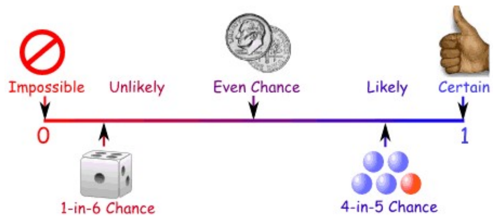
Chapter 2.   Probability review



사건의 확률

◆ : 확률적 실험

- 그 결과에 변이가 나타나는 현상을 관측하는 과정
  - 동전/주사위 던지기, 섞여진 카드 한 벌에서 하나의 카드 뽑기, 의견조사에서 많은 고객을 표본으로 추출하기, 생산 공정에서 제품의 품질조사…

◆ : S

- 실험의 서로 다른 모든 가능한 결과들(*outcomes*)의 모임(**set**, 집합)
- 각각의 결과 : 표본공간의 원소 **(element of the sample space)**
  - 표본공간의 기본결과 **(elementary outcome)**, 단순사건 **(simple event)**
  - 표현법 : $e_1, e_2, \ldots$

◆ : $A, B, \ldots$

- 표본공간의 부분집합**(subset)** : $A, B \subseteq S$
- 특정한 성질을 가지는 기본결과들의 집합

◆ 실험에서 사건 A 에 속하는 기본결과 중 어느 하나라도 발생하면 "사건 A 가 발생했다"고 한다.

◆ 예제 : 표본공간과 셀 수 있는 사건

❍ 토요일 오후에 계산하는 135명의 고객을 관찰하여, 카드(신용카드 또는 직불카드)로 계산한 고객의 수를 기록한다. (a) 표본공간을 표시하고 (b) 구매자의 50% 이상이 카드로 계산할 사건을 표시하라.

➢ 카드로 구매한 고객의 수는 **0, 1, 2, ..., 135** 중의 하나

$$S = \{0,\ 1,\ 2,\ \ldots,\ 135\}$$
$$S = \{e_0,\ e_1,\ e_2,\ \ldots,\ e_{135}\}$$

➢ 사건 A를 구매고객의 **50%** 이상이 카드로 계산할 사건이라고 하자.

✓ $0.5 \times 135 = 67.5$이므로, 사건 A를 다음과 같이 표시할 수 있다.

무한개의 원소를 가지는 표본공간의 예

◆ 실험 : 한 도박사가 카지노에서 첫 번째 잭팟이 나올 때까지 계속 슬롯머신의 핸들을 당기는 횟수

❍ $S = \{ 1, 2, 3, \ldots \}$: 무한개의 원소를 가짐

❍

◆ 이산 표본공간(***discrete*** sample space)

❍ 유한개의 원소를 가지거나 가산무한인 유형의 표본공간

◆ 연속 표본공간 (***continuous*** sample space)

❍ 일반적으로 실수(**real number**)값을 가지는 표본공간

❍ 예 : 휘발유를 가득 채운 자동차가 기름이 없어질 때까지 주행한 거리

➢ $S = \{ d : d \geq 0 \}$

- 사건(event) : 표본공간 S의 부분집합
- 근원사건(elementary event), 단순사건(simple event) : 오직 한 개의 원소로 이루어진 사건
- 사건 A가 일어남 $\Leftrightarrow$ 사건 A의 한 원소가 조사의 결과로 관측됨
- $A \cup B$: 합사건, $A \cap B$: 곱사건, A^c : 여사건
- 두 사건 A, B가 $\boxed{}$ $\Leftrightarrow A \cap B = \varnothing$
  (두 사건 A, B가 동시에 일어날 수 없음)

확률의 공리적 정의

- **Kolmogorov(1903 – 1987)**
  - 확률을 상대도수의 극한 개념으로 파악하여 공리적으로 정의
- 확률의 공리적 정의 (Axioms of Probability)
  - 다음을 만족할 때, **P(*A*)**를 사건 ***A***의 확률이라 함

$\boxed{}$

- 확률의 성질
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $P(\bar{A}) = 1 - P(A)$
  - $A \subseteq B$ 이면, $P(A) \leq P(B)$

$$
\begin{aligned}
P(A \cup B) &= P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) \\
P(A) &= P(A \cap B^c) + P(A \cap B) \\
P(B) &= P(A \cap B) + P(A^c \cap B)
\end{aligned}
$$

확률을 배정하는 방법

◆ 확률모형의 결정

❍ 표본공간의 모든 사건에 확률을 배정하는 것

➢ 올바른 확률모형이 되기 위해서는 확률의 성질 1, 2, 3을 만족해야 한다.

➢ 실제적인 의미를 갖기 위해서는 사건에 배정된 확률이 그 사건이 일어날 것이라고 기대되는 발생 횟수의 비율로서의 확률의 개념과 일치해야 한다.

◆

❍ 일어날 가능성이 동일한 기본결과들 (***equally likely*** elementary outcomes)

➢ 예 : 공정한 주사위를 1회 던져서 윗면에 나온 숫자를 기록하는 실험

$$S = \{e_1, e_2, e_3, e_4, e_5, e_6\}$$

$$P(e_1) = P(e_2) = \cdots = P(e_6) = \frac{1}{6}$$

➢ 4보다 큰 숫자가 나올 사건 $A = \{e_5, e_6\}$ ⇒ $P(A) = P(e_5) + P(e_6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

모든 기본결과가 나타날 가능성이 동일하게 모형화하면, 균일확률모형을 따른다. 만일 S 에 k개의 기본결과가 있으면, 각각에 확률 $1/k$을 배정한다.

m 개의 기본결과로 구성되어 있는 사건 A 의 확률은

$$P(A) = \frac{m}{k} = \frac{A\text{에 속하는 기본결과의 개수}}{S\text{ 에 속하는 기본결과의 개수}}$$

로 배정된다.

공정한 동전을 던지는 균일확률모형

공정한 동전을 두 번 던질 때 앞면이 정확히 한 번 나올 확률을 구하라.

예제 1에서 나열한 것과 같이 표본공간에는 4개의 기본결과가 포함되어 있다. 즉,

이다. 공정한 동전이라는 말은 S에 속하는 4개의 기본결과가 일어날 가능성이 모두 동일함을 의미한다. 따라서 이들 각각에 확률 $1/4$을 할당한다. 사건 A = [앞면이 한 번]는 2개의 기본결과, 즉 HT와 TH로 구성된다. 따라서 $P(A) = 2/4 = 0.5$이다.

오랜 시행에서의 상대돗수인 확률

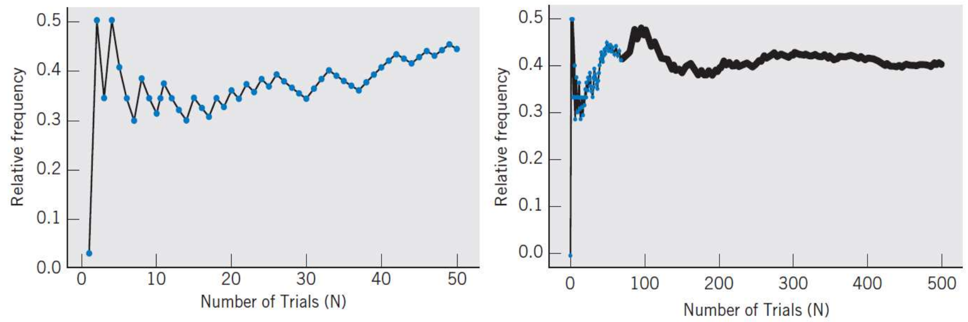
◆ 기본결과들이 동일한 가능성을 갖도록 표본공간을 구성하는 것이 불가능한 경우

- 주사위의 한쪽 모서리를 잘라내면, 각 면이 나타날 가능성이 동일하다는 가정은 불합리
- 한 남자가 30대에 죽을 확률을 이야기할 때, **기본결과**로써 10년 단위 또는 1년 단위의 사망 발생을 생각할 수 있음
  - 그러나 균일확률모형을 선호할 합당한 근거는 없음
  - 인구통계학자의 연구 결과 : 연령별로 죽음의 위험이 상당히 불균형함을 발견.

◆ 기본결과가 발생할 가능성이 동일하다고 가정할 수 없을 때, 사건의 확률을 어떻게 구할 것인가?

- 실험을 여러 번 반복해서 그 사건이 발생하는 비율을 관측.

상대돗수의 안정성



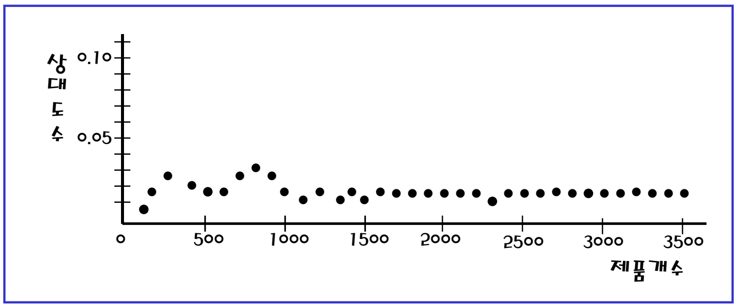
(*a*) Relative frequency versus number of trials. First 1–50. (*b*) Relative frequency versus number of trials. First 500 trials.

Probability as Long-Run Relative Frequency

We define $P(A)$, the probability of an event A, as the value to which the relative frequency stabilizes with increasing number of trials.

Although we will never know $P(A)$ exactly, it can be estimated accurately by repeating the experiment many times.

• 일정한 조건 아래에서 생산되는 불량품의 비율은 생산되는 제품의 수가 증가함에 따라 어떤 상수에 가까워지며, 만약 이 수를 안다면 이것을 불량품이 생산될 확률로 정의할 수 있음



< 불량품의 상대도수 변화 >

조건부 확률과 독립

◈ 한 사건이 일어났다는 조건 하에 다른 사건이 일어날 확률을 고려

- 사건 A가 일어났다면, A^c에 속하는 결과는 일어날 수 없음
- 따라서, 사건 A를 모든 가능한 결과의 집합, 즉 표본공간으로 간주할 수 있음

◈ 조건부 확률의 정의

- 사건 A가 주어졌을 때, 사건 B의 조건부 확률 :

B가 주어졌을 때 A의 조건부 확률은 $P(A|B)$로 나타내며

$$P(A|B) = \frac{P(AB)}{P(B)}$$

로 정의한다. 이 공식은

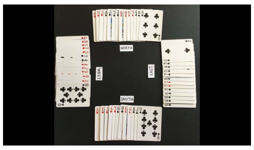
$$P(AB) = P(B)P(A|B)$$

와 동치이며 이 공식을 **확률의 곱셈법칙 (multiplication law of probability)** 이라고 한다.

◆ The theory of probability has always been associated with gambling.

- ❍ A fair coin gives you Heads (***H***) or Tails (***T***) with *equal* probability.
- ❍ A fair die will give you 1, 2, 3, 4, 5, or 6 with *equal* probability.
- ❍ A shuffled deck of cards means that any ordering of cards is *equally likely*.



◆ Example

- ❍ **(1)** Start with a shuffled deck of cards and distribute all 52 cards to 4 players, 13 cards to each.
  - ➢ *What is the probability that **each player gets an Ace**?*
- ❍ **(2)** Next, assume that you are a player and you get a single Ace.
  - ➢ *What is the probability now that **each player gets an Ace**?*

◆ *Answers.*

- ❍ **(1)** If any ordering of cards is equally likely, then any position of the four Aces in the deck is also equally likely.
  - ➢ There are [ ] possibilities for the positions (slots) for the 4 aces.
  - ➢ Out of these, the number of positions that give each player an Ace is [ ]:
    - ✓ pick the first slot among the cards that the first player gets, then the second slot among the second player's cards, then the third and the fourth slot.
  - ➢ Therefore, the answer is [ ] (**probability** that each player gets an Ace).
- ❍ **(2)** After you see that you have a single Ace, the probability goes up:
  - ➢ the previous answer needs to be divided by the probability that you get a single Ace, which is [ ] [ ]
  - ➢ The answer then becomes [ ] (**conditional probability)**

◈ *With additional information the probability must be adjusted.*

Example. Here is a question asked on Wall Street job interviews.

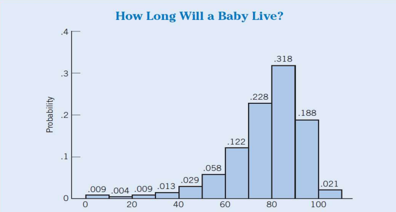
"Let's play a game of Russian roulette. You are tied to your chair. Here's a gun, a revolver. Here's the barrel of the gun, six chambers, all empty. Now watch me as I put *two* bullets into the barrel, into *two adjacent chambers*. I close the barrel and spin it. I put a gun to your head and pull the trigger. Click. Lucky you! Now I'm going to pull the trigger one more time. Which would you prefer: that I *spin the barrel first* or that I *just pull the trigger* ? "

- Assume that the barrel rotates clockwise after the hammer hits and is pulled back.
- You are given the choice between an ***unconditional*** and a ***conditional*** *probability of death*.
- The former, if the barrel is spun again, remains $\boxed{}$
- The latter, if the trigger is pulled without the extra spin, equals the probability that the hammer clicked on an empty slot, which is next to a bullet in the counterclockwise direction, and equals $\boxed{}$

◈ 예제 : 생존의 조건부 확률

a. 새로 태어나는 아기가 **90**세 이상 생존할 확률은 얼마인가?

b. **80**세가 된 사람이 **90**세 이상 생존할 확률은 얼마인가?



The probabilities for life length of a baby born in the United States. (Obtained from the *National Vital Statistics Reports* 54 [2006]).

(a) 90세 이상 생존할 사건을 A로 표기하자. 90−100세, 100세 초과 집단의 사망 확률을 더하여 사건 A의 확률을 구할 수 있다.

$$\boxed{}$$

(b) 80세 이상 생존할 사건을 B로 표기하면, 구하고자 하는 확률은 조건부 확률 $P(A|B)$이다. $AB = A$, $P(A) = 0.209$,

$$P(B) = 0.318 + 0.188 + 0.021 = 0.527$$

$$\boxed{}$$

독립사건

◆ 사건 A와 사건 B가 서로 독립(mutually independent)

- $P(B|A) = P(B)$
- 사건 A가 일어났다고 하더라도 사건 B가 일어날 **확률**에 아무런 영향을 미치지 않음

독립사건
사건 A와 사건 B가 독립이면

가 성립한다.

$$P(B) = P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \Rightarrow \quad P(A \cap B) = P(A)P(B)$$

$$P(A)P(B) = P(A \cap B) = P(A|B)P(B) \quad \Rightarrow \quad P(A|B) = P(A)$$

- 어떤 사건 B의 확률을 직접 계산하기 어려운 경우에

$$A_i \cap A_j = \varnothing \ (i \neq j) \text{ 이고, } A_1 \cup A_2 \cup \cdots \cup A_n = S$$

인 사건 $A_1, A_2, \cdots, A_n$으로 표본공간 S를 분할하면 편리

- 이 때에는

$$(B \cap A_i) \cap (B \cap A_j) = \varnothing \ (i \neq j) \text{ 이고,}$$

$$\begin{aligned} B &= B \cap S = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n) \end{aligned}$$

이므로, 사건 B의 확률을 다음과 같이 계산 가능

베이즈 정리 (Bayes' Theorem)

-
  - 두 사건 A와 B가 동시에 발생할 수 있고, 어느 사건을 관찰하기 전에 $P(B)$를 알 수 있다고 가정하면 $P(B^c)=1-P(B)$ 역시 알 수 있음
  - 이 두 확률을 사전확률(prior probabilities)이라 부름
  - 두 개의 조건부 확률 $P(A|B)$와 $P(A|B^c)$를 알고 있다면, A의 상태를 관찰할 때 B의 확률을 갱신할 수 있음
-
  - 사건 A 가 발생했다는 것을 알게 된 후에, 사건 B 의 갱신된 사후확률은 다음 조건부 확률로 구할 수 있음

Bayes' Theorem

The posterior probability of $\overline{B}$ is then $P(\overline{B}\,|\,A) = 1 - P(B\,|\,A)$

실험결과의 수치적 측면

- 실험을 확률로 모형화 하기 위한 2가지 기본 요소
  - : 모든 기본결과들의 모임
  - : 각 기본결과에 적절한 확률을 할당하는 것 (사건의 확률)
- 실험 결과의 분류
  - 질적 결과
    - 동전 2번 던지는 실험의 기본 결과 : **HH, HT, TH, TT**.
    - 백신 접종 후의 반응(메스꺼움) 정도 : 심하다, 조금 있다, 전혀 없다.
  - 수치적 척도의 측정값 결과
    - 어느 도시에서 하루에 발생하는 도난 건수, 아르바이트 학생의 시급, 대학 입학 성적, 키, 몸무게 등
  - 실험의 기본결과가 질적인 경우에도 수치를 구할 필요가 있는 경우
    - 새로운 백신의 부작용에 대한 정보를 얻기 위해 100명에게 접종
      - ✓ 백신 평가에 관련된 정보 : 세 개의 범주 —심한 메스꺼움, 약간의 메스꺼움, 메스꺼움이 없음 —에 속하는 반응자의 수
    - 새로운 조례에 대한 시민의 의견 조사 : 500명에서 찬성자의 수, 반대자의 수

확률변수 (random variable)

◆ 확률변수 (random variable)

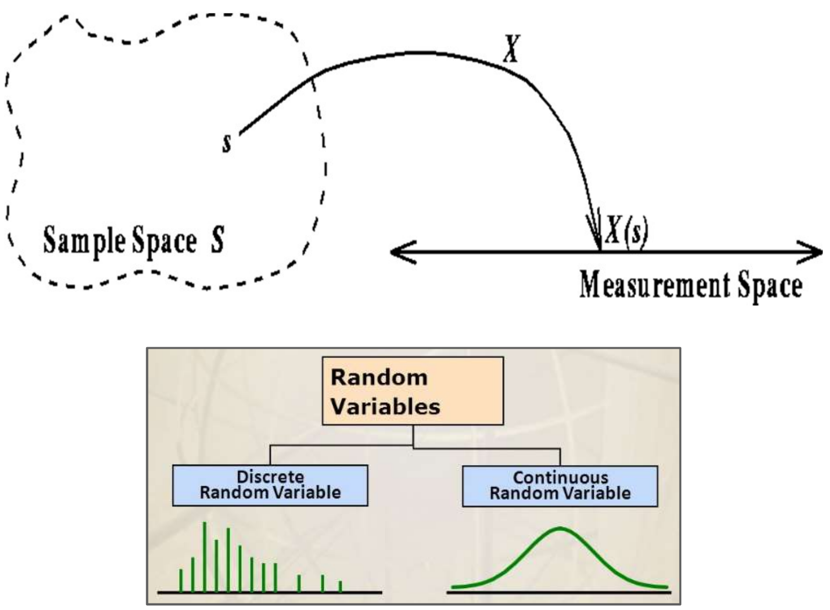
- 확률변수 X는 실험의 기본결과 각각에 결과의 특성을 나타내는 수치를 대응시키는 함수
- 확률변수 X는 표본공간에서 정의된 실숫값 함수
  - "확률"이라는 말은 실험의 결과 또는 이에 관련된 X의 값을 미리 알 수 없음을 암시

◆ 예제 : 확률변수로서 앞면의 개수

- 동전을 세 번 던져 나타난 앞면의 개수 X를 생각하자. X의 수치적 값과 대응되는 기본결과를 나열하라.

기본결과	X의 값
HHH	3
HHT	2
HTH	2
THH	2
HTT	1
THT	1
TTH	1
TTT	0

사건으로서의 X의 값	사건의 구성
$[X=0]$	$=\{$TTT$\}$
$[X=1]$	$=\{$HTT, THT, TTH$\}$
$[X=2]$	$=\{$HHT, HTH, THH$\}$
$[X=3]$	$=\{$HHH$\}$



유한 최댓값을 갖는 개수를 나타내는 확률변수

100마일 자동차 경주에 50대의 자동차가 참여했다. X를 실제로 경주를 마친 자동차의 수라고 하면, X는 0, 1, ..., 50의 값을 취할 수 있다.

상한이 없는 개수를 나타내는 확률변수

어떤 학생은 일주일에 한 번 복권 하나를 구입한다. X를 최소한 $1000짜리에 당첨될 때까지 구매한 복권의 수라고 하자. X의 가능한 값은 1, 2, 3, ...으로 무한히 계속된다.

◈ 이산 (discrete) 확률변수

- 유한개의 값을 갖거나 *수열로 나열할 수 있는 무한개*의 값을 갖는 확률변수
- ***At most countable***

◈ 연속(continuous)확률변수

- 어느 구간의 어떠한 실수값이라도 가질 수 있는 확률변수
- 연속적인 척도는 추상적 의미 : 측정 도구의 정확도는 한계가 있음
  - 성인 남자의 키, 젖소의 우유 생산량, 심장 발작 환자의 생존기간 등

◈ X의 확률분포(probability distribution)

- 확률변수는 표본공간 위의 확률을 수의 집합 위의 확률에 대응시켜 줌
- 확률변수 X의 값에 따라 확률이 어떻게 흩어져 있는지를 합이 1인 양수로 나타낸 것을 X의 확률분포 또는 간단히 분포라고 함

  - 균일한 동전을 1회 던질 때 앞면 개수 X(확률변수)의 확률분포

X	0	1
P(X=x)	1/2	1/2

◈ 이산확률변수(random variable of discrete type)

- 확률변수 X가 취할 수 있는 모든 값을 $x_1, x_2, \ldots$ 로 셀 수 있을 때, X를 이산확률변수(discrete random variable)라고 함
- $p(x)$: X의 $\boxed{}$
  - $p(x) = P(X = x_i)$, $x = x_i$ 일 때(i=1,2, …)
    - ✓ $p(x) = 0$, x가 기타의 값일 때
  - $0 \le p(x) \le 1$, $\sum_{\text{모든 } x} p(x) = 1$
  - $P(a < X \le b) = \sum_{a < X \le b} p(x)$

- X를 공정한 동전을 세 번 던져 나타난 앞면의 수라 할 때, X 의 확률분포를 구하라.

X의 값	확률
0	1/8
1	3/8
2	3/8
3	1/8
합 계	1

이산확률분포의 형태

x의 값	확률 $f(x)$
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_k	$f(x_k)$
합계	1

이산확률변수 X의 **확률분포** (probability distribution)는 함수

$$f(x_i) = P[X = x_i]$$

로 나타내며, 다음을 만족한다.

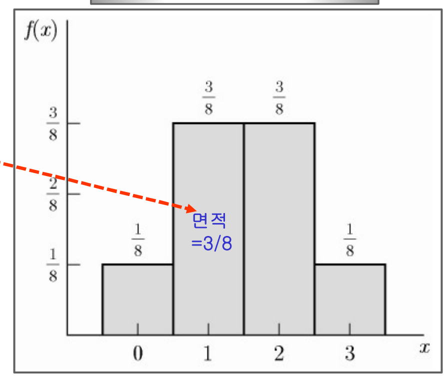
1. X의 모든 값 x_i에 대해, $0 \le f(x_i) \le 1$

2. $\sum_{i=1}^{k} f(x_i) = 1$

- X 를 공정한 동전을 세 번 던져 나타난 앞면의 수라 할 때, X 의 확률분포

X의 값	확률
0	1/8
1	3/8
2	3/8
3	1/8
합 계	1

X 의 확률 히스토그램 (probability histogram)



확률분포의 기댓값(평균)과 표준편차

◆ 데이터 집합의 평균 계산

- 주사위를 20번 던져 얻어진 데이터(표본)의 평균(표본평균)

4,	3,	4,	2,	5,	1,	6,	6,	5,	2
2,	6,	5,	4,	6,	2,	1,	6,	2,	4

$$\bar{x} = \frac{\text{Sum of the observations}}{\text{Sample size}} = \frac{76}{20} = 3.8$$

$$\bar{x} = 1\left(\frac{2}{20}\right) + 2\left(\frac{5}{20}\right) + 3\left(\frac{1}{20}\right) + 4\left(\frac{4}{20}\right) + 5\left(\frac{3}{20}\right) + 6\left(\frac{5}{20}\right) = 3.8$$

$$\boxed{\text{Sample mean } \bar{x} = \sum (\text{Value} \times \text{Relative frequency})}$$

$$\text{표본평균} \quad \bar{x} = \sum(\text{값} \times \text{상대도수})$$

◆ 주사위를 매우 많이 던진다고 하면…

- 상대도수는 공정한 주사위의 경우 모두 확률 1/6에 접근할 것
- 공정한 주사위를 무한 번 던질 때의 평균값 :

$$1 \times \left(\frac{1}{6}\right) + 2 \times \left(\frac{1}{6}\right) + \cdots + 6 \times \left(\frac{1}{6}\right) = \sum(\text{값} \times \text{확률}) = 3.5$$

◆ 확률변수 X의 평균/기댓값 (X의 확률분포에 대한 평균/기댓값)

- **확률변수 X의 평균 = X의 기댓값 (expected value)**
- 확률분포의 평균을 확률변수 X의 모평균이라고도 함

The mean of X or population mean

$$E(X) = \mu$$
$$= \sum (\text{Value} \times \text{Probability}) = \sum x_i f(x_i)$$

Here the sum extends over all the distinct values x_i of X.

◆ 동전을 세 번 던져 나오는 앞면의 수 X의 평균을 구하라.

x	$f(x)$	$xf(x)$
0	$\frac{1}{8}$	0
1	$\frac{3}{8}$	$\frac{3}{8}$
2	$\frac{3}{8}$	$\frac{6}{8}$
3	$\frac{1}{8}$	$\frac{3}{8}$
Total	1	$\frac{12}{8} = 1.5 = \mu$

◆ 확률분포의 평균 $\mu =$ $\boxed{}$

❍ 물리적 의미로 μ는 히스토그램의 무게 중심

◆ X를 게임이나 복권구입에서 획득한 돈이라고 하면,

❍ $E[X]$: 기대이익 또는 평균이익

◆ 통계학에서는 "기댓값"과 "평균" 둘 다 널리 사용됨

X의 분산과 표준편차

Variance of X = $\sum$ (Deviation)^2 × (Probability)
$= \sum (x_i - \mu)^2 f(x_i)$

X의 분산은 $\mathrm{Var}(X)$ 또는 σ^2으로 나타낸다. X의 표준편차 (standard deviation) 는 분산의 양의 제곱근이고 $\mathrm{sd}(X)$ 또는 σ로 나타낸다.

X의 분산 σ^2을 모분산 (population variance) 이라 하고 X의 표준편차 σ를 모표준편차 (population standard deviation) 라고 한다.

X의 분산과 표준편차

$$\sigma^2 = \mathrm{Var}(X) = \sum (x_i - \mu)^2 f(x_i)$$
$$\sigma = \mathrm{sd}(X) = +\sqrt{\mathrm{Var}(X)}$$

Variance and Standard Deviation of X

$$\sigma^2 = \mathrm{Var}(X) = \sum (x_i - \mu)^2 f(x_i)$$
$$\sigma = \mathrm{sd}(X) = +\sqrt{\mathrm{Var}(X)}$$

베르누이분포

◆ 베르누이 시행(Bernoulli trial)

- 어떤 실험이 두 가지 가능한 결과만을 가질 경우
  - 표본공간 : **S = { s, f }**
  - 성공 확률을 p=**P(s)**, 실패 확률을 q=**P(f)**로 나타내고, $0 \le p \le 1$, $q=1-p$.

◆ 베르누이 확률변수(Bernoulli random variable)

- 베르누이 시행의 결과에 따라서 **0** 또는 **1**의 값을 대응시키는 확률변수
  - S={s, f}에서 $X(s) = 1$, $X(f) = 0$ 인 확률변수 X
- 베르누이 분포 : 베르누이 확률변수의 확률분포

x	0	1
$p(x)$	$1-p$	p

< "1"의 확률이 p인 베르누이 분포 >

이항분포 (Binomial distribution)

◆ 오직 두 가지 가능한 결과가 얻어지는 실험을 반복할 때 나타나는 우연적인 변동을 모형화한 기본 분포

-
- 관심 확률변수 X는 정해진 수만큼의 반복실험에서 한 범주에 속하는 결과의 도수

확률모형(probability model)은 확률변수 X에 대한 우연적 현상을 기술하는 확률분포의 가정된 형태이다. 확률은 **모수**(parameter)라고 하는 모집단과 관련된 어떤 수로 표현된다.

이항분포

다음 기호를 정의하자.

n : 베르누이 시행의 고정된 횟수
p : 각 시행에서 성공할 확률
X : n번의 시행에서 성공의 횟수

확률변수 X를 **이항확률변수**(binomial random variable)라 한다. 그것의 분포를 **이항분포**(binomial distribution)라 한다.

The **binomial distribution** with n trials and success probability p is described by the function

for the possible values $x = 0, 1, \ldots, n$.

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

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$
0	1							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	
7	1	7	21	35	35	21	7	1

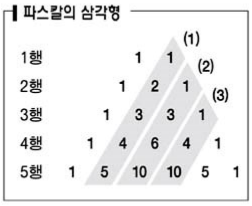
1
1 1
1 2 1
1 3 3 1
⋮

$(a+b)^1 = a+b$

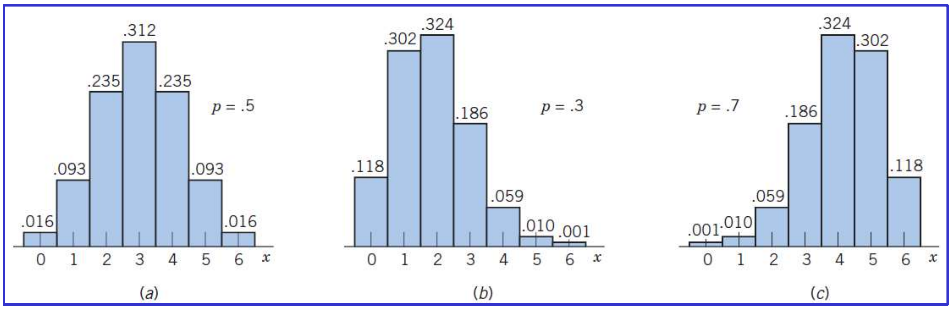
$(a+b)^2 = a^2+2ab+b^2$

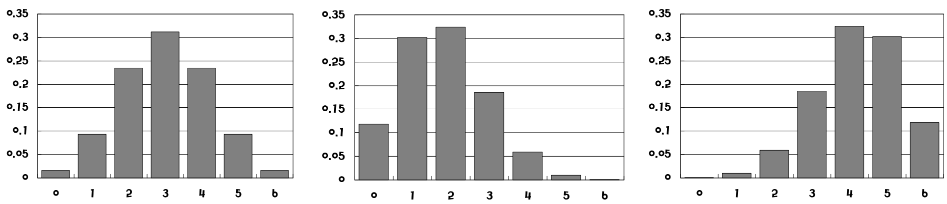
$(a+b)^3 = a^3+3a^2b+3ab^2+b^3$

⋮

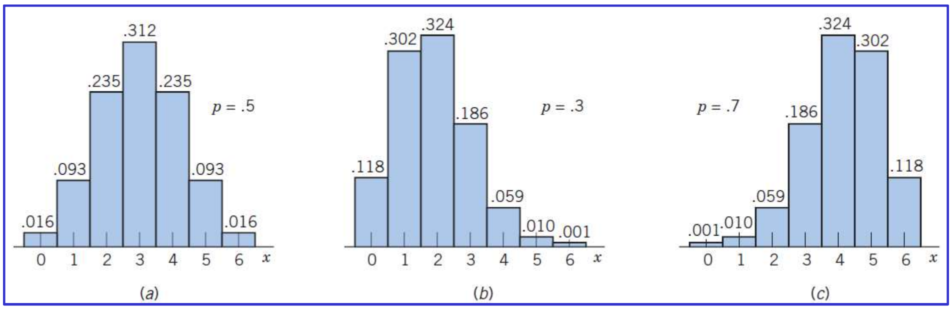


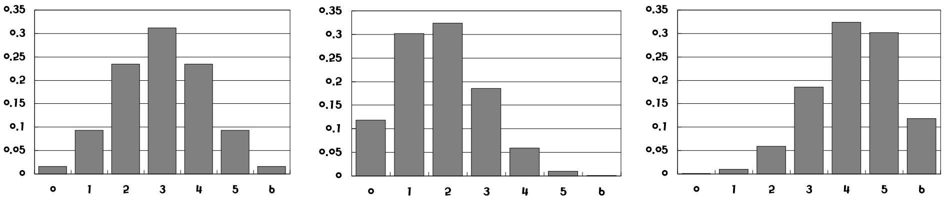
이항분포 B(n=6, p)의 확률분포





이항분포 **B(*n*=6, *p*)**의 확률분포





이항분포의 평균과 표준편차

- 확률변수 $X_1, X_2, \cdots, X_n$이 서로 독립이고, "1"의 확률이 p인 베르누이분포 즉, $B(1,p)$를 따른다면,

$$X = X_1 + X_2 + \cdots + X_n \sim B(n,p)$$

- 이 성질을 이용하면 이항분포의 평균과 분산은 쉽게 유도 가능. 즉,

$$E[X] = E[X_1] + \cdots + E[X_n] = np$$

$$Var(X) = Var(X_1) + \cdots + Var(X_n) = np(1-p)$$

◈ 이항분포 **B(*n*,*p*)**의 평균과 분산 : ***X* ~ B(*n*,*p*)**

- **E[*X*] = *np***
- **Var(*X*) = *np*(1-*p*) = *npq* , *q*=1-*p***

공분산과 상관계수

◈ 두 확률변수 X와 Y의 상관계수(correlation coefficient), 공분산(covariance)

(1) $Cov(X,Y) = E[(X-\mu_1)(Y-\mu_2)]$

(2) $Corr(X,Y) = E\left[\left(\frac{X-\mu_1}{\sigma_1}\right)\left(\frac{Y-\mu_2}{\sigma_2}\right)\right] = \frac{Cov(X,Y)}{sd(X)sd(Y)}$

$$-1 \le Corr(X,Y) \le 1$$

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_1)(Y-\mu_2)] \\ &= E[XY - \mu_2 X - \mu_1 Y + \mu_1\mu_2] \\ &= E[XY] - \mu_2 E[X] - \mu_1 E[Y] + \mu_1\mu_2 \\ &= E[XY] - \mu_1\mu_2 \end{aligned}$$

◈ 공분산과 상관계수의 성질

◈ 두 확률변수의 합의 분산

(1) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

(2) $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

◈ 두 확률변수가 서로 독립인 경우

(1) $E[XY] = E[X]\cdot E[Y]$

(2) $Cov(X,Y) = 0\,,\ \ Corr(X,Y) = 0$

(2) $Var(X \pm Y) = Var(X) + Var(Y)$

$$\begin{aligned} E[XY] &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j p(x_i, y_j) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j p_1(x_i) p_2(y_j) \\ &= \sum_{i=1}^{\infty} x_i p(x_i) \sum_{j=1}^{\infty} y_j p_2(y_j) \\ &= E[X]\cdot E[Y] \end{aligned}$$

$Cov(X,Y) = E[XY] - E[X]\cdot E[Y] = 0$

연속확률변수의 확률모형

◆ 연속확률변수의 확률분포

- 한 구간에 있는 임의의 실수값을 취할 수 있는 확률변수
  - 몸무게, 힘, 수명, 온도와 같이 연속적인 척도에 의해 측정된 변수들
- 상대도수 히스토그램의 극한 형태로 표현 가능

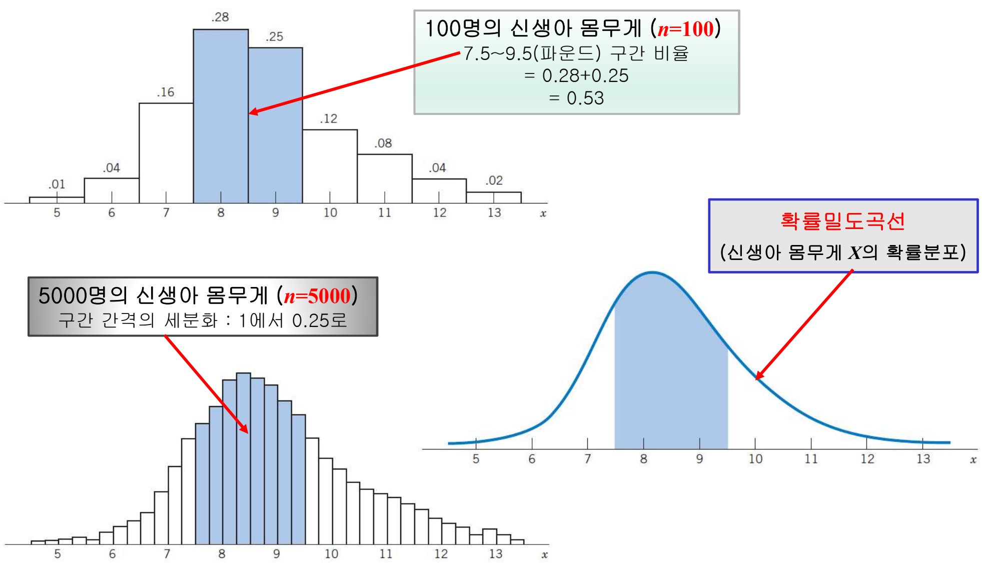
◆ 상대도수 히스토그램의 두 가지 중요한 성질

1. 히스토그램으로 나타나는 전체 면적은 1이다.
2. 어떤 계급의 경계점인 두 점 a와 b 사이의 구간에 포함된 측정값들의 상대도수는 이 구간의 히스토그램의 면적이다.

◆ 상대도수 히스토그램의 극한 형태 :

- 측정대상의 수를 증가시키고 계급간격을 줄임으로써 더욱 상세한 분포상태를 나타내는 히스토그램을 얻을 수 있음
- 히스토그램을 세분화하는 과정에 있어서 잇따른 직사각형들 간의 면적의 차이들은 점점 줄어들고 히스토그램의 형태는 완만한 곡선의 모양에 가까워짐

확률밀도곡선 : 상대도수 히스토그램의 극한 형태



연속확률변수 X의 확률밀도함수

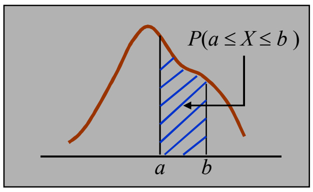
확률밀도함수 $f(x)$는 연속확률변수에 대한 확률의 분포를 나타낸다. 이 확률밀도함수는 다음 성질을 만족한다.

1. 확률밀도곡선 아래의 전체 면적은 1이다.
2. $P[a \le X \le b]$는 a와 b 사이의 구간에서 확률밀도곡선 아래의 면적이다.
3. 모든 x에 대해 $f(x) \ge 0$이다.

연속확률변수 X에 대해, $X = x$일 확률은 **항상 0**이다. 그러므로 X가 어떤 구간에 포함될 확률에 대해 말하는 것만이 의미가 있다.

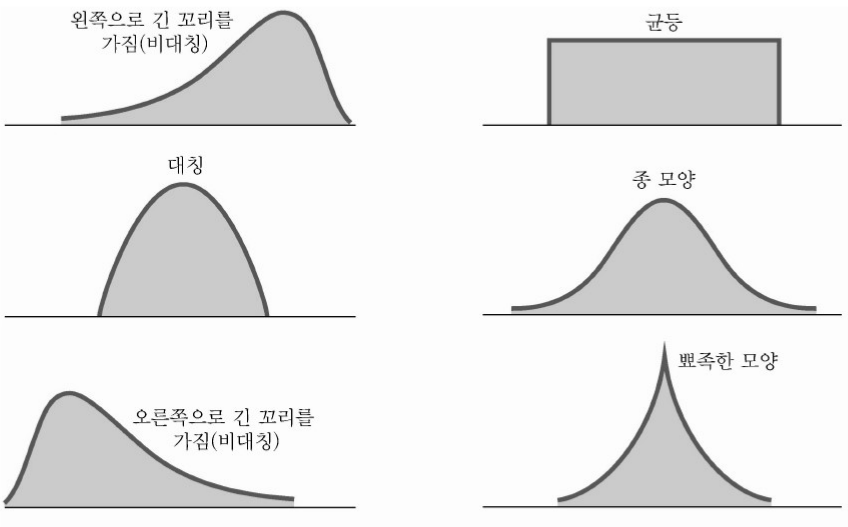
연속확률변수 X 에 대하여 함수 $p(x)$ 가

를 만족하면, 함수 $p(x)$를 X 의 확률밀도함수라고 함



- ✓ 임의의 c에 대해서 $P(X = c) = 0$
- ✓ $P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b)$

연속확률변수의 확률밀도곡선



연속확률변수의 기댓값

◆ 확률변수 X 의 기대값(expectation) 또는 평균(mean)

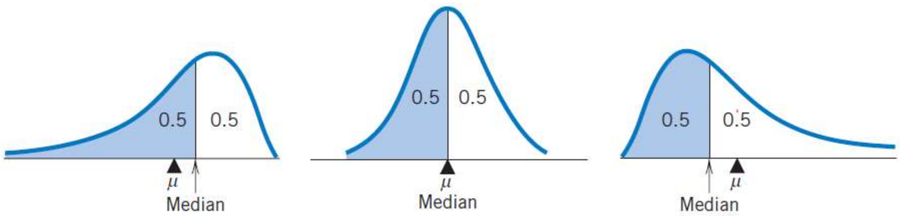
❍ 이산확률변수 : $E[X] = \sum_{i=1}^{\infty} x_i \, p(x_i)$

❍ 연속확률변수 : $E[X] = \int_{-\infty}^{\infty} x \, p(x) \, dx$

◆ 연속확률변수의 기댓값(평균)과 중앙값

❍ $\boxed{}$: 확률질량의 균형점(무게중심)

❍ $\boxed{}$: 확률밀도함수 면적의 이등분점



◆ 정규분포(normal distribution)

❍ 가우스에 의해서 제시된 분포로서 일명 가우스분포(Gaussian distribution)로 불림

❍ 물리학 실험 등에서 오차에 대한 확률분포를 연구하는 과정에서 발견

❍ 초기의 통계학자들은 모든 자료의 히스토그램이 정규분포의 형태와 유사하지 않으면 비정상적인 자료로 생각하여 이 분포에 "정규(normal)"라는 이름이 붙게 됨

- 평균이 μ이고, 표준편차가 σ인 정규분포의 확률밀도함수

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

◆ 정규분포의 확률밀도 곡선

❍ 모평균 μ 를 중심으로 **좌우 대칭**인 그래프

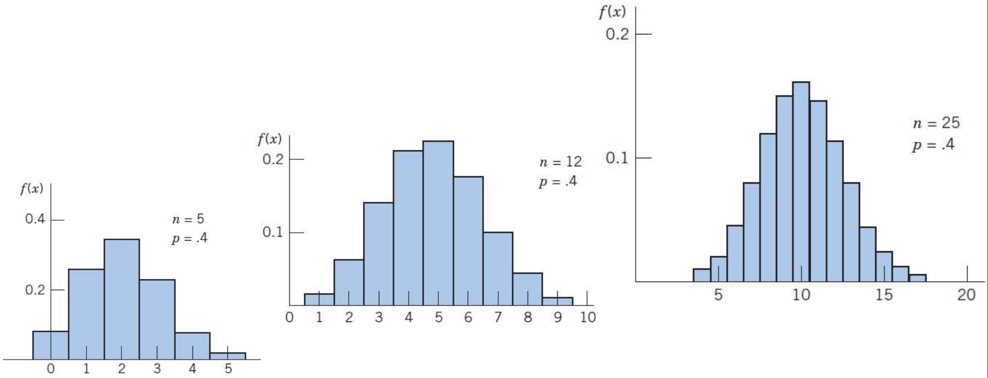
❍ 대칭점에서 그 높이가 가장 높고, 대칭점에서 멀어질수록 점점 낮아짐

❍ 변곡점과 대칭점 사이의 거리가 바로 **모표준편차** σ

이항분포의 정규근사

◆ 이항분포

- ❍ 어떤 실험을 n번 독립적으로 시행하여 얻어지는 성공의 횟수인 X의 분포
- ❍ 성공 확률 p가 **0이나 1에 너무 가깝지 않고, 시행횟수가 크면…**
  - ➢ **정규분포를 이항확률의 좋은 근사 분포로 활용 가능**
  - ➢ 중심극한정리로 증명 가능



확률표본(랜덤표본, random sample)

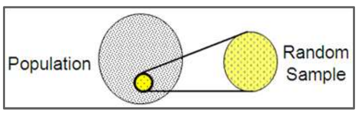
만일 $X_1, X_2, \ldots, X_n$이 독립적으로 추출되고 모두 모집단과 같은 분포를 갖는다면, $X_1, X_2, \ldots, X_n$은 **모집단 분포로부터의 크기** n **인 확률표본** (random sample of size n from the population distribution)이라 한다.

◆ 모집단이 매우 크고 표본크기가 상대적으로 작은 경우

- ❍ **복원추출**과 **비복원추출**에 관한 확률값에 별 차이가 없음
- ❍ 표본추출을 복원추출로 간주하면, 각 표본의 독립성과 동일분포성 확보
  - ➢

◆ $X_1, \ldots, X_n$: 확률표본(random sample, 랜덤표본)

- ❍ 독립성과 동일분포성(**모집단 분포와 동일**)을 가지는 표본
  - ➢ $X_1, \ldots, X_n$: 서로 독립
  - ➢ 모든 X_i 의 분포는 모집단 분포와 동일



표본평균의 분포와 중심극한정리

Population mean $= \mu$

Population standard deviation $= \sigma$

$X_1, \dots, X_n$: 확률표본(random sample, 랜덤표본)

(i) $X_1, \dots, X_n$: 서로 독립

(ii) 모든 X_i 의 분포는 모집단 분포와 동일 $(1 \le i \le n)$

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

$$E(X_1) = \cdots = E(X_n) = \mu$$
$$\mathrm{Var}(X_1) = \cdots = \mathrm{Var}(X_n) = \sigma^2$$

Mean and Standard Deviation of $\overline{X}$

The distribution of the sample mean, based on a random sample of size n, has

$$E(\overline{X}) = \mu \qquad (= \text{Population mean})$$
$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \qquad \left(= \frac{\text{Population variance}}{\text{Sample size}}\right)$$
$$\mathrm{sd}(\overline{X}) = \frac{\sigma}{\sqrt{n}} \qquad \left(= \frac{\text{Population standard deviation}}{\sqrt{\text{Sample size}}}\right)$$

중심극한정리(Central Limit Theorem)

- 정규모집단이 아닌 모집단에서 표본을 추출하면…
  - $\overline{X}$ 의 분포는 모집단 분포의 특정한 모양에 의존
- 중심극한정리로 알려진 놀라운 결과는…
  - 표본의 크기 n 이 크면 모집단 분포의 모양에 상관없이 표본평균 $\overline{X}$ 의 분포는 근사적으로 정규분포를 따른다는 것
    - ✓ 실제로 n 이 30보다 크면 보통 정규근사가 적절

중심극한정리

모집단에 관계없이 n이 크면, $\overline{X}$의 분포는 근사적으로 정규분포를 따른다.

평균이 μ이고 표준편차가 σ인 임의의 모집단에서 확률표본을 추출하면, n이 클 때 $\overline{X}$의 분포는 평균이 μ이고 표준편차가 $\sigma / \sqrt{n}$ 인 정규분포를 근사적으로 따른다. 따라서

이다.

Central Limit Theorem

Whatever the population, the distribution of $\overline{X}$ is approximately normal when n is large.

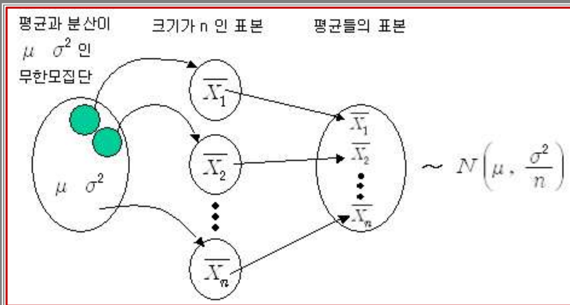
In random sampling from an arbitrary population with mean μ and standard deviation σ, when n is large, the distribution of $\overline{X}$ is approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$. Consequently,

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \quad \text{is approximately } N(0, 1)$$

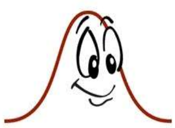
Central Limit Theorem: Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2 < \infty$. Define

$$U_n = \frac{\sum_{i=1}^{n} Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \quad \text{where } \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Then the distribution function of U_n converges to the standard normal distribution function as $n \to \infty$. That is,

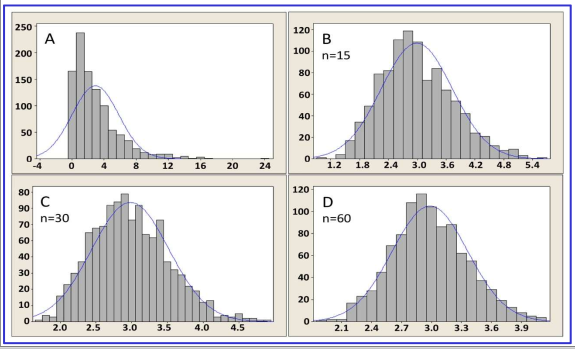


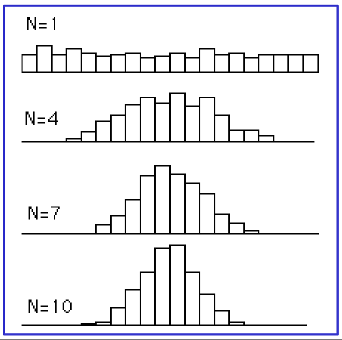
Central Limit Theorem



The distribution of the average of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution.

표본평균 $\overline{X}$ 의 극한분포는 정규분포





표본평균의 분포 & 중심극한정리

◈ 모집단의 특성값이 "0" 또는 "1"로 이원적인 경우

- 모평균은 모비율이 되고, 표본평균은 표본비율이 됨. 표본비율의 표본분포는 앞에서 다루었음.
- 여기에서는 일반적인 모집단에서 표본평균의 표본분포에 대한 몇 가지 기본적인 사

- $X_1, X_2, \cdots, X_n$이 모평균 μ, 모분산 σ^2인 무한모집단으로부터의 랜덤표본인 경우, 표본 평균

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

의 기대값과 분산은 각각 $\bar{X}$의 표본분포에서 중심 위치와 산포의 크기를 나타냄

- $\bar{X}$의 기대값과 분산을 구해보면, $E[X_i] = \mu$, $Var(X_i) = \sigma^2$ 이고, $X_1, X_2, \cdots, X_n$이 서로 독립이므로

$$E[\bar{X}] = \frac{1}{n}E[X_1 + \cdots + X_n] = \frac{1}{n}\{E[X_1] + \cdots + E[X_n]\} = \mu \ ,$$

$$Var(\bar{X}) = \frac{1}{n^2}Var(X_1 + \cdots + X_n) = \frac{1}{n^2}\{Var(X_1) + \cdots + Var(X_n)\} = \frac{\sigma^2}{n} \ .$$

- 모평균이 μ, 모분산이 σ^2인 무한모집단에서 크기 n인 랜덤표본의 표본평균 $\bar{X}$에 대하여 다음이 성립

$$E[\bar{X}] = \mu \ , \quad Var(\bar{X}) = \frac{\sigma^2}{n} \ , \quad sd(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

- 모평균이 μ, 모분산이 σ^2, 크기가 N인 유한모집단의 경우, 다음이 성립

$$E[\bar{X}] = \mu \ , \quad Var(\bar{X}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n} \ .$$

- 위의 결과로부터 표본평균 $\bar{X}$의 평균은 모집단의 평균과 같고, 표본의 크기 n이 클수록 그 분산이 0에 가까워져서, 결국 표본의 크기가 클 때, $\bar{X}$는 모집단의 평균인 μ 근처에 밀집되어 분포한다는 사실을 알 수 있음

- 특히, 모집단의 분포가 정규분포 $N(\mu, \sigma^2)$인 경우에는 정규분포의 성질로부터 $\bar{X}$의 분포가 정규분포임을 알 수 있음. 즉,

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

- 모집단의 분포가 정규분포 $N(\mu, \sigma^2)$일 때, 표본평균 $\bar{X}$는 정규분포 $N(\mu, \sigma^2/n)$을 따른다.
- 모집단의 분포가 정규분포가 아닌 경우 위의 사실은 성립하지 않음
- 그러나 표본의 크기 n이 충분히 클 때에는 임의의 모집단으로부터의 표본평균이라 하더라도 그 분포가 정규분포에 가깝다는 사실이 알려져 있음
- 중심극한정리(central limit theorem) : 평균이 μ이고, 분산이 σ^2인 임의의 무한 모집단에서 표본의 크기 n이 충분히 크면, 랜덤표본의 표본평균 $\bar{X}$는 근사적으로 정규분포 $N(\mu, \sigma^2)$을 따른다. 즉, 다음이 성립한다.

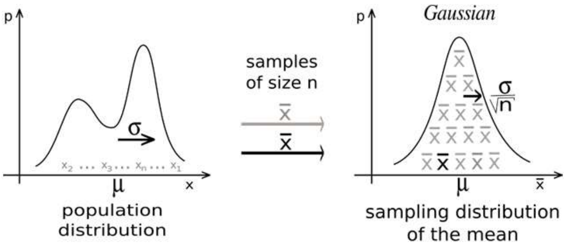
$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1)$$

- 유한모집단인 경우에도 N과 n이 충분히 크면 다음이 성립

$$\frac{\bar{X}-\mu}{\sigma\sqrt{\frac{N-n}{n(N-1)}}} \sim N(0,1)$$

중심극한정리 (Central Limit Theorem)

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables each having mean μ and variance σ^2. Then the distribution of $\frac{X_1+\ldots+X_n-n\mu}{\sigma\sqrt{n}}$ tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,



Law of Large Numbers

◆ **Theorem** (*Weak* Law of Large Numbers)

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$ and variance σ^2. Then, for any $\epsilon > 0$,

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \qquad \text{as} \qquad n \to \infty$$

Theorem (Strong Law of Large Numbers)

Let $X_1, X_2, \ldots$ be iid random variables with a finite first moment, $\mathbb{E}X_i = \mu$. Then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu$$

Modes of convergence

1. if for all $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$.
2. $X_n \to X$ *in distribution*, if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ as $n \to \infty$, for all $x \in \mathbb{R}$ at which $x \mapsto \mathbb{P}(X \leq x)$ is continuous.
3. $X_n \to X$ *in* L^1, if $\mathbb{E}(|X_n|) < \infty$ for all $n \geq 1$ and $\mathbb{E}(|X_n - X|) \to 0$ as $n \to \infty$.
4. if $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$.

$$\begin{array}{ccc}
X_n \to X \textit{ almost surely} & & \\
\Downarrow & & \\
X_n \to X \textit{ in probability} & \Rightarrow & X_n \to X \textit{ in distribution} \\
\Uparrow & & \\
X_n \to X \textit{ in } L^1 & \Rightarrow & \mathbb{E}(X_n) \to \mathbb{E}(X)
\end{array}$$

Summary

◆ 사건의 확률

- 확률적 실험, 표본공간, 사건

◆ 확률의 공리

- Kolmogorov's Axiom, 극한 개념, 상대돗수
- 조건부 확률과 독립성, Bayes' theorem

◆ 확률변수

- 이산확률변수, 연속확률변수

◆ 확률의 극한정리

- 중심극한정리(C.L.T.)
- 대수정리(Laws of large numbers)