금융인을 위한

통계와 데이터 분석 입문

확률변수와 확률분포 알아보기

한국금융연수원
KOREA BANKING INSTITUTE

학습 내용

1. 확률변수와 확률분포
2. 이산형 확률변수와 분포
   - (예)베르누이분포, 이항분포
3. 연속형 확률변수와 분포
   - (예)정규분포, T-분포

확률변수와 확률분포

확률변수(random variable)

모집단에 대해 어떤 구하고자 하는 것이 있을 때,
어떤 값을 가질지에 대한 가능성을 확률로
표현할 수 있음

- 확률변수에는 이산형, 연속형 확률 변수가 있음

확률분포(probability distribution)

확률변수가 특정 값을 취할 가능성을 수학적으로
표현한 것

- 확률변수 X가 확률 분포 f_x를 따를 때,
  $X \sim f_x(x)$라 표현함

이산형 확률분포

이산형(discrete) 확률변수

- **이산형(discrete) 확률변수의 예시 :**

$$X = \begin{cases} 1 & \text{투자적격 신용등급에 속하는 경우} \\ 0 & \text{투자적격 신용등급에 속하지 못하는 경우} \end{cases}$$

- 그 밖에도 만족도 등급(1-5), 부도여부(0-1) 등이 있음

이산형 확률분포

이산형(discrete) 확률분포

- **이산형(discrete) 확률변수의 분포 :**
  - $f_X(x_1) = P(X = x_1) = p_1, f_X(x_2) = P(X = x_2) = p_2, \ldots, f_X(x_k) = P(X = x_k) = p_k$
  - $0 \le p_i \le 1$
  - $\sum p_i = 1$
- **표로도 표현 가능함**

X	x_1	x_2	$\cdots$	x_k	sum
$f_X(x)$	p_1	p_2	$\cdots$	p_k	1

이산형 확률분포

이산형(discrete) 확률분포의 평균과 분산

- 평균(mean)
  - 기대값(expectation)이라고도 부름
  - 이산형 확률변수의 **기대값** : $E(X) = \sum x_i f_X(x) = \sum x_i p_i$
- 분산(variance)
  - 이산형 확률변수의 분산
    : $Var(X) = E[(X - E(X))^2] = E(X^2) - E^2(X)$
    $= \sum x_i^2 p_i - E^2(X)$

이산형 확률분포

베르누이분포

- X=1인 확률이 p, X=0인 확률이 (1-p)인 확률변수를 베르누이분포를 따른다고 함
  - $P(X = 1) = p, \quad P(X = 0) = 1 - p$
- 가격의 상승 또는 하락, 신용등급의 상승 또는 하락, 성공 또는 실패 등과 같이 오직 2개의 사건만 있는 경우 활용함

예

상승인 경우 X = 1, 하락인 경우 X = 0

- 베르누이분포를 따르는 확률변수의 평균과 분산 :
  - $E(X) = 1 \times p + 0 \times (1 - p) = p$
  - $Var(X) = 1 \times p + 0 \times (1 - p) - p^2 = p(1 - p)$

이산형 확률분포

이항분포

- 베르누이와 같이 두 범주로 나눠진 경우, **n번의 독립적**인 시행에서 **하나의 범주**에 속하는 사건이 **관측된 횟수**
- n번의 독립적인 **베르누이 시행의** 합이라고 생각할 수 있음
- $X \sim B(n, p)$라 표현함
- $P(X = k) = \dfrac{n!}{k!\,(n-k)!} p^k (1-p)^{(n-k)}$

이산형 확률분포

이항분포의 활용

예

어떤 기업의 영업이익이 전년보다 증가할 확률이 0.6이며, 영업이익의 증감 여부는 매해 독립적이라고 하자.

- 향후 5년 동안 두 해의 영업이익이 전년보다 증가할 확률은?
  - $X \sim B(5, 0.6)$이고 구해야 할 것은 $P(X = 2)$
- 향후 5년 동안 적어도 한 해의 영업이익이 전년도 대비 증가할 확률은?
  - $X \sim B(5, 0.6)$이고 구해야 할 것은 $P(X \geq 1)$

- X가 이항분포 $B(n, p)$를 따를 때, 확률변수 X의 평균과 분산 :
  - $E(X) = np$
  - $Var(X) = np(1 - p)$

연속형 확률분포

연속형(continuous) 확률변수

- 연속형(continuous) 변수의 예시 : 키, 몸무게, 펀드의 수익률 등
- 연속형(continuous) 확률변수의 분포는 확률밀도함수 $f_x(x)$로 수식으로 표현됨
  - $f_x(x) \geq 0$
  - $\int_{-\infty}^{\infty} f_x(x)dx = 1$: 확률의 합은 1
- 어떤 사건이 일어날 확률은 확률밀도함수 밑의 면적으로 계산됨
  - $P(a \leq X \leq b) = \int_a^b f_X(x)dx$

정규분포

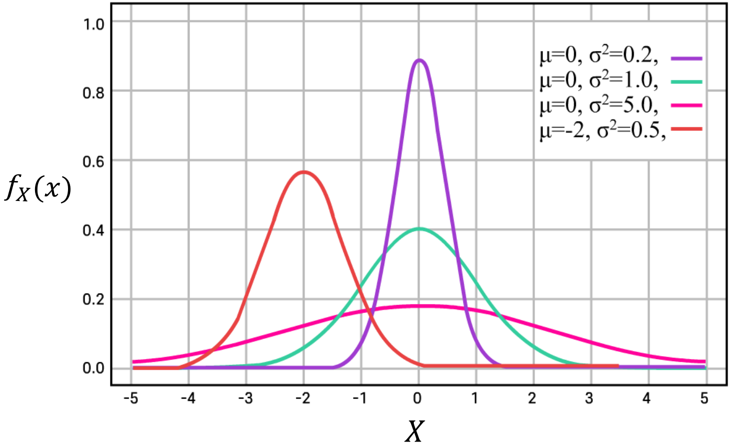
- 정규 분포(normal Distribution)는 **좌우대칭(symmetric)이고 하나의 피크(unimodal)를 가졌으며 벨 모양(bell-shape)의** 특징을 가지고 있음
  - **평균(μ)과 분산(σ^2) 두 개의 모수로 분포가 결정됨**
  - $X \sim N(\mu, \sigma^2)$인 X의 확률 분포 함수 :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

연속형 확률분포

정규분포

- 평균과 분산에 따른 정규분포의 **확률밀도함수**



- 모집단에 대한 가정, 중심극한정리 등에서 활용할 중요한 확률분포

정규분포의 표준화

- 정규분포의 표준화는 정규분포를 따르는 확률변수를 **평균이 0이고 분산이 1인 표준정규분포로 변환하는 것**
- $X \sim N(\mu, \sigma^2)$일 때 표준화는

$$Z = \frac{x-\mu}{\sigma} \rightarrow Z \sim N(0,1)$$

  - $X \sim N(10, 25)$일 때 $\mathrm{Z} = \dfrac{X - 10}{5}$는 $\mathrm{N}(0, 1)$을 따름

연속형 확률분포

T-분포

- T-분포는 **좌우대칭(symmetric)**이고 **하나의 피크(unimodal)**를 가졌으며 **벨 모양(bell-shape)**의 특징을 가진다는 점에서 정규분포와 유사하지만 **자유도**를 가진다는 점이 다름

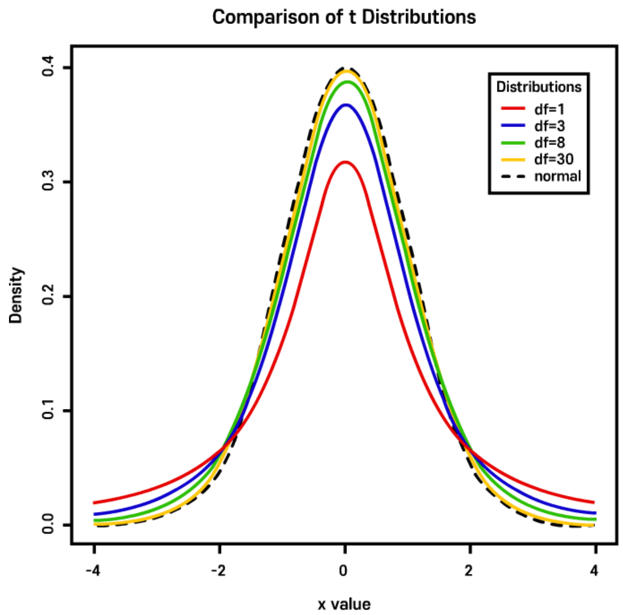
- 평균은 0이고 분산은 자유도 v가 2보다 클 때

  $\dfrac{v}{v-1}$로 정의됨

연속형 확률분포

T-분포

- 표준정규분포에 비해서 두꺼운 꼬리(heavy tail)를 가짐



- 구간추정, 평균에 대한 가설 검정 등에서 활용할 분포

학습정리

- 어떠한 사건이 일어날 가능성을 **확률**이라 하며 만족시켜야 할 기본 성질들이 있음

- 확률분포를 가진다고 가정된 값을 **확률변수**라 함

- 대표적인 **이산형** 확률분포로 **베르누이분포, 이항분포**가 있음
  > 두 분포에서 모두 일어날 수 있는 **사건은 두 종류**로 가정함
  > **이항분포**를 따르는 변수는 **베르누이분포를 따르는 변수들의 합**이라고 볼 수 있음

- 대표적인 **연속형** 확률분포로 **정규분포, T-분포**가 있음
  > **정규분포**는 **평균과 분산**으로 모양이 결정되며 표준화 작업을 통해 표준정규분포를 따르게 할 수 있음
  > **T-분포**는 정규분포와 유사한 모양이나 언제나 **평균이 0**이고 **꼬리부분이 두껍다**는 특징이 있음