Lecture 03 - Discrete Distribution (2)

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr

국립 서울과학기술대학교 IT정책전문대학원

I. 베르누이분포 (Bernoulii Distribution)

Motivation

- 동전을 던지는 시행 나오는 면을 X라 하였을 때,
  - 앞면과 뒷면이 나오는 확률이 같다면, X는 이산일양분포를 따른다고 한다.
  - 만약에 동전이 찌그러져 있어서 앞면과 뒷면이 나오는 확률이 다르다면 어떻게 될까? 이 경우에는 이산일양분포라고 할 수 없다.

Definition

- 동전을 던지는 시행의 결과와 같이 두 가지 결과중에 하나가 나오는 **이진 결과(binary outcome)**를 가진 경우를 나타내는 분포가 **베르누이(Bernoulii) 분포**라고 한다.
- 그리고 이진결과가 나오게 되는 한번의 확률변수에 대한 관찰을 **베르누이 시행**이라고 한다.
- 지난번에 다룬 양면의 확률이 같은 동전(fair coin)의 경우에는 이산일양변수이며 또한 베르누이변수이다.

수식으로 표현

- 이산확률변수 X가 1과 0과 같이 binary outcome을 가지고 있다. 만약에 1이 결과로 나올 확률을 p라고 한다면, 아래와 같이 수식(pmf)으로 표현할 수 있다.

$$p(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

- 이진 결과를 가진 다른 많은 시행들이 위와 같은 형식으로 표현될 수 있다.
  - 오늘 하루중에서 비가 오고 안오는 것,
  - 비행기가 취소되는 것과 아닌것,
  - 집에가는 길에서 만원 지폐를 발견하는 것 모두 이진 결과이므로
  - 이들을 각기 다른 p값을 가지는 베르누이 분포로 표현할 수 있다.

모수 (parameter)와 분포의 표기

- 위의 pmf 수식에서 p값을 알면 분포를 정확하게 알고 있는 것이 된다.
- 베르누이 분포를 이야기 할 때에는 p값이 반드시 동반된다.
- 이에 대해서 **베르누이 분포의 모수(parameter)는 p이다**라고 한다.
- 그리고 어떤 확률 변수 X가 모수가 p인 베르누이 분포를 따른다면 아래와 같이 표기하며, 아래의 표기를 보면 위의 pmf를 떠올릴 수 있어야 한다.

$$X \sim Ber(p)$$

이진결과와 통계학의 convention

이진결과의 이해

- 베르누이 분포의 기본인 **이진 결과(binary outcome)**는 흔히 성공과 실패로 표현한다.
  - 예를 들면 동전의 앞면이 나온다면 **성공(success)**, 뒷면이 나온다면 **실패(failure)**이다.
  - 오늘 하루종일 비가 온다면, 비행기가 취소된다면, 집에 가는 길에 만원짜리 지폐를 발견한다면 성공이다.
  - 성공과 실패는 일반적인 의미의 성공과 실패와는 연관이 없을 수도 있다. (비행기가 취소되었을 때에 성공이라고 표현하는 것은 공항 근처의 숙박업소를 운영하지 않는한 일반적인 의미는 아니므로⋯)

통계학의 convention

- 성공의 결과에 1이라 하고 실패의 결과를 0이라고 흔히 해석한다.
- 동전의 앞면이 나온 것을 1이라고 하고 그것의 확률이 p라면, 동전을 1회 던져서 앞면이 나온 횟수가 1회일 확률이 p이다.
- 성공과 실패로 나누고 한 쪽 결과의 갯수를 세는것이 쉽다.
- **바꾸어 말하면, 베르누이 시행은 1회의 시행을 해서 시행에 성공할 횟수를 측정하는 것이다.**

베르누이의 확장

- 베르누이 시행을 1회가 아니라 여러번 같은 시행을 반복했을 경우로 확장할 수 있다.
- 이진 결과에 대한 1회의 시행으로 정의되는 베르누이 시행을 여러번 반복하여 성공한 횟수를 모두 더하는 것을 이항시행이라고 한다.
- 이를 표현할 수 있는 분포가 **이항 분포(binary distribution)**이며, 이어서 다룬다.

II. 순열과 조합 (permutation and combination)

- 베르누이 분포를 이항 분포로 연결시키기 위해서는
  - 확률에 관련된 개념인 **경우의 수**와
  - **순열**과 **조합**에 대한 이해가 필요하다.
  - 이번 절에서는 이에 대해서 논의한다.

경우의 수 (number of cases)

- 경우의 수는 어떤 사건이 일어날 수 있는 경우의 가짓수를 수로 표현한 것이다.
  - 동전을 던진다면 앞면과 뒷면이 나오는 2가지 경우의 결과가 있으므로 동전을 던지는 것의 경우의 수는 2이다.
  - 마찬가지로 주사위를 한 번 던지는 시행의 경우의 수는 6이다.

순열 (Permutation)

두개의 카드

- 숫자가 적힌 카드 2개를 생각해보자. 각 카드에는 숫자 1과 2가 적혀있고 이를 나열하는 경우의 수는 아래의 2가지이다.
  1. 숫자 1의 카드를 먼저 나열하고 그 다음에 숫자 2의 카드를 나열한다.
  2. 숫자 2의 카드를 먼저 나열하고 그 다음에 숫자 1의 카드를 나열한다.

세개의 카드

- 이번에는 숫자 1,2,3이 각각 적혀있는 카드 3개를 생각해보자. 이들을 나열하는 경우의 수는 아래의 6가지 이다.
  1. 숫자 1 카드 - 숫자 2 카드 - 숫자 3 카드
  2. 숫자 1 카드 - 숫자 3 카드 - 숫자 2 카드
  3. 숫자 2 카드 - 숫자 1 카드 - 숫자 3 카드
  4. 숫자 2 카드 - 숫자 3 카드 - 숫자 1 카드
  5. 숫자 3 카드 - 숫자 1 카드 - 숫자 2 카드
  6. 숫자 3 카드 - 숫자 2 카드 - 숫자 1 카드

사전식 배열

- 위의 나열들은 중복되지 않는 경우라면 가급적 낮은 숫자의 카드가 먼저나오도록 배열을 한 것이다. 이를 사전식 배열(lexicographical order)이라고 하며, 경우의 수를 산출하는 데에 자주 사용되는 방법이다.

A more clever approach

- 사전식 배열을 사용하지 않고 다른 방식으로 이해할 수도 있다.
  1. 위의 3개 카드를 이용한 경우를 보면 먼저 처음에 올 숫자 카드를 3카드 중에서 하나를 정한다.
  2. 그리고 그 다음에 올 숫자 카드를 나머지 2카드 중에서 정한다.
  3. 마지막으로 하나 남은 1개의 카드를 이용해서 나열을 완성한다.
  4. 따라서 경우의 수는 $3 \times 2 \times 1 = 6$이 된다.
- 숫자 1,2,3,4의 카드를 이용한다면 같은 원리로,
  1. 처음에 올 숫자 카드를 4카드 중에서 하나를 정하고,
  2. 다음 카드는 나머지 3카드 중에서 정하고,
  3. 그 다음은 2카드 중에서 정한다.
  4. 마지막으로 남은 1개의 카드를 이용해서 나열을 완성한다.
  5. 따라서 경우의 수는 $4 \times 3 \times 2 \times 1 = 24$가 된다.

Exercise 1

- 숫자 1,2,3,4의 카드 4장을 이용하여 사전식 배열의 경우의 수 24가지를 제시하라.

Factorial (팩토리얼, 계수)

- 같은 원리로 숫자 1-5가 적혀있는 5장의 카드를 이용한 나열은 $5 \times 4 \times 3 \times 2 \times 1 = 120$이 된다.
- 어떤 자연수 n에서 시작해서 1씩 빼면서 계속 곱해서 1까지 곱하는 연산을 팩토리얼 (factorial)이라고 하며, $n!$이라고 적는다. 팩토리얼은 아래의 식으로 정의된다.

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 1$$

- 지금까지 등장한 나열 문제들의 경우의 수는 $2! = 2, 3! = 6, 4! = 24, 5! = 120$으로 계산할 수 있다.
- 숫자 $1 \sim n$이 적힌 카드를 모두 이용하여 배열을 만드는 경우의 경우의 수는 $n!$이다.

- 그렇다면 숫자 $1 \sim 5$가 적혀있는 5장의 카드를 모두 이용하지 않고, 이중에 2장의 카드를 이용해서 나열하는 경우의 수는 몇 가지일까? 이를 사전식으로 배열하면 다음과 같다.

1. 숫자 1 카드 - 숫자 2 카드
2. 숫자 1 카드 - 숫자 3 카드
3. 숫자 1 카드 - 숫자 4 카드
4. 숫자 1 카드 - 숫자 5 카드
5. 숫자 2 카드 - 숫자 1 카드
6. 숫자 2 카드 - 숫자 3 카드
7. 숫자 2 카드 - 숫자 4 카드
8. 숫자 2 카드 - 숫자 5 카드
9. 숫자 3 카드 - 숫자 1 카드
10. 숫자 3 카드 - 숫자 2 카드
11. 숫자 3 카드 - 숫자 4 카드
12. 숫자 3 카드 - 숫자 5 카드
13. 숫자 4 카드 - 숫자 1 카드
14. 숫자 4 카드 - 숫자 2 카드
15. 숫자 4 카드 - 숫자 3 카드
16. 숫자 4 카드 - 숫자 5 카드
17. 숫자 5 카드 - 숫자 1 카드
18. 숫자 5 카드 - 숫자 2 카드
19. 숫자 5 카드 - 숫자 3 카드
20. 숫자 5 카드 - 숫자 4 카드

- 사전식 배열을 이용하면 경우의 수가 20가지 임을 알 수 있다.
- 즉, 첫 번째 카드로서 5개 중에 하나를 선택하고 두 번째 카드로서 나머지 4개의 카드 중에서 하나를 선택하기 때문에 $5 \times 4 = 20$이 된다.

Using factorial

- 즉, 팩토리얼의 연산을 이용한 모든 카드의 나열에서 5부터 하나씩 작은 숫자를 곱하면서 1까지 진행하였던 것과 대조적으로
- 나열해야할 카드의 갯수가 2개 이므로, 첫 번째 카드의 경우의 수 5와 두 번째 카드의 경우의 수 4를 곱하여 20이 계산된다. 즉, **5**부터 시작해서 1씩 줄여가며 총 **2**개의 숫자를 곱하여 $5 * (5-1) = 20$을 구한것이다.
- 경우의 수의 계산에서 자주 쓰이는 팩토리얼을 이용해서도 이를 표현할 수 있다! 아래의 수식을 잘 살펴보라.

$$5 \times 4 = \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1} = \frac{5!}{3!} = \frac{5!}{(5-2)!}$$

- 위의 수식에서는 숫자 5와 2만을 이용해서 $\frac{5!}{(5-2)!}$의 수식을 도출해 내었고, $5! = 120, 3! = 6$을 이용하여 경우의 수가 20임을 알 수 있다.

Permutation (순열)

- 이를 일반화 시키면 숫자 $1 \sim n$이 적힌 카드에서 총 r장의 카드를 이용하여 나열을 만드는 경우의 경우의 수는 $\frac{n!}{(n-r)!}$과 같다.
- 이처럼 총 n개의 서로 다른 원소중에서 r개를 이용해서 나열하는 것을 **순열 (permutation)**이라고 하며, 순열의 앞글자인 "P"를 따서 ${}_nP_r$과 같이 적는다. 순열의 공식을 아래에 다시 한번 적는다.

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

Problem 2

- 공식의 적용을 연습하기 위해서 $1 \sim 7$의 숫자가 적혀있는 카드중에서 3개를 나열하는 경우의 수를 계산해보자.

$$_7P_3 = \frac{7!}{(7-3)!} = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1} = 7 \cdot 6 \cdot 5 = 210$$

조합 (combination)

Motivation

- 순열에서는 원소를 나열함에 있어서 순서를 고려한다. (숫자 1 카드 - 숫자 2 카드로 나열하는 것과 숫자 2 카드 - 숫자 1 카드로 나열하는 것은 순열 관점에서 서로 다르다.)
- 순서에 상관없이 나열하는 경우에는 어떻게 되는지 알아보자.

Development

- 숫자 $1 \sim 5$가 각각 적혀있는 5장의 카드를 다시 생각해보자.
- 이중에 2장의 카드를 선택하여 나열하는 경우의 수는 위의 공식을 이용하여 ${}_5P_2 = \frac{5!}{(5-2)!} = \frac{120}{6} = 20$이 된다.
- 만약에 순서를 고려하지 않고 어떤 카드를 사용했나만 고려한다면, 앞의 사전식 나열의 1번과 (숫자 1 카드 - 숫자 2 카드)와 5번이 (숫자 2 카드 - 숫자 1 카드)가 동일하다.
- 이를 표로 정리하면 다음과 같다.

No.	순열 (순서 고려)	조합 (순서를 고려하지 않음)
1.	숫자 1 카드 - 숫자 2 카드	숫자 1 카드, 숫자 2 카드
5.	숫자 2 카드 - 숫자 1 카드	(위와 같음)
2.	숫자 1 카드 - 숫자 3 카드	숫자 1 카드, 숫자 3 카드
9.	숫자 3 카드 - 숫자 1 카드	(위와 같음)
3.	숫자 1 카드 - 숫자 4 카드	숫자 1 카드, 숫자 4 카드
13.	숫자 4 카드 - 숫자 1 카드	(위와 같음)
4.	숫자 1 카드 - 숫자 5 카드	숫자 1 카드, 숫자 5 카드
17.	숫자 5 카드 - 숫자 1 카드	(위와 같음)
6.	숫자 2 카드 - 숫자 3 카드	숫자 2 카드, 숫자 3 카드
10.	숫자 3 카드 - 숫자 2 카드	(위와 같음)
7.	숫자 2 카드 - 숫자 4 카드	숫자 2 카드, 숫자 4 카드
14.	숫자 4 카드 - 숫자 2 카드	(위와 같음)
8.	숫자 2 카드 - 숫자 5 카드	숫자 2 카드, 숫자 5 카드
18.	숫자 5 카드 - 숫자 2 카드	(위와 같음)
11.	숫자 3 카드 - 숫자 4 카드	숫자 3 카드, 숫자 4 카드
15.	숫자 4 카드 - 숫자 3 카드	(위와 같음)
12.	숫자 3 카드 - 숫자 5 카드	숫자 3 카드, 숫자 5 카드
19.	숫자 5 카드 - 숫자 3 카드	(위와 같음)
16.	숫자 4 카드 - 숫자 5 카드	숫자 4 카드, 숫자 5 카드
20.	숫자 5 카드 - 숫자 4 카드	(위와 같음)

조합(combination)

- 위의 표와 같이 순서를 고려하지 않고 선택하는 것의 경우의 수를 **조합(combination)** 이라고 한다.
- 위의 표의 오른쪽 컬럼은 숫자 카드 5개 중에서 2개의 카드를 선택하는 경우의 수를 의미한다.
- 왼쪽 컬럼에 비해서 2배로 경우의 수가 작은 것을 알 수 있는데, 그 이유는 왼쪽 컬럼에서 2개의 수를 나열하는 데 있어서의 경우의 수가 각각 2개이기 때문이다.
  - 예를 들어 왼쪽에서 [12. 숫자 3 카드 - 숫자 5 카드]와 [19. 숫자 5 카드 - 숫자 3 카드]의 두 경우는 순서를 고려하지 않았을 때에는 [숫자 3 카드, 숫자 5 카드]를 선택한 것과 동일하다.
- 조합의 경우의 수는 순열의 경우의 수를 먼저 구하고, 선택하는 카드의 수를 나열하는 경우의 수로 나누어 주면 된다.
- 정리하자면, 5장의 카드에서 2장을 선택하는 경우의 수는 5장으로 2장을 나열하는 경우의 수 ${}_5P_2 = \frac{5!}{(5-2)!} = \frac{120}{6} = 20$를 2장을 이용해 나열하는 경우의 수 $2! = 2$로 나누어 10이 된다.

Problem 3

- 5장의 카드에서 3장을 선택하는 경우의 수는 아래의 수식처럼 계산된다.

$$\frac{{}_5P_3}{3!} = \frac{5!/(5-3)!}{3!} = \frac{5!}{(5-3)!3!} = \frac{120}{2 \cdot 6} = 10$$

조합(combination)

- n장의 카드에서 r장을 선택하는 경우의 수를 **조합(combination)**의 경우의 수라 하며, 조합의 “C”를 따서 ${}_nC_r$이라고 한다.
- 순열 기호를 이용해서 아래와 같이 계산할 수 있다.

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

순열과 조합의 비교

- 5장의 카드중에서 r장을 선택하는 경우에 대해서 순열과 조합을 모두 고려해서 정리하면 아래와 같다.

[표 1] 5개의 카드를 이용한 순열과 조합의 경우의 수

r	순열(permutation)	조합(combination)
0	${}_5P_0 = \frac{5!}{(5-0)!} = 1$	${}_5C_0 = \frac{{}_5P_0}{0!} = 1$
1	${}_5P_1 = \frac{5!}{(5-1)!} = 5$	${}_5C_1 = \frac{{}_5P_1}{1!} = 5$
2	${}_5P_2 = \frac{5!}{(5-2)!} = 20$	${}_5C_2 = \frac{{}_5P_2}{2!} = 10$
3	${}_5P_3 = \frac{5!}{(5-3)!} = 60$	${}_5C_3 = \frac{{}_5P_3}{3!} = 10$
4	${}_5P_4 = \frac{5!}{(5-4)!} = 120$	${}_5C_4 = \frac{{}_5P_4}{4!} = 5$
5	${}_5P_5 = \frac{5!}{(5-5)!} = 120$	${}_5C_5 = \frac{{}_5P_5}{5!} = 1$

- 팩토리얼의 계산에서 $0! = 1$이 사용되었음을 유의하라.
- 조합의 경우의 수는 순열의 경우의 수를 $r!$로 나눈 것과 같다.
- 조합의 컬럼에서 확인할 수 있듯이, ${}_5C_0 = {}_5C_5$, ${}_5C_1 = {}_5C_4$, ${}_5C_2 = {}_5C_3$이다.
  - 5장의 카드에서 3장을 뽑는 경우의 수와 2장을 뽑는 경우의 수가 같다.
  - 이는 뽑을 카드 3장을 선택하는 것과 뽑지 않은 카드 2장을 선택하는 것이 결국 같은 행위이기 때문이다.
  - 일반화 시키면 ${}_nC_r = {}_nC_{n-r}$이 항상 성립한다.

- 조합의 ${}_nC_r$은 n개 중에서 r개를 선택하는 경우의 수이며,
- "n chooses r"이라 읽는 아래와 같은 표기로도 사용한다.

$${}_nC_r = \binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Example

만약에 나에게 셔츠가 7벌 있다고 하자. 다음주 월요일, 화요일, 수요일에 각각 다른 셔츠를 입는 경우의 수는 ${}_7P_3$ 이 된다. 만약에 여행을 가기 위해 3벌을 가져간다고 하면, 3벌을 선택하여 가방에 넣는 경우의 수는 ${}_7C_3$ 가 된다.

III. 이항분포 (Binomial Distribution)

III. 이항분포 (Binomial Distribution)

Motivation

- 앞의 베르누이 분포에서는 동전을 1회만 던지는 시행에 대해서 다루었다.
- 이번에는 동전을 3회 던지는 시행을 생각해보자. 동전의 앞면을 H, 뒷면을 T라고 한다면 동전을 던져서 나올 수 있는 경우의 수는 아래 8가지로 정리된다. 역시 사전식 배열을 사용한 점을 확인하라.
    1. HHH
    2. HHT
    3. HTH
    4. HTT
    5. THH
    6. THT
    7. TTH
    8. TTT

- 만약에 동전의 앞면이 나올 확률이 p라고 한다면 뒷면이 나올 확률은 $1-p$이다.
  - 그렇다면 1번의 사건(HHH)이 일어날 확률은 $p \times p \times p = p^3$이고,
  - 2번의 사건(HHT)이 일어나 확률은 $p \times p \times (1-p)$이다.
  - 이들 확률을 표로 정리하면 아래와 같다.

[표 2] 동전을 3번 던지는 시행의 결과와 확률1

	사건	확률	앞면이 나온 횟수 (성공횟수)
1.	HHH	ppp	3
2.	HHT	$pp(1-p)$	2
3.	HTH	$p(1-p)p$	2
4.	HTT	$p(1-p)(1-p)$	1
5.	THH	$(1-p)pp$	2
6.	THT	$(1-p)p(1-p)$	1
7.	TTH	$(1-p)(1-p)p$	1
8.	TTT	$(1-p)(1-p)(1-p)$	0

- 위의 표는 사전식 배열을 이용하였다. 앞면이 나온 횟수로 정렬하면 아래의 표를 만들수 있다.

[표 3] 동전을 3번 던지는 시행의 결과와 확률2

앞면의 수	사건	확률
3	HHH	ppp
2	HHT	$pp(1-p)$
2	HTH	$p(1-p)p$
2	THH	$(1-p)pp$
1	HTT	$p(1-p)(1-p)$
1	THT	$(1-p)p(1-p)$
1	TTH	$(1-p)(1-p)p$
0	TTT	$(1-p)(1-p)(1-p)$

- 앞면이 2번 나오는 사건들은 HHT, HTH, 혹은 THH에 해당하므로 앞면이 2번 나올 확률은 $pp(1-p)+p(1-p)p+(1-p)pp = 3p^2(1-p)$ 이다.
- 개별 사건이 아니라 앞면이 나오는 총 횟수에 대해서만 정리하면 아래의 표를 만들 수 있다.

[표 4] 동전을 3번 던지는 시행의 결과와 확률3

앞면의 수	확률
3	p^3
2	$3p^2(1-p)$
1	$3p(1-p)^2$
0	$(1-p)^3$

- 만약에 동전의 앞면이 나올 확률과 뒷면이 나올 확률이 같다면 (fair coin 이라면),
- 즉 $p = 1 - p$이고 $p = 0.5$라면, 위의 표에 대입하여 아래의 결과를 얻는다.

[표 5] Fair한 동전을 3번 던지는 시행의 결과와 확률

앞면의 수	확률
3	$1/8$
2	$3/8$
1	$3/8$
0	$1/8$

Discussion

- 페어한 동전(앞면이 나올 확률이 0.5인 동전)을 4번 던져서 앞면이 2번 나올 확률은 얼마인가?
  - 만약에 2/4이니까 50%라고 답했다면,
  - 동전을 100번 던졌을 때에 정확히 50번의 앞면이 나오는 것이 얼마나 어려운지 생각해보라.
  - 만약에 상상이 잘 안된다면 우리나라 사람이 각자 자신이 가지고 있는 500원 짜리를 던졌고 그 동전의 갯수가 2억개일때, 정확히 1억개가 앞면이고 1억개가 뒷면일 확률을 생각해보라. 매우 낮지 않은가?
  - 이런 이유로 동전을 불과 4번을 던졌을 때에도 앞면이 2번 나올 확률은 50%보다 낮다.

- 이항분포는 베르누이 시행의 성공확률이 p라고 할때에, 같은 시행을 n번하여 r번의 성공을 기록할 확률에 관한 분포이다.
  - 위의 수치 1/8은 성공확률(앞면이 나올 확률)이 0.5인 3번의 시행에서 3번의 성공(앞면이 나옴)을 기록할 확률이고
  - 3/8은 성공확률이 0.5인 3번의 시행에서 2번의 성공(앞면이 나옴)을 기록할 확률이다.
  - 일반화된 이항 분포의 도출을 위해서 다시한번 앞에서 나온 [표 3]을 살펴보자.

[표 3] 동전을 3번 던지는 시행의 결과와 확률2

앞면의 수	사건	확률
3	HHH	ppp
2	HHT	$pp(1-p)$
2	HTH	$p(1-p)p$
2	THH	$(1-p)pp$
1	HTT	$p(1-p)(1-p)$
1	THT	$(1-p)p(1-p)$
1	TTH	$(1-p)(1-p)p$
0	TTT	$(1-p)(1-p)(1-p)$

- 앞면이 나온 횟수가 3인 경우에는 1가지 경우 (HHH) 밖에 없으므로 확률은 p^3이 된다.
- 앞면이 나온 횟수가 2인 경우는 3가지 경우 (HHT, HTH, THH)가 있으므로 확률은 각각의 확률 $p^2(1-p)$를 3번 더하여 $3p^2(1-p)$가 된다.
  - 여기에서 3이라는 숫자는 어떻게 나온 것일까? 이는 조합을 이용해서 설명할 수 있다.
  - 3번의 시행중에서 앞면이 2회 나오므로, 3개의 시행중에서 2개를 뽑는 것과 경우의 수가 같다. 즉, ${}_3C_2 = 3$이다.

[표 6] 동전을 3번 던지는 시행의 결과와 확률4

앞면의 수	확률
3	${}_3C_3p^3 = \frac{3!}{0!3!}p^3 = p^3$
2	${}_3C_2p^2(1-p) = \frac{3!}{1!2!}p^2(1-p) = 3p^2(1-p)$
1	${}_3C_1p(1-p)^2 = \frac{3!}{2!1!}p(1-p)^2 = 3p(1-p)^2$
0	${}_3C_0(1-p)^3 = \frac{3!}{3!0!}(1-p)^3 = (1-p)^3$

- **즉, 성공확률이 p 인 개별 시행을 n 번 반복하여 총 r 번의 성공을 기록할 확률은 아래와 같다.**

$$ {}_nC_rp^r(1-p)^{n-r} = \binom{n}{r}p^r(1-p)^{n-r} $$

Problem 4

1. Fair coin을 4회 던져서 2회 앞면이 나올 확률을 구하라.
2. Fair coin을 6회 던져서 3회 앞면이 나올 확률을 구하라.

- **성공확률이 p인 개별 시행을 n번 반복하여 총 r번의 성공을 기록할 확률은 아래와 같다.**

$$ {}_nC_r p^r (1-p)^{n-r} = \binom{n}{r} p^r (1-p)^{n-r} $$

- 확률 변수를 등장시켜 좀 더 정교하게 위의 확률을 표현해보자.
  - 성공확률이 p이고 실패확률이 $1-p$인 이진 결과를 나타내는 시행(베르누이 시행)이 있다.
  - 이 시행을 n번 반복하여 성공한 횟수를 확률 변수 X라고 하자.
  - 확률 변수 X가 r과 같을 확률은, (그러니까 총 n번의 시행에서 r번을 성공할 확률은) 아래와 같다.

$$ \mathbb{P}(X=r) = p(r) = \begin{cases} {}_nC_r p^r (1-p)^{n-r} = \binom{n}{r} p^r (1-p)^{n-r} & \text{if } r \in \{0,1,2,...,n\} \\ 0 & \text{otherwise} \end{cases} $$

- 이항 분포를 따른 변수를 이항 변수라고 부르는데, 이항 변수는 베르누이 시행을 n번 반복한 결과에 해당한다. 그러므로 이항 변수를 베르누이 변수의 합으로 표현할 수 있다.
  1. 모수가 모두 p로 같은 n개의 베르누이 시행을 생각해보자.
  2. $B_1, B_2, \cdots, B_n$은 각각 $Ber(p)$를 따르는 확률 변수이다. 즉, $B_1 \sim Ber(p)$, $B_2 \sim Ber(p), \cdots, B_n \sim Ber(p)$이다.
  3. B_1부터 B_n을 모두 더한 변수를 X라고 한다면 X는 모수가 n과 p인 이항변수이고, 아래와 같이 쓴다.

$$X = B_1 + B_2 + \cdots + B_n \sim bin(n, p)$$

이항 분포의 사례

- 어떤 동전이 있는데 약간 찌그러져 있어서 앞면이 많이 나올 것 같다.
- 이런 경우에 앞면이 많이 나오는 동전이라고 확신하기 위해서는 어떻게 해야할까? 여러번 던져서 결과를 관찰하는 방법이 가장 적합할 것이다.
- 10번을 던졌더니 앞면이 6번 나왔다. 이 동전은 앞면이 많이 나오는 동전이라고 확신할 수 있을까? 즉, 여러분이 가진 전재산의 상당 부분을 앞면이 나온다에 걸 수 있는가? 아마도 10번 밖에 안 던져 보았기 때문에 매우 높은 확신을 가지고 큰 돈을 걸수는 없을 것이다.
- 만약에 10번을 던졌더니 앞면이 8번 나왔다면 아까보다 더 확신을 가질수 있겠지만 여전히 매우 높은 확신을 가지고 큰 돈을 걸 수는 없을 것이다.
- 이번에는 실험횟수를 늘려서 10000번을 던져보았더니 8000번 앞면이 나왔다. 이 경우에는 이 동전이 정말로 앞면이 많이 나오는 동전임을 확신할 수 있을 것이다. 여유자금으로 가지고 있는 돈이 있다면 걸어볼만 하지 않을까?

- 위에서 소개한 3번의 실험 결과들은 아래와 같다.
  1) 10번을 던져서 앞면이 6번
  2) 10번을 던져서 앞면이 8번
  3) 10000번을 던져서 앞면이 8000번
- 3)-2)-1) 순서대로 찌그러진 동전에서 앞면이 많이 나올 것이라는 확신을 강하게 준다.
- 이런 확신의 정도를 수치로 제공하는 것이 기초통계학의 추정과 검정에 해당하며 이후의 챕터에서 다루게 된다. 그리고 동전에 문제에 관한 답을 내리기 위해서 사용하는 분포가 바로 이항분포이다.

IV. Exercises

Problem 5

앞면과 뒷면이 나올 확률이 동일한 동전을 던지는 게임을 한다. 아래의 두 종류의 게임이 있다.

- **Game 1**: 시행의 45%-55% 사이에서 앞면이 나온다면 게임을 이긴다.
- **Game 2**: 시행의 60% 이상에서 앞면이 나온다면 게임을 이긴다.

각각의 게임에 대해서 아래의 선택이 주어진다.

- **Choice 1**: 동전을 20회 던진다. $X \sim bin(20, 0.5)$
- **Choice 2**: 동전을 200회 던진다. $X \sim bin(200, 0.5)$

각각의 게임에서 Choice 1과 Choice 2중에서 어떤 것을 선택하는 것이 좋은 전략인지 서술하라.

Partial Solution to Problem 5

For Game 1,

- Choice 1의 승리확률
  - $X \sim bin(20, 0.5)$일 때에 $\mathbb{P}(9 \leq X \leq 11)$를 계산한다.

```
sum(dbinom(x = 9:11, size = 20, prob = 0.5))
```

```
## [1] 0.4965553
```

- Choice 2의 승리확률
  - $X \sim bin(200, 0.5)$, $\mathbb{P}(90 \leq X \leq 110)$를 계산한다.

```
sum(dbinom(x = 90:110, size = 200, prob = 0.5))
```

```
## [1] 0.8626333
```

- 그러므로 정답은 Choice 2이다.

Problem 6

- 한 야구선수의 타율이 0.325이다. (타율은 규정타석에서 안타를 기록할 확률을 의미한다.) 이 선수가 5개의 규정타석에서 1개 이상의 안타를 기록할 확률을 구하라.

Solution to Problem 6

- $X \sim bin(5, 0.325)$, $\mathbb{P}(X \neq 0) = 1 - \mathbb{P}(X = 0) = 1 - (1 - 0.325)^5$

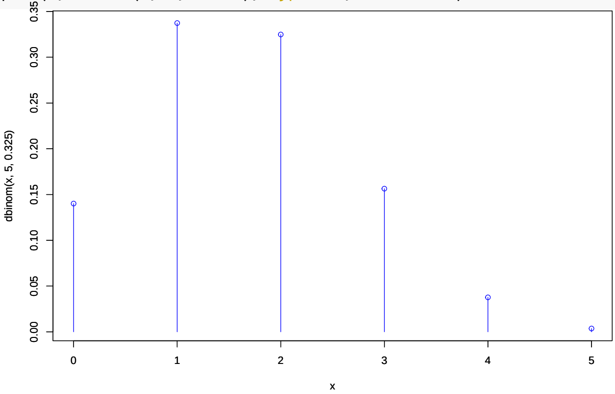
```
sum(dbinom(x = 1:5, size = 5, prob = 0.325))
```

```
## [1] 0.859874
```

- 85.98%의 확률로 1개 이상의 안타를 기록할 것이다.

- (참고: 김선희 원우님이 2019년에 그린 binomial에 대한 차트)

```
x <- seq(0, 5, by=1)
plot(x, dbinom(x, 5, 0.325), type='p', col='blue')
par(new=TRUE)
plot(x, dbinom(x, 5, 0.325), type='h', col='blue')
```



Problem 7

한 상점에 방문하는 고객의 35%가 쿠폰을 사용한다. 7명의 고객이 방문하여 4명 이하의 고객이 쿠폰을 사용할 확률을 구하라.

Problem 8

한 제조시설에서 박스단위로 검수가 이루어진다. 한 박스에서 무작위로 3개의 물건을 검수하여 2개가 합격이면 그 박스 전체를 통과시킨다. 어떤 한 박스에 실제로 10%의 불량품이 들어있다. 이 박스가 검수에서 통과할 확률을 구하라.

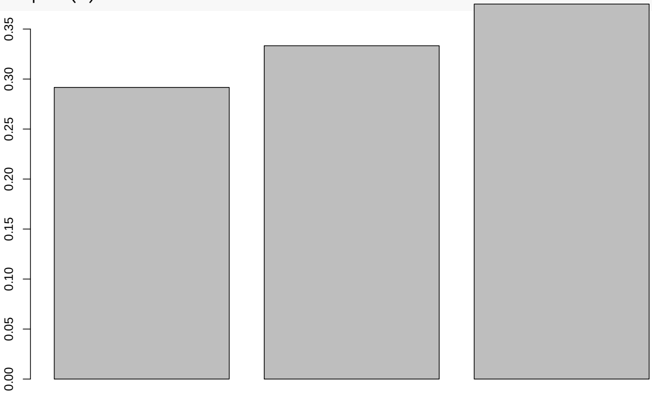
- Hint
  - 불량품의 갯수를 X라 하고, $X \sim bin(3, 0.1)$, $\mathbb{P}(X \leq 1)$
  - 양품의 갯수를 X라 하고, $X \sim bin(3, 0.9)$, $\mathbb{P}(X \geq 2)$

Problem 9

- 이산 확률변수의 pmf가 $\mathbb{P}(X = x) = p(x) = (x + 6)/24$, for $x = 1, 2$, and 3 이가 될 수 있는가?
- 즉, 해당 함수는 pmf 함수의 특징을 만족하는가?
  1) $p(x) \geq 0$ for all x
  2) $\sum p(x) = 1$

- (참고: 노석현 원우님이 2019년에 그린 차트)

```
x <- (7:9)/24
barplot(x)
```



Problem 10

- 어떤 이산 확률변수 X는 0,1,2,$\cdots$,10의 값을 가질 수 있으며 pmf가 아래와 같다.

$$P(X = k) = p(k) = \binom{10}{k}(0.37)^k(0.63)^{10-k}$$

- 이 확률 변수의 평균은 얼마인가?

Solution to Problem 10

- 확률 변수의 분포는 $X \sim bin(10, 0.37)$이다.
- 그러므로 평균은 10*0.37=3.7이다.