이론 통계학

5장 랜덤샘플

2021년 봄학기

전북대학교 통계학과

모집단과 표본

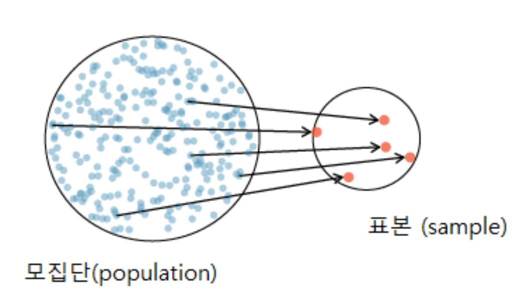
표본적률

정규분포로부터의 표본추출

- 카이제곱분포
- t-분포
- F-분포

순서통계량

통계학



- 여론조사, 제품의 수명조사, 새로운 의약품의 효능에 관한 연구
- 통계적 추론 방법들은 불확실성를 계량화하기 위하여 확률을 사용.
- **표본분포**: 확률분포이론을 통계이론에 적용하는 데에 있어서 중요한 다리 역할

모집단과 표본

랜덤샘플

모집단 (population
연구 또는 관측 대상이 되는 전체

랜덤표본(random sample)
확률변수 $X_1, X_2, \ldots, X_n$의 결합 확률밀도함수

$$f_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

$\Rightarrow X_1, X_2, \ldots, X_n$을 모확률밀도함수가 $f(\cdot)$인 크기가 n인 **랜덤표본(random sample)**

random sample $\Leftrightarrow$ iid (independent and identically distributed)

모집단 (population
연구 또는 관측 대상이 되는 전체

랜덤표본(random sample)
확률변수 $X_1, X_2, \ldots, X_n$의 결합 확률밀도함수

$$f_{X_1,X_2,\ldots,X_n}(x_1,\ldots,x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

$\Rightarrow X_1, X_2, \ldots, X_n$을 모확률밀도함수가 $f(\cdot)$인 크기가 n인 **랜덤표본(random sample)**

random sample $\Leftrightarrow$ iid (independent and identically distributed)

모집단 (population
연구 또는 관측 대상이 되는 전체

랜덤표본(random sample)
확률변수 $X_1, X_2, \ldots, X_n$의 결합 확률밀도함수

$$f_{X_1, X_2, \ldots, X_n}(x_1, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

$\Rightarrow X_1, X_2, \ldots, X_n$을 모확률밀도함수가 $f(\cdot)$인 크기가 n인 **랜덤표본(random sample)**

random sample $\Leftrightarrow$ iid (independent and identically distributed)

예제 5.1
모분포가 '성공'확률이 p인 베르누이 분포라고 하고, 서로 독립인 베르누이시행을 10회 반복하였을 때의 표본을 $X_1, X_2, \ldots, X_{10}$으로 표기하자.

- 베르누이 확률밀도함수
- $X_1, X_2, \ldots, X_{10}$의 결합 확률밀도함수

통계량 (statistic)과 표본분포(sampling distribution)

- 미지의 모수를 포함하지 않는 **랜덤표본의 함수** $T = T(X_1, X_2, \ldots, X_n)$
- 통계량 자체도 확률변수임
- 통계량의 분포, 즉 랜덤표본 또는 랜덤표본의 함수의 분포를 **표본분포**라고 함

$X_1, X_2, \ldots, X_n$이 확률밀도함수 $f(x;\theta)$ (θ 미지)로부터 얻은 랜덤샘플이다. 다음 중 통계량은?

- 표본평균 $\bar{X}_n = \sum_{i=1}^{n} X_i/n$
- 표본최댓값 $\max\{X_1, \ldots, X_n\}$
- $\bar{X}_n - \theta$
- $\max\{X_1/\theta, \ldots, X_n/\theta\}$

표본적률

표본적률

$X_1, X_2, \ldots, X_n$이 확률밀도함수 $f(x;\theta)$로부터 얻은 랜덤샘플이다.

r차 적률(rth moment)

- 모 r차 적률 : $\mu_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x;\theta)dx$
- 표본 r차 적률 : $m_r = \dfrac{1}{n}\sum_{i=1}^{n} X_i^r$

- 표본적률은 모수를 포함하지 않는 랜덤표본의 함수 $\Rightarrow$ 표본적률은 통계량
- $E(m_r) = \mu_r \Rightarrow$ 모적률 추정에 표본적률을 사용
- 모평균과 모분산 추정에 표본평균과 표본분산 사용.

표본적률

표본평균 $\overline{X}_n$과 표본분산 S_n^2

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

정리

$X_1, X_2, \ldots, X_n$이 $E(X_i) = \mu, Var(X_i) = \sigma^2$인 랜덤샘플일 때 다음이 성립한다.

- $E(\overline{X}_n) = \mu, \quad Var(\overline{X}_n) = \dfrac{\sigma^2}{n}$
- $E(S_n^2) = \sigma^2$