금융인을 위한

통계와 데이터 분석 입문

표본평균의 분포 이해하기

한국금융연수원
KOREA BANKING INSTITUTE

학습 내용

1. 표본평균의 분포
2. 중심극한정리 이해
3. (실습)파이썬 numpy의 확률변수 생성 함수를 활용하여 중심극한정리를 확인

표본평균의 표본분포

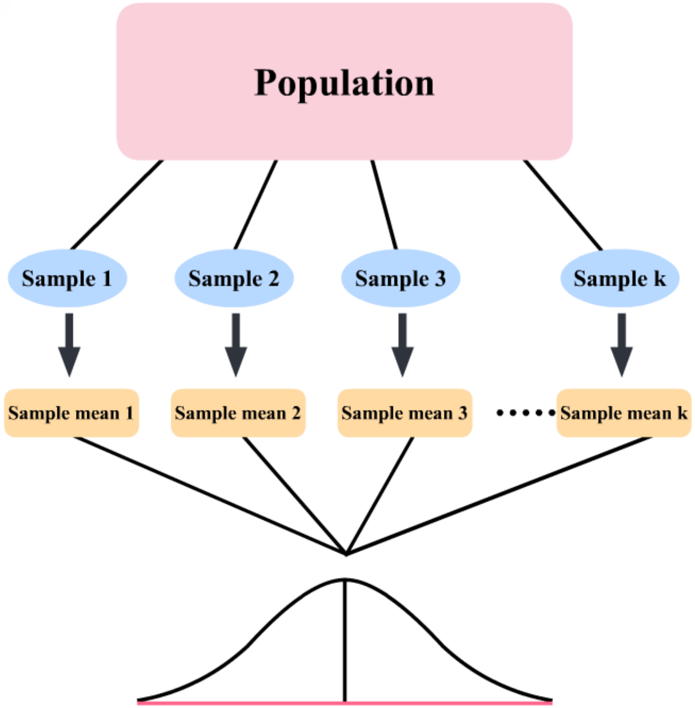
- 평균이 μ이고 분산이 σ^2 인 임의의 모집단으로부터 크기가 n인 표본으로부터 산출되는 **표본평균**

$$\boxed{\bar{X} = \sum_{1}^{n} X_i / n}$$

- **표본평균의 평균** $E(\bar{X}) = \mu$
- **표본평균의 분산** $Var(\bar{X}) = \dfrac{\sigma^2}{n}$
  - n이 커질수록 $\bar{X}$의 분산은 작아지게 됨
    : 모평균인 μ를 중심으로 밀집되게 됨
- **표본평균의 표준편차** $SD(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$
  : **표준오차(standard error)**라고 함

표본평균의 표본분포

- 모집단의 분포와 상관없이 성립함



중심극한정리

Q. 표본평균은 어떠한 분포에서 나왔다고 가정해야 할까?

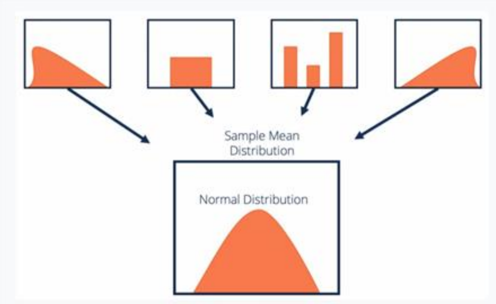
중심극한정리

표본평균 $\bar{X}$ 의 분포는 **표본크기** n이 충분히 커짐에 따라 평균이 μ, 분산이 σ^2/n인 **정규분포로 근사**함

- 이를 응용하면 $\bar{X}$ 를 **표준화한**

확률변수 $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 은

n이 충분히 커짐에 따라 **표준정규분포** $N(0, 1)$에 근사하다는 것을 알 수 있음



표본평균의 표본분포

- 얼마나 큰 표본크기 n? 30?
  - 모집단의 분포형태에 따라 달라지므로 절대적인 기준은 없음
- $\bar{X}$와 $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$의 분포
  - 모집단이 정규분포인 경우는 정확히 정규분포를 따름
  - 모집단이 정규분포가 아닌 경우는 정규분포에 근사함

학습정리

- **모집단의 분포와 상관없이** 표본평균의 평균과 분산은

  $E(\bar{X}) = \mu, \qquad Var(\bar{X}) = \dfrac{\sigma^2}{n}$ 를 만족함

- **표본크기** n이 충분히 커짐에 따라 표본평균의 분포는 평균이 μ, 분산이 σ^2/n인 **정규분포로 근사함** (중심극한정리)

- 정규분포와 거리가 있는 **균일분포**와 **이항분포**를 따르는 모집단에서의 여러 개의 표본을 추출하고 **표본크기가** 증가함에 따라 표본평균의 분포가 **정규분포**에 근사한다는 것을 파이썬 실습을 통해 확인하였음