}\right)^4 \left(\dfrac{1}{4}\right)^0 = \dfrac{81}{256}$

따라서 3발 이상 10점 과녁에 맞힐 확률은 $\dfrac{27}{64} + \dfrac{81}{256} = \dfrac{189}{256}$ 이다.

통계

3

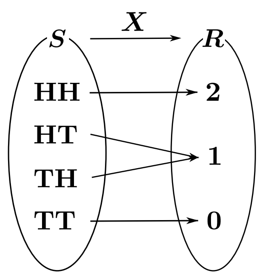
1. 확률분포
2. 통계적 추정

1 확률변수와 확률분포

확률변수

어떤 시행에서 표본공간 S의 각 원소에 하나의 실숫값이 대응되는 함수를 확률변수라 한다.

➤ 동전 두 개를 동시에 던지는 시행에서 앞면이 나오는 것을 H, 뒷면이 나오는 것을 T라 할 때, 이 시행의 표본공간은 {HH, HT, TH, TT}가 된다 .이 표본공간의 각 근원사건에서 앞면이 나온 동전의 개수를 X라고 하면, 오른쪽 그림과 같은 대응 관계가 성립한다. 이러한 대응 관계를 나타내는 함수를 확률변수라고 한다. 즉, 확률변수는 표본공간을 정의역으로 하고 실수의 집합을 공역으로 하는 함수이다.



➤ 확률변수는 대개 알파벳 대문자 $X,\ Y,\ Z$ 등으로 나타내고, 확률변수가 갖는 값들은 대개 알파벳 소문자 $x,\ y,\ z$ 등으로 나타낸다.

이산확률변수의 확률분포

(1) 이산확률변수

확률변수 X가 갖는 값을 셀 수 있을 때, 그 확률변수 X를 이산확률변수라고 한다.

(2) 이산확률분포

이산확률변수 X의 값 $x_1,\ x_2,\ x_3,\ \cdots,\ x_n$ 각각에 대한 확률 $p_1,\ p_2,\ p_3,\ \cdots,\ p_n$의 대응관계를 이산확률변수 X에 대한 확률분포라 하고, 이 대응 관계를

$$\mathrm{P}(X=x_i)=p_i \quad (i=1,\ 2,\ 3,\ \cdots,\ n)$$

과 같이 나타낸다. 이때, $\mathrm{P}(X=x_i)$를 이산확률변수 X의 확률질량함수라고 한다. 또한, 이 확률분포를 표로 나타낸 것을 확률분포표라고 한다.

X	x_1	x_2	$\cdots$	x_n
$\mathrm{P}(X=x_i)$	p_1	p_2	$\cdots$	p_n

➤ $\mathrm{P}(X=x)$는 이산확률변수 X가 어떤 값 x를 가질 확률을 나타낸다.

➤ 이산확률변수 X가 x_i이상 x_j이하의 값을 가질 확률은 다음과 같이 계산하다.

$$\mathrm{P}(x_i \le X \le x_j)=\sum_{k=i}^{j} p_k \quad (\text{단},\ i,\ j=1,\ 2,\ \cdots,\ n,\ i<j)$$

예제1

주사위를 던져서 나오는 눈의 수를 4로 나눈 나머지를 확률변수 X라고 할 때, 다음 물음에 답하시오.

(1) X의 확률분포를 구하시오.

(2) $\mathrm{P}(1 \le X \le 2)$를 구하시오.

(1) 주사위를 던져서 나오는 눈의 수는 $1,\ 2,\ 3,\ 4,\ 5,\ 6$의 여섯 가지이고, 이들을 4로 나눈 나머지는 각각 $1,\ 2,\ 3,\ 0,\ 1,\ 2$이다. 따라서 확률변수 X가 가질 수 있는 값들은 $0,\ 1,\ 2,\ 3$이고, 이들에 대응되는 확률분포는 다음 표와 같다.

X	0	1	2	3
$\mathrm{P}(X = x_i)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

(2) $\mathrm{P}(1 \le X \le 2) = \mathrm{P}(X = 1) + \mathrm{P}(x = 2) = \frac{2}{6} + \frac{2}{6} = \frac{2}{3}$

확률질량함수의 성질

이산확률변수 X가 취하는 값이 $x_1,\ x_2,\ x_3,\ \cdots,\ x_n$일 때, 확률질량함수

$$\mathrm{P}(X = x_1) \quad (i = 1,\ 2,\ 3,\ \cdots,\ n)$$

에 대해서 다음이 성립한다.

(1) $0 \le \mathrm{P}(X = x_i) \le 1$

(2) $\sum_{i=1}^{n} \mathrm{P}(X = x_i) = 1$

- 확률질량함수의 성질은 확률의 기본성질로부터 성립한다.
- 표본공간은 시행에서 일어날 수 있는 모든 결과들의 집합이다. 따라서 표본공간의 원소에 대한 모든 확률값들을 더하면 1이 된다.
- 예제1 에서 $\mathrm{P}(X = 0) = \frac{1}{6}$, $\mathrm{P}(X = 1) = \frac{2}{6}$, $\mathrm{P}(X = 2) = \frac{2}{6}$, $\mathrm{P}(X = 3) = \frac{1}{6}$로 (1)을 만족하는 것을 확인할 수 있다. 또한 $\frac{1}{6} + \frac{2}{6} + \frac{2}{6} + \frac{1}{6} = 1$로 (2)도 만족하는 것을 확인할 수 있다.

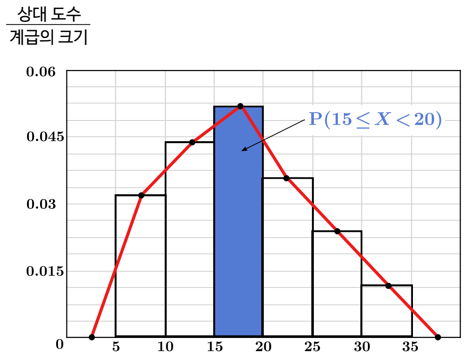
연속확률변수의 확률분포

(1) 연속확률변수
사람의 몸무게, 통학시간, 각 물질의 끓는 점 등과 같이 확률변수 X가 어떤 구간의 모든 실숫값을 가질 때, 그 확률변수 X를 연속확률변수라고 한다.

(2) 연속확률분포 - 아래 설명 참조

➤ 다음은 큐브 동호회 회원 100명의 기록을 조사한 표와 $\dfrac{\text{상대도수}}{\text{계급의 크기}}$에 대한 히스토그램 및 분포다각형을 보여준다.

기록(초)	도수	상대도수	$\dfrac{(\text{상대도수})}{(\text{계급의 크기})}$
$5^{\text{이상}} \sim 10^{\text{미만}}$	16	0.16	0.032
$10 \sim 15$	22	0.22	0.044
$15 \sim 20$	26	0.26	0.052
$20 \sim 25$	18	0.18	0.036
$25 \sim 30$	12	0.12	0.024
$30 \sim 35$	6	0.06	0.012
합계	100	1	

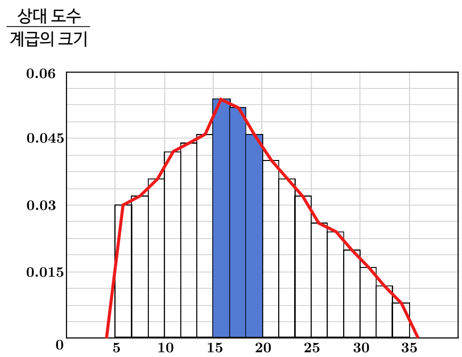


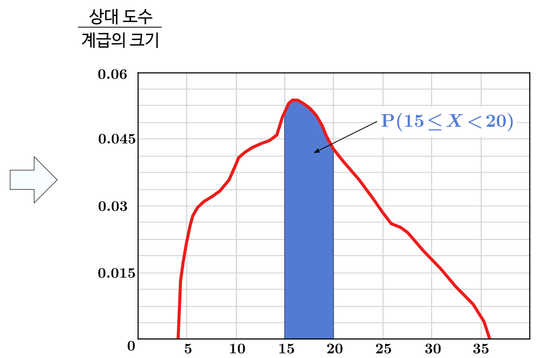
➤ 히스토그램에서 각 직사각형의 넓이는

$$(\text{계급의 크기}) \times \frac{(\text{상대도수})}{(\text{계급의 크기})} = (\text{상대도수})$$

이므로, 직사각형의 넓이의 합은 상대도수의 합 1과 같다. 즉, 분포다각형의 내부의 넓이는 1이다.

➤ 더 많은 수의 회원의 기록을 조사하고, 계급의 크기를 더 작게하여 히스토그램과 분포다각형을 그리게 되면, 분포다각형이 다음과 같이 점점 곡선에 가까워진다.





➤ 이러한 곡선을 그래프로 갖는 함수 $f(x)$를 연속확률변수 X의 확률밀도함수라고 한다.

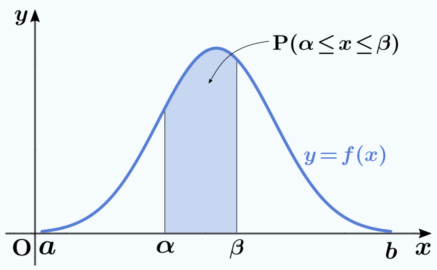
연속확률변수의 확률분포

연속확률변수 X의 확률밀도함수 $f(x)$ $(a \le X \le b)$에 대하여 다음이 성립한다.

(1) $f(x) \ge 0$

(2) $f(x)$의 그래프와 X축 및 두 직선 $x=a$, $x=b$로 둘러싸인 부분의 넓이는 1이다.

(3) 두 상수 α, β $(a \le \alpha \le \beta \le b)$에 대하여 $\mathrm{P}(\alpha \le X \le \beta)$는 $f(x)$의 그래프와 x축 및 두 직선 $x=\alpha$, $x=\beta$로 둘러싸인 부분의 넓이와 같다.



➤ 어떤 함수 $f(x)$가 위의 성질 (1), (2)를 모두 만족하면 $f(x)$는 확률밀도함수다.

➤ 연속확률변수 X가 어떤 특정한 값 x를 취할 확률 $\mathrm{P}(X=x)=0$이다.
따라서 다음이 성립한다.

$$\mathrm{P}(\alpha \le X \le \beta) = \mathrm{P}(\alpha \le X < \beta) = \mathrm{P}(\alpha < X \le \beta) = \mathrm{P}(\alpha < X < \beta)$$

예제2

연속확률변수 X의 확률밀도함수 $f(x)=kx$ $(0 \le X \le 3)$일 때, 다음 물음에 답하시오.
(단, k는 상수)

(1) k의 값을 구하시오.

(2) $\mathrm{P}(1 \le X \le 2)$를 구하시오.

(1) $\displaystyle\int_0^3 kx\,dx = \left[\frac{k}{2}x^2\right]_0^3 = \frac{9}{2} = 1$에서 $k=\dfrac{2}{9}$

(2) $\displaystyle\mathrm{P}(1 \le X \le 2) = \int_1^2 \frac{2}{9}x\,dx = \left[\frac{1}{9}x^2\right]_1^2 = \frac{4-1}{9} = \frac{3}{9} = \frac{1}{3}$

2 이산확률변수의 기댓값과 표준편차

이산확률변수의 기댓값

이산확률변수 X의 확률분포가 오른쪽 표와 같을 때, $x_1p_1+x_2p_2+\cdots+x_np_n$을 이산확률변수 X의 평균 또는 기댓값이라 하고, 기호 $\mathrm{E}(X)$로 나타낸다.

X	x_1	x_2	$\cdots$	x_n
$\mathrm{P}(X=x_i)$	p_1	p_2	$\cdots$	p_n

$$\mathrm{E}(X)=\sum_{i=1}^{n}x_ip_i=x_1p_1+x_2p_2+\cdots+x_np_n$$

예제3

아래 표와 같이 상금이 걸려 있는 50장의 복권에서 상금의 평균을 구하시오.

상금(원)	$10,000$	$5,000$	$1,000$	0	합계
복권 수(장)	1	5	15	29	50

복권 한 장에 대한 상금을 확률변수 X라고 하면, X의 확률분포는 아래 표와 같다.

X	$10,000$	$5,000$	$1,000$	0	합계
$\mathrm{P}(X)$	$\frac{1}{50}$	$\frac{5}{50}$	$\frac{15}{50}$	$\frac{29}{50}$	1

따라서 확률변수 X의 기댓값은 다음과 같다.

$$\mathrm{E}(X)=\left(10000\times\frac{1}{50}\right)+\left(5000\times\frac{5}{50}\right)+\left(1000\times\frac{15}{50}\right)+\left(0\times\frac{29}{50}\right)=1000\ (\text{원})$$

이산확률변수의 분산과 표준편차

(1) 분산

이산확률변수 X의 기댓값 $\mathrm{E}(X)$를 m이라고 할 때, $\mathrm{E}\left((X-m)^2\right)$을 확률변수 X의 분산이라 하고, 이것을 기호 $\mathrm{V}(X)$로 나타낸다.

$$\mathrm{V}(X) = \mathrm{E}\left((X-m)^2\right) = \sum_{i=1}^{n}(x_1 - m)^2 p_i$$

(2) 표준편차

분산 $\mathrm{V}(X)$의 양의 제곱근을 표준편차라 하고, 이것을 기호로 $\sigma(X)$로 나타낸다.

$$\sigma(X) = \sqrt{\mathrm{V}(X)}$$

➤ 이산확률변수 X의 분산과 표준편차를 다음과 같이 구할 수도 있다.

$$\begin{aligned}
\mathrm{V}(X) &= \sum_{i=1}^{n}(x_1 - m)^2 p_i \\
&= \sum_{i=1}^{n} x_i^2 p_i - 2m\sum_{i=1}^{n} x_i p_i + m^2 \sum_{i=1}^{n} p_i \qquad \left(\sum_{i=1}^{n} x_i p_i = m,\ \sum_{i=1}^{n} p_i = 1\right) \\
&= \sum_{i=1}^{n} x_i^2 p_i - 2m^2 + m^2 \\
&= \sum_{i=1}^{n} x_i^2 p_i - m^2 \\
&= \mathrm{E}\left(X^2\right) - \{\mathrm{E}(X)\}^2 \\
\sigma(X) &= \sqrt{\mathrm{E}\left(X^2\right) - \{\mathrm{E}(X)\}^2}
\end{aligned}$$

예제4

검은 공이 3개, 흰 공이 4개 들어 있는 주머니에서 두 개의 공을 임의로 꺼낼 때, 흰 공의 개수를 확률변수 X라 하자. 확률변수 X의 기댓값과 분산을 구하시오.

$\mathrm{P}(X=0)=\dfrac{{}_3\mathrm{C}_2}{{}_7\mathrm{C}_2}=\dfrac{3}{21}=\dfrac{1}{7}$, $\mathrm{P}(X=1)=\dfrac{{}_3\mathrm{C}_1\times{}_4\mathrm{C}_1}{{}_7\mathrm{C}_2}=\dfrac{12}{21}=\dfrac{4}{7}$,

$\mathrm{P}(X=2)=\dfrac{{}_4\mathrm{C}_2}{{}_7\mathrm{C}_2}=\dfrac{6}{21}=\dfrac{2}{7}$ 이므로 확률변수 X의 확률분포는 다음 표와 같다.

X	0	1	2	합계
$\mathrm{P}(X)$	$\frac{1}{7}$	$\frac{4}{7}$	$\frac{2}{7}$	1

따라서 확률변수 X의 기댓값과 분산은 다음과 같이 구할 수 있다.

$$\mathrm{E}(X)=\left(0\times\frac{1}{7}\right)+\left(1\times\frac{4}{7}\right)+\left(2\times\frac{2}{7}\right)=\frac{8}{7}$$

$$\mathrm{E}\left(X^2\right)=\left(0^2\times\frac{1}{7}\right)+\left(1^2\times\frac{4}{7}\right)+\left(2^2\times\frac{2}{7}\right)=\frac{12}{7}$$

$$\mathrm{V}(X)=\mathrm{E}\left(X^2\right)-\{\mathrm{E}(X)\}^2=\frac{12}{7}-\left(\frac{8}{7}\right)^2=\frac{20}{49}$$

이산확률변수의 평균, 분산, 표준편차의 성질

이산확률변수 X와 임의의 상수 $a,\ b\ (a \neq 0)$에 대하여

(1) $\mathrm{E}(aX+b) = a\mathrm{E}(X)+b$

(2) $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$

(3) $\sigma(aX+b) = |a|\sigma(X)$

➤ 확률변수 X의 분포가 오른쪽 표와 같을 때, X에 대한 일차식 $aX+b$로 정해지는 확률변수를 Y라고 하자.

X	x_1	x_2	$\cdots$	x_n
$\mathrm{P}(X=x_i)$	p_1	p_2	$\cdots$	p_n

$$Y = aX+b$$

이때, $y_i = ax_i + b$라고 하면 $\mathrm{P}(Y=y_i) = \mathrm{P}(X=x_i) = p_i$이므로 확률변수 Y의 확률분포는 다음과 같다.

Y	y_1	y_2	$\cdots$	y_n
$\mathrm{P}(Y=y_i)$	p_1	p_2	$\cdots$	p_n

따라서 확률변수 Y의 평균과 분산은 다음과 같다.

(1) $$\begin{aligned}\mathrm{E}(Y) &= \sum_{i=1}^{n} y_i p_i \\ &= \sum_{i=1}^{n} (ax_i+b)p_i \\ &= a\sum_{i=1}^{n} x_i p_i + b\sum_{i=1}^{n} \\ &= a\mathrm{E}(X)+b\end{aligned}$$

(2) $\mathrm{E}(X) = m$이라고 하면 $\mathrm{E}(Y) = am+b$이므로

$$\begin{aligned}\mathrm{V}(Y) &= \sum_{i=1}^{n} \{y_i - \mathrm{E}(Y)\}^2 p_i \\ &= \sum_{i=1}^{n} \{(ax_i+b)-(am+b)\}^2 p_i \\ &= a^2\sum_{i=1}^{n} (x_1-m)^2 p_i \\ &= a^2\mathrm{V}(X)\end{aligned}$$

(3) $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$의 양변에 $\sqrt{\ }$를 취하면

$$\sqrt{\mathrm{V}(aX+b)} = \sqrt{a^2\mathrm{V}(X)} = |a|\sqrt{\mathrm{V}(X)}$$

$$\therefore\ \sigma(aX+b) = |a|\sigma(X)$$

예제5

확률변수 X에 대하여 $\mathrm{E}(X)=m,\ \mathrm{V}(X)=\sigma^2$이라고 할 때, 확률변수 $Z=\dfrac{X-m}{\sigma}$의 평균과 표준편차를 구하시오.

$$\mathrm{E}(Z)=\mathrm{E}\left(\frac{X-m}{\sigma}\right)=\mathrm{E}\left(\frac{1}{\sigma}X-\frac{m}{\sigma}\right)=\frac{1}{\sigma}\mathrm{E}(X)-\frac{m}{\sigma}=\frac{m}{\sigma}-\frac{m}{\sigma}=0$$

$$\mathrm{V}(Z)=\mathrm{V}\left(\frac{1}{\sigma}X-\frac{m}{\sigma}\right)=\frac{1}{\sigma^2}\mathrm{V}(X)=\frac{\sigma^2}{\sigma^2}=1$$

$$\sigma(Z)=\sqrt{\mathrm{V}(Z)}=\sqrt{1}=1$$

3 이항분포와 그 성질

이항분포

한 번의 시행에서 사건 A가 일어날 확률이 p, 일어나지 않을 확률이 q일 때, n회의 독립시행에서 사건 A가 일어난 횟수를 확률변수 X로 하는 확률분포를 이항분포라고 하고, $\mathrm{B}(n,\ p)$로 나타낸다. n번의 독립시행 중 사건 A가 r번 일어날 확률은 다음과 같다.

$$\mathrm{P}(X=r)={}_n\mathrm{C}_r p^r q^{n-r} \quad (\text{단, } p+q=1,\ r=0,\ 1,\ 2,\ \cdots,\ n)$$

➤ 한 번의 시행에서 사건 A가 일어날 확률이 p, 사건 A가 일어나지 않을 확률이 q로 일정하고, 각각의 시행의 결과가 서로 영향을 주지 않는 독립시행일 때, 이 시행을 n번 반복하는 독립시행에서 사건 A가 r번 일어날 확률은 독립시행의 확률로부터

$${}_n\mathrm{C}_r p^r q^{n-r} \quad (r=0,\ 1,\ 2,\ \cdots,\ n)$$

임을 알 수 있다. 이때, n번의 독립시행에서 사건 A가 일어난 횟수를 확률변수 X라고 하면 X는 $0, 1, 2, \cdots, n$의 값을 갖는 이산확률변수이고, X의 확률분포는 다음과 같다.

X	0	1	2	$\cdots$	x	$\cdots$	n	합계
$\mathrm{P}(X=x)$	${}_n\mathrm{C}_0 q^n$	${}_n\mathrm{C}_1 pq^{n-1}$	${}_n\mathrm{C}_2 p^2 q^{n-2}$	$\cdots$	${}_n\mathrm{C}_x p^x q^{n-x}$	$\cdots$	${}_n\mathrm{C}_n p^n$	1

➤ 이항정리를 이용하면 $\sum\limits_{x=0}^{n} {}_n\mathrm{C}_x p^x q^{n-x}=1$임을 알 수 있다.

예제6

한 개의 주사위를 세 번 던지는 시행에서 2이하의 눈이 나오는 횟수를 확률변수 X라고 할 때, 다음 물음에 답하시오.

(1) 확률변수 X의 확률분포를 구하시오.

(2) $\mathrm{P}(X \geq 2)$를 구하시오.

(1) 확률변수 X는 이항분포 $\mathrm{B}\left(3,\ \frac{1}{3}\right)$을 따르므로 X의 확률분포는 다음과 같다.

X	0	1	2	3	합계
$\mathrm{P}(X=x)$	${}_3\mathrm{C}_0\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^3$	${}_3\mathrm{C}_1\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2$	${}_3\mathrm{C}_2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1$	${}_3\mathrm{C}_3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^0$	1

(2) $$\begin{aligned}\mathrm{P}(X \geq 2) &= \mathrm{P}(X=2)+\mathrm{P}(X=3)\\ &= {}_3\mathrm{C}_2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1 + {}_3\mathrm{C}_3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^0\\ &= \frac{6}{27}+\frac{1}{27}\\ &= \frac{7}{27}\end{aligned}$$

이항분포의 평균, 분산, 표준편차

확률변수 X가 이항분포 $\mathrm{B}(n\ p)$를 따를 때, 확률변수 X의 평균, 분산, 표준편차는 각각 다음과 같다. (단, $q = 1 - p$)

(1) $\mathrm{E}(X) = np$

(2) $\mathrm{V}(X) = npq$

(3) $\sigma(X) = \sqrt{npq}$

➤
$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{r=0}^{n} \{r \times \mathrm{P}(X = r)\} \\
&= \sum_{r=0}^{n} \{r \times {}_n\mathrm{C}_r p^r q^{n-r}\} \\
&= \sum_{r=0}^{n} \{n \times {}_{n-1}\mathrm{C}_{r-1} p^r q^{n-r}\} \\
&= np \times \sum_{r=0}^{n} \{{}_{n-1}\mathrm{C}_{r-1} p^{r-1} q^{n-r}\} \\
&= np(p+q)^{n-1} \\
&= np
\end{aligned}
$$

$$
\begin{aligned}
r \times {}_n\mathrm{C}_r &= r \times \frac{n!}{(n-r)! \times r!} \\
&= n \times \frac{(n-1)!}{(n-r)! \times (r-1)!} \\
&= n \times {}_{n-1}\mathrm{C}_{r-1}
\end{aligned}
$$

예제6

어떤 질병에 대한 치료율이 60%인 의약품으로 200명의 환자가 치료를 받고 있다. 이 중에서 치료될 환자의 수를 X라고 할 때, 확률변수 X의 평균과 표준편차를 구하시오.

확률변수 X는 이항분포 $\mathrm{B}\left(200, \dfrac{3}{5}\right)$를 따르므로, 확률변수 X의 평균, 분산, 표준편차는 다음과 같이 구할 수 있다.

$$\mathrm{E}(X) = 200 \times \frac{3}{5} = 120$$

$$\mathrm{V}(X) = 200 \times \frac{3}{5} \times \frac{2}{5} = 48$$

$$\sigma(X) = \sqrt{48} = 4\sqrt{3}$$

이항분포의 분산과 표준편차

➤ $$\mathrm{E}\left(x^2\right) = \sum_{r=0}^{n}\left\{r^2 \times \mathrm{P}(X=r)\right\}$$

$$= \sum_{r=0}^{n}\left\{r^2 \times {}_n\mathrm{C}_r p^r q^{n-r}\right\}$$

$$= \sum_{r=0}^{n}\left\{\left(r^2 - r + r\right) \times {}_n\mathrm{C}_r p^r q^{n-r}\right\}$$

$$= \sum_{r=0}^{n}\left\{r(r-1) \times {}_n\mathrm{C}_r p^r q^{n-r}\right\} + \sum_{r=0}^{n}\left\{r \times {}_n\mathrm{C}_r p^r q^{n-r}\right\}$$

$$= \sum_{r=0}^{n}\left\{r(r-1) \times {}_n\mathrm{C}_r p^r q^{n-r}\right\} + np$$

$$= \sum_{r=2}^{n}\left\{r(r-1) \times \frac{n!}{(n-r)! \times r!} \times p^r q^{n-r}\right\} + np$$

$$= n(n-1)p^2 \times \sum_{r=2}^{n}\left\{\frac{(n-2)!}{(n-r)! \times r!} \times p^{r-2} q^{(n-2)-(r-2)}\right\} + np$$

$$= n(n-1)p^2 \times \sum_{k=0}^{n-2}\left\{{}_{n-2}\mathrm{C}_k p^k q^{n-2-k}\right\} + np \qquad \leftarrow k = r-2$$

$$= n(n-1)p^2(p+q)^{n-2} + np$$

$$= n^2p^2 - np^2 + np$$

➤ $$\mathrm{V}(X) = \mathrm{E}\left(X^2\right) - \{\mathrm{E}(X)\}^2$$

$$= n^2p^2 - np^2 + np - (np)^2$$

$$= np(1-p)$$

$$= npq$$

➤ $$\sigma(X) = \sqrt{\mathrm{V}(X)} = \sqrt{npq}$$

큰 수의 법칙

어떤 한 시행에서 사건 A가 일어날 수학적 확률이 p일 때, n번의 독립시행에서 사건 A가 일어나는 횟수를 X라고 하면, 임의의 작은 양수 h에 대해서 다음이 성립한다.

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{X}{n} - p\right| < h\right) = 1$$

➤ 한 개의 주사위를 n회 던지는 시행에서 1의 눈이 나오는 횟수를 X라고 하자. n이 10, 30, 50인 경우에 대해서 $\mathrm{P}\left(\left|\frac{X}{n} - p\right| < 0.1\right)$를 구해보면 다음과 같다.

(1) $n = 10$

$\left|\frac{X}{10} - \frac{1}{6}\right| < 0.1 \quad \Leftrightarrow \quad \frac{2}{3} < X < \frac{8}{3}$이므로

$\mathrm{P}\left(\left|\frac{X}{10} - \frac{1}{6}\right| < 0.1\right) = \mathrm{P}(X = 1) + \mathrm{P}(X = 2) = 0.323 + 0.291 = 0.614$

(2) $n = 30$

$\left|\frac{X}{30} - \frac{1}{6}\right| < 0.1 \quad \Leftrightarrow \quad 2 < X < 8$이므로

$\mathrm{P}\left(\left|\frac{X}{30} - \frac{1}{6}\right| < 0.1\right) = \sum_{k=3}^{7} \mathrm{P}(X = k) = 0.784$

(3) $n = 50$

$\left|\frac{X}{50} - \frac{1}{6}\right| < 0.1 \quad \Leftrightarrow \quad \frac{10}{3} < X < \frac{40}{3}$이므로

$\mathrm{P}\left(\left|\frac{X}{50} - \frac{1}{6}\right| < 0.1\right) = \sum_{k=4}^{13} \mathrm{P}(X = k) = 0.946$

➤ 위의 결과로부터 시행 횟수 n이 커짐에 따라 상대도수 $\frac{X}{n}$와 수학적 확률 $\frac{1}{6}$의차가 0.1 미만일 확률이 1에 점점 가까워짐을 확인할 수 있다.

➤ 큰 수의 법칙에 의하면 시행의 횟수가 충분히 클 때, 통계적 확률은 수학적 확률에 가까워지므로 사건 A의 상대도수 $\frac{X}{n}$를 사건 A가 일어날 확률 $\mathrm{P}(A)$의 근삿값으로 사용할 수 있다. 그러므로 자연 현상이나 사회 현상에서 수학적 확률을 구하기 곤란한 경우, 통계적 확률을 대신 사용할 수 있다.

4 정규분포와 그 성질

정규분포

연속확률변수 X의 확률밀도함수 $f(x)$가

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (\text{단}, \ -\infty < x < \infty, \ e = 2.71828\cdots)$$

으로 주어질 때, X의 확률분포를 정규분포라 하고, $\mathrm{N}\left(m, \ \sigma^2\right)$으로 나타낸다. m과 σ는 각각 확률변수 X의 평균과 표준편차를 나타낸다.

정규분포 곡선의 성질

정규분포의 확률밀도함수의 그래프인 정규분포 곡선은 다음과 같은 성질을 갖는다.

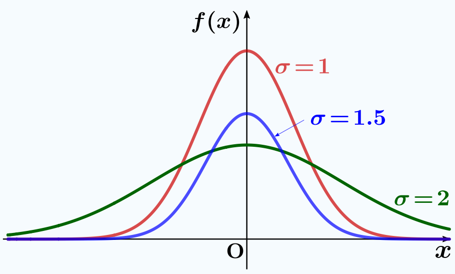
(1) 직선 $x = m$에 대하여 대칭이고, x축이 점근선이다.

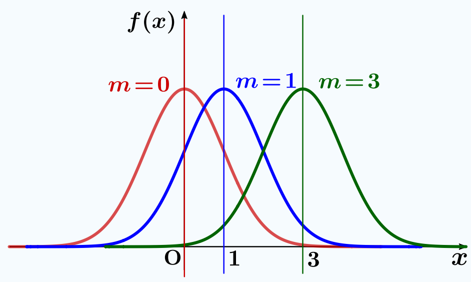
(2) $x = m$일 때, 최댓값 $\dfrac{1}{\sqrt{2\pi}\sigma}$를 갖는다.

(3) 곡선과 x축 사이의 넓이는 1이다.

(4) m이 일정할 때, 표준편차 σ가 클수록 곡선의 높이는 낮아지고, 폭은 넓어진다.
반대로 σ가 작을수록 곡선의 높이는 높아지고 폭은 좁아진다.

(5) σ가 일정할 때, m이 변하면 대칭축의 위치는 변하지만 곡선의 모양은 변하지 않는다.





➤ 강수량, 기온, 물체의 길이, 무게, 학생들의 성적 등과 같이 자연 현상이나 사회 현상에서 나타나는 연속확률변수의 확률밀도함수의 그래프는 정규분포 곡선과 같이 좌우가 대칭인 곡선에 가까운 경우가 많다. 따라서 이러한 현상들을 설명하는데 정규분포가 유용하게 사용된다.

표준정규분포

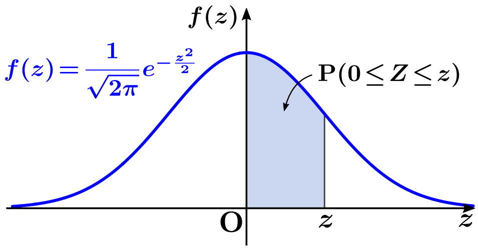
평균이 0, 표준편차가 1인 정규분포 $\mathrm{N}\left(0,\ 1^2\right)$을 표준정규분포라고 한다. 표준정규분포의 확률변수는 Z로 나타내며, 표준정규분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \quad (-\infty < z < \infty,\ e = 2.71828\cdots)$$

➤ 임의의 양수 z에 대하여 확률

$$\mathrm{P}(0 \le Z \le z)$$

는 오른쪽 그림에서 색칠된 영역의 넓이와 같다. 이 확률을 구하여 표로 나타낸 것이 표준정규분포표다.



➤ 표준정규분포표로부터 확률 구하기
표준정규분포표의 가장 왼쪽 세로줄에서 소수 첫 번째 자리 수까지 찾은 다음, 가장 위쪽 가로줄에서 소수 두 번째 자리 수를 찾아 세로줄과 가로줄이 만나는 곳의 값을 읽는다.

$$\mathrm{P}(0 \le Z \le 1.96) = 0.4750$$

$$\mathrm{P}(0 \le Z \le 2.58) = 0.4951$$

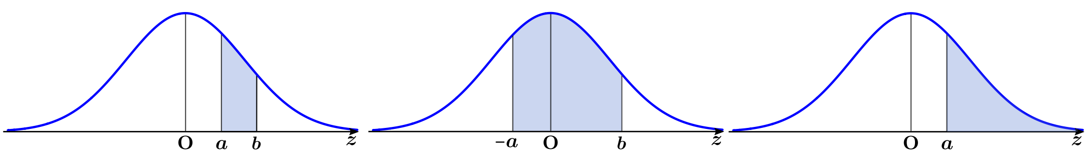
	0	$\cdots$	6	7	8	9
0.0	0.0000	$\cdots$	0.0239	0.0279	0.0319	0.0359
0.1	0.0398		0.0636	0.0675	0.0714	0.0753
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1.9	0.4713	$\cdots$	0.4750	0.4756	0.4761	0.4767
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2.5	0.4938	$\cdots$	0.4948	0.4949	0.4951	0.4952
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

➤ $a > 0,\ b > 0$일 때, 표준정규분포표의 이용법

① $\mathrm{P}(a \le Z \le b) = \mathrm{P}(0 \le Z \le b) - \mathrm{P}(0 \le Z \le a)$

② $\mathrm{P}(-a \le Z \le b) = \mathrm{P}(-a \le Z \le 0) + \mathrm{P}(0 \le Z \le b) = \mathrm{P}(0 \le Z \le a) + \mathrm{P}(0 \le Z \le b)$

③ $\mathrm{P}(Z \ge z) = 0.5 - \mathrm{P}(0 \le Z \le a)$

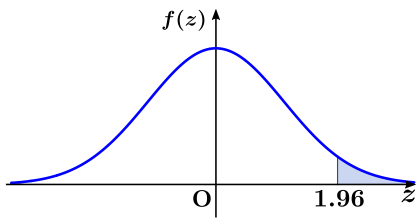


예제7

확률변수 Z가 표준정규분포를 따를 때, 표준정규분포표를 이용하여 $\mathrm{P}(Z > 1.96)$을 구하시오.

오른쪽 표준정규분포곡선과 z축 사이의 넓이는 1이고, 표준정규분포곡선은 $z = 0$에 대하여 대칭이므로 $\mathrm{P}(Z \geq 0) = 0.5$이다. 또한, 표준정규분포표에서 $\mathrm{P}(0 \leq z \leq 1.96) = 0.4750$이므로

$$\begin{aligned}\mathrm{P}(Z > 1.96) &= 0.5 - \mathrm{P}(0 \leq Z \leq 1.96) \\ &= 0.5 - 0.4750 \\ &= 0.025\end{aligned}$$



예제8

확률변수 Z가 표준정규분포를 따를 때, 표준정규분포표를 이용하여 $\mathrm{P}(|Z| < 2)$을 구하시오.

표준정규분포표에서 $\mathrm{P}(0 \leq Z \leq 2) = 0.4772$이다.

$$\begin{aligned}\therefore\ \mathrm{P}(|Z| < 2) = \mathrm{P}(-2 < Z < 2) &= \mathrm{P}(-2 < Z < 0) + \mathrm{P}(0 \leq Z < 2) \\ &= \mathrm{P}(0 < Z < 2) + \mathrm{P}(0 \leq Z < 2) \\ &= 2 \times \mathrm{P}(0 \leq Z \leq 2) \\ &= 0.9544\end{aligned}$$

예제9

확률변수 Z가 표준정규분포를 따를 때, 표준정규분포표를 이용하여 $\mathrm{P}(Z > -2.58)$을 구하시오.

표준정규분포표에서 $\mathrm{P}(0 \leq Z \leq 2.58) = 0.4951$이다.

$$\begin{aligned}\therefore\ \mathrm{P}(Z > -2.58) &= \mathrm{P}(-2.58 < Z < 0) + \mathrm{P}(Z \geq 0) \\ &= \mathrm{P}(0 < Z < 2.58) + 0.5 \\ &= 0.9951\end{aligned}$$

정규분포의 표준화

확률변수 X가 정규분포 $\mathrm{N}\left(m,\ \sigma^2\right)$을 따르면 확률변수 $Z = \dfrac{X-m}{\sigma}$은 표준정규분포 $\mathrm{N}\left(0,\ 1^2\right)$을 따른다는 사실이 알려져 있다. 이와 같이 정규분포 $\mathrm{N}\left(m,\ \sigma^2\right)$을 표준정규분포 $\mathrm{N}\left(0,\ 1^2\right)$으로 변환하는 것을 정규분포의 표준화라고 한다.

➤ X가 정규분포 $\mathrm{N}\left(m,\ \sigma^2\right)$을 따를 때, $Z = \dfrac{X-m}{\sigma}$이라고 하면

$$\mathrm{E}(Z) = \mathrm{E}\left(\frac{X-m}{\sigma}\right) = \frac{\mathrm{E}(X)}{\sigma} - \frac{m}{\sigma} = \frac{m}{\sigma} - \frac{m}{\sigma} = 0$$

$$\sigma(Z) = \sqrt{\mathrm{V}\left(\frac{X-m}{\sigma}\right)} = \sqrt{\frac{\mathrm{V}(X)}{\sigma^2}} = \frac{\sigma(X)}{\sigma} = \frac{\sigma}{\sigma} = 1$$

➤ 정규분포 $\mathrm{N}\left(m,\ \sigma^2\right)$을 따르는 확률변수 X에 대하여 확률 $\mathrm{P}(a \le X \le b)$는 표준화를 통하여 다음과 같이 구할 수 있다.

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

예제10

확률변수 X가 정규분포 $\mathrm{N}\left(70,\ 3^2\right)$을 따를 때, 오른쪽 표준정규분포표를 이용하여 다음을 구하시오.

(1) $\mathrm{P}(70 \le X \le 76)$ (2) $\mathrm{P}(67 \le X \le 73)$

(3) $\mathrm{P}(73 \le X \le 76)$

Z	$\mathrm{P}(0 \le Z \le z)$
1	0.3413
2	0.4772

(1) $\mathrm{P}(70 \le X \le 76) = \mathrm{P}\left(\dfrac{70-70}{3} \le Z \le \dfrac{76-70}{3}\right) = \mathrm{P}(0 \le Z \le 2) = 0.4772$

(2) $\mathrm{P}(67 \le X \le 73) = \mathrm{P}\left(\dfrac{67-70}{3} \le Z \le \dfrac{73-70}{3}\right) = \mathrm{P}(-1 \le Z \le 1) = 2 \times 0.3413$

$= 0.6826$

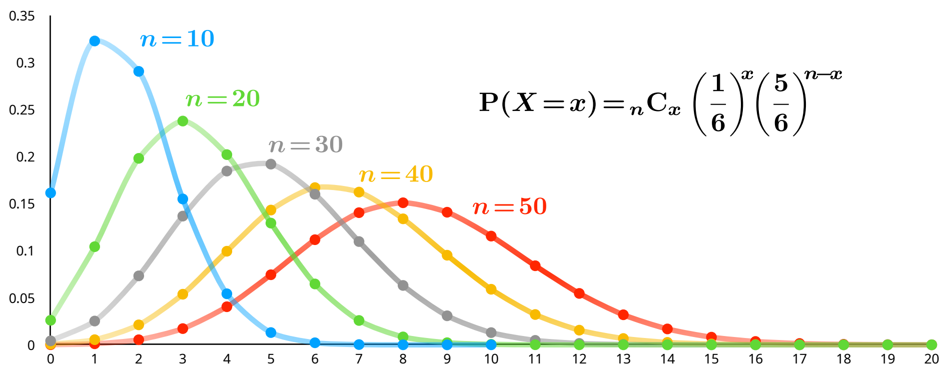
(3) $\mathrm{P}(73 \le X \le 76) = \mathrm{P}\left(\dfrac{73-70}{3} \le Z \le \dfrac{76-70}{3}\right) = \mathrm{P}(1 \le Z \le 2) = 0.4772 - 0.3413$

$= 0.1359$

이항분포와 정규분포의 관계

확률변수 X가 이항분포 $\mathrm{B}(n,\ p)$를 따르고, 시행 횟수 n이 충분히 크다면 확률변수 X의 분포는 근사적으로 정규분포 $\mathrm{N}(np, npq)$를 따른다는 것이 알려져 있다. (단, $p+q=1$)

➤ 주사위를 n회 던지는 시행에서 1의 눈이 나오는 횟수를 확률변수 X라고 하면, X는 이항분포 $\mathrm{B}\left(n, \frac{1}{6}\right)$을 따르게 된다. 이때, 시행 횟수 n이 커짐에 따라 이항분포의 확률값들의 형태가 아래 그림에서처럼 정규분포의 형태(bell-shpace)를 보임을 알 수 있다.



예제11

한 개의 주사위를 720번 던지는 시행에서 1의 눈이 95회 이상 130회 이하로 나올 확률을 구하시오. (단, $\mathrm{P}(0 \le Z \le 1) = 0.3413$, $\mathrm{P}(0 \le Z \le 2.5) = 0.4938$)

720회의 독립시행에서 1의 눈이 나오는 횟수를 확률변수 X라고 하면,
X는 이항분포 $\mathrm{B}\left(720, \frac{1}{6}\right)$을 따르므로

$$\mathrm{E}(X) = 720 \times \frac{1}{6} = 120, \quad \sigma(X) = \sqrt{720 \times \frac{1}{6} \times \frac{5}{6}} = 10$$

이 된다. 이때, $n = 720$은 충분히 큰 수이므로 X는 근사적으로 정규분포 $\mathrm{N}\left(120,\ 10^2\right)$을 따른다고 볼 수 있다.

$$\begin{aligned}\therefore \mathrm{P}(95 \le X \le 130) &= \mathrm{P}\left(\frac{95-120}{10} \le \frac{X-120}{10} \le \frac{130-120}{10}\right)\\ &= \mathrm{P}(-2.5 \le Z \le 1)\\ &= \mathrm{P}(-2.5 \le Z \le 0) + \mathrm{P}(0 \le Z \le 1)\\ &= 0.4938 + 0.3413 = 0.8351\end{aligned}$$

1 모집단과 표본

모집단과 표본

(1) 모집단과 표본
통계조사에서 조사 대상 전체를 모집단이라 하고, 모집단 중 통계조사를 위해 추출된 조사 대상을 표본이라고 한다. 이때, 표본에 포함되는 자료의 개수를 표본의 크기라고 하고, 모집단에서 표본을 뽑는 것을 추출이라고 한다.

(2) 전수조사와 표본조사
통계조사의 대상이 되는 자료 전체를 조사하는 것을 전수조사라 하고, 조사 대상의 일부만을 추출하여 조사함으로써 전체를 추정하는 조사를 표본조사라고 한다.

(3) 표본의 추출

① 복원추출과 비복원추출
하나의 자료를 뽑고, 뽑은 자료를 다시 넣어 재추출이 가능하도록 하는 방법을 복원추출이라 한다. 한 번 뽑은 자료는 다시 넣지 않아 재추출이 불가능하게 하는 방법을 비복원추출이라 한다.

② 임의추출
각 대상이 추출될 확률이 동일하게 되도록 하는 방법을 임의추출이라 한다.

모평균, 모분산, 모표준편차

(1) 모집단의 분포
모집단의 특성을 나타내는 확률변수 X의 확률분포를 모집단의 분포라고 한다.

(2) 모평균, 모분산, 모표준편차
모집단의 특성을 나타내는 확률변수 X의 평균, 분산, 표준편차를 각각 모평균, 모분산, 모표준편차라 하고, 이것을 각각 m, σ^2, σ로 나타낸다.

표본평균, 표본분산, 표본표준편차

어떤 모집단에서 임의추출한 크기 $n\ (n \geq 2)$인 표본을 $X_1,\ X_2,\ X_3,\ \cdots,\ X_n$이라 할 때, 이 표본의 평균, 분산, 표준편차를 각각 표본평균, 표본분산, 표본표준편차라 하고, 이것을 각각 $\overline{X},\ S^2,\ S$로 나타낸다.

(1) 표본평균 : $\overline{X} = \dfrac{1}{n}(X_1 + X_2 + X_3 + \cdots + X_n)$

(2) 표본분산 : $S^2 = \dfrac{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \cdots + \left(X_n - \overline{X}\right)^2}{n-1}$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

(3) 표본표준편차 : $S = \sqrt{S^2}$

➤ 표본분산을 구할 때, n이 아닌 $n-1$로 나누는 이유는 S^2의 기댓값이 σ^2이 되도록 하기 위함이다. (고등학교 수학의 범위를 벗어나므로 자세한 내용은 생략한다.)

예제12

자연수 $1,\ 2,\ \cdots,\ 10$으로 구성된 모집단에서 크기 3인 표본 $6,\ 8,\ 10$이 추출되었다고 할 때, 다음 물음에 답하시오.

(1) 모평균, 모분산, 모표준편차를 구하시오.

(2) 표본평균, 표본분산, 표본표준편차를 구하시오.

(1) 모평균 $m = \dfrac{1+2+\cdots+10}{10} = 5.5$

모분산 $\sigma^2 = \dfrac{1^2+2^2+\cdots+10^2}{10} - (5.5)^2 = 8.25$

모표준편차 $\sigma = \sqrt{8.25}$

(2) 표본평균 $\overline{X} = \dfrac{6+8+10}{3} = 8$

표본분산 $S^2 = \dfrac{1}{3-1}\left\{(6-8)^2 + (8-8)^2 + (10-8)^2\right\} = 4$

표본표준편차 $S = \sqrt{4} = 2$

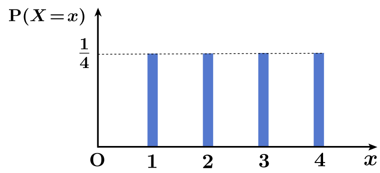
표본평균의 평균, 분산, 표준편차

평균이 m이고, 표준편차가 σ인 모집단에서 복원으로 임의추출한 크기 n인 표본의 표본평균 $\overline{X}$의 평균, 분산, 표준편차에 대하여 다음이 성립한다는 것이 알려져 있다.

$$\mathrm{E}\left(\overline{X}\right)=m, \quad \mathrm{V}\left(\overline{X}\right)=\frac{\sigma^2}{n}, \quad \sigma\left(\overline{X}\right)=\frac{\sigma}{\sqrt{n}}$$

- 모평균 m은 상수이지만, 표본평균 $\overline{X}$는 표본의 관측값에 따라 다른 값을 갖게 되므로 확률변수이다.
- 어느 모집단의 확률변수 X가 다음과 같은 확률분포를 갖는다면 확률변수 X의 모평균 m과 모분산 σ^2은 각각 $m=\frac{5}{2}$, $\sigma^2=\frac{5}{4}$이다.

X	1	2	3	4	합계
$\mathrm{P}(X=x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

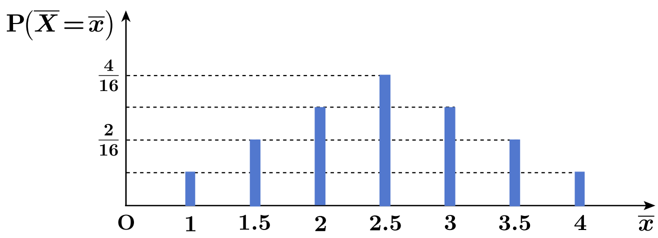


- 위 모집단에서 크기 $n=2$인 표본을 복원추출한 관측값을 X_1, X_2라고 하면 가능한 관측값들은 다음 표와 같다.

(X_1, X_2)	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$
표본평균 $\overline{X}$	1	1.5	2	2.5	1.5	2	2.5	3
(X_1, X_2)	$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$
표본평균 $\overline{X}$	2	2.5	3	3.5	2.5	3	3.5	4

따라서 표본평균 $\overline{X}=\frac{1}{2}(X_1+X_2)$의 확률분포는 다음과 같다.

$\overline{X}$	1	1.5	2	2.5	3	3.5	4	합계
$\mathrm{P}\left(\overline{X}=\overline{x}\right)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	1



이때, 표본평균 $\overline{X}$의 평균 $\mathrm{E}\left(\overline{X}\right)=\frac{5}{2}$, 분산 $\mathrm{V}\left(\overline{X}\right)=\frac{5}{8}$ 가 된다.

- 위의 모평균 m, 모분산 σ^2, 표본평균의 평균 $\mathrm{E}\left(\overline{X}\right)$, 표본평균의 분산 $\mathrm{V}\left(\overline{X}\right)$으로 다음의 관계가 성립함을 알 수 있다.

$$\mathrm{E}\left(\overline{X}\right)=m, \quad \mathrm{V}\left(\overline{X}\right)=\frac{\sigma^2}{n}$$

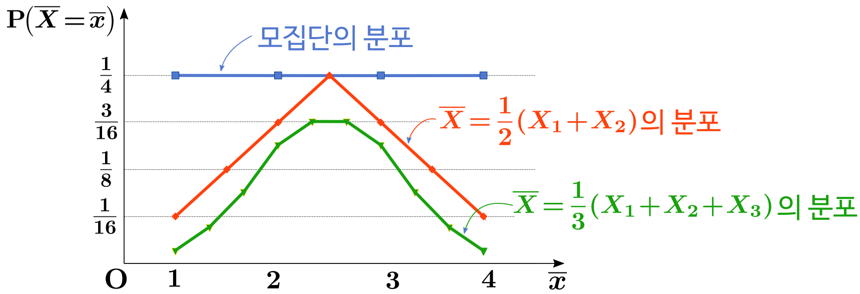
표본평균의 분포

모평균이 m, 모표준편차가 σ인 모집단에서 크기가 n인 표본을 임의추출할 때의 표본평균을 $\overline{X}$라 하면

(1) 모집단의 분포가 정규분포일 때, $\overline{X}$의 분포는 정규분포 $\mathrm{N}\left(m,\ \dfrac{\sigma^2}{n}\right)$을 따른다는 것이 알려져 있다.

(2) 모집단의 분포가 정규분포가 아닐 때에도 n이 충분히 크면 $\overline{X}$의 분포는 근사적으로 정규분포 $\mathrm{N}\left(m,\ \dfrac{\sigma^2}{n}\right)$을 따른다는 것이 알려져 있다.

➤ 표본평균 $\overline{X}$의 확률분포는 오른쪽 그림과 같이 모집단의 분포와는 상관없이 표본의 크기가 커짐에 따라 정규분포에 가까워지는 것을 알 수 있다.



예제13

정규분포 $\mathrm{N}(100,\ 4^2)$을 따르는 모집단에서 크기가 4인 표본을 임의추출할 때, 표본평균 $\overline{X}$의 평균과 분산 및 $\mathrm{P}(\overline{X} \le 104)$을 구하시오. (단, $\mathrm{P}(0 \le Z \le 2) = 0.4772$)

표본평균 $\overline{X}$의 평균과 분산은 각각 $\mathrm{E}(\overline{X}) = 100,\ \mathrm{V}(\overline{X}) = \dfrac{4^2}{4} = 4$이다.

이때, $\overline{X}$는 정규분포 $\mathrm{N}(100,\ 2^2)$을 따르므로 다음과 같이 $\mathrm{P}(\overline{X} \le 104)$를 구할 수 있다.

$$\mathrm{P}(\overline{X} \le 104) = \mathrm{P}\left(Z \le \frac{104-100}{2}\right) = \mathrm{P}(Z \le 2)$$

$$= 0.5 + \mathrm{P}(0 \le Z \le 2) = 0.5 + 0.4772 = 0.9772$$

2 모평균의 추정

모평균의 추정

(1) 추정

모집단의 평균, 표준편차 등을 알지 못할 때, 표본을 이용하여 이들 값을 추측하는 방법을 추정이라고 한다.

(2) 모평균 m의 신뢰구간

모표준편차가 σ이고, 표본의 크기 n이 충분히 클 때의 표본평균을 $\overline{X}$라고 하면, 모평균 m의 신뢰구간은 다음과 같이 구할 수 있다.

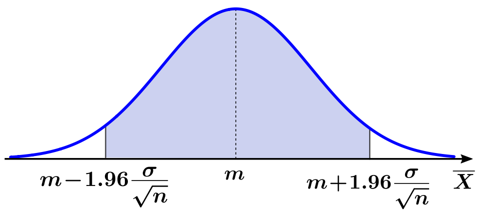
① 신뢰도 95%의 신뢰구간 : $\overline{X}-1.96\frac{\sigma}{\sqrt{n}} \le m \le \overline{X}+1.96\frac{\sigma}{\sqrt{n}}$

② 신뢰도 99%의 신뢰구간 : $\overline{X}-2.58\frac{\sigma}{\sqrt{n}} \le m \le \overline{X}+2.58\frac{\sigma}{\sqrt{n}}$

➤ 모평균 m이 알려져 있지 않고, 모분산 σ^2이 알려져 있는 정규분포를 따르는 모집단에서 임의추출한 크기 n인 표본을 $X_1,\ X_2,\ \cdots,\ X_n$이라고 할 때, 표본평균 $\overline{X}$는 정규분포 $\mathrm{N}\left(m,\ \frac{\sigma^2}{n}\right)$을 따른다. 이때, 확률변수 $Z=\dfrac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$은 표준정규분포 $\mathrm{N}(0,\ 1^2)$을 따르게 된다.

➤ 표준정규분포표에서 $\mathrm{P}(-1.96 \le Z \le 1.96)=\mathrm{P}\left(-1.96 \le \dfrac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}} \le 1.96\right)=0.95$이므로 다음이 성립한다.

$$\mathrm{P}\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}} \le m \le \overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)=0.95$$

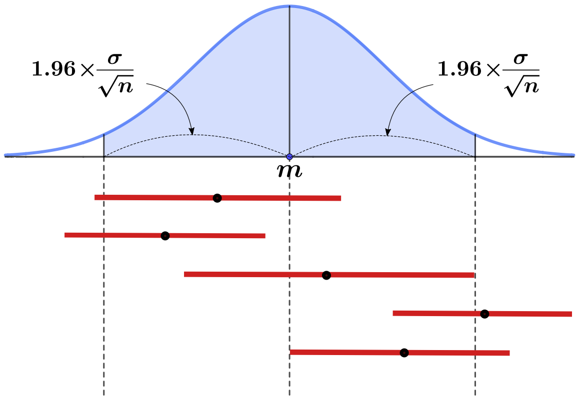


➤ 위 식은 모평균 m이 구간 $\left[\overline{X}-1.96\frac{\sigma}{\sqrt{n}},\ \overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right]$에 속할 확률이 0.95임을 나타내므로, 이 구간을 모평균 m의 신뢰도 95%인 신뢰구간이라고 한다.

➤ $\mathrm{P}(-2.58 \le Z \le 2.58)=0.99$임을 이용하여 위와 같은 방법으로 모평균 m의 신뢰도 99%인 신뢰구간을 구하면 $\left[\overline{X}-2.58\frac{\sigma}{\sqrt{n}},\ \overline{X}+2.58\frac{\sigma}{\sqrt{n}}\right]$이 된다.

➤ 모표준편차 σ가 주어지지 않고, 표본의 표준편차 S가 주어진 경우 σ 대신 S를 사용하여 신뢰구간을 구할 수 있다.

➤ 표본의 관측값 $\overline{X}$는 추출되는 표본에 따라 그 값이 달라지므로 신뢰구간도 표본에 따라 달라지게 된다. 따라서 모평균 m에 대한 신뢰도가 95%라는 의미는 크기가 n인 표본을 계속 추출하여 각 표본에 대한 신뢰구간을 구할 때, 이 중 95%는 모평균 m을 포함하고 있다는 뜻이다. 즉, 100번의 표본을 추출하여 100개의 신뢰구간을 구했을 때, 이 중 95개 정도의 신뢰구간이 모평균 m을 포함하고 있을 것으로 기대된다는 뜻이다. 신뢰도 95%의 의미가 한 번의 표본추출에서 얻은 신뢰구간이 모평균을 포함하고 있을 확률이 0.95가 된다는 뜻이 아님에 주의해야 한다.



➤ 일반적으로 신뢰구간의 크기 혹은 길이는 신뢰구간 양 끝값의 차를 일컫는다. 따라서 신뢰도 95%, 99% 신뢰구간의 길이는 각각 다음과 같다.

$$2\times 1.96\frac{\sigma}{\sqrt{n}},\quad 2\times 2.58\frac{\sigma}{\sqrt{1.96n}}$$