확률과 통계

# 맑은 개념

INTO THE MATH

Collector's Edition

송지은 편저

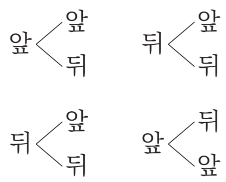
smart is sexy
Orbi.kr

orbibooks

## 수형도를 잘 그리는 원칙

경우의 수를 빠뜨림 없이, 중복 없이, 가독성이 높도록 수형도에 나타내기 위한 원칙을 알아봅시다.

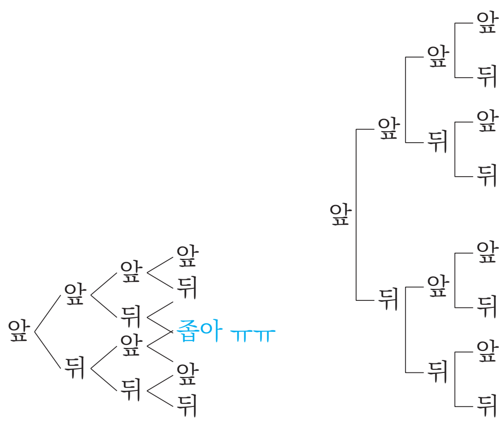
### 순서에 대한 규칙을 정하자




(a) (b)

수형도를 혼동 없이 그리기 위해서는 일정한 규칙을 정하는 것이 좋습니다. 가령 그림 (a)와 같은 수형도에서는 항상 '앞'을 마디보다 위쪽에 적고, '뒤'를 마디보다 아래쪽에 적는 규칙을 따라 그리고 있습니다. 그림 (b)와 같이 순서에 일관성이 없다면, 수형도를 그리는 과정에서 몇 가지 경우를 빠뜨리거나 중복하여 적을 우려가 있습니다. 나중에 검토할 때에도 수형도에 틀린 부분이 있는지 검증하기 어려움을 겪을 수 있습니다.

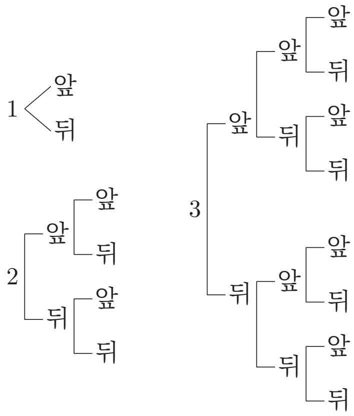
### 첫 마디는 넓게 뻗어나가자




(a) (b)

수형도의 특성상 뒤의 가지가 많이 퍼질 수밖에 없습니다. 그래서 그림 (a)와 같이 첫 마디에서 좁게 뻗어나가면 나중에 그림의 한가운데에서 위쪽 가지와 아래쪽 가지가 겹치게 됩니다. 그림 (b)와 같이 첫 마디에서 위아래로 멀리 뻗어나가면 이러한 문제를 예방할 수 있습니다.

원칙 적용의 예



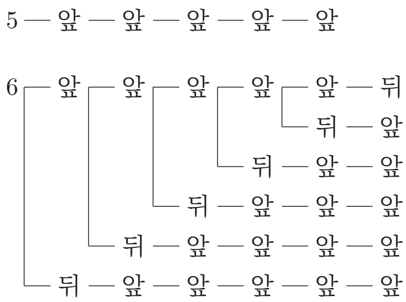

주사위를 던져 나온 눈의 수만큼 동전을 던지는 경우의 수형도를 일부만 그려보면 위 그림과 같습니다. 가장 작은 수인 1부터 위에 적어나가면서, 동전은 앞을 항상 뒤보다 위에 적어주고 있음을 알 수 있습니다. 4, 5, 6일 때는 일부러 생략했으므로, 반드시 직접 빈 종이에 수형도를 그려보시기 바랍니다.

예제 1. 주사위를 던져 나온 눈의 수만큼 동전을 던진다. 이때 앞면이 나온 횟수가 5인 경우의 수는?

### 예제 1 풀이

주사위를 던져 나온 눈의 수만큼 동전을 던진다. 이때 앞면이 나온 횟수가 5인 경우의 수는?

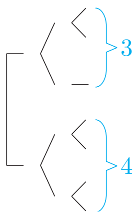

주사위를 던져 나온 눈의 수가 1, 2, 3, 4인 경우는 셀 필요가 없으므로, 수형도는 '첫 마디가 5인 경우'와 '첫 마디가 6인 경우' 두 개만 그리면 됩니다.




5에서는 앞앞앞앞앞 1가지, 6에서는 앞앞앞앞앞뒤, 앞앞앞앞뒤앞, 앞앞앞뒤앞앞, 앞앞뒤앞앞앞, 앞뒤앞앞앞앞, 뒤앞앞앞앞앞 6가지가 있습니다. 따라서 구하는 경우의 수는 $1+6=7$가지입니다.

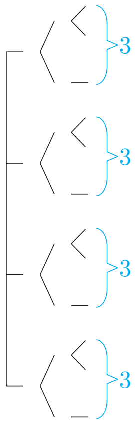
## 합의 법칙과 곱의 법칙 : 수형도를 그리는 데 활용되는 기본 원리

합의 법칙과 곱의 법칙은 수형도를 그리는 데 활용되는 기본 원리입니다. 수형도를 그리다 보면 수형도가 매 마디에서 갈라질 때, 갈라져 나온 이후의 구조가 다른지 같은지를 따지게 됩니다. 예상하셨겠듯이, 다르다면 합의 법칙을, 같다면 곱의 법칙을 씁니다.



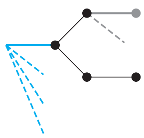

$3+4=7$

갈라진 각 마디 이후 단계에서 수형도의 구조가 다르다면, 구조가 각기 다른 수형도를 각각 그리고, 각각의 경우의 수를 더하여 셉니다. 이것이 곧 합의 법칙입니다.




$3 \times 4 = 12$

갈라진 각 마디 이후 단계에서 수형도의 구조가 같다면, 하나만 그리고 나머지 수형도는 그리지 않아도 됩니다. 그저 수형도를 한 번만 그려 세고, 구조가 동일한 수형도가 몇 개인지를 파악하여 곱해주면 되는 것입니다. 이것이 곧 곱의 법칙입니다.



$4 \times \{(2 \times 1) + 1\}$

곱의 법칙에서 주어진 수형도를 압축하여 그려봅시다. 합의 법칙과 곱의 법칙을 융합하여 그림과 같이 한 개의 수형도만 그린 후 아래에는 점선으로 표시하거나 곱하기로 표시하여 수형도를 생략할 수 있고, 수형도의 가지가 달라지는 마디는 덧셈으로 표기할 수 있습니다.

결국 합의 법칙과 곱의 법칙을 쓰는 것은, 수형도가 그려지는 구조를 파악하여 가급적 수형도를 덜 그리기 위한 것이라 생각할 수 있습니다. 수형도의 패턴이 반복되면 곱의 법칙을 이용해 수형도를 한 번만 그려 생략하고, 패턴이 다른 경우에는 어쩔 수 없이 각각의 수형도를 그린 후 합의 법칙으로 더합니다.

지금까지 배운 내용인 수형도, 합의 법칙, 곱의 법칙만을 이용하여 아래의 문제를 풀어봅시다.[8)]

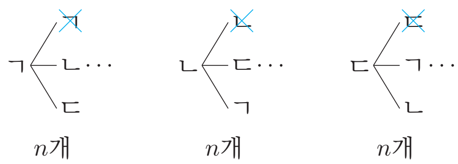
8) 같은 것이 있는 순열(같있순)은 아직 배우지 않았으므로, 같있순을 이용한 풀이는 뒤에서 다룰 것입니다.

예제 2. 여섯 개의 자음 ㄱ, ㄱ, ㄴ, ㄴ, ㄷ, ㄷ을 일렬로 나열하여 문자열을 만든다. ㄱ은 ㄱ과 서로 이웃하지 않고, ㄴ은 ㄴ과 서로 이웃하지 않고, ㄷ은 ㄷ과 서로 이웃하지 않도록 배열된 문자열의 개수를 구하시오.

## 예제 2 풀이

여섯 개의 자음 ㄱ, ㄱ, ㄴ, ㄴ, ㄷ, ㄷ을 일렬로 나열하여 문자열을 만든다. ㄱ은 ㄱ과 서로 이웃하지 않고, ㄴ은 ㄴ과 서로 이웃하지 않고, ㄷ은 ㄷ과 서로 이웃하지 않도록 배열된 문자열의 개수를 구하시오.

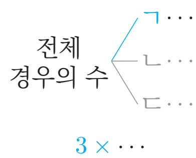
처음에 ㄱ, ㄴ, ㄷ으로 분류하여 수형도를 3개 그리겠다고 생각했다면, 그것만으로도 절반은 성공한 것입니다. 이때 ㄱ으로 시작하든, ㄴ으로 시작하든, ㄷ으로 시작하든, 이후 수형도의 구조가 같은지 곰곰이 생각해봅시다.



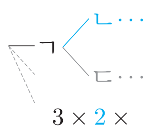

ㄱ의 입장에서 남은 것은 ㄱㄴㄴㄷㄷ이므로, 두 개씩 남은 ㄴ, ㄷ 중 하나를 선택하면 됩니다. ㄴ의 입장에서 남은 것은 ㄴㄷㄷㄱㄱ이므로, 두 개씩 남은 ㄱ, ㄷ 중 하나를 선택하면 됩니다. ㄷ의 입장에서 남은 것은 ㄷㄱㄱㄴㄴ이므로, 두 개씩 남은 ㄱ, ㄴ 중 하나를 선택하면 됩니다.

즉 서로 입장만 다를 뿐 완전히 동일한 상황임을 알 수 있습니다. 그래서 ㄱ으로 시작하는 경우의 수인 $n$을 구한 후, 여기에 3을 곱하면 전체 경우의 수인 $3n$을 구할 수 있습니다.[9)]

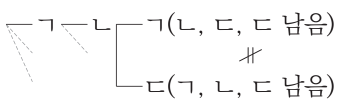
9) 이것이 곧 곱의 법칙입니다.




여기까지 우리가 풀이하며 작성하는 식은 $3\times$ 입니다.

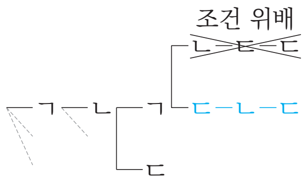



이제 ㄱ에만 집중하여 수형도를 그려나가봅시다. 앞서 말한 것과 같은 원리로, ㄴ을 고르든 ㄷ을 고르든 뒤의 수형도의 구조가 동일합니다. 따라서 ㄱ-ㄴ 이후만 그려나가면 됩니다. 여기까지 우리가 풀이하며 작성하는 식은 $3 \times 2\times$ 입니다.




그런데 ㄱ-ㄴ-ㄱ과 ㄱ-ㄴ-ㄷ은 상황이 달라집니다. ㄱ-ㄴ-ㄱ에서는 ㄱ이 이미 다 쓰여 ㄴㄷㄷ만 남고, ㄱ-ㄴ-ㄷ은 ㄱ,ㄴ,ㄷ이 하나씩 남아 있는 상황이기 때문입니다. 따라서 이런 경우는 각각의 경우의 수인 $a$, $b$를 구한 후 더하여야 합니다.[10)] 여기까지 우리가 풀이하며 작성하는 식은 $3 \times 2 \times (\quad + \quad)$입니다. 더하기의 왼쪽에는 $a$의 값을, 더하기의 오른쪽에는 $b$의 값을 적으면 됩니다.

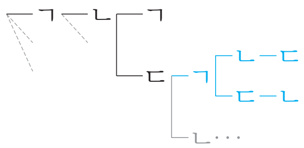
10) 이것이 곧 합의 법칙입니다.




$$3 \times 2 \times (1 + \quad\quad )$$

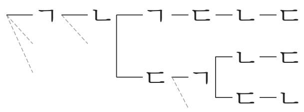
이제 $a$를 구하기 위해 ㄱ-ㄴ-ㄱ를 마저 그려봅시다. ㄱ-ㄴ-ㄱ-ㄴ-ㄷ-ㄷ는 ㄷ끼리 이웃하여 문제의 조건에 위배되므로 세지 않습니다. ㄱ-ㄴ-ㄱ-ㄷ-ㄴ-ㄷ는 문제의 조건을 만족시킵니다.[11] 따라서 $a = 1$입니다.

11) 여기서 ㄱ-ㄴ-ㄱ-ㄷ-ㄷ-ㄴ은 왜 그리지 않았는지 의문이 들 수 있습니다. 앞서 다룬 예제 1에서 주사위의 눈이 1, 2, 3, 4가 나온 경우를 그리지 않은 것처럼, 확실하게 조건에 위배되므로 해당 경우는 그리지 않았습니다.



$$3 \times 2 \times (1 + 2 \times 2)$$

이제 $b$를 구하기 위해 ㄱ-ㄴ-ㄷ를 마저 그려봅시다. ㄱ-ㄴ-ㄷ-ㄱ와 ㄱ-ㄴ-ㄷ-ㄴ는 뒤의 수형도의 구조가 같습니다. 따라서 ㄱ-ㄴ-ㄷ-ㄱ만 풀이하고 2를 곱하면 $b$를 구할 수 있습니다. ㄱ-ㄴ-ㄷ-ㄱ에서는 ㄱ-ㄴ-ㄷ-ㄱ-ㄴ-ㄷ와 ㄱ-ㄴ-ㄷ-ㄱ-ㄷ-ㄴ의 2가지가 있으므로, $b = 2 \times 2$입니다.



$$3 \times 2 \times (1 + 2 \times 2)$$

따라서 최종적으로 정답은 $3 \times 2 \times (\ 1 \ + \ 2 \times 2\ ) = 30$이 됩니다.

## 마치며

이와 같이 경우의 수를 잘 하는 것은, 수형도를 얼마나 잘 그리는지에 달려 있습니다. 그리고 수형도를 잘 그리는 것은 '몇 개의 수형도를 그릴 것이며', '언제 더하고 언제 곱하는지'를 아는 것과 같습니다.

이에 더하여, 앞으로 다룰 여러 가지 순열과 조합 공식들은 자주 나오는 수형도를 그리지 않고도 한 방에 구하는 테크닉일 뿐입니다. 결국 이러한 테크닉을 적재적소에 활용한다면, 수형도를 그리지 않고도 수형도의 가짓수를 알 수 있을 것입니다.

자, 이제는 이 단원을 시작하며 말했던 아래의 문장이 다시금 와닿을 것입니다.

> 경우의 수 단원의 실력은 '수형도를 얼마나 잘 그리는 지'에 달려 있습니다. 그런데 역설적이게도, 수형도를 잘 그릴 수 있는 실력을 갖추면, 수형도를 간소하게 그리거나, 수형도를 전혀 그리지 않고도 경우의 수 문제를 풀 수 있게 됩니다.

### 예제 1 풀이

> 딸기맛 사탕 3개, 포도맛 사탕 3개, 레몬맛 사탕 3개가 있다. 중복을 허용하여 3개의 사탕을 임의로 동시에 선택하였을 때, 세 사탕을 한 개씩 선택했을 확률을 구하시오. (단, 같은 맛 사탕끼리는 서로 구별되지 않고, 선택되지 않은 사탕이 있을 수도 있다.)

확률의 세계관을 배울 때 강조했듯이, 문제에 구별되지 않는다고 적혀 있어도 항상 구별해야 하고, 문제에 동시에 선택한다고 적혀 있어도 항상 순서대로 선택해야 합니다. 중복조합의 상황에서 확률을 사용하는 것이 그 대표적인 예시입니다. 풀이에서 중복조합을 사용했다면 $\frac{1}{{}_3H_3} = \frac{1}{{}_5C_3} = \frac{1}{10}$이라는 답을 구했을 것입니다. 이를 다른 방법으로 구한 값과 비교해보도록 합시다.

경우의 수로 풀어봅시다. 전체 경우의 수는 9개의 사탕 중에서 3개의 사탕을 고르는 ${}_9C_3$이고, 해당 경우의 수는 $3 \times 3 \times 3$입니다. 따라서 $\frac{3 \times 3 \times 3}{{}_9C_3} = \frac{27}{84} = \frac{9}{28}$입니다.[73)]

확률로 풀어봅시다. 9개의 사탕을 차례대로 3번 선택한다고 생각해야 합니다. 첫 번째 사탕은 아무거나 골라도 상관없으므로 $\frac{9}{9} = 1$입니다. 두 번째 사탕은 남은 8개의 사탕 중에서 고르되 첫 번째 사탕과 다른 종류여야 하므로 $\frac{6}{8}$입니다. 마지막 사탕은 남은 7개의 사탕 중에서 고르되 앞에서 고르지 않은 한 종류의 사탕을 고르면 되므로 $\frac{3}{7}$입니다. 따라서 곱의 법칙에 의해 $1 \times \frac{6}{8} \times \frac{3}{7} = \frac{9}{28}$입니다.

73) 확률을 계산할 때에는 마지막 계산 과정에 약분하게 되므로, 중간과정에서는 곱셈을 남겨두는 것이 편리합니다.

### 표준편차 : 제곱으로 인한 과장된 값을 루트를 씌워 보정하고, 단위를 맞춘다.

분산을 통해 각각의 확률변수의 분포가 어떤지를 수치로 나타낼 수 있었지만, 계산 과정에서 제곱이 쓰이다보니 그 값이 원래 다루던 값들에 비해 과장되는 면이 있습니다. 또한 분산을 구하는 과정에서 제곱이 사용되어 단위가 달라진 상태입니다.[95)]

이렇게 제곱으로 인한 과장을 보정함과 동시에 단위를 맞추어주기 위한 목적으로 분산에 루트를 씌운 값을 **표준편차**라 합니다. 홍규의 표준편차는 $17.88\cdots$, 선경의 표준편차는 $6.32\cdots$, 남호의 표준편차는 $0$이므로, 분산을 비교할 때보다는 값들이 작아져서 다루기 편하고, 원래의 데이터와 단위가 동일하여 다루기 편리합니다.[96)]

## $aX+b$의 평균, 분산, 표준편차

$\mathrm{E}(aX+b)=a\mathrm{E}(X)+b$는 $\sum$의 성질로 쉽게 증명할 수 있고, $\mathrm{V}(aX+b)=a^2\mathrm{V}(X)$는 $\mathrm{V}(X)=\mathrm{E}\left(X^2\right)-m^2$으로 쉽게 증명할 수 있고, $\sigma(aX+b)=|a|\,\sigma(X)$는 정의에 의해 자명합니다. 이 수식이 무엇을 의미하는지 선경, 홍규, 남호의 성적에 $a$와 $b$의 값을 구체적으로 넣어 직접 계산해보는 것도 좋습니다.

## 이항분포 : 표도 그리지 않고, 증명도 필요 없는 특수한 이산확률분포

지금까지 알아본 바와 같이, 이산확률변수는 주로 표를 그려 해결합니다. 표를 그려야 평균, 분산, 표준편차를 구할 수 있기 때문입니다. 그런데 표를 그리지 않는 특이한 이산확률분포가 있습니다. 바로 이항분포입니다.

한 번의 어떤 사건 $A$가 일어날 확률이 $p$, 일어나지 않을 확률이 $q$일 때, $n$번의 독립시행에서 사건 $A$가 몇 번 일어났는지를 확률변수로 $X$라 하면, 직관적으로 $\mathrm{E}(X)=np$임을 알 수 있습니다. 또한 $\mathrm{V}(X)=npq$, $\sigma(X)=\sqrt{npq}$임을 증명 없이 받아들입시다.

### 어떤 이항분포는 표나 확률질량함수를 줍니다!

| $X$ | $0$ | $1$ | $2$ | $\cdots$ | $n-1$ | $n$ | 합 |
|---|---|---|---|---|---|---|---|
| $\mathrm{P}(X=x)$ | ${}_n\mathrm{C}_0q^n$ | ${}_n\mathrm{C}_1pq^{n-1}$ | ${}_n\mathrm{C}_2p^2q^{n-2}$ | $\cdots$ | ${}_n\mathrm{C}_{n-1}p^{n-1}q$ | ${}_n\mathrm{C}_np^n$ | $1$ |

비록 이항분포의 확률분포를 표로 잘 나타내지 않는다고 하더라도, 이항분포 또한 태생이 이산확률분포임을 잊지 말아야 합니다. 따라서 위와 같은 표나 $\mathrm{P}(X=x)={}_n\mathrm{C}_xp^xq^{n-x}$와 같은 확률질량함수를 이용하여 $X\sim\mathrm{B}(n,\ p)$라는 정보를 간접적으로 제시할 수 있습니다.

### $n$이 충분히 크면 $X\sim\mathrm{B}(n,\ p)$인 $X$는 근사적으로 $X\sim\mathrm{N}(np,\ npq)$이다.

$n$이 충분히 큰 경우[97)] 이항분포 대신 정규분포를 이용하여 쉽게 계산할 수 있음이 알려져 있습니다.

95) 각 학생들의 과목별 점수, 각 학생의 평균 점수, 각 학생의 과목별 편차의 단위를 '길이'로 생각하면, 분산의 단위는 '넓이'가 된다고 재해석할 수 있습니다. 그러면 '편차의 제곱'은 사각형(square)의 넓이가 되고, 편차의 제곱의 평균인 '분산'은 '여러 사각형의 넓이의 평균'이 됩니다. 이러한 재해석과 관련하여, '사각형(square)의 넓이의 합을 최소로(least) 만들어주는 값'이 평균임이 알려져 있습니다. 이것을 이용하여 평균을 구하거나 평균의 근삿값을 구하는 기법이 최소제곱법(least square method)입니다.

96) 사실 표준편차는 이산확률분포에서 존재감이 별로 없습니다. 그러나 연속확률분포, 그 중에서도 정규분포에서 매우 중요한 역할을 하며, 이후 통계적 추정에서도 아주 중요한 역할을 할 것입니다.

97) $np\geq 5$, $nq\geq 5$인 경우

# 02 부록 : 찾아보기