

통계학의 개요

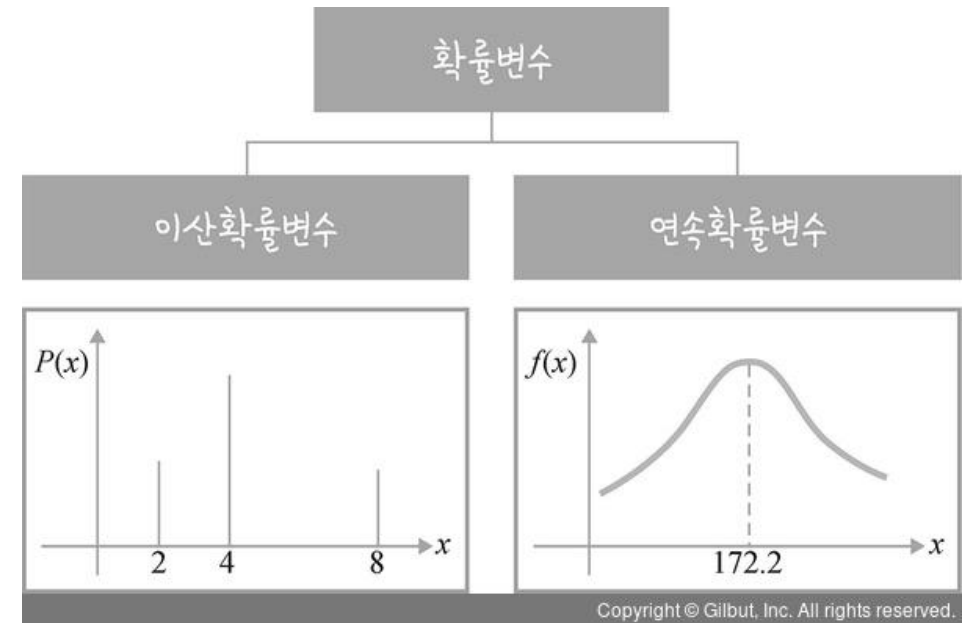
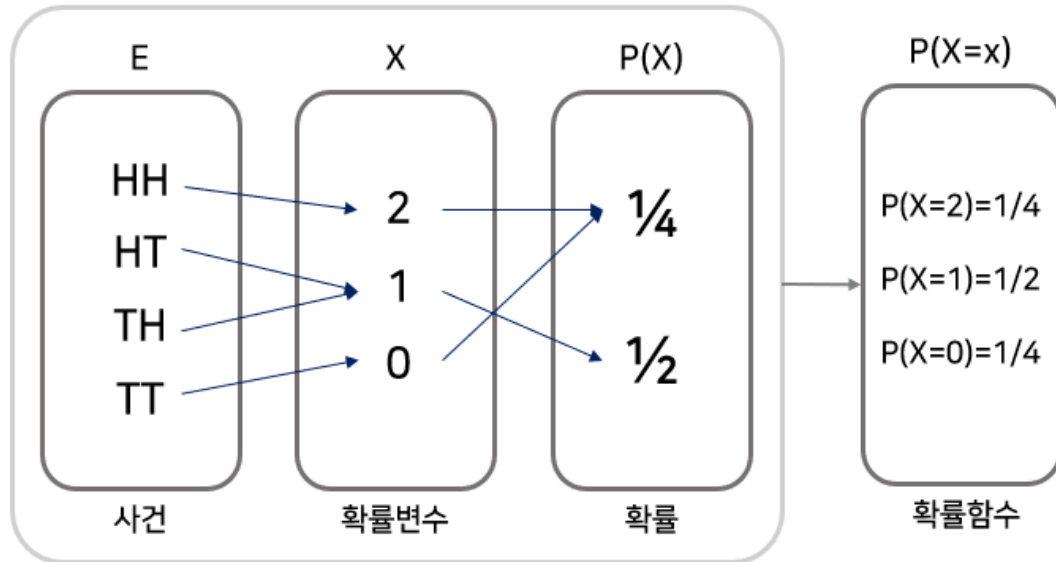
유튜브 : <https://youtu.be/7abHNnLhpQw>

IT융합공학부 이준희

‘확률 변수’란?

- ‘확률 변수’

➤ 정의 : 어떤 사건의 결과를 수치로 나타내는 변수



‘확률 변수’의 두 가지 유형

• ‘이산형 확률 변수’

➤ 정의 : 서로 구별되는 개별적인 값을 가질 수 있는 확률변수
즉, 이산확률변수는 유한 개의 값을 가진다.

✓ 특징

- 이산확률변수는 특정 값들의 집합을 가진다. 예를 들어, 정수 집합 $\{0, 1, 2, 3, \dots\}$ 처럼 **명확하게 구분되는 값**만 가질 수 있다.
- 각 값이 발생할 확률을 **확률질량함수(Probability Mass Function, PMF)**로 나타낼 수 있다. (이산확률변수의 확률분포함수)

ex) $P(X=3) = 1/6$

예시) 주사위를 던질 때: 주사위 눈의 값은 1, 2, 3, 4, 5, 6 중 하나. => ‘주사위 눈의 값’이 ‘이산확률변수’
동전을 던질 때: 동전의 결과가 앞면(1) 또는 뒷면(0) 중 하나. => ‘동전의 결과’가 ‘이산확률변수’

$$P(X = x) = \frac{1}{6}, \quad x \in \{1, 2, 3, 4, 5, 6\}$$



‘확률변수’의 두 가지 유형

• ‘연속형 확률 변수’

➤ 정의 : 연속적인 값을 가질 수 있는 확률변수

즉, 연속형 확률 변수는 무한히 많은 값을 가질 수 있으며, 이 값들은 특정 구간 내에서 연속적으로 존재한다.

✓ 특징

- 연속확률변수는 특정 값이 아닌 **값의 구간**에 대해 확률을 할당한다.
- 각 값이 발생할 확률은 0에 가깝기 때문에, **확률밀도함수(Probability Density Function, PDF)**로 나타낸다. (연속확률변수의 확률분포함수)
- 특정 값이 아니라 **값의 범위**에 대한 확률을 계산한다. ex) $P(165 < X < 180) = 0.85$

예시) **사람의 키**: 150cm에서 200cm 사이의 모든 실수 값이 가능하므로, 사람의 키는 연속확률변수이다.

시간: 특정 구간에서 시간(예: 0초에서 10초 사이)은 연속적인 값을 가질 수 있으므로, 시간은 연속확률변수이다.

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$



‘확률분포’란?

- ‘확률분포’

- 정의 : 확률변수가 취할 수 있는 모든 가능한 값과 그 값들이 발생할 확률을 포괄적으로 나타내는 개념
즉, 확률분포는 확률변수가 특정한 값을 가질 확률을 전체적으로 설명하는 것
- 형태 : 확률분포는 ‘표, 그래프, 함수’ 또는 수학적 표현 등 다양한 형태

확률변수 X 의 확률분포표

| 확률 변수 X 값과 각 값이 가질 확률 $P(X = x)$ 을 표로 표현

X	2	4	6	8	10	12	14	16	18
$P(X = x)$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{5}{25}$	$\frac{4}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{1}{25}$



‘확률분포함수’란?

- ‘확률분포함수’

- 정의 : ‘확률분포’를 수학적으로 나타내는 함수
- 확률변수 값이 발생할 확률을 나타내는 함수

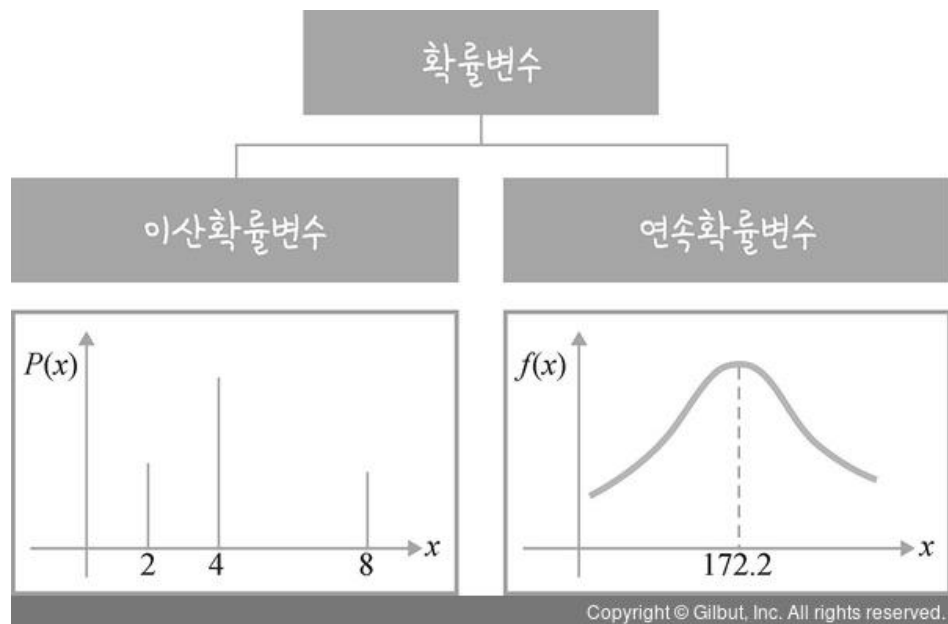
- ‘확률분포함수’의 종류

1. 확률질량함수(Probability Mass Function)

이산확률변수의 확률분포를 나타내는 함수 $\sum F(X)$

2. 확률밀도함수(Probability Density Function)

연속확률변수의 확률분포를 나타내는 함수 $\int F(X)$



‘이산확률분포’의 종류

- 베르누이 분포 (Bernoulli Distribution)

- 정의 : 성공과 실패 두 가지 결과만을 가지는 실험(베르누이 시행)에서, 성공 확률 p 와 실패 확률 $1-p$ 를 나타내는 이산 확률분포

- 이항분포 (Binomial Distribution)

- 정의 : 베르누이 시행을 n 번 반복했을 때, 성공이 k 번 나올 확률을 나타내는 분포

- 다항분포 (Multinomial Distribution)

- 정의 : 여러 개의 값을 가질 수 있는 확률 변수들에 대한 분포로, 여러 번의 독립적 시행에서 각 값이 특정 횟수가 나타날 확률을 정의한 분포

- 포아송분포 (Poisson Distribution)

- 정의 : 단위 시간 안에 어떤 사건이 몇 번 발생할 것인지 표현하는 분포

- 기하분포 (Geometric Distribution)

- 정의 : 베르누이 시행에서 첫 성공까지의 실패 횟수의 분포

- 초기하분포 (Hypergeometric Distribution)

- 정의 : 비복원추출에서 크기가 N 인 모집단에서 표본을 n 번 추출했을 때 원하는 것 k 개가 뽑힐 확률의 분포



‘연속확률분포’의 종류

• 정규분포 (Normal Distribution)

- 가장 널리 알려진 연속확률분포로, 통계적 추정 및 가설 검정의 핵심이다.
- 종 모양의 대칭적인 곡선**을 가진다. 평균 μ 와 표준편차 σ 로 정의된다.
- 자연 현상, 시험 점수, 키 등 많은 실제 데이터를 모델링하는 데 사용됩니다.
- 사회적/자연적 현상의 여러 자료가 정규분포를 따른다.
- 표준정규분포(Standard Distribution)**: 평균이 0이고, 표준편차가 1인 정규분포

• 지수분포 (Exponential Distribution)

- 특정 사건이 연속적으로 발생할 때, 두 사건 사이의 시간 간격에 대한 분포
- ex) 콜센터에서 두 통화 사이의 시간 간격, 고객이 도착하는 시간 간격, 방사성 물질이 붕괴할 때까지의 시간

• 스튜던트 t-분포 (Student's t-Distribution)

- 표본의 크기가 작고, 모집단의 분산이 알려져 있지 않은 경우 평균에 대한 가설 검정에 사용된다.
- 표본 평균이 정규분포를 따를 때, 분포의 꼬리가 두꺼워지는 특성을 가진다.

• 카이제곱분포 (Chi-Square Distribution)

- 카이제곱 검정에서 사용되며, 표본분산을 기반으로 모집단의 분산을 추정하는 데 사용된다.
- 정규분포를 따르는 K개의 서로 독립적인 확률변수들의 제곱의 합으로 분포를 정의
- N이 30이상이면 정규분포에 근사한다. (by 중심극한정리)

• F-분포(F-distribution)

- 두 개의 서로 독립적인 카이제곱 분포를 따르는 확률변수의 비율로 정의
- 두 집단 간 분산의 동일성 검정에 사용되는 검정 통계량 분포이다. (두 집단의 분산비를 비교)



Q & A

