빅데이터의 이해와 활용

Understanding and Using Big Data

08

확률의 개념

학습 내용

01 확률이란

02 확률변수

03 확률분포

04 실습

학습 목표

- 확률을 정의하고 확률 법칙에 대해 이해하고 설명할 수 있다.
- 확률 용어 및 변수에 대해 이해하고 설명할 수 있다.
- 확률분포의 종류와 특성에 대해 이해하고 설명할 수 있다.

생각
해보기

순열과 조합

색깔 볼펜 5자루와 연필 3자루가 있습니다.

1 이 중에서 필기구 한 자루를 선택할 수 있는 경우의 수(합의 법칙)

2 이 중에서 필기구 두 자루를 선택하는데 볼펜과 연필 각각 한 자루씩을 선택하는 경우의 수(곱의 법칙)

3 이 중에서 필기구 두 자루를 선택하여 배열 하는 경우의 수(순열)

4 이 중에서 필기구 두 가지를 선택하여 가지는 경우의 수(조합)

01

확률이란

1) 확률의 사전적 의미
2) 확률의 기원
3) 확률의 정의
4) 확률 용어
5) 확률 법칙
6) 확률 문제

1) 확률의 사전적 의미

- 확률이라는 단어

確	率
굳을 확	비율 률

(어떤 결정 등을) 굳힐 비율

1) 확률의 사전적 의미

- 확률이라는 단어

"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

2) 확률의 기원

- 17세기 도박사 슈발리에 드메리와 파스칼의 서신 내용

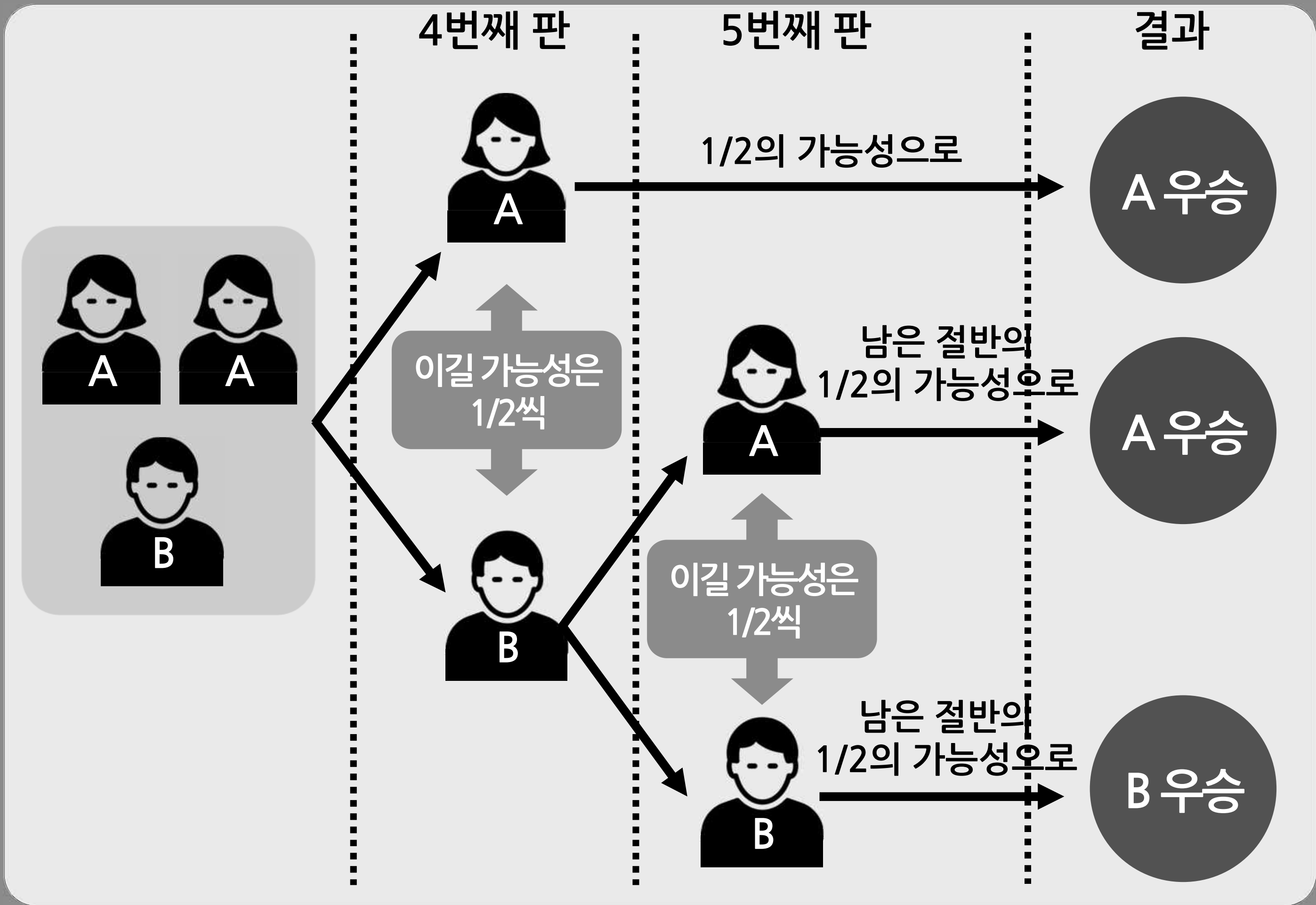
서로 능력이 동일한 갬블러 A와 B가 있습니다.

5판 3선승제의 게임을 통해 우승자를 가리는 게임이 시작되었습니다.
각각 32 피스톨씩 내고 우승자가 64개의 피스톨을 가져가기로 하였습니다.
3번의 게임을 한 결과 A가 두 번을 이겼고, B가 한 번을 이겼습니다.

4번째 판이 시작되려는 순간 천재지변으로 더 이상 게임을 할 수 없게 된 상황이 발생하였습니다. A가 우승했다고 하기도 곤란하고, B가 우승했다고 하기도 곤란한 상황입니다.

상품을 어떻게 나누는 것이 좋을까요?

▷ 결과



A가 이길 확률

A가 4번째 판을
이길 확률

A가 4번째 판을 지고
5번째 판을 이길 확률

B가 이길 확률

B가 4번째 판과
5번째 판을 이길 확률

3) 확률의 정의

확률(確率)

어떤 사건이 실제로 일어날 것인지 혹은 일어났는지에 대한 지식 혹은 믿음을 표현하는 방법이며 같은 원인에서 특정한 결과가 나타나는 비율

수학에서는 확률론에서 설명하고 있음

수학, 통계학, 회계, 도박, 과학과 철학에서 어떤 잠재적 사건이 일어날 경우의 가능성과 이 가능성 안에 있는 복잡한 시스템의 구조에 대한 답을 이끌어내기 위해 사용되고 있음

〈출처 : 위키피디아〉

3) 확률의 정의

빈도주의

- 모든 가능성을 지닌 경우의 수에 대해 원하는 경우의 수의 비

예 6면체 주사위에서 2의 배수가 나올 확률은 1/2이고, 3의 배수가 나올 확률은 1/3임

베이즈 확률

- 어떤 사건에 대한 사전 확률을 활용하여 이 사건이 발생할 확률을 계산하는 것
- 사전 확률은 이전에 발생한 동일 사건의 확률 값을 사용하거나 개인의 주관적인 판단에 의해 정할 수 있음

예 Monty Hall 문제

3) 확률의 정의

- **수학적 확률**

01 어떤 시행의 결과로 나타날 수 있는 가능한 결과의 수 : O

02 각 결과들이 나타날 가능성은 동일하다는 가정

03 동일한 각 결과들의 확률 : 1 / O

3) 확률의 정의

● 수학적 확률

임의의 사건 A가 발생할 수학적 확률

- 표본 공간의 원소의 개수(O) 중 사건 A에 해당하는 근원사건의 개수(n)임(n/O)

예 주사위를 굴려 홀수가 나올 확률

- 주사위의 각 눈이 나올 확률은 전체 6개의 눈으로 구성되어 있으며 각각이 나올 확률은 동일하다고 가정하면 확률은 1/6임
- 홀수인 사건을 구성하는 근원사건의 수는 {1, 3, 5} 세 개임
- 전체 눈의 개수는 6이고 이로부터 홀수 눈의 확률은 3/6 = 1/2임

3) 확률의 정의

● 통계적 확률

01 동일한 조건에서 같은 실험을 N번 반복

02 사건 A가 모두 몇 번 발생했는지를 조사 : n

03 사건 A가 발생할 확률 $P(A) = n / N$

3) 확률의 정의

통계적 확률

- 실험의 반복횟수 N은 매우 커야 그 값을 받아 들일 수 있음
- 반복 횟수가 커짐에 따라 사건 A의 상대 도수 (n/N)가 상수 P(A)로 접근해 가는 경향을 보임

3) 확률의 정의

● 통계적 확률

▶주사위 실험

시행 횟수	1의 눈	2의 눈	3의 눈	4의 눈	5의 눈	6의 눈
12	1	3	3	2	2	1
1,200	211	214	196	204	202	173
12,000,000	2,002,632	1,999,749	2,0003,28	1,999,958	1,996,037	2,001,296

3) 확률의 정의

- 확률 공리

확률이 만족해야 하는 기본 성질을 정리함

- 러시아 수학자 안드레이 콜모고로프
(Andrey Nikolaevich Kolmogorov, 1903~1987)

3) 확률의 정의

● 확률 공리

확률이 만족해야 하는 기본 성질

1. P(A)는 0과 1사이의 값을 갖고($0 \le P(A) \le 1$)
2. 반드시 일어나는 사건(표본 공간 전체)의 값은 1이며 ($P(\Omega) = 1$)
3. 서로 배반인 사건 $A_1, A_2, \cdots, A_n, \cdots$의 합집합에 대해 다음을 만족하면

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

➡ 함숫값 P(A)를 사건 A의 확률이라 함

4) 확률 용어

확률실험	Experiment

성립조건

- 결과를 구하기 위한 어떤 실험을 통해 나타나는 결과를 알지 못함
- 결과는 알지 못하지만 결과로 나타날 수 있는 가능한 경우를 알고 있음
- 동일한 실험을 몇 번이고 반복할 수 있음

예제

- 동전을 던지기 전에 '앞면'이 나올지 '뒷면'이 나올지 알 수 없음
- 가능한 결과는 '앞면'과 '뒷면' 중에 하나임을 알고 있음
- 동전을 던지는 실험은 몇 번이고 반복할 수 있음

4) 확률 용어

표본 공간

- Ω (Omega)
- 확률실험으로부터 출현 가능한 모든 결과들의 모임

예 동전 던지기의 경우 Ω = {앞면, 뒷면} = {H, T}

4) 확률 용어

사건(Event)

표본 공간의 각 원소(출현 가능한 개별 결과)들의 부분집합

- 사건을 기호로 나타낼 때는
  영어 대문자(A, B, C, … 등)로 표현함

근원사건 (Elementary Event)	어떤 사건이 표본 공간상의 하나의 원소로 구성된 사건

5) 확률 법칙

- **덧셈 법칙**

임의의 사건 A와 사건 B의 합 사건에 대한 확률

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

만일 두 사건 A와 B가 서로 배반이라면 ($A \cap B = \phi$)

➡ $P(A \cup B) = P(A) + P(B)$

5) 확률 법칙

● 곱셈 법칙

조건부 확률

두 사건 A와 B에 대해…

P(A | B) : 사건 B가 발생했을 때 사건 A가 발생할 확률

➡ $$P(A|B)=\frac{P(A\cap B)}{P(B)}, P(B)>0$$

P(B | A) : 사건 A가 발생했을 때 사건 B가 발생할 확률

➡ $$P(B|A)=\frac{P(A\cap B)}{P(A)}, P(A)>0$$

5) 확률 법칙

● 곱셈 법칙

▶ 예제

Q1 주사위를 던지는 실험에서 주사위의 눈이 짝수인 사건을 A,
주사위의 눈이 4 이상인 사건을 B라 할 때 P(A|B)를 구해 봅시다.

풀이

- $P(A|B) = \frac{P(A\cap B)}{P(B)}$ 이므로 P(B)와 $P(A\cap B)$를 구함

① $P(B)$: 주사위의 눈이 4 이상이 나올 확률은 1/2임

② $P(A\cap B)$: 주사위의 눈이 짝수이고 4 이상인 경우는 {4, 6}이므로 확률은 1/3임

③ $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{1/3}{1/2} = \frac{2}{3}$

5) 확률 법칙

● 곱셈 법칙

▶ 예제

Q1 주사위를 던지는 실험에서 주사위의 눈이 짝수인 사건을 A,
주사위의 눈이 4 이상인 사건을 B라 할 때 P(A|B)를 구해 봅시다.

풀이 : 표본 공간의 변화

- 표본공간의 수정으로 주어진 사건 재정의
  - ① $\Omega = \{1, 2, 3, 4, 5, 6\} \rightarrow \Omega_B = \{4, 5, 6\}$
  - ② 변화된 표본 공간 Ω_B에서 사건 A가 발생할 확률
  - ③ Ω_B 상에서 짝수의 눈은 {4, 6}이므로 2/3

5) 확률 법칙

● 곱셈 법칙

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, P(A) > 0$$

✔ 두 사건 A와 B에 대한 조건부 확률을 아래와 같이 변형할 수 있음

$$P(A \cap B) = \begin{cases} P(B)P(A|B), & P(A)>0 \\ P(A)P(B|A), & P(B)>0 \end{cases}$$

5) 확률 법칙

● 독립 사건과 조건부 확률

만일 두 사건 A와 B가 독립이라면…

$$P(A\cap B)=P(A)P(B)$$

➡ 두 사건이 서로의 발생에 영향을 끼치지 않는다면 곱 사건의 확률은 두 사건의 곱이 됨

예 동전을 두 번 던져 처음 앞면이 나온 것이 두 번째 던졌을 때 영향을 끼치지 않음

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 발생 여부가 사건 A의 발생에 영향을 끼치지 않음

5) 확률 법칙

여사건의 확률

사건 A의 여사건 A^c의 사건 $P(A^c)$는…

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

$$P(A) + P(A^c) = 1$$

5) 확률 법칙

● 예제

Q2

주사위를 두 번 던지는 실험에서 '첫 번째 던진 주사위 눈이 짝수인 사건'과 '두 번째 던진 주사위 눈이 3의 배수인 사건'에 대해 합사건, 곱사건의 확률을 구해 봅시다.

풀이

- 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
  - $P(B) = \frac{12}{36} = \frac{1}{3}$

6) 확률 문제

● 예제

Q2 주사위를 두 번 던지는 실험에서 '첫 번째 던진 주사위 눈이 짝수인 사건'과 '두 번째 던진 주사위 눈이 3의 배수인 사건'에 대해 합사건, 곱사건의 확률을 구해 봅시다.

사건 A

두 번째 \ 첫 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

6) 확률 문제

● 예제

Q2 주사위를 두 번 던지는 실험에서 '첫 번째 던진 주사위 눈이 짝수인 사건'과 '두 번째 던진 주사위 눈이 3의 배수인 사건'에 대해 합사건, 곱사건의 확률을 구해 봅시다.

풀이 **: 사건 A와 사건 B의 표본 공간**

- **합사건**
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $\mathrm{P}(A \cap B) = \{ (2,3),(2,6),(4,3),(4,6),(6,3),(6,6)\}$ 이므로 $\mathrm{P}(A \cap B) = {}^{1}/_{6}$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B) = {}^{1}/_{2} + {}^{1}/_{3} - {}^{1}/_{6} = {}^{2}/_{3}$

6) 확률 문제

● 예제

Q2 주사위를 두 번 던지는 실험에서 '첫 번째 던진 주사위 눈이 짝수인 사건'과 '두 번째 던진 주사위 눈이 3의 배수인 사건'에 대해 합사건, 곱사건의 확률을 구해 봅시다.

풀이 : 사건 A와 사건 B의 표본 공간

- 곱사건
  - $P(A \cap B) = P(B)P(A|B)$
  - $P(A|B) = {}^{P(A\cap B)}/_{P(B)}$ 를 사건 B가 조건인 새로운 표본 공간으로 변화시켜 봄

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

6) 확률 문제

● 예제

Q2 주사위를 두 번 던지는 실험에서 '첫 번째 던진 주사위 눈이 짝수인 사건'과 '두 번째 던진 주사위 눈이 3의 배수인 사건'에 대해 합사건, 곱사건의 확률을 구해 봅시다.

풀이 : 사건 A와 사건 B의 표본 공간

- 곱사건
  - 앞의 표로부터 $P(A|B) = {}^{6}/_{12} = {}^{1}/_{2}$임
  - $P(A \cap B) = P(B)P(A|B) = {}^{1}/_{3} \cdot {}^{1}/_{2} = {}^{1}/_{6}$

02

확률변수

1) 확률변수
2) 평균과 기댓값

1) 확률변수

- 동전을 두 번 던져 앞면이 나오는 횟수 관찰

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

1) 확률변수

확률변수

동전을 두 번 던져 앞면이 나오는 횟수처럼
표본 공간의 각 원소를 실숫값에 대응시키는 함수

확률변수	알파벳 대문자 X, Y, Z, …로 표기함
확률변수가 취하는 실숫값	알파벳 소문자 x, y, z …로 표기함
확률변수 X가 값 실숫값 x를 가질 때	X = x로 표기함

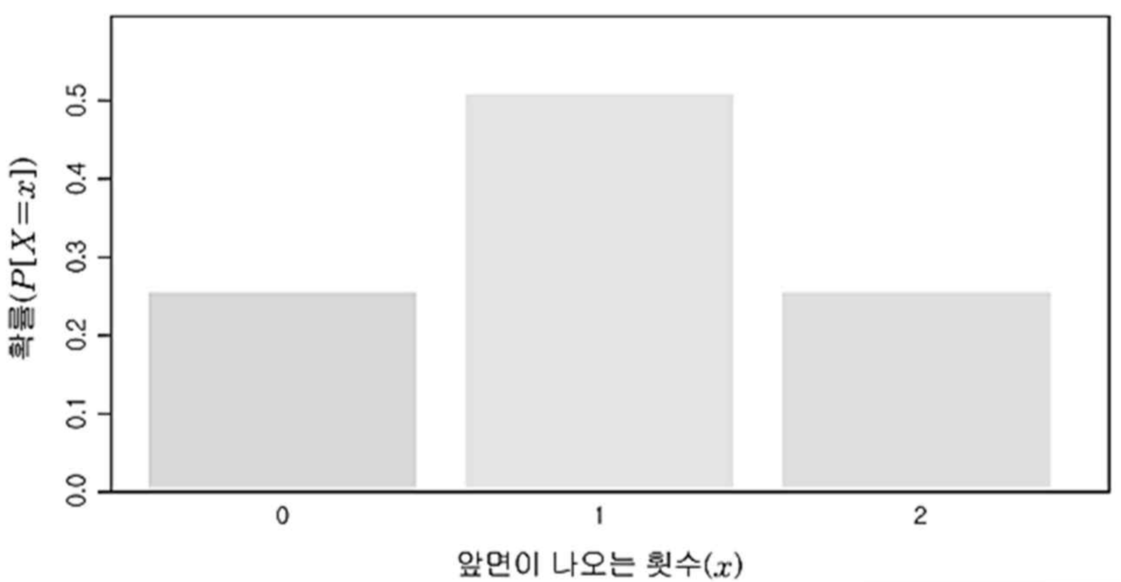
1) 확률변수

- 확률변수 X가 가질 수 있는 모든 x_i들에 확률이 대응되고, 확률변수는 이 확률에 따라 실숫값을 가짐
- 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)

1) 확률변수

<table>
<tr><th>표본 공간</th><th>표본 공간에서의 확률</th><th>X = x</th><th>P(X = x)</th></tr>
<tr><td>{H, H}</td><td>1/4</td><td>2</td><td>1/4</td></tr>
<tr><td>{H, T}</td><td>1/4</td><td rowspan="2">1</td><td rowspan="2">1/2</td></tr>
<tr><td>{T, H}</td><td>1/4</td></tr>
<tr><td>{T, T}</td><td>1/4</td><td>0</td><td>1/4</td></tr>
</table>



1) 확률변수

● 확률변수의 평균과 분산

Q3 확률변수 X가 동전을 두 번 던져 앞면이 나오는 횟수일 때의 평균과 분산을 구해 봅시다.

풀이 : 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 $1/n$ 은 각 자료들이 모두 동일하게 $1/n$의 비중을 갖고 있음을 나타냄

➡ 확률변수에서는 $1/n$에 해당하는 비중이 각 값이 나타날 확률로 바뀜

1) 확률변수

● 확률변수의 평균과 분산

Q3 확률변수 X가 동전을 두 번 던져 앞면이 나오는 횟수일 때의 평균과 분산을 구해 봅시다.

풀이 : 확률변수의 평균

- **확률변수 X의 평균**
  - $E(X) = \sum_{모든\ x} x \cdot P(X = x)$
- **$E(X)$는 확률변수 X의 평균을 나타내는 기호로, 확률변수의 평균을 기댓값이라고 함**
- **확률변수 X의 평균 즉, 기댓값을 구해보고자 함**
  - $E(X) = \sum_{i=1}^{x} x_i \cdot P(X = x_i) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = \frac{(0 \cdot 1 + 1 \cdot 2 + 2 \cdot 1)}{4} = \frac{4}{4} = 1$

1) 확률변수

● 확률변수의 평균과 분산

Q3 확률변수 X가 동전을 두 번 던져 앞면이 나오는 횟수일 때의 평균과 분산을 구해 봅시다.

풀이 : 확률변수의 분산

- 분산은 편차 제곱의 평균임
- 확률변수의 분산에 그대로 적용해 보고자 함
  - $Var(X) = E[(X - E(X))^2]$

➡ 확률변수의 분산은 $Var(X)$ 혹은 $V(X)$로 나타냄

➡ 기댓값을 구할 때와 마찬가지로 상수 자료들의 분산을 구할 때 사용한 $1/n$이 확률로 바뀐다는 점 외에 나머지는 동일함

1) 확률변수

● 확률변수의 평균과 분산

Q3 확률변수 X가 동전을 두 번 던져 앞면이 나오는 횟수일 때의 평균과 분산을 구해 봅시다.

풀이 : 확률변수의 분산

- 확률변수 X의 분산을 구해보고자 함
  - 분산의 간편식을 이용함
  - 앞서 구한 기댓값($E(X)$)은 1임
  - 확률변수의 제곱의 기댓값을 구해보고자 함

$$\Rightarrow E(X^2) = \sum_{i=1}^{3} {x_i}^2 \cdot P(X = x_i) = 0^2 \frac{1}{4} + 1^2 \frac{2}{4} + 2^2 \frac{1}{4} = \frac{(0^2 \cdot 1 + 1^2 \cdot 2 + 2^2 \cdot 1)}{4} = \frac{6}{4}$$

  - $Var(X) = E(X^2) - [E(X)]^2 = \frac{6}{4} - 1^2 = \frac{2}{4} = \frac{1}{2}$

2) 평균과 기댓값

● 평균

평균

어떤 자료가 있을 때, 이 자료를 대표하는 값 중의 하나로,
모든 자료 값을 더한 것을 자료의 개수로 나눈 값

N개의 자료가 있고, I번째 자료의 값을
X_i 라고 할 때, 평균 m(mean)은?

$$m= \frac{1}{N}\sum_{i=1}^{N} xi$$

2) 평균과 기댓값

● 기댓값

기댓값

확률적 사건에 대한 평균값으로 사건이 벌어졌을 때의 이득과 그 사건이 벌어질 확률을 곱한 것을 전체 사건에 대해 합한 값

N 개의 자료가 있고, I 번째 자료의 값을 X_i라 하고,
X_i 에 대한 확률 값을 P(X_i)라고 할 때, 기댓값 E(x)은?

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} xi\,P(x_i)$$

※ 평균과 기댓값은 같은 값을 가지지만 사용되는 성격과 정의가 다름

〈참고 사이트: http://blog.daum.net/rhaoslikesan/293〉

03

확률분포

1) 확률분포
2) 베르누이 시행
3) 이항분포

1) 확률분포

확률분포

확률변수가 취할 수 있는 값과 발생할 확률을 대응한 관계

누적분포함수

- 확률변수가 X 가질 수 있는 임의의 실측 값 x에 대해 다음과 같이 정의된 함수 F를 확률변수 X의 누적분포함수(Cumulative Density Function) 또는 간략히 분포함수라고 함

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

1) 확률분포

● 확률질량함수, 확률밀도함수

확률변수 X가 실측값 x를 가질 확률($P(X=x)$)에 대한 함수를 $f(x)$로 나타냄

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

확률질량함수(PMF)

- Probability Mass Function
- 확률변수가 취하는 값이 이산형일 경우

확률밀도함수(PDF)

- Probability Density Function
- 확률변수가 취하는 값이 연속형일 경우

2) 베르누이 시행

베르누이 시행

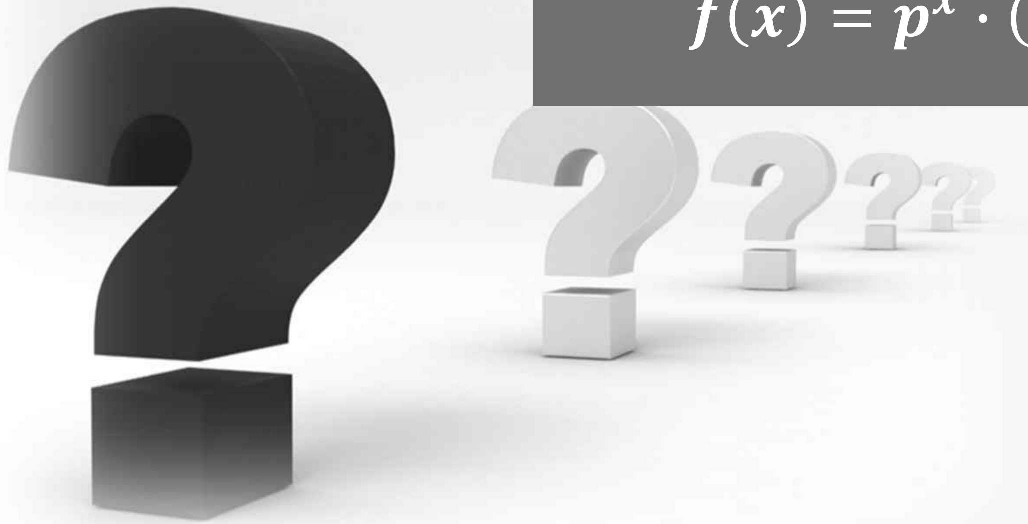
p의 확률로 원하는 결과가 나타났을 때 '성공'으로,
1-p의 확률로 그렇지 않은 결과가 나타났을 때 '실패'로
하는 두 가지 결과가 나타나는 확률실험

✔ 성공 확률 p가 베르누이 시행의 모수임

2) 베르누이 시행

확률변수 X가 베르누이 시행에 따라
성공일 때 1, 실패일 때 0을 가질 경우 확률질량함수는?

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



▷ 주사위를 던져 3의 배수의 눈이 나오면 상금을 얻는 게임의 예

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

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

2) 베르누이 시행

- 기댓값 : p

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

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

2) 베르누이 시행

- 분산 : p · (1-p)

$Var(X)$	$= E(X^2) - (EX)^2$
	$= \sum_{\text{모든 } x} \{x^2 \cdot f(x)\} - p^2$
	$= \sum_{\text{모든 } x} \{(0^2 \cdot (p^0 \cdot (1-p)^1) + 1^2 \cdot (p^1 \cdot (1-p)^0)\} - p^2$
	$= p - p^2 = p(1-p)$

3) 이항분포

개요

성공 확률이 p로 동일한 베르누이 시행을
n번 반복해서 실험하는 경우

- 실험이 n번 반복되더라도 성공 확률 p는 변하지 않고 동일함
- 각 실험이 서로 독립적으로 시행함
(IID, Independent and Identically Distributed)

3) 이항분포

● 개요

이항분포

n번 반복 실험에서 성공의 횟수가 따르는 분포

이항분포의 모수	• n : 시행의 횟수 • p : 성공의 확률
이항분포의 표기	두 모수를 이용하여 $B(n, p)$으로 표기함

→ 확률변수 X가 이항분포를 따를 때 $X \sim B(n, p)$와 같이 나타냄

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4

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.

풀이

- 확률변수 X를 성공의 확률이 1/3인 베르누이 시행을 3번 독립으로 반복해서 실험하였을 때 성공의 횟수라 할 때 확률변수 X는 다음과 같은 이항분포를 따름

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

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

〈1/12〉

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4

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.

풀이 **X=0일 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 구함 (성공의 확률 p=1/3)
  - $f_{Ber}(x_1 = 0) = \frac{1}{3}^{x_1=0} \frac{2}{3}^{1-(x_1=0)}$
  - $f_{Ber}(x_2 = 0) = \frac{1}{3}^{x_2=0} \frac{2}{3}^{1-(x_2=0)}$
  - $f_{Ber}(x_3 = 0) = \frac{1}{3}^{x_3=0} \frac{2}{3}^{1-(x_3=0)}$

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4

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.

풀이 X=0일 때 : 성공의 횟수가 0일 때

- 모두 실패한 경우는 $x_1 = 0$, $x_2 = 0$, $x_3 = 0$의 한 가지 경우만 있음
- 각 실험이 독립이므로 각 확률들의 곱으로 나타낼 수 있음
  - $f_{Ber}(x_1 = 0) \cdot f_{Ber}(x_2 = 0) \cdot f_{Ber}(x_3 = 0)$
  - $P(X = 0) = \frac{1}{3}^{x_1=0} \frac{2}{3}^{1-(x_1=0)} \cdot \frac{1}{3}^{x_2=0} \frac{2}{3}^{1-(x_2=0)} \cdot \frac{1}{3}^{x_3=0} \frac{2}{3}^{1-(x_3=0)} = \frac{1}{3}^{0} \frac{2}{3}^{3}$

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4

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.

풀이 **X=1일 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 가짐

  ① 첫 번째 성공했을 때 : $x_1 = 0, x_2 = 0, x_3 = 0$

  - $\left(\frac{1}{3}\right)^{x_1=1}\left(\frac{2}{3}\right)^{1-(x_1=1)} \cdot \left(\frac{1}{3}\right)^{x_2=0}\left(\frac{2}{3}\right)^{1-(x_2=0)} \cdot \left(\frac{1}{3}\right)^{x_3=0}\left(\frac{2}{3}\right)^{1-(x_3=0)} = \left(\frac{1}{3}\right)^{1}\left(\frac{2}{3}\right)^{2}$

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4 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.

풀이 **X=1일 때 : 성공의 횟수가 1일 때**

② 두 번째 성공했을 때 : $x_1 = 0, x_2 = 1, x_3 = 0$

- $\frac{1}{3}^{x_1=0}\ \frac{2}{3}^{1-(x_1=0)} \cdot \frac{1}{3}^{x_2=1}\ \frac{2}{3}^{1-(x_2=1)} \cdot \frac{1}{3}^{x_3=0}\ \frac{2}{3}^{1-(x_3=0)} = \frac{1}{3}^{1}\ \frac{2}{3}^{2}$

③ 세 번째 성공했을 때 : $x_1 = 0, x_2 = 0, x_3 = 1$

- $\frac{1}{3}^{x_1=0}\ \frac{2}{3}^{1-(x_1=0)} \cdot \frac{1}{3}^{x_2=0}\ \frac{2}{3}^{1-(x_2=0)} \cdot \frac{1}{3}^{x_3=1}\ \frac{2}{3}^{1-(x_3=1)} = \frac{1}{3}^{1}\ \frac{2}{3}^{2}$

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4

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.

풀이 **X=1일 때 : 성공의 횟수가 1일 때**

- **세 가지 경우의 $\frac{1}{3}^{1}\frac{2}{3}^{2}$이 있는 것으로 $P(X=1)=3\cdot\frac{1}{3}^{1}\frac{2}{3}^{2}$임**

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4 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.

풀이 **X=2일 때 : 성공의 횟수가 2일 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 가짐
  - ① 첫 번째와 두 번째가 성공했을 때 : $x_1 = 1,\ x_2 = 1,\ x_3 = 0$
    - $\left(\frac{1}{3}\right)^{x_1=1}\left(\frac{2}{3}\right)^{1-(x_1=1)} \cdot \left(\frac{1}{3}\right)^{x_2=1}\left(\frac{2}{3}\right)^{1-(x_2=1)} \cdot \left(\frac{1}{3}\right)^{x_3=0}\left(\frac{2}{3}\right)^{1-(x_3=0)} = \left(\frac{1}{3}\right)^{2}\left(\frac{2}{3}\right)^{1}$

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4 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.

풀이 **X=2일 때 : 성공의 횟수가 2일 때**

② 두 번째와 세 번째가 성공했을 때 : $x_1 = 0$, $x_2 = 1$, $x_3 = 1$

- $\frac{1}{3}^{x_1=0}\frac{2}{3}^{1-(x_1=0)} \cdot \frac{1}{3}^{x_2=1}\frac{2}{3}^{1-(x_2=1)} \cdot \frac{1}{3}^{x_3=1}\frac{2}{3}^{1-(x_3=1)} = \frac{1}{3}^{2}\frac{2}{3}^{1}$

③ 첫 번째와 세 번째가 성공했을 때 : $x_1 = 1$, $x_2 = 0$, $x_3 = 1$

- $\frac{1}{3}^{x_1=1}\frac{2}{3}^{1-(x_1=1)} \cdot \frac{1}{3}^{x_2=0}\frac{2}{3}^{1-(x_2=0)} \cdot \frac{1}{3}^{x_3=1}\frac{2}{3}^{1-(x_3=1)} = \frac{1}{3}^{2}\frac{2}{3}^{1}$

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4

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.

풀이 **X=2일 때 : 성공의 횟수가 2일 때**

- X=1일 때와 마찬가지로 세 가지 경우가 있어, $P(X=2) = 3 \cdot \frac{1}{3}^{2}\frac{2}{3}^{1}$임

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4

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.

풀이 **X=3일 때 : 성공의 횟수가 3일 때**

- $x_1 = 1$, $x_2 = 1$, $x_3 = 1$인, 즉 모두 성공한 경우로 다음과 같은 한 가지 경우임
  - 즉, $\frac{1}{3}^{x_1=1}\ \frac{2}{3}^{1-(x_1=1)} \cdot \frac{1}{3}^{x_2=1}\ \frac{2}{3}^{1-(x_2=1)} \cdot \frac{1}{3}^{x_3=1}\ \frac{2}{3}^{1-(x_3=1)} = \frac{1}{3}^{3}\ \frac{2}{3}^{0}$

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4 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.

풀이 모두 모았을 경우

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

이항계수 = $\binom{n}{x}=\frac{n!}{(n-x)!x!}$

확률 $p^{x}(1-p)^{n-x}$

3) 이항분포

● 이항분포의 확률질량함수(Probability Mass Function)

Q4

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.

풀이

- **이상으로부터 확률변수 X가 이항분포를 따를 때의 확률질량함수는 다음과 같음**
  - $P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \dots, n$

04

실습

1) 실습

1) 실습

● 이항분포의 확률 계산

```
(dbinom(2, size = n, prob = p))
```

▶ # 확률변수 X = 2일 확률 P(X = 2)

```
(dbinom(4, size = n, prob = p))
```

▶ # 확률변수 X = 4일 확률 P(X = 4)

〈소스코드 : https://github.com/LEESUAJE1978/r_statistics/blob/master/lecture2.R〉

1) 실습

● 이항분포의 확률 계산

```
px <- dbinom(x, size = n, prob = p)
```

▶ # 확률변수(성공 횟수)별 확률

```
plot(x, px, type = 'h', xlab = '성공 횟수(x)', ylab
= "확률(P[X=x])", main = "B(6, 1/3)")
```

▶ # 성공 횟수와 각각의 확률 그래프로 보기

```
help("dbinom") #dbinom
```

▶ # dbinom 상세 내용 보기

〈소스코드 : https://github.com/LEESUAJE1978/r_statistics/blob/master/lecture2.R〉

1) 실습

● 누적질량함수

#pbinorm, probability, the cumulative distribution function
확률변수 X가 특정 값 이하일 확률 P(X<=2)을 구함

```
pbinom(2, size = n, prob = p)
```

▶ # 성공 횟수가 2 이하일 확률

```
pbinom(4, size = n, prob = p)
```

▶ # 성공 횟수가 4 이하일 확률

```
pbinom(4, size = n, prob = p) -
pbinom(2, size = n, prob = p)
```

▶ # 성공 횟수가 2 초과(3 이상) 4 이하인 확률

〈소스코드 : https://github.com/LEESUAJE1978/r_statistics/blob/master/lecture2.R〉

1) 실습

● 사분위수 구하기

#qbinom, quantile, the inverse c.d.f
사분위 수의 값 구하기

✓ C.D.F의 반대, 분위수에 해당하는 확률변수 X의 값 x를 구함

1) 실습

● 사분위수 구하기

```
qbinom(0.1, size = n, prob = p)
```

▶ # 확률변수 X의 값이 0일 경우 CDF가 0.09, 1일 경우 CDF가 0.35임, 0~0.09 미만은 0, 0.09부터 0.35미만까지는 1이 선택됨

```
qbinom(0.5, size = n, prob = p)
```

▶ # 확률변수 x의 값이 2일 경우 CDF가 0.68임

▶ # 0.35~0.68 미만은 확률변수의 값은 2가 선택됨

〈소스코드 : https://github.com/LEESUAJE1978/r_statistics/blob/master/lecture2.R〉

1) 실습

- 사분위수 구하기

#rbinom, 이항분포를 따르는 확률표본 추출

확률변수 X(성공 횟수)의 값을 임의로 추출함

```
rbinom(15, size = n, prob = p)
```

1) 실습

● 기댓값과 분산

```
(ex <- sum(px * x))
```

```
ex2 <- sum(x^2 * px)
```

▶ # 확률변수 각각의 값에 대한 제곱 값

```
(varx <- ex2 - ex^2)
```

▶ # 이항분포의 분산

〈소스코드 : https://github.com/LEESUAJE1978/r_statistics/blob/master/lecture2.R〉

1) 실습

● ggplot을 활용한 이항분포 그리기

```
require(ggplot2)
```

```
nn = 10000
```

▶ #실험 반복 횟수 또는 실험 관찰 횟수

```
pp = 1/10
```

▶ #성공 확률

```
ss = 10
```

▶ #시행횟수

〈소스코드 : https://github.com/LEESUAJE1978/r_statistics/blob/master/lecture2.R〉

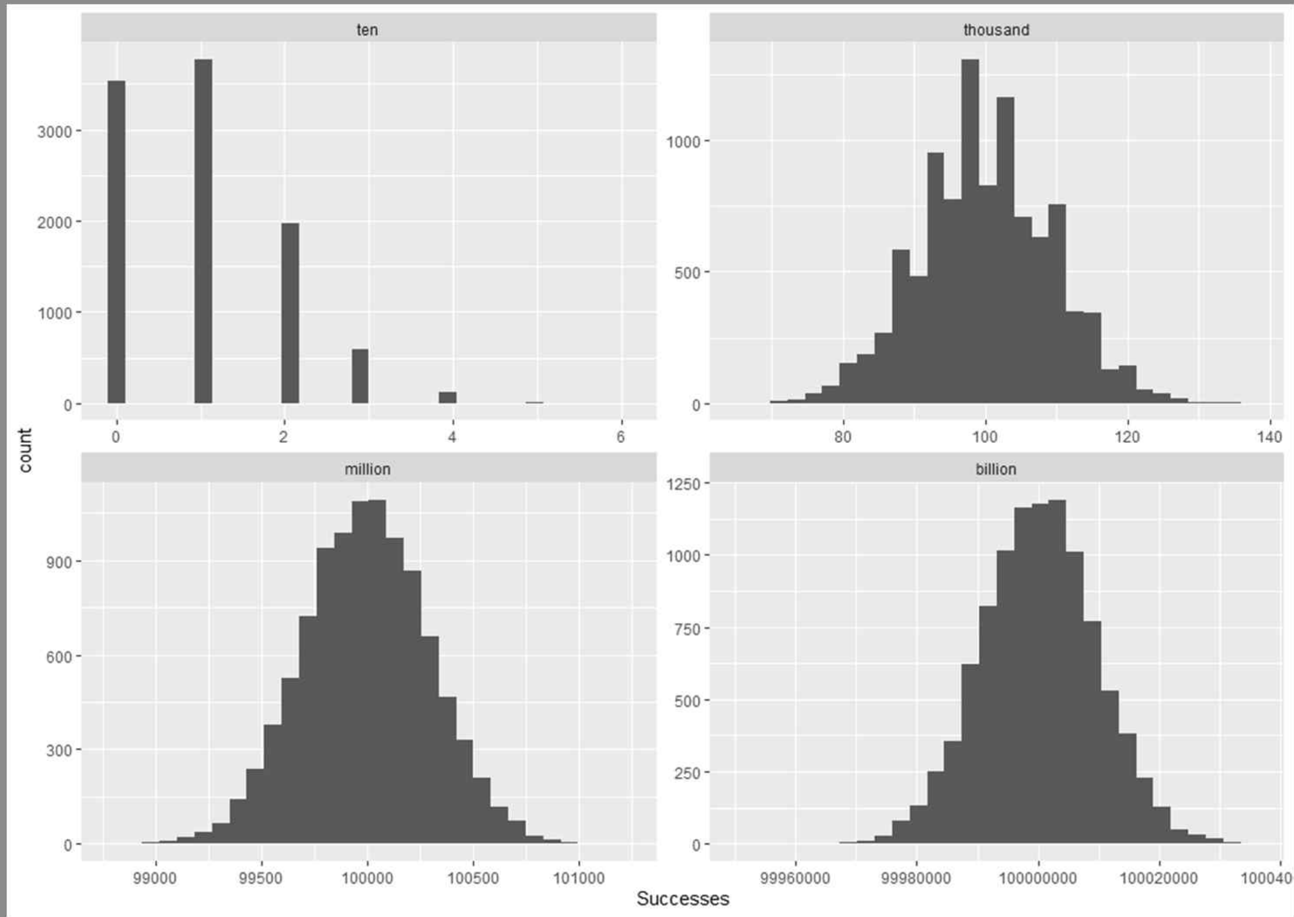
▷ ggplot을 활용한 이항분포 그리기

```
binomdata_ten <-data.frame(Successes = rbinom(n = nn, size = ss, prob = p), size ='ten')
binomdata_thousand <-data.frame(Successes = rbinom(n = nn, size = ss*100, prob = p), size ='thousand')
binomdata_million <-data.frame(Successes = rbinom(n = nn, size = ss*100000, prob = p), size = 'million')
binomdata_billion <-data.frame(Successes = rbinom(n = nn, size = ss*100000000, prob = p), size = 'billion')
binomall<- rbind(binomdata_ten,binomdata_thousand, binomdata_million, binomdata_billion)


ggplot2::ggplot(binomall, aes(x=Successes))+geom_histogram(bins=30)+
  facet_wrap(~size, scales = 'free')
```

〈소스코드 : https://github.com/LEESUAJE1978/r_statistics/blob/master/lecture2.R〉

▷ 중심 극한 정리에 의한 이항분포의 정규 분포화



⟨소스코드 : https://github.com/LEESUAJE1978/r_statistics/blob/master/lecture2.R⟩

학습 평가

Q1 Q2

Q1

표본 공간의 각 원소를 실숫값에 대응시키는 함수는 무엇인가?

1 표본 평균

2 분산

3 확률변수

4 확률분포

학습
평가

Q1 Q2

Q1

표본 공간의 각 원소를 실숫값에 대응시키는 함수는 무엇인가?

1 표본 평균

2 분산

3 확률변수

4 확률분포

정답 3번

해설 표본 평균은 추출된 표본 들의 평균 값이고 분산은 확률변수가 기댓값으로부터 얼마나 떨어진 곳에 분포하는지 가능하는 지표이고, 확률분포는 확률변수가 특정한 값을 가질 확률입니다.

학습 평가

Q1 Q2

Q2

은/는 임의의 결과가 ‘성공’ 또는 ‘실패와 같이 두 가지 중 하나인 실험이며,

은/는 이것을 여러 번 수행한 확률 분포의 모형이다.

Q1 Q2

학습 평가

Q2

베르누이 시행 은/는 임의의 결과가 '성공' 또는 '실패와 같이 두 가지 중 하나인 실험이며,
이항분포 은/는 이것을 여러 번 수행한 확률 분포의 모형이다.

정답

베르누이 시행, 이항분포

해설

베르누이 시행은 임의의 결과가 '성공' 또는 '실패와 같이 두 가지 중 하나인 실험이며, 이항분포는 베르누이 시행을 여러 번 수행한 확률 분포의 모형입니다.

정리하기

확률

✓ **확률이란**

– 확률(確率)

: 어떤 사건이 실제로 일어날 것인지 혹은 일어났는지에 대한 지식 혹은 믿음을 표현하는 방법이며 같은 원인에서 특정한 결과가 나타나는 비율

✓ **확률실험 조건**

– 결과를 구하기 위한 어떤 실험을 통해 나타나는 결과를 알지 못함

– 결과는 알지 못하지만 결과로 나타날 수 있는 가능한 경우를 알고 있음

– 동일한 실험을 몇 번이고 반복할 수 있음

정리
하기

확률

✓ **표본공간 : Ω (Omega)**

- 확률 실험으로부터 출현 가능한 모든 결과들의 모임
- 동전던지기의 경우 Ω ={앞면, 뒷면} = {H,T}

✓ **사건 : (Event)**

- 사건(Event) : 표본 공간의 각 원소 (출현 가능한 개별 결과)들의 부분집합
- 사건을 기호로 나타낼 때는 영어 대문자 (A, B, C, … 등)로 표현
- 근원사건(Elementary Event) : 어떤 사건이 표본 공간상의 하나의 원소로 구성된 사건

정리
하기

확률

✓ **확률변수(Random Variable)**

- 동전을 두 번 던져 앞면이 나오는 횟수처럼 표본공간의 각 원소를 실숫값에 대응시키는 함수

✓ **확률분포(Probability Distribution)**

- 확률 변수가 취할 수 있는 값과 발생할 확률을 대응한 관계

- 다음 시간에 살펴 볼 내용 -

09강 추론통계학

수고하셨습니다.