강의교안 이용 안내

- 본 강의교안의 저작권은 **이윤환과 한빛아카데미㈜**에 있습니다.
- 이 자료를 무단으로 전제하거나 배포할 경우 저작권법 136조에 의거하여 벌금에 처할 수 있고 이를 병과(倂科)할 수도 있습니다.

CHAPTER 03

확률과 확률분포

한빛아카데미 Hanbit Academy, Inc.

Contents

01. 확률

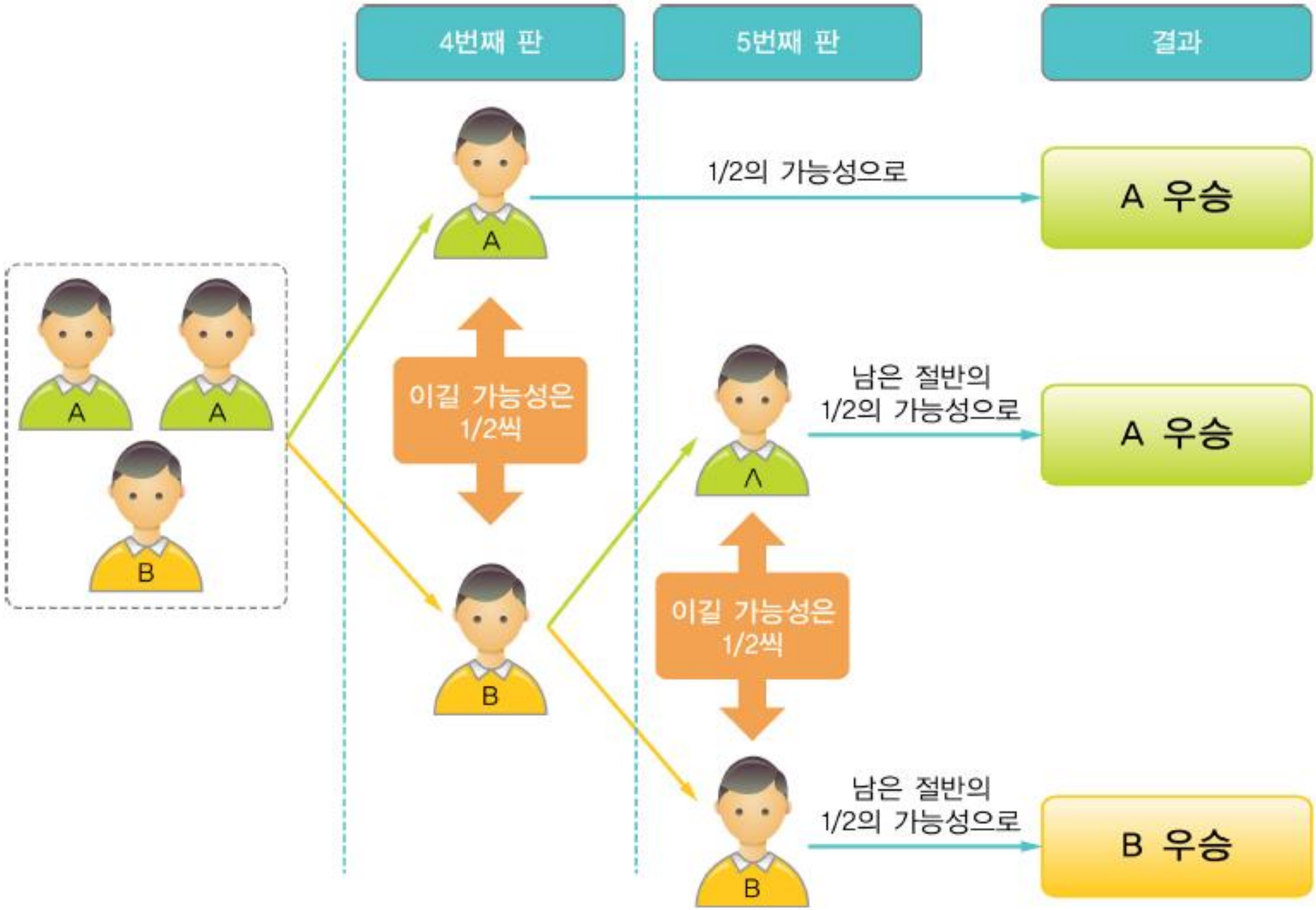
: 우연이 아닌 과학

1. 확률을 정의하고 확률법칙을 학습한다.
2. 확률변수에 대해 학습한다.

확률

- **어떻게 나눠야 불만이 없을까요?**
  - 서로 능력이 동일한 두 명의 게이머 A와 B가 있습니다.
  - 5판 3선승제의 게임을 통해 우승자를 가리는 게임이 시작되었습니다.
    - 우승자에게는 64개의 게임 캐릭터 배지를 상품으로 주는 이벤트 대결입니다.
  - 3판이 지난 지금 A가 두 번을 이겼고 B가 한 번을 이겼습니다.
  - 4번째 판이 시작되려는 순간 게임 서버의 이상으로 더 이상 게임을 진행할 수 없게 되었습니다.
  - A가 우승했다고 하기도 곤란하고, B가 우승했다고 하기도 곤란한 상황입니다.
  - 상품을 어떻게 나누는 것이 좋을까요?

확률


4번째 판
5번째 판
결과
A
A
A
B
1/2의 가능성으로
A 우승
이길 가능성은
1/2씩
A
B
남은 절반의
1/2의 가능성으로
A 우승
A
이길 가능성은
1/2씩
B
남은 절반의
1/2의 가능성으로
B 우승

확률

- **확률이라는 단어**
  - 確率(확률) : 굳을 확, 비율 률
    - ‘(어떤 결정 등을) 굳힐 비율’
  - probability
    - probable : (명) ‘(어떤 일이) 있을 것 같은’, ‘개연성 있는’
    - ‘개연성’ 혹은 ‘개연성 있는 일’
      - 개연성의 사전적 의미 : 어떤 일이 일어날 수 있는 확실성의 정도

확률

- **용어**
  - 확률실험 : E
    - 다음의 세가지를 만족할 때 확률실험 혹은 확률시행이라고 합니다.

❶ (결과를 구하기 위한) 어떤 실험을 통해 나타나는 **결과를 알지 못한다**.
❷ 결과는 알지 못하지만 결과로 나타날 수 있는 **가능한 경우를 알고 있다**.
❸ 동일한 실험을 몇 번이고 **반복할 수 있다.**

- 예제 : 동전던지기

❶ 동전을 던지기 전에 '앞면'이 나올지 '뒷면'이 나올지 **알 수 없습니다.**
❷ 가능한 결과는 **'앞면'과 '뒷면' 중에 하나**임을 알고 있습니다.
❸ 동전을 던지는 실험은 몇 번이고 **반복**할 수 있습니다.

확률

- **용어**
  - 표본공간 : Ω
    - 확률실험으로부터 출현 가능한 모든 결과들의 모임을 표본공간이라 합니다.
    - 예제 : 동전던지기
      - Ω = {앞면, 뒷면} = {H, T}
  - 사건 : 기호 알파벳 대문자
    - 표본공간의 각 원소(즉, 출현 가능한 개별 결과)들의 부분집합을 사건이라 합니다.
    - “(관심 있는)사건이 발생했다 ” : 시행 결과가 (관심 있는)사건에 속하는 경우
    - 근원사건
      - 어떤 사건이 표본공간상의 하나의 원소로 구성된 사건

확률

- 사건의 연산 : 임의의 두 사건 A, B에 대해
  - 합사건
    - 어떤 사건의 발생이 사건 A에서 일어나거나 혹은 사건 B에서 일어나는 사건
    - $A \cup B = \{\omega \mid \omega \in A \ or \ \omega \in B\}$
  - 곱사건
    - 어떤 사건의 발생이 사건 A와 사건 B에서 동시에 일어나는 사건
    - $A \cap B = \{\omega \mid \omega \in A \ or \ \omega \in B\}$
  - 여사건 : A^c
    - 사건 A가 발생하지 않을 사건
    - $A^c = \{\omega \mid \omega \notin A\}$
  - 배반사건
    - 두 사건이 겹치는 부분이 없는 즉, 동시에 발생하지 않는 사건($A \cap B = \phi$)
  - ❖ 독립사건
    - ❖ 두 사건이 서로의 발생에 영향을 끼치지 않는 사건

확률

- 예 : 확률실험, 표본공간, 사건
  - E_1 : 동전을 2번 던져 나오는 면 관찰
    - 표본공간 : $\Omega_1 = \{HH, HT, TH, TT\}$
    - 사건 : 첫 번째 동전이 앞면이 나오는 사건
      - $A_1 = \{HH, HT\}$
  - E_2 : 하루 중 인터넷 사용시간 관찰
    - 표본공간 : $\Omega_2 = \{0 \leq t \leq 24\}$
    - 사건 : 사용시간이 1시간 이하인 사건
      - $A_2 = \{0 \leq t \leq 1\}$

확률

- **확률의 정의**
  - 수학적 확률 (고전적 확률)

❶ 어떤 시행의 결과로 나타날 수 있는 가능한 결과의 수 : O
❷ 각 결과들이 나타날 가능성은 동일하다는 가정
❸ 동일한 각 결과들의 확률 : $\frac{1}{O}$

  - 임의의 사건 A가 발생할 수학적 확률은 표본공간의 원소의 개수(O) 중 사건 A에 해당하는 근원사건의 개수(n)입니다 ($\frac{n}{O}$).
  - 예) 주사위를 굴려 홀수가 나올 확률
    - 주사위의 각 눈이 나올 확률은 전체 6개의 눈으로 구성되어 있으며 각각이 나올 확률은 동일하다고 가정하면 확률은 1/6입니다.
    - 홀수인 사건을 구성하는 근원사건의 수는 {1, 3, 5}으로 세 개가 있습니다.
    - 전체 눈의 개수는 6이고 이로부터 홀수 눈의 확률은 3/6 = 1/2 입니다.

확률

- 통계적 확률 (고전적 확률)

❶ 동일한 조건에서 같은 실험을 N번 반복

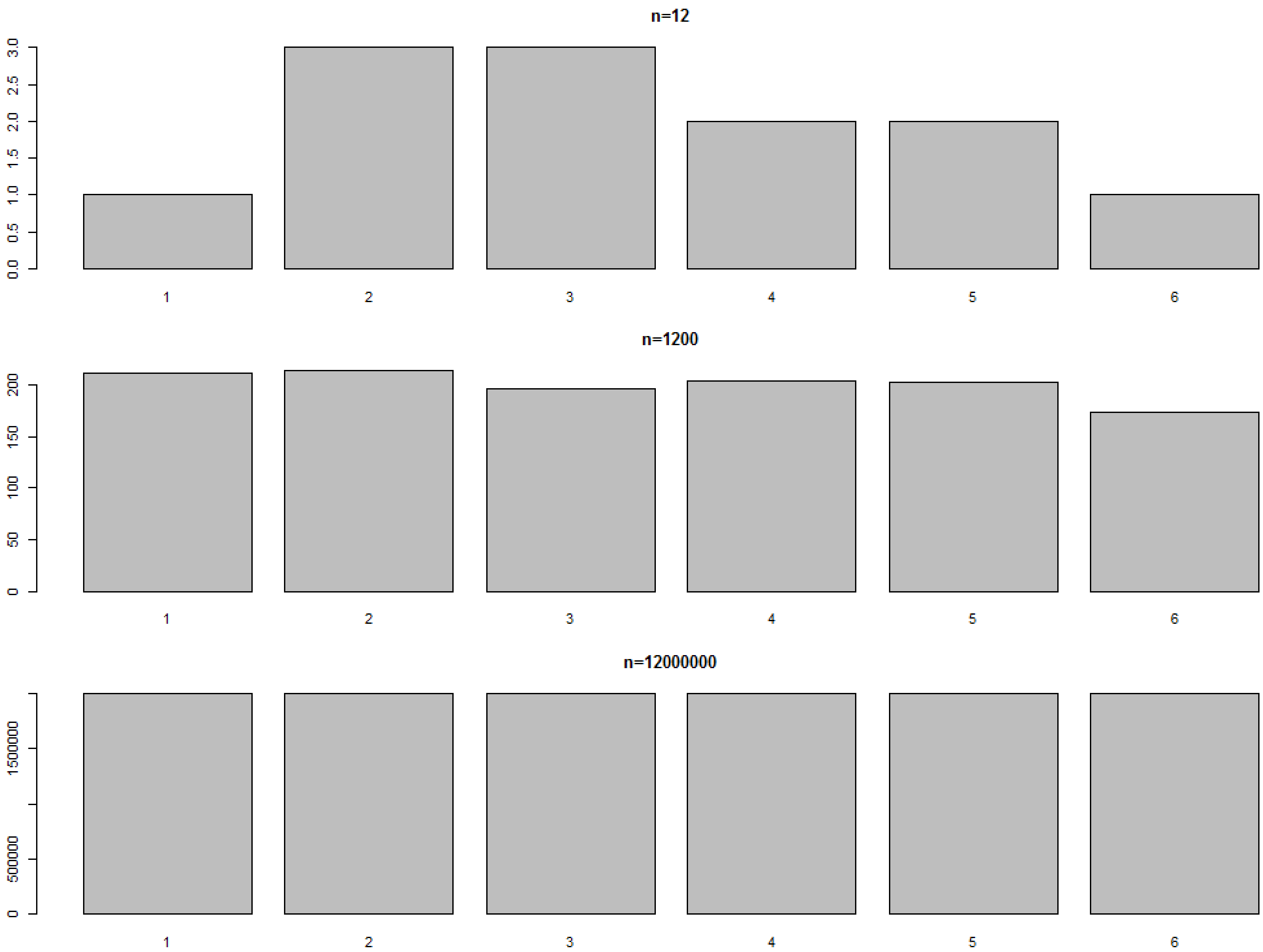
❷ 사건 A가 모두 몇 번 발생했는지를 조사 : n

❸ 사건 A가 발생할 확률 : $\mathrm{P(A)} = \frac{n}{N}$

- 실험의 반복횟수 N은 매우 커야 그 값을 받아들일 수 있으며, 반복횟수가 커짐에 따라 사건 A의 상대도수($\frac{n}{N}$)가 상수 P(A)로 접근해가는 경향을 보입니다.
- 예) 주사위를 여러 번 굴려 나온 눈을 관찰해 봅시다.

시행횟수	1의 눈	2의 눈	3의 눈	4의 눈	5의 눈	6의 눈
12	1	3	3	2	2	1
1200	211	214	196	204	202	173
12000000	2002632	1999749	2000328	1999958	1996037	2001296

확률



n=12
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3
4
5
6
n=1200
0
50
100
150
200
1
2
3
4
5
6
n=12000000
0
500000
1500000
1
2
3
4
5
6

확률

- 확률 공리 (공리적 확률)
  - 표본공간 Ω상의 임의의 사건 A에 대한 실수치 함수에 대해

❶ P(A)는 0과 1사이의 값을 갖고(0 ≤ P(A) ≤ 1),

❷ 반드시 일어나는 사건(표본공간 전체)의 값은 1이며(P(Ω) = 1),

❸ 서로 배반인 사건 $A_1, A_2, \ldots, A_n, \ldots$ 의 합집합에 대해 다음을 만족하면,

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

함숫값 P(A)를 사건 A의 확률이라 합니다.

  - 확률 공리는 확률이 만족해야 하는 기본 성질이며 이를 통해 확률 계산을 합니다.

확률

- **확률법칙**
  - 덧셈법칙
    - 임의의 사건 A와 사건 B의 합사건에 대한 확률
      - $\mathrm{P}(\mathrm{A} \cup B) = P(A) + P(B) - P(A \cap B)$
      - 만일 두 사건 A와 B가 서로 배반이라면 $(A \cap B = \phi)$
        - $\mathrm{P}(\mathrm{A} \cup B) = P(A) + P(B)$
  - 곱셈법칙
    - 조건부 확률
      - 두 사건 A와 B에 대해
        - P(A | B) : 사건 B가 발생했을 때 사건 A가 발생할 확률
          - $P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0$
        - P(B | A) : 사건 A가 발생했을 때 사건 B가 발생할 확률
          - $P(B|A) = \frac{P(A \cap B)}{P(A)}, \qquad P(A) > 0$

확률

- 조건부 확률 예)
  - 주사위를 던지는 실험에서 주사위의 눈이 짝수인 사건을 A, 주사위의 눈이 4이상인 사건을 B라 할 때 P(A|B)를 구해봅시다.
    - $P(A|B) = \frac{P(A \cap B)}{P(B)}$ 이므로 P(B)와 $P(A \cap B)$를 구합니다.
      - P(B) : 주사위의 눈이 4이상이 나올 확률 = 1 / 2
      - $P(A \cap B)$: 주사위의 눈이 짝수이고 4이상인 경우는 {4, 6} 이므로 확률은 1/3 입니다.
      - $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/3}{1/2} = \frac{2}{3}$
    - 표본공간의 변화 : 표본공간으로 조건으로 주어진 사건으로 변화
      - $\Omega = \{1, 2, 3, 4, 5, 6\} \rightarrow \Omega_B = \{4, 5, 6\}$
      - 변화된 표본공간 Ω_B에서 사건 A가 발생할 확률
      - Ω_B 상에서 짝수의 눈은 {4, 6} 이므로 2 / 3

확률

- 곱셈법칙
  - 두 사건 A와 B에 대해 조건부 확률을 이용하여
    - $P(A \cap B) = \begin{cases} P(A)P(B|A), & P(A)>0 \\ P(B)P(A|B), & P(B)>0 \end{cases}$
  - 만일 두 사건 A와 B가 독립이라면,
    - $P(A \cap B) = P(A)P(B)$
    - 두 사건이 서로의 발생에 영향을 끼치지 않는다면 곱사건의 확률은 두 사건의 곱이 됩니다. (예, 동전을 두 번 던져 처음 앞면이 나온 것이 두 번째 던졌을 때 영향을 끼치지 않습니다.)
- 독립사건과 조건부 확률
  - 위에서 두 사건 A와 B가 독립이면, $P(A \cap B) = P(A)P(B)$ 이므로
    - $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$
    - 즉, 사건 B의 발생여부가 사건 A의 발생에 영향을 끼치지 않습니다.

확률

- 여사건의 확률
  - 사건 A의 여사건 A^c의 사건 $P(A^c)$는,
    - $P(A^c) = 1 - P(A)$
    - $P(A) + P(A^c) = 1$

- 예제) 주사위를 두 번 던지는 실험에서 다음의 두 사건에 대해 합사건, 곱사건의 확률을 구해 봅시다.
  - 사건 A : 첫 번째 던진 주사위의 눈이 짝수인 사건, $P(A) = \frac{18}{36} = \frac{1}{2}$
  - 사건 B : 두 번째 던진 주사위의 눈이 3의 배수인 사건, $P(B) = \frac{12}{36} = \frac{1}{3}$

확률

사건 A

첫 번째 두 번째	1	2	3	4	5	6
1	(1, 1)	(2, 1)	(3, 1)	(4, 1)	(5, 1)	(6, 1)
2	(1, 2)	(2, 2)	(3, 2)	(4, 2)	(5, 2)	(6, 2)
3	(1, 3)	(2, 3)	(3, 3)	(4, 3)	(5, 3)	(6, 3)
4	(1, 4)	(2, 4)	(3, 4)	(4, 4)	(5, 4)	(6, 4)
5	(1, 5)	(2, 5)	(3, 5)	(4, 5)	(5, 5)	(6, 5)
6	(1, 6)	(2, 6)	(3, 6)	(4, 6)	(5, 6)	(6, 6)

사건 B

확률

- 합사건
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $A \cap B = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$ 이므로 $P(A \cap B) = {}^{1}/_{6}$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B) = {}^{1}/_{2} + {}^{1}/_{3} - {}^{1}/_{6} = {}^{2}/_{3}$
- 곱사건 : 사건 B를 조건으로
  - $P(A \cap B) = P(B)P(A|B)$
  - $P(A|B) = {}^{P(A \cap B)}/_{P(B)}$로 사건 B를 새로운 표본공간으로 변화시켜 봅시다.

첫 번째 / 두 번째	1	2	3	4	5	6
3	(1, 3)	$\boxed{(2, 3)}$	(3, 3)	$\boxed{(4, 3)}$	(5, 3)	$\boxed{(6, 3)}$
6	(1, 6)	$\boxed{(2, 6)}$	(3, 6)	$\boxed{(4, 6)}$	(5, 6)	$\boxed{(6, 6)}$

- 위의 표로부터 $P(A|B) = {}^{6}/_{12} = {}^{1}/_{2}$ 입니다.
- $P(A \cap B) = P(B)P(A|B) = {}^{1}/_{3} \cdot {}^{1}/_{2} = {}^{1}/_{6}$

확률

예제 3-1 prob 패키지를 이용한 확률 계산

준비파일 | 01.package_prob.R

- prob 패키지를 이용해 봅시다.
  - 동전던지기, 주사위 굴리기, 주머니 속에서 공 뽑기 등의 확률실험
  - 표본공간을 만들고 확률을 구해 봅시다.

```
① library(prob)
② tosscoin(1)
③ rolldie(1)
④ urnsamples(1:3, size=2)
⑤ urnsamples(1:3, size=2, replace=T)
⑥ urnsamples(c( rep("R", 3), rep("B", 2)), size=2)
⑦ tosscoin(2, makespace=T)
```

확률

- **코드 설명**
  - 1줄 : 설치한 패키지 prob을 현재 작업공간에서 사용하기 위해 library() 함수를 이용합니다.
  - 2줄 : 동전을 던지는 실험
    - tosscoin() 함수는 동전던지기 실험의 표본공간을 생성합니다.
    - 함수수행을 위해 두 개의 전달인자를 갖는데, 필수 전달인자는 times에 동전을 던지는 횟수를 지정합니다. 코드에서는 동전을 한 번 던질 때의 표본공간을 생성합니다
      - 동전을 한번 던지는 실험에서 표본공간은 {H, T} 입니다.

```
> tosscoin(1)
  toss1
1     H
2     T
```

확률

- 3줄 : 주사위를 굴리는 실험
  - rolldie() 함수는 주사위를 굴리는 실험의 표본공간을 생성합니다.
  - times 전달인자는 tosscoin()과 마찬가지로 필수로 주사위를 굴리는 횟수를 지정합니다.
  - 위의 코드에서는 주사위를 한 번 굴릴 때의 표본공간을 생성합니다.
    - 주사위를 한번 굴리는 실험에서 표본공간은 {1의 눈, 2의 눈, 3의 눈, 4의 눈, 5의 눈, 6의 눈} 입니다. (각 눈의 숫자를 출력합니다.)

```
> rolldie(1)
  X1
1  1
2  2
3  3
4  4
5  5
6  6
```

확률

- 4줄 : 주머니 속에 있는 공을 꺼내는 실험
  - urnsamples()는 사용자가 입력한 벡터 값의 개별 원소들로 구성된 표본공간을 생성하는 함수입니다.
    - size에 전달인자로 실험 횟수를 전달합니다.
    - 코드에서는 1, 2, 3(벡터 1:3)으로 구성된 주머니 속의 공을 꺼내는 실험을 2회 실시하였을 때의 표본공간을 생성합니다.
    - 비복원추출입니다.

```
> urnsamples(1:3, size=2)
  X1 X2
1  1  2
2  1  3
3  2  3
```

확률

- 5줄 : 비복원추출
  - urnsamples()의 전달인자 replace를 통해 복원추출(replace=T)을 실시합니다.
  - replace 전달인자를 사용하지 않은 경우 기본값은 F로 비복원추출입니다 (4줄)

```
> urnsamples(1:3, size=2, replace=T)
  X1 X2
1  1  1
2  1  2
3  1  3
4  2  2
5  2  3
6  3  3
```

확률

- 6줄
  - 문자R 3개, 문자B 2개로 구성된 추출실험을 실시합니다.
  - 붉은 구슬3개, 파란 구슬 2개로 구성된 주머니에서 구슬을 추출하는 실험을 가정한 상황으로 size=2이므로 두 개의 구슬을 추출하는 실험의 표본공간을 생성합니다

```
> urnsamples(c( rep(""R"", 3), rep(""B"", 2)), size=2)
   X1 X2
1   R  R
2   R  R
3   R  B
4   R  B
5   R  R
6   R  B
7   R  B
8   R  B
9   R  B
10  B  B
```

확률

- 7줄 : 확률 생성
  - 동전을 두 번 던지는 실험에서 생성된 표본공간에 확률을 부여합니다. (makespace=T(=TRUE))
    - 앞서 사용한 rolldie(), urnsamples() 에서도 사용합니다.
    - makespace의 기본값은 FALSE 입니다.
  - 결과로 표본공간을 구성하는 각 결과들이 발생한 확률은 모두 0.25로 동일합니다

```
> tosscoin(2, makespace=T)
  toss1 toss2 probs
1     H     H  0.25
2     T     H  0.25
3     H     T  0.25
4     T     T  0.25
```

확률변수

- 동전을 두 번 던지는 실험을 생각해봅시다.
  - 이 실험에서 나올 수 있는 결과는 {H, H}, {H, T}, {T, H}, {T, T}입니다.
  - '(동전을 두 번 던져 앞면이 나오는 횟수) × 1,000원'의 상금이 주어지는 게임을 한다고 하면,
    - 우리의 관심사는 동전을 두 번 던져 앞면이 나오는 횟수가 될 것입니다.
    - 동전을 두 번 던져 앞면이 나오는 횟수는 다음의 표 3-3과 같습니다.

확률변수

[표 3-3] 동전을 두 번 던져 앞면이 나오는 횟수 관찰

첫 번째 던진 동전	두 번째 던진 동전	표본공간	앞면이 나오는 횟수
		{H, H}	2
		{H, T}	1
		{T, H}	1
		{T, T}	0

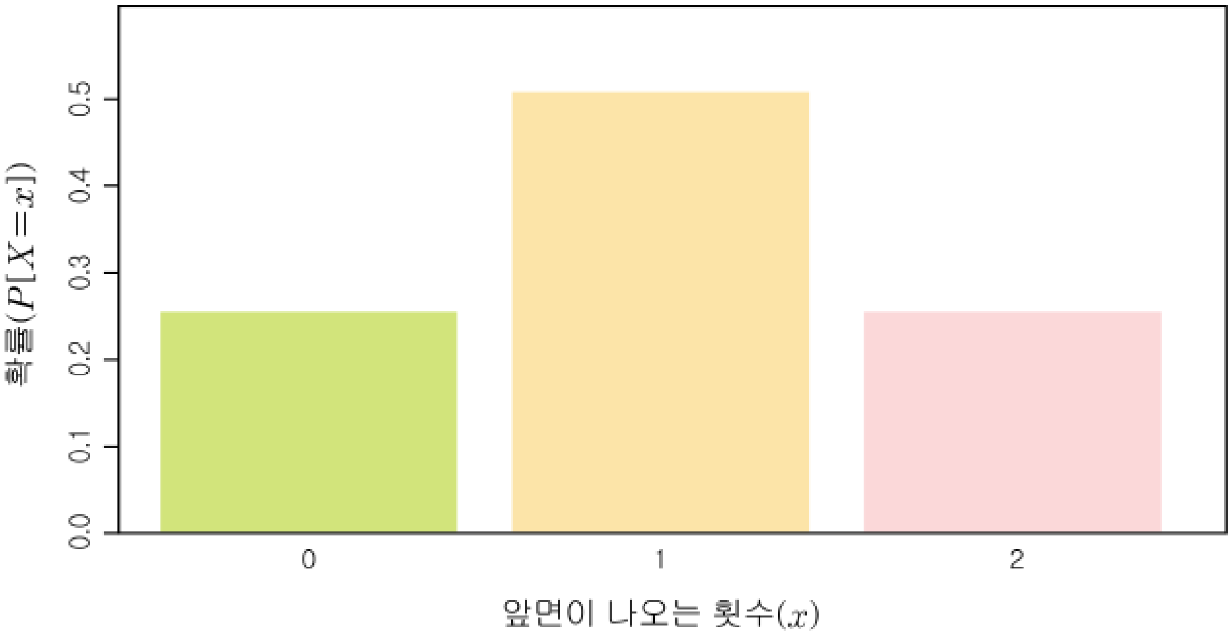
확률변수

- 확률변수
  - '동전을 두 번 던져 앞면이 나오는 횟수'처럼 표본공간의 각 원소를 실숫값에 대응시키는 함수를 확률변수라 합니다.
    - 확률변수는 알파벳 대문자 X, Y, Z, …
    - 확률변수가 취하는 실숫값은 알파벳 소문자 x, y, z ⋯
    - '확률변수 X가 값 실숫값 x를 가질 때'는 X=x로 표기합니다.
  - 확률변수 X가 가질 수 있는 모든 x_i들에 확률이 대응되고, 확률변수는 이 확률에 따라 실숫값을 갖습니다.
  - 확률변수 X가 특정 값 x를 가지는 사상 X=x의 확률을 P(X=x)로 표기합니다.
  - 확률분포
    - 확률변수가 취할 수 있는 값과 각 값이 나타날 확률을 대응시킨 관계(함수) (표 3-4)
  - 확률변수가 가질 수 있는 값에 따라 이산형 확률변수와 연속형 확률변수가 있습니다.

확률변수

[표 3-4] 확률변수의 확률

표본공간	표본공간에서의 확률	$X=x$	$P(X=x)$
{H, H}	1/4	0	1/4
{H, T}	1/4	1	1/2
{T, H}	1/4		
{T, T}	1/4	2	1/4



확률변수

- **확률변수의 평균과 분산**
  - 확률변수 X가 '동전을 두번 던져 앞면이 나오는 횟수'일 때의 평균과 분산을 구해 봅시다.
  - 확률변수의 평균
    - 확률변수 X가 가질 수 있는 값은 {0, 1, 2}
    - 상수 0, 1, 2에 대한 평균
      - $\bar{X} = \sum_{i=1}^{n} \frac{1}{n} \cdot x_i = \frac{1}{3}(0+1+2) = 1$
      - 상수의 평균에서 $\frac{1}{n}$ 은 각 자료들이 모두 동일하게 $\frac{1}{n}$의 비중을 갖고 있음을 나타냅니다.
      - 확률변수에서는 $\frac{1}{n}$에 해당하는 비중이 각 값이 나타날 확률로 바뀝니다.
    - 확률변수 X의 평균 : $E(X) = \sum_{\text{모든 } x} x \cdot P(X = x)$ (식 3.15)

확률변수

- $E(X)$ 는 확률변수 X의 평균을 나타내는 기호로, 확률변수의 평균을 **기댓값이라** 합니다.
- 확률변수 X의 평균 즉, 기댓값을 구해봅시다. (식 3.15 이용)
  - $E(X) = \sum_{i=1}^{3} x_i \cdot P(X = x_i) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = \frac{(0\cdot1+1\cdot2+2\cdot1)}{4} = \frac{4}{4} = 1$

확률변수의 분산

- 2장에서 살펴보았듯이, 분산은 편차 제곱의 평균입니다.
- 확률변수의 분산에 그대로 적용해 봅시다.
  - $Var(X) = E[(X - E(X))^2]$ (식 3.17)
    - 확률변수의 분산은 $Var(X)$ 혹은 $V(X)$로 나타냅니다.
  - 기댓값을 구할 때와 마찬가지로 상수 자료들의 분산을 구할때 사용한 $1/_n$이 확률로 바뀐다는 점 외에 나머지는 동일합니다.
    - $E\left[(X - E(X))^2\right] = \sum_{모든\ x}(x - E(X))^2 \cdot P(X = x)$ (식 3.18)

확률변수

- 분산의 간편식
  - $Var(X) = E(X^2) - [E(X)]^2 = \sum_{모든\ x} x^2 \cdot P(X = x) - [E(X)]^2$ (식 3.19)
  - '확률변수의 제곱의 기댓값($E(X^2)$)'을 구한 후 '기댓값의 제곱($[E(X)]^2$)'을 뺍니다.
- 확률변수 X의 분산을 구해봅시다.
  - 분산의 간편식을 이용합니다.
  - 앞서 구한 기댓값($E(X)$)은 1입니다.
  - 확률변수의 제곱의 기댓값을 구해봅시다.
    - $E(X^2) = \sum_{i=1}^{3} {x_i}^2 \cdot P(X = x_i) = 0^2 \frac{1}{4} + 1^2 \frac{2}{4} + 2^2 \frac{1}{4} = \frac{(0^2 \cdot 1 + 1^2 \cdot 2 + 2^2 \cdot 1)}{4} = \frac{6}{4}$
  - $Var(X) = E(X^2) - [E(X)]^2 = \frac{6}{4} - 1^2 = \frac{2}{4} = \frac{1}{2}$

- 기댓값과 분산은 확률변수의 특성을 파악하는 좋은 모수입니다.
- R을 이용해 확률변수의 기댓값과 분산을 구해 봅시다.

확률변수

예제 3-2 확률변수의 평균(기댓값)과 분산

준비파일 | 02.rv_e_var.R

- **내용**
  - R로 확률변수의 평균과 분산을 구해 봅시다.
  - 이를 통해 R에서 벡터 계산에 대해 알아봅시다.
  - 확률변수 X는 앞서 사용한 동전을 '두 번 던져 앞면이 나오는 횟수'로 X가 가지는 값은 {0, 1, 2}이고 각각 등장할 확률은 {1/4, 1/2, 1/4} 입니다.
- **Step #1) 확률변수의 기댓값**

```
① x <- c(0, 1, 2)
② px <- c(1/4, 2/4, 1/4)
③ EX <- sum( x * px )
④ EX
```

확률변수

- 코드 설명
  - 1줄 : 확률변수가 취할 수 있는 값을 벡터 x에 저장합니다.
  - 2줄 : 각 확률변수가 나타날 확률을 벡터 px에 저장합니다.
  - 3줄 : 식 (3.15)를 이용하여 이산형 확률변수의 기댓값을 구하고, 그 값을 EX에 저장합니다.
    - (확률변수가 취할 수 있는 값) * (확률변수가 나타날 확률) : x * px

```
> x * px
[1] 0.0 0.5 0.5
```

    - 위의 값을 모두 더함 : sum(x * px)
  - 4줄 : 3줄에서 계산한 기댓값이 저장된 변수 EX의 값을 출력합니다(기댓값은 1).

```
> EX
[1] 1
```

확률변수

- **Step #2) 벡터간의 연산**

```
⑤ x * 2
⑥ x * (1:6)
⑦ x * (1:4)
```

- 5줄 : 벡터와 스칼라(R에서는 원소가 하나인 벡터입니다) 연산
  - * 연산자를 이용한 곱셈(다른 연산자도 동일)에서 스칼라 값이 벡터의 모든 원소와 곱한 벡터를 결과로 반환합니다.
    - 스칼라인 원소가 벡터의 원소 수만큼 반복하면서 계산합니다.
    - 코드에서는 벡터 x, 스칼라 2와의 '곱(*)'이므로 벡터 x의 모든 원소에 2를 '곱'한 벡터를 반환합니다.

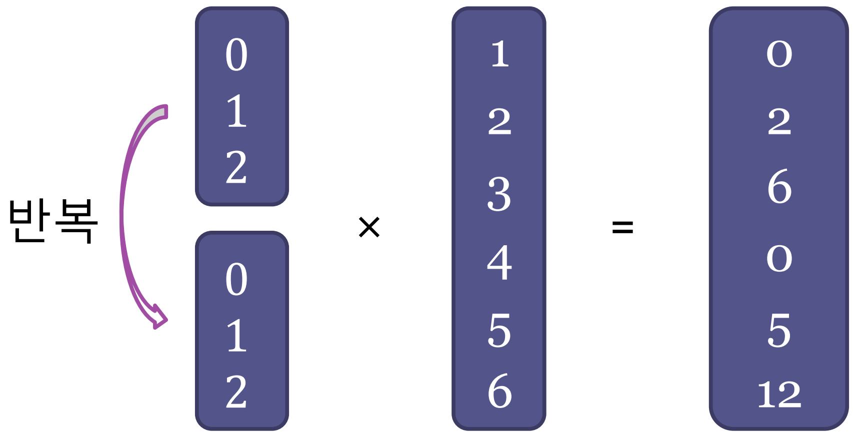
```
> x * 2
[1] 0 2 4
```

$$\begin{bmatrix}0\\1\\2\end{bmatrix} \times 2 = \begin{bmatrix}0\\2\\4\end{bmatrix}$$

확률변수

- 6줄 : 두 벡터간의 연산
  - 두 벡터의 크기가 다를(원소의 개수가 다른) 경우에는 크기가 작은 벡터가 반복하면서 계산을 합니다.
    - 이는 바로 위에서 원소가 1개인 스칼라가 벡터의 원소만큼 반복한 것과 같습니다.
    - 코드에서는 원소가 작은 벡터 x의 원소 수는 3, 큰 벡터(1:6)의 원소 수는 6으로, 6이 3의 두 배이므로 연산이 가능합니다

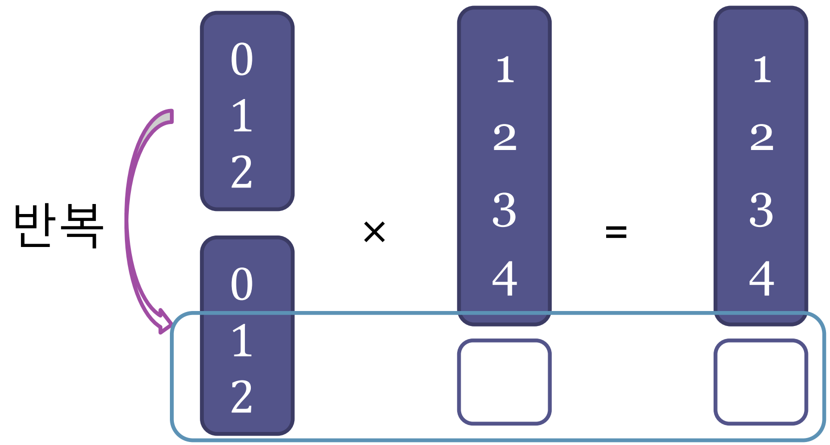
```
> x * (1:6)
[1]  0  2  6  0  5 12
```



확률변수

- 7줄 : 원소의 개수가 정수 배수가 아닌 경우
  - 원소 수가 다른 경우 두 벡터의 원소 수의 관계가 정수 배수여야 작은 원소가 반복되면서 계산을 하는데, 그렇지 않을 경우 R은 어떻게든 계산을 한 후에 경고 메시지를 보냅니다.
  - 경고메시지를 무시할 수 있지만, 이런 경우에는 계산이 올바로 되지 않는 문제가 있음을 생각합니다.

```
> x * (1:4)
[1]  0  2  6  0
Warning message: In x * (1:4) :
longer object length is not a
multiple of shorter object length
```



확률변수

- 7줄 : 원소의 개수가 정수 배수가 아닌 경우
  - 원소 수가 다른 경우 두 벡터의 원소 수의 관계가 정수 배수여야 작은 원소가 반복되면서 계산을 하는데, 그렇지 않을 경우 R은 어떻게든 계산을 한 후에 경고 메시지를 보냅니다.

확률변수

- **Step #3) 확률변수의 분산**

```
⑧ VX <- sum(x^2 * px) - EX^2
⑨ VX
```

- 8줄 : 분산의 간편식으로 구합니다.
  - 이산형 확률변수의 분산을 구하고, 그 값을 VX에 저장합니다.
    - (확률변수가 취하는 값의 제곱)확*률(변수가 나타날 확률)
    - 위의 값을 모두 더함 : sum(x^2 * px)
    - 앞서 구한 기댓값의 제곱을 빼줌
- 9줄 : 8줄에서 계산한 분산이 저장된 변수 VX의 값을 출력합니다.

```
> VX
[1] 0.5
```

02. 확률분포

: 잘 알려진 확률변수의 분포

1. 다양한 분포함수 중 가장 기초가 되는 분포함수에 대해 학습한다.
2. R 을 이용하여 분포함수를 이용한 확률 계산을 실습한다.

개요

- **확률분포**
  - 확률변수가 취할 수 있는 값과 발생할 확률을 대응한 관계
  - 확률변수가 X 가질 수 있는 임의의 실측값 x에 대해

$$F(x) = P(X \le x)$$

와 같이 정의된 함수 F를 확률변수 X의 누적분포함수, 또는 간략히 분포함수라고 합니다

  - 분포의 특성인 모수에 따라 분포의 모양이 결정됩니다.
  - 확률질량함수, 확률밀도함수
    - 확률변수 X가 실측값 x를 가질 확률($P(X = x)$)에 대한 함수를 $f(x)$로 나타냅니다.

$$f(x) = P(X = x)$$

    - 확률변수가 취하는 값이 이산형일 경우에는 확률질량함수, 연속형일 경우에는 확률밀도함수라 부릅니다.

베르누이 시행

- **베르누이 시행**
  - p의 확률로 원하는 결과가 나타났을 때 '성공'으로, 1-p의 확률로 그렇지 않은 결과가 나타났을 때 '실패'로 하는 두 가지 결과가 나타나는 확률실험입니다.
  - 성공 확률 p가 베르누이 시행의 모수입니다.
  - 확률변수 X가 베르누이 시행에 따라 성공일 때 1, 실패일 때 0을 가질 경우 확률질량함수는 다음과 같습니다.

$$f(x) = p^x \cdot (1-p)^{1-x}, \qquad x = \begin{cases} \text{성공} & 1 \\ \text{실패} & 0 \end{cases}$$

  - 예) 주사위를 던져 3의 배수의 눈이 나오면 상금얻는 게임
    - 성공 : 3의 눈, 6의 눈이 나오는 경우, X=1

$$P(X=1) = p^{x=1} \cdot (1-p)^{1-(x=1)} = p$$

    - 실패 : 성공의 경우가 아닌 눈이 나오는 경우, X=0

$$P(X=0) = p^{x=0} \cdot (1-p)^{1-(x=0)} = 1-p$$

베르누이 시행

▫ 기댓값과 분산

- 기댓값 : p

$$E(X) = \sum_{\text{모든 } x} x \cdot P(X=x)$$

$$= \sum_{\text{모든 } x} x \cdot f(x) = 0 \cdot (p^0 \cdot (1-p)^1) + 1 \cdot (p^1 \cdot (1-p)^0) = p$$

- 분산 : $p \cdot (1-p)$

$$Var(X) = E(X^2) - (EX)^2$$

$$= \sum_{\text{모든 } x} \{x^2 \cdot f(x)\} - p^2$$

$$= \sum_{\text{모든 } x} \{(0^2 \cdot (p^0 \cdot (1-p)^1) + 1^2 \cdot (p^1 \cdot (1-p)^0)\} - p^2$$

$$= p - p^2 = p(1-p)$$

이항분포

- **개요**
  - 성공 확률이 p로 동일한 베르누이 시행을 n번 반복해서 실험하는 경우
    - 실험이 n번 반복되더라도 성공 확률 p는 변하지 않고 동일
    - 각 실험이 서로 독립적으로 시행 (iid: independent and identically distribution)
  - n번 반복 실험에서 성공의 횟수가 따르는 분포를 이항분포라고 합니다.
  - 이항분포의 모수
    - n : 시행의 횟수
    - p : 성공의 확률
  - 이항분포의 표기 : 위의 두 모수를 이용하여 $B(n,p)$
    - 확률변수 X가 이항분포를 따를 때 $X \sim B(n,p)$와 같이 나타냅니다.

이항분포

- **이항분포의 확률질량함수**
  - 주사위를 굴려 3의 배수가 나올 때를 성공으로 하는 실험(성공의 확률이 1/3인 베르누이 시행)을 3번 독립으로 반복해서 실험할 때의 성공 횟수의 확률질량함수를 구해 봅시다.
    - 확률변수 X를 성공의 확률이 1/3인 베르누이 시행을 3번 독립으로 반복해서 실험하였을 때 성공의 횟수라 할 때 확률변수 X는 다음과 같은 이항분포를 따릅니다.

$$X \sim B\left(n = 3, p = \frac{1}{3}\right)$$

    - 확률변수 X가 가질 수 있는 값은 {0, 1, 2, 3}이고 각각의 확률들을 구해봅시다.
      - x_i는 i번째 주사위를 굴렸을 때 성공과 실패를 각각 1과 0을 가짐을 나타냅니다.

이항분포

- X=0 일 때 : 성공의 횟수가 0일 때
  - $x_1 = 0, x_2 = 0, x_3 = 0$일 때의 베르누이 시행의 확률질량함수($f_{Ber}(x_i = 0)$)를 구합니다. (성공의 확률 p=1/3)
    - $f_{Ber}(x_1 = 0) = \frac{1}{3}^{x_1=0} \frac{2}{3}^{1-(x_1=0)}$
    - $f_{Ber}(x_2 = 0) = \frac{1}{3}^{x_2=0} \frac{2}{3}^{1-(x_2=0)}$
    - $f_{Ber}(x_3 = 0) = \frac{1}{3}^{x_3=0} \frac{2}{3}^{1-(x_3=0)}$
  - 모두 실패한 경우는 $x_1 = 0, x_2 = 0, x_3 = 0$의 한 가지 경우만 있습니다.
  - 각 실험이 독립이므로 각 확률들의 곱으로 나타낼 수 있습니다.
    - $f_{Ber}(x_1 = 0) \cdot f_{Ber}(x_2 = 0) \cdot f_{Ber}(x_3 = 0)$
    - $P(X = 0) = \frac{1}{3}^{x_1=0} \frac{2}{3}^{1-(x_1=0)} \cdot \frac{1}{3}^{x_2=0} \frac{2}{3}^{1-(x_2=0)} \cdot \frac{1}{3}^{x_3=0} \frac{2}{3}^{1-(x_3=0)} = \frac{1}{3}^{0} \frac{2}{3}^{3}$

이항분포

- X=1 일 때 : 성공의 횟수가 1일 때
  - x_1, x_2, x_3 중 한 경우는 1을 갖고(성공) 나머지 두 경우는 0을 가지는(실패) 경우로 다음과 같이 세 가지 상태를 갖습니다.
  - 첫 번째 성공했을 때 : $x_1 = 1, x_2 = 0, x_3 = 0$
    - $\frac{1}{3}^{x_1=1}\frac{2}{3}^{1-(x_1=1)} \cdot \frac{1}{3}^{x_2=0}\frac{2}{3}^{1-(x_2=0)} \cdot \frac{1}{3}^{x_3=0}\frac{2}{3}^{1-(x_3=0)} = \frac{1}{3}^1\frac{2}{3}^2$
  - 두 번째 성공했을 때 : $x_1 = 0, x_2 = 1, x_3 = 0$
    - $\frac{1}{3}^{x_1=0}\frac{2}{3}^{1-(x_1=0)} \cdot \frac{1}{3}^{x_2=1}\frac{2}{3}^{1-(x_2=1)} \cdot \frac{1}{3}^{x_3=0}\frac{2}{3}^{1-(x_3=0)} = \frac{1}{3}^1\frac{2}{3}^2$
  - 세 번째 성공했을 때 : $x_1 = 0, x_2 = 0, x_3 = 1$
    - $\frac{1}{3}^{x_1=0}\frac{2}{3}^{1-(x_1=0)} \cdot \frac{1}{3}^{x_2=0}\frac{2}{3}^{1-(x_2=0)} \cdot \frac{1}{3}^{x_3=1}\frac{2}{3}^{1-(x_3=1)} = \frac{1}{3}^1\frac{2}{3}^2$
  - 세가지 경우의 $\frac{1}{3}^1\frac{2}{3}^2$ 이 있는 것으로 $P(X = 1) = 3 \cdot \frac{1}{3}^1\frac{2}{3}^2$ 입니다.

이항분포

- X=2 일 때 : 성공의 횟수가 2일 때
  - x_1, x_2, x_3 중 두 경우는 1을 갖고(성공) 나머지 한 경우는 0을 가지는(실패) 경우로 다음과 같이 세 가지 상태를 갖습니다.
  - 첫 번째와 두 번째가 성공했을 때 : $x_1 = 1, x_2 = 1, x_3 = 0$
    - $\frac{1}{3}^{x_1=1}\frac{2}{3}^{1-(x_1=1)} \cdot \frac{1}{3}^{x_2=1}\frac{2}{3}^{1-(x_2=1)} \cdot \frac{1}{3}^{x_3=0}\frac{2}{3}^{1-(x_3=0)} = \frac{1}{3}^{2}\frac{2}{3}^{1}$
  - 두 번째와 세 번째가 성공했을 때 : $x_1 = 0, x_2 = 1, x_3 = 1$
    - $\frac{1}{3}^{x_1=0}\frac{2}{3}^{1-(x_1=0)} \cdot \frac{1}{3}^{x_2=1}\frac{2}{3}^{1-(x_2=1)} \cdot \frac{1}{3}^{x_3=1}\frac{2}{3}^{1-(x_3=1)} = \frac{1}{3}^{2}\frac{2}{3}^{1}$
  - 첫 번째와 세 번째가 성공했을 때 : $x_1 = 1, x_2 = 0, x_3 = 1$
    - $\frac{1}{3}^{x_1=1}\frac{2}{3}^{1-(x_1=1)} \cdot \frac{1}{3}^{x_2=0}\frac{2}{3}^{1-(x_2=0)} \cdot \frac{1}{3}^{x_3=1}\frac{2}{3}^{1-(x_3=1)} = \frac{1}{3}^{2}\frac{2}{3}^{1}$
  - X=1일때와 마찬가지로 세가지 경우가 있어, $P(X = 2) = 3 \cdot \frac{1}{3}^{2}\frac{2}{3}^{1}$ 입니다.

이항분포

- X=3 일 때 : 성공의 횟수가 3일 때
  - $x_1 = 1, x_2 = 1,\ x_3 = 1$ 인, 즉 모두 성공한 경우로 다음과 같은 한가지 경우입니다.
  - 즉, $\frac{1}{3}^{x_1=1}\frac{2}{3}^{1-(x_1=1)} \cdot \frac{1}{3}^{x_2=1}\frac{2}{3}^{1-(x_2=1)} \cdot \frac{1}{3}^{x_3=1}\frac{2}{3}^{1-(x_3=1)} = \frac{1}{3}^{3}\frac{2}{3}^{0}$
- 모두 모아 봅시다.

$$P(X=x) = f(x) = \begin{cases} X=0, & \boxed{\begin{matrix}1\\3\\3\\1\end{matrix}} \cdot \boxed{\begin{matrix}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^3\\ \left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2\\ \left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1\\ \left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^0\end{matrix}} \begin{matrix}= \frac{1\cdot 8}{27}\\ = \frac{3\cdot 4}{27}\\ = \frac{3\cdot 2}{27}\\ = \frac{1\cdot 1}{27}\end{matrix} \\ X=1, & \\ X=2, & \\ X=3, & \end{cases}$$

이항계수 = $\binom{n}{x} = \frac{n!}{(n-x)!x!}$ 　　　확률 $p^x(1-p)^{n-x}$

이항분포

- 이상으로 부터 확률변수 X가 이항분포를 따를 때의 확률질량함수는 다음과 같습니다.
  - $P(X=x)=f(x)=\binom{n}{x}p^x(1-p)^{n-x}, \qquad x=0,1,2,\ldots,n$

이항분포의 기댓값과 분산

- 앞서 사용한 주사위를 굴려 3의 배수가 나오는 경우의 성공 횟수에 대한 분포 ($B(n=3, p=\frac{1}{3})$)를 이용하여 기댓값의 정의에 따라 확률변수 X가 이항분포를 따를 때의 기댓값을 구해 봅시다.

$$E(X)=\sum_{\text{모든 } x} x\cdot P(X=x)=\sum_{x=0}^{n} x\cdot\binom{n}{x}p^x(1-p)^{n-x}$$

이항분포

- $X \sim B\left(3, \frac{1}{3}\right)$ 이고, $E(X) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$ 이므로,

$$E(X) = \sum_{x=0}^{n} x \cdot \binom{n}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{n-x}$$

$$= \left(0 \times 1 \cdot \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^3\right) + \left(1 \times 3 \cdot \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^2\right) + \left(2 \times 3 \cdot \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^1\right) + \left(3 \times 1 \cdot \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^0\right)$$

$$= 0 \times \frac{8}{27} + 1 \times \frac{12}{27} + 2 \times \frac{6}{27} + 3 \times \frac{1}{27} = \frac{0+12++12+3}{27} = \frac{27}{27} = 1$$

  - 여기서 구한 값인 1은 이항분포의 두 모수 n=3과 p=1/3 의 곱 np와 같습니다.

- 이항분포의 기댓값($E(X)$) : np

이항분포

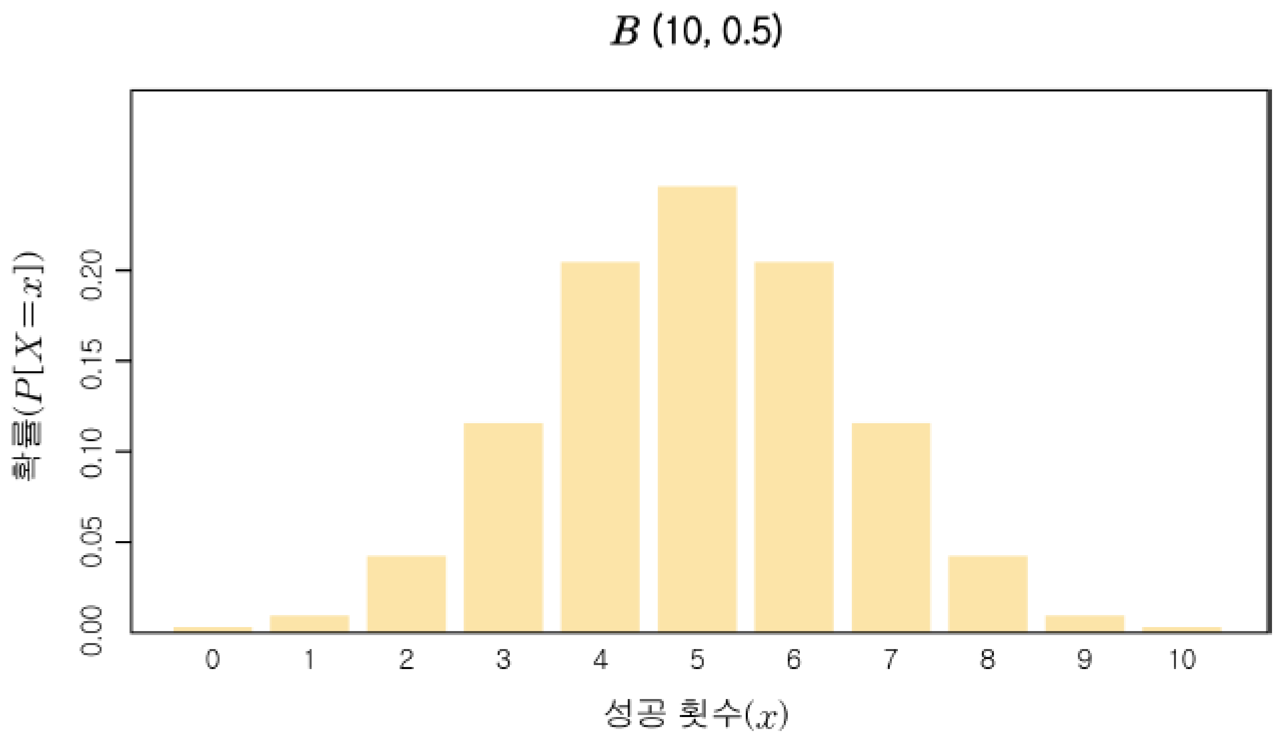
- 이항분포의 분산
  - 확률분수의 분산을 구하는 간편식을 이용하여 (교재에 잘못된 표현) 구할 수 있으며 그 값은 $np(1-p)$로 알려져 있습니다.
  - 다음 식의 전개 과정은 QR 코드(위키백과)를 통해 확인할 수 있습니다.

$$Var(X) = E(X^2) - \big(E(X)\big)^2 = \sum_{\text{모든 } x} x^2 \cdot P(X = x) - \big(E(X)\big)^2$$

$$= \sum_{x=0}^{n} x^2 \cdot \binom{n}{x} p^x (1-p)^{n-x} - (np)^2 \qquad \text{(식 3.26)}$$

이항분포

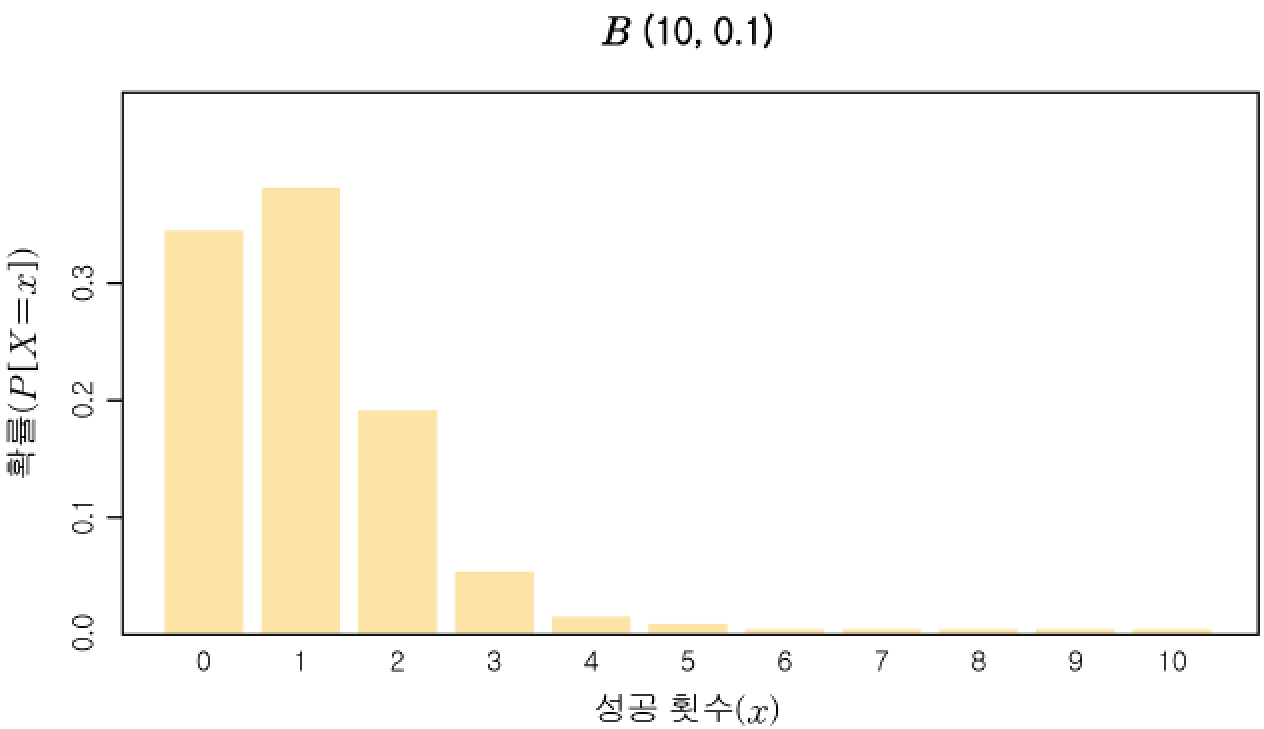
- **성공 확률 p의 변화에 따른 이항분포의 모양 변화**
  - 어떤 확률변수 X가 시행 횟수가 10, 성공 확률이 0.5인 이항분포를 따른다고 할 때 $(X \sim B(10, 0.5))$, 확률변수 X의 분포도는 다음과 같습니다.



- 기댓값인 5를 중심으로 좌우대칭을 하고 있습니다.

이항분포

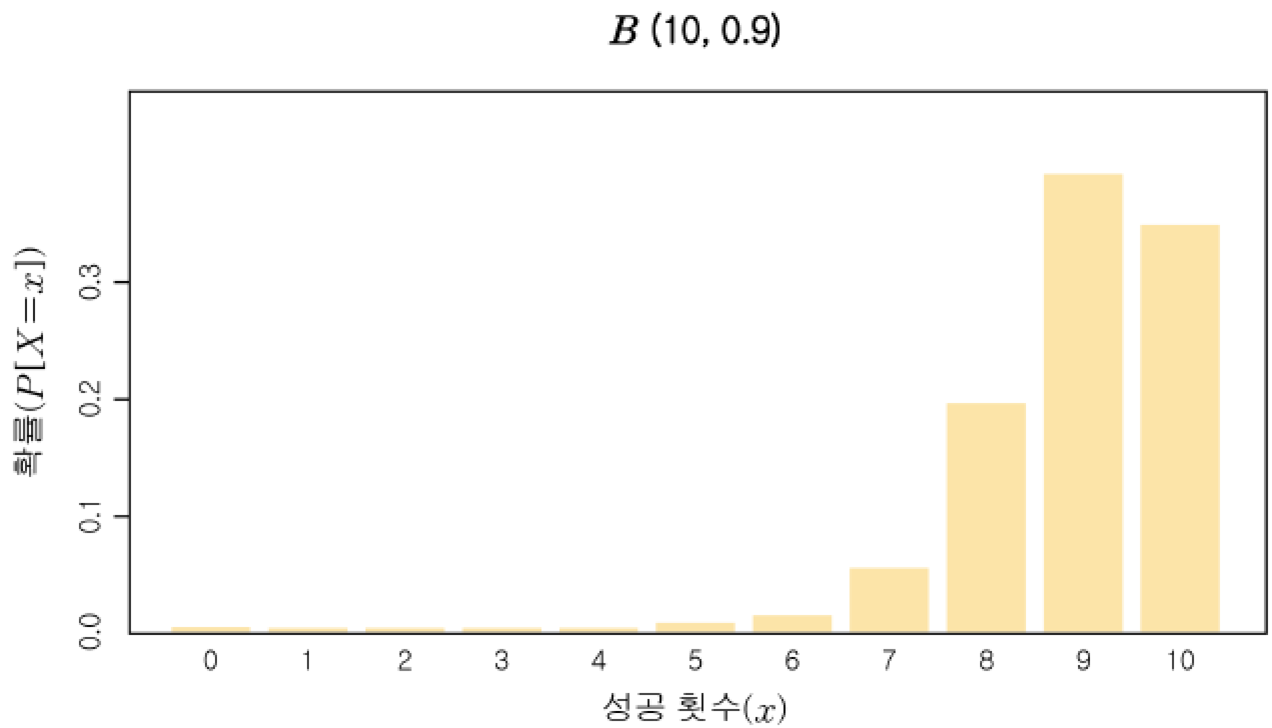
- 성공의 확률이 0.1인 경우에는 다음 그림과 같습니다.



- 기댓값인 1이 최빈값이며 꼬리가 오른쪽으로 길어집니다.
- 성공의 확률이 0.5보다 작은 경우 이러한 모양을 갖습니다.

이항분포

- 성공의 확률이 0.9인 경우에는 다음 그림과 같습니다.



- 기댓값인 9가 최빈값이며 꼬리가 왼쪽으로 길어집니다.
- 성공의 확률이 0.5보다 큰 경우 이러한 모양을 갖습니다.

이항분포

예제 3-3 R을 이용한 이항분포 계산 준비파일 | 03.binomial_dist.R

- **개요**
  - 확률변수 X가 시행의 횟수가 6이고 성공의 확률이 1/3인 이항분포를 따를 때 R이 내장하고 있는 이항분포와 관련된 함수를 이용하여 각종 확률을 계산해 봅시다.
- **Step #1) 이항분포 함수 사용을 위한 모수 준비**

```
① n <- 6
② p <- 1/3
③ x <- 0:n
```

  - 1줄 : 시행 횟수는 6이고 이를 변수 n에 저장합니다.
  - 2줄 : 성공 확률은 1/3이고 이를 변수 p에 저장합니다.
  - 3줄 : 확률변수 X가 가질 수 있는 값은 {0, 1, 2, 3, 4, 5, 6} 으로 이를 벡터로 변수 x에 저장합니다.

이항분포

- **Step #2) 확률질량함수($P(X=x)$) 를 구합니다.**
  - 이항분포의 확률질량함수를 구하는 R 함수인 dbinom()을 사용합니다.

```
dbinom(x, size, prob)
- x : 이항분포의 성공의 횟수의 벡터(원소 1개짜리 포함)
- size : 시행의 횟수
- prob : 성공의 확률
```

```
⑤ ( dbinom(2, size=n, prob=p) )
⑥ ( dbinom(4, size=n, prob=p) )
⑦ ( px <- dbinom(x, size=n, prob=p) )
⑧ plot(x, px, type="s", xlab="성공 횟수(x)", ylab="확률
   (P[X=x])", main="B(6, 1/3)")
```

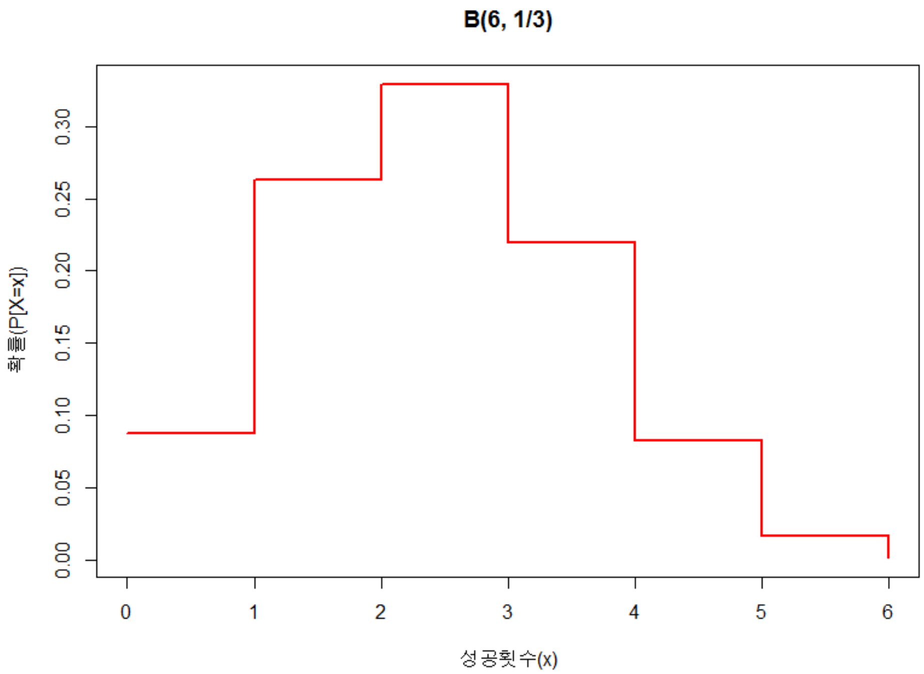
이항분포

- 5, 6줄 : 확률변수 X가 2와 4를 가질 확률을 각각 계산하고 출력합니다. R에서 변수 할당문의 앞과 뒤를 괄호( )로 둘러싸면 바로 결과를 출력해줍니다.
- 7줄 : x를 통해 전달된 성공의 횟수별 확률을 계산한 벡터를 반환합니다.
  - 코드에서는 x가 확률변수 X가 가질 수 있는 값인 0부터 6을 저장한 벡터로, 각각의 확률을 계산한 결과를 변수 px에 저장하고 출력합니다.

```
> (dbinom(2, size=n, prob=p))
[1] 0.3292181
> (dbinom(4, size=n, prob=p))
[1] 0.08230453
> (px <- dbinom(x, size=n, prob=p))
[1] 0.087791495 0.263374486 0.329218107 0.219478738 0.082304527
[6] 0.016460905 0.001371742
```

이항분포

- 8줄 : 위에서 계산한 벡터 x를 이용하여 이항분포 그래프를 작성합니다. plot() 함수의 전달인자로 type에 “s”(소문자)를 사용할 경우, 시작 값을 수평으로 먼저 그리는 계단 형태의 그래프를 작성합니다.
  - type으로 “s”를 가질 때의 plot() 함수의 모양을 설명하기 위한 예제로 이산형 확률분포에는 각 값들이 떨어져 있는 그래프가 적절합니다. (type에 “h”를 넣어보세요)



이항분포

- **Step #3) 누적분포함수($P(X \le x)$) 를 구합니다.**
  - 이항분포의 누적분포함수를 구하는 R 함수인 pbinom()을 사용합니다.

```
pbinom(x, size, prob)
- x : 이항분포의 성공의 횟수의 벡터(원소 1개짜리 포함)
- size : 시행의 횟수
- prob : 성공의 확률
```

```
10.(pbinom(2, size=n, prob=p))
11.(pbinom(4, size=n, prob=p))
12.(pbinom(4, size=n, prob=p) - pbinom(2, size=n,
   prob=p))
```

이항분포

- 10줄 : 성공 횟수가 2 이하일 확률을 구하고 출력합니다.
- 11줄 : 성공 횟수가 4 이하일 확률을 구하고 출력합니다.
- 12줄 : (성공 횟수가 4 이하일 확률) - (성공 횟수가 2 이하일 확률)
  - 성공 횟수가 4 이하 2 초과(3 이상)일 확률($P(2 < X \leq 4)$)을 구합니다.

```
> (pbinom(2, size=n, prob=p))
[1] 0.6803841
> (pbinom(4, size=n, prob=p))
[1] 0.9821674
> (pbinom(4, size=n, prob=p) - pbinom(2, size=n, prob=p))
[1] 0.3017833
```

이항분포

- **Step #4) 분위 p에 해당하는 확률변수 X의 값 x를($P(X \leq x) = p$) 를 구합니다.**
  - 이항분포의 분위 p에 해당하는 값을 찾는 R 함수인 qbinom()을 사용합니다.

```
qbinom(p, size, prob)
- p : 알고자 하는 분위 벡터(원소 1개짜리 포함)
- size : 시행의 횟수
- prob : 성공의 확률
```

```
14. (qbinom(0.1, size=n, prob=p))
15. (qbinom(0.5, size=n, prob=p))
```

이항분포

- 14, 15줄 : 확률변수 X가 $B(6, \frac{1}{3})$을 따를때, 확률변수 X의 확률분포에서 10%(0.1)와 50%(중앙값, 0.5)에 해당하는 x의 값을 출력합니다.
  - 이산형 분포함수에서는 x가 서로 떨어져 있으므로 누적확률값이 전달하는 분위를 포함하는 가장 작은 값이 출력됩니다.
  - p 가 0.1인 경우 X가 0이하일 확률이 0.09, 1 이하일 확률이 0.35, 2이하일 확률이 0.68로, 이 중 0은 해당하지 않고 0.1 이상의 확률값을 갖는 x 중({1, 2, 3, ...}) 가장 작은 값인 1이 이에 해당합니다.

이항분포

- **Step #4) 이항분포를 따르는 모집단으로부터 n개의 표본 추출**
  - 이항분포로부터 난수를 생성하는 R 함수인 rbinom()을 사용합니다.

```
rbinom(n, size, prob)
- n : 생성하고자 하는 난수의 개수
- size : 시행의 횟수
- prob : 성공의 확률
```

```
17. rbinom(10, size=n, prob=p)
```

  - 17줄 : $B\left(6, \frac{1}{3}\right)$ 인 모집단으로부터 10개의 확률표본을 추출합니다
    - 난수는 실행할 때마다 다르게 나타납니다.

```
> (rbinom(10, size=n, prob=p))
[1] 4 1 2 2 3 1 1 3 2 3
```

이항분포

예제 3-4 R의 분포함수를 이용한 기댓값과 분산

준비파일 | 04.ex_varx.R

- **개요**
  - 기댓값과 분산을 구하는 식을 이용하여 확률변수 X가 이항분포 $B(6, \frac{1}{3})$를 따를 때의 기댓값과 분산을 R에서 구해봅시다.
- **Step #1) 기댓값과 분산을 위한 정보 생성**

```
① n <- 6
② p <- 1/3
③ x <- 0:n
④ px <- dbinom(x, size=n, prob=p)
```

  - 1~3줄 : 앞서 사용한 것과 같이 시행의 횟수(n), 성공의 확률(p), 확률변수가 가질 수 있는 모든 값(x)를 설정합니다.
  - 4줄 : 확률변수 X가 가질 수 있는 값인 0부터 6에 대해 dbinom() 함수를 이용해 각각의 확률을 계산하고, 변수 px에 저장합니다.

이항분포

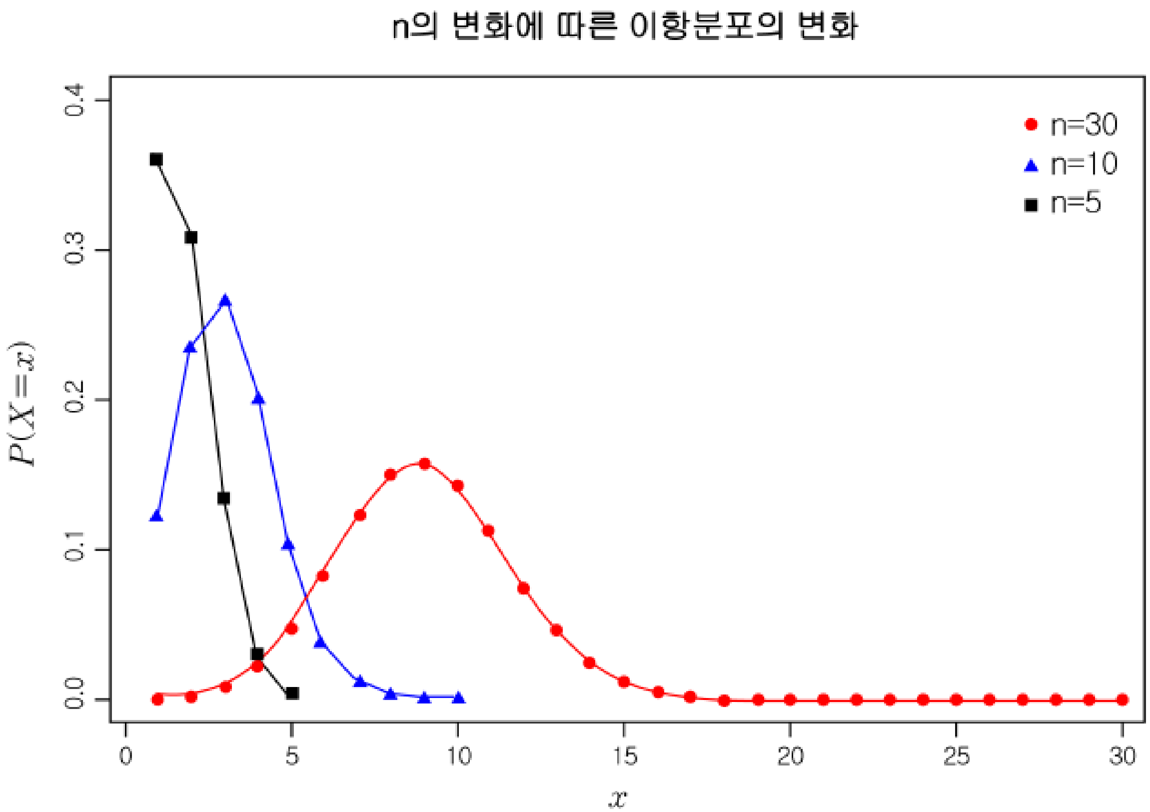
- **Step #2) 기댓값과 분산의 계산식을 이용하여 R로 구합니다.**
  - $E(X) = \sum_{모든\ x} x \cdot \boxed{P(X = x)}$ → dbinom()
  - $Var(X) = E(X^2) - \left(E(X)\right)^2 = \sum_{모든\ x} x^2 \cdot \boxed{P(X = x)} - \left(E(X)\right)^2$

```
⑤ (ex <- sum( x * px ))
⑥ ex2 <- sum( x^2 * px )
⑦ (varx <- ex2 - ex^2)
```

  - 5줄 : 확률변수의 기댓값의 정의를 이용하여 기댓값을 구합니다.
    - x는 확률변수 X가 가질 수 있는 값들이 저장된 벡터
    - px는 확률변수 X가 실측값 x를 가질 때의 이항분포의 확률값이 저장된 벡터
    - 두 벡터를 곱하고 sum() 함수를 이용하여 이를 더합니다.
  - 6, 7줄 : 분산의 간편식을 이용하여 분산을 구합니다.
    - 6줄($E(X^2)$) : x를 제곱하고 이를 px와 곱한후 모두 더한 값을 ex2에 저장합니다.
    - 7줄 : 6줄에서 구한 $E(X^2)$에서 5줄에서 구한 $E(X)$의 제곱값을 빼고 varx에 저장하고 출력합니다.

정규분포

- **개요**
  - p가 0.3일 때 시행횟수 n의 변화에 따른 이항분포의 변화
    - 다음 그림은 p가 0.3으로 고정하고 n=5, n=10, n=30 일 때의 이항분포입니다.



정규분포

- 이항분포에서 시행 횟수 n이 커지면, 그에 따라 이를 따르는 확률변수 X가 갖는 확률($P(X = x)$) 계산은 복잡해집니다.
- 프랑스 태생의 수학자 드무아브르(1667~1754) 가 성공 확률이 0.5이고 시행 횟수 n이 아주 큰 이항분포가 어떤 함수와 비슷해지는 것을 발견하였습니다.
  - 이 함수의 모양은 앞선 그림에서 n이 30인 경우와 많이 닮았습니다.
  - 좌우가 대칭인 종모양(확률분포의 확률값이 x축에 가까이 다가가나 확률이 0이되지 않는)의 형태와 유사합니다.
  - n 이 충분히 크다면 이산형이 아닌 연속형처럼 다루는 것이 가능합니다.
  - 이런 형태를 갖는 분포는, 이항분포가 아닌 다른 분포에서도 이와 닮아감을 밝혔습니다. (라플라스(1749~1827))
  - 관측 오차가 이러한 분포를 따른다는 점이 발견되어 폭넓게 사용되었습니다.(가우스(1777~1855))

정규분포

- **정규분포**

  ① 종모양의 형태를 가집니다.

  - 양 끝이 아주 느린 속도로 감소하지만, 축에 닿지 않고 -∞와 ∞까지 계속됩니다.

  ② 평균을 중심으로 좌우대칭입니다

  ③ 평균 주변에 많이 몰려 있으며 양 끝으로 갈수록 줄어듭니다.

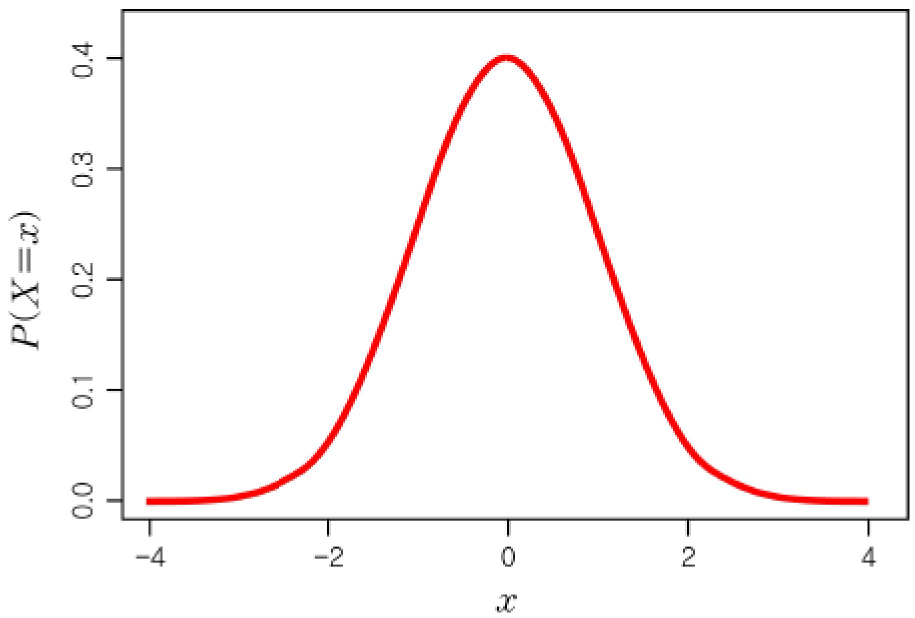
  ④ 평균과 표준편차로 분포의 모양을 결정합니다.

  - 정규분포의 모수는 평균 μ와 표준편차 σ(분산 σ^2)로, $N(\mu, \sigma^2)$으로 나타냅니다.

- 정규분포의 확률밀도함수

  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \ -\infty \le x \le \infty$

평균이 0이고 표준편차가 1인 정규분포



정규분포

- **표준정규분포**
  - 평균이 0이고 표준편차가 1인 정규분포($N(0,1^2)$) 를 **표준정규분포**라하고 대문자 Z로 표시합니다.
  - 모든 정규분포는 표준정규분포로 변환할 수 있습니다.
    - 확률변수 X가 평균 μ와 표준편차 σ인 정규분포를 따른다고 할 때,

$$Z = \frac{X - \mu}{\sigma}, \qquad Z \sim N(0, 1^2)$$

  - 표준정규분포를 사용하면 보다 손쉬운 계산이 가능합니다
    - 예 : 어느 대학교 남학생들 키의 평균은 170cm, 표준편차는 6cm입니다. 이 대학교에서 남학생의 키가 182cm 이상일 확률은 다음과 같이 구합니다.(남학생의 키는 정규분포를 따로 있는 것으로 가정합니다.)

$$P(X \geq 182) = 1 - P(X \leq 182) = 1 - \int_{-\infty}^{182} \frac{1}{6\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-170}{6}\right)^2} dt$$

정규분포

  - 수도 없이 많은 정규분포별로 이런 계산을 해야만 했습니다.
  - 만일, 어느 한 정규분포로 각 확률 값들을 계산해 놓고 다른 정규분포들을 이 분포로의 변환이 가능하며 필요에 따라 다시 또 원래대로 돌아갈 수 있게 한다면 참으로 편리할 것입니다.
- 평균이 0이고 표준편차가 1인 표준정규분포로 각 값을 계산해 표로 만들고, 다음과 같은 과정을 통해 그 값을 구해보았습니다.

① 임의의 정규분포를 표준정규분포로 변환합니다.

② 구하고자 하는 값을 미리 계산된 표준정규분포의 분포표를 통해 구합니다(부록 C).

③ 구한 값을 원래의 정규분포로 변환합니다.

정규분포

- 표준화 변환을 통한 표준정규분포로 계산
  - $z = \frac{x-\mu}{\sigma} = \frac{182-170}{6} = \frac{12}{6} = 2$
  - 이를 이용하여 표준정규분포에서 구하면 다음과 같습니다.
    - $P(Z \geq 2) = 1 - P(Z \leq 2)$
    - [부록 C]의 표준정규분포표에서 z값이 2가 되는 값, 즉 행에서 2.0을 찾고 열에서 0.00을 찾은 값은 0.977(유효숫자 셋째 자리)입니다. 표로부터 표준정규분포에서 2보다 작을 확률은 0.977이고, z가 2보다 클 확률은 1-0.977≈0.023입니다.
- 이제 다시 원래의 정규분포로 돌아가서 z 값으로 변환하여 2가 된 원래의 값을 구해보면 182입니다. 이를 통해 182cm보다 클 확률은 0.023이 됨을 알 수 있습니다.

정규분포

예제 3-5 R을 이용한 정규분포 계산

준비파일 | 07.normal_dist_R.R

- **개요**
  - 표준정규분포를 이용하여 분포표를 읽어 정규분포를 계산할 수 있습니다.
  - 우리는 R을 이용하여 정규분포와 관련된 각종 계산과 그래프를 그려봅시다.
  - 확률변수 X가 평균이 170이고 표준편차가 6인 정규분포($N(170, 6^2)$)를 따를 때로 각종 값들을 계산해 봅시다.
- **Step #1) 각종 모수 및 그래프를 위한 작업**

```
1. options(digits=3)
2. mu <- 170
3. sigma <- 6
4. ll <- mu - 3*sigma
5. ul <- mu + 3*sigma
```

정규분포

- 1줄 : 출력물이 세 자릿수가 되도록 합니다.
- 2줄 : 평균은 170이고, 이를 변수 mu에 저장합니다.
- 3줄 : 표준편차는 6이고, 이를 변수 sigma에 저장합니다.
- 4, 5줄 : 정규분포에서 확률변수가 가질 수 있는 값의 범위가 -∞ ≤ X ≤ ∞로 그래프를 그리기에 너무 넓습니다.
  - 전체 구간보다는 평균 중심으로 세 배의 표준편차 범위로 한정해서 구하려고 합니다.
  - 변수 ll에는 '평균-(3×표준편차)'를, ul에는'평균+(3×표준편차)'를 저장합니다.

- Step #2) dnorm() 함수를 이용하여 $N(170, 6^2)$의 분포도를 작성합니다.

```
7. x <- seq(ll, ul, by=0.01)
8. nd <- dnorm(x, mean=mu, sd=sigma)
9. plot(x, nd, type="l", xlab="x", ylab="P(X=x)", lwd=2,
   col="red")
```

정규분포

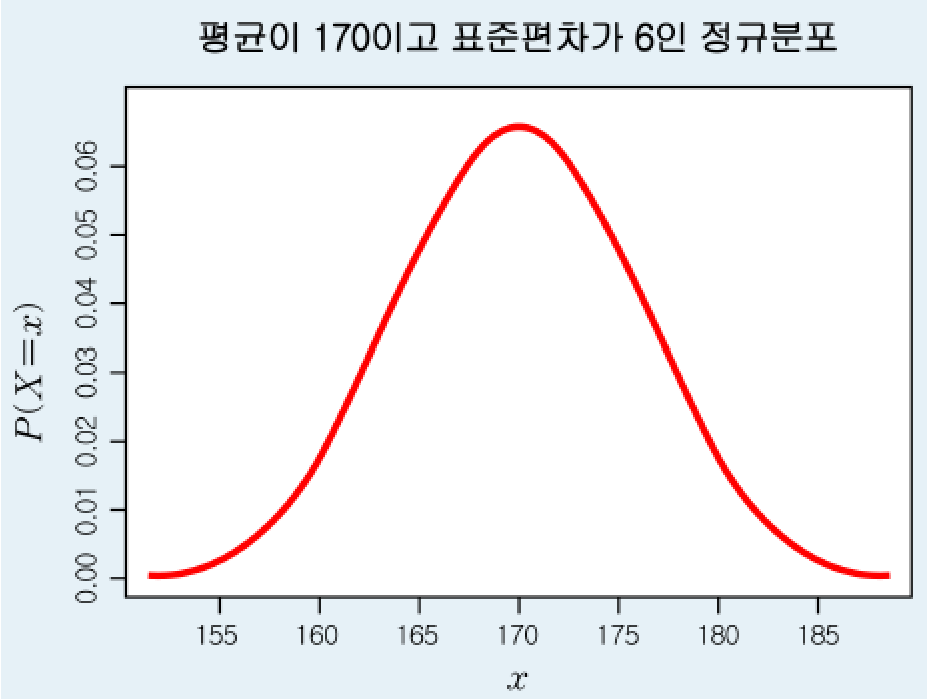
- 7줄 : 확률변수 X가 갖는 값을 ll부터 ul까지 0.01씩 증가하는 값으로 하여 벡터 x에 저장합니다.
- 8줄 : dnorm() 함수는 정규분포의 확률밀도함수를 구하는 함수입니다.

```
dnorm(x, mean, sd)
- x : 정규분포가 가질 수 있는 값의 벡터(원소 1개짜리 포함)
- mean : 정규분포의 평균
- sd : 정규분포의 표준편차
```

- dnorm() 함수의 첫 번째 전달인자로 코드에서와 같이 벡터를 전달할 경우 각 값의 확률밀도함수 값을 갖는 벡터를 전달합니다.
- 여기서 구한 확률밀도함수는 우리가 구하고자 하는 범위 내의 모든 x에 대한 확률 ($P(X = x)$)로, 이를 변수 nd 에 벡터로 저장합니다.

정규분포

- 9줄 : 위에서 구한 x를 x축의 값으로, 각 x의 확률값 nd를 높이로 하여 산점도를 그립니다.
  - 산점도의 형태를 type="l"로 하여 각 값을 선으로 연결합니다.
  - 앞서 구한 x 값은 0.01씩 증가하는 이산형 자료이나, 이 증가분을 작게 하여 연속형 자료처럼 표현합니다



정규분포

- **Step #3) pnorm() 함수를 이용하여 $N(170, 6^2)$ 의 분포함수를 구합니다**

```
pnorm(x, mean, sd)
- x : 정규분포가 가질 수 있는 값의 벡터(원소 1개짜리 포함)
- mean : 정규분포의 평균
- sd : 정규분포의 표준편차
```

```
11.pnorm(mu, mean=mu, sd=sigma)
12.pnorm(158, mean=mu, sd=sigma)
13.pnorm(180, mean=mu, sd=sigma) -
                      pnorm(160, mean=mu, sd=sigma)
```

- 11줄 : 확률변수 X가 $N(170, 6^2)$ 을 따를 때 $P(X \leq 170)$을 구합니다.
  - 변수 mu에 평균값 170이 저장되어 있으며, 정규분포에서 평균 이하일 확률은 0.5입니다(평균을 중심으로 좌우대칭).

정규분포

- 12줄 : 확률변수 X가 $N(170, 6^2)$ 을 따를 때 $P(X \le 158)$ 을 구합니다.
  - $N(170, 6^2)$ 에서 158은 '170-(2×6)'인 값으로, 정규분포에서 이 확률은 약 0.0228입니다.
  - 정규분포는 좌우대칭이므로 '170+(2×6)'인 182이 이상일 확률 $P(X \ge 182)$ 또한 약 0.0228입니다.
- 13줄 : 확률변수 X가 $N(170, 6^2)$ 을 따를 때 $P(X \le 180) - P(X \le 160)$ 즉, $P(160 \le X \le 180)$의 확률을 구합니다. 구한 값은 약 0.904 입니다.

```
> pnorm(mu, mean=mu, sd=sigma)
[1] 0.5
> pnorm(158, mean=mu, sd=sigma)
[1] 0.0228
> pnorm(180, mean=mu, sd=sigma) - pnorm(160, mean=mu, sd=sigma)
[1] 0.904
```

정규분포

- **Step #4) qnorm() 함수를 이용하여 $N(170, 6^2)$ 의 분위 p에 해당하는 값을 구합니다**

```
qnorm(p, mean, sd)
- p : 구하고자 하는 분위 벡터(원소 1개짜리 포함)
- mean : 정규분포의 평균
- sd : 정규분포의 표준편차
```

```
15.qnorm(0.25, mean=mu, sd=sigma)
16.qnorm(0.5, mean=mu, sd=sigma)
17.qnorm(0.75, mean=mu, sd=sigma)
```

- 15~17줄 : 확률변수 X가 $N(170, 6^2)$ 을 따를 때 1사분위수(0.25), 중앙값(0.5), 3사분위수(0.75)를 구합니다.
  - 정규분포는 좌우대칭으로 평균과 중앙값이 같습니다.

정규분포

```
> qnorm(0.25, mean=mu, sd=sigma)
[1] 166
> qnorm(0.5, mean=mu, sd=sigma)
[1] 170
> qnorm(0.75, mean=mu, sd=sigma)
[1] 174
```

- **Step #5) rnorm() 함수를 이용하여** $N(170, 6^2)$ **에서 400개의 난수를 생성하고 모집단과 비교해 봅니다.**
  - 이 과정은 모집단이 $N(170, 6^2)$ 을 따를 때 400개의 확률표본을 추출하는 것을 R로 구현해보는 과정입니다.(이항분포에서도 마찬가지입니다.)

```
rnorm(n, mean, sd)
- n : 정규분포로 부터 추출할 난수의 개수
- mean : 정규분포의 평균
- sd : 정규분포의 표준편차
```

정규분포

```
19.options(digits=5)
20.set.seed(5)
21.smp <- rnorm(400, mean=mu, sd=sigma)
22.c(mean(smp), sd(smp))
23.hist(smp, prob=T, main="N(170, 6^2)으로부터 추출한 표본의 분포
   (n=400)", xlab="", ylab="", col="white", border="black")
24.lines(x, nd, lty=2, lwd=2, col="red")
```

- 19줄 : 표시할 자릿수를 5자리로 변경합니다.
- 20줄 : 난수의 초기값을 5로 설정합니다.
  - R을 비롯한 컴퓨터에서 난수는 유사난수로 초깃값에 따라 발생하는 난수가 결정됩니다.
    - 난수 생성시 일정하지 않은 초깃값을 사용하여 완전 난수처럼 보이게 할 뿐입니다.
  - set.seed() 함수는 초깃값을 사용자가 원하는 대로 지정합니다.
    - 예제와 동일한 초깃값을 사용한다면 동일한 난수가 발생합니다.

정규분포

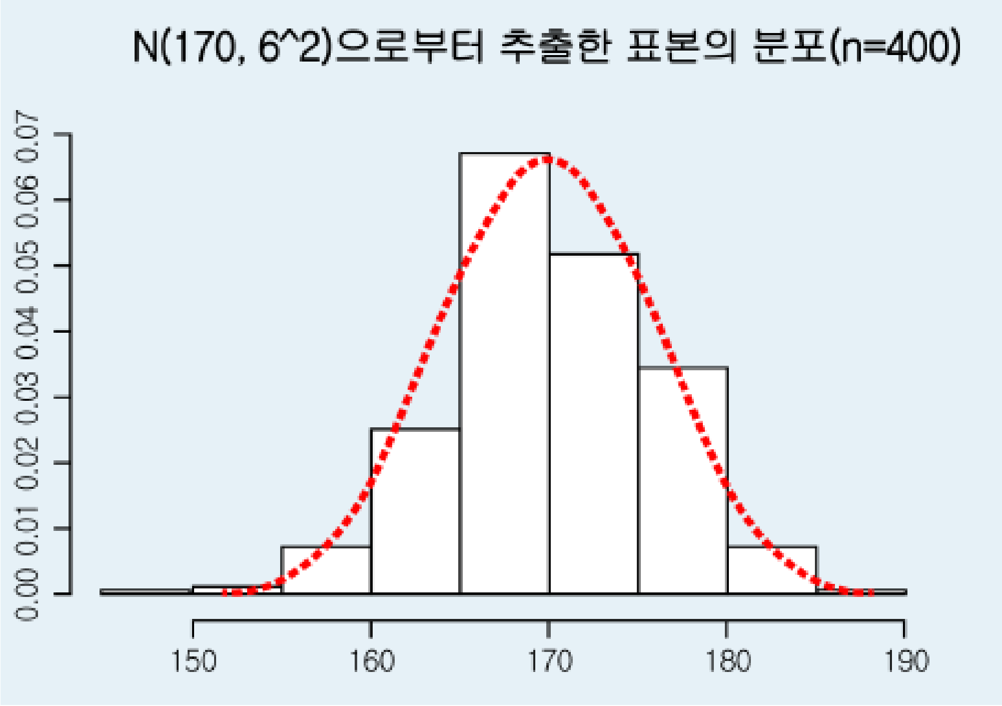
- 21줄 : 변수 smp에 $N(170, 6^2)$ 로부터 400개의 표본을 추출하고 저장합니다.
- 22줄 : 생성한 smp의 평균과 표준편차를 벡터로 출력했습니다.
  - 표본의 개수가 충분히 크다면 모집단의 평균과 표준편차와 비교해보면 크게 차이가 나지 않습니다.

```
> c(mean(smp), sd(smp))
[1] 170.0165   6.0054
```

- 23줄 : 추출한 표본의 상대도수(prob=T)로 히스토그램을 작성합니다
- 24줄 : 표본의 히스토그램 위에 점선(lty=2)으로, 선굵기는 2(lwd=2)인 붉은색(col="red")의 선으로 $N(170, 6^2)$ 의 분포를 표시했습니다.
  - 모집단의 분포와 차이는 있지만 많이 닮았음을 알 수 있습니다
  - lines() 함수와 같이 기존에 작성된 그래프 위에 각종 표현을 하는 함수를 **저수준 그래프 함수**라 합니다. lines() 함수는 전달된 좌표들을 선으로 연결합니다.
    - 첫번째와 두번째 전달인자인 x와 nd의 각 순서쌍 (x_1, nd_1), (x_2, nd_2), ... 를 좌표로 하여 각 점들을 선으로 연결합니다.

정규분포

```
lines(x, y, …)
- x : x좌표로 사용할 벡터(원소 1개짜리 포함)
- y : y좌표로 사용할 벡터(원소 1개짜리 포함)
- … : 그래프에서 사용하는 공통 전달인자 (lty, lwd, col 등등)
```



정규분포

예제 3-6 R을 이용해 정규분포의 특징 알아보기 준비파일 | 08.Z.to.Normal.R

- **개요**
  - R을 이용하여 정규분포의 특징을 알아봅시다.
    - 표준정규분포를 이용하여 하위 2.5%, 5%에 해당하는 값을 찾아봅시다.
    - 정규분포의 대칭성을 이용하여 상위 2.5%, 5%에 해당하는 값은 부호만 바뀝니다.
    - 정규분포의 분포함수를 이용하여 확률계산을 해 봅시다.
      - 그래프를 통해 어떤 면적인지 확인해 봅시다.
- **Step#1) 표준정규분포의 모수 준비**

```
1. options(digits = 4)
2. mu <- 0
3. sigma <- 1
```

정규분포

- 1줄 : 앞서 사용한 바와 같이 출력물이 네 자릿수가 되도록 합니다.
  - 다음의 코드를 실행 하기 전에 ex1, ex2, ex3가 어떻게 출력될지 예측하고 R에 입력하여 본인 생각과 일치하는지 확인해 봅시다.

```
ex1 <- 12.356
ex2 <- 0.001234
ex3 <- c(ex1, ex2)
```

  - 출력상 값이 digits로 정해진 방식에 따라 출력될 뿐 원래의 값은 유지하고 있습니다. 즉, options(digits=n) 으로는 값 자체를 바꾸지 않습니다.
- 2줄 : 표준정규분포의 평균인 0을 변수 mu에 저장합니다.
- 3줄 : 표준정규분포의 표준편차인 1을 변수 sigma에 저장합니다.

정규분포

- **Step #2)** 표준정규분포의 특별한 값을 찾습니다.

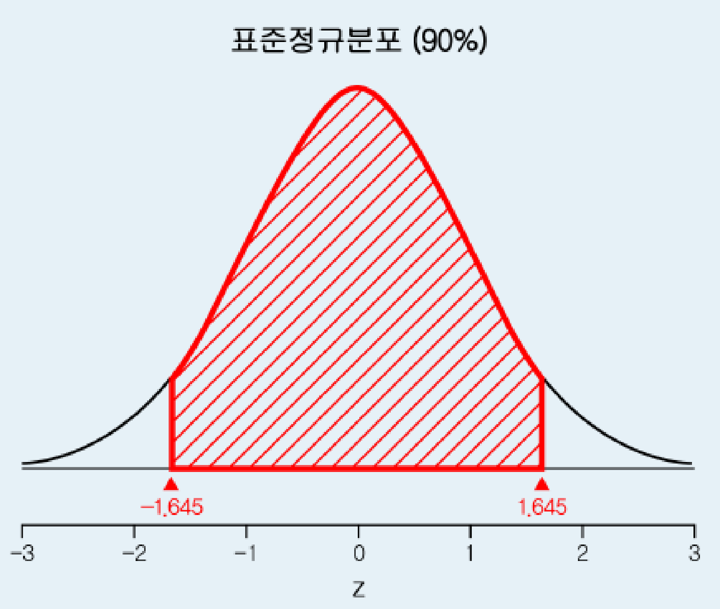
```
5. (p0.05 <- qnorm(0.05, mean=mu, sd=sigma))
6. (p0.025 <- qnorm(0.025, mean=mu, sd=sigma))
```

- 5줄 : $P(Z \le z) = 0.05$인 z값을 qnorm() 함수로 구합니다.
  - 표준정규분포에서 z가 -1.645보다 작을 확률은 약 0.05(5%)입니다.
  - 표준정규분포는 좌우대칭으로 z가 1.645보다 클 확률 역시 약 0.05(5%)입니다.
- 5줄 : $P(Z \le z) = 0.025$인 z값을 qnorm() 함수로 구합니다.
  - 표준정규분포에서 z가 -1.96보다 작을 확률은 약 0.025(2.5%)입니다.
  - 표준정규분포는 좌우대칭으로 z가 1.96보다 클 확률 역시 약 0.025(2.5%)입니다.

```
> (p0.05 <- qnorm(0.05, mean=mu, sd=sigma))
[1] -1.645
> (p0.025 <- qnorm(0.025, mean=mu, sd=sigma))
[1] -1.96
```

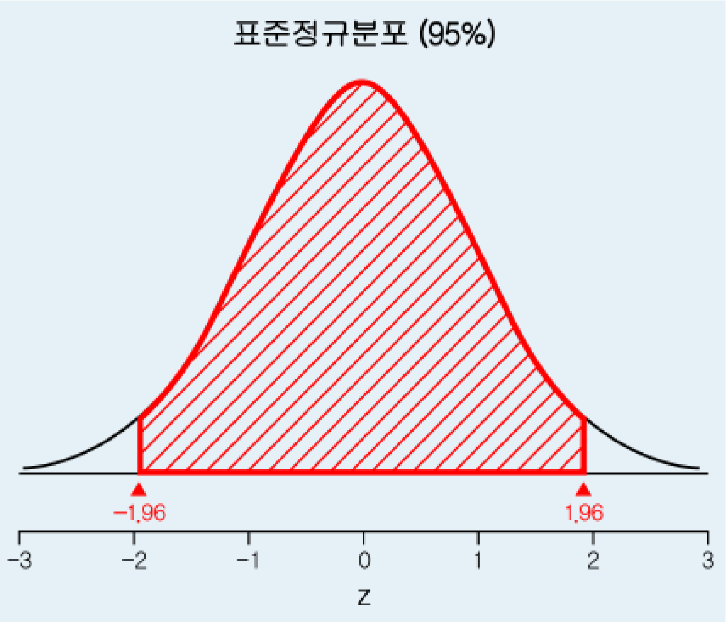
정규분포

- **Step #3)** 분포함수를 통해 원하는 구간의 면적을 찾습니다
  - 8 줄
    - 앞서 -1.645보다 작은 쪽의 면적은 0.05이고 1.645보다 큰 쪽의 면적 또한 0.05 이므로 그 사이의 면적은 0.9임을 알 수 있습니다.
    - pnorm() 함수를 이용하여 $P(-1.645 \le Z \le 1.645)$를 확인합니다.
    - 정규분포에서 (평균 - 1.645×표준편차)부터 (평균 + 1.645×표준편차) 사이에 들어갈 확률은 약 90%입니다.



정규분포

- 9 줄
  - 앞서 -1.96보다 작은 쪽의 면적은 0.025이고 1.96보다 큰 쪽의 면적 또한 0.025 이므로 그 사이의 면적은 0.95임을 알 수 있습니다.
  - pnorm() 함수를 이용하여 $P(-1.96 \leq Z \leq 1.96)$를 확인해합니다.
  - 정규분포에서 (평균 - 1.96×표준편차)부터 (평균 + 1.96×표준편차) 사이에 들어갈 확률은 약 95%입니다.
- pnorm() 함수를 이용하여 다양한 정규분포의 확률을 계산할 수 있습니다.



4장을 위한 준비

: 반복문(for)

반복문

- **반복문**
  - 프로그래밍의 가장 기본적인 요소로 특정한 행위를 데이터만 바꿔가면서 여러 번 반복하는 문장입니다.
    - 코드의 흐름을 위에서 아래로 내려가는 구조에서 벗어나 특정 코드들을 반복합니다.
  - 반복문은 통계학에서 모의실험을 할 때 필수적인 요소입니다.
    - 4장에서 간단한 모의실험을 할 때 이 반복문을 사용합니다.
  - R에서 사용하는 반복문은 여러 가지가 있지만, 가장 기본적인 for 에 대해 알아봅시다.
    - for문의 수행속도가 상당히 느린 것으로 알려져 가급적 사용하지 않는 사용자들도 있지만, 아주 큰 자료가 아니라면 큰 문제는 없습니다.
    - 반복문은 프로그래밍의 중요한 요소로 다같이 익혀봅시다.

반복문

- **for**
  - for 문은 특정 영역을 원하는 횟수만큼 반복하는 R의 문장으로 다음과 같이 작동합니다.

① for 문은 중괄호 { }로 둘러싸여진 R 코드들을 반복합니다.

② in 이후의 '벡터'의 원소 수만큼 반복합니다.

③ 한 번 반복할 때마다 벡터의 원소가 반복의 정보로 '반복정보가 저장될 변수'에 저장됩니다

반복문

예제 3-7 for 문의 사용 예

준비파일 | 10.for_ex.R

- **for 문이 어떻게 사용되는지 예를 통해 살펴봅시다.**
- **벡터의 원소값 순회하기**

```
1. v <- c(1, 4, 5)
2. for( i in v ) {
3.   print( i )
4. }
```

```
[1] 1
[1] 4
[1] 5
```

- 1줄 : 1, 4, 5로 구성된 벡터를 변수 v에 저장합니다.
- 2, 4줄 : 벡터 변수 v의 원소 수만큼, 즉 3번 3번 줄을 반복합니다.
- 3줄 : 반복정보가 저장되는 변수 i의 값을 출력합니다.
  - 벡터 변수 v의 첫 번째, 두 번째, 세 번째 값이 반복마다 변수 i에 저장되고 그 값을 출력합니다.
  - print() 함수는 전달되는 변수의 값을 출력하는 함수입니다.

반복문

- **벡터를 구성하는 원소들의 합 구하기**
  - R이 내장하고있는 sum() 함수와 같은 기능을 흉내내어 봅시다.
  - 합을 구하는 과정보다 합을 구하기 위해 필요로 하는 추가 정보(변수)를 생성하는 것을 눈여겨 봐주시기 바랍니다.
  - 벡터의 위치정보(인덱스)를 위해 1:10 을 반복문에 사용한 점을 주의깊게 봐주세요.
  - 우리가 구한 벡터합과 sum() 함수와 결과를 비교해 봅시다.

```
r.n <- rnorm(10)
sum <- 0
for(i in 1:10) {
  sum <- sum + r.n[i]
}
print(sum)
sum(r.n)
```

반복문

- 6줄 : 표준정규분포로부터 10개의 난수를 생성하여 벡터 변수 r.n에 저장합니다.
- 7줄 : 합을 구할 변수를 합의 항등원인 0으로 초기화합니다.
- 8, 10 줄
  - in 이후의 벡터는 1부터 10까지 1씩 증가하는 벡터로 총 10개의 원소를 갖는 벡터입니다.
  - **이 벡터의 값을 이용하여 다른 벡터의 i번째 원소에 접근합니다.**
- 9줄
  - 매 반복마다 i의 값은 1부터 1씩 증가합니다.
  - 첫 번째 반복에서는
    - sum 변수의 값이 0 입니다.
    - r.n[i]는 위에서 생성한 난수가 저장된 r.n 변수의 첫 번째, 즉 r.n[1]의 값을 가져옵니다.
    - sum에 r.n[1]의 값을 더한 후 그 결과를 sum에 다시 넣습니다.
    - for 문을 풀어서 보면 다음과 같습니다.

반복문

```
 1번째 반복 : sum <- sum + r.n[1]
 2번째 반복 : sum <- sum + r.n[2]
 3번째 반복 : sum <- sum + r.n[3]
                                   ...
 8번째 반복 : sum <- sum + r.n[8]
 9번째 반복 : sum <- sum + r.n[9]
10번째 반복 : sum <- sum + r.n[10]
```

- 11줄 : print() 함수는 주어진 변수의 값을 출력하는 함수로 위에서 구한 난수의 합(sum)을 출력합니다. (난수이므로 동일하지 않을 수 있습니다.)
- 12줄 : 우리가 구한 합과 R이 내장하고 있는 벡터의 합을 구하는 sum() 함수와 비교해봅시다

```
> print(sum)
[1] 0.1402
> sum(r.n)
[1] 0.1402
```

반복문

- **구구단의 2단 구하기**

```
14.dan <- 2
15.for( i in 2:9 ) {
16.   times <- dan * i
17.   print( paste(dan, "곱하기", i, "=", times) )
18.}
```

- 14줄 : 변수 dan에 우리가 구하려는 단인 2를 넣습니다.
- 15, 18줄 : in 이후의 벡터는 2부터 9까지 1씩 증가하는 벡터(2:9)로 2단에서 곱해지는 값입니다.
- 16줄 : 변수 times는 dan의 값과 반복정보가 저장되는 변수 i의 곱을 저장합니다.
  - 변수 i는 첫 번째 반복에서 2:9 벡터의 첫 번째 원소가 들어가고, 그 다음 반복에서는 2:9 벡터의 두 번째 원소가 들어가는 식으로 매 반복에 맞춰 2:9 벡터의 값이 순서대로 들어갑니다.
- 17줄 : paste() 변수를 이용하여 결과를 문자열로 만든 값을 print() 함수를 이용하여 출력합니다. (2단이 출력됩니다.)

반복문

```
[1] "2 곱하기 2 = 4"
[1] "2 곱하기 3 = 6"
[1] "2 곱하기 4 = 8"
[1] "2 곱하기 5 = 10"
[1] "2 곱하기 6 = 12"
[1] "2 곱하기 7 = 14"
[1] "2 곱하기 8 = 16"
[1] "2 곱하기 9 = 18"
```

반복문에 의해
in 이후 벡터 안의 원소만큼 반복하고
반복마다 벡터 안의 원소를 순회하여
값이 변경되는 것을 주의해서 봐주세요.

paste() 함수
전달인자로 변수와 문자열 등을 받아 **하나의 문자열로 합칩니다.**
paste() 함수에서 합칠 때 기본값으로 빈 칸 하나(" ")를 사용하지만,
sep=" " 전달인자를 사용하여 우리가 원하는 문자열로 합칠 수 있습니다.

예제)

```
> paste("You", "I", sep="&")
[1] "You&I"
```

반복문

- **for 문의 중첩 사용 : 행렬의 각 원소 출력하기**
  - 행과 열로 구성된 행렬을 순회해 봅시다. (행렬은 구성하는 모든 자료가 동일한 자료구조입니다. 부록 B 참고)
    - 행렬을 만드는 함수는 matrix() 입니다.
  - 행 요소 순회와 열 요소 순회에 따라 두 개의 반복문이 필요합니다.
  - 두 개의 반복은 하나의 반복문이 다른 반복문을 감싸고 있습니다.(중첩)
    - 바깥의 반복문에 의해 안에 있는 반복문 자체도 반복됩니다.
    - 23줄의 i와 j의 반복값을 출력을 통해 확인해 주세요.

```
20.(m <- matrix(1:12, ncol=3))
21.for(i in 1:nrow(m)) {
22.  for(j in 1:ncol(m)) {
23.    cat( i, "행", j, "열 =", m[i,j], "\n")
24.  }
25.}
```

반복문

- 20줄 : 변수 m에 열의 수가 3인 1부터 12까지의 행렬을 만들어 저장합니다.
  - matrix() 함수는 기본 상태에서는 값을 열 우선으로 채웁니다.
  - 만일 행부터 채우려면 byrow 전달인자의 값으로 논리값 TRUE를 넣습니다 (byrow=TRUE).

```
> (m <- matrix(1:12, ncol=3))
     [,1] [,2] [,3]
[1,]    1    5    9
[2,]    2    6   10
[3,]    3    7   11
[4,]    4    8   12
```

반복문

- 21, 25 줄
  - nrow() 함수는 행렬 혹은 데이터 프레임의 행의 수를 구하는 함수로 1:4의 결과를 가져와 22~24줄을 4번 반복합니다.
  - 각 반복마다 변수 i에 1부터 4까지 순차적으로 값을 넣습니다(행의 수 변화).
- 22, 24줄 : 중첩된 반복문으로 앞서 반복에 의해 4번 반복합니다.
  - 각 반복마다 1:3의 반복을 합니다.
  - 반복의 정보로 변수 j에 1부터 3까지 순차적으로 값을 넣습니다
  - 열의 수 변화, ncol()은 행렬 혹은 데이터 프레임의 열의 수를 구합니다
- 23줄: cat() 함수는 paste() 함수와 동일하나 그 결과를 자료로 만드는 것이 아니라 화면에 출력하는 역할을 합니다.
  - 앞에서 paste() 함수와 결과물의 차이를 보면 큰 따옴표로 묶이지 않는 것을 확인할 수 있는데, 이는 데이터로서의 역할은 하지 않는 것으로 보면 됩니다.(단순 출력)

반복문

바깥 반복문 : 행 변화

i = 1

i = 2

i = 3

i = 4

```
> m
     [,1] [,2] [,3]
[1,]    1    5    9
[2,]    2    6   10
[3,]    3    7   11
[4,]    4    8   12
```

j = 1   j = 2   j = 3

안쪽 반복문 : 열 변화
각각의 i에 대해

m[i,j]

i=1 일 때 j=1, j=2, j=3 반복

i=2 일 때 j=1, j=2, j=3 반복

i=3 일 때 j=1, j=2, j=3 반복

i=4 일 때 j=1, j=2, j=3 반복

반복문

- cat 함수 마지막 전달에서 "\n"으로 했는데, 여기서 역 슬래시(\)는 키보드 상에서 ₩ 키를 눌러 입력합니다.
  - 문자열에서 역 슬래시는 탈출문자로 탈출문자 이후 한 문자에 대해 문자열이 아닌 키보드 상의 기능을 나타냅니다.
- 문자열에서는 탈출문자의 역할로 n은 새로운 줄을 의미합니다.
  - paste()에서는 이 또한 데이터로 저장합니다.
- 출력은 행렬의 행과 열별로 어떤 원소가 있는지를 보여줍니다

```
1 행 1 열 = 1
1 행 2 열 = 5
        …
3 행 3 열 = 11
4 행 1 열 = 4
4 행 2 열 = 8
4 행 3 열 = 12
```

Q & A

수고하셨습니다.