

I. 순열과 조합

1. 순열과 조합

- ☑ 시행(trial) : 어떤 실험, 관찰, 행위.
 사건(event) : 시행의 결과로서 나타나는 현상.
 경우의 수 : 어떤 시행에서 가능한 사건의 수.
 $n(A)$: 사건 A 가 일어나는 경우의 수

① 경우의 수

↳ a number of occasions

{ 사건 A 가 일어나는 경우의 수 : m 일 때,
 사건 B 가 일어나는 경우의 수 : n

1. 합의 법칙 : A, B 가 동시에 일어나지 않는 경우

$$(1) (A \text{ 또는 } B \text{ 가 일어나는 경우의 수}) = m + n$$

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

$$\square A \cap B = \emptyset \Rightarrow n(A \cup B) = n(A) + n(B)$$

(2) (A 또는 B 또는 C 가 일어나는 경우의 수)

$$= n(A \cup B \cup C)$$

$$= n(A) + n(B) + n(C)$$

$$- n(A \cap B) - n(B \cap C) - n(C \cap A)$$

$$+ n(A \cap B \cap C)$$

2. 곱의 법칙 : A 와 B 가 동시에 일어나는 경우

$$(A, B \text{ 가 동시에 일어나는 경우의 수}) = m \times n$$

$$n(A \cap B) = n(A) \times n(B)$$

☑ 한 사건 A 가 m 가지의 경우로 일어나고,
 그 각각에 대하여 다른 사건 B 가 n 가지의 경우로
 일어날 때

$$\square (\text{동시에}) = (\text{잇달아}) = (\text{연속적으로})$$

3. 집합과 경우의 수

$$(1) n(A^C) = n(U) - n(A)$$

$$(2) n(A - B) = n(A \cap B^C)$$

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

$$(3) n(A^C \cap B^C) = n\{(A \cup B)^C\}$$

$$= n(U) - n(A \cup B)$$

$$(4) n(A \cup B \cup C) = n(A) + n(B) + n(C)$$

$$- n(A \cap B) - n(B \cap C) - n(C \cap A)$$

$$+ n(A \cap B \cap C)$$

4. 주의해야할 것

(1) 모든 경우를 빠짐없이, 중복되지 않게 세어야 한다.
 \Rightarrow 사전식 배열법 or 수형도(樹型圖)

$$(2) n(\text{적어도 } \sim) = n(\text{전체}) - n(\text{반대인 경우})$$

(3) 정수를 만드는 문제

\Rightarrow 맨 앞자리(최고자리 숫자) 0 이 올 수 없다.

5. 화폐의 지불 방법과 지불 금액의 가지 수

A 원 권 p 장, B 원 권 q 장, C 원 권 r 장으로 지불

$$(1) \text{지불 방법의 수} : (p+1)(q+1)(r+1) - 1$$

(2) 지불 금액의 수

① 화폐 액면이 중복되지 않을 때

$$\Rightarrow (p+1)(q+1)(r+1) - 1$$

② 화폐 액면이 중복될 때

\Rightarrow 큰 액면을 작은 액면으로 바꾸어 \rightarrow ①

6. 경우의 수의 응용

(1) $ax + by + cz = k$ 의 꼴의 부정방정식의 해

\Rightarrow 계수가 큰 항을 기준으로 각각 분류한다.

(2) 양의 약수의 개수

$$N = a^p \cdot b^q \cdot c^r \quad (a, b, c \text{ 는 서로 다른 소수})$$

$$\textcircled{1} \text{ 양의 약수의 개수 } \Rightarrow (p+1)(q+1)(r+1)$$

② 양의 약수의 총합

$$\Rightarrow \frac{a^{p+1}-1}{a-1} \times \frac{b^{q+1}-1}{b-1} \times \frac{c^{r+1}-1}{c-1}$$

(3) $n(a, b$ 의 공약수)

$$= n(a, b \text{ 의 최대공약수의 약수})$$

7. 비둘기 집의 원리

(1) 9 개의 비둘기 집에 10 마리의 비둘기가 들어갔다면 적어도 어느 한 집에는 두 마리 이상의 비둘기가 있다.

(2) n 개의 비둘기 집에 $(n+1)$ 마리 이상의 비둘기가 들어갔다면 적어도 어느 한 집에는 두 마리 이상의 비둘기 있다.

(3) n 개의 비둘기 집에 $(nk+1)$ 마리 이상의 비둘기가 들어갔다면 적어도 어느 한 집에는 $(k+1)$ 마리 이상의 비둘기가 있다.

(4) $1 < n < a$ 일 때, n 개의 방에 a 이상의 사람을 넣으려고 하면 어딘가에는 $\left\lceil \frac{a-1}{n} \right\rceil + 1$ 명 이상의 사람이 들어가 있는 방이 있다.

※ 순열과 조합

	중복 허용 ×	중복 허용 ○
순서 ○	순열 ${}_nP_r$	중복순열 ${}_n\Pi_r$
순서 ×	조합 ${}_nC_r$	중복조합 ${}_nH_r$

② 순열 $\left\{ \begin{array}{l} \text{서로 다른 } n \text{ 개 중에서} \\ \text{서로 다른 } r \text{ 개를 뽑아서 } \iff {}_nP_r \\ \text{일렬로 나열 (중복 ×)} \end{array} \right.$

1. 순열(Permutation) \Rightarrow (순서 ○, 중복 ×)

(o) 서로 다른 n 개의 물건 중에서 r 개를 뽑아서
순서 있게 나열 하는 것 \Rightarrow (기호) ${}_nP_r$
↳ 순서가 바뀌면 다른 것으로 취급.

(1) 서로 다른 n 개에서 r 개를 택하여 일렬로 배열하는 순열의 수

$${}_nP_r = n(n-1)(n-2) \cdots (n-r+1)$$

$\underbrace{\hspace{10em}}_{r \text{ 개}} \quad \text{(단, } 0 \leq r \leq n)$

☑ n 부터 시작하여 1 만큼씩 작은 수를 차례로 r 개 곱한다.

(2) 계승(factorial) $n! = {}_nP_n = n(n-1) \cdots 2 \cdot 1$

$$\begin{array}{lll} \textcircled{1} 1! = 1 & \textcircled{2} 2! = 2 & \textcircled{3} 3! = 6 \\ \textcircled{4} 4! = 24 & \textcircled{5} 5! = 120 & \textcircled{6} 6! = 720 \\ \textcircled{7} 7! = 5040 & \textcircled{8} 8! = 40320 & \\ \textcircled{9} 9! = 362880 & \textcircled{10} 10! = 3628800 & \\ & & = 2^8 \times 3^4 \times 5^2 \times 7 \end{array}$$

(3) 순열의 공식

$$\textcircled{1} {}_nP_r = \frac{n!}{(n-r)!} \quad (0 \leq r \leq n)$$

$$\textcircled{2} 0! = 1, {}_nP_0 = 1 \quad \underbrace{{}_nP_r = n(n-1) \cdots (n-r+1)}_{r \text{ 개}}$$

$$\textcircled{3} {}_nP_r = n \cdot {}_{n-1}P_{r-1} \quad (1 \leq r \leq n)$$

☑ 먼저 특정한 1 개를 뽑아 첫 번째에 고정시킨 후,
나머지 $(n-1)$ 개에서 $(r-1)$ 개를 택하여 나열

$$\textcircled{4} {}_nP_r = {}_{n-1}P_r + r \cdot {}_{n-1}P_{r-1} \quad (1 < r \leq n)$$

☑ i) 특정한 1 개가 포함되지 않는 경우 ;

나머지 $(n-1)$ 개에서 r 개를 택하여 나열

ii) 특정한 1 개가 반드시 포함되는 경우 ;

먼저 특정한 1 개를 정하는 방법 r 가지

나머지 $(n-1)$ 개에서 $(r-1)$ 개를 택하여 나열

$$\textcircled{5} {}_nP_r = (n-r+1) \cdot {}_nP_{r-1} \quad (1 < r \leq n)$$

$$\textcircled{6} {}_nP_{r+s} = {}_nP_r \cdot {}_{n-r}P_s$$

$$\textcircled{7} n \cdot n! = \{(n+1)-1\} \cdot n! = (n+1)! - n!$$

(4) 피보나치 순열 $\Rightarrow a_{n+2} = a_n + a_{n+1}$

$$\textcircled{1} \{a_n\} : 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

$$\textcircled{2} a_1 = 1, a_2 = 1, a_{n+2} = a_n + a_{n+1}$$

$$\Rightarrow a_{m+n} = a_{m-1} \cdot a_n + a_m \cdot a_{n+1}$$

$$\textcircled{3} a_n = \frac{1}{\sqrt{5}} \left\{ \left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right\}$$

④ 이웃한 두 항의 비의 극한은 황금분할의 비.

$$\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = \frac{1 + \sqrt{5}}{2}$$

⑤ 자연수를 1 과 2 의 합으로 나타내는 방법의 수
= 계단을 1 계단 또는 2 계단씩 오르는 방법의 수

⑥ 앵무조개(Nautilus)의 단면의 길이

⑦ 흰 바둑돌과 검은 바둑돌 n 개를 일렬로 나열
하되, 흰 바둑돌끼리는 이웃하지 않도록 나열
하는 방법의 수

⑧ 벌이 각 방에서 오른쪽 또는 오른쪽 위, 오른쪽
아래의 방향으로 이동할 수 있다고 할 때, 현재의
위치에서 출발하여 각 방에 도달할 수 있는 경로
의 수

⑨ 하나의 정사각형을 서로 다른 정사각형으로 분할
하는 문제와 같은 분할수수께끼, 그림, 잎차레,
해바라기 꽃, 솔방울의 소용돌이 등에 적용.

(5) Mon-Mort (완전 or 교란) 순열

$f(n) \neq n$ 인 함수의 개수 : a_n

$$\Rightarrow a_{n+2} = (n+1) \{a_n + a_{n+1}\}$$

2. 이웃할 때와 이웃하지 않을 때의 순열

(1) 이웃하게 나열하는 순열의 수

① 이웃하는 것들 \Rightarrow 한 묶음으로 생각

② (묶음을 하나로 보고 구한 순열의 수)
 \times (묶음 안에서 순열의 수)

(2) 2 명이 이웃하지 않게 나열하는 순열의 수

$$= n(\text{전체}) - n(\text{이웃하여 나열})$$

(3) 3 명 이상이 이웃하지 않게 나열하는 순열의 수

① 이웃할 수 있는 것을 먼저 나열

② 그 양끝과 사이사이에 나머지를 배열

(4) 같은 수를 교대로 나열하는 순열의 수

\Rightarrow 각각을 일렬로 배열하여 곱한다.

3. 원순열

(a) 서로 다른 n 개를 원형으로 배열하는 순열

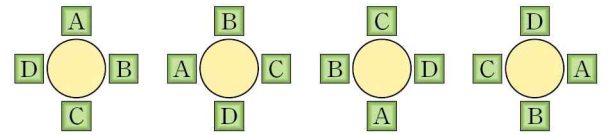
(1) 서로 다른 n 개를 원형으로 배열하는 원순열

$$\Rightarrow \frac{{}_n P_n}{n} = \frac{n!}{n} = (n-1)!$$

☑ ① 합동인 도형의 수 n 으로 나눈다.

② 기준 1 명을 제외시키고 나머지를 일렬로 배열

예) A, B, C, D 의 네 사람이 원탁에 앉는 방법의 수



$$\Rightarrow \frac{4!}{4} = (4-1)! = 3!$$

(2) 서로 다른 n 개에서 r 개를 택하여 원형으로
배열하는 원순열

$$\Rightarrow \frac{{}_n P_r}{r} = {}_n C_r \cdot \frac{r!}{r}$$

(3) 서로 같은 것이 각각 p 개, q 개, r 개 들어 있는
 n 개를 원형으로 배열하는 원순열

$$\Rightarrow \frac{(n-1)!}{p! q! r!} \quad (\text{단, } p, q, r \text{ 은 서로소})$$

(4) 서로 다른 n 개의 구슬을 실에 꿰는 목걸이순열

\Leftrightarrow 뒤집어 놓아도 같은 것이 되는 순열

$$\Rightarrow \frac{1}{2} (n-1)!$$

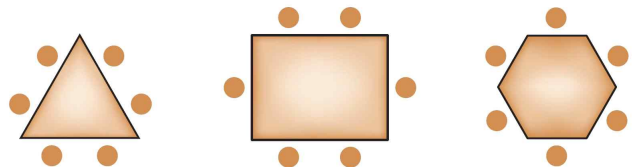
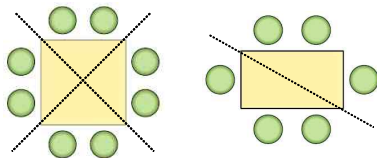
(5) 다각형의 순열의 수(대칭인 경우)

① 정사각형 $\Rightarrow \frac{n!}{4}$

② 직사각형 $\Rightarrow \frac{n!}{2}$

③ 정삼각형 $\Rightarrow \frac{n!}{3}$

\Rightarrow 합동인 도형
의 수로
나눈다



3 중복순열과 같은 것을 포함하는 순열

1. 중복순열

복원 추출

(o) 서로 다른 n 개 중에서 중복을 허용하여 r 개를 뽑아서 순서있게 나열하는 것

\Rightarrow (기호) ${}_n\Pi_r$

☑ Π 는 Product(곱)의 첫 글자 P에 해당하는 그리스 문자로 '파이' 라고 읽는다.

(1) 서로 다른 n 개에서 r 개를 택하여 일렬로 배열하는 중복순열

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

(2) 중복순열의 공식 암기법

$\Rightarrow {}_n\Pi_r$ 에서 중복 가능한 것이 n

$\begin{cases} n \text{ 은 받는 쪽 (고정 숫자, 수동적)} \\ r \text{ 은 주는 쪽 (선택 숫자, 능동적)} \end{cases}$

☑ (종류) Π (개수), (고정) Π (이동)

(3) 중복순열인 경우(순서○, 중복○)

- ① 중복을 허용하는 자릿수 ② 모스 부호(·, -)
- ③ 기명투표 ④ 우체통
- ⑤ 반 편성 ⑥ 집합의 원소 배정
- ⑦ 함수의 개수

(4) 중복순열의 유형

- ① 중복을 허용하여 1, 2, 3 [종류]를 사용하여 만들 수 있는 다섯 자리 수 [개수] $\Rightarrow {}_3\Pi_5 = 3^5$
- ② 모스 부호 '·'과 '-' [종류]에서 8 개를 뽑아 만들 수 있는 방법의 수 [개수] $\Rightarrow {}_2\Pi_8 = 2^8$
- ③ 3 명의 후보자 [고정]에게 5 명의 유권자 [이동]가 기명 투표하는 방법의 수 $\Rightarrow {}_3\Pi_5 = 3^5$
- ④ 3 개의 우체통 [고정]에 5 통의 편지 [이동]를 넣는 방법의 수 $\Rightarrow {}_3\Pi_5 = 3^5$
- ⑤ 5 명의 학생 [이동]을 3 개 반 [고정]으로 편성하는 방법의 수 $\Rightarrow {}_3\Pi_5 = 3^5$
- ⑥ 5 개의 원소 [이동]를 $A \subset B \subset U$ [고정]에 배정한 두 집합 (A, B)의 순서쌍의 수 $\Rightarrow {}_3\Pi_5$
- ⑦ $n(X) = 5$ [이동], $n(Y) = 3$ [고정]일 때, 집합 X 에서 Y 로의 함수의 개수 $\Rightarrow {}_3\Pi_5 = 3^5$

2. 같은 것을 포함하는 순열

(1) 같은 것을 포함하는 순열

서로 같은 것이 각각 p 개, q 개, r 개 들어 있는 n 개를 모두 택하여 만든 순열

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

예) a, a, a, b, b, c, c 를 일렬로 배열하는 방법의 수

$$\begin{array}{|c|c|c|c|c|c|c|} \hline \square & a & a & \square & a & \square & \square \\ \hline \end{array} a : {}_7C_3 = \frac{7!}{3!4!}$$

$$\begin{array}{|c|c|c|c|c|c|c|} \hline \square & a & a & b & a & \square & b \\ \hline \end{array} b : {}_4C_2 = \frac{4!}{2!2!}$$

$$\begin{array}{|c|c|c|c|c|c|c|} \hline c & a & a & b & a & c & b \\ \hline \end{array} c : {}_2C_2 = \frac{2!}{2!}$$

$$\therefore {}_7C_3 \times {}_4C_2 \times {}_2C_2 = \frac{7!}{3!2!2!} = 210 \text{ (가지)}$$

(2) 순서가 정해진 경우의 순열

서로 다른 n 개 중 r 개의 순서가 일정할 때, 이들 n 개를 모두 일렬로 배열하는 순열의 수

$$\Rightarrow \frac{n!}{r!}$$

☑ 순서가 일정한 문자는 모두 '같은 문자' 로 생각

$\begin{cases} \text{일부만 순서가 일정} \Rightarrow \text{같은 문자로 치환} \\ \text{전체의 순서가 일정} \Rightarrow \text{조합} \\ \text{묶음으로 순서가 일정} \Rightarrow \text{분할} \end{cases}$

(3) 바둑판과 같은 도로망의 최단 거리

- ① 이전 지점까지 가는 방법의 수를 계속 더한다.
- ② 가로 한 구간 a , 세로 한 구간 b
- ③ P 를 지나지 않는 최단거리
 \Rightarrow (모든 경우의 수) - (P 를 지나는 경우)

(4) 오른쪽과 같은 모양의 도로

① A 에서 C 로 가는

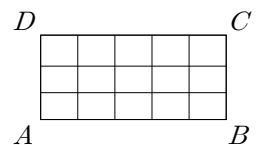
최단경로의 수

$$\Rightarrow \frac{8!}{5!3!} = 56$$

② A 에서 B 로 가는 경로의 수 (\rightarrow 중복순열)

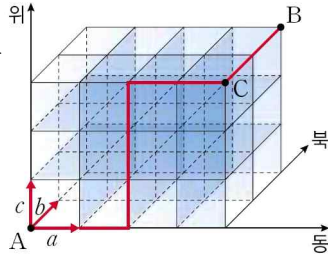
(단, 동·남·북쪽으로는 가능, 서쪽은 불가능)

$$\Rightarrow 4 \times 4 \times 4 \times 4 \times 4 = 4^5 = 2^{10}$$



- (5) 정육면체들의 모서리를 위
따라 A에서 B로 가는
최단경로의 수

$$\Rightarrow \frac{9!}{4!2!3!} = 1260$$



2. 조합

1 조합

1. 조합(Combination) $\nwarrow \rightarrow$ =(순서가 이미 결정)

(o) 서로 다른 n 개에서 순서에 관계없이 r 개를
선택하는 경우의 수 $\Rightarrow {}_n C_r$

☑ 순서를 생각하지 않고 그 일부를 뽑기만 한다.

(순서 \times) = (순서가 일정) = (크기가 고정)

(1) 서로 다른 n 개에서 r 개를 택하는 조합의 수

$$\begin{aligned} {}_n C_r &= \frac{{}_n P_r}{r!} = \frac{n(n-1) \cdots (n-r+1)}{r!} \\ &= \frac{n!}{r!(n-r)!} \quad (\text{단, } 0 \leq r \leq n) \end{aligned}$$

$$\square (\text{순열}) = (\text{조합}) \times (\text{나열}) \iff {}_n P_r = {}_n C_r \times r!$$

(2) 조합의 공식

$$\textcircled{1} 0! = 1, {}_n C_0 = 1, {}_n C_n = 1, {}_n C_1 = n$$

$$\textcircled{2} r \cdot {}_n C_r = n \cdot {}_{n-1} C_{r-1} \quad (1 \leq r \leq n)$$

☑ n 명의 학생 중에서 반장을 포함한 r 명의 임원을 뽑는
경우의 수는 다음과 같이 두 가지 방법으로 구할 수 있다.

i) n 명의 학생 중에서 r 명의 임원을 뽑는 경우의 수는

${}_n C_r$ 이고, 뽑힌 임원 r 명 중에서 반장 1 명을 뽑는
경우의 수는 r 이므로 구하는 경우의 수는 $r \cdot {}_n C_r$

ii) n 명의 학생 중에서 반장 1 명을 뽑는 경우의 수는 n
이고, 나머지 $(n-1)$ 명의 학생 중에서 $(r-1)$ 명의
임원을 뽑는 경우의 수는 ${}_{n-1} C_{r-1}$ 이므로 구하는 경우
의 수는 $n \cdot {}_{n-1} C_{r-1}$ 따라서 $r \cdot {}_n C_r = n \cdot {}_{n-1} C_{r-1}$

$$\textcircled{3} {}_n C_{r+1} = \frac{n-r}{r+1} {}_n C_r$$

2. 조건이 있는 조합의 수(순서 \times , 중복 \times)

(1) n (적어도 \sim) = n (전체) - n (반대인 경우)

(2) 특별한 것이 반드시 포함되어야 하는 경우

\Rightarrow 이미 뽑았다고 생각하고 나머지만을 계산

$$(3) n \text{ 개 팀} \rightarrow \text{경기수는} \begin{cases} \text{리그전} \Rightarrow {}_n C_2 \\ \text{토너먼트} \Rightarrow n-1 \end{cases}$$

3. 도형에 관한 문제

(1) 직선의 개수

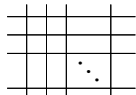
- ① 어느 세 점도 같은 직선 위에 있지 않은 서로 다른 n 개의 점 중에서 두 개의 점을 이어서 만든 직선의 개수
 $\Rightarrow {}_n C_2$
- ② 같은 직선 위에 있는 서로 다른 m 개의 점 중 두 개의 점을 이어서 생기는 직선의 개수
 $\Rightarrow 1$ (\leftarrow 점들이 있는 직선 자체)
- ③ n 개의 서로 다른 점 중에서 m 개의 점은 같은 직선 위에 있을 때 두 개의 점을 이어서 생기는 직선의 개수
 $\Rightarrow {}_n C_2 - {}_m C_2 + 1$
- ④ 볼록 n 각형의 대각선의 총수 $\Rightarrow {}_n C_2 - n$

(2) 삼각형의 개수

- ① 어느 세 점도 같은 직선 위에 있지 않은 서로 다른 n 개의 점 중에서 세 개의 점을 이어서 만든 삼각형의 개수
 $\Rightarrow {}_n C_3$
- ② 같은 직선 위에 있는 서로 다른 m 개의 점 중 세 개의 점을 이어서 생기는 삼각형의 개수
 $\Rightarrow 0$
- ③ n 개의 서로 다른 점 중에서 m 개의 점은 같은 직선 위에 있을 때 세 개의 점을 이어서 생기는 삼각형의 개수
 $\Rightarrow {}_n C_3 - {}_m C_3$

- (3) 원주 위의 n 개의 점에 대하여 원의 내부에서 현들이 만나서 생기는 교점의 개수
 $\Rightarrow {}_n C_4$ (\because 4 개를 뽑아 엇갈리게 연결)

- (4) 서로 다른 m 개의 평행선과 서로 다른 n 개의 평행선이 만날 때 평행사변형의 개수
 $\Rightarrow {}_m C_2 \times {}_n C_2$



(5) 수평선 m 개, 수직선 n 개에서 정사각형의 개수

- ① 한 변의 길이가 1 $\Rightarrow (m-1) \times (n-1)$
- ② 한 변의 길이가 2 $\Rightarrow (m-2) \times (n-2)$
- ③ 한 변의 길이가 3 $\Rightarrow (m-3) \times (n-3)$
- ④ 개수가 0 이 나오기 전까지 계속하여 더한다.

- (6) n 개의 직선에 의하여 평면을 분할할 때

$$\textcircled{1} \text{ 최대 개수} \Rightarrow 2 + 2 + 3 + 4 + \dots + n = 1 + \frac{n(n+1)}{2}$$

$$\textcircled{2} \text{ 최소 개수} \Rightarrow 2 + 1 + 1 + \dots + 1 = n + 1 \quad (\because \text{모두 평행선})$$

2 중복조합

1. 중복조합

- (o) 서로 다른 n 개에서 중복을 허용하여 r 개를 택하는 조합 \Rightarrow (기호) ${}_n H_r$

☐ H 는 Homogeneous product(동차의 곱)의 첫 글자이다.

(1) 중복조합의 예

- ① 두 숫자 1, 2 에서 세 개를 뽑는 중복조합의 수
 $(1, 1, 1) \rightarrow (+0, +1, +2) \rightarrow (1, 2, 3)$
 $(1, 1, 2) \rightarrow (+0, +1, +2) \rightarrow (1, 2, 4)$
 $(1, 2, 2) \rightarrow (+0, +1, +2) \rightarrow (1, 3, 4)$
 $(2, 2, 2) \rightarrow (+0, +1, +2) \rightarrow (2, 3, 4)$
 네 숫자 1, 2, 3, 4 에서 세 개를 뽑는 조합의 수
 $\therefore {}_2 H_3 = {}_4 C_3 = {}_{2+3-1} C_3 = 4$

- ② 두 숫자 1, 2 에서 네 개를 뽑는 중복조합의 수
 각각의 경우는 4 개의

● 사이에 한 개의 막대

■ 를 놓아 막대 앞에는 1 을,

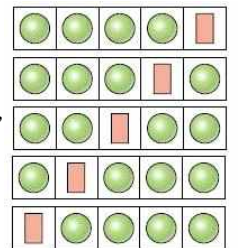
막대 뒤에는 2 를 놓아 대응

\therefore 구하는 경우의 수는

5 개의 자리에서 ● 를

놓을 자리 4 개를 선택하는 경우의 수

$${}_2 H_4 = {}_5 C_4 = 5$$



(2) 서로 다른 n 개에서 r 개를 택하는 중복조합

$\Leftrightarrow r$ 개의 ● 와 $(n-1)$ 개의 ■ 를

일렬로 배열하는 방법의 수

$\Leftrightarrow (n+r-1)$ 개의 자리에서 ● 를 놓을

r 개의 자리를 택하는 조합의 수

$${}_n H_r = {}_{n+r-1} C_r \quad (r > n \text{ 이어도 된다})$$

n 개의 수 $1, 2, 3, \dots, n$ 중에서 r 개의 수를 택하는 중복조합을 수의 크기의 순서로 정리하면

1, 1, 1, ..., 1
1, 1, 1, ..., 2
⋮
2, 2, 2, ..., 2
⋮
 $n, n, n, \dots, n-1$
 n, n, n, \dots, n

왼쪽의 각 조합의 첫 번째 수에 0, 두 번째 수에 1, 세 번째 수에 2, ..., r 번째 수에 $r-1$ 을 각각 더하면

1, 2, 3, ..., r
1, 2, 3, ..., $r+1$
⋮
2, 3, 4, ..., $r+1$
⋮
 $n, n+1, n+2, \dots, n+r-2$
 $n, n+1, n+2, \dots, n+r-1$

(3) 중복조합의 공식

$$\textcircled{1} \quad {}_{n+1}H_r + {}_nH_{r+1} = {}_{n+1}H_{r+1}$$

$$\textcircled{2} \quad {}_nH_0 + {}_nH_1 + {}_nH_2 + \dots + {}_nH_r = {}_{n+1}H_r$$

2. 중복조합인 경우 \Rightarrow 중복 가능한 것이 n

(1) 중복을 허용하여 선택하는 방법의 수

예 포도, 복숭아, 자두의 3 가지의 과일[종류]이 바구니 안에 들어 있을 때, 중복을 허용하여 5 개의 과일[개수]을 선택하는 방법의 수
(단, 각 과일은 5 개 이상 있다)

$$\Rightarrow {}_3H_5 = {}_7C_5 = {}_7C_2 = 21$$

(2) n 명의 후보자[고정]에게 r 명의 유권자[이동]가 무기명 투표하는 방법의 수 $\Rightarrow {}_nH_r$

(3) 똑같은 구슬 r 개[이동]를 A, B, \dots 의 n 명[고정]에게 나누어 주는 방법의 수 $\Rightarrow {}_nH_r$

(4) 똑같은 r 개의 물건[이동]을 서로 다른 n 개의 상자[고정]에 넣는 방법의 수 $\Rightarrow {}_nH_r$

(5) 부정방정식 $x_1 + x_2 + \dots + x_n = r$ 에서

① 음이 아닌 정수해 $\Rightarrow {}_nH_r$

② 양의 정수해 $\Rightarrow {}_nH_{r-n} \quad (r \geq n)$

(6) $(x+y+z)^n$ 을 전개할 때 생기는 서로 다른 종류의 항의 개수(전개항수) $\Rightarrow {}_3H_n$

3 분할

1. 집합의 분할

(1) 유한집합을 공집합이 아닌 서로소인

몇 개의 부분집합의 합집합으로 나타내는 것

(2) 원소의 개수가 n 인 집합을 k 개의 부분집합으로 분할
 \Rightarrow (기호) $S(n, k)$

☐ S 는 Stirling number(스털링 수)의 첫 글자이다.

(3) 똑같은 k 개의 상자에 서로 다른 n 개의 공을 넣을 때, 빈 상자가 없도록 넣는 방법의 수는 n 개의 공을 k 개의 상자에 분할하여 넣는 방법의 수와 같으므로 $S(n, k)$ 이다.

예 서로 다른 5개의 빵을 똑같은 3개의 접시에 나누어 담을 때, 빈 접시가 없도록 담은 모든 방법의 수는?

(3개, 1개, 1개)로 분할 : ${}_5C_3 \times {}_2C_1 \times {}_1C_1 \times \frac{1}{2!} = 10$

(2개, 2개, 1개)로 분할 : ${}_5C_2 \times {}_3C_2 \times {}_1C_1 \times \frac{1}{2!} = 15$

$$\therefore S(5, 3) = 10 + 15 = 25$$

2. 분할·분배의 방법의 수(조편성)

(1) 서로 다른 n 개의 물건을 p 개, q 개, r 개

($p+q+r=n$)의 세 묶음으로 나눈다(집합의 분할)

① p, q, r 이 모두 다른 수

$$\Rightarrow {}_nC_p \times {}_{n-p}C_q \times {}_rC_r$$

② p, q, r 중 어느 두 수가 같으면

$$\Rightarrow {}_nC_p \times {}_{n-p}C_q \times {}_rC_r \times \frac{1}{2!}$$

③ p, q, r 이 모두 같은 수이면

$$\Rightarrow {}_nC_p \times {}_{n-p}C_q \times {}_rC_r \times \frac{1}{3!}$$

예 토너먼트 경기의 대진표 작성 \Rightarrow 집합의 분할

(\because 나누기만 하면 자기들끼리 알아서 경기한다)

(2) 세 사람에게 p 개, q 개, r 개씩 나누어 준다(분배)

\Rightarrow (분할의 방법) \times (나누어 주는 대상의 수)!

$\hookrightarrow 3!$

예 (공역) = (치역)인 함수

⇒ 공역의 원소에 정의역의 원소를 적어도 한 개씩 분배

(3) 함수 $f: A \rightarrow A$ 에서 $f = f^{-1} \iff f \circ f = I$
를 만족하는 함수 f 의 개수

$$\textcircled{1} \quad n(A) = 1 \quad \Rightarrow 1$$

$$\textcircled{2} \quad n(A) = 2 \quad \Rightarrow 1 + {}_2C_2 = 2$$

$$\textcircled{3} \quad n(A) = 3 \quad \Rightarrow 1 + {}_3C_2 = 4$$

$$\textcircled{4} \quad n(A) = 4 \quad \Rightarrow 1 + {}_4C_2 + {}_4C_2 \cdot \frac{1}{2!} = 10$$

$$\textcircled{5} \quad n(A) = 5 \quad \Rightarrow 1 + {}_5C_2 + {}_5C_2 \cdot {}_3C_2 \cdot \frac{1}{2!} = 26$$

3. 자연수의 분할

(1) 자연수를 순서에 관계없이 몇 개의 자연수의 합으로 나타내는 것

(2) 자연수 n 을 k 개의 자연수로 분할

⇒ (기호) $P(n, k)$

☐ P 는 Partition(분할)의 첫 글자이다.

(3) 똑같은 k 개의 상자에 똑같은 n 개의 공을 넣을 때,
빈 상자가 없도록 넣는 방법의 수는 n 개의 공을
순서에 관계없이 k 개의 상자에 분할하여 넣는
방법의 수와 같으므로 $P(n, k)$ 이다.

예 똑같은 9 개의 구슬을 똑같은 4 개의 상자에 담을 때,
빈 상자가 없도록 담는 모든 방법의 수는?

⇒ 자연수 9 를 4 개의 자연수로 분할하는 것과 같다.

$$9 = 6 + 1 + 1 + 1 = 5 + 2 + 1 + 1 = 4 + 3 + 1 + 1$$

$$= 4 + 2 + 2 + 1 = 3 + 3 + 2 + 1 = 3 + 2 + 2 + 2$$

$$\therefore P(9, 4) = 6$$

5. 조합의 활용

(1) 1 부터 9 까지 자연수 중에서 세 수 a, b, c 와
자연수 n 에 대하여 등식 $a + b + c = n$ 을 만족
하는 순서쌍 (a, b, c) 의 개수를 a_n 이라 하면,

$$\Rightarrow n = 1 + 1 + 1 + \dots + 1$$

$(n - 1)$ 개의 $+$ 중에서 $+$ 를 2 개 선택하면 되므로

$$a_n = {}_{n-1}C_2$$

(2) 자연수 n 을 한 개 또는 두 개 이상의 자연수의 합
으로 나타내는 방법의 수를 a_n 이라 하면,

$$\Rightarrow n = 1 + 1 + 1 + \dots + 1$$

$(n - 1)$ 개의 $+$ 중에서 $+$ 를 0 개, 1 개, 2 개 ...,
 $(n - 1)$ 개를 선택(고정)하는 경우를 더하면 되므로

$$a_n = {}_{n-1}C_0 + {}_{n-1}C_1 + {}_{n-1}C_2 + \dots + {}_{n-1}C_{n-1} \\ = 2^{n-1}$$

4 함수에 관한 문제

$f: X \rightarrow Y, n(X) = r, n(Y) = n$ 일 때,

1. 함수의 개수 $\Rightarrow {}_n\Pi_r = n^r$

2. 일대일 함수 $\iff x_1 \neq x_2$ 이면 $f(x_1) \neq f(x_2)$

$$\Rightarrow {}_nP_r = \frac{n!}{(n-r)!}$$

일대일 대응 \iff 일대일 함수 & (공역) = (치역)

$$\Rightarrow {}_nP_n = n! \quad (\because n = r)$$

3. $x_1 < x_2 \Rightarrow f(x_1) < f(x_2) \Rightarrow {}_nC_r$
[\because 크기 고정]

4. $x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2) \Rightarrow {}_nH_r$
[\because 크기 고정 & 중복 허용]

5. $f(X) = Y$ 인 함수의 개수 \Rightarrow 서로 다른 r 개
의 물건을 n 명에게 적어도 한 개씩 나누어
주는 방법의 수 [분배]

5 공을 바구니에 넣는 방법의 수

다음 각각의 경우에 5 개의 공을 3 개의 바구니에 넣는
방법의 수는? (단, 비어 있는 바구니가 있을 수 있다.)

1. 5 개의 공에 1, 2, 3, 4, 5 가 적혀 있고,
3 개의 바구니에 각각 A, B, C 가 적혀 있다.

$$\Rightarrow \text{[중복순열]} : {}_3\Pi_5 = 3^5 = 243$$

2. 5 개의 공은 구분하지 않고,
3 개의 바구니에 각각 A, B, C 가 적혀 있다.

$$\Rightarrow [\text{중복조합}] \quad a + b + c = 5, \quad a \geq 0, \quad b \geq 0, \quad c \geq 0 \\ \therefore {}_3H_5 = {}_7C_5 = {}_7C_2 = 21$$

3. 5 개의 공에 1, 2, 3, 4, 5가 적혀 있고,
3 개의 바구니는 구분하지 않는다.

$$\Rightarrow [\text{집합의 분할}] : n(5, 0, 0) + n(4, 1, 0) + \\ n(3, 2, 0) + n(3, 1, 1) + n(2, 2, 1) \\ = 1 + 5 + 10 + 10 + 15 = 41$$

4. 5 개의 공은 구분하지 않고,
3 개의 바구니도 역시 구분하지 않는다.

$$\Rightarrow [\text{자연수의 분할}] : n\{(5, 0, 0), (4, 1, 0), \\ (3, 2, 0), (3, 1, 1), (2, 2, 1)\} = 5$$

3. 이항정리

① 파스칼의 삼각형

$$1. \quad {}_nC_r = {}_nC_{n-r} \quad (0 \leq r \leq n) \\ \Rightarrow {}_nC_r = {}_nC_s \iff r = s \quad \text{또는} \quad r + s = n$$

$$\square \quad n \text{ 개 중 } (r \text{ 개를 뽑는다}) = (n - r \text{ 개를 제낀다})$$

$$2. \quad {}_nC_r = {}_{n-1}C_{r-1} + {}_{n-1}C_r \quad (1 \leq r \leq n-1) \\ \Rightarrow {}_nC_r + {}_nC_{r+1} = {}_{n+1}C_{r+1}$$

$$\square \quad n \text{ 개 중 } (r \text{ 개를 뽑는다}) \\ = (\text{특정한 } A \text{ 를 포함 } \bigcirc) + (A \text{ 를 포함 } \times)$$

3. 파스칼의 삼각형(Pascal's triangle) : $(a + b)^n$ 에서

$$(a + b)^0 = 1$$

$$(a + b)^1 = 1a + 1b$$

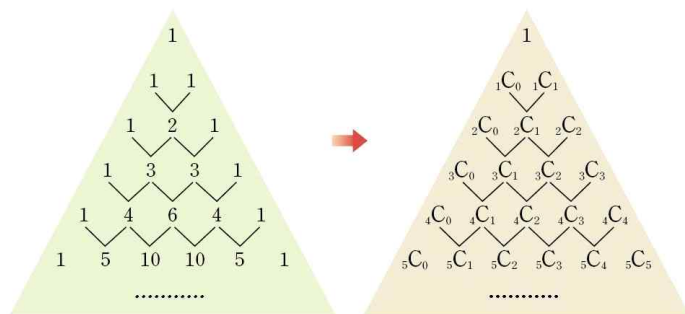
$$(a + b)^2 = 1a^2 + 2ab + 1b^2$$

$$(a + b)^3 = 1a^3 + 3a^2b + 3ab^2 + 1b^3$$

$$(a + b)^4 = 1a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1b^4$$

$$(a + b)^5 = 1a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + 1b^5$$

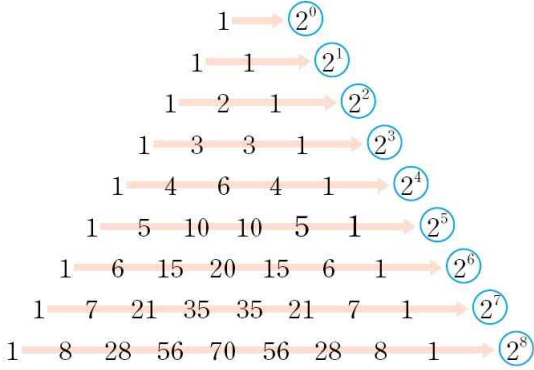
$$\square \quad {}_nC_r + {}_nC_{r+1} = {}_{n+1}C_{r+1} \quad \text{를 이용}$$



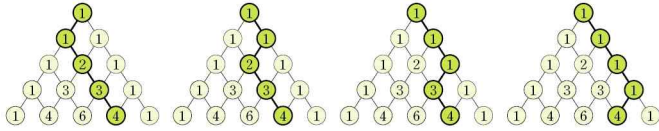
- (1) 각 행에서 양 끝의 수는 1이다.
- (2) 각 행의 수는 중앙에 대하여 좌우 대칭이다.
- (3) 각 수는 그 수의 왼쪽 위와 오른쪽 위에 있는 수의 합과 같다.

4. 파스칼의 삼각형에서 나타나는 성질

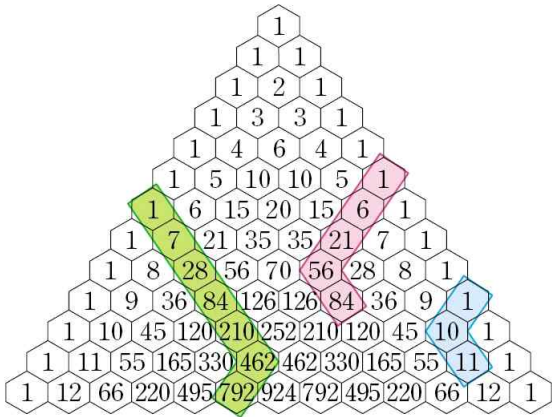
(1) 행의 수의 합 : n 번째 행의 수의 합은 항상 2^{n-1}



(2) 첫 번째 행의 1에서 시작하여 연결된 선을 따라 어떤 숫자 n 에 도달하는 최단 거리의 수는 바로 n 가지 [골턴(Galton) 관]



(3) 하키 스틱 패턴 : 바깥쪽의 1에서 시작하여 대각선 방향으로 수들을 더하면 아래 행의 안쪽 하키 스틱 모양에 있는 수가 된다.



$$(6) {}_nC_0 + {}_nC_1 a + {}_nC_2 a^2 + \cdots + {}_nC_n a^n \\ = (1+a)^n$$

$$(7) \textcircled{1} {}_nC_0 - {}_nC_2 + {}_nC_4 - {}_nC_6 + \cdots \\ = (\sqrt{2})^n \cos \frac{n\pi}{4}$$

☑ $(1+i)^n$ 의 “실수부”

$$\textcircled{2} {}_nC_1 - {}_nC_3 + {}_nC_5 - {}_nC_7 + \cdots \\ = (\sqrt{2})^n \sin \frac{n\pi}{4}$$

☑ $(1+i)^n$ 의 “허수부”

3. 이항정리의 활용

$$(1) (a+b)^n = \sum_{r=0}^n {}_nC_r a^{n-r} b^r \\ \hookrightarrow \frac{n!}{r!(n-r)!}$$

(2) $(a+b+c)^n$ 의 전개식에서 $a^p b^q c^r$ 의 계수

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

$$(3) (a+b+c)^n = \sum \frac{n!}{p!q!r!} a^p b^q c^r \\ \hookrightarrow {}_nC_p \times {}_{n-p}C_q$$

3 이항정리의 활용

$$1. ({}_nC_0)^2 + ({}_nC_1)^2 + ({}_nC_2)^2 + \cdots + ({}_nC_n)^2 \\ = \sum_{r=0}^n ({}_nC_r)^2 = {}_{2n}C_n = \frac{(2n)!}{n!n!}$$

☑ $(1+x)^n (1+x)^n = (1+x)^{2n}$ 에서 x^n 의 계수

$$2. ({}_{2n}C_0)^2 - ({}_{2n}C_1)^2 + ({}_{2n}C_2)^2 - \cdots + ({}_{2n}C_{2n})^2 \\ = \sum_{r=0}^{2n} (-1)^r ({}_{2n}C_r)^2 = (-1)^n {}_{2n}C_n$$

☑ $(1-x)^n (1+x)^n = (1-x^2)^n$ 에서 x^{2n} 의 계수

$$3. {}_mC_0 \times {}_nC_k + {}_mC_1 \times {}_nC_{k-1} + \cdots + {}_mC_k \times {}_nC_0 \\ = \sum_{r=0}^k {}_mC_r \times {}_nC_{k-r} = {}_{m+n}C_k$$

☑ $(1+x)^m (1+x)^n = (1+x)^{m+n}$ 에서 x^k 의 계수

$$4. {}_nC_1 + 2{}_nC_2 + 3{}_nC_3 + \cdots + r{}_nC_r + \cdots + n{}_nC_n \\ = \sum_{r=0}^n r \cdot {}_nC_r = n \cdot 2^{n-1}$$

☑ $r \cdot {}_nC_r = n \cdot {}_{n-1}C_{r-1}$ 이용 or $(1+x)^n$ 을 미분

$$5. {}_nC_0 + 2{}_nC_1 + 3{}_nC_2 + \cdots + (n+1){}_nC_n \\ = \sum_{r=0}^n (r+1) \cdot {}_nC_r = (n+1) \cdot 2^{n-1}$$

☑ 두 번 더하여 $2S = (n+1) \cdot 2^n$

$$6. {}_nC_0 + \frac{{}_nC_1}{2} + \frac{{}_nC_2}{3} + \cdots + \frac{{}_nC_r}{r+1} + \cdots + \frac{{}_nC_n}{n+1} \\ = \sum_{r=0}^n \frac{{}_nC_r}{r+1} = \frac{2^{n+1}-1}{n+1}$$

☑ $(r+1) \cdot {}_{n+1}C_{r+1} = (n+1) \cdot {}_nC_r$ 이용 or 적분

$$7. {}_kC_k + {}_{k+1}C_k + {}_{k+2}C_k + \cdots + {}_nC_k \\ = {}_{n+1}C_{k+1}$$

$$\textcircled{1} (1+x)^k + (1+x)^{k+1} + \cdots + (1+x)^n \\ = \frac{(1+x)^{n+1} - (1+x)^k}{x} \text{ 에서 } x^k \text{ 의 계수} \\ \iff (1+x)^{n+1} \text{ 에서 } x^{k+1} \text{ 의 계수}$$

$$8. 1 \cdot ({}_nC_1)^2 + 2 \cdot ({}_nC_2)^2 + \cdots + n \cdot ({}_nC_n)^2 \\ = \sum_{k=1}^n k \cdot ({}_nC_k)^2 = n \times {}_{2n-1}C_{n-1} \\ = \frac{n}{2} \times {}_{2n}C_n$$

☑ $(1+x)^{n-1} (1+x)^n = (1+x)^{2n-1}$ 에서 x^{n-1} 의 계수

$$\sum_{k=1}^n ({}_{n-1}C_{k-1} \times {}_nC_{n-k}) = {}_{2n-1}C_{n-1}$$

한편, $k \cdot {}_nC_k = n \cdot {}_{n-1}C_{k-1}$ 이므로

$$\begin{aligned} \sum_{k=1}^n k \cdot ({}_nC_k)^2 &= \sum_{k=1}^n n ({}_{n-1}C_{k-1} \times {}_nC_k) \\ &= n \times {}_{2n-1}C_{n-1} = \frac{n}{2} \times {}_{2n}C_n \end{aligned}$$

$$\begin{aligned} 9. \quad &1^2 \cdot {}_nC_1 + 2^2 \cdot {}_nC_2 + 3^2 \cdot {}_nC_3 + \cdots + n^2 \cdot {}_nC_n \\ &= n(n+1) \cdot 2^{n-2} \end{aligned}$$

☐ $(1+x)^n$ 을 미분 $\Rightarrow (\times x)$
 \Rightarrow 다시 미분한 후 $x=1$ 을 대입

$$10. \quad \sum_{k=2}^{98} {}_{k-1}C_1 \times {}_{100-k}C_2 = {}_{100}C_4$$

☐ 1 부터 100 까지의 자연수 중에서 서로 다른 4 개의 수를 선택할 때, 4 개의 수 중에서 두 번째로 작은 수가 k 인 경우의 수를 a_k 라 하면,
 $\Rightarrow a_k = {}_{k-1}C_1 \times {}_{100-k}C_2$

$\Rightarrow A$ 가 1 점 획득 $\rightarrow a$, B 가 1 점 획득 $\rightarrow b$

$$n(U) = 2^4 = 16$$

$$1. A \text{ 가 승} : {}_4C_4 + {}_4C_3 + {}_4C_2 = 1 + 4 + 6 = 11$$

$aaaa, aaab, aaba, aabb, abaa$

$abab, abba, baaa, baab, baba$

$bbaa$

$$2. B \text{ 가 승} : {}_4C_1 + {}_4C_0 = 4 + 1 = 5$$

$abbbb, babb, bbab, bbbb, bbbb$

\therefore 상금은 11 : 5 로 배분하면 된다.

< 갈릴레이의 문제 >

3 개의 주사위를 던졌을 때 눈의 수의 합이 9 가 되는 경우와 10 이 되는 경우는 모두 6 가지로 같은데, 실제로는 10 이 되는 경우가 더 많이 나온다. 그 이유는 무엇인가?

1. 합이 9 가 되는 경우 :

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

$(\underline{2, 5, 5}), (\underline{2, 3, 4}), (\underline{3, 3, 3})$

합이 10 이 되는 경우 :

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

$(\underline{2, 3, 5}), (\underline{2, 4, 4}), (\underline{3, 3, 4})$

이므로 경우의 수가 같다고 생각하면 안 된다.

2. 각각의 경우마다 순열의 수를 생각하면 합이 9 가 되는 경우는 25 가지, 합이 10 이 되는 경우는 27 가지이다.

< 페르마 문제 >

두 사람 A, B 가 동전을 던져 앞면이 나오면 A 가 1 점, 뒷면이 나오면 B 가 1 점을 얻는다고 한다. 동전을 번갈아 던져 10 점을 먼저 얻는 사람이 상금을 독차지하기로 하였다. 그런데 급한 사정이 생겨 A 는 8 점, B 는 7 점을 얻은 상황에서 동전 던지기를 중단하게 되었다.

상금을 어떻게 배분해야 할까?

< 순열과 조합의 역사 >

순열과 조합의 개념을 누가 처음으로 생각했는지는 분명하지 않다. 그러나 역사적으로는 순열의 개념 보다 조합의 개념이 먼저 발달하였다.

조합의 개념은 3~5 세기 무렵 고대 그리스에서 쓰여진 “Sefer Yetzirah(창조의 책)”에 처음으로 기록되어 있다고 한다. 이 책에는 22 개의 글자 중에서 2 개를 순서를 고려하지 않고 뽑는 방법(조합)의 수가

$$\frac{n(n-1)}{2} = \frac{22 \times 21}{2} = 231$$

이라고 적혀 있다.

그 후 순열과 조합의 개념은 인도의 수학자들에 의하여 발달되었는데, 9 세기 무렵 마하비라(Mahavira ; 800~870)가 순열과 조합의 개념을 체계화 하였고, 12 세기 무렵 바스카라(Bhaskara, A. ; 1114~1185)는 그의 책 “릴라바티”에서 순열과 조합에 관한 문제를 다루었으며 ${}_nC_r$ 과 ${}_nP_r$ 에 관한 중요한 정리를 제시하였다고 한다.

< 순열과 조합 문제의 진술 유형 >

순열과 조합 문제의 진술은 학생들의 문제 해결에 크게 영향을 끼치는 주요한 변인이다. 순열과 조합 문제는 그 진술 방법에 따라 크게 선택 문제, 배열 문제, 분할 문제의 세 가지 유형으로 분류될 수 있다.

(1) 선택 문제

‘꺼내다.’, ‘선택하다.’, ‘뽑다.’ 등의 표현이 있는 문제는 선택 문제라고 할 수 있다. 예를 들어 ‘주머니에 4 개의 숫자 1, 2, 3, 4 가 각각 적힌 공이 있다. 이 주머니에서 3 개의 공을 꺼내는 방법의 수는 모두 몇 가지인가?’ 라는 문제는 선택 문제이다.

일반적으로 ‘서로 다른 n 개의 사물 중에서 서로 다른 r 개를 뽑는 방법의 수를 구하여라.’ 라는 조합 문제와 ‘서로 다른 n 개의 사물 중에서 중복을 허락하여 r 개를 뽑는 방법의 수를 구하여라.’라는 중복조합 문제는 선택 문제이다.

(2) 배열 문제

‘배열하다.’, ‘배치하다.’, ‘나열하다.’, ‘분배하다.’, ‘놓다.’ 등의 표현이 있는 문제는 배열 문제라고 할 수 있다.

예를 들어 ‘4 개의 숫자 1, 2, 3, 4 를 배열하여 네 자리 자연수를 만드는 방법의 수는 모두 몇 가지인가?’ 라는 문제는 배열 문제이다.

일반적으로 ‘서로 다른 n 개의 사물 중에서 r 개를 뽑아 배열하는 방법의 수를 구하여라.’라는 순열 문제와 ‘서로 다른 n 개의 사물 중에서 중복을 허락하여 r 개를 뽑아 배열하는 방법의 수를 구하여라.’라는 중복순열 문제는 배열 문제이다.

그러나 ‘똑같은 r 개의 사물을 서로 다른 n 개의 상자에 넣는 방법의 수를 구하여라.’라는 배열 문제는 ‘서로 다른 n 개의 사물 중에서 중복을 허락하여 r 개를 뽑는 방법의 수를 구하여라.’라는 선택 문제로 바꾸어 진술할 수 있다.

(3) 분할 문제

‘나누다.’, ‘분할하다.’ 등의 표현이 있는 문제는 분할 문제라고 할 수 있다. 예를 들어 ‘7 명이 2 대의 승용차에 나누어 타는 방법은 모두 몇 가지인가?’라는 문제는 분할 문제이다.

일반적으로 ‘서로 다른 n 개의 사물을 서로 다른 r 개의 그룹으로 나누는 경우의 수를 구하여라.’라는 분할 문제는 서로 다른 n 개의 사물을 서로 다른 r 개의 상자에 넣는 배열 문제에서 상자를 구별하지 않고, 사물이 나누어진 결과만 보는 것과 같다.

따라서 분할 문제와 배열 문제 사이에는 일대일 대응이 존재한다.

선택 문제, 배열 문제, 분할 문제는 모두 공·상자 문제로 표현될 수 있다. 공·상자 문제는 n 개의 상자에 r 개의 공을 넣는 경우의 수를 구하는 문제로서 상자와 공의 구별 여부, 빈 상자의 여부 등에 따라 선택 문제, 배열 문제, 분할 문제가 된다.

일반적으로 선택 문제, 배열 문제는 다음과 같이 공·상자 문제와 동형이다.

	선택 문제, 배열 문제	공·상자 문제
순열 ${}_nP_r$	서로 다른 n 개에서 r 개 ($r \leq n$) 을 택하여 일렬로 배열하는 방법의 수	서로 다른 n 개의 상자에서 서로 다른 r 개의 공을 많아야 한 개씩 넣는 방법의 수
중복 순열 ${}_n\Pi_r$	서로 다른 n 개에서 중복을 허락하여 r 개를 택하여 일렬로 배열하는 방법의 수	서로 다른 n 개의 상자에서 서로 다른 r 개의 공을 넣는 방법의 수
조합 ${}_nC_r$	서로 다른 n 개에서 r 개 ($r \leq n$) 를 택하는 방법의 수	서로 다른 n 개의 상자에 똑같은 r 개의 공을 많아야 한 개씩 넣는 방법의 수

중복 조합 ${}_n H_r$	서로 다른 n 개에서 중복을 허락하여 r 개를 택하는 방법의 수	서로 다른 n 개의 상자에 똑같은 r 개의 공을 넣는 방법의 수
------------------	---	---

< 집합의 분할 >

원소의 개수가 n 인 집합 A 에 대하여 그 부분집합 $A_i \subset A$ ($i = 1, 2, \dots, k$)가

- i) $A_i \neq \phi$ ($i = 1, 2, \dots, k$)
- ii) $i \neq j$ 이면 $A_i \cap A_j = \phi$

iii) $\bigcup_{i=1}^k A_i = A$

를 만족할 때, $\{A_1, A_2, \dots, A_k\}$ 를 집합 A 의 분할이라 하고, 이 분할의 수를 기호로

$$S(n, k)$$

와 같이 나타낸다. 이 수를 이계 스털링 수(Stirling numbers of the second kind)라고도 한다.

$n = 1, 2, 3, \dots, k = 1, 2, 3, \dots$ 일 때 $S(n, k)$ 의 값을 구하면 다음 표와 같은데, 이를 삼각형 모양으로 배열한 것을 스털링 수의 삼각형이라고 한다.

$k \backslash n$	1	2	3	4	...
1	1				
2	1	1			
3	1	3	1		
4	1	7	6	1	
5	1	15	25	10	1
6	1	31	90	65	15
⋮	⋮	⋮	⋮	⋮	⋮

위의 표에서 $S(n, 1) = 1, S(n, n) = 1$ 임을 알 수 있다. 또 $S(n, 0) = 0$ 으로 정한다.

[정리] $1 \leq k \leq n$ 인 정수 n, k 에 대하여 다음이 성립한다.

$$S(n, k) = S(n-1, k-1) + S(n-1, k) \times k \quad \cdots \textcircled{1}$$

[증명] n 개의 원소를 k 개의 부분집합으로 나누는 경우는 다음 두 가지가 있다.

- i) 원소 하나가 한 개의 부분집합을 이루는 경우 ;
나머지 $(n-1)$ 개의 원소를 $(k-1)$ 개의 부분집합으로 분할하면 되므로 그 경우의 수는 $S(n-1, k-1)$ 이다.
- ii) 원소 하나가 다른 원소와 함께 한 개의 부분집합을 이루는 경우 ;

나머지 $(n-1)$ 개의 원소를 k 개의 부분집합으로 분할한 다음 그 원소 하나를 k 개의 부분집합 중 어느 한 부분집합에 포함시키면 되므로 그 경우의 수는

$$S(n-1, k) \times k \text{ 이다.}$$

따라서 i), ii)에 의하여 ①이 성립한다.

< 자연수의 분할 >

자연수 n 을 k 개의 자연수의 합, 즉

$$n = n_1 + n_2 + \dots + n_k$$

($n_1 \geq n_2 \geq \dots \geq n_k \geq 1, n_1, n_2, \dots, n_k$ 는 자연수)

와 같이 나타내는 것을 자연수의 분할이라 하고, 이 분할의 수를 기호로

$$P(n, k)$$

와 같이 나타낸다. $P(n, k)$ 는 똑같은 n 개의 공을 똑같은 k 개의 상자에 빈 상자가 없도록 넣는 방법의 수와 같다.

따라서 $P(n, 1) = 1, P(n, n) = 1$ 이다.

[정리] $1 \leq k \leq n$ 인 정수 n, k 에 대하여 다음이 성립한다.

$$P(n, k) = P(n-k, 1) + P(n-k, 2) + \dots + P(n-k, k) \quad \cdots \textcircled{2}$$

[증명 1] 똑같은 n 개의 공을 똑같은 k 개의 상자에 빈 상자가 없도록 넣는 방법의 수는 어떤 상자에 공을 한 개만 넣는 경우와 모든 상자에 2 개 이상의 공을 넣는 경우로 나누어 생각할 수 있다.

- i) 어떤 상자에 공을 한 개만 넣는 경우 ;
한 개의 공만 들어 있는 상자 한 개를 제외한 나머지 $(k-1)$ 개의 상자에 $(n-1)$ 개의 공을 빈 상자가 없도록 넣어야 하므로 그 경우의 수는 $P(n-1, k-1)$ 이다.
- ii) 모든 상자에 2 개 이상의 공을 넣는 경우 ;
각 상자에 한 개의 공을 넣고 나머지 $(n-k)$ 의 공을 k 개의 상자에 빈 상자가 없도록 넣어야 하므로 그 경우의 수는 $P(n-k, k)$ 이다.

즉, i), ii)에 의하여

$$P(n, k) = P(n-1, k-1) + P(n-k, k) \quad \cdots \textcircled{3}$$

한편 $P(n-1, k-1)$ 도 위와 같은 방법으로 하면

$$P(n-1, k-1) = P(n-2, k-2) + P(n-k, k-1) \quad \cdots \textcircled{4}$$

을 만족한다.

④를 ③에 대입하면

$$\begin{aligned}
 P(n-1, k-1) \\
 &= P(n-2, k-2) + P(n-k, k-1) \\
 &\quad + P(n-k, k) \quad \dots \textcircled{5}
 \end{aligned}$$

가 된다.

따라서 같은 방법으로 계속하면 ②가 성립한다.

[증명 2] 자연수 n 을 k 개의 자연수로 분할하는 방법은 1이 포함되는 경우와 포함되지 않는 경우로 나누어 생각할 수 있다.

i) 1이 포함되는 경우 ;

$n = 1 + n_1 + n_2 + \dots + n_{k-1}$ (n_1, n_2, \dots, n_{k-1} 은 1 이상의 자연수)에서 $n-1 = n_1 + n_2 + \dots + n_{k-1}$ 이므로 자연수 $(n-1)$ 을 $(k-1)$ 개의 자연수로 분할하는 방법과 같으므로 그 방법의 수는 $P(n-1, k-1)$ 이다.

ii) 1이 포함되지 않는 경우 ;

$n = n_1 + n_2 + \dots + n_k$ (n_1, n_2, \dots, n_k 는 2 이상의 자연수)에서 $n_i = n'_i + 1$ ($i = 1, 2, \dots, k$)이라고 하면 $n-k = n'_1 + n'_2 + \dots + n'_k$ (n'_i 는 1 이상의 자연수)이므로 자연수 $(n-k)$ 를 k 개의 자연수로 분할하는 방법과 같으므로 그 방법의 수는 $P(n-k, k)$ 이다.

즉, i), ii)에 의하여

$$P(n, k) = P(n-1, k-1) + P(n-k, k) \quad \dots \textcircled{3}$$

위와 같은 방법으로 하면 [증명 1]의 ④, ⑤가 되므로 ②가 성립한다.

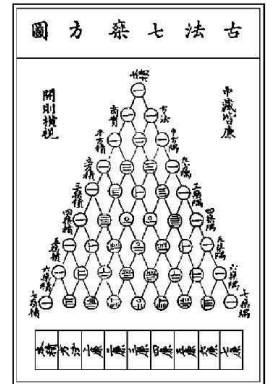
집합과 자연수의 분할의 수를 공·상자 문제로 나타내면 다음 표와 같다.

	분할의 정의	공·상자 문제
집합의 분할 $S(n, k)$	원소의 개수가 n 인 집합을 공집합이 아닌 k 개의 서로소인 집합으로 나누는 방법의 수	서로 다른 n 개의 공을 똑같은 k 개의 상자에 빈 상자가 없도록 넣는 방법의 수
자연수의 분할 $P(n, k)$	자연수 n 을 순서에 관계없이 k 개의 자연수의 합으로 나타내는 방법의 수	똑같은 n 개의 공을 똑같은 k 개의 상자에 빈 상자가 없도록 넣는 방법의 수

< 파스칼의 삼각형 >

오른쪽 그림은 11세기 중국의 수학자 진산의 업적으로 그의 저서 <진경>에 나오며, 주세걸의 <사원옥감(1303)>에 나온다.

이러한 배열은 1527년 아피아누스(Apianus, P.)와 1544년 슈티펠(Stifel, M. : 1487~1567)에 의해 독립적으로 발견되었다. 그러나 서양에서는 프랑스의 수학자이자 철학자인 파스칼(Pascal, B. : 1623~1662)의 이름을 따서 '파스칼의 삼각형'이라고 부른다. 파스칼은 이것을 '산술적 삼각형(arithmetical triangle)'이라고 불렀다.



순열의 개념은 서기 200년에서 600년 사이에 쓰여진 헤브류의 <창조의 책>에서 처음으로 나타난다. 그 후 17세기에 와서 베르누이는 <추측의 기술>이란 책에서 순열과 조합에 관한 문제를 체계적으로 다루었고 확률에 대한 개념을 처음 소개하기도 했다.

다항식 $(x+y)^n$ 의 전개식에 관한 이항정리도 기원전 300년경에 유클리드가 간단한 문제에서 다루고 있다. 그러나 현재와 같은 형태의 이항계수의 개념은 16세기에 스티펠에 의해 처음 소개되었고, 이후 17세기에 파스칼에 의해 연구되었다.

프랑스의 수학천재 파스칼을 1623년에 태어났다. 그는 허약한 체질 때문에 어릴 때 집에만 갇혀 있었다. 당연히 수학공부도 해서는 안되었는데, 어느 날 기하학 공부를 하고 있는 파스칼을 우연히 발견한 아버지는 소년의 능력에 놀라 결국 그에게 수학공부를 허락했다는 얘기가 전해진다.

파스칼의 <수삼각형론>은 1653년에 쓰여졌는데, 오른쪽 그림과 같은 수삼각형이 나온다. 이 수삼각형에서 1행과 1열은 모두 1이고, 1을 제외한 각 수들은 자신의 바로 위와 바로 오른쪽 수를 더하여 얻어진 것이다.

1	1	1	1	1	1	...
1	2	3	4	5	6	...
1	3	6	10	15	21	...
1	4	10	20	35	56	...
1	5	15	35	70	126	...
1	6	21	56	126	252	...
.

< 라이프니츠의 조화삼각형 >

파스칼의 삼각형에서는 위 행의 두 항의 합이 아래 행의 항이 된다. 이와 달리 아래 그림과 같이 아래 행의 두 항의 합이 위 행의 항이 되도록 수를 배열한 것을 조화삼각형이라고 한다. 예를 들어 세 번째 행의 $\frac{1}{12}$ 은 네 번째 행의 $\frac{1}{20}$ 과 $\frac{1}{30}$ 의 합이다.

$$\begin{array}{cccccccc}
 & & & & 1 & & & \\
 & & & \frac{1}{2} & & \frac{1}{2} & & \\
 & & \frac{1}{3} & & \frac{1}{6} & & \frac{1}{3} & \\
 & \frac{1}{4} & & \frac{1}{12} & & \frac{1}{12} & & \frac{1}{4} \\
 & \frac{1}{5} & & \frac{1}{20} & & \frac{1}{30} & & \frac{1}{20} & & \frac{1}{5} \\
 & \frac{1}{6} & & \frac{1}{30} & & \frac{1}{60} & & \frac{1}{60} & & \frac{1}{30} & & \frac{1}{6} \\
 & \frac{1}{7} & & \frac{1}{42} & & \frac{1}{105} & & \frac{1}{140} & & \frac{1}{105} & & \frac{1}{42} & & \frac{1}{7} \\
 & \frac{1}{8} & & \frac{1}{56} & & \frac{1}{168} & & \frac{1}{280} & & \frac{1}{280} & & \frac{1}{168} & & \frac{1}{56} & & \frac{1}{8} \\
 & \vdots & & & & & & & & & & & & & & \vdots
 \end{array}$$

조화삼각형은 삼각수(1, 3, 6, 10, 15, ...)의 역수의 합을 구하는 문제를 해결하는 과정에서 라이프니츠가 고안한 것이다.

조화삼각형에서 대각선 방향의 수를 모두 더한 무한급수의 합은 바로 위 행의 첫 번째 수가 된다.

< MBTI검사와 16가지 유형 >

MBTI(Myers-Briggs Type Indicator)는 성격 유형 선호 지표로서 자신과 타인의 성격을 이해하는 데 아주 유용하게 쓰이는 도구이다. MBTI는 다음 표와 같이 4가지 선호 지표에 의해 각각을 두 가지로 구분하여 알파벳 네 글자로 코드화된 16가지의 유형으로 성격을 분류한다.

선호지표	선호경향
외향(E) - 내향(I)	에너지의 방향은 어느 쪽인가?
감각(S) - 직관(N)	무엇을 인식하는가?
사고(T) - 감정(F)	어떻게 결정하는가?
판단(J) - 인식(P)	채택하는 생활양식은 무엇인가?

ISTJ 세상의 소금형	ISFJ 임금뒀 편의 권력형	INFJ 예언자형	INTJ 과학자형
ISTP 백과사전형	ISFP 성인군자형	INFP 잔다르크형	INTP 아이디어 뱅크형
ESTP 수완좋은 활동가형	ESFP 사교적인 유형	ENFP 스파크형	ENTP 발명가형
ESTJ 사업가형	ESFJ 친선도모형	ENFJ 연변능숙형	ENTJ 지도자형

< 바코드와 ISBN >

1. 바코드(Bar code) : 세계상품코드(UPC : Universal Product Code)와 한국상품코드(KAN : Korean Artical Number) 등이 있다.



- (1) 바코드에 담긴 정보 : 위의 그림에서 KAN을 보면 양 끝과 중앙의 줄무늬는 다른 것에 비해 긴 것을 알 수 있는데, 양 끝은 시작과 끝을 나타내고, 중앙의 것은 상품 정보가 들어 있는 영역을 나타내고 있다. 13자리 숫자 중 3자리는 국가 식별 코드, 4자리는 제조 업체 코드, 5자리는 상품 코드, 1자리는 검사 숫자로 구성되어 있으며, 한 나라에서는 $10^4 \times 10^5 = 10^9$ (10 억)가지의 바코드 종류를 가질 수 있다.

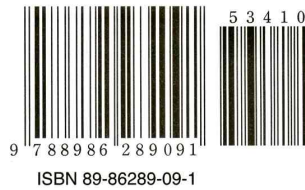
- (2) 바코드 판독과 검사수 : 바코드 판독기는 30개의 줄무늬와 함께 13개의 숫자를 읽어서 옳게 읽었는지를 판단한다. 13개의 숫자 중 맨 끝의 검사 숫자는 바코드 배열의 오류를 검사하는 숫자로,

앞의 12 자리 숫자를 일정한 방식으로 계산하여 나온 숫자이다. 이 숫자가 계산의 결과와 일치해야 배열이 올바른 것이다.

- (3) 바코드와 가격 : 실제로 바코드는 물건에 대한 직접적인 가격을 나타내는 것이 아니라 이미 등록된 바코드와 그 물건의 가격이 입력되어 있는 상태에서 통신회로를 통하여 특정한 바코드에 정해진 가격을 얻어내는 것이다.

2. 국제표준책번호(ISBN ; International Standard Book Number)

- (1) ISBN의 의미 : 국가 - 출판사 - 책번호 - 검사수



- (2) 국가 : 89는 대한민국, 독일권은 3, 영어권은 0, 프랑스는 2, 중국은 7, 일본은 4

- (3) 검사수 : 0에서 9까지의 수가 오게 되며, 10을 써야 할 때는 로마 숫자 X를 사용한다.

- (4) ISBN의 검사 방법 : 첫 번째 숫자는 10을 곱하고 두 번째 숫자는 9, 세 번째 숫자는 8, ... 이런 식으로 끝에서 두 번째 숫자에 2를 곱할 때까지 계산한다. 각 곱셈 결과를 모두 더해 마지막 자리에 있는 번호에 더한다. 최종 결과는 반드시 11로 나누어지는 수이어야 한다. 그렇지 않다면 어디선가 오류가 발생한 것이다.

< 60갑자 >

- 60 갑자(六十甲子) : 10 간(十干)과 12 지(十二支)를 조합한 것으로 60 간지(六十干支)라고도 한다.
- 10 간 : 갑(甲), 을(乙), 병(丙), 정(丁), 무(戊), 기(己), 경(庚), 신(辛), 임(壬), 계(癸)
- 12 지 : 자(子;쥐), 축(丑;소), 인(寅;호랑이), 묘(卯;토끼), 진(辰;용), 사(巳;뱀), 오(午;말), 미(未;양), 신(申;원숭이), 유(酉;닭), 술(戌;개), 해(亥;돼지)

4. 회갑·환갑 : 10 간과 12 지의 최소공배수가 60 이므로 사람이 태어난 해의 육십갑자가 60 세 되는 해 다시 돌아오게 되어 60 세 생일을 갑자가 돌아온다는 뜻의 회갑(回甲) 또는 환갑(還甲)이라 부른다.

< 윤년 >

- 지구가 태양을 한 번 공전하는 데 걸리는 시간은 정확하게 365 일 5 시간 48 분 46 초이므로 1년을 365 일로 하고, 나머지 5 시간 48 분 46 초를 4년 동안 모으면 23 시간 15 분 정도가 되어 4년마다 2월에 하루(윤일)을 늘리게 되는 것이다.

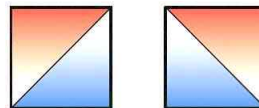
2. 그레고리력의 윤년 규칙

- 서력 기원 연수가 4로 나누어 떨어지는 해 ⇨ 윤년
- (1)에서 100으로 나누어 떨어지는 해 ⇨ 평년
- (1)에서 400으로 나누어 떨어지는 해 ⇨ 윤년

< 카탈란 수(catalan Number) >

- $(n+2)$ 개의 변을 가진 볼록 다각형을 서로 만나지 않는 대각선을 이용하여 n 개의 삼각형으로 나누는 방법의 수 ⇨ (기호) c_n

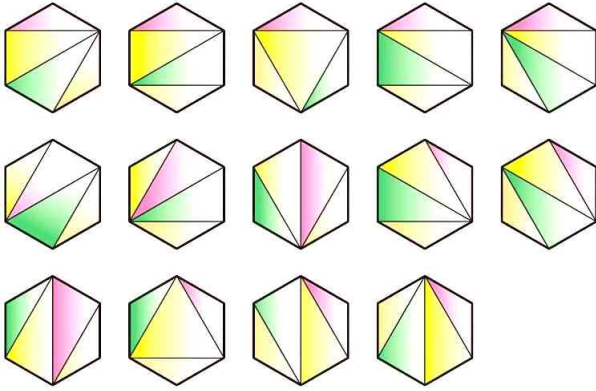
- (1) 사각형을 2 개의 삼각형으로 나누는 방법 : $c_2 = 2$



- (2) 오각형을 3 개의 삼각형으로 나누는 방법 : $c_3 = 5$



(3) 육각형을 4개의 삼각형으로 나누는 방법 : $c_4 = 14$



(4) $c_1 = 1$ 이라 하고 카탈란 수를 차례로 나열하면
1, 2, 5, 14, 42, 132, 429, ...

(5) 카탈란 수의 일반항은

$$c_2 = {}_{2n}C_n - {}_{2n}C_{n-1} = \frac{(2n)!}{n!(n+1)!} = \frac{{}_{2n}C_n}{n+1}$$

2. n 쌍의 괄호가 만들 수 있는 가능한 모든 괄호 형태의 개수 $\Rightarrow c_n$ (단, $c_0 = 1$ 로 정의)

쌍의 개수	가능한 모든 괄호 형태	경우의 수
$n=0$	없음	1
$n=1$	()	1
$n=2$	() (), (())	2
$n=3$	() () (), () (()), (()) (), ((()))	5

(1) $c_4 = 14$ 를 구하는 방법

- ① () {3 쌍의 괄호} $\Rightarrow c_0 c_3$
- ② ({1 쌍의 괄호}) {2 쌍의 괄호} $\Rightarrow c_1 c_2$
- ③ ({2 쌍의 괄호}) {1 쌍의 괄호} $\Rightarrow c_2 c_1$
- ④ ({3 쌍의 괄호}) $\Rightarrow c_3 c_0$

$$\begin{aligned} \therefore c_4 &= c_0 c_3 + c_1 c_2 + c_2 c_1 + c_3 c_0 \\ &= 5 + 2 + 2 + 5 = 14 \end{aligned}$$

(2) $c_{n+1} = c_0 c_n + c_1 c_{n-1} + \dots + c_{n-1} c_1 + c_n c_0$

(3) 카탈란수 $c_0, c_1, c_2, \dots, c_n, \dots$ 을 계수로 하는 (무한) 다항식

$$f(x) = c_0 + c_1 x + c_2 x^2 + \dots + c_n x^n + \dots$$

$$\begin{aligned} \textcircled{1} \{f(x)\}^2 &= c_0 c_0 + (c_1 c_0 + c_0 c_1)x \\ &\quad + (c_2 c_0 + c_1 c_1 + c_0 c_2)x^2 + \dots \\ &= c_1 + c_2 x + c_3 x^2 + \dots + c_{n+1} x^n + \dots \end{aligned}$$

$$\textcircled{2} f(x) = x \{f(x)\}^2 + 1$$

$$\textcircled{3} f(x) = \frac{1 - \sqrt{1 - 4x}}{2x}$$

< 주령구(酒令具) >



[다빈치 14면체]

1975년 경주 안압지에서 출토된 정사각형 모양의 면 6개와 육각형 모양의 면 8개로 이루어진 14면체 주사위이다. 재질은 참나무이다. 각 면에는 다양한 별칙이 적혀 있어 신라인들의 음주 습관의 풍류를 보여주고 있다. 출토된 진품은 유물 보존 처리도중 불타버렸고, 복제품만 남아있다.

금성작무(禁聲作舞) : 소리내지 않고 춤추기

중인타비(衆人打鼻) : 여러 사람 코 두드리기

음진대소(飲盡大笑) : 술 한잔 다 마시고 크게 웃기

삼잔일거(三盞一去) : 한번에 술 석 잔 마시기

유범공과(有犯空過) : 덤벼드는 사람이나 별난 짓으로 골려도 가만히 있기

자창자음(自唱自飲) : 스스로 노래 부르고 마시기

곡비즉진(曲臂則盡) : 팔뚝을 구부려 다 마시기

농면공과(弄面孔過) : 얼굴 간질어도 꿈쩍하지 않기

임의청가(任意請歌) : 누구에게나 마음대로 노래시킴

월경일곡(月鏡一曲) : 월경 한 곡조 부르기

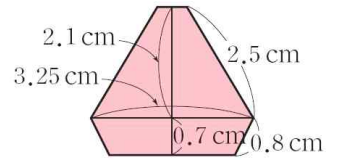
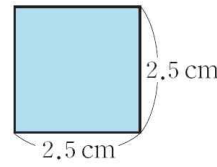
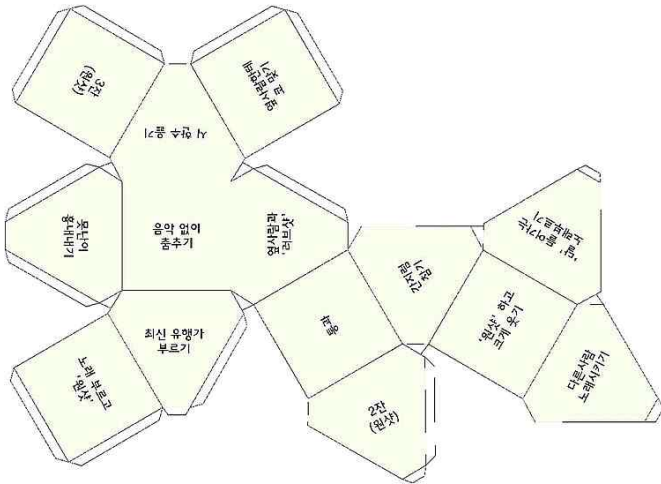
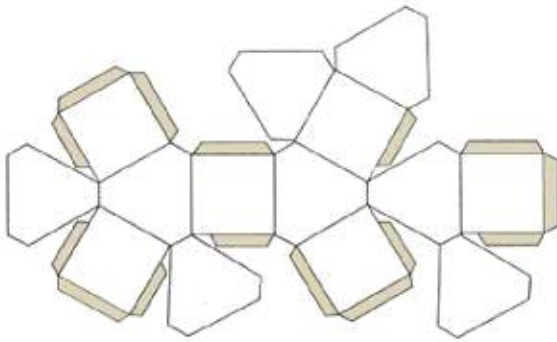
(달이란 여자에 관한 내용일 듯..)

공영시과(空詠詩過) : 시 한 수 읊기

양잔즉방(兩盞則放) : 술 두 잔이면 즉각 마시기

추물막방(醜物莫放) : 더러워도 버리지 않기

자창괴래만(自唱怪來晩) : 스스로 괴래만(밤늦게 곤드레 되어 들어오는 모양새)으로 부르기



실제로 이 주사위를 만들어 던지는 실험을 한 결과는 다음과 같다.

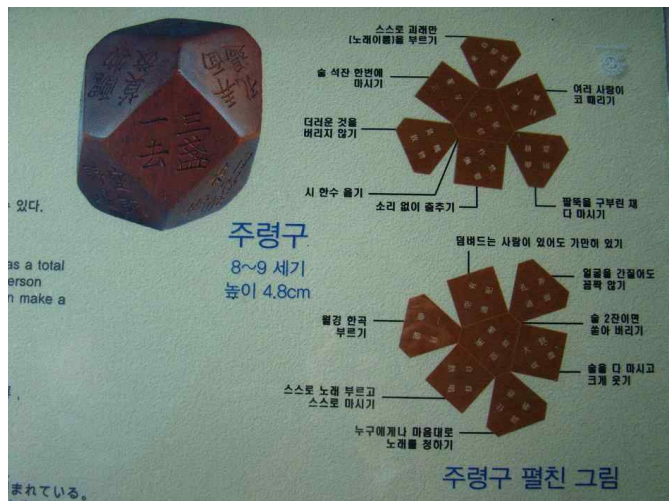
면	1	2	3	4	5	6	7	8	9	10	11	12	13	14	합계
도수	73	74	66	39	59	56	31	48	34	39	60	39	42	40	700
상대도수	0.104	0.106	0.094	0.056	0.084	0.080	0.044	0.068	0.049	0.056	0.086	0.056	0.060	0.057	1

< 모든 지도는 4가지 색으로 구분할 수 있다. >

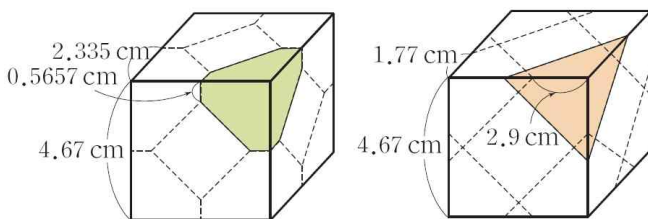
1852년 구드리(Guthrie, F. ; 1831~1899)는 영국의 지도를 색칠하다가 ‘인접한 구획들을 구분하여 칠하려면 최소한 몇 가지의 색이 필요할까?’라는 의문을 갖게 되었다. 그는 4가지 색만 사용하면 각 구획을 구분할 수 있을 것으로 생각하였다. 이 문제를 ‘4색 문제’라고 하는데 그 후 100여 년 동안 많은 수학자들이 이 문제를 증명하기 위하여 노력한 결과 다음이 밝혀졌다.

1. 1890년 하우드(Heawood, P. J.) : 5색을 쓰면 어떠한 지도라도 구분할 수 있다.
2. 1920년 프랭크린(Franklin, D.) : 구획의 수가 25개 이하이면 4색으로 구분할 수 있다.
3. 1968년 오어(Ore, O.)와 스템플(Stemple, J.) : 구획의 수가 40개 이하이면 4색으로 구분할 수 있다.

1976년 미국의 아펠(Appel, K.)과 하켄(Haken, W.)은 수백 시간 동안 컴퓨터를 이용하여 4색 문제를 증명하였다. 그러나 수학자들은 컴퓨터를 이용한 증명법에 별로 공감하지 않고 있다. 좀 더 간단하고 우아한 방법이 존재하리라는 것이 일반적인 수학자의 생각이다.



[참고] 목재주령구는 다음 그림과 같이 두 가지 방법으로 정육면체의 꼭짓점 부근을 잘라서 만들 수 있다.



목재주령구에서 정사각형 모양의 면의 넓이는 6.25 cm^2 이고, 육각형 모양의 면의 넓이는 6.265 cm^2 이다.

< 논리를 키우는 수학 >

다음 문장을 읽고, 물음에 답하여라.

어느 박람회에서는 행운의 번호가 적힌 입장권을 소지한 사람에게 상품을 준다고 한다. 입장권의 번호는 10 개의 숫자



0, 1, 2, 3, 4, 5, 6, 7, 8, 9
를 중복하여 사용하고, 이 중에서 행운의 번호는 앞에서 절반까지의 숫자의 합과 나머지 절반의 숫자의 합이 같은 경우이다. 예를 들어 두 자리 수가 적힌 10^2 (장)의 입장권 중에는

00, 11, 22, 33, 44, 55, 66, 77, 88, 99
와 같은 10 개의 행운의 번호가 있다. 같은 방법으로 여섯 자리 수가 적힌 10^6 (장)의 입장권 중에서 행운의 번호를 일일이 나열하면 그 개수는 55252 개임을 알 수 있다.

1. 두 자리 수가 적힌 $10^2 = 100$ (장)의 입장권 중에서 행운의 번호의 개수는 방정식

$$x + y = 9 \quad (x, y \in \{0, 1, 2, \dots, 9\})$$

의 해의 개수와 같음을 설명하여라.

2. 네 자리 수가 적힌 $10^4 = 10000$ (장)의 입장권 중에서 행운 번호의 개수는 방정식

$$x + y + z + w = 18 \quad (x, y, z, w \in \{0, 1, 2, \dots, 9\})$$

의 해의 개수와 같음을 설명하고, 그 개수를 구하여라.

[해설] 1. 두 자리 수의 행운의 번호 $\square x \square \square x \square$ ($x = 0, 1, 2, \dots, 9$)는 $\square x \square \square 9 - x \square$ 에 하나씩 대응시킬 수 있으므로 $9 - x = y$ 라고 하면 x, y 는 $x + y = 9$ 를 만족한다.

따라서 두 자리 수가 적힌 입장권 중에서 행운의 번호의 개수는 방정식 $x + y = 9$ ($x, y \in \{0, 1, 2, \dots, 9\}$)의 해의 개수와 같다.

[해설] 2. 네 자리 수의 행운의 번호를 $\square a \square \square b \square \square c \square \square d \square$ 라고 하면

$$a + b = c + d \quad (a, b, c, d \in \{0, 1, 2, \dots, 9\})$$

이때 $\square a \square \square b \square \square c \square \square d \square$ 는 $\square a \square \square b \square \square 9 - c \square \square 9 - d \square$ 에 하나씩 대응시킬 수 있으므로 $a = x, b = y, 9 - c = z, 9 - d = w$ 라고 하면 x, y, z, w 는

$$x + y + z + w = 18$$

을 만족한다.

따라서 네 자리 수가 적힌 행운의 번호의 개수는 방정식

$$x + y + z + w = 18 \quad (x, y, z, w \in \{0, 1, 2, \dots, 9\})$$

의 해의 개수와 같다.

방정식 $x + y + z + w = 18$ 의 음이 아닌 정수해의 개수는

$${}_4H_{18} = {}_{21}C_{18} = {}_{21}C_3 = 1330$$

이때 $x \geq 10$ 인 경우의 음이 아닌 정수해의 개수는 $x = x' + 10$ 이라고 할 때 $x' + y + z + w = 8$ 의 음이 아닌 정수해의 개수와 같으므로

$${}_4H_8 = {}_{11}C_8 = {}_{11}C_3 = 165$$

같은 방법으로 $y \geq 10, z \geq 10, w \geq 10$ 인 경우의 음이 아닌 정수해의 개수도 각각 165 이므로 구하는 해의 개수는

$$1330 - 4 \times 165 = 670$$

II. 확 률

1. 확 률

① 시행과 사건

1. 뜻

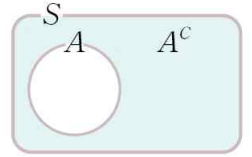
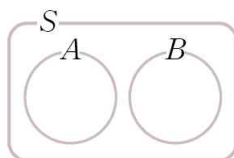
- (1) 시행 : 동일한 조건에서 여러 번 반복할 수 있고, 그 결과가 우연에 의하여 지배되는 실험 or 관찰
- (2) 표본공간(전사건) : 어떤 시행에서 나올 수 있는 모든 결과들의 집합 $\Rightarrow S$ or Ω
- (3) 사건(Event) : 시행의 결과 or 표본공간의 부분집합
- (4) 근원사건 : 사건 중에서 더 이상 세분할 수 없는 기본적인 사건 or 표본공간의 한 원소로만 이루어진 집합
- (5) 전사건(전체집합) : 반드시 일어나는 사건
- (6) 공사건(공집합) : 절대로 일어날 수 없는 사건 $\Rightarrow \emptyset$

2. 합사건, 곱사건 : 두 사건 A, B 에 대하여

- (1) 합사건 : A 또는 B 가 일어나는 사건
 $\Rightarrow A \cup B$
- (2) 곱사건 : A 와 B 가 동시에 일어나는 사건
 $\Rightarrow A \cap B$

3. 배반사건, 여사건 : 두 사건 A, B 에 대하여

- (1) 배반사건 : A, B 가 동시에 일어나지 않을 때
 $\Rightarrow A \cap B = \emptyset$



- (2) 여사건 : 어떤 시행에서, 사건 A 가 일어나지 않는 사건 $\Rightarrow A^c$

$\square A$ 와 A^c 는 서로 배반 $\iff A \cap A^c = \emptyset$

② 확률의 뜻

\hookrightarrow Probability

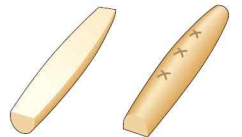
0. 확률 : 어떤 사건이 일어날 수 있는 가능성을 수의 값으로 나타낸 것
 \Rightarrow 사건 A 가 일어날 확률 : $P(A)$

★ $n(A)$: 사건 A 가 일어나는 경우의 수

1. 수학적 확률(선형적 확률)

$$\begin{aligned}
 P(A) &= \frac{(\text{사건 } A \text{가 일어나는 경우의 수})}{(\text{일어날 수 있는 모든 경우의 수})} \\
 &= \frac{(\text{사건 } A \text{의 원소의 개수})}{(\text{표본공간 } S \text{의 원소의 개수})} \\
 &= \frac{n(A)}{n(S)}
 \end{aligned}$$

- \square 각 근원사건의 일어날 가능성이 모두 같을 때 즉, 균등확률로 정의한다.



2. 통계적 확률(경험적 확률)

$$P(A) = \lim_{n \rightarrow \infty} \frac{r_n}{n} = p \quad (\text{일정한 값} \approx \text{수학적 확률})$$

- $\square \frac{r_n}{n}$ 는 상대도수 (n : 전체도수, r_n : A 의 도수)

- \square 통계적 확률 : 수집한 자료를 바탕으로 경향을 예측한 확률

3. 기하학적 확률

$$P(A) = \frac{(A \text{ 영역의 넓이})}{(\text{전체 영역의 넓이})}$$

☑ 선분의 길이, 실수 · 면적 · 부피의 크기 등

4. 확률의 기본 성질

- (1) 임의의 사건 A 에 대하여 $0 \leq P(A) \leq 1$
 $\because 0 \leq n(A) \leq n(S)$
- (2) 표본공간 S 에 대하여 $P(S) = 1$
- (3) 공사건 \emptyset 에 대하여 $P(\emptyset) = 0$

3 확률의 덧셈 정리

1. 확률의 덧셈 정리

- (1) $P(A \cup B)$ 의 의미
 - ① A 또는 B 가 일어날 확률
 - ② A, B 중 적어도 하나가 일어날 확률
- (2) 사건 A, B 에 대하여

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

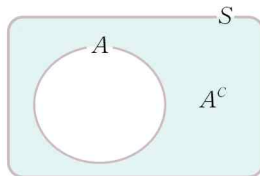
$$\because n(A \cup B) = n(A) + n(B) - n(A \cap B)$$
- (3) 사건 A 와 B 가 서로 배반사건 $\Leftrightarrow A \cap B = \emptyset$
 $\Leftrightarrow P(A \cup B) = P(A) + P(B)$
- (4) 사건 A, B, C 가 서로 배반
 $\Leftrightarrow P(A \cup B \cup C) = P(A) + P(B) + P(C)$
- (5) 사건 A_1, A_2, \dots 가 서로 배반
 $\Leftrightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

2. 여사건의 확률

$$P(A^C) = 1 - P(A)$$

$$\because 1 = P(S) = P(A \cup A^C) = P(A) + P(A^C)$$

- ☑ $P(\text{적어도 하나가 } \bigcirc) = 1 - P(\text{모두 } \times)$
 ☑ $P(X \geq a) = 1 - P(X < a)$



3. 확률의 계산

- (1) $P(A^C \cup B^C) = P\{(A \cap B)^C\} = 1 - P(A \cap B)$
- (2) $P(A^C \cap B^C) = P\{(A \cup B)^C\} = 1 - P(A \cup B)$
- (3) $P(A \cap B^C) = P(A) - P(A \cap B)$
 $= P(A \cup B) - P(B)$

4. 확률의 일반적 정의

표본공간 $S = \{u_1, u_2, \dots, u_n\}$ 에 대하여
 각 원소가 나올 확률이 $P(\{u_i\}) = p_i$ 일 때,

- (1) $0 \leq p_i \leq 1$ ($i = 1, 2, \dots, n$)

$$(2) \sum_{i=1}^n p_i = 1$$

$$(3) P(A) = \sum_{u_i \in A} P(\{u_i\})$$

- (4) $p_i = \frac{1}{n}$ 일 때, 균등확률이라 한다.

< 수학적 확률 >

어떤 시행에서

- ① 일어날 수 있는 모든 경우의 수가 n 가지이고
- ② 이 n 가지 경우는 어느 두 가지도 동시에 일어나지 않으며,
- ③ n 가지 경우가 일어날 가능성이 같은 경우로 기대될 때

n 가지 경우 중 사건 A 가 일어날 경우가 r 가지이면

$$\Rightarrow \text{사건 } A \text{ 가 일어날 수학적 확률 } P(A) = \frac{r}{n}$$

< 통계적 확률에서 고려해야 할 사항 >

주사위를 n 번 던질 때, 1의 눈이 나오는 횟수가 r_n 이면 그 상대도수는 다음과 같다.

$$\frac{(1 \text{의 눈이 나오는 횟수})}{(\text{전체 시행 횟수})} = \frac{r_n}{n}$$

여기서 n 을 충분히 크게 하면 상대도수 $\frac{r_n}{n}$ 은 어떤 값 (이를테면 $\frac{1}{6}$) 에 가까이 가게 될 것이고, $\frac{r_n}{n}$ 이 가까이 가는 값을 통계적 확률이라고 한다.

이러한 통계적 확률의 뜻에서 고려해야 할 사항 세 가지를 알아보자.

첫째, 시행 횟수 n 을 충분히 크게 한다는 것은 무슨 뜻인가? 어느 정도의 n 이면 충분히 크다고 볼 수 있는가? n 이 100 이면 충분히 큰 것인가? 아니면 n 이 백 만 정도는 되어야 충분히 큰 것인가?

이러한 의문에 대한 답(누구나 그렇다고 인정하는 답)은 없다. 가장 그럴듯한 답은 ‘충분히 크다고 느낄 만큼 큰 것’이다. 그러나 이러한 답에 선뜻 수긍하기는 어렵다. 그러므로 통계적 확률을 구할 때, 시행 횟수 n 을 정하는 일을 신중히 생각하여야 한다.

둘째, 상대도수 $\frac{r_n}{n}$ 이 어떤 값에 가까이 간다는 뜻은 무엇인가? $\frac{r_n}{n}$ 이 어떤 값에 가까이 간다고 할 수 있는가? 어떤 수열 $\{a_n\}$ 에서 $\{a_n\}$ 이 α 에 가까이 가느냐, 가지 않느냐를 수학적으로 판정하는 것은 그리 어렵지 않은 일이다.

그러나 상대도수 $\frac{r_n}{n}$ 으로 이루어지는 수열에서는 일반항을 찾을 수 없기 때문에 이 수열의 수렴 여부를 판정하는 것은 쉬운 일이 아니다.

셋째, 상대도수 $\frac{r_n}{n}$ 이 가까이 가는 구체적인 값은 무엇인가? 즉, $\frac{r_n}{n}$ 의 극한값은 무엇인가? 주사위를 던질 때, 1의

눈이 나오는 상대도수 $\frac{r_n}{n}$ 의 극한값은 $\frac{1}{6}$ 이라고 많은 사람들은 의심 없이 믿고 있다. 그러나 윷을 던질 때, 윷짝의 불룩한 면이 나오는 상대도수 $\frac{r_n}{n}$ 에 대한 극한값은 아무도

언급을 하지 않고 있다. 그 이유는 주사위 던지기에서는 수학적 확률을 생각하기가 쉽고, 윷 던지기에서는 어렵기 때문이다. 다시 말하면 통계적 확률의 구체적인 값을 구하기는 매우 어려운 것이다.

< 공리적 확률 >

수학자 콜모고로프는 다음과 같은 공리를 만족하는 것을 확률이라고 정의하였다.

표본공간 S 에서 어떤 사건 A 에 대하여

[공리1] $P(A) \geq 0$

[공리2] $P(S) = 1$

[공리3] A_1, A_2, A_3, \dots 의 표본공간 S 에서 정의된

서로 배반인 사건일 때,

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum_{i=1}^{\infty} P(A_i)$$

위의 세 가지 공리를 만족할 때, $P(A)$ 를 사건 A 의 공리적 확률이라 한다.

공리적 확률은 수학적 확률 또는 통계적 확률과 같이 어떤 구체적인 식으로 정의된 것이 아니라 확률이 만족해야 하는 기본 성질을 형식적으로 나타낸 것이다. 따라서 공리적 확률로는 확률의 값을 구할 수 없다. 표본공간이 유한이고, 각 원소가 같은 정도로 일어날 것이 기대되면 수학적 확률이 위의 공리를 만족하므로 수학적 확률로 확률의 값을 구하고, 그렇지 않으면 통계적 확률로 확률의 값을 구하면 된다.

< 두 사람이 만날 확률 >

[문제] A, B 두 사람이 11시와 12시 사이에 어떤 장소에서 만나기로 하였다. 누가 먼저 도착하던지 10분 동안만 기다리기로 하였을 때, A 와 B 가 이 장소에서 만날 확률을 구하여라.

[1단계] 문제를 이해하여 보자.

(1) A 와 B 가 약속 장소에 도착하는 시각을 각각 11시 x 분, 11시 y 분이라고 할 때, x 가 가질 수 있는 값의 범위와 y 가 가질 수 있는 값의 범위를 구하여라.

$$\Rightarrow 0 \leq x \leq 60, 0 \leq y \leq 60$$

(2) A 가 11시 30분에 도착한다고 할 때, 두 사람이 만나려면 B 는 언제 도착하면 되는가?

$$\Rightarrow 11시 20분에서 11시 40분까지 도착하면 된다.$$

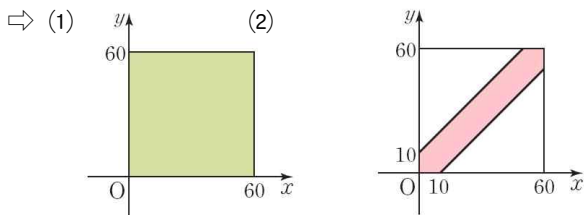
(3) 두 사람이 만나기 위한 x, y 사이의 관계식을 구하여라.

$$\Rightarrow |x - y| \leq 10$$

[2단계] 계획을 세워 보자.

(1) x, y 가 가질 수 있는 값의 범위를 나타내어라.

(2) (1)의 영역 안에서 [1단계]의 (3)에서 구한 x, y 사이의 관계식을 만족하는 영역을 나타내어라.



[3단계] 문제를 풀어 보자.

(1) [2단계]의 (1)에서 나타낸 영역의 넓이를 구하여라.

$$\Rightarrow 60 \times 60 = 3600$$

(2) [2단계]의 (2)에서 나타낸 영역의 넓이를 구하여라.

$$\Rightarrow 3600 - 2 \cdot \frac{1}{2} \cdot 50 \cdot 50 = 1100$$

(3) 두 사람이 만날 확률을 구하여라.

$$\Rightarrow \frac{1100}{3600} = \frac{11}{36}$$



< 몬티 홀 문제(Monty Hall Problem) >

미국의 한 TV 프로그램인 'Let's make a deal'의 진행자인 몬티 홀과 출연자와의 게임에서 유래되었다.

[문제] 출연자 앞에 3개의 문이 있고, 출연자가 하나의 문을 선택하여 그 뒤에 있는 상품을 받게 된다. 3개의 문 중 하나의 문 뒤에는 멋진 자동차가 있고, 다른 2개의 문 뒤에는 염소가 있다. 출연자가 하나의 문을 선택하면 이미 모든 상황을 알고 있는 사회자는 나머지 2개의 문 중에서 염소가 있는 문을 하나 열어 보인다.

그리고 출연자에게 “자, 당신이 선택한 문을 바꾸셔도 좋습니다. 바꾸시겠습니까?”라고 질문한다. 이 상황에서 출연자는 어떤 선택을 하여야 할까?

[마릴린의 답] 결론은 ‘바꾸는 것이 좋다’이다.

선택을 바꾸면 행운을 잡을 확률이 $\frac{2}{3}$ 이고,

바꾸지 않으면 그 확률은 $\frac{1}{3}$ 이기 때문이다.

[해설] 장막에 가려진 세 곳을 각각

1번, 2번, 3번이라고 할 때, 그곳에 자동차 또는 염소가 들어 있는 경우는 오른쪽 표와 같이

1번	2번	3번
자동차	염소	염소
염소	자동차	염소
염소	염소	자동차

3가지이다. 이 중에서 참여자가

처음에 어떤 곳을 선택하더라도 안 바꾸는 것이 유리한 경우는 1가지이고, 바꾸는 것이 유리한 경우는 2가지이다.

[결론] 인간의 직관과 수학적 확률이 다를 수 있다는 것을 잘 보여 주는 예.

< 우수한 인재를 뽑아라 >

- 비서류 문제 또는 지참금 문제 -

어느 회사에서 채용 공고를 내었을 때 위와 같이 지원자가 많이 몰려오면 지원자들을 모두 면접하기가 매우 어려운 일이 된다. 이럴 때는 약간의 위험이 따르더라도 가장 우수한 인재를 뽑을 확률이 높은 방법을 생각하게 된다. 즉, 모든 지원자를 면접하지 않고 어느 정도의 인원만 면접한 후에 채용 결정을 내리는 것이다. 이제 다음과 같은 전제 조건으로 가장 우수한 인재를 뽑는 경우를 생각해보자.

- 지원자는 무작위로 분포되어 있다.
- 지원자 전체 인원수는 알고 있지만 그 외 사항은 전혀 모른다.
- 한 사람씩 면접하면서 바로 채용 여부를 결정한다.
- 채용하지 않기로 결정한 후에는 다시 결정을 바꿀 수 없다.

위의 전제 조건으로 가장 우수한 인재를 뽑는 확률을 극대화시키는 방법은 어느 정도 인원까지는 면접 후 무조건 보내고 그 다음부터 이미 면접한 지원자보다 우수한 지원자를 만나면 채용하는 것으로 알려져 있다. 문제는 전체 N 명 중에서 무조건 r 명을 보낸다고 할 때, r 의 값을 어떻게 구하느냐이다.

예를 들어, 지원자의 우수한 정도를 순서대로 1, 2, 3으로 표시하자. 세 명을 면접하는 순서는 세 수를 늘어놓는 순서와 같으므로 6가지의 경우가 생긴다. 먼저 한 명을 보내고 그 다음 지원자 중 이미 면접한 지원자보다 우수한 경우 지원자를 뽑는 경우를 생각해보자.

- 123 : 1을 보내고 1보다 높은 2를 뽑으면 실패
 132 : 1을 보내고 1보다 높은 3을 뽑으면 성공
 213 : 2를 보낸 후 1을 2보다 낮으므로 또 보내고
 그 다음 3이 앞의 2, 1보다 높으므로 뽑으면 성공

이와 같은 식으로 면접을 진행하면 132, 213, 231 세 가지 경우에 성공하는 것이므로 확률은 $1/2$ 이다. 이번에는 두 명을 무조건 보내는 경우를 생각해보면 123, 213의 경우만 성공하는 것이므로 확률은 $1/3$ 이다. 따라서 지원자가 세 명인 경우에는 한 명을 무조건 보내고 그 다음에 결정하는 것이 최선이고, 그 확률은 0.5이다.

1. 지원자가 4명일 때, 지원자의 능력을 1, 2, 3, 4로 나타내자. 네 명의 면접 순서는 4!로 24가지가 가능하다.

1234 1243 1324 1342 1423 1432

2134 2143 2314 2341 2413 2431
 3124 3142 3214 3241 3412 3421
 4123 4132 4213 4231 4312 4321

- ① 한 명을 무조건 보내고 그 다음 지원자 중 이미 면접한 지원자보다 우수한 지원자를 뽑을 때 가장 우수한 4를 뽑는 경우는 1423, 1432, 2143, 2413, 2431, 3124, 3142, 3214, 3241, 3412, 3421로 그 확률은 $11/24$
- ② 두 명을 무조건 보내고 나서 뽑는 경우는 1243, 1324, 1342, 2143, 2314, 2341, 3124, 3142, 3214, 3241로 그 확률은 $10/24$ 이다.
- ③ 세 명을 무조건 보내고 나서 뽑는 경우는 1234, 1324, 2134, 2314, 3124, 3214로 그 확률은 $6/24$ 이다. 따라서 지원자가 네 명인 경우에 한 명을 보내고 그 다음에 결정하는 것이 최선이다. 즉, $r = 1$ 이다.

2. n 명의 지원자가 있다고 할 때, n 이 증가하면서 $n!$ 의 값이 매우 급속히 커져 이 과정을 계속 되풀이할 수는 없으므로 이제 일반적인 경우로 생각해보자.

전체 지원자는 n 명, 처음부터 r 명까지는 면접 후 바로 보내고, $(r+1)$ 번째 사람부터 그때까지 면접한 지원자보다 우수한(여기에서는 1, 2, 3, ..., n 으로 점수로 표현) 지원자가 나타나면 뽑는다고 할 때, 가장 우수한(점수가 높은) 지원자를 뽑을 확률을 구하고, 그 확률을 최대로 하는 r 의 값을 구해보자.

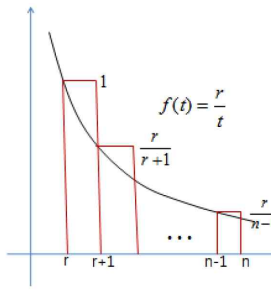
지원자들의 점수가 알려져 있지 않으므로 모든 지원자가 가장 높은 점수를 가질 확률은 모두 $1/n$ 로 같다고 본다. 1부터 r 번째까지 지원자를 보낸 후 그 다음 지원자부터 가장 높은 점수를 가진 사람이 뽑힐 확률은 다음과 같다.

① $(r+1)$ 번째에 가장 높은 능력을 지닌 지원자가 있어서 뽑힐 확률은 $1/n$ 이다. 그런데 이 지원자가 뽑히는 경우는 1번째부터 r 번째 지원자 중 가장 높은 점수를 갖는 사람이 1번째부터 r 번째 사이에 있는 경우로 확률은 r/r 이다. 따라서 $(r+1)$ 번째 지원자가 뽑힐 확률은 $r/nr = 1/n$ 이다.

② $(r+2)$ 번째에 가장 높은 능력을 지닌 지원자가 있을 확률은 $1/n$ 이다. 그런데 이 지원자가 뽑히는 경우는 1번째부터 $(r+1)$ 번째 지원자 중 가장 높은 점수를 갖는 사람이 1번째부터 r 번째 사이에 있는 경우로 확률은 $r/(r+1)$ 이다. 따라서 $(r+2)$ 번째 지원자가 뽑힐 확률은 $r/n(r+1)$ 이다.

- ③ 마찬가지로 생각하면, $(r+3)$ 번째 지원자가 가장 높은 점수를 갖고 있으면서 뽑힐 확률은 $r/n(r+2)$ 이고, 마지막 지원자가 뽑힐 확률은 $r/n(n-1)$ 이다.

위 사건들은 동시에 일어나지 않으므로 각각의 경우를 모두 더하면

$$\begin{aligned}
 P(r) &= \frac{1}{n} + \frac{r}{n(r+1)} + \frac{r}{n(r+2)} + \cdots \\
 &\quad \cdots + \frac{r}{n(n-1)} \\
 &= \frac{1}{n} \sum_{t=r}^{n-1} \frac{r}{t} \\
 &\doteq \frac{1}{n} \int_r^n \frac{r}{t} dt \\
 &\quad \left| \begin{array}{l} u = \frac{t}{r} \\ \Rightarrow r du = dt \end{array} \right. \\
 &= \frac{r}{n} \int_1^{n/r} \frac{1}{u} du \\
 &= \frac{r}{n} \ln\left(\frac{n}{r}\right) = -\frac{r}{n} \ln\left(\frac{r}{n}\right)
 \end{aligned}$$


이다. 이제 확률 $P(r)$ 이 최댓값을 갖는 r 을 구하기 위하여 $P(r)$ 을 미분하면

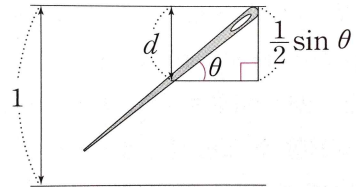
$$P'(r) = -\frac{1}{n} \left\{ 1 + \ln\left(\frac{r}{n}\right) \right\}$$

이 값이 0이 되는 경우는 $\ln\left(\frac{r}{n}\right) = -1$, 즉 $r = \frac{n}{e}$ 일 때이다. 따라서 지원자가 n 명인 경우 n/e 명은 인터뷰 후 바로 보내고 그 다음부터 이제까지 인터뷰한 지원자 중 가장 우수한 지원자를 뽑으면 가장 성공할 확률이 높다. 한편, 급수를 적분으로 고치는 과정에서 급수에서의 의 범위와 적분 구간이 일치하지 않는 것에 대해서는 적절히 판단하도록 한다.

< 뷔퐁(Buffon)의 바늘 문제 >

프랑스의 수학자 뷔퐁(Buffon; 1707~1788)은 다음과 같은 문제를 제시하였다.

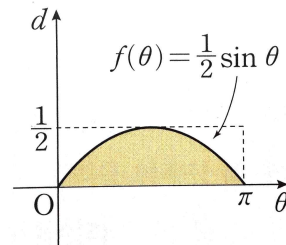
“평행하고 동일한 간격의 줄이 그려진 마룻바닥에 바늘을 하나 떨어뜨렸다. 이때, 바늘이 두 평행선 사이에 떨어질 확률은 얼마인가?”



평행선 사이의 간격을 1, 바늘의 길이를 1 이라고 하고, 바늘의 중점과 가장 가까운 직선 사이의 거리를 d , 바늘과 직선이 이루는 각의 크기를 θ 라 하자.

이때, 표본공간 S 는

$$S = \left\{ (d, \theta) \mid 0 \leq d \leq \frac{1}{2}, 0 \leq \theta \leq \pi \right\}$$



구하는 확률은 $d \leq \frac{1}{2} \sin \theta$ 일 확률이므로

$$\begin{aligned}
 P(A) &= \frac{(A \text{ 영역의 넓이})}{(\text{전체 영역의 넓이})} \\
 &= \frac{\int_0^\pi \frac{1}{2} \sin \theta d\theta}{\frac{1}{2} \pi} = \frac{2}{\pi}
 \end{aligned}$$

뷔퐁은 이러한 원리를 이용하면 바늘을 마룻바닥에 던지는 실험을 계속하여 평행선과 만나는 횟수를 조사하여 상대도수를 구하면 원주율 π 의 값을 계산할 수 있다고 주장하였다.

실제로, 바늘을 던지는 실험을 계속하여 반복하면 원주율 π 의 근삿값을 구할 수 있다. $\Rightarrow \frac{355}{113}$

< 도박과 확률 >

17세기 중반 상류층 귀족인 드메레(de Mere, C. 1607~1684)는 내기를 좋아했다.

1. 주사위 한 개를 네 번 던질 때, 6의 눈이 적어도 한 번은 나온다는 데 돈을 걸어 상당한 이익을 보았다. 그는 경험을 통해 주사위 한 개를 네 번 던질 때, 6의 눈이 적어도 한 번은 나올 확률이 그렇지 않을 확률보다 더 크다는 것을 알고 있었다. 수학적으로 계산하면 실제로 그 확률은 얼마쯤 될까?

⇒ 여사건의 확률을 이용하면 주사위 한 개를 네 번 던질 때, 6의 눈이 적어도 한 번은 나올 확률은 $1 - \left(\frac{5}{6}\right)^4 \approx 0.518$ 이다. 즉, 주사위 한 개를 네 번 던질 때, 6의 눈이 적어도 한 번은 나올 확률은 약 51.8%였던 것이다.

2. 주사위 두 개를 동시에 24번 던질 때 나온 두 눈의 수의 합이 12가 되는 경우가 적어도 한 번 있을 확률은 0.5보다 클까? 작을까? 같을까? 드메레는 0.5보다 크다고 생각하여 돈을 걸었다. 처음에는 자신에게 유리한 듯 보였지만 시간이 흐를수록 점점 돈을 잃기 시작했다. 그래서 절친한 친구인 파스칼(1623~1662)에게 편지를 보내 이 내기를 분석해달라고 부탁했다. 이것이 확률론이 시작으로 전해진다.

⇒ 주사위 두 개를 동시에 24번 던질 때 나온 두 눈의 수의 합이 12가 되지 않을 확률이 $\left(\frac{35}{36}\right)^{24} \approx 0.509$ 이므로 주사위 두 개를 동시에 24번 던질 때 나온 두 눈의 수의 합이 12가 될 확률은 약 0.491 즉, 49.1%였던 것이다.

< 카오스 게임 >

주사위를 던질 때, 윗면에 나오는 눈의 수를 항상 정확히 알아맞히는 사람은 아무도 없다. 윗면에 나오는 눈의 수는 어떤 규칙에 의하여 나타나는 것이 아니고 단지 우연에 의한 것이기 때문이다. 즉, 주사위 던지기에서 각 시행의 결과에 대하여 우리는 ‘혼돈(混沌, chaos)’의 상태에 있는 것이

다.

그러나 이 시행을 여러 번 반복하면 규칙성을 찾을 수 있다.

다음과 같은 게임을 하여 보자.

[1단계] 정삼각형 ABC 를 그리고, 삼각형의 변 위에 임의의 한 점을 찍는다.

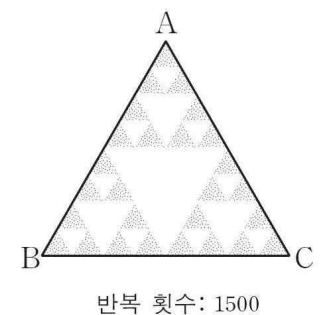
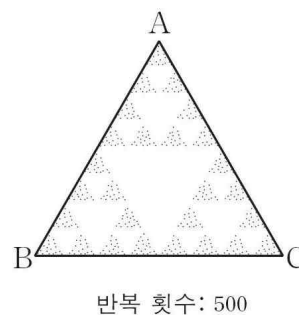
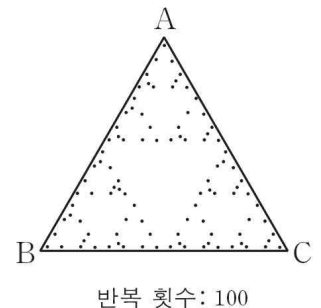
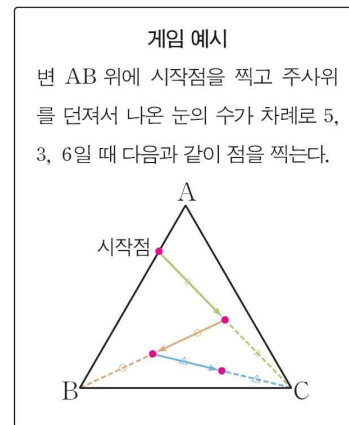
[2단계] 주사위를 던져서 나오는 눈의 수에 따라

다음과 같이 새로운 점을 찍는다.

- ① 눈의 수가 1 또는 2일 때 : 주어진 점과 점 A 의 중점
- ② 눈의 수가 3 또는 4일 때 : 주어진 점과 점 B 의 중점
- ③ 눈의 수가 5 또는 6일 때 : 주어진 점과 점 C 의 중점

[3단계] 새로운 점에 대하여 위의 [2단계]를 반복한다.

다음 그림은 위의 시행에서 반복하는 횟수를 각각 100, 500, 1500으로 하였을 때의 예시 그림이다.



< 확률의 역사 >

확률의 개념은 15세기 말에서 16세기 초 무렵 처음으로 도입되었다. 1487년 파촐리(Pacioli, L. ; 1445~1517)는 그의 논문 “산수, 기하학, 비례와 비례적인 것들의 대전 (Summa de arithmetica, geometria, proportioni et proportionalita)”에 다음과 같은 득점 문제를 제시하였다.

‘두 사람 A와 B가 같은 금액의 판돈을 걸고 먼저 다섯 게임을 이긴 사람이 판돈을 모두 갖는 게임을 하였다. 그런데 두 사람이 게임을 하다가 불가피한 이유로 A가 2승 1패한 상황에서 게임을 중단할 수밖에 없었다. 이때 두 사람이 승률이 같다면 판돈을 어떻게 나누는 것이 공정한가?’

파촐리는 이 득점 문제의 해법을 제시하였지만 그의 해법은 틀린 것이었다.

1526년 카르다노(Cardano, G. ; 1501~1576)가 쓴 “우연 게임에 관한 책”은 도박 안내 책자이자 확률의 기본 개념을 확립한 최초의 책으로 그의 사후 1663년에 발표되었다. 카르다노는 이 책에서 ‘circuit’(오늘날의 표본공간)이라는 개념을 처음으로 언급하였고, 두 개의 주사위를 던져서 나오는 눈의 수의 합이 10일 경우의 확률을 계산하는 방법을 제시하였다. 또 득점 문제의 해법도 제시하였지만 그의 해법도 틀렸다고 한다.

갈릴레이(Galilei, G. ; 1564~1642)는 그의 사후 1718년에 발표된 “주사위 게임에 관한 생각”이라는 책에서 세 개의 주사위를 던질 때 나오는 눈의 수의 합이 10이 되는 경우가 합이 9가 되는 경우보다 더 자주 일어난다는 것을 밝혔다.

확률의 개념은 1654년 파스칼(Pascal, B. ; 1623~1662)과 페르마(Fermat, P. ; 1601~1665)가 주사위 문제와 득점 문제를 해결하는 과정에서 정립되었다. 프랑스의 귀족이었던 드메레(deMere, C. ; 1607~1684)는 파스칼에게 다음과 같은 주사위 문제를 질문하였다.

‘두 개의 주사위를 던져서 적어도 한 번 이상 (6, 6)이 나올 우연이 $\frac{1}{2}$ 보다 크려면 최소 몇 번을 던져야 하는가?’

드메레는 주사위를 24번 던지면 된다고 생각했지만 그의 생각은 잘못되었고, 파스칼은 페르마와의 서신 왕래를 통하여 시행의 수가 25번 되어야 함을 확인하였다.

[참고] 적어도 한 번 이상 (6, 6)의 눈이 나올 확률 p 는 $p = 1 - \left(\frac{35}{36}\right)^n$ 이고, $n = 24$ 일 때 $p = 0.4914$, $n = 25$ 일 때 $p = 0.5055$ 이다.

파스칼이 해결한 두 번째 문제는 득점 문제였다. 그는 게임이 계속된다고 할 때, 일반화된 득점 문제의 해법은 각 도박사의 승률에 비례해서 몫을 배분해야 한다고 하면서 이항계수의 산술 삼각형(파스칼의 삼각형)을 이용하여 그 몫을 계산하였다.

이와 같은 파스칼과 페르마의 해법은 확률의 기원을 이루었고, 많은 수학자로 하여금 확률에 관한 연구를 가능하게 하였다.

< 확률 이론의 발달 >

파스칼과 페르마 이후 확률 이론이 수학의 원리로 굳건히 자리 잡게 된 것은 하위헌스, 베르누이, 드무아브르, 라플라스, 가우스 등과 같은 수학자의 공헌에 의한 것이다. 이들의 공헌을 간단하게 살펴보면 다음과 같다.

(1) 하위헌스(Huygens, C. ; 1629~1695)

1657년 네덜란드의 수학자 하위헌스는 확률 이론에 관한 논문 “우연 게임에 관한 추론”을 발표하였다. 이 논문은 실제로 우연 게임에 관한 첫 번째 논문으로(카르다노의 책은 그의 사후 1663년에 발표됨) 3개의 명제와 11개의 정리로 구성되어 있다.

그가 제시한 3개의 명제는 기댓값에 관한 것이고, 11개의 정리는 파스칼과 페르마의 주사위 문제를 체계적으로 정리, 확장한 것이다. 그리고 그의 논문의 끝 부분에는 파스칼과 페르마가 제시한 5개의 확률 문제를 첨가하였고, 그의 논문은 반세기 동안 확률 이론의 입문서가 되었다. 파스칼, 페르마, 하위헌스의 연구로부터 17세기 중엽까지 확률의 덧셈법칙, 확률의 곱셈법칙 등이 발달하였으나 18세기 중엽까지 확률 이론은 수학적으로 큰 진전이 없었다.

(2) 베르누이(Bernoulli, J. ; 1654~1705)

1713년 베르누이는 “추측술”이라는 책을 출간했는데 그는 이 책에서 초기 확률 이론의 개념적 혁신이라고 할 수 있는 큰 수의 법칙을 증명하였다.

이 책은 4개의 장으로 구성되어 있는데 제1장에서는 베르누이의 가장 중요한 공헌인 하위헌스의 정리 1, 2, 3, 4, 14와 하위헌스의 문제 5를 조합론, 덧셈법칙, 곱셈법칙, 무한수열의 이론을 이용하여 해결한 내용을 다루었다. 제2장에서는 여러 가지 순열과 조합 이론을 다루었고, 제3장에서는 여러 가지 우연 게임과 주사위 게임에 대한 문제를 합의 법칙, 곱의 법칙, 조합론, 기댓값의 계산, 조건부 기댓값 등으로 체계적으로 해결하였고, 귀납법을 사용하였

다. 제 4 장에서는 사회문제, 윤리문제, 경제문제에 대한 확률의 응용에 대해 논의하였고, 큰 수의 법칙을 다루었다. 큰 수의 법칙은 확률 이론의 실제적 응용과 이론 발달에 큰 공헌을 하였다.

(3) 드무아브르(de Moivre, A. ; 1667~1754)

드무아브르의 확률 이론에 관한 첫 번째 논문은 1711 년 “우연의 측정이나 우연성에 의존하는 게임에서 사건이 일어날 확률”이다. 이 논문은 공 굴리기 게임에서의 득점 문제와 도박사의 패망 문제 등을 포함하고 있으며 확률의 발달사에 중요한 공헌을 하였다. 이 논문의 대부분의 내용은 1718 년에 발표한 드무아브르의 “우연론”에 포함되어 있다. 이 책의 제 3 판은 1756 년에 출간되었으며 과거 반세기 동안의 확률 이론을 총망라한 최초의 현대적 확률 교재이다.

“우연론”에서는 확률의 정의(같은 정도로 일어난다는 가정이 없음)와 기본 정리, 그리고 일련의 문제들로 구성되어 있다. “우연론”의 첫 번째 개정판에는 합성 확률, 도박사의 패망 확률 등 53 개의 확률 문제가 실려 있고, 두 번째 개정판에는 득점 문제, 주사위 확률, 확률의 곱셈정리, 기댓값 등에 대한 75 개의 문제와 15 개의 보험 수학 문제가 실려 있다. 또 세 번째 개정판에는 1718 년부터 1756 년까지 다루었던 확률 문제를 망라하였고, 이전 문제를 확장한 새로운 문제와 새로운 주석, 정리를 추가하여 이항분포, 대기시간 분포 등 74 개의 확률 문제와 33 개의 보험 수학 문제가 실려 있다.

(4) 라플라스(Laplace, P. S. ; 1749~1827)

확률 이론에 지대한 공헌을 하여 현대 확률론의 아버지로 불리는 라플라스는 1812 년에 “확률의 해석적 이론”을 발표하였다. 그는 이 책에서 베르누이 정리의 새로운 증명, 득점 문제, 뷔퐁의 바늘문제, 상트페테르부르크의 역설 등 여러 확률 문제의 해법을 제시하였고, 수학적 확률을 정의하였다.

라플라스는 확률의 추론에 대한 4 권의 저서를 남겼다. 이 중에서 18 세기 수학적 확률론의 역사에서 가장 의미 있는 연구로는 1773 년의 “원인 사건에 대한 고찰”, 1780 년의 “확률에 대한 고찰”을 손꼽는다.

라플라스는 1774 년부터 1786 년까지의 일련의 논문에서 이항분포나 기하분포의 정규근사를 탐구하였고, 이 논문들의 대부분은 “확률의 해석적 이론(1812)”에 수록되어 있다. 라플라스의 가장 중요한 업적은 1810 년에 발표한 중심 극한정리이다. 그는 드무아브르의 극한정리를 일반화하여 이항분포의 극한분포는 정규분포에 따른다는 것을 증명하였다. 이 정리는 현재 드무아브르 - 라플라스 정리로서 알려져 있다.

(5) 가우스(Gauss, K. F. ; 1777~1855)

가우스는 1809 년에 “태양 주위로 움직이는 행성 운동 이론”이라는 논문에서 행성의 궤도를 확률적 개념으로 접근하여 처음으로 정규분포를 제시하였다. 그는 행성궤도에 관한 미지의 방정식과 관측값 사이의 차를 오차 E 라고 하면 이에 대한 확률값 $f(\epsilon)$ 은 곡선을 이루며 다음을 만족하는 것을 밝혔다.

$$f(\epsilon E) = \frac{h}{\sqrt{\pi}} e^{-\frac{\epsilon^2}{h^2}}$$

($h > 0$, e 는 2.71828... 인 무리수)

여기서 h 는 관측값의 정도에 대한 척도로써 이 곡선을 정규곡선이라고 부른다.

(6) 콜모고로프(Kolmogorov, A. N. ; 1903~1987)

20 세기 초까지는 모든 확률 교재에서 라플라스의 고전적 확률의 정의를 사용하였다. 이 정의는 같은 정도로 일어나는 유한 표본공간에서 정의된 확률이기 때문에 조건이 만족되지 않는 상황에서는 적용할 수가 없다.

확률 이론에도 공리적 방법에 의한 새로운 논리 구조의 필요성이 대두되기 시작하였다. 20 세기 초반 수학 분야에서는 유클리드의 기하학 원론에서의 공리, 로바체프스키의 공리(평행선 공준)를 이어 받아 힐베르트의 공리와 페아노의 공리가 발표되던 시기이다. 확률 이론의 공리화는 베른슈타인(Bernstein, S. N. ; 1880~1968)에 의해 최초로 도입되었고, 그는 1917 년에 발표한 논문 “확률 이론의 공리적 기초에 관한 에세이”에서 확률을 3가지 공리, 즉 확률의 비교 공리, 배반사건 공리, 결합사건 공리로 구체화하였다. 그 후 콜모고로프는 1933 년에 발표한 책 “Grundbegriffe der Wahrscheinlichkeitsrechnung”에서 집합론의 개념을 이용한 공리로 확률을 정의하였다.

공리적 확률의 정의는 다음과 같다.

‘표본공간 S 의 사건 A 에 대하여

공리 1. $0 \leq P(A) \leq 1$

공리 2. $P(S) = 1$

공리 3. $A_1, A_2, \dots, A_n, \dots$ 이 서로 배반사건일 때

$$P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = \sum_{i=1}^{\infty} P(A_i)$$

를 만족할 때, $P(A)$ 를 사건 A 의 확률이라고 한다.’

콜모고로프의 공리적 확률의 정의로 확률론은 현대의 수학 원리로 발달하게 되었고 많은 분야에서 응용되었다. 1654 년 파스칼과 페르마가 제시했던 확률 개념이 1933 년 콜모고로프의 공리적 확률의 정의로 명확하게 정의되기까지 무려 400 년 이상이 걸렸다.

공리적 확률의 정의로부터 확률은 다음 성질을 만족한다.

i) 공사건 ϕ 에 대하여

$$P(\phi) = 0$$

ii) 두 사건 A, B 에 대하여

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

iii) 사건 A 의 여사건 A^C 에 대하여

$$P(A^C) = 1 - P(A)$$

iv) 두 사건 A, B 에 대하여 A 가 B 의 부분집합이면

$$P(A) \leq P(B)$$

v) 부울의 부등식(Boole's inequality)

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

vi) 본페로니 부등식(Bonferroni's inequality)

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) \leq 1 - \sum_{i=1}^{\infty} P(A_i^C)$$

[증명] i) 표본공간 S 에서 $S \cap \phi = \phi$ 이므로 S 와 ϕ 은 서로 배반사건이다.

따라서 공리 3에 의하여 $P(S \cup \phi) = P(S) + P(\phi)$

그런데 $S \cup \phi = S$ 이고, 공리 2에 의하여

$$P(S \cup \phi) = P(S) = 1$$

따라서 $1 = 1 + P(\phi)$ 이므로 $P(\phi) = 0$

iv) $A \subset B$ 이므로 $B = A \cup (B - A)$, $A \cap (B - A) = \phi$

따라서 공리 3에 의하여 $P(B) = P(A) + P(B - A)$

그런데 공리 1에 의하여 $P(B - A) \geq 0$

따라서 $P(B) - P(A) \geq 0$ 이므로 $P(A) \leq P(B)$

< 창의 사고력 UP >

이탈리아의 수학자 파촐리(Pacioli, L. ; 1445~1517)는 다음과 같은 문제를 해결하려고 하였다. 물음에 답하여라.

갑과 을이 같은 금액을 걸고 게임을 하고 있다. 한 게임을 승리하면 1 점을 얻고, 먼저 5 점을 얻는 사람이 건 돈을 모두 가지기로 하였다. 이때 갑과 을이 각 게임에서 이길 확률은 서로 같으며, 비기는 경우는 없다고 한다.

1. 갑이 4 점을 얻고 을이 3 점을 얻은 상황에서 게임을 중단해야 했을 때, 돈을 어떻게 분배하는 것이 공정한지 구하고, 그 이유를 설명하여라.

2. 갑이 3 점을 얻고 을이 2 점을 얻은 상황에서 게임을 중단

해야 했을 때, 돈을 어떻게 분배하는 것이 공정한지 구하고, 그 이유를 설명하여라.

[해설] 1. 점수를 얻는 사람을 순서대로 적으면

갑이 먼저 5 점을 얻는 경우는 갑, 을-갑이므로 그 확률은

$$\frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{3}{4}$$

을이 먼저 5 점을 얻는 경우는 을-을이므로 그 확률은

$$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

따라서 갑, 을이 이길 확률은 각각 $\frac{3}{4}$, $\frac{1}{4}$ 이므로 돈은 3 : 1로 분배하는 것이 공정하다.

[해설] 2. 갑이 먼저 5 점을 얻는 경우는 갑-갑, 갑-을-갑, 을-갑-갑, 갑-을-을-갑, 을-갑-을-갑, 을-을-갑-갑이므로 그 확률은

$$\begin{aligned} & \frac{1}{2} \times \frac{1}{2} + 2\left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) + 3\left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) \\ &= \frac{11}{16} \end{aligned}$$

을이 먼저 5 점을 얻는 경우는 을-을-을, 갑-을-을-을, 을-갑-을-을, 을-을-갑-을이므로 그 확률은

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} + 3\left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) = \frac{5}{16}$$

따라서 갑, 을이 이길 확률은 각각 $\frac{11}{16}$, $\frac{5}{16}$ 이므로 돈은 11 : 5로 분배하는 것이 공정하다.

< 세상 곳곳에는 '지프의 법칙'이 숨어 있다. >

미국의 언어학자 지프(Zipf, G. K. ; 1902~1950)는 영어에서 자주 사용하는 단어의 빈도를 조사하다가 흥미로운 법칙을 발견하였다.

그는 영어 문장에서 가장 많이 사용하는 단어는 'the'로서 이 빈도를 1로 볼 때, 그 다음으로 가장 많이 사용하는 단어 'of'의 빈도는 약 $\frac{1}{2}$ 이고, 세 번째로 많이 사용하는

단어 'and'의 빈도는 약 $\frac{1}{3}$ 임을 알았다.

이 관계를 일반화하면 '가장 많이 나오는 단어의 빈도를 1이라고 할 때, r 번째로 많이 나오는 단어의 빈도는 근사적으로 $\frac{1}{r}$ 이다.'가 되는데 이것을 '지프의 법칙'이라고 한다.

지프의 법칙이 놀라운 건 이와 같은 관계가 단어 사용 빈도뿐만 아니라 도시의 인구수, 성씨 비율 등에서도 잘 들어맞는다는 것이다.

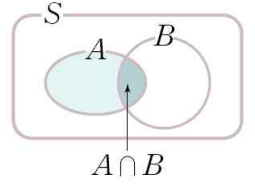
예를 들어 미국에서 가장 인구가 많은 도시는 뉴욕인데 두 번째 도시인 로스앤젤레스의 인구는 뉴욕의 약 $\frac{1}{2.2}$, 세 번째 도시인 시카고의 인구는 뉴욕의 약 $\frac{1}{2.9}$,

	인구(명)	비율
뉴욕	831만	1
로스앤젤레스	383만	$\frac{1}{2.2}$
시카고	283만	$\frac{1}{2.9}$
휴스턴	220만	$\frac{1}{3.8}$
피닉스	155만	$\frac{1}{5.4}$

네 번째 도시인 휴스턴의 인구는 뉴욕의 약 $\frac{1}{3.8}$, 다섯 번째 도시인 피닉스의 인구는 뉴욕의 약 $\frac{1}{5.4}$ 이라고 한다.

2. 조건부확률

1 조건부 확률



☆ A가 일어났을 때, B가 일어날 확률

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

☑ 사건 A를 새로운 표본공간으로 생각하고 A 안에서 $A \cap B$ 가 일어날 확률을 뜻한다.

1. 조건부확률 : 사건 A가 일어났다는 조건 아래에서 사건 B가 일어날 확률

$$\Rightarrow P(B|A)$$

2. 사건 A가 일어났을 때, 사건 B가 일어날 조건부확률

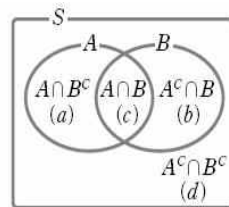
$$P(B|A) = \frac{n(A \cap B)}{n(A)} = \frac{P(A \cap B)}{P(A)}$$

(단, $P(A) > 0$)

(1) $P(B|A)$ 는 A가 먼저 일어났을 때,

$A \cap B$ 가 일어날 확률

$$\begin{aligned} \Rightarrow P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{n(A \cap B)}{n(A)} = \frac{c}{a+c} \end{aligned}$$



(2) $P(A|B)$ 는 B가 먼저 일어났을 때,

$A \cap B$ 가 일어날 확률

$$\begin{aligned} \Rightarrow P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{n(A \cap B)}{n(B)} = \frac{c}{b+c} \end{aligned}$$

예) A 주머니 : 빨간 공 3개, 파란 공 2개

B 주머니 : 빨간 공 2개, 파란 공 2개

동전을 던져 그 결과에 따라 한 주머니를 선택한 후, 그 주머니에서 임의로 공을 한 개 꺼냈더니 파란 공이었을 때, 그 공이 A 주머니에서 나온 공일 확률 [풀이] 동전 → 파란공

$$\begin{aligned} A &: \frac{1}{2} \times \frac{2}{5} = \frac{1}{5} = \left(\frac{2}{10} \right) \\ B &: \frac{1}{2} \times \frac{2}{4} = \frac{1}{4} = \left(\frac{5}{10} \right) + \left(\frac{7}{10} \right) \therefore \frac{2}{7} \end{aligned}$$

4. 확률의 곱셈 정리(종속사건의 곱셈정리)

$P(A) > 0, P(B) > 0$ 일 때,

A, B 가 동시에 일어날 확률

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B|A) \\ &= P(B) \cdot P(A|B) \end{aligned}$$

5. 조건부확률의 성질

(1) $0 \leq P(B|A) \leq 1$

(2) $P(S|A) = 1, P(A|S) = P(A)$

(3) $P(\phi|A) = 0$

(4) B_1, B_2 가 배반사건

$$\Rightarrow P(B_1 \cup B_2 | A) = P(B_1 | A) + P(B_2 | A)$$

(5) $P(B|A) + P(B^C|A) = 1$

$$P(B|A^C) + P(B^C|A^C) = 1$$

6. 전확률의 정리

(1) 두 사건 A, B 에 대하여

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^C \cap B) \\ &= P(A) \cdot P(B|A) + P(A^C) \cdot P(B|A^C) \end{aligned}$$

(2) 표본공간 S 가 서로 배반인 사건 A_1, A_2, \dots, A_n 의 합집합으로 표시될 때,

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B|A_i)$$

(3) [베이즈의 법칙] 표본공간 S 가 서로 배반인 사건 A_1, A_2, \dots, A_n 의 합집합으로 표시될 때,

$$\begin{aligned} P(A_k|B) &= \frac{P(A_k \cap B)}{P(B)} \\ &= \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^n P(A_i) \cdot P(B|A_i)} \end{aligned}$$

2 사건 독립과 종속

1. 독립사건과 종속사건

두 사건 A, B 에 대하여

(1) 독립사건 : 한 사건이 일어나는 것이 다른 사건이 일어날 확률에 아무런 영향을 주지 않을 때
(= 복원추출) $\Leftrightarrow A$ 와 B 는 서로 독립

(2) 종속사건 : 서로 독립이 아닐 때 (= 비복원추출)
 $\Leftrightarrow A$ 와 B 는 서로 종속

2. A 와 B 가 서로 독립사건

(단, $P(A) > 0, P(B) > 0$)

$$\begin{aligned} \Leftrightarrow P(A \cap B) &= P(A) \cdot P(B) \\ &\hookrightarrow (\text{독립사건의 곱셈 정리}) \end{aligned}$$

$$\Leftrightarrow P(B|A) = P(B) = P(B|A^C)$$

$$\Leftrightarrow P(A|B) = P(A) = P(A|B^C)$$

3.(1) A, B 가 독립 $\Leftrightarrow A, B^C$ 도 독립

$$\Leftrightarrow A^C, B \text{도 독립} \Leftrightarrow A^C, B^C \text{도 독립}$$

(2) $P(A) \neq 0, P(B) \neq 0$ 일 때,

A, B 가 독립 $\Rightarrow A, B$ 는 배반이 아니다.

A, B 가 배반 $\Rightarrow A, B$ 는 독립이 아니다(종속)
($\because P(A \cap B) = 0$)

(3) 배반사건의 확률 \Leftrightarrow 덧셈

종속 or 독립사건의 확률 \Leftrightarrow 곱셈

4. 세 사건 A, B, C 가 서로 독립

$$\Leftrightarrow \textcircled{1} P(A \cap B) = P(A) \cdot P(B)$$

$$\textcircled{2} P(B \cap C) = P(B) \cdot P(C)$$

$$\textcircled{3} P(C \cap A) = P(C) \cdot P(A)$$

$$\textcircled{4} P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

3 독립시행의 확률

1. 독립시행

- (1) 매회 영향 없이 반복되는 시행
- (2) 매회 일어나는 사건이 서로 독립
- (3) 매회 일어날 확률이 일정한 시행
- (4) 복원추출
- (5) 주사위, 동전, 명중률, 정답률 등

2. 독립시행의 정리(베르누이의 정리) ⇐ 이항정리

매회의 시행에서 사건 A 가 일어날 확률이 p 로 일정할 때, n 회의 독립시행에서 사건 A 가 r 번 일어날 확률

$$P_r(A) = {}_n C_r p^r (1-p)^{n-r}$$

$$= {}_n C_r p^r q^{n-r}$$

(단, $p+q=1$, $r=0, 1, 2, \dots, n$)

예 한 개의 주사위를

4 회 던질 때,
1 의 눈이 2 회
나올 확률은

$${}_4 C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2$$

1회	2회	3회	4회	확률
○	○	×	×	$\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$
○	×	○	×	$\frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6}$
○	×	×	○	$\frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$
×	○	○	×	$\frac{5}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6}$
×	○	×	○	$\frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$
×	×	○	○	$\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$

(1의 눈: ○, 1 이외의 눈: ×)

3. 1 회 시행에서 사건 A, B, C 가 일어날 확률을 각각 a, b, c 라 하자. n 회의 독립시행에서 A 가 p 회, B 가 q 회, C 가 r 회 일어날 확률

$$\frac{n!}{p! q! r!} a^p b^q c^r$$

(단, $p+q+r=n$, $a+b+c=1$)

4. 독립시행의 확률 계산

- (1) 주사위를 A 는 m 번, B 는 n 번 던질 때,
1 의 눈이 나온 횟수의 합이 k 번일 확률

$$\sum_{r=0}^k {}_m C_r \left(\frac{1}{6}\right)^r \left(\frac{5}{6}\right)^{m-r} \times {}_n C_{k-r} \left(\frac{1}{6}\right)^{k-r} \left(\frac{5}{6}\right)^{n-k+r}$$

$$= \sum_{r=0}^k \{ {}_m C_r \times {}_n C_{k-r} \} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{(m+n)-k}$$

$$= \sum_{r=0}^k {}_{m+n} C_k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{(m+n)-k}$$

⇨ A & B 가 $(m+n)$ 번 중 k 번 1 의 눈 출현

- (2) 한 개의 주사위를 n 번 던져 나온 눈의 수 중 최소인 것을 확률변수 X 라 할 때, $P(X=k)$ 는?

⇨ 한 번의 시행에서 k 보다 작은 눈이 나올 확률 :

$$\frac{k-1}{6}$$

한 번의 시행에서 k 이상의 눈이 나올 확률 :

$$1 - \frac{k-1}{6} = \frac{7-k}{6}$$

n 번의 시행에서 k 이상의 눈이 나올 확률 :

$$P(X \geq k) = \left(\frac{7-k}{6}\right)^n$$

$$\therefore P(X=k) = P(X \geq k) - P(X \geq k+1)$$

$$= \left(\frac{7-k}{6}\right)^n - \left(\frac{7-k-1}{6}\right)^n$$

$$= \left(\frac{7-k}{6}\right)^n - \left(\frac{6-k}{6}\right)^n$$

< 하디(Hardy)-바이베르크(Weinberg)의 법칙 >

1. 생물의 특징을 나타내는 어떤 형질에 우성인 유전자 A 와 열성인 유전자 a 가 있을 때, 잡종 Aa 끼리 교배하면 오른쪽 표와 같이 우성 형질(AA, Aa, Aa)과 열성 형질(aa)이 3 : 1의 비로 분리되어 나타나는 것처럼 보인다.

여 \ 남	A	a
A	AA(우성)	Aa(우성)
a	Aa(우성)	aa(열성)

2. 우성 유전자(A)의 비율을 p , 열성 유전자(a)의 비율을 q 라고 하면 $p + q = 1$ 이다. 이 유전자들이 무작위로 교배된다면, 유전형이 AA 가 될 확률은 p^2 , aa 가 될 확률은 q^2 , 그리고 Aa 가 될 확률은 $2pq$ 이다. 이때, 우성 유전자(A)와 열성 유전자(a)의 비율은

(1) 우성 유전자(A)의 비율 :

$$p^2 \times 1 + 2pq \times \frac{1}{2} + q^2 \times 0 = p(p + q) = p$$

(2) 열성 유전자(a)의 비율 :

$$p^2 \times 0 + 2pq \times \frac{1}{2} + q^2 \times 1 = q(p + q) = q$$

즉, 우성 유전자와 열성 유전자의 비율에는 변화가 없음을 알 수 있다.

< 심프슨의 역설(Simpson's Paradox) >

1. 심프슨의 역설

동일하지 않은 가중치를 적용함에 따라 부분에 대한 분석 결과와 전체에 대한 분석 결과가 일치하지 않는 현상.

2. 어느 대학의 남녀별 합격자 수에 대한 분석

		응시자 수(명)	합격자 수(명)	합격률(%)
A 학과	여자	30	15	50
	남자	70	35	50
B 학과	여자	70	21	30
	남자	30	9	30
전체	여자	100	36	36
	남자	100	44	44

(1) A 학과와 B 학과의 남녀별 합격률은 모두 같다.

(2) A 학과와 B 학과의 여자의 합격률의 평균은

$$\frac{50 + 30}{2} = 40(\%) \text{이고, 마찬가지로 남자의 합격률}$$

의 평균도 40(%)로 같다.

(3) 하지만, 전체적으로는 여자의 합격률은 36(%)이고 남자의 합격률은 44(%)로 남녀간 차이가 있다.

(4) 차이가 생기는 이유 : 남자들이 합격률이 높은 A 학과에 여자보다 더 많이 지원했고, 여자들은 합격률이 낮은 B 학과에 남자보다 더 많이 지원했기 때문이다.

3. 어느 대학의 A 학과와 B 학과에 지망한 수험생의 자료가 다음 표와 같다고 하자.

구분	여학생			남학생		
	지원자 수	합격자 수	합격률	지원자 수	합격자 수	합격률
A 학과	50	20	40 %	30	10	약 33 %
B 학과	40	30	75 %	70	50	약 71 %
A, B 학과	90	50	약 56 %	100	60	60 %

이 표에서 보면 A 학과의 경우 여학생의 합격률은 40(%)이고, 남학생의 합격률은 약 33(%)이다.

또 B 학과의 경우 여학생의 합격률은 75(%)이고 남학생의 합격률은 약 71(%)이다. 두 학과 모두 여학생의 합격률이 높다. 그러나 두 학과 A, B를 합쳐서 보면 여학생의 합격률은 약 56(%)이고, 남학생의 합격률은 60(%)로 남학생의 합격률이 높다.

이것은 우리의 직관과는 상충되는 것으로써 합격률이 높은 B 학과의 남학생 수가 많아서 일어나는 현상이다.

< 확률의 역사 >

확률의 역사는 중세 시대 도박에서 유래하였다고 볼 수 있다. 이탈리아의 파촐리(Pacioli, L. ; 1450? ~ 1520)는 그의 저서 <산술, 기하, 비 및 비례 대전>에서 'Problem of point'라고 불리는 도박 문제를 소개하여 확률의 시초로서 평가받고 있다. 이 문제는 능력이 같은 두 도박사가 도박 경기 도중에 경기를 중단해야 할 경우 두 도박사가 건 돈을 어떻게 분배해야 하는가에 대한 문제이다.

이러한 문제는 여러 사람들에 의하여 오랫동안 연구되어 왔지만 해결에 큰 발전을 가져온 사람은 프랑스의 파스칼(Pascal, B. ; 1623~1662)과 페르마(Fermat, P. ; 1601~1665)에 의해서이다.

이들로부터 시작된 확률론의 연구는 스위스의 베르누이(Bernoulli, J. ; 1667~1748), 프랑스의 드 무아브르(De Moivre, A. ; 1667~1754)에 의하여 급속한 발전이 이루어졌다.

< 라플라스의 확률의 정의 >

19세기 초기에 프랑스의 라플라스(Laplace, P.S.; 1749~1827)는 확률에 관한 저서 <확률의 해석적 이론>을 통하여 고전적인 확률론의 체계를 확립하였다.

라플라스는 다음과 같이 확률을 정의하였다.

‘어떤 시행에서 일어날 수 있는 경우의 수가 n 가지이고 어느 두 경우도 동시에 일어나지 않으며 각 경우가 일어나는 정도가 동등할 때, 사건 A 가 일어나는 경우의 수를 r 가 지라고 하면 사건 A 가 일어날 확률은 $P(A) = \frac{r}{n}$ 이다.’

이와 같이 확률을 엄밀하게 수학적으로 정의함으로써 그 당시에 상대도수의 극한으로 정의된 확률의 개념과 많은 논쟁을 일으켰다. 또한, 라플라스의 확률의 정의는 같은 정도로 확실한 근원사건으로 이루어진 표본공간에서만 계산이 가능한 단점을 갖고 있었다.

따라서 이 정의에 대하여 많은 비판이 제기되었다.

첫 번째 일어날 수 있는 모든 경우의 수가 무한인 경우에도 확률을 생각해야 되는 경우가 있는데 이 경우 위의 정의를 적용할 수 없다.

두 번째 각 경우가 일어나는 정도가 동등하다는 것은 확률이 같다는 뜻인데 아직 정의되지 않은 이러한 용어를 확률의 정의에 사용하는 것은 적합하지 않다는 것이다.

이 선형적 확률(수학적 확률)에 대하여 경험적 확률(통계적 확률)이 논쟁을 벌이며 20 세기에 이르게 된다.

< 콜모고로프의 확률의 정의 >

이 밖에도 확률에 대한 여러 가지 해석과 견해들이 있었으나 한결같이 그 어떤 것도 보편적인 설득력을 가지지 못했다. 결국 수학자들이 동시에 후퇴하면서 모두의 승리를 선언하는 것으로 끝났다.

‘지금까지 나온 합리적인 확률의 정의만으로도 결국 확률이 가지고 있는 형식과 성질을 충분히 알 수 있으므로 확률은 그런 형식과 성질(확률의 공리)을 만족하는 어떤 것이라고 정의하자.’고 마침내 협정을 맺은 것이다.

이 논쟁의 종지부를 찍은 사람은 바로 힐베르트(Hilbert, D.; 1862~1943)의 공리적 이론에 영향을 받은 러시아의 콜모고로프(Kolmogorov, A.; 1903~1987)이었다. 콜모고로프에 의하여 확립된 공리론적 확률의 정의는 전 표본공간의 측도가 1 인 르베그(Lebesgue, H.L.; 1875~1941) 측도로서 정의되며, 확률의 계산은 르베그 적분에 의하여 간단히 해결할 수 있다.

콜모고로프에 의하여 도입된 공리론적 확률의 정의는 확률의 기본 성질을 규명한 것으로 다음과 같다.

[공리 1] $P(S) = 1, P(\emptyset) = 0$

[공리 2] 어떤 사건 A 에 대해서도 $0 \leq P(A) \leq 1$

[공리 3] A, B 가 배반사건이면

$$P(A \cup B) = P(A) + P(B)$$

따라서 이러한 정의는 선형적(수학적) 정의와 경험적(통계적) 정의를 포함하게 되었고, 또한 확률의 고전적 정의(라플라스의 정의)를 공리적 정의(콜모고로프의 정의)의 한 특수한 경우로 만들어 버렸다.

< 조건부 확률과 베이즈 정리 >

사건 A 가 일어났다는 조건 하에서 사건 B 가 일어날 확률을 조건부확률이라 하고, $P(B|A)$ 로 나타낸다. 조건부확률 $P(B|A)$ 는 $P(A) > 0$ 일 경우 $P(A)$ 에 대한 $P(A \cap B)$ 의 비율로 정의된다. 즉,

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

표본공간 S 의 사건 $A_1, A_2, A_3, \dots, A_n$ 에 대하여

$$(1) A_i \cap A_j = \emptyset \quad (i \neq j)$$

$$(2) A_1 \cup A_2 \cup \dots \cup A_n = S$$

이면 사건 $A_1, A_2, A_3, \dots, A_n$ 을 표본공간 S 의 분할(partition)이라 한다.

[전확률정리, Total Probability Law]

사건 $A_1, A_2, A_3, \dots, A_n$ 이 표본공간 S 의 한 분할이고, $P(A_i) > 0 \quad (i = 1, 2, \dots, n)$ 이면 임의의 사건 B 에 대하여

$$P(B) = P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n)$$

$$= \sum_{i=1}^n P(A_i)P(B|A_i)$$

전확률정리는 주어진 사건 B 의 확률을 계산할 때, 그 사건의 원인을 여러 가지로 나누어서 각 원인에 대한 조건부확률 $P(B|A_i)$ 와 그 원인이 되는 확률 $P(A_i)$ 의 가중합으로 구할 수 있다는 것을 보여준다.

전확률정리를 이용하여 다음의 베이즈정리를 얻을 수 있다.

[베이즈정리(Bayes theorem)]

사건 A_1, A_2, \dots, A_n 이 표본공간의 분할이고

$$P(A_i) > 0, P(B) > 0 \text{ 이면}$$

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

베이즈정리는 형태는 간단하지만 이를 이용하여 다양한 설명을 할 수 있다. 먼저 $A_i \quad (i = 1, 2, \dots, n)$ 가 어떤 사건 B 에 대한 원인들이라면 $P(A_i|B)$ 는 B 라는 사건이 발생하였을 때 그 원인의 확률을 의미한다. 또, $P(A_i)$ 를 어떤 사건이 일어나지 않은 상태에서의 주관적인 신뢰도라 하면 $P(A_i|B)$ 는 사건이 발생한 후 그 사건에 의하여 변화된 신뢰도를 의미할 수 있다.

이와 같은 이유에서 $P(A_i)$ 를 사전적 확률(prior probability), $P(A_i|B)$ 를 사후적 확률(posterior probability)이라고 부른다.

< 검색 엔진의 정확성은 ‘베이즈 정리’에 달려 있다. >

18세기 영국의 수학자 베이즈(Bayes, T.; 1702~1761)는 ‘베이즈 정리’라는 조건부확률 이론을 발표하였다. 이 이론의 핵심은 확률 값이 항상 고정불변한 것이 아니라 기존의 통계 자료를 적용하면 바뀐다는 것이다.

예를 들어 어느 환자가 갑상선암을 진단하는 초음파 검사에서 양성 반응이 나오면 실제로 자신이 암에 걸렸을 확률이 초음파 검사 장비의 정확도인 0.95라고 생각하는 경향이 있다. 그러나 기존의 통계 자료에 의하면 우리나라 국민의 갑상선암 발생률이 0.01 정도밖에 안 되고 암에 걸리지 않았음에도 양성으로 잘못 판정할 확률이 0.01이라고 하면 비록 검사에서 양성 반응이 나왔다고 해도 실제로 암에 걸렸을 확률은 0.49에 불과하다.

베이즈 정리의 장점은 기존의 통계 자료가 많을수록 확률 값이 정확해지고, 자료가 바뀌면 확률 값도 자동적으로 수정된다는 ‘자가 수정’이론에 있다. 이 이론은 인터넷 검색 엔진에서 중요한 수학적 기반을 제공한다. 최근에 각종 포털 사이트가 사용자의 요구에 최대한 부합하는 검색 결과를 제공하고, 초기에는 오답 투성이었던 지식 검색이 갈수록 정확해지는 이유가 여기에 있다.

세계 최대의 컴퓨터 회사인 마이크로소프트가 개발하고 있는 ‘컨텍스트(context) 서버’는 사람들의 일상 습관을 분석하여 다양한 상황에서 작업의 순서를 정해주고, 생활을 도와주는 전자 비서 개념의 소프트웨어이다. 소프트웨어 개발자는 “나는 베이즈 이론을 이용하여 작업의 우선 순위를 정한다. 모든 것을 알 수는 없는 불확실성의 세계에서 지적 활동의 토대는 바로 확률이라고 생각한다.”고 말했다.

< 스포츠 과학의 천령, 스포츠 애널리스트 >

최근 스포츠 과학을 기반으로 한 경기 분석의 중요성이 강조되면서 전문적인 전력 분석가, 스포츠 해설가 등의 수요가 점점 커지고 있다.

스포츠 애널리스트(Sports Analyst)란 스포츠 경기 자료를 체계적으로 기록, 수집, 분석, 평가하여 게임 전략에 활용하거나 선수들의 경기력 향상에 도움을 주는 전문가를 말한다.

우리나라에 처음으로 이름을 알린 스포츠 애널리스트는 2002년 한일 월드컵 4강 신화의 주역인 히딩크 감독의 든든한 조력자였던 고트비 코치였다. 그는 우리 선수들의 움직임과 상대 팀의 경기 비디오 자료를 철저히 분석하여 선수 기용 및 작전 전략에 활용하였다. 축구 감독 중에는 스포츠 애널리스트 출신이 꽤 있다. 이란 축구 감독이 된 고트비 감독뿐만 아니라 스페인 레알 마드리드의 무리뉴 감독, 우리나라의 대표 팀 코치였던 아디 수석 코치도 스포츠 애널리스트 출신이다.

스포츠 애널리스트들이 가장 많이 진출해 있는 종목은 야구이다. 야구는 기록의 스포츠라고 불릴 정도로 타율, 방어율 등 공격과 수비의 모든 기록이 선수 기용 및 전략에 활용된다. 다년간 쌓인 야구 데이터를 이용하여 선수들을 평가하고 연봉 협상의 기초 자료를 제공하는 전문가를 세이버메트릭션(Sabermetrician)이라고 한다. 이들이 분석한 정밀한 통계 자료를 바탕으로 데이터 야구를 구사하는 전략 덕분에 우리나라의 프로야구는 질적인 향상과 함께 관객 동원에 성공하였다고 할 수 있다.

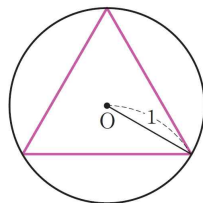
스포츠 애널리스트는 스포츠 분석을 수행하는 전문 인력으로 축구, 야구, 무용 등의 다양한 종목에서 스포츠 언론 전문가(평론가, 기자), 스포츠 정보 전문가(스카우터), 전력분석관으로 진출할 수 있다. 스포츠 애널리스트 자격증을 취득하기 위해서는 스포츠 심리, 분석 프로그램, 통계 등의 교육과정을 이수해야 한다.

< 논리를 키우는 수학 >

다음 문장을 읽고, 물음에 답하여라.

프랑스의 수학자 베르트랑(Bertrand, J. L. F. ; 1822~1900)은 그의 저서 “확률론”에서 다음과 같은 문제를 제시하였다.

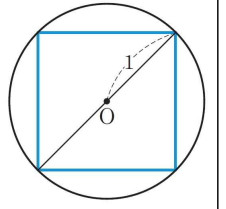
‘반지름의 길이가 1인 원에서 임의의 현을 그을 때, 그 현의 길이가 내접 정삼각형의 한 변의 길이 $\sqrt{3}$ 보다 클 확률을 구하여라.’



이 문제는 현을 긋는 방법에 따라 확률이 달라지므로 ‘베르트랑의 역설’이라고 불린다.

베르트랑의 역설 문제는 다음과 같이 변형될 수 있다.

‘반지름의 길이가 1인 원에서 임의의 현을 그을 때, 그 현의 길이가 내접 정사각형의 한 변의 길이 $\sqrt{2}$ 보다 클 확률을 구하여라.’



1. 변형된 베르트랑의 문제에서 다음 각 경우의 표본공간을 구하여라.

- (1) 좌표평면 위의 원점에 원의 중심을 놓고 내접 정사각형을 세로 변과 y 축이 평행하게 놓는다. 내접 정사각형의 세로 변과 평행하게 현을 그을 때, 현과 x 축의 교점의 x 좌표
- (2) 내접 정사각형의 한 꼭짓점에서 현을 그을 때, 그 꼭짓점에서 그은 원의 접선과 현이 이루는 각 x°
- (3) 좌표평면 위의 원점에 원의 중심을 놓고 내접 정사각형의 가로 변, 세로 변과 각각 평행하게 현을 그을 때, 두 현의 교점의 좌표 (x, y)

2. 위의 각 경우의 확률을 구하고, 변형된 베르트랑의 문제에 대한 답을 말하여라. 답을 말할 수 없다면 그 이유는 무엇인지 설명하여라.

- [해설] 1. (1) $\{x \mid -1 < x < 1, x \text{는 실수}\}$
 (2) $\{x \mid 0 < x < 180, x \text{는 실수}\}$
 (3) $\{(x, y) \mid x^2 + y^2 \leq 1, x, y \text{는 실수}\}$

[해설] 2. 각 경우의 확률은 (1) $\frac{\sqrt{2}}{2}$ (2) $\frac{1}{2}$ (3) $\frac{2}{\pi}$

반지름의 길이가 1인 원에서 임의의 현을 그을 때, 현을 긋는 방법에 따라 표본공간이 다르므로 확률이 다르다. 즉, 임의의 현을 긋는다는 것이 하나의 시행이 아니라 ‘임의’를 어떻게 정의하느냐에 따라 여러 시행이 나타난다.

베르트랑의 문제는 다음과 같이 현을 긋는 세 가지의 방법에 따라 확률이 달라지기 때문에 베르트랑의 역설이라고 불린다.

- (1) 원의 반지름에 수직인 현을 긋는 경우(삼각형의 한 변과 평행한 현을 긋는 경우)
- (2) 내접 정삼각형의 한 꼭짓점을 지나는 현을 긋는 경우
- (3) 내접 정삼각형에 내접하는 원을 지나는 현을 긋는 경우

III. 통 계

0. 도수분포

1 도수분포

1. 도수분포 : 키, 몸무게와 같은 변량 x_i 에
인원수와 같은 도수 f_i 가 대응되는 분포

2. 평균과 분산

$$(1) \text{ 평균 } m = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{N}$$

$$= \frac{1}{N} \sum_{i=1}^n x_i f_i \quad \left(\text{단, } N = \sum_{i=1}^n f_i \right)$$

☑ (평균) = {(계급값) × (도수)}의 합

$$(2) \text{ 분산 } \sigma^2 = \frac{1}{N} \sum_{i=1}^n (x_i - m)^2 f_i$$

$$= \frac{1}{N} \sum_{i=1}^n x_i^2 f_i - m^2$$

☑ (분산) = {(편차)의 제곱}의 평균
= {(제곱}의 평균} - {(평균)의 제곱}

(3) 표준편차 σ : 분산 (σ^2)의 양의 제곱근

☑ (표준편차) = $\sqrt{\text{분산}}$

2 가평균을 이용한 평균과 표준편차

☑ 가평균 : 평균에 가깝게 택한 값

① 중앙값 ② 도수가 가장 많은 계급값(최빈값)

가평균 x_0 , 계급의 크기를 c 라 할 때,
계급값 x_i ($i = 1, 2, \dots, n$) 에 대하여,
$$u_i = \frac{x_i - x_0}{c} = \frac{(\text{계급값}) - (\text{가평균})}{(\text{계급의 크기})}$$

1. 평균

$$m = x_0 + c \bar{u} \quad \left(\text{단, } \bar{u} = \frac{1}{N} \sum_{i=1}^n u_i f_i \right)$$

☑ (평균) = (가평균) + (과부족의 평균)

↳ {(계급값) - (가평균)}의 평균

2. 분산

$$\sigma^2 = c^2 \left(\frac{1}{N} \sum_{i=1}^n u_i^2 f_i - \bar{u}^2 \right)$$

$$\left(\text{단, } N = \sum_{i=1}^n f_i, \bar{u} = \frac{1}{N} \sum_{i=1}^n u_i f_i \right)$$

3. 표준편차 = $\sqrt{\text{분산}}$

3 표준편차의 의미

1. 표준편차(분산 or 산포도)가 작을수록

(1) 자료가 고르게 분포되어 있다.

(2) 그래프의 폭이 좁다.

(3) 변량이 평균을 중심으로 집중되어 있다.

4 도수분포와 이산확률분포

1. 도수분포표에서

(1) 변량 $X : x_1, x_2, \dots, x_n$

변량	도수	상대도수
x_1	f_1	$\frac{f_1}{N}$
x_2	f_2	$\frac{f_2}{N}$
\vdots	\vdots	\vdots
x_n	f_n	$\frac{f_n}{N}$
합계	N	1

(2) 상대도수 $\frac{f_i}{N} : X$ 가 x_i

의 값을 가질 확률, 즉

$$P(X = x_i) = \frac{f_i}{N}$$

2. 평균과 표준편차

$$(1) \text{평균 } m = \sum_{i=1}^n x_i \frac{f_i}{N} = \sum_{i=1}^n x_i P(X = x_i)$$

☐ (평균) = {(변량) × (확률)}의 합

$$(2) \text{분산 } \sigma^2 = \sum_{i=1}^n (x_i - m)^2 \frac{f_i}{N}$$

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

☐ (분산) = {(편차)의 제곱} × (확률)의 합
= {(제곱)의 평균} - {(평균)의 제곱}

(3) 표준편차 $\sigma : \text{분산 } (\sigma^2)$ 의 양의 제곱근

☐ (표준편차) = $\sqrt{(\text{분산})}$

< 통계학 >

통계학의 기원은 인구조사로부터 시작되었다고도 할 수 있다. 기록 상 인류 최초의 인구조사는 구약성서에 나타난 것으로 모세가 애굽 땅에서 나온 유대인을 대상으로 20세 이상의 성인의 수를 조사한 것이었다. 그러나 실질적인 인구조사는 이보다 훨씬 더 이전 시대인 기원전 3000년경 고대 바빌로니아, 중국, 이집트에서 군사적인 징병과 세금 징수를 목적으로 인구조사가 실시되었다고 한다.

인구조사를 영어로는 census라고 하는데 census는 과세한다는 뜻의 라틴어인 censere로부터 유래된 것이다. 로마시대에는 국가의 조직이 발달되면서 과세의 목적과 군사적인 징병 목적을 위하여, 로마의 6대왕 투리우스(534 ~ 378 B.C.)시대부터 서기 74년까지 5년 주기의 센서스가 실시되었다. 그러나 로마제국이 붕괴되면서 17세기 초반까지 서구 세계에서는 인구조사가 실시된 기록이 없다가 17세기 이후에 통계가 다시 등장하기 시작하였다.

17세기 중반 독일의 콘링(Corning, H. : 1606 ~ 1681)이 국가의 중요 사항인 국토, 군사, 인구 및 행정에 관하여 강의한 것을 국상학(staatenkunde)이라고 불렀는데, 이 국상학은 국가의 중요 사항을 통계 수치 없이 문장으로 기술한 것으로 통계학의 모체라고 할 수 있다.

17세기 중반 영국에서는 흑사병으로 많은 사상자가 발생하자 출생과 사망에 관한 통계를 매주 발표하였는데, 이를 토대로 영국 상인 그란트(Grant, J. : 1620 ~ 1674)는 1662년에 「사망표에 관한 자연적, 정치적 관찰」이라는 연구에서 인구, 사망, 출생에 관한 법칙성을 규명하고 인구의 측정, 연령 계층별 연구, 인구의 증가율 등을 추정하였다.

그 후 18세기에 파스칼과 페르마 등에 의하여 확률론이 정립되어 우연히 일어나는 현상을 수학적으로 분석하여 처리하는 방법이 개발되었고, 이를 기초로 19세기 초 통계학이 싹트기 시작하였다. 그러나 그때까지의 통계학은, 많은 양의 관측값을 얻을 수 없는 경우에는 추론이 어려웠다. 이후 피셔(Fisher, R. A. : 1890 ~ 1962)는 이를 보완하여 현대추측통계학의 기초를 확립하였고, 금세기 확률론과 통계학의 두 분야는 컴퓨터의 역할에 힘입어 여러 분야에서 크게 발전하고 있다.

< 호이겐스와 수학적 기댓값 >

호이겐스는 파스칼과 페르마의 서신 왕래를 기초로 1657년 확률에 관한 첫 논문을 썼다. 그 논문에서 '수학적 기댓값'이라는 중요한 개념이 등장한다.

상금 S 를 받을 확률이 p 라면 상금의 수학적 기댓값은 pS 이다. 또, 상금 A 를 받을 확률이 p 이고, 상금 B 를 받을 확률이 q 라면 상금의 수학적 기댓값은 $pA + qB$ 이다.

< 확률분포 >

※ 확률분포

$$\left\{ \begin{array}{l} \text{이산확률분포} \rightarrow \text{이항분포 } X : B(n, p) \\ \sum_i p_i = 1 \\ m = E(X) = \sum_i x_i p_i \\ \sigma^2 = \sum_i (x_i - m)^2 p_i \\ = E(X^2) - m^2 \end{array} \right. \quad \begin{array}{l} m = np \\ \sigma^2 = npq \end{array}$$

연속확률분포 \rightarrow 정규분포 $X : N(m, \sigma^2)$
확률밀도함수 $y = f(x)$

$$m = E(X) = \int_a^b x f(x) dx$$

$$\sigma^2 = \int_a^b (x - m)^2 f(x) dx$$

$$= \int_a^b x^2 f(x) dx - m^2$$

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

표준정규분포 $Z : N(0, 1)$

$$P(a \leq X \leq b)$$

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

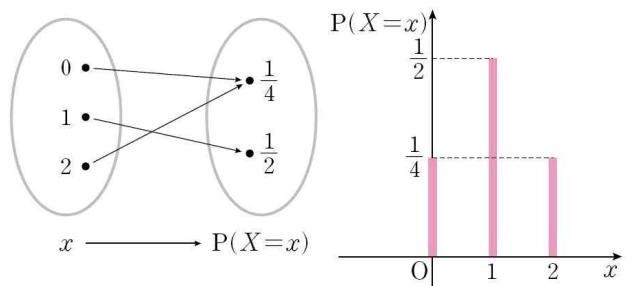
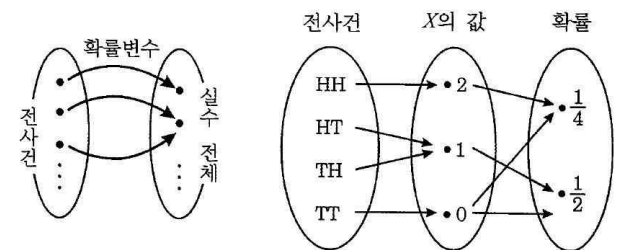
1. 이산확률분포

1 확률변수

1. 확률변수

- (1) 어떤 시행에서 표본공간의 각 원소에 하나의 수를 대응시킬 때 $\Rightarrow X$ 는 확률변수
- (2) 어떤 시행의 표본공간 S 에서 실수 전체의 집합 R 로의 함수 $X : S \rightarrow R$

예) 동전 2개 \Rightarrow 앞면이 나오는 횟수 X



X	0	1	2	합계
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

- (3) X 가 어떤 값 x 를 가질 확률 \Rightarrow (기호) $P(X = x)$
- (4) 확률변수는 보통 X, Y, Z, \dots 로 나타내고 확률변수가 취하는 값은 x, y, z, \dots 으로 나타낸다.

2. 이산확률변수와 확률분포

- (1) 확률변수 X 가 유한 개이거나 자연수와 같이 셀 수 있을 때 $\Rightarrow X$ 는 이산확률변수 : x_1, x_2, \dots, x_n
- (2) $P(X = x_i) = p_i \quad (i = 1, 2, \dots, n)$
 $\Rightarrow X$ 의 확률질량함수

(3) 이산확률변수 X 가 취하는 값 x_i 와 그 값을 취할 확률 p_i 의 대응관계 $\Rightarrow X$ 의 확률분포

(4) 이산확률분포표

X	x_1	x_2	x_3	\cdots	x_n	합계
$P(X=x)$	p_1	p_2	p_3	\cdots	p_n	1

(5) X 가 x_i 이상 x_j 이하의 값을 가질 확률
 $\Rightarrow P(x_i \leq X \leq x_j)$

3. 확률질량함수(probability mass function)의 성질

(1) $0 \leq p_i \leq 1$ (단, $i = 1, 2, \cdots, n$)

(2) $\sum_{i=1}^n p_i = 1 \quad \Rightarrow$ 확률의 총합은 항상 1

(3) $P(x_i \leq X \leq x_j) = \sum_{k=i}^j P(X = x_k) = \sum_{k=i}^j p_k$
 (단, $i, j = 1, 2, \cdots, n, i < j$)

예 남학생 4명과 여학생 3명 중에서 임의로 대표 3명을 뽑을 때, 뽑힌 남학생의 수를 확률변수 X 라고 하자.

(1) 확률변수 X 가 취하는 값은 0, 1, 2, 3이고, 각 값을 취할 확률은

$$P(X=0) = \frac{{}_4C_0 \times {}_3C_3}{{}_7C_3} = \frac{1}{35}$$

$$P(X=1) = \frac{{}_4C_1 \times {}_3C_2}{{}_7C_3} = \frac{12}{35}$$

$$P(X=2) = \frac{{}_4C_2 \times {}_3C_1}{{}_7C_3} = \frac{18}{35}$$

$$P(X=3) = \frac{{}_4C_3 \times {}_3C_0}{{}_7C_3} = \frac{4}{35}$$

$\therefore X$ 의 확률질량함수는

$$P(X=x) = \frac{{}_4C_x \times {}_3C_{3-x}}{{}_7C_3} \quad (x=0, 1, 2, 3)$$

(2) 남학생이 적어도 2명 뽑힐 확률은 $X \geq 2$ 이므로

$$\begin{aligned} P(X \geq 2) &= P(X=2) + P(X=3) \\ &= \frac{18}{35} + \frac{4}{35} = \frac{22}{35} \end{aligned}$$

2 이산확률분포의 평균과 표준편차

1. 이산확률분포의 평균(기댓값)

(1) (평균) = {(변량) \times (확률)}의 합

$$(2) \text{평균(기댓값)} : m = E(X) = \sum_{i=1}^n x_i p_i$$

☑ 평균(mean) $\Rightarrow m$

기댓값(Expected value) $\Rightarrow E(X)$

2. 이산확률분포의 분산과 표준편차

(1) (분산) = {(편차)의 제곱}의 평균
 $= \{(\text{제곱})의 \text{평균}\} - \{(\text{평균})의 \text{제곱}\}$

(2) 분산 : σ^2

$$\begin{aligned} V(X) &= E\{(X-m)^2\} = \sum_{i=1}^n (x_i - m)^2 p_i \\ &= E(X^2) - m^2 = \sum_{i=1}^n x_i^2 p_i - m^2 \end{aligned}$$

(3) 표준편차 : $\sigma(X) = \sqrt{V(X)} = \sigma$

☑ 분산(Variance) $\Rightarrow V(X) = \sigma^2$

표준편차(Standard deviation) $\Rightarrow \sigma(X) = \sigma$

3. 분산 구하는 식의 변형

$$\begin{aligned} V(X) &= \sum_{i=1}^n (x_i - m)^2 p_i \\ &= \sum_{i=1}^n x_i^2 p_i - 2m \sum_{i=1}^n x_i p_i + m^2 \sum_{i=1}^n p_i \\ &= \sum_{i=1}^n x_i^2 p_i - 2m \cdot m + m^2 \\ &= \sum_{i=1}^n x_i^2 p_i - m^2 = E(X^2) - m^2 \end{aligned}$$

4. 확률변수의 변환

확률변수 X 와 상수 a, b 에 대하여

(1) 평균 : $E(aX+b) = aE(X) + b$

(2) 분산 : $V(aX+b) = a^2 V(X)$

$$\begin{aligned}\square E(aX+b) &= \sum_{i=1}^n (ax_i+b)p_i \\ &= a \sum_{i=1}^n x_i p_i + b \sum_{i=1}^n p_i \\ &= aE(X) + b \\ \square V(aX+b) &= \sum_{i=1}^n \{(ax_i+b) - (am+b)\}^2 p_i \\ &= a^2 \sum_{i=1}^n (x_i-m)^2 p_i = a^2 V(X)\end{aligned}$$

$$(3) \text{ 표준편차 : } \sigma(aX+b) = |a|\sigma(X)$$

5. 이산확률변수의 평균과 표준편차 구하기

$$(1) \text{ 확률변수 } X \text{의 확률분포표 작성} \Rightarrow \sum p_i = 1$$

$$(2) \text{ 평균} \Rightarrow E(X) = m$$

$$(3) \text{ 제곱의 평균} \Rightarrow E(X^2)$$

$$(4) \text{ 분산} \Rightarrow \sigma^2 = V(X) = E(X^2) - m^2$$

$$(5) \text{ 표준편차} \Rightarrow \sigma = \sqrt{V(X)}$$

6. 표준점수(표준화된 점수)

(1) 확률변수 X 의 평균과 표준편차가 각각 m, σ 일 때,

$$Z = \frac{X-m}{\sigma} \Rightarrow E(Z) = 0, \sigma(Z) = 1$$

(2) 평균이 m , 표준편차가 σ 인 시험 점수 X 에 대하여

① 언어·수리·외국어 표준점수 T 는

평균 100, 표준편차 20으로 변환하므로

$$\begin{aligned}Z &= \frac{X-m}{\sigma} = \frac{T-100}{20} \\ \therefore T &= 100 + 20\left(\frac{X-m}{\sigma}\right)\end{aligned}$$

② 탐구영역 각 과목 표준점수 T 는

평균 50, 표준편차 10으로 변환하므로

$$\begin{aligned}Z &= \frac{X-m}{\sigma} = \frac{T-50}{10} \\ \therefore T &= 50 + 10\left(\frac{X-m}{\sigma}\right)\end{aligned}$$

7. 두 확률변수 X, Y 가 서로 독립이면

$$(1) E(X+Y) = E(X) + E(Y)$$

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

$$(2) V(X+Y) = V(X) + V(Y)$$

(3) 함수 $f(x), g(x)$ 에 대하여

$$f(x) \leq g(x) \Rightarrow E(f(x)) \leq E(g(x))$$

③ 이항분포 $\Rightarrow X : B(n, p)$

\hookrightarrow Binomial distribution

1. 이항분포 (\iff 독립시행의 확률분포) $\Rightarrow B(n, p)$

(1) n 회 독립시행에서 사건 A 가 일어나는 횟수 X

$\Rightarrow X$ 는 0, 1, 2, \dots , n 의 값을 갖는 확률변수

(2) 1회 시행에서 사건 A 가 일어날 확률 p

\Rightarrow 확률변수 X 의 확률질량함수는

$$P_x = P(X=x) = {}_n C_x p^x q^{n-x}$$

(단, $p+q=1, x=0, 1, 2, \dots, n$)

(3) 확률변수 X 의 확률분포

X	0	1	2	\dots	n	합계
$P(X=x)$	${}_n C_0 q^n$	${}_n C_1 p^1 q^{n-1}$	${}_n C_2 p^2 q^{n-2}$	\dots	${}_n C_n p^n$	1

$$(4) \sum_{x=0}^n P_x = \sum_{x=0}^n {}_n C_x p^x q^{n-x} = (p+q)^n = 1$$

(5) 이와 같은 확률변수 X 의 확률분포

\Rightarrow 이항분포 (기호) $X : B(n, p)$

2. 이항분포의 평균과 표준편차

확률변수 X 가 이항분포 $B(n, p)$ 를 따를 때,

$$(o) \sum_{x=0}^n {}_n C_x p^x q^{n-x} = (p+q)^n = 1 \quad (\because \text{이항정리})$$

$$\begin{aligned}(1) E(X) &= \sum_{x=0}^n x \cdot {}_n C_x p^x q^{n-x} = np \\ &\quad \left(\because {}_n C_x = \frac{n}{x} \times {}_{n-1} C_{x-1} \right)\end{aligned}$$

$$(2) V(X) = npq = np(1-p) \quad (\text{단, } p+q=1)$$

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

$$(4) E(X^2) = \sum_{x=0}^n x^2 \cdot {}_n C_x p^x q^{n-x} = npq + (np)^2$$

$$(\because V(X) = E(X^2) - m^2)$$

$$\text{예} \quad ① \sum_{x=0}^9 {}_9 C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x} = \left(\frac{1}{3} + \frac{2}{3}\right)^9 = 1$$

$$② \sum_{x=0}^9 x \cdot {}_9 C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x} = 9 \cdot \frac{1}{3} = 3$$

$$③ \sum_{x=0}^9 x^2 \cdot {}_9 C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x}$$

$$= 9 \cdot \frac{1}{3} \cdot \frac{2}{3} + \left(9 \cdot \frac{1}{3}\right)^2 = 2 + 9 = 11$$

3. 이항 분포의 특징

(1) 시행 결과는 성공(p) 또는 실패(q) 뿐이다.

(2) 실험은 n 번의 독립시행으로 이루어진다.

즉, 한 시행에서의 결과가 다음 시행의 성공 확률에 영향을 미치지 않는다.

(3) $p + q = 1$ 의 관계를 갖는다.

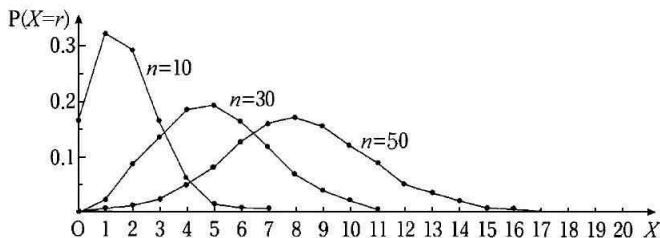
4. 이항분포의 그래프와 성질

(1) p 를 일정하게 하고 n 을 크게 ($n \geq 50$)

(2) n 을 일정하게 하고 p 를 0.5 에 가깝게

\Rightarrow 점차로 종 모양(정규분포곡선)에 가까워진다.

예 주사위를 n 번 던질 때, 1의 눈이 나올 횟수 X



5. 특수한 경우의 평균

(1) $a \times 100$ 원 동전 n 개를 모두 꺼낼 때, 앞면(or 뒷면)이 나오는 개수만큼 상금을 받는다면 기댓값은?

$$E(X) = \sum_{x=0}^n {}_n C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} \times x \times a \cdot 100$$

$$= a \cdot 100 \times \sum_{x=0}^n x \cdot {}_n C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x}$$

$$= a \cdot 100 \times n \left(\frac{1}{2}\right)$$

(2) n 개의 동전을 던져서 앞면이 r 개가 나오면 a^r 원을 상금으로 받는다고 할 때, 상금의 기댓값은

$$E(X) = \sum_{x=0}^n a^x {}_n C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x}$$

$$= \sum_{x=0}^n {}_n C_x \left(\frac{a}{2}\right)^x \left(\frac{1}{2}\right)^{n-x}$$

$$= \left(\frac{a+1}{2}\right)^n$$

6. 확률변수 X 가 이항분포 $B(n, p)$ 를 따른다고 할 때,

$$(1) E(X) = \sum_{x=0}^n x \cdot P(X=r)$$

$$= \sum_{x=0}^n x \cdot {}_n C_x p^x q^{n-x}$$

$$= \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x}$$

$$= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}$$

$$= np \sum_{x=1}^n {}_{n-1} C_{x-1} p^{x-1} q^{(n-1)-(x-1)}$$

$$= np \sum_{x=0}^{n-1} {}_{n-1} C_r p^x q^{n-1-x}$$

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

(1) [미분을 이용한 다른 증명]

$$(px+q)^n = \sum_{r=0}^n {}_n C_r (px)^r q^{n-r}$$

$$= \sum_{r=0}^n {}_n C_r p^r q^{n-r} x^r$$

양변을 x 에 관하여 미분하면

$$n(px+q)^{n-1} \cdot p = \sum_{r=0}^n r \cdot {}_n C_r p^r q^{n-r} x^{r-1}$$

양변에 $x=1$ 을 대입하면

$$np(p+q)^{n-1} = \sum_{r=0}^n r \cdot {}_n C_r p^r q^{n-r}$$

$$\therefore np = E(X)$$

(2) $x^2 = x(x-1) + x$ 를 이용하여

$$\begin{aligned}
 E(X^2) &= \sum_{x=0}^n x^2 \cdot P(X=x) \\
 &= \sum_{x=0}^n x^2 \cdot {}_n C_x p^x q^{n-x} \\
 &= \sum_{x=0}^n x(x-1) {}_n C_x p^x q^{n-x} \\
 &\quad + \sum_{x=0}^n x \cdot {}_n C_x p^x q^{n-x} \\
 &= \sum_{x=2}^n \frac{n!}{(x-2)!(n-x)!} p^x q^{n-x} + m \\
 &= n(n-1)p^2 \sum_{x=2}^n \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} + np \\
 &= n(n-1)p^2 \sum_{x=0}^{n-2} \frac{(n-2)!}{x!(n-2-x)!} p^x q^{n-2-x} + np \\
 &= n(n-1)p^2(p+q)^{n-2} + np \\
 &= n(n-1)p^2 + np \\
 &= n^2p^2 + np(1-p) = (np)^2 + npq
 \end{aligned}$$

$$\begin{aligned}
 (3) \quad V(X) &= E(X^2) - \{E(X)\}^2 \\
 &= n^2p^2 + np(1-p) - n^2p^2 \\
 &= np(1-p) = npq
 \end{aligned}$$

3 큰 수의 법칙

1. 큰 수의 법칙

(1) n 회의 독립시행에서 사건 A 가 일어나는 횟수 X , 매회의 시행에서 사건 A 가 일어날 확률이 p 일 때, 아주 작은 임의의 양수 h 에 대하여

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

(2) 큰 수의 법칙에 의하여 시행 횟수 n 이 충분히 클 때 상대도수 $\frac{X}{n}$, 즉 통계적 확률은 수학적 확률 p 에 가까우므로 자연현상 및 사회현상에서 수학적 확률을 알 수 없는 경우에는 통계적 확률을 이용할 수 있다.

(3) 큰 수의 법칙에 대한 엄밀한 정의

모집단에서 임의 추출한 표본 X_1, X_2, \dots, X_n 에 대하여 표본평균 \bar{X} 는 표본의 크기 n 이 커질수록 모평균 m 에 가까워진다는 것이다. 즉, 충분히 작은

양수 h 에 대하여

$$\lim_{n \rightarrow \infty} P(|\bar{X} - m| < h) = 1 \quad \cdots \textcircled{1}$$

이 성립한다는 것이다.

여기서 X_i ($i = 1, 2, \dots, n$)가 베르누이 시행이면 X_i 는 0과 1의 값만 취하므로

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n} = \frac{X}{n}$$

이고, X 는 이항분포에 따르므로 평균 m 은

$$E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} \cdot np = p$$

가 된다.

따라서 $\textcircled{1}$ 은

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

이다.

이 식은 상대도수(통계적 확률)가 시행횟수 n 이 커질수록 수학적 확률 p 에 가까워진다는 것을 의미한다.

(4) 체비쇼프의 부등식

임의의 확률변수 X 와 $h > 0$ 에 대하여

$$P(|X - m| \geq h) \leq \frac{V(X)}{h^2}$$

(단, $m = E(X)$)

2. 큰 수의 약법칙

(1) X_1, X_2, \dots, X_n 은 서로 독립이고 같은 분포를

가질 때, $S_n = \sum_{k=1}^n X_k$, $m = E(X_k)$ 라 하면,

임의의 $h > 0$ 에 대하여

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - m\right| > h\right) = 0$$

(2) X_1, X_2, \dots, X_n 은 서로 독립이고

$$X_k = \begin{cases} 1 & (k\text{번째 사건 } A \text{ 가 일어나는 경우}) \\ 0 & (k\text{번째 사건 } A \text{ 가 일어나지 않는 경우}) \end{cases}$$

(단, $k = 1, 2, \dots, n$) [베르누이 분포]이면

$E(X_k) = p$ 이고 S_n 은 이항분포 $B(n, p)$ 를 따르므로

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - p\right| > h\right) = 0$$

(단, 여기서 $\frac{S_n}{n}$ 은 사건 A 의 상대도수)

3. 큰 수의 강법칙

(1) X_1, X_2, \dots, X_n 은 서로 독립이고 같은 분포를

가질 때, $S_n = \sum_{k=1}^n X_k$, $m = E(X_k)$ 라 하면,

임의의 $h > 0$ 에 대하여

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - m\right| < h\right) = 1$$

(2) X_1, X_2, \dots, X_n 은 서로 독립이고

$$X_k = \begin{cases} 1 & (k\text{번째 사건 } A \text{가 일어나는 경우}) \\ 0 & (k\text{번째 사건 } A \text{가 일어나지 않는 경우}) \end{cases}$$

(단, $k = 1, 2, \dots, n$) [베르누이 분포]이면

$E(X_k) = p$ 이고 S_n 은 이항분포 $B(n, p)$ 를

따르므로

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

< 대규모 시험에서의 표준 점수 >

우리나라의 대학 수학 능력 시험이나 미국의 SAT 등과 같이 수험생이 많고, 과목 수도 여러 가지인 경우에는 항상 과목별 난이도의 차이로 인한 문제가 생긴다.

예를 들어 어떤 학생의 국어 점수가 80 점, 수학 점수가 75 점일 때, 80 과 75 의 단순 비교만으로는 국어 성적이 좋은지, 수학 성적이 좋은지 알 수 없다.

특히 선택 과목의 경우 선택에 따른 유리함 또는 불리함을 없애고 각 과목 사이의 난이도 조정을 위해서 표준 점수의 도입이 필수적이다. 표준 점수는 전체 응시자의 과목별 평균 점수와 표준편차를 이용하여 각 학생의 과목별 점수가 과목별 전체 평균 점수보다 얼마나 높은지 혹은 얼마나 낮은지를 비교하는 것이다.

표준 점수 중에서 T 점수는 전체 평균이 50 점, 표준편차가 10 점이 되도록 변환한 것이다.

예를 들어 각 학생의 수학에 대한 T 점수는 다음과 같이 계산한다.

$$T = 50 + 10 \times \frac{(\text{각 학생의 수학 원점수}) - (\text{수학 점수의 평균})}{(\text{수학 점수의 표준편차})}$$

이 점수는 과목에 상관없이 약 99.7 %의 학생들의 점수가 15 점에서 85 점 사이에 있게 된다.

[보기] 전국 규모의 성취도 평가에서 A, B 두 학생의 과목별 점수가 다음과 같다고 하자.

구분	국어	영어	수학	과학	사회	합계
평균	70	60	50	75	80	
표준편차	8	6	10	7	5	
A 학생	86	75	52	82	90	385
B 학생	78	72	61	89	85	385

여기서 A, B 두 학생의 총점은 385 점으로 같음을 알 수 있다.

그러나 다음과 같이 T 점수로 바꾸면 A 학생의 점수가 B 학생의 점수보다 높음을 알 수 있다.

구분	국어	영어	수학	과학	사회	합계
A 학생	70	75	52	60	70	327
B 학생	60	70	61	70	60	321

< 큰 수의 법칙에 대한 오해 >

주사위 던지기를 하여 나오는 눈의 수를 맞추는 게임을 한다고 하자. 주사위를 충분히 많이 던진 후, 이제까지 나온 눈을 보았을 때 6의 눈이 가장 적게 나왔다고 한다. 그렇다면 다음번에 던질 때는 주사위의 눈이 6이 나올 가능성이 가장 클 것이라고 생각할 수 있을까?

대답은 ‘아니오.’이다. 6의 눈이 나올 확률은 조금도 높아지거나 낮아지지 않는다. 여전히 6의 눈이 나올 확률은 다른 눈의 수와 마찬가지로 $\frac{1}{6}$ 이다. 그렇지만 많은 사람들은 시행의 횟수가 많아지면 통계적 확률이 수학적 확률에 점점 근접한다는 큰 수의 법칙에 근거하여 이제 슬슬 6의 눈이 나올 때가 되었다고 생각한다. 그래서 이 게임에서 이번에는 6이 나올 것이라고 말한다. 정말 그럴까?

실제로, 주사위를 계속 던지다 보면 6의 눈이 나오는 경우의 수는 어느 정도 차이는 있겠지만 이론상의 수학적 확률 $\frac{1}{6}$ 에 근접하게 된다. 예를 들어, 주사위를 60회 던져서 6의 눈이 13회 나왔다고 하자. 또, 주사위를 600회 던져서 6의 눈이 94회가, 6000회를 던져서 986회 나왔다고 하자. 주사위를 한 번 던져 6의 눈이 나올 수학적 확률

$$\frac{1}{6} = 0.166 \dots \text{에 대하여}$$

$$\frac{13}{60} = 0.216 \dots, \frac{94}{600} = 0.156 \dots, \frac{986}{6000} = 0.164 \dots$$

로 그 확률은 점점 $\frac{1}{6}$ 에 가까워지고 있다.

하지만 실제로 6의 눈이 나오는 횟수와 이론적으로 나와야 될 횟수의 차는

$$13 - 10 = 3, 100 - 94 = 6, 1000 - 986 = 14$$

로 점점 커진다. 즉, 시행 횟수가 커질 때 실제로 6의 눈이 나오는 횟수와 이론적으로 나와야 할 횟수 사이의 간격은 작아지는 것이 아니라, 일반적으로 점점 더 커진다.

즉, n 회의 시행 중 6의 눈이 나오는 횟수를 X 라고 하면 큰 수의 법칙에 의하여 n 의 값이 점점 커지면 통계적 확률 $\frac{X}{n}$ 는 점점 수학적 확률 $\frac{1}{6}$ 에 가까워지지만, 이론적으로

6이 나와야 할 횟수 $\frac{n}{6}$ 과 6이 실제로 나온 횟수 X 의 차가 줄어드는 것은 아니다.

< 여러 가지 분포 >

1. 기하분포

한 시행에서 어떤 사건이 일어날 확률이 p 일 때, 그 사건이 처음으로 일어날 때까지의 시행 횟수를 확률변수 X 라고 하면 X 의 확률질량함수는

$$P(X = x) = pq^{x-1} \quad (\text{단, } x = 1, 2, \dots)$$

이다. 이 확률질량함수가 이루는 확률분포를 기하분포라고 한다.

2. 초기하분포

검은 구슬 M 개를 포함하여 총 N 개의 구슬이 들어 있는 주머니에서 n 개의 구슬을 꺼낼 때 나오는 검은 구슬의 개수를 확률변수 X 라 하면 X 의 확률질량함수는

$$P(X = x) = \frac{{}^M C_x \cdot {}^{N-M} C_{n-x}}{{}^N C_n} \quad (\text{단, } x = 1, 2, \dots)$$

이다. 이 확률질량함수가 이루는 확률분포를 초기하분포라고 한다.

3. 적률(Moment) : 확률변수 X 에 대하여 X^k 의 기댓값 $E(X^k)$ 를 확률변수 X 의 k

2. 연속확률분포

1 확률밀도함수

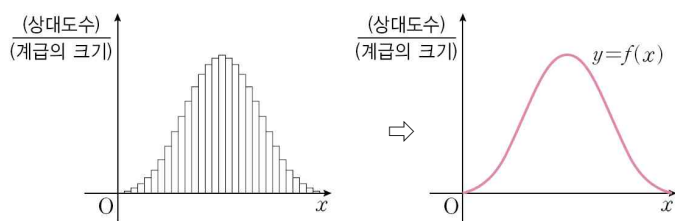
1. 연속확률변수

- (1) 시간, 길이, 무게, 온도 등과 같이 어떤 구간의 모든 실수값을 가지는 확률변수 X
 $\Rightarrow X$ 는 연속확률변수

- (2) 연속확률변수 X 의 $\frac{(\text{상대도수})}{(\text{계급의 크기})}$ 를 히스토그램으로 나타내면 히스토그램에서 각 구간에 세워진 직사각형의 넓이는 각 구간의 상대도수를 나타낸다.
 $\frac{(\text{상대도수})}{(\text{계급의 크기})} \times (\text{계급의 크기}) = (\text{직사각형의 넓이})$

- (3) 상대도수의 합이 1이기 때문에 직사각형들의 넓이의 합은 항상 1이 된다.

- (4) 총 도수를 늘리고 계급의 폭을 더욱 좁게 해서 히스토그램을 그리면 곡선 모양에 가까워진다.



2. 확률밀도(density)함수 $y = f(x)$ 의 성질

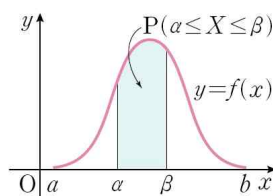
- (o) 연속확률변수 X 가 하나의 값을 가질 확률은 0

$$P(\alpha \leq X \leq \beta)$$

$$= P(\alpha \leq X < \beta)$$

$$= P(\alpha < X < \beta)$$

$$= P(\alpha < X \leq \beta)$$



- (1) $f(x) \geq 0$ (단, $a \leq x \leq b$)
 \Leftrightarrow 곡선 $y = f(x)$ 는 x 축의 위쪽에 있다.

$$(2) \int_a^b f(x) dx = 1$$

$$\Leftrightarrow \begin{cases} y = f(x) \\ y = 0 \text{ (x 축)} \end{cases} \text{사이의 넓이는 1}$$

$$(3) P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x) dx$$

$$\Leftrightarrow \begin{cases} y = f(x) \\ y = 0 \text{ (x 축)} \end{cases} \text{와 } \alpha \leq x \leq \beta \text{의 넓이}$$

3. 연속확률분포(\Leftrightarrow 확률밀도함수)의 평균과 표준편차

연속확률변수 X 의 확률밀도함수가 $y = f(x)$ 일 때

$$(1) m = E(X) = \int_a^b x f(x) dx$$

$$\Leftrightarrow \begin{cases} y = x \cdot f(x) \\ y = 0 \text{ (x 축)} \end{cases} \text{사이의 넓이}$$

$$(2) V(X) = \int_a^b (x - m)^2 f(x) dx$$

$$= \int_a^b x^2 f(x) dx - m^2$$

$$\Leftrightarrow \begin{cases} y = x^2 \cdot f(x) \\ y = 0 \text{ (x 축)} \end{cases} \text{사이의 넓이} - (\text{평균})^2$$

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

4. 확률변수의 변환

확률변수 X 와 상수 a, b 에 대하여

$$(1) \text{평균} : E(aX + b) = aE(X) + b$$

$$(2) \text{분산} : V(aX + b) = a^2 V(X)$$

$$(3) \text{표준편차} : \sigma(aX + b) = |a| \sigma(X)$$

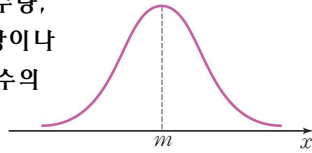
$$\begin{aligned} \square E(aX + b) &= \int_a^b (ax + b) f(x) dx \\ &= a \int_a^b x f(x) dx + b \int_a^b f(x) dx \\ &= aE(X) + b \end{aligned}$$

$$\begin{aligned} \square V(aX + b) &= \int_a^b \{ (ax + b) - (am + b) \}^2 f(x) dx \\ &= a^2 \int_a^b (x - m)^2 f(x) dx \\ &= a^2 V(X) \end{aligned}$$

2 정규분포 $\Rightarrow X : N(m, \sigma^2)$

↳ Normal distribution

☑ 키, 몸무게, 측정오차, 강수량,
IQ, 시험 점수 등 자연 현상이나
사회 현상과 관련된 확률변수의
확률밀도함수의 그래프는
좌우 대칭인 종 모양의
곡선에 가까운 경우가 많다.



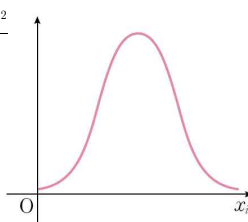
1. 정규분포 \Rightarrow (기호) $N(m, \sigma^2)$

연속확률변수 X 의 확률밀도함수 $y = f(x)$ 가

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

(단, $-\infty < x < \infty$)

$\Rightarrow X$ 는 정규분포를 따른다
 $y = f(x)$ 는 정규분포곡선



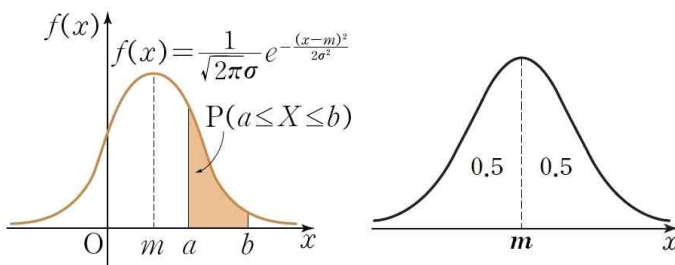
☑ e 는 $e = 2.7182818286 \dots$ 인 무리수

2. 정규분포곡선 $y = f(x)$ 의 기본 성질

- (1) $f(x) \geq 0$ (단, $-\infty < x < \infty$)
 \Leftrightarrow 곡선 $y = f(x)$ 는 x 축의 위쪽에 있다.

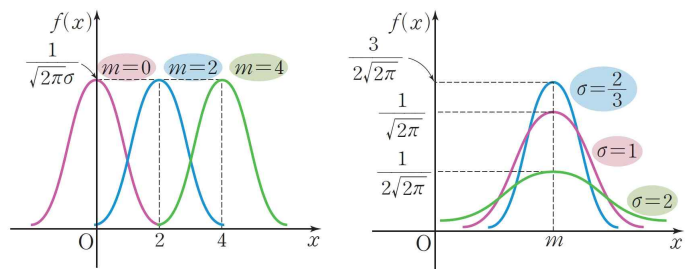
(2) $\int_{-\infty}^{\infty} f(x) dx = 1$
 $\Leftrightarrow \begin{cases} y = f(x) \\ y = 0 \text{ (} x \text{ 축)} \end{cases}$ 사이의 넓이는 1.

(3) $P(a \leq X \leq b) = \int_a^b f(x) dx$
 $\Leftrightarrow \begin{cases} y = f(x) \\ y = 0 \text{ (} x \text{ 축)} \end{cases}$ 와 $a \leq x \leq b$ 의 넓이



(4) $P(X \geq m) = P(X \leq m) = 0.5$

2. 정규분포곡선 $y = f(x)$ 의 성질



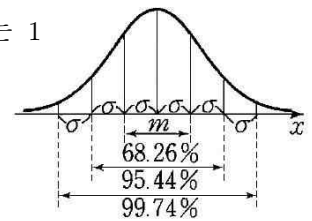
(1) 직선 $x = m$ 에 대하여 대칭이고, 종 모양의 곡선

(2) 점근선은 x 축이며, $x = m$ 에서 최대값 $\frac{1}{\sqrt{2\pi}\sigma}$

(3) σ 가 일정할 때, m 이 변하면
곡선의 모양은 같고 대칭축의 위치만 바뀐다.

(4) m 이 일정할 때,
 σ 가 작아질수록 곡선은 폭이 좁고 높아진다.

(5) 곡선과 x 축 사이의 넓이는 1



3. 정규분포의 확률

확률변수 X 가 정규분포 $N(m, \sigma^2)$ 를 따를 때,

- (1) $P(m - \sigma \leq X \leq m + \sigma) \approx 0.6826$
(2) $P(m - 2\sigma \leq X \leq m + 2\sigma) \approx 0.9544$
(3) $P(m - 3\sigma \leq X \leq m + 3\sigma) \approx 0.9974$

4. 정규분포에서 확률을 구하는 방법

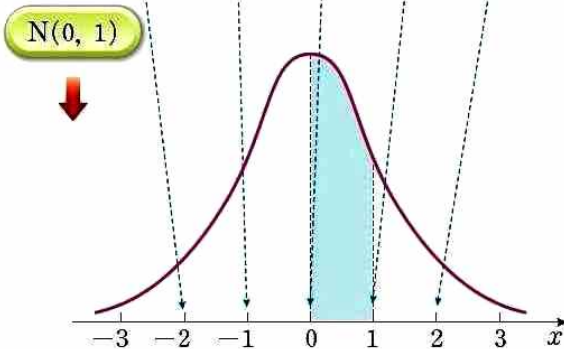
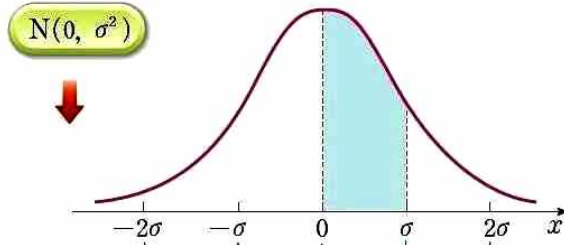
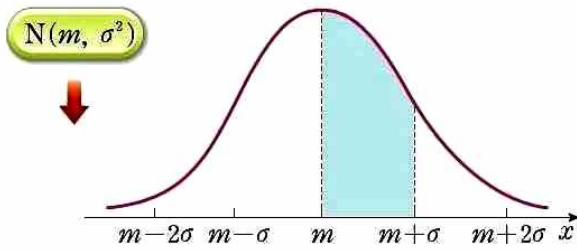
확률변수 X 가 정규분포 $N(m, \sigma^2)$ 을 따른다고 할 때
 $P(m \leq X \leq m + \sigma)$

$$= \int_m^{m+\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

그런데, m 과 σ 는 분포에 따라 여러 가지 값은
가지므로 매번 확률을 계산하려면 매우 복잡하다.
따라서 미리 계산해 놓은 표준정규분포표를 이용한다.

$$\begin{aligned} \therefore P(m \leq X \leq m + \sigma) &= P\left(\frac{m-m}{\sigma} \leq Z \leq \frac{m+\sigma-m}{\sigma}\right) \\ &= P(0 \leq Z \leq 1) = 0.3413 \end{aligned}$$

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



③ 표준정규분포 $\Rightarrow Z : N(0, 1)$

1. 표준정규분포 : 정규분포 $N(m, \sigma^2)$ 을 표준화

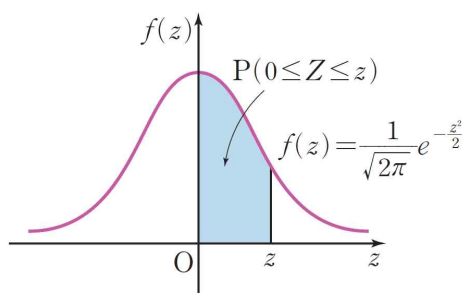
$$(1) f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

(단, $-\infty < z < \infty$)

$$(2) P(Z \leq 0) = P(Z \geq 0) = 0.5$$

$$(3) P(0 \leq Z \leq a) = \int_0^a f(z) dz$$

= (위쪽 그림에서 색칠된 부분의 넓이)



(4) 표준정규분포
 $P(0 \leq Z \leq 1.46)$
 $= 0.4279$

②			
Z	0.00	...	0.06
⋮			
① 1.4			.4279
⋮			

2. 확률변수의 표준화

(1) 정규분포 $\Rightarrow \Rightarrow \Rightarrow$ 표준정규분포
 $X : N(m, \sigma^2) \Rightarrow \Rightarrow \Rightarrow Z : N(0, 1)$

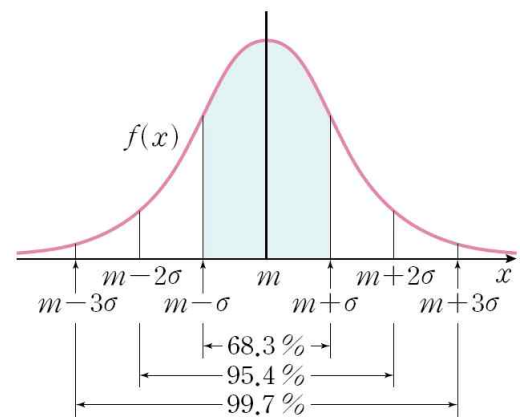
$$Z = \frac{X - m}{\sigma}$$

$$(2) P(a \leq X \leq b) = P\left(\frac{a - m}{\sigma} \leq Z \leq \frac{b - m}{\sigma}\right)$$

$$(3) \textcircled{1} P(m - \sigma \leq X \leq m + \sigma) = P(|Z| \leq 1) = P(-1 \leq Z \leq 1) = 0.6826 \quad (\approx 68.3\%)$$

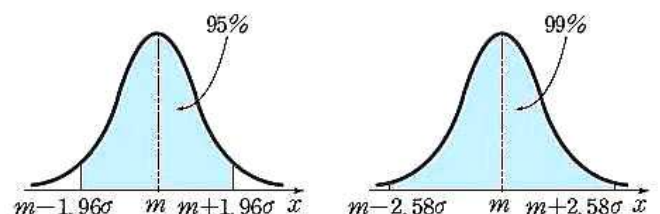
$$\textcircled{2} P(m - 2\sigma \leq X \leq m + 2\sigma) = P(|Z| \leq 2) = P(-2 \leq Z \leq 2) = 0.9544 \quad (\approx 95.4\%)$$

$$\textcircled{3} P(m - 3\sigma \leq X \leq m + 3\sigma) = P(|Z| \leq 3) = P(-3 \leq Z \leq 3) = 0.9974 \quad (\approx 99.7\%)$$



$$(4) \textcircled{1} P(m - 1.96\sigma \leq X \leq m + 1.96\sigma) = P(-1.96 \leq Z \leq 1.96) = 0.95 \quad (\approx 95\%)$$

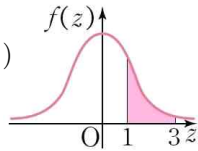
$$\textcircled{2} P(m - 2.58\sigma \leq X \leq m + 2.58\sigma) = P(-2.58 \leq Z \leq 2.58) = 0.99 \quad (\approx 99\%)$$



3. 표준정규분포표를 이용한 확률의 계산

(1) $P(a \leq Z \leq b)$

$$= P(0 \leq Z \leq b) - P(0 \leq Z \leq a)$$



예) $P(1 \leq Z \leq 3)$

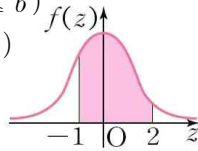
$$= P(0 \leq Z \leq 3) - P(0 \leq Z \leq 1)$$

$$= 0.4987 - 0.3413 = 0.1574$$

(2) $P(-a \leq Z \leq b)$

$$= P(-a \leq Z \leq 0) + P(0 \leq Z \leq b)$$

$$= P(0 \leq Z \leq a) + P(0 \leq Z \leq b)$$

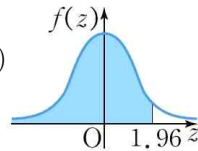


예) $P(-1 \leq Z \leq 2)$

$$= P(0 \leq Z \leq 1) + P(0 \leq Z \leq 2)$$

$$= 0.3413 + 0.4772 = 0.8185$$

(3) $P(Z \leq a) = 0.5 + P(0 \leq Z \leq a)$

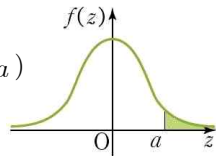


예) $P(Z \leq 1.96)$

$$= 0.5 + P(0 \leq Z \leq 1.96)$$

$$= 0.5 + 0.4750 = 0.9750$$

(4) $P(Z \geq a) = 0.5 - P(0 \leq Z \leq a)$

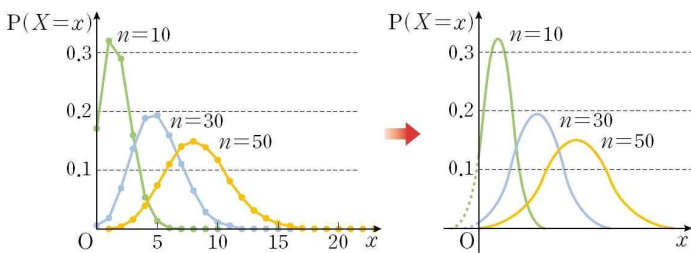


4. 이항분포와 정규분포의 관계

(1) 확률변수 X 가 이항분포 $B(n, p)$ 를 따를 때,

n 이 충분히 크면 X 는 근사적으로

정규분포 $N(np, npq)$ 를 따른다. (단, $p + q = 1$)

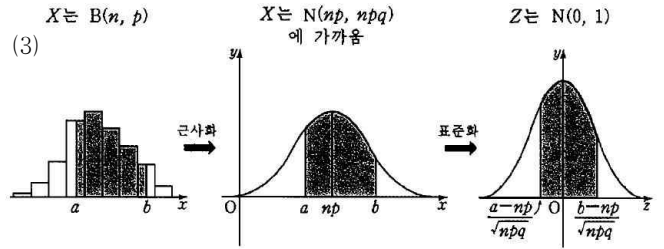


(2) $B(n, p) \Rightarrow N(np, npq) \Rightarrow N(0, 1)$

$n \rightarrow \text{大}$

$$Z = \frac{X - np}{\sqrt{npq}}$$

☑ $np \geq 5, nq \geq 5$ 일 때, n 이 충분히 크다고 간주



$$(4) P(a \leq X \leq b) = \sum_{x=a}^b {}_n C_x p^x q^{n-x}$$

$$= P\left(\frac{a - np}{\sqrt{npq}} \leq Z \leq \frac{b - np}{\sqrt{npq}}\right)$$

< 합격자의 최저 점수 구하기 >

[문제] 어느 대학에서는 400 점 만점의 입학 전형 자료 점수로 신입생 242 명을 선발한다. 이 대학에 지원한 학생 1,000 명의 입학 전형 자료 점수는 정규분포 $N(300, 20^2)$ 을 따른다고 할 때, 합격자의 최저 점수를 구하여 보자.

[1단계] 문제를 이해하여 보자.

(1) 수험생의 점수를 확률변수 X 라고 할 때, X 의 확률분포를 말하여라.

⇒ X 는 정규분포 $N(300, 20^2)$ 을 따른다.

(2) 신입생 1,000 명 중에서 242 명이 선발될 확률을 구하여라.

$$\Rightarrow \frac{242}{1000} = 0.242$$

[2단계] 계획을 세워 보자.

(1) 합격자의 최저 점수를 c 라고 놓을 때, 확률 $P(X \geq c)$ 를 말하여라.

$$\Rightarrow 0.242$$

(2) 확률 $P(X \geq c)$ 를 표준화하여라.

$$\Rightarrow P(X \geq c) = P\left(Z \geq \frac{c - 300}{20}\right)$$

[3단계] 문제를 풀어 보자.

(1) [2단계]의 (1), (2)를 이용하여 식을 세우고, 표준정규분포표를 이용하여 c 의 값을 구하여라.

$$\Rightarrow P\left(Z \geq \frac{c - 300}{20}\right) = 0.242$$

$$\frac{c - 300}{20} = k \quad (> 0)$$

$$0.5 - P(0 \leq Z \leq k) = 0.242$$

$$\therefore P(0 \leq Z \leq k) = 0.258$$

$$P(0 \leq Z \leq 0.7) = 0.258 \text{ 이므로}$$

$$k = \frac{c - 300}{20} = 0.7 \quad \therefore k = 314$$

(2) 합격자의 최저 점수는 몇 점인지 말하여라.

$$\Rightarrow 314 \text{ 점}$$

< 평균으로의 회귀(회귀) - 골턴(Galton) >

1. 의미 : 맨 처음에는 평균을 훨씬 뛰어넘지만

두 번째는 평균값 이하로 되돌아오는 현상.

2. 설명 : 주사위를 한 번 던졌을 때 나오는 눈의 기댓값은 3.5 이다. 주사위를 처음 던졌을 때 기댓값보다 높은 숫자가 나왔다고 하면 다음 번엔 작은 숫자가 나올 확률이 높다. 예를 들어, 4 의 눈이 나왔다고 하자. 다음 번에 4 보다 작은 수 1, 2, 3 이 나올 확률은 $\frac{1}{2}$ 이고,

5 나 6 이 나올 확률은 $\frac{1}{3}$ 로 작은 숫자가 나올 확률이

크다. 즉, 한 번 평균을 웃도는 값이 나오면 다음 번엔 평균 이하의 값이 나오는 것이 일반적인 이야기이다.

반대로, 처음 던졌을 때 2 와 같이 평균에 못 미치는 값이 나오면 두 번째는 더 큰 숫자가 나올 가능성이 커지게 된다.

2. 어떤 식당에서 식사를 하였는데, 그 맛이 아주 좋아 다시 그곳을 찾았다가 실망한 경험.

3. 골턴(Galton) : 아들의 키와 아버지의 키의 관계에서 아들의 키는 평균 키로 회귀하려는 경향이 있음을 알아내었다.

아버지의 키가 클 때, 유전적 영향으로 자식의 키도 크게 된다면 몇 세대만 내려가더라도 인류는 한쪽 극단에 키가 무척 큰 사람들과 다른 극단에 키가 무척 작은 사람들로 양분될 것이다.

하지만 세대를 거듭하면서 안정적인 상태를 유지하는 '평균으로의 회귀'때문에 그러한 현상은 벌어지지 않는다.

< 정규분포의 역사 >

정규분포를 최초로 발견한 사람은 드 무아브르(De Moivre, A. : 1667 ~ 1754)이다. 1733 년 그가 라틴어로 쓴 논문 중 「이항식 $(a + b)^n$ 을 급수로 전개하여 이 항들의 합을 근사하는 방법」에는, 이항분포의 확률 ${}_nC_r p^r q^{n-r}$ 은 n 과 r 이 클 때 계산하기가 매우 곤란하므로 그 근사값을 정규분포를 이용하여 구한 내용이 들어 있다.

한편, 가우스(Gauss, K. : 1777~1855)는 이항분포의 극한분포가 아니라 ‘오차분포’로서 정규분포를 제시하였다. 정규분포는 평균값을 중앙으로 하여 좌우대칭인 종 모양을 이루는 것으로 신장(身長)의 분포, 지능(知能)의 분포 등 그 예는 많다. 가우스가 측정오차의 분포에서 그 중요성을 강조하였기 때문에 이것을 가우스분포·오차분포라고도 하며, 그 곡선을 가우스곡선 또는 오차곡선이라고도 한다. 또한 케틀레가 정규분포를 통계에 이용하였으므로 이것을 케틀레곡선이라고도 한다.

< 확률분포 >

이산확률변수 X 의 확률질량함수 $P(X=x)$ 가 나타내는 확률분포를 이산확률분포, 연속확률변수 X 의 확률밀도함수 $f(x)$ 가 나타내는 확률분포를 연속확률분포라고 한다. 이산확률분포에는 이항분포, 초기하분포, 기하분포, 푸아송분포 등이 있고, 연속확률분포에는 균등분포, 지수분포, 정규분포 등이 있다. 이 중에서 고등학교 교육과정에서는 이항분포와 정규분포를 주로 다룬다.

1. 이항분포

이항분포는 베르누이 시행으로부터 얻어진다. 베르누이 시행은 시행의 결과가 두 가지로 구분되는 시행이다. 예를 들어 동전을 던지는 시행은 그 결과가 앞면 또는 뒷면 뿐이므로 베르누이 시행이다. 또 농구 경기에서 자유투를 던지는 시행은 그 결과가 성공 또는 실패 뿐이므로 베르누이 시행이다. 반면에 주사위를 던지는 시행은 그 결과가 6가지이므로 베르누이 시행이 아니지만 1의 눈이 나오느냐 안 나오느냐로 구분한다면 베르누이 시행이 된다. 베르누이 시행은 스위스의 수학자 베르누이(Bernoulli, J. ; 1654~1705)의 이름에서 유래되었다.

일반적으로 베르누이 시행에서의 두 결과(단순 사건)를 성공 s 와 실패 f 로 생각하면 베르누이 시행의 표본공간 S 는 $S = \{s, f\}$ 로 나타낼 수 있다. 이때 한 번의 베르누이 시행에서 성공 횟수를 확률변수 X 라고 하면

$$s \xrightarrow{X} 1 \quad (X(s) = 1)$$

$$f \xrightarrow{X} 0 \quad (X(f) = 0)$$

과 같이 대응시킬 수 있다.

일반적으로 성공 확률이 p 인 베르누이 시행을 n 번 독립적으로 반복시행하였을 때, 성공 횟수를 확률변수 X 라고

하면 X 가 취하는 값은 $0, 1, 2, \dots, n$ 이고 각 값을 취할 확률은 다음과 같다.

$$P(X=x) = {}_nC_x p^x q^{n-x} \quad (q=1-p, x=0, 1, 2, \dots, n)$$

이와 같은 확률질량함수를 갖는 확률변수 X 는 이항분포에 따른다고 한다.

2. 정규분포

정규분포를 최초로 발견한 사람은 드무아브르(de Moivre, A. ; 1667~1754)이다. 1733년 그가 라틴어로 쓴 논문은 “우연론” 제2판(1738) 및 제3판(1756)에 영문으로 수록되어 있는데 그 논문의 제목은 “이항식 $(a+b)^n$ 을 급수로 전개하여 이 항들의 합을 근사하는 방법”이다.

이항분포의 확률 ${}_nC_x p^x q^{n-x}$ 의 계산은 n 과 p 가 클 때 매우 복잡하다. 따라서 이 확률의 근삿값을 구할 필요가 있었는데 드무아브르와 라플라스(Laplace, P. S. ; 1749~1827)는 n 이 충분히 클 때, 이항분포는 정규분포에 근사한다는 것을 발견하였다. 즉,

$${}_nC_x p^x q^{n-x} = \frac{1}{\sqrt{2\pi} \sqrt{npq}} e^{-\frac{(x-np)^2}{2npq}}$$

한편 가우스(Gauss, K. F. ; 1777~1855)는 이항분포의 극한 분포가 아니라 ‘오차 분포’로써 정규분포를 제시하였다. 가우스는 1809년 그의 논문 “Theoria motus corporum coelestium in sectionibus conicis solem ambientium”에서 다음과 같이 언급하였다.

“행성 궤도의 이론적인 값과 실제 값의 차이인 오차는 참값(이론 값) 주위에 분포되어 있고 같은 정도로 일어날 것이 예상되므로 각 오차 ϵ 이 일어날 확률을 확률밀도함수 $\phi(\epsilon)$ 로 나타내면 $\phi(\epsilon)$ 의 값은 $\epsilon=0$ 에서 최대가 되고, ϵ 의 부호는 다르지만 절댓값이 같은 ϵ 의 값에 대해서는 $\phi(\epsilon)$ 의 값은 같고, ϵ 이 최대 오차 또는 그것보다 큰 값을 가질 때는 $\phi(\epsilon)$ 은 0이 될 것이라고 가정할 수가 있다.”

이와 같은 가정 하에서 가우스는

$$\phi(\epsilon) = \frac{h}{\sqrt{\pi}} e^{-h^2 \epsilon^2} \quad \left(\text{단, } h = \frac{1}{\sqrt{2\sigma^2}} \right)$$

을 유도하였다. 이때 h 는 관측값의 정도이다.

19세기에 천문학에서의 방대한 관측 자료가 정규분포에 잘 적합화된다는 경험적 사실에 기초하여 관측값의 편차는 정규분포에 의하여 표현될 것이라는 정규분포 신앙이 지배적이었다. 정규분포에 잘 적합화 되지 않으면 자료가 부족하거나 편의되어 있기 때문이라 생각하게 되었고 정규분포(normal distribution)가 아닌 그 밖의 것은 비정상적인 분포(abnormal distribution)라고 생각하였다.

우생학의 창시자 골턴(Galton, F. ; 1822~1911)은 1889년 그의 책 “자연 유전”에서 다음과 같이 정규분포를 찬미

하였다.

“오차의 법칙으로 나타내지는 우주 질서의 놀라운 형식만큼 감명을 주는 것은 아무 것도 없다. 만약 고대 그리스인이 이 법칙을 알고 있었다면 그것은 의인화되고 신격화되었을 것이다. 이 법칙은 매우 무질서한 가운데서 조용히 군림하고 있다. 집단이 클수록 또 외견상 무질서가 심할수록 이 법칙의 지배력은 크다. 그것은 이유 없는 가장 고귀한 법칙이다. 무질서한 요소에서 대 표본을 추출하여 크기순으로 나열하면 생각지도 않는 매우 아름다운 질서의 형식이 숨어 있었다는 것이 분명해진다.”

그러나 웰던(Weldon, W. F. R. ; 1860~1906), 피어슨(Pearson, K. ; 1857~1936) 등은 “도수분포 중에는 비틀림이 있는 것도 있고, 비대칭인 것도 있다. 도수분포가 종모양이 되지 않는 것은 자료 수집 방법이 나쁘거나 자료수가 부족한 것이 아니라 도수분포 자체의 기본 특성이다.”라고 하였고, 여러 분야에서 정규분포에 적합화 되지 않는 예가 많이 발견되면서 정규분포에 대한 신앙은 점차 무너져갔다.

< 정규분포의 평균과 분산 >

정규분포의 평균과 분산은 다음과 같이 확인할 수 있다. 확률변수 Z 의 확률밀도함수

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

에서 다음 관계가 성립한다.

$$\phi'(z) = -z\phi(z)$$

$$\phi''(z) = -\phi(z) - z\phi'(z) = \phi(z)(z^2 - 1)$$

따라서

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} z\phi(z)dz = -\int_{-\infty}^{\infty} \phi'(z)dz \\ &= [\phi(z)]_{-\infty}^{\infty} = 0 \end{aligned}$$

$$\begin{aligned} V(Z) &= \int_{-\infty}^{\infty} z^2\phi(z)dz - 0^2 \\ &= \int_{-\infty}^{\infty} \{\phi''(z) + \phi(z)\}dz = 0 + 1 = 1 \end{aligned}$$

따라서 표준정규분포의 확률변수 $Z = \frac{X-m}{\sigma}$ 에서

$X = m + \sigma Z$ 이므로 다음이 성립한다.

$$E(X) = E(m + \sigma Z) = m + \sigma E(Z) = m$$

$$V(X) = V(m + \sigma Z) = \sigma^2 V(Z) = \sigma^2$$

< 통계적 추정 >

1. 모집단 $X : N(m, \sigma^2)$

↓ 크기가 n 인 임의표본을 복원추출

표본평균 $\bar{X} : N\left(m, \frac{\sigma^2}{n}\right)$

2. 모집단 X 의 표준편차가 σ 일 때, 모평균 m 의 범위?

↓ 표본의 크기가 n

$$\text{표본평균 } \bar{X} \xrightarrow{\quad} Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}}$$

신뢰도 $\alpha\%$ 의 모평균 m 의 신뢰구간

(1) $P(|Z| \leq 1.96) = 0.95$

$$\hookrightarrow \left| \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \right| \leq 1.96$$

$$\Leftrightarrow |\bar{X} - m| \leq 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow |m - \bar{X}| \leq 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq m - \bar{X} \leq 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

(2) 모평균 m 의 신뢰구간

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

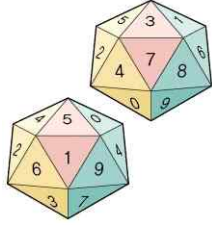
☑ 모평균 m 이 이 범위에 포함될 확률 95%

(3) 신뢰구간의 길이 : $l = 2 \times 1.96 \cdot \frac{\sigma}{\sqrt{n}}$

(4) 오차 $|m - \bar{X}|$ 의 한계 : $1.96 \cdot \frac{\sigma}{\sqrt{n}}$

3. 통계적 추정

1 모집단과 표본

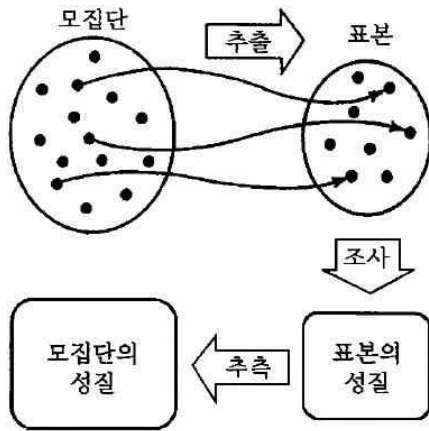


1. 통계조사

(1) 전수조사 : 조사 대상의 자료 전체를 조사하는 것

(2) 표본조사 : 조사 대상 중에서 일부만을 조사하여 전체에 대한 특성을 추측

- ① 모집단 : 조사하고자 하는 집단 전체
- ② 추출 : 모집단에서 표본을 뽑는 것
- ③ 표본 : 모집단에서 뽑은 대상의 모임
- ④ 표본의 크기 : 표본에 포함된 대상의 개수
- ⑤ 표본평균 : 모집단에서 추출된 n 개 표본의 평균



⑥ 임의추출 : 모집단의 각 대상이 추출될 확률이 동일하게 되도록 표본을 추출하는 방법
⇨ 임의표본

⑦ 복원추출 : 한 번 추출된 원소를 다시 되돌려 놓은 후 다음 원소를 뽑는 방법
↔ 비복원추출

2. 표본평균의 분포

모평균 m , 모분산 σ^2 , 모표준편차 σ 인 모집단 X 에서
⇨ 크기가 n 인 임의표본을 복원추출할 때,
표본평균 \bar{X} 에 대하여

(a) 크기가 n 인 표본 X_1, X_2, \dots, X_n 을 임의추출

① 표본평균 : $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$

$$\textcircled{2} \text{ 표본분산 : } S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

$$\textcircled{3} \text{ 표본표준편차 : } S (\geq 0)$$

☑ 표본분산 S^2 은 자료의 분산과는 달리 편차의 제곱의

합 $\sum_{k=1}^n (X_k - \bar{X})^2$ 을 $n-1$ 로 나눈다. 이는

표본분산과 모분산의 차이를 줄이기 위한 것이다.

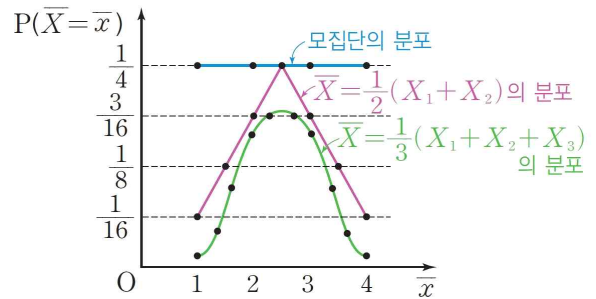
(1) 표본평균 \bar{X} 의 평균 : $E(\bar{X}) = m$

(2) 표본평균 \bar{X} 의 분산 : $V(\bar{X}) = \frac{\sigma^2}{n}$

(3) 표본평균 \bar{X} 의 표준편차 : $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

3. 모평균 m , 모분산 σ^2 , 모표준편차 σ 인 모집단 X 에서
⇨ 크기가 n 인 임의표본을 복원추출할 때,

(1) 표본평균 \bar{X} 의 확률분포는 표본의 크기가 커질수록 분산이 작아지므로 정규분포에 가까워진다.



(2) 모집단이 정규분포 $\Rightarrow n$ 의 크기에 관계없이

\bar{X} 는 항상 정규분포 $N\left(m, \frac{\sigma^2}{n}\right)$ 을 따른다.

(3) [중심극한정리] 표본의 크기 n 이 충분히 큰 수

$\Rightarrow \bar{X}$ 는 근사적으로 정규분포 $N\left(m, \frac{\sigma^2}{n}\right)$ 을 따른다.

☑ 표본의 크기 n 이 충분히 크다 $\Leftrightarrow n \geq 30$

(4) 표본평균 \bar{X} 의 표준화

표본평균 $\bar{X} : N\left(m, \frac{\sigma^2}{n}\right) \Rightarrow \Rightarrow \Rightarrow Z : N(0, 1)$

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}}$$

4. 표본비율의 분포

- (o) ① 모비율 : 시청률, 실업률, 불량률 등과 같은 어떤 사건에 대한 모집단에서의 비율
 \Rightarrow (기호) p

☐ p 는 **population ratio**(모집단 비율)의 첫글자

- ② 표본비율 : 그 사건에 대한 모집단에서 임의추출한 표본에서의 비율
 \Rightarrow (기호) $\hat{p} \rightarrow$ "피햇(p hat)"

- ③ 확률변수 X : 크기가 n 인 표본에서 어떤 특성을 갖는 사건의 개수(횟수)라고 할 때, 그 사건에 대한 표본비율은
 $\Rightarrow \hat{p} = \frac{X}{n}$

- ④ 표본비율 $\hat{p} = \frac{X}{n}$ 인 확률변수 $X : B(n, p)$

☐ X : 확률이 p 인 n 번의 독립시행에서의 횟수
 $E(X) = np, V(X) = npq$

- (1) 표본비율 \hat{p} 의 평균 :

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} \cdot np = p$$

- (2) 표본비율 \hat{p} 의 분산 :

$$V(\hat{p}) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} \cdot npq = \frac{pq}{n}$$

(단, $p+q=1$)

- (3) 표본비율 \hat{p} 의 표준편차 :

$$\sigma(\hat{p}) = \sqrt{V(\hat{p})} = \sqrt{\frac{pq}{n}}$$

5. 모비율이 p 인 모집단에서

↓ 크기가 n 인 표본을 임의추출했을 때,

표본비율 $\hat{p} = \frac{X}{n}$ 의 확률변수 $X : B(n, p)$

- (1) n 이 충분히 클 때 $\Leftrightarrow np \geq 5, nq \geq 5$

$\Rightarrow \hat{p}$ 의 분포는 근사적으로

정규분포 $N\left(p, \frac{pq}{n}\right)$ 을 따른다.

(단, $p+q=1$)

- (2) 표본비율 \hat{p} 의 표준화

$$\text{표본비율 } \hat{p} : N\left(p, \frac{pq}{n}\right) \Rightarrow \Rightarrow \Rightarrow Z : N(0, 1)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

- 예) 어느 도시 성인 남성의 약 20%가 복부 비만이라고 한다. 그 도시에 있는 한 기업의 직원 100 명을 임의 추출하였을 때, 이 중 복부 비만인 사람이 20% 이상 30%이하일 확률을 구하여라.

\Rightarrow 직원 100 명 중 복부 비만인 사람의 비율 : \hat{p}

$$E(\hat{p}) = p = 0.2$$

$$V(\hat{p}) = \frac{pq}{n} = \frac{0.2 \times 0.8}{100} = \frac{0.16}{100}$$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \frac{0.4}{10} = 0.04$$

n 이 충분히 크면 \hat{p} 의 분포는 근사적으로 정규분포를 따르고 이것을 표준화하면

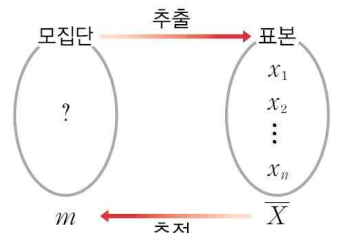
$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{\hat{p} - 0.2}{0.04}$$

$$\therefore P(0.2 \leq \hat{p} \leq 0.3)$$

$$= P\left(\frac{0.2 - 0.2}{0.04} \leq Z \leq \frac{0.3 - 0.2}{0.04}\right)$$

$$= P(0 \leq Z \leq 2.5) = 0.4938$$

2 모평균의 추정



0. 추정과 신뢰도

- (1) 추정 : 표본에서 얻은 자료를 이용하여 모평균 또는 모비율 등과 같은 모집단의 참값을 확률적으로 추측하는 것

- (2) 신뢰도 : 추정이 적중할 확률

- (3) 신뢰구간 : 모평균 m 또는 모비율 p 가 존재할 것으로 추정되는 구간

1. 표본평균 \bar{X} 의 분포

$$\bar{X} : N\left(m, \frac{\sigma^2}{n}\right) \Rightarrow \Rightarrow \Rightarrow Z : N(0, 1)$$

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}}$$

☑ 모표준편차 σ 를 모르는 경우 \Rightarrow 모집단이 정규분포를 따르거나 n 이 클 때 ($n \geq 30$) 표본표준편차를 사용!

\because 표본의 크기가 충분히 클 때, 모표준편차 대신 표본표준편차를 이용하여 모평균의 신뢰구간을 구할 수 있는데 이는 표본의 크기가 커지면 큰 수의 법칙에 의하여 표본표준편차가 모표준편차에 가까워지기 때문이다.

2. 신뢰도 95 %의 모평균 m 의 신뢰구간

$$(1) P(|Z| \leq 1.96) = 0.95$$

$$\hookrightarrow \left| \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \right| \leq 1.96$$

$$\Leftrightarrow |\bar{X} - m| \leq 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

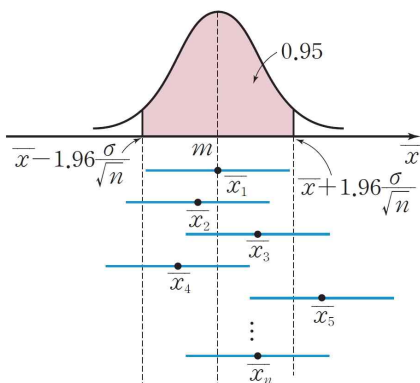
$$\Leftrightarrow |m - \bar{X}| \leq 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq m - \bar{X} \leq 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

(2) 모평균 m 의 신뢰구간

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

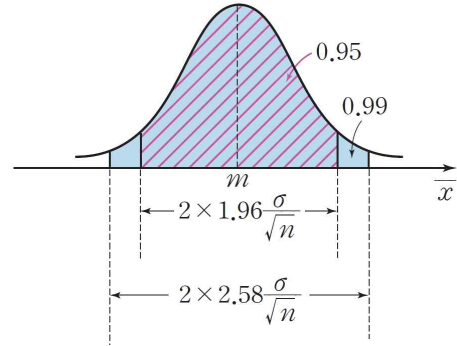
☑ 모평균 m 이 이 신뢰구간에 포함될 확률 95 %



\Leftrightarrow 크기가 n 인 표본을 여러 번 추출하여 신뢰구간을 각각 구해보면 그 중에서 약 95 %는 모평균 m 을 포함할 것으로 기대된다는 뜻.

$$(3) \text{신뢰구간의 길이} : l = 2 \times 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$(4) \text{오차 } |m - \bar{X}| \text{ 의 한계} : 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$



3. 신뢰도 99 %의 모평균 m 의 신뢰구간

$$(1) P(|Z| \leq 2.58) = 0.99$$

$$\hookrightarrow \left| \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \right| \leq 2.58$$

$$\Leftrightarrow |\bar{X} - m| \leq 2.58 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow |m - \bar{X}| \leq 2.58 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow 2.58 \cdot \frac{\sigma}{\sqrt{n}} \leq m - \bar{X} \leq 2.58 \cdot \frac{\sigma}{\sqrt{n}}$$

(2) 모평균 m 의 신뢰구간

$$\bar{X} - 2.58 \cdot \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + 2.58 \cdot \frac{\sigma}{\sqrt{n}}$$

☑ 모평균 m 이 이 신뢰구간에 포함될 확률은 99 %

$$(3) \text{신뢰구간의 길이} : l = 2 \times 2.58 \cdot \frac{\sigma}{\sqrt{n}}$$

$$(4) \text{오차 } |m - \bar{X}| \text{ 의 한계} : 2.58 \cdot \frac{\sigma}{\sqrt{n}}$$

4. 신뢰도 α %의 모평균 m 의 신뢰구간

$$(1) P(|Z| \leq k) = \frac{\alpha}{100}$$

$$\hookrightarrow \left| \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \right| \leq k$$

$$\Leftrightarrow |\bar{X} - m| \leq k \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow |m - \bar{X}| \leq k \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow k \cdot \frac{\sigma}{\sqrt{n}} \leq m - \bar{X} \leq k \cdot \frac{\sigma}{\sqrt{n}}$$

(2) 모평균 m 의 신뢰구간

$$\bar{X} - k \cdot \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + k \cdot \frac{\sigma}{\sqrt{n}}$$

☐ 모평균 m 이 이 신뢰구간에 포함될 확률은 $\alpha\%$

(3) 신뢰구간의 길이(폭) : $l = 2k \cdot \frac{\sigma}{\sqrt{n}}$

(4) 오차 $|m - \bar{X}|$ 의 한계 : $k \cdot \frac{\sigma}{\sqrt{n}}$

(5) 크기가 n 인 표본의 추출을 반복할 때,
그 신뢰구간들의 약 $\alpha\%$ 가 모평균 m 을 포함할
것으로 기대할 수 있다.

5. 신뢰도, 표본의 크기, 신뢰구간 사이의 관계
 $\alpha\%$ 신뢰도의 모평균 m 의 신뢰구간 l 에서

$$l = 2k \cdot \frac{\sigma}{\sqrt{n}} \quad \left(\text{단, } P(|Z| \leq k) = \frac{\alpha}{100} \right)$$

(1) 표본의 크기 n 이 일정할 때,

$$\alpha \uparrow \iff k \uparrow \iff l \uparrow$$

∴ [신뢰도가 높아지면 신뢰구간은 길어진다.
[신뢰구간을 짧게 하려면 신뢰도를 낮춘다.

(2) 신뢰도 $\alpha\%$ 가 일정할 때,

$$n \uparrow \iff l \downarrow$$

∴ [표본의 크기가 커지면 신뢰구간은 짧아진다.
[신뢰구간을 짧게 하려면 표본을 크게 잡는다.

3 모비율의 추정

1. 모비율이 p 인 모집단에서

↓ 크기가 n 인 표본을 임의추출했을 때,

표본비율 $\hat{p} = \frac{X}{n}$ 의 확률변수 $X : B(n, p)$

(단, $\hat{p} + \hat{q} = 1$)

(1) n 이 충분히 클 때 $\Leftrightarrow n\hat{p} \geq 5, n\hat{q} \geq 5$

$\Rightarrow \hat{p}$ 의 분포는 근사적으로

정규분포 $N\left(p, \frac{pq}{n}\right)$ 을 따른다.

(2) 표본비율 \hat{p} 의 표준화

횃수 $X : N\left(m, \frac{\sigma^2}{n}\right) \Rightarrow \Rightarrow \Rightarrow Z : N(0, 1)$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

(3) n 이 클 때에는

모비율 p, q 대신에 표본비율 \hat{p}, \hat{q} 를 사용

횃수 $X : N\left(m, \frac{\sigma^2}{n}\right) \Rightarrow \Rightarrow \Rightarrow Z : N(0, 1)$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

2. 신뢰도 95%의 모비율 p 의 신뢰구간

$$(1) P(|Z| \leq 1.96) = P\left(|\hat{p} - p| \leq 1.96 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 0.95$$

(2) 모비율 p 의 신뢰구간

$$\hat{p} - 1.96 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + 1.96 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

☐ 모비율 p 가 이 신뢰구간에 포함될 확률은 95%

(3) 신뢰구간의 길이 : $l = 2 \times 1.96 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$

(4) 최대 허용 오차 한계 : $1.96 \cdot \sqrt{\frac{1}{4n}}$

☐ 표본비율의 표준편차 $\sqrt{\frac{\hat{p}\hat{q}}{n}}$ 은 n 의 값이 일정하고

$$\hat{p} = \hat{q} = \frac{1}{2} \text{ 일 때, 변동폭의 최댓값은 } \sqrt{\frac{1}{4n}}$$

∴ $\hat{p} + \hat{q} = 1$ 일 때, $\hat{p}\hat{q}$ 의 최댓값은 산술·기하평균 사용

3. 신뢰도 99%의 모비율 p 의 신뢰구간

$$(1) P(|Z| \leq 2.58) = P\left(|\hat{p} - p| \leq 2.58 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 0.99$$

(2) 모비율 p 의 신뢰구간

$$\hat{p} - 2.58 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + 2.58 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

☑ 모비율 p 가 이 범위에 포함될 확률은 99 %

(3) 신뢰구간의 길이 : $l = 2 \times 2.58 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$

(4) 최대 허용 오차 한계 : $2.58 \cdot \sqrt{\frac{1}{4n}}$

4. 신뢰도 99 %의 모비율 p 의 신뢰구간

$$(1) P(|Z| \leq k) = P\left(|\hat{p} - p| \leq k \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 0.99$$

(2) 모비율 p 의 신뢰구간

$$\hat{p} - k \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + k \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

☑ 모비율 p 가 이 범위에 포함될 확률은 99 %

(3) 신뢰구간의 길이 : $l = 2 \times k \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$

(4) 최대 허용 오차 한계 : $k \cdot \sqrt{\frac{1}{4n}}$

(5) 크기가 n 인 표본의 추출을 반복할 때,
그 신뢰구간들의 약 α %가 모비율 p 를 포함할 것으로 기대할 수 있다.

5. 신뢰도, 표본의 크기, 신뢰구간 사이의 관계

α % 신뢰도의 모비율 p 의 신뢰구간 l 에서

$$l = 2k \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \left(\text{단, } P(|Z| \leq k) = \frac{\alpha}{100} \right)$$

(1) 표본의 크기 n 이 일정할 때,

$$\alpha \uparrow \iff k \uparrow \iff l \uparrow$$

∴ | 신뢰도가 높아지면 신뢰구간은 길어진다.
| 신뢰구간을 짧게 하려면 신뢰도를 낮춘다.

(2) 신뢰도 α %가 일정할 때,

$$n \uparrow \iff l \downarrow$$

∴ | 표본의 크기가 커지면 신뢰구간은 짧아진다.
| 신뢰구간을 짧게 하려면 표본을 크게 잡는다.

< 전수조사와 표본조사의 경제적 가치 >

상대적으로 작은 표본에 의한 표본조사가 전수조사와 거의 동일한 결과를 얻을 수 있다는 사실을 일반인들은 인정하기 어렵다. 그러나 실제 표본조사에 의한 오차가 그리 크지 않고, 전수조사에 의한 비용과 시간을 고려하면 표본조사의 경제적 가치는 훨씬 크다는 것은 여러 연구에서 입증된 사실이다.

미국의 철도는 우리나라의 철도와는 달리 지역과 노선에 따라 철도의 소유주가 다르기 때문에 어떤 화물을 여러 노선을 경유하여 운송하게 되면 총화물 운임을 각 노선을 소유한 철도 회사가 배분하여 가진다. 이때, 각 회사의 수익액은 화물의 경로, 운임, 품목 등이 자세하게 기록된 화물 운임장으로 계산한다. 그러나 모든 화물 운임장으로부터 각 회사의 수익액을 계산한다는 것은 비용도 많이 들고 매우 번거롭기 때문에 표본을 추출하여 그에 따라 수익액을 배분한다.

C 철도 회사에서는 표본에 의해서도 수익액이 정확히 계산되는지 분석해 보기로 하고, 표본조사와 전수조사를 모두 시행하여 그 결과를 비교 분석하였다. 다음은 6개월 동안 총화물 운임장인 23,000 장에서 2,000 장을 임의 추출하여 두 결과를 비교한 결과이다.

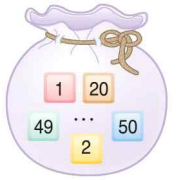
전수조사에 의한 C 회사의 배분액	64,651 불
표본조사에 의한 C 회사의 배분액	64,568 불
차액	83 불

위의 표에서 보면 전체의 약 9% 정도의 표본에 의한 수익액 추정액과 전수조사에 의한 수익액과의 차이는 불과 83 불이다. 전수조사에 들어간 비용은 약 5,000 불이었고, 표본조사에 들어간 비용은 1,000 불 미만이었다고 한다. C 회사의 경우는 표본조사로 83 불을 손해봤지만 다음 표본조사에서는 이익을 보는 경우도 있을 수 있으므로 표본조사로 인해 입게 되는 손해는 크게 없을 것이다.

따라서, 전수조사보다는 표본조사에 의한 수익액 배분이 경제적 측면에서 훨씬 유리한 방법이라 할 수 있다.

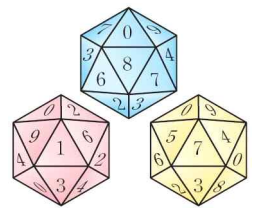
1. 제비뽑기를 이용한 임의추출

학생 수가 50 명인 동아리에서 5 명을 임의추출하기 위하여 50 장의 종이에 1 에서 50 까지의 번호를 적은 제비를 주머니에 넣는다. 이 제비를 잘 섞은 다음, 임의로 5 장을 추출하고, 추출된 제비에 해당하는 번호를 가진 학생을 표본으로 택하면 이것은 임의추출법에 의하여 추출된 것이다.



2. 난수 주사위를 이용한 임의추출

난수 주사위는 오른쪽 그림과 같이 0 에서 9 까지 숫자를 두 번씩 사용하여 정이십면체의 각 면에 숫자 하나씩을 새긴 것이다.



난수 주사위를 이용하여 300 명 중에서 5 명을 임의추출하여 보자.

- (1) 각 사람에게 000 에서 299 까지의 번호를 붙인다.
- (2) 서로 다른 색의 난수 주사위 3 개를 준비한다.
이를테면 빨간색, 파란색, 노란색을 준비한다. 이때, 빨간색, 파란색, 노란색 주사위에서 나오는 숫자를 각각 100 의 자리 숫자, 10 의 자리 숫자, 1 의 자리 숫자로 정하고 3 개의 주사위를 동시에 던지면 000 에서 999 까지의 수를 얻을 수 있다.
- (3) 299 이하의 수가 5 개 나오면 그 번호를 가진 다섯 사람을 순서대로 택한다.

3. 난수표를 이용한 임의추출

난수표는 0 에서 9 까지의 숫자를 임의로 배열한 표이다. 부록에 있는 난수표를 이용하여 50 명 중에서 5 명을 임의추출하는 방법을 알아보자.

41 10 50 81 22	94 80 71 10 68	23 58 20
13 49 57 94 72	78 92 78 78 04	17 00 92
33 87 89 24 77	65 37 12 38 63	76 49 69
15 91 02 97 10	37 14 47 47 79	81 63 34
37 94 89 58 24	29 22 39 42 66	95 14 63
48 06 32 88 07	06 19 13 11 04	45 95 73
92 65 65 69 32	05 63 75 76 57	26 10 31
48 66 49 80 78	34 30 47 61 73	44 31 65
23 50 07 82 24	34 88 84 90 39	20 46 32
47 02 38 86 81	59 77 46 17 55	54 59 00
39 65 34 38 46	26 95 15 80 70	40 06 89
90 36 99 74 53	71 05 53 69 01	49 59 53
46 60 38 92 08	09 16 06 33 02	13 60 78
62 67 74 04 84	75 68 64 11 42	22 88 64
21 17 44 02 71	21 59 79 73 18	24 74 77

- (1) 각 사람에게 00 에서 49 까지의 번호를 붙인다.
- (2) 제비뽑기나 난수 주사위를 이용하여 난수표의 시작하는 행, 열을 정한다. 이를테면 이들이 14 행 6 열일 때 14 행을 따로 쓰면 다음과 같다.

62 67 74 04 84 75 68
64 11 42 22 88 64 ...

(3) 이 행의 6 번째 숫자인 4 부터 두 자리씩 오른쪽으로 나아가면서 쓴다. 이때, 이것이 나타내는 두 자리 수 중 50 이상인 것을 지우면 다음과 같다.

40 48 47 56 86 41 14 22 28 86 ...

(4) 이와 같이 얻은 수 중에서 처음 5 개의 수 40, 48, 47, 41, 14 의 번호를 가진 사람을 택한다.

< KS(Korea Standard) 난수표 >

난수표를 최초로 만든 사람은 팀페트이다. 그 후 여러 종류의 난수표가 나왔는데 이 책의 부록에 있는 KS 난수표는 1963년 박한식 교수가 만든 것으로 우리나라 최초의 난수표이다. KS 난수표는 부록과 같은 것이 모두 40 장으로 구성되어 있다. 따라서 숫자의 개수는 40,000 개이다.

이 난수표를 만들 당시 우리나라에 전자계산기가 없었기 때문에 오른쪽 그림과 같은 룰렛을 돌려서 나온 숫자를 하나씩 기록해 나가는 수작업으로 만들었다. 그런데 난수표는 숫자가 임의로 배열된 것이므로 그 임의성을 검정해야 한다. KS 난수표의 임의성을 검정하기 위하여 한 자리 검정, 두 자리 검정, 게프 검정, 포커 검정을 하였다.



1. 한 자리 검정

이것은 숫자 각각을 한 자리 수로 볼 때 난수표의 숫자들이 임의로 배열되어 있는가를 보는 것이다. 이들 속에 들어 있는 0, 1, ..., 9의 개수를 세어서 이상적인 개수와의 차이를 검정한다. 한장에 1,000 개의 숫자가 들어 있으므로 각각 10 개씩 들어 있는 것이 이상적이다.

2. 두 자리 검정

이것은 숫자를 두 개씩 묶어서 두 자리 수로 볼 때 난수표의 수들이 임의로 배열되어 있는가를 보는 것이다. 이들 속에 있는 00, 01, ..., 99의 개수를 세어서 이상적인 개수

와의 차이를 검정한다. 한 장에 500 개의 수가 들어 있으므로 각각 5 개씩 들어 있는 것이 이상적이다.

3. 게프 검정

이를테면 1,000 개의 숫자의 나열에서 0 과 999 사이에 있는 숫자의 개수에 대한 통계를 내서 이들 개수가 나오는 확률과 대비하여 검정하는 것이다.

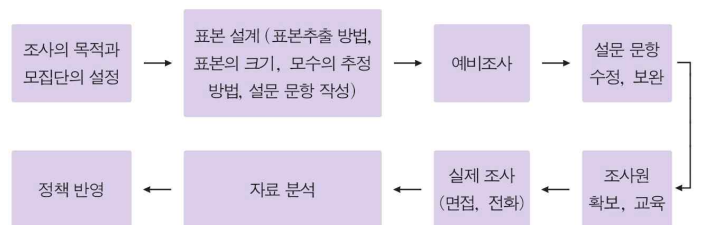
< 여론 조사란 무엇인가? >

민주주의의 기본은 국민의 뜻에 따라 사회와 국가를 운영하는 것이다. 이때, 국민의 뜻을 객관적이고도 구체적으로 파악하는 것이 중요한데, 그 수단으로써 여론 조사가 많이 쓰인다.

영국의 정치학자인 브라이스(Bryce, J.)는 이미 19 세기 말에, 민주주의의 발전에서 마지막 단계는 국민의 의지가 즉시 파악되는 것이라고 예견하였다. 오늘날은 통계적 기법의 발전, 정보 통신기술의 발달, 시민 의식의 성숙 등에 힘입어 어떤 사건에 대한 여론 조사는 거의 순식간에 매우 정확하게 이루어지고 있다. 투표의 종료 시각과 동시에 발표되는 당선자의 예측에서 이러한 사실을 확인할 수 있다. 이러한 면에서 볼 때, 우리 사회는 브라이스가 예견한 민주주의 발전의 마지막 단계에 이르렀다고 할 수 있다.

이와 같이 여론 조사는 처음에는 정치적 문제에서 출발하였으나 현재는 사회·과학 분야뿐만 아니라 기업 경영에서도 많이 활용되고 있다. 오늘날 소비자의 여론에 귀를 기울이지 않는 기업이 하나도 없다고 하여도 과언이 아니다.

여론 조사의 대부분은 표본조사로 이루어지는데, 그 과정을 살펴보면 다음과 같다.



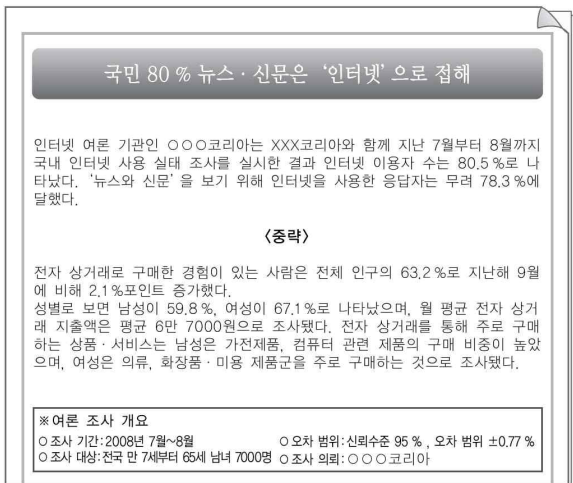
여론 조사의 결과는 정확해야 하고, 신뢰할 수 있어야 한다. 이러한 정확성과 신뢰성을 확보하기 위하여 여론 조사를 시행하고 그 결과를 발표할 때 다음 사항을 제시하여야 한다.

- 필수 사항 : 조사 기관명, 조사 대상, 조사 시기, 유효 표본의 크기와 구체적인 조사 지역

- 권장 사항 : 표본추출 방법, 조사 방법(면접, 전화, 인터넷 등), 설문지, 무응답자의 비율

< 월 평균 전자 상거래 지출액 >

[문제] 다음 기사는 국내의 인터넷 사용 실태를 조사한 결과의 보도 자료의 일부분이다. 월 평균 전자 상거래 지출액에 대한 신뢰도 95 %인 신뢰구간을 구하여 보자. (단, 전자 상거래 지출액은 표준편차가 14,000 원인 정규분포를 따른다.)



[1단계] 문제를 이해하여 보자.

(1) 표본의 크기를 구하여라.

⇒ 7,000

(2) 표본평균 \bar{x} 의 값을 구하여라.

⇒ 67,000

[2단계] 계획을 세워 보자.

표본의 크기가 n 이고, 표본평균의 값이 \bar{x} 일 때, 모평균 m 에 대하여 신뢰도 95 %인 신뢰구간을 말하여라.

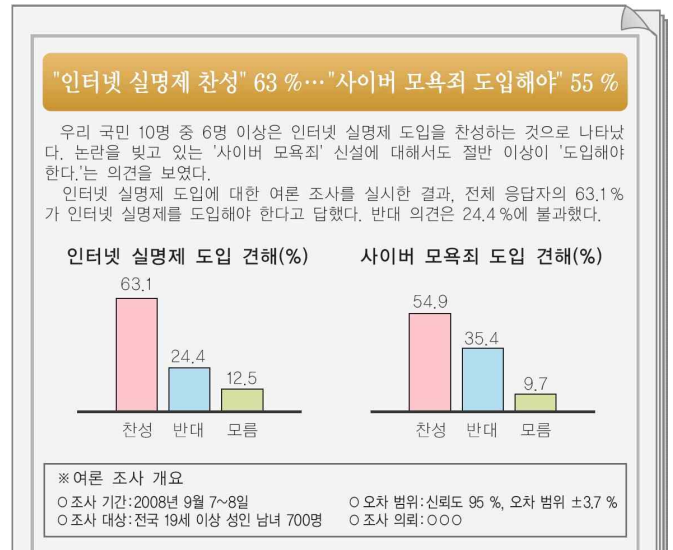
$$\Rightarrow \bar{x} - 1.96 \cdot \frac{14000}{\sqrt{n}} \leq m \leq \bar{x} + 1.96 \cdot \frac{14000}{\sqrt{n}}$$

[3단계] 문제를 풀어 보자.

[2단계]의 신뢰구간에 $n = 7,000$, $\bar{x} = 67,000$ 을 대입하여 신뢰구간을 구하여 보자. (단, 소수 첫째 자리에서 반올림한다.)

$$\Rightarrow 66,672 \leq m \leq 67,328$$

< 인터넷 실명제 도입에 대한 찬성률 >



[문제] 위의 기사는 인터넷 실명제 도입에 대한 여론 조사를 다룬 것이다. 인터넷 실명제 도입을 찬성하는 19 세 이상 전체 국민의 모비율 p 에 대하여 신뢰도 95 %인 신뢰구간을 추정하여라.

[1단계] 문제를 이해하여 보자.

(1) 표본의 크기를 구하여라.

⇒ 700

(2) 인터넷 실명제 도입을 찬성하는 표본비율 \hat{p} 의 값을 구하여라.

⇒ 0.631

[2단계] 계획을 세워 보자.

표본의 크기를 n , 표본비율을 \hat{p} , $\hat{q} = 1 - \hat{p}$ 이라고 할 때, 모비율 p 에 대하여 신뢰도 95 %인 신뢰구간을 말하여라.

$$\Rightarrow \hat{p} - 1.96 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + 1.96 \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

[3단계] 문제를 해결하여 보자.

(1) [2단계]의 신뢰구간에 $n = 700$, $\hat{p} = 0.631$, $\hat{q} = 0.369$ 를 대입하여 신뢰구간을 구하여라. (단, 소수 넷째 자리에서 반올림한다.)

$$\Rightarrow 0.595 \leq p \leq 0.667$$

(2) $1.96 \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq 1.96 \sqrt{\frac{0.5 \times 0.5}{700}}$ 임을 확인하고

$1.96 \sqrt{\frac{0.5 \times 0.5}{700}} \times 100 (\%)$ 의 값을 구하여라.

(단, 소수 둘째 자리에서 반올림한다.)

$\Rightarrow 3.7 \%$

(3) 기사에 주어진 오차 범위 $\pm 3.7 \%$ 를 사용하여 모비율 p 에 대한 신뢰구간을 구하여라.

$\Rightarrow 0.594 \leq p \leq 0.668$

< 통계적 추정 >

통계학은 어떤 현상에 대하여 관측한 자료를 통해 그 현상 속에 내재된 여러 가지 정보를 과학적이고도 객관적으로 추측하는 방법이라고 할 수 있다. 예를 들어 선거에서 후보의 지지도, 국가 정책에 대한 국민의 찬반 정도 등 전체 국민의 의견을 조사하기 위해서는 통계적인 방법을 이용해야 한다.

일반적으로 모평균, 모분산, 모비율과 같은 모집단의 정보에 대한 어떤 판단을 하기 위해서 표본을 추출하고, 이 표본을 바탕으로 통계 이론에 의한 추측을 하는 일련의 과정을 통계적 추론(statistical inference)이라고 한다. 통계적 추론은 크게 추정(estimation)과 가설검정(test of hypothesis)이 있다. 통계적 추정이란 표본을 추출하여 모평균, 모비율 등과 같은 모집단의 어떤 미지의 값을 추측하는 과정을 말하고, 가설검정이란 표본을 추출하여 모집단에 대한 어떤 예상 또는 주장의 옳고 그름을 판정하는 과정을 말한다. 통계적 추정에는 점추정 방법과 구간추정 방법이 있다. 고등학교 교육과정에서는 구간추정만을 다루지만 구간추정은 점추정을 기초로 한다.

1. 점추정

점추정이란 표본에서 얻어진 정보를 이용하여 알지 못하는 모수의 참값으로 생각되는 하나의 값을 추측하는 방법이다. 이때 이 하나의 값을 추정량이라고 한다.

표본을 추출하여 추정량을 얻는 과정은 마치 궁수가 하나의 화살을 쏘아 과녁에 맞추는 것과 같다. 만일 한 궁수가 하나의 화살을 쏘아 과녁을 정확히 맞추었다고 해서 누구도 그 궁수가 최고의 명궁이라고 단정하지는 못할 것이다. 왜냐하면 한 번의 적중은 우연일 수도 있기 때문이다. 그러나 궁수가 수천 번을 쏘아 모두 과녁 가까이 맞추었다면 누구나 그를 명궁이라 할 것이다.

마찬가지로 추정량의 좋고 나쁨은 특정한 하나의 추정량이 취하는 값을 기준으로 판단할 수는 없다. 이때 이 추정량이 취하는 값을 추정값이라고 하는데 추정값을 여러 번 관측하여 이들이 참값 주위에 얼마나 가까이 분포되는가에 따라 추정량의 좋고 나쁨은 결정되는 것이다. 예를 들어 모평균 또는 모비율과 같이 모집단에서 추정하고자 하는 값을 θ 라고 하면 θ 에 대한 어떤 추정량 $\hat{\theta}$ 의 좋고 나쁨은 그 표본분포가 참값 θ 주위에 어떤 형태로 나타나는가에 따라 결정된다고 할 수 있다.

2. 구간추정

점추정에 의해서 추정된 하나의 값은 참값에 가깝기는 하지만 참값과 같을 확률은 매우 작다. 즉, 참값은 추정값을 중심으로 하는 어느 구간에 속해 있을 확률이 크다고 할 수 있다. 이와 같이 참값이 속해 있을 것이라고 생각되는 구간을 추측하는 것을 구간추정이라고 한다.

일반적으로 구간추정은 모집단에서 추정하고자 하는 값을 θ 라고 할 때, θ 를 크기 n 인 표본 X_1, X_2, \dots, X_n 으로 계산된 어떤 추정량을 중심으로 하는 구간 (L, U) 에 속하도록 L 과 U 를 정하는 것이다. 이때 이 구간의 길이가 커질수록 θ 가 이 구간 안에 속해 있을 확률은 커진다. 이 확률을 신뢰도 또는 신뢰수준이라 하고, 기호로 γ 또는 $1-\alpha$ 로 나타낸다. 그리고 구간 (L, U) 를 θ 에 대한 $100(1-\alpha) \%$ 신뢰구간이라 하고, 흔히 $1-\alpha$ 는 0.9, 0.95, 0.99 등이 사용된다.

< 우리나라에서는 세종대왕이 최초로 여론조사를 하였다. >

1430년 세종대왕은 신하들에게 다음과 같이 명하였다.

“정부, 육조와 각 도의 감사, 수령 및 품관으로부터 여영의 세민에 이르기까지 새로운 조세제도에 대하여 모두 가부를 물어서 아뢰게 하라.”

이 여론조사의 대상은 172648명, 조사를 하기 위해 걸린 시간은 약 5개월이었다. 왕의 명령은 하늘의 뜻이고 법과 같았던 조선 시대에 세종대왕은 왜 국민의 의견을 묻는 파격적인 정책을 실시하였을까?

세종대왕이 하고자 했던 것은 세제 개혁이었다. 기존의 세제는 농지 소득에 대한 세금을 거둘 때 그 해에 곡식이 얼마나 산출되었는가를 보고 세율을 정하여 흉년이면 조금

거두고, 풍년이면 많이 거두는 제도였다. 제도의 의도는 좋으나 흉년과 풍년을 가르는 기준이 명확하지 않았고, 땅을 소유한 지주들이 조사관에게 뇌물을 바쳐 수확량을 축소해 보고하는 것이 관행이었기 때문에 이를 타파하기 위한 제도가 필요하였다.

세종대왕은 공법 제도를 도입하고자 하였다. 공법은 정액세로서 흉작이든 풍작이든 토지의 크기에 따라 동일한 세금을 내 조세의 투명성을 높이는 제도이었다.

그러나 신하들의 반대가 만만치 않았다. 그 당시 조정 대신들은 기본적으로 대지주였기 때문에 공법 제도를 찬성할 수 없었다. 청렴한 정승으로 유명한 황희 정승조차 반대하였다. 세종대왕은 이를 여론조사 결과로 돌파하기 위해서 여론조사 방법을 택하였던 것이다.

< 논리를 키우는 수학 >

다음 문장을 읽고, 물음에 답하여라.

두 봉투 A, B 속에 각각 돈이 들어 있다. 한 봉투 속의 돈은 다른 봉투 속의 돈의 2 배이다. 이 두 봉투 중에서 하나를 무심코 골랐을 때, 나올 수 있는 금액의 기댓값을 생각해 보자.

만일 무심코 고른 봉투가 A 봉투이고, 그 속에 1000 원이 들어 있다고 하자. 그러면 B 봉투 속에는 2000 원이 들어 있거나 500 원이 들어 있다. 만일 B 봉투로 바꾸어 골랐더라면 1000 원의 이익을 보거나 500 원의 손해를 보았을 것이다. 즉, 기댓값이 $\frac{1}{2} \times 1000 - \frac{1}{2} \times 500 = 250$ 이므로 250 원의 이익을 보았을 것이다.

이번에는 B 봉투를 골랐다고 하고, 마찬가지로 그 속에 1000 원이 들어 있다고 하자. 그러면 A 봉투 속에는 2000 원이 들어 있거나 500 원이 들어 있다. 따라서 A 봉투로 바꾸어 골랐더라면 마찬가지로 250 원의 이익을 보았을 것이다.

즉, 먼저 고른 봉투보다는 남아 있는 봉투를 고르는 것이 항상 이익이라고 할 수 있다. 이것은 모순이다.

1. 두 봉투 속의 금액 중에서 작은 액수를 a 라고 하자. 두 봉투 중에서 하나를 무심코 골랐을 때, 나올 수 있는 금액의 기댓값을 구하여라.

2. 1의 결과에 근거하여 위의 모순이 어떻게 일어났는지 설명하여라.

[해설] 1. 두 봉투 속의 금액 중에서 작은 액수가 a 이므로 큰 액수는 $2a$ 이다. 따라서 두 봉투 중에서 하나를 무심코 골랐을 때 나올 수 있는 금액의 기댓값은

$$\frac{1}{2} \times a + \frac{1}{2} \times 2a = \frac{3}{2}a$$

[해설] 2. 두 봉투 중에서 무심코 고른 봉투가 A 봉투이고, 그 속에 1000 원이 들어 있다고 하였을 때, B 봉투 속의 돈이 2000 원 또는 500 원의 두 개의 값을 취할 수 있다고 가정하고 기댓값을 구한 것이 잘못되었다.