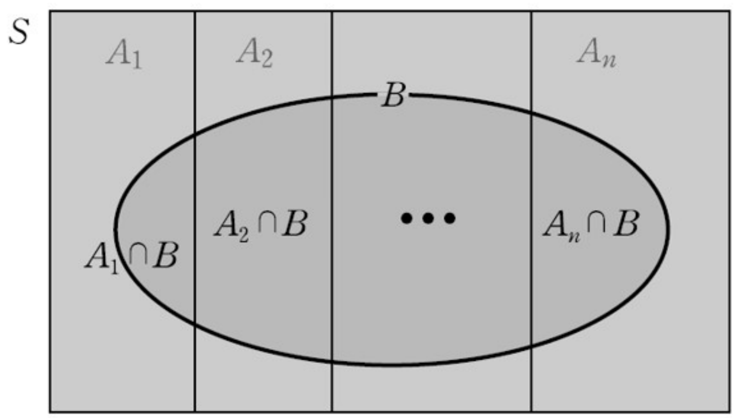
1 베이즈정리

1.1 전확률 공식

Bayes(1763)가 증명한 확률에 관한 정리이자 확률변수의 조건부(conditional) 확률분포와 주변부(marginal) 확률분포를 연관 짓는 확률이론으로, 서로 배반적인(exclusive) n개 사건 $A_1, \ldots, A_n$이 을 만족할 때, 어떤 사건 E가 일어났다는 가정에서의 조건확률(conditional probability) $P(A_i|E)$에 관한 정리이다.



위 그림에서 볼 수 있듯이 $P(A_i) > 0, (i = 1, 2, \ldots n)$인 사건 A_i들이 표본공간 S의 분할이라고 하자. 이 때 다음의 두 조건을 만족한다.

- $A_i \cap A_j = \phi, \quad i \neq j, \quad , i, j = 1, 2, \ldots n$
- $S = A_1 \cup A_2 \cup \cdots \cup A_n$

이 때, 임의의 사건 B는 그림과 같이 $A_i \cap B$, $i = 1, 2, \ldots n$으로 분할되며, 다음과 같이 표현 가능하다.

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

1.2 베이즈정리

사건 $A_1, A_2, \ldots A_n$이 표본공간 S의 분할이고 $P(B) > 0$인 어떤 사건 B가 발생하였을 때, 사건 A_i의 조건부 확률은 다음과 같다.

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

여기서 $P(A_i|B)$는 사후확률 (posterior probability), $P(A_i)$를 사전확률 (prior probability) 라고 한다.

조금 더 쉽게 설명하면, 사건 A와 B가 있을 때, 사건 B가 일어난 것을 전제로 한 사건 A의 조건부 확률을 구하고 싶다고 하자. 그런데 지금 알고 있는 것은 사건 A가 일어난 것을 전제로 한 사건 B의 조건부 확률, A의 확률, B의 확률뿐이다. 그럴 때, 원래 구하고자 했던 사건 B가 일어난 것을 전제로 한 사건 A의 조건부 확률은 다음과 같이 구할 수가 있다.

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

베이즈 정리는 위의 정의에서 알 수 있듯, 본래 역확률(inverse probability) 문제를 해결하기 위한 방법이었다. 즉, 조건부 확률 $P(B|A)$를 알고 있을 때, 전제와 관심 사건이 관계가 정반대인 조건부 확률 $P(A|B)$을 구하는 방법이었다.
하지만 베이즈 정리를 이전의 경험과 현재의 증거를 토대로 어떤 사건의 확률을 추론하는 알고리즘으로 보고 관심을 가지는 사람들도 있었다. 이 때는 어떤 사건이 일어날 확률에 대한 임의의 가정 $P(A)$에 실제로 발견된 자료나 증거 B를 반영해서, 자료로 미루어보아 어떤 사건이 일어날 확률 $P(A|B)$을 구하는 것이 관심의 대상이 된다.

- Example #1 : 의학보고서에 따르면 전체 국민의 7%가 폐질환을 앓고 있으며, 그들 중 85%가 흡연가라고 한다. 그리고 폐질환을 갖지 않은 사람 중에 25%가 흡연가라고 한다. 이 때, 임의로 선정한 사람이 흡연가일 확률을 구해보자.

- Example #2 : 위의 예제에서 임의로 선정한 사람이 흡연가라고 할 때, 이 사람이 폐질환을 앓고 있을 확률을 구해보자.

- Example #3 : 코로나 간이진단 검사를 실시하는데 이 검사는 코로나에 걸렸을 때 양성반응이 나올 확률은 0.96이고 코로나에 걸리지 않았을 때 양성반응이 나올 확률이 0.05라고 하자. 그리고 코로나에 걸릴 확률을 0.001이라고 하자. 만약 검사에서 양성반응이 나왔다면, 코로나에 걸렸을 확률을 구해보자.

- Example #4 : 크기와 모양이 같은 공이 상자 A에는 검은 공 2개와 흰 공 2개, 상자 B에는 검은 공 1개와 흰 공 2개가 있다. 두 상자 A, B 중 임의로 선택한 하나의 상자에서 공을 1개 꺼냈더니 검은 공이 나왔을 때, 그 상자에 남은 공이 모두 흰 공일 확률을 구해보자.

- Example #5 : 주머니에 들어있는 세장의 카드가
  - 앞과 뒤가 모두 검은색 : 1장
  - 앞과 뒤가 각각 검은색과 흰색 : 1장
  - 앞과 뒤가 모두 흰색 : 1장

  주머니에서 1장을 뽑아서 본 면이 검은색일 때, 그 뒷면이 검은색이 나올 확률을 구해보자.

- Example #6 : ¡몬티홀 문제¿ 하나의 문 위에 스포츠카가 있고 두개의 문 뒤에는 염소가 있을 때, 세 개의 문 중에 하나를 선택하여 그 문뒤에 스포츠카가 있으며 가질 수 있는 게임이 있다.

  출연자가 한 문을 선택한다. 게임 진행자가 나머지 문 중에 염소가 있는 문을 보여준다. 그리고 나서 처음 선택한 문을 바꿀 의사가 있는지 출연자에게 묻는다.

  출연자가 처음 선택을 고수할 때, 스포츠카를 갖게 될 확률을 구해보자.

- Example #7 : 볼트를 만드는 어느 공장에서 두 대의 기계 A, B가 각각 전체 생산량의 40%, 60%를 만들며, 또 각각의 제품의 5%, 3%가 불량품이라고 한다. 제품 가운데 꺼낸 하나의 볼트가 불량품이었다고 할 때, 그것이 기계 A에서 만들어졌을 확률을 구해보자.

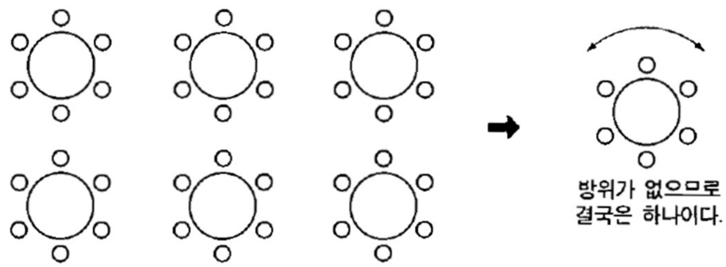
- Example #8 : 어느 보안전문회사에서 바이러스 감염 여부를 진단하는 프로그램을 개발하였다. 그 진단 프로그램은 바이러스에 감연된 컴퓨터를 감염되었다고 진단할 확률이 94%이고, 바이러스에 감염되지 않은 컴퓨터를 감염되지 않았다고 진단할 확률이 98%이다. 실제로 바이러스에 감염된 컴퓨터 200대와 바이러스에 감염되지 않은 컴퓨터 300대에 대해 이 진단 프로그램으로 바이러스 감염 여부를 검사하려고 한다. 이 500대의 컴퓨터 중 임의로 한 대를 택하여 이 진단 프로그램으로 감염 여부를 검사하였더니 바이러스에 감염되었다고 진단하였을 때, 이 컴퓨터가 실제로 감염된 컴퓨터일 확률을 구해보자.

- Example #9 : 어느 청량 음료 회사의 연간 청량음료 판매량은 그 해 여름의 평균 기온에 크게 좌우된다. 과거 자료에 따르면, 한 해의 판매 목표액을 달성할 확률은 그 해 여름에 평균 기온이 예년보다 높을 경우에 0.8, 예년과 비슷할 경우에 0.6, 예년보다 낮을 경우에 0.3이다. 일기예보에 따르면 내년 여름의 평균기온이 예년보다 높을 확률이 0.4, 예년과 비슷할 확률이 0.5, 예년보다 낮을 확률이 0.1이라고 한다. 이 회사가 내년에 판매 목표액을 달성할 확률을 구해보자.

2 여러가지 순열

2.1 원순열

순열이 서로 다른 것들을 일렬로 나열하는 것이라면, **원순열** (circular permutation)은 말 그대로 서로 다른 것들을 원형으로 배열하는 것이다.



위 그림에서 볼 수 있듯이 원탁에 6명을 앉힌다고 가정할 때, 원탁에 A를 먼저 앉힌 후 나머지 5명 (B, C, D, E, F)을 앉히는 경우의 수를 생각해보자. 아무도 앉지 않은 상태에서 원탁의 6개의 자리는 서로 구분할 수가 없기 때문에 A를 먼저 앉히는 경우의 수는 오진 한 가지뿐이다. 그리고 나머지 5명을 원탁의 남은 5개의 자리에 앉히는 경우의 수는 서로 다른 5개를 일렬로 나열하는 순열의 수 5!와 같다. 따라서 총 경우의 수는 120 가지이다.

$$1 \times 5! = 120$$

따라서, 서로 다른 n개를 원형으로 배열하는 원순열의 수는 $(n-1)!$ 이다.

- Example #10 : 8명의 학생 A, B, C, D, E, F, G, H 가 있다. 8명이 원탁에 일정한 간격으로 둘러앉는 방법의 수를 구해보자. 그리고 8명 중 4명을 뽑아 원탁에 일정한 간격으로 둘러앉히는 방법의 수를 구해보자.

2.2 중복순열

서로 다른 n개에서 순서를 생각하지 않고 중복을 허락하여 r개를 택하여 일렬로 나열하는 것을 n개에서 r개를 택하는 **중복순열**이라 하고, 이를 아래와 같이 나타낸다.

$$ {}_{n}\Pi_{r} = n \times n \times \cdots \times n = n^{r} $$

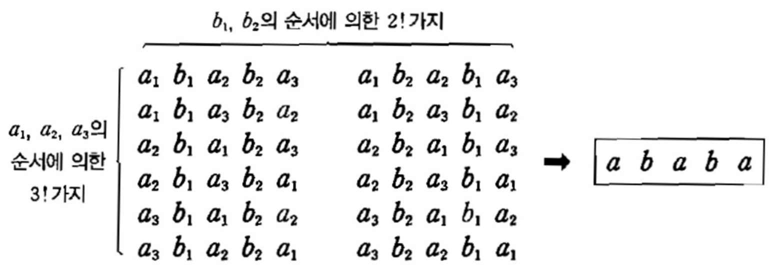
중복순열에서는 한 번 택한 것을 또 다시 택할 수 있으므로 각 단계에서 택할 수 있는 경우의 수가 항상 동일하게 유지된다.

- Example #11 : 0, 1, 2, 3, 4, 5의 6개의 숫자에 대하여 다음을 구해보자.
  - 중복을 허용하여 만들 수 있는 다섯 자리 수의 개수
  - 중복을 허용하여 만들 수 있는 30000보다 큰 다섯 자리 수의 개수
  - 중복을 허용하여 만들 수 있는 짝수인 다섯 자리 수의 개수

- Example #12 : 다섯 개의 알파벳 a, b, c, d, e 중에서 중복을 허락하여 4개를 뽑아 만들 수 있는 문자열 중 양 끝의 영문자가 모두 자음인 것의 개수를 구해보자.

2.3 같은 것이 있는 순열

순열이 서로 다른 것들을 일렬로 나열하는 것이라면 같은 것이 있는 순열은 같은 것이 몇몇 포함되어 있는 것들을 일렬로 나열하는 것이다.
예를 들어 5개의 문자 a, a, a, b, b를 일렬로 나열하는 경우에 대해 생각해보자. 이 때, 모두 다른 것으로 간주하여 일렬로 나열한 뒤 같은 경우를 없애는 방법으로 그 수를 확인할 수 있다. 먼저 3개의 a와 2개의 b에 각각 번호를 붙여 a_1, a_2, a_3, b_1, b_2로 놓으면 이들을 일렬로 나열하는 모든 경우의 수는 5!이다.



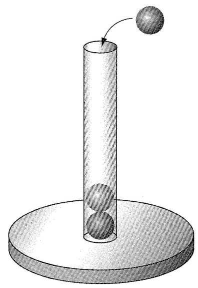
그런데 각 나열에 속해 있는 a_1, a_2, a_3, b_1, b_2의 번호를 모두 지워 a, b로 바꾸어 보면 위와 같이 $3! \times 2!$ 가지의 나열은 모두 동일한 경우가 된다. 따라서, 5개의 문자 a, a, a, b, b를 일렬로 나열하는 경우의 수는 10가지이다.

$$\frac{5!}{3!2!} = 10$$

이를 일반화해 보면, n개 중에서 서로 같은 것이 p개, q개, $\dots$ r개씩 있을 때, n개를 일렬로 나열하는 방법의 수는 다음과 같다.

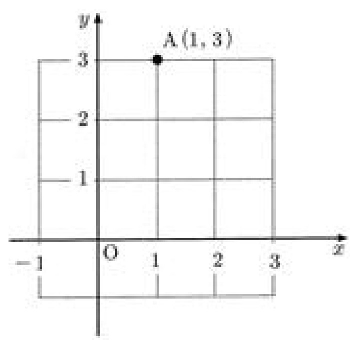
$$\frac{n!}{p!q!\dots r!} \quad (p + q + \cdots + r = n)$$

- Example #13 : 투명한 원기둥 모양의 관에 크기와 모양이 같은 구슬을 넣어 쌓아올리는 장난감이 있다. 빨간 구슬이 4개, 파란 구슬이 2개, 노란 구슬이 1개 있을 때, 이 구슬 7개를 모두 사용하여 일렬로 쌓아 올릴 수 있는 모든 경우의 수를 구해보자.



- Example #14 : 1, 1, 2, 2, 2, 3, 4의 7개의 수를 모두 일렬로 배열할 때, 양 끝에 2가 오도록 배열하는 방법의 수는 a가지, 두 개의 1이 이웃하도록 배열하는 방법의 수는 b가지이다. $a+b$의 값을 구해보자.

- Example #15 : 좌표평면 위에서 상하 또는 좌우방향으로 한 번에 1만큼씩 움직이는 점 P가 있다. 이 때, 원점을 출발한 점 P가 6번 움직여서 최종 위치가 점 A(1, 3)이 되는 경우의 수를 구해보자.



3 중복조합

서로 다른 n개에서 순서를 생각하지 않고 중복을 허락하여 r개를 선택하는 것을 서로 다른 n개에서 r개를 택하는 **중복조합** (repeated combination)이라 하고, 이를 아래와 같이 나타낸다.

$$ {}_nH_r = {}_{n+r-1}C_r $$

중복조합의 수를 계산하기 위해서는 중복조합을 지난 시간에 배운 (일반)조합으로 변형하여 생각하는 것이 핵심이다. 3개의 숫자 1, 2, 3에서 2개를 선택하는 중복조합의 경우를 생각해보자.

$$ \{1,1\},\{1,2\},\{1,3\},\{2,2\},\{2,3\},\{3,3\} $$

이 때, 각 경우의 첫 번째, 두 번째 수에 각각 0과 1을 더해보면,

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

가 되는데, 이는 4개의 숫자 1, 2, 3, 4에서 2개를 선택하는 조합의 각 경우의 수와 같다. 위의 두 경우가 서로 일대일 대응되므로 그 개수는 ${}_3H_2 = {}_4C_2 = 6$으로 같다.
이를 일반화하면, 서로 다른 n개에서 r개를 택하는 중복조합은 서로 다른 $(n+r-1)$개에서 r개를 선택하는 조합으로 변형될 수 있다.

$$ {}_nH_r = {}_{n+r-1}C_r $$

- Example #16 : 크기와 모양이 같은 검은 구슬 5개와 흰 구슬 2개를 서로 다른 세 상자에 모두 넣는 방법의 수를 구해보자.

이번에는 서로 다른 n개에서 r개를 택하는 중복조합의 수를 n개의 미지수를 가진 일차방정식의 해의 개수와 연관지어 접근해 보자. 서로 다른 n개에서 r개를 선택하는 중복조합의 수 ${}_nH_r$는 n개의 미지수를 가지는 일차 방정식

$$x_1 + x_2 + \cdots + x_n = r$$

의 음이 아닌 정수해의 개수로 바꾸어 생각할 수 있다. 예를 들어, 방정식 $x+y+z=2$의 음이 아닌 정수해를 순서쌍 (x,y,z)로 나타내면 다음과 같다.

$$(2,0,0),\ (1,1,0),\ (1,0,1),\ (0,2,0),\ (0,1,1),\ (0,0,2)$$

이를 미지수 x, y, z의 사용 개수로 해석한다면,

$$xx,\quad xy,\quad xz,\quad yy,\quad yz,\quad zz$$

가 되며, 이는 3개의 문자 x, y, z에서 2개를 택하는 중복조합의 각 경우와 같다.

- Example #17 : 방정식 $x+y+z=6$의 음이 아닌 정수해의 개수를 구해보자.

- Example #18 : $x+y+z+10u=39$를 만족하는 자연수인 홀수 x,y,z,u에 대하여 순서쌍 (x,y,z,u)의 개수를 구해보자.

- Example #19 : $(a+b+c)^5$의 전개식에서 서로 다른 항의 개수를 구해보자.

- Example #20 : 사과 주스, 포도 주스, 감귤 주스 중에서 8병을 선택하려고 한다. 사과 주스, 포도 주스, 감귤 주스를 각각 적어도 1병 이상씩 선택하는 경우의 수를 구해보자. (단, 각 종류의 주스는 8병 이상씩 있다.)

- Example #21 : 네 종류의 사탕 중에서 15개를 선택하려고 한다. 초콜릿사탕은 4개 이하, 박하사탕은 3개 이상, 딸기사탕은 2개 이상, 버터사탕은 1개 이상을 선택하는 경우의 수를 구해보자. (단, 각 종류의 사탕은 15개 이상씩 있다.)

e-mail : kimtwan21@dongduk.ac.kr