## 표본의 분포

- 통계량 : 확률표본의 측정값들인 확률변수($X_1, X_2, \ldots, X_n$)와 확률표본의 크기인 상수 n의 함수(표본평균)
- 표본분포 : 통계량들의 확률분포 -> 모수 추정/가설검정에 주로 사용

**1. 정규모집단의 표본분포**

**-> 모집단의 평균($\mu$)과 분산($\sigma^2$)을 알 때**

**> 표본평균의 확률분포 :** $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ , $z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

-> $z^2$은 카이제곱분포(자유도:1)를 따름

**-> 모집단의 분산을 알지 못할 때**

**> 표본평균의 확률분포 :** $\overline{X} \sim N(\mu, \frac{S^2}{n})$ ($n \geqq 30$), $\frac{\overline{X} - \mu}{S / \sqrt{n}} \sim T(n-1)$

: 정규모집단일 경우, 표본분산과 모집단의 분산의 값이 서로 비슷할 것이기 때문에, 표본분산을 정규모집단의 분산처럼 사용. but, 표본분산이 원래 모분산 보다 큰 값일 것. -> t분포(원래는 정규분포를 따르는데 표본분산을 사용함으로써 정규분포보다 꼬리가 좀 더 두꺼운 분포를 따를 것이다.) 만약 표본의 크기가 충분히 크다면 표본분산과 모분산이 비슷할 것이기 때문에 정규분포를 따름)

**t분포*

- $\frac{\overline{X} - \mu}{S / \sqrt{n}} = \frac{(\overline{X} - \mu)/(\sigma / \sqrt{n})}{\sqrt{S^2 / \sigma^2}} = \frac{Z}{\sqrt{V(n-1)/(n-1)}}$의 확률분포
- 0을 중심으로 좌우대칭, 종모양을 띄고 있음
- 표준정규분포보다 꼬리가 두꺼운 분포(n->∞일수록 표준정규분포와 비슷해짐)
- $\frac{Z}{\sqrt{V(v)/v}} \sim T(v)$(v가 커질수록 정규분포 모양에 가까워짐)($Z, V$는 서로 독립 확률변수)
- 표본정규분포와 카이제곱분포에 의해 만들어지는 분포
- $T$분포의 자유도는 $\chi^2$분포의 자유도를 따라감

**▶ 표본분산의 분포 :** $\frac{(n-1)S^2}{\sigma^2} \sim X^2$(자유도 n-1)

*감마함수 : $\Gamma(\alpha) = \int_0^\infty y^{(\alpha-1)} e^{-y} dy, \alpha > 0$

-> 성질

1. $\Gamma(1/2) = \sqrt{\pi}$
2. $\alpha$가 양의 정수 인 경우, $\Gamma(\alpha) = (\alpha - 1)!$
3. 모든 $\alpha$에 대해 $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

*감마분포 : $\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{(\alpha-1)}e^{-x/\beta}, \alpha>0, \beta>0, x>0$

*카이제곱분포 : $\alpha = v/2(v=\text{자유도}), \beta=2$

$\frac{1}{\Gamma(v/2)2^{v/2}}x^{(v/2-1)}e^{-x/2}, \alpha>0, x>0$

-> 평균 : v, 분산 : 2v

**-> 카이제곱분포를 따르는 서로 독립인 확률변수들의 합의 확률분포** $\sim \chi^2(v_1+v_2+...+v_k)$

**-> 표준정규분포 z의 $z^2$은 자유도 1의 카이제곱분포를 따름.**

=> 평균과 분산을 아는 정규모집단으로부터 추출한 $X_1, X_2, X_3, ..., X_n$ 확률변수들의

$Z_i = \frac{X_i-\mu}{\sigma}, i=1,2,3,...,n$ 는 각각 독립이며 모평균, 모분산을 따르는 정규분포-> $\Sigma_{i=1}^{n} Z_i^2$는 자유도 n인 카이제곱분포를 따름.

***F분포 : 카이제곱분포를 따르는 서로 독립인 2개의 확률변수의 비율**

$F(v_1, v_2) = \frac{U/v_1}{V/v_2}$

$F(n_1-1, n_2-1) = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$

**2. 정규모집단이 아닌 표본분포**

**-> 모집단의 분산($\sigma^2$)을 아는 경우**

-> 중심극한정리 : 모집단의 평균, 분산 : $\mu, \sigma^2$, 표본의 크기가 충분히 크다면($n \geqq 30$) 표본평균은 모집단의 확률분포에 상관없이(이산확률분포도 가능) **근사적으로 $\overline{X} \sim N(\mu, \sigma^2/n)$을 따름**

(Y=$X_1+X_2+...+X_n$, Y=$n\overline{X}$ => $Y \sim N(n\mu, n\sigma^2)$) *(각 확률변수가 정규분포를 따르지 않지만, 표본평균이 근사적으로 정규분포를 따르기 때문에 다음과 같은 결과)*

**경험적으로는 n이 30이상이면 문제가 없다고 알려져 있음. 모집단이 좌우대칭인 확률분포를 띌수록 표본의 크기가 작더라도 잘 맞음. 만약 모집단이 한쪽으로 크게 치우쳤다면 표본의 크기가 커야 함*

**-> 모분산의 분산($\sigma^2$)을 모르는 경우**

- n>=30이면 표본분산과 모분산이 비슷할 것이기 때문에 표본분산을 모분산 대신 사용 가능

$\overline{X} \sim N(\mu, S^2/n)$

- n이 충분히 크지 못하면 표본분산과 모분산이 다를 것임. -> 무슨 분포인지는 구할 수 없음

*좌우대칭인 확률분포

- 정규분포, T분포

$\therefore t_{1-\alpha} = -t_{\alpha}$

$z_{1-\alpha} = -z_{\alpha}$