정규분포로부터의 표본추출

정규분포로부터의 표본추출

- 통계적 추정이론이나 방법 등에서 정규분포의 역할은 매우 큼
- 추정량들이 흔히 확률변수의 합의 꼴로 표현되는데 중심극한정리에 의하여 이들의 분포가 정규분포로 근사될 수 있음
- 모분포에 대한 정규성 가정은 통계적 추론에 필요한 수리적 접근을 용이하게 함
- 정규분포에서 얻어진 표본과 그의 함수들에 대한 성질 파악은 통계적 추론에 매우 중요하고 유용함

정리 5.4

서로 독립인 확률변수 $X_i, i = 1, \ldots, n$들이 정규분포 $N(\mu_i, \sigma_i^2)$을 따르면

$$\sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$$

- 정규분포를 따르는 확률변수의 선형결합은 정규분포를 따름
- 정규분포의 적률생성함수를 이용하여 증명

예제

어떤 종류의 제품을 1개 생산하는 데 걸리는 시간 X가 $N(6, 2^2)$을 따른다. 10개의 제품을 생산하는 데 걸리는 시간이 70 이상일 확률을 구하시오.

정리 5.5:표본평균 $\overline{X}_n$의 분포

$X_1, X_2, \ldots, X_n$ 이 $N(\mu, \sigma^2)$로부터의 랜덤샘플일 때 표본평균 $\overline{X}_n$의 분포

$$\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

카이제곱분포

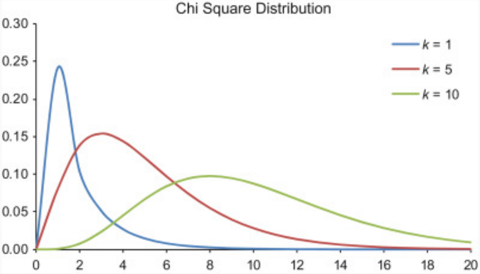
모분산 σ^2의 추정량으로 사용되는 표본분산 S_n^2의 표본분포를 이해하는데 중요한 역할

자유도(degree of freedom)가 n인 카이제곱분포, χ_n^2

확률밀도함수: $Gamma(n/2, 2)$

$$f(x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2} I(x > 0)$$

평균$=n$, 분산$=2n$



카이제곱분포와 정규분포의 연관성

표준정규확률변수 Z에 대하여 $X = Z^2$의 확률밀도함수

$$f(x) = \frac{1}{\sqrt{\pi}2^{1/2}} x^{n/2-1} e^{-x/2} I(x > 0) \quad \Rightarrow \quad X \sim \chi_1^2$$

정리 5.2와 5.3

- $X_1 \sim \chi_m^2,\ X_2 \sim \chi_n^2$이고 서로 독립 $\Rightarrow X_1 + X_2 \sim \chi_{m+n}^2$
- 서로 독립인 X_1과 X_2에 대하여 $Y = X_1 + X_2$라고 할 때

$$Y \sim \chi_m^2, X_1 \sim \chi_n^2 \Rightarrow X_2 \sim \chi_{m-n}^2$$

- 서로 독립인 카이제곱 확률변수의 합은 카이제곱 확률변수
- 카이제곱 확률변수의 차이도 적절한 조건 하에서 카이제곱 확률변수

표본평균 $\overline{X}_n$과 표본 분산 S_n^2의 독립

- 이변량정규분포를 따르는 두 확률변수의 공분산이 0이면 두 확률변수는 서로 독립
- 정규분포를 따르는 여러 확률변수 사이에도 공분산이 0이면 서로 독립

표본평균 $\overline{X}_n$과 표본 분산 S_n^2의 독립

- $\overline{X}_n$, $X_k - \overline{X}_n$: $X_1, \ldots, X_n$의 선형결합 $\Rightarrow$ 정규분포
- $Cov(\overline{X}_n, X_k - \overline{X}_n) = 0$
- $\overline{X}_n$와 $X_k - \overline{X}_n$는 서로 독립
- 표본분산 S_n^2은 $(X_1 - \overline{X}_n, X_2 - \overline{X}_n, \ldots, X_n - \overline{X}_n)$의 함수
- $\overline{X}_n$와 S_n^2은 서로 독립

정리 5.6:표본 분산 S_n^2의 분포

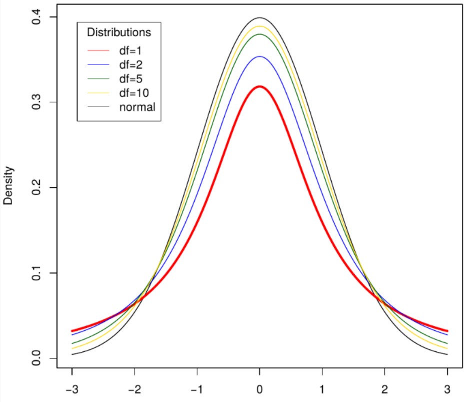
$X_1, X_2, \ldots, X_n$ 이 $N(\mu, \sigma^2)$로부터의 랜덤샘플일 때

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_n, \quad \frac{n(\overline{X}_n-\mu)^2}{\sigma^2} \sim \chi^2_1$$

t-분포 (Student's t distribution)

- 정규분포의 모평균 가설검정 등을 포함한 검정론에서 중요한 역할
- 표준정규분포와 유사 (0을 중심으로 대칭)하지만 꼬리가 더 두껍다.
- 자유도가 증가할수록 t-분포는 표준정규분포에 가까이 간다.



t_k의 확률밀도함수

$X \sim t_k$, 자유도 (degree of freedom) k인 t분포의 pdf

$$f(x) = \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} (1 + x^2/k)^{-(k+1)/2}$$

정의 5.2

$Z \sim N(0,1), U \sim \chi_k^2$이며 서로 독립일 때

$$T = \frac{Z}{\sqrt{U/k}}$$

의 확률분포는 자유도가 k인 t-분포

$T \sim t_n$일 때 $E(T)$와 $Var(T)$?

$X_1, X_2, \ldots, X_n$ 이 $N(\mu, \sigma^2)$로부터의 랜덤샘플일 때

정리 5.5, 5.6, 5.7

- $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- $\overline{X}_n$ and S_n^2 are independent
- $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{(n-1)}$
- $\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$

정규분포로부터 구한 독립인 두 표본의 분산비에 대한 분포를 설명하는 데 중요한 역할

정의 5.3: $F \sim F_{m,n}$ **자유도** m, n**인** F**-분포**

$U \sim \chi^2_m, V \sim \chi^2_n$이며 독립일 때 F의 확률분포

$$F = \frac{U/m}{V/n}$$

F-분포의 성질

- $X \sim t_n \Rightarrow X^2 \sim F_{1,n}$
- $X \sim F_{m,n} \Rightarrow X^{-1} \sim F_{n,m}$

$F \sim F_{m,n}$일 때 $E(F)$와 $Var(F)$?

Homework 7주차

- 다음을 보이시오.
  - (1) $Z \sim N(0,1)$일 때 $Z^2 \sim \chi^2_1$
  - (2) $X_1, \ldots, X_n$이 랜덤샘플일 때 $Cov(\overline{X}_n, X_k - \overline{X}_n) = 0$
- 다음을 구하시오.
  - (3) $X_1, \ldots, X_n$이 Gamma(α, β)로부터의 랜덤샘플일 때 $\overline{X}_n$의 확률분포
  - (4) $X_1, \ldots, X_n$이 $N(\mu, \sigma^2)$로부터 구한 랜덤샘플일 때 $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2/n$의 분산
  - (5) $X_1, \ldots, X_{16}$과 $Y_1, \ldots, Y_{25}$이 각각 $N(0,9)$와 $N(2,16)$으로부처 구한 서로 독립인 랜덤샘플일 때 $\overline{X}_{16} - \overline{Y}_{25}$의 분포

교재 5장 연습문제 1,4,12

순서통계량

순서통계량 (order statistics)

$X_1, X_2, \ldots, X_n$이 랜덤샘플일 때 작은 것부터 크기순으로 나열한 통계량

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

최대값, 최소값, 중앙값, 범위, 제1사분위수...

- $X_{(k)}$는 $X_1, X_2, \ldots, X_n$ 중의 하나의 값
- $X_{(k)}$의 확률분포는 $X_1, X_2, \ldots, X_n$ 모두의 영향을 받음
- $X_{(k)}$의 확률분포는 X_i의 확률분포와 같지 않음

예제

X_1, X_2, X_3: 1∼5까지 정수 값을 가지는 이산형 확률분포의 랜덤샘플

- $P(X_{(1)} = 2, X_{(2)} = 3, X_{(3)} = 5)$
- $P(X_{(3)} \leq 3)$

$$\{X_{(1)} = 2, X_{(2)} = 3, X_{(3)} = 5\} = \{X_1 = 2, X_2 = 3, X_3 = 5\}?$$

$$\{X_{(3)} \leq 3\} = \{X_1 \leq 3, X_2 \leq 3, X_3 \leq 3\}$$

순서통계량 $X_{(k)}$의 확률분포

$X_1, X_2, \ldots, X_n$: pdf $f(x)$, cdf $F(x)$를 가지는 확률분포로부터의 랜덤샘플

최소값($X_{(1)}$)과 최대값 ($X_{(n)}$)의 분포

- $f_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x)$
- $f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x)$

$$f(x) = \frac{dF(x)}{dx}$$

$$\{X_{(n)} \leq x\} = \{X_1 \leq, X_2 \leq x, \cdots, X_n \leq x\}$$

$$\{X_{(1)} > x\} = \{X_1 > x, X_2 > x, \cdots, X_n > x\}$$

$X_1, X_2, \ldots, X_n$: pdf $f(x)$, cdf $F(x)$를 가지는 확률분포로부터의 랜덤샘플

정리 5.8: $X_{(k)}$의 분포

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}f(x)[1-F(x)]^{n-k}$$

$X_{(k)}$의 cdf $F_{X_{(k)}}(x) = P(X_{(k)} \leq x)$를 구하여 미분

$\{X_{(k)} \leq x\} = \{X_1, X_2, \ldots, X_n$ 중 x보다 작은 것의 수가 k개 이상$\}$

순서통계량 $X_{(k)}$의 확률분포

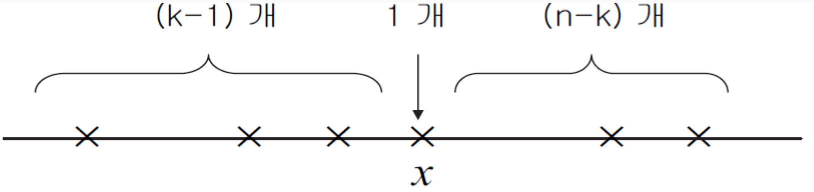
$$\{X_{(k)} \le x\} = \{X_1, X_2, \ldots, X_n \text{ 중 } x\text{보다 작은 것의 수가 } k\text{개 이상}\}$$

$$I_i = \begin{cases} 1, & X_i \le x \\ 0 & X_i > x \end{cases} \Rightarrow \sum_{i=1}^{n} I_i \sim B(n, F(x))$$

$$\begin{aligned} F_{X_{(k)}}(x) &= P(X_{(k)} \le x) = P(X_1, X_2, \ldots, X_n \text{ 중 } x\text{보다 작은 갯수} \ge k) \\ &= P(\sum_{i=1}^{n} I_i \ge k) = \sum_{l=k}^{n} \binom{n}{l} [F(x)]^l [1 - F(x)]^{n-l} \end{aligned}$$

순서통계량 $X_{(k)}$의 확률분포

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}f(x)[1-F(x)]^{n-k}$$



예제 5.1, 5.2, 5.4

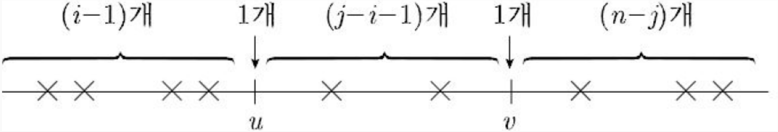
$X_1, X_2, \ldots, X_n$: Uniform(0,1)로부터의 랜덤샘플일 때 $X_{(k)}$의 확률분포

예제 5.3

각 부품의 수명이 평균이 $1/\lambda$인 지수분포를 따를 때 독립인 n개의 부품이 직렬로 연결된 시스템의 수명에 대한 확률분포

정리 5.9
$i < j$, $u < v$인 경우

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}$$
$$[F(u)]^{i-1} f(u)[F(v) - F(u)]^{j-i-1} f(v)[1 - F(v)]^{n-j}$$



예제 5.5

$X_1, X_2, \ldots, X_n$: Uniform(0,1)로부터의 랜덤샘플일 때 $X_{(n)} - X_{(1)}$의 확률분포