[7강]

표본분포 (2)

정보통계학과 이긍희 교수

학습목표

1. 표본평균 차의 분포를 이해한다.
2. 표본분산 비의 분포를 이해한다.
3. 확률적 수렴을 이해한다.
4. 중심극한정리를 이해한다.

[7강]

표본분포 (2)

1 표본평균 차의 확률분포

1.1 표본평균의 차에 대한 분포

- $X_i \sim N(\mu_1, \sigma_1^2),\ Y_j \sim N(\mu_2, \sigma_2^2),\ \ X_i, Y_j$ 독립
  - $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2,\ \sigma_1^2/m + \sigma_2^2/n)$

1.1 표본평균의 차에 대한 분포

- $X_i \sim N(\mu_1, \sigma^2)$, $Y_j \sim N(\mu_2, \sigma^2)$, X_i, Y_j 독립,
  - $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2,\ \sigma^2[1/m + 1/n])$
  - $\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim N(0,1)$

1.2 합동 표본분산

- $S_p^2 = \frac{1}{m+n-2}\{\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2\}$
- $\frac{(m+n-2)S_p^2}{\sigma^2} \sim \chi^2(m+n-2)$

1.3 표본평균의 차에 대한 분포

- 두 집단 모집단 분산을 모르지만 같다고 가정

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m + n - 2)$$

[7강]

표본분포 (2)

2 표본분산 비의 확률분포

2.1 두 집단 분산의 비교

- **표본분산비** : 두 개 독립인 카이제곱 통계량 비
  - F분포로 표현

2.2 F분포

- $V_1 \sim \chi^2(r_1)$, $V_2 \sim \chi^2(r_2)$, 독립

$$F = \frac{V_1/r_1}{V_2/r_2} \sim F(r_1, r_2)$$

- 분산분석(analysis of variance) 이용 분포

2.2 F분포

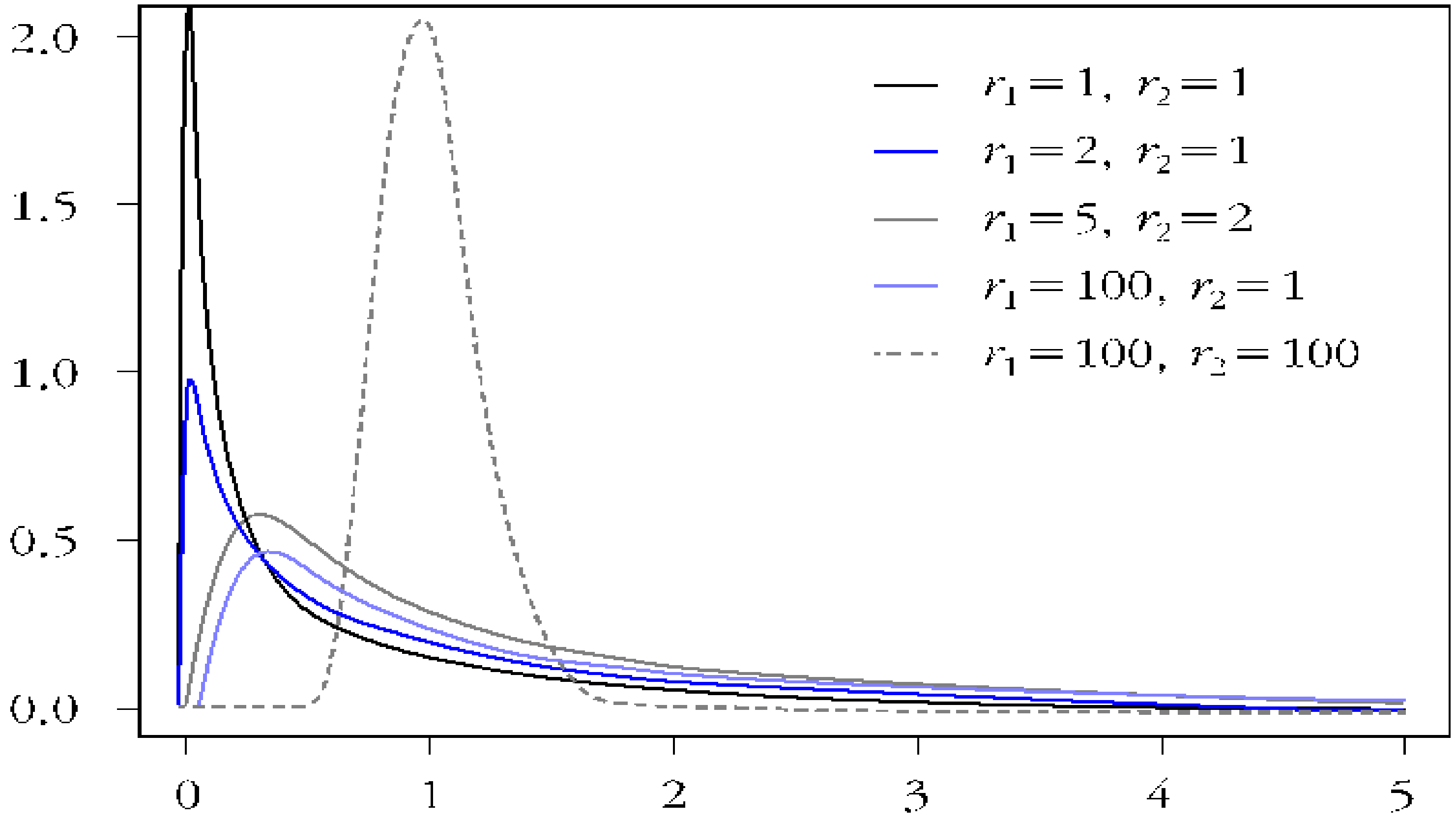
- F분포 확률밀도함수

$$f(x) = \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \frac{r_1}{r_2} \left(\frac{r_1}{r_2} x\right)^{\frac{r_1}{2}-1} \left(1 + \frac{r_1}{r_2} x\right)^{-\frac{r_1+r_2}{2}},$$

$$x > 0$$

2.2 *F*분포

- *F*분포 확률밀도함수



2.3 *F*분포의 특성

- $F \sim F(r_1, r_2) \rightarrow \frac{1}{F} \sim F(r_2, r_1)$
- $T \sim t(n) \;\rightarrow T^2 \sim F(1, n)$

2.4 표본분산의 비교

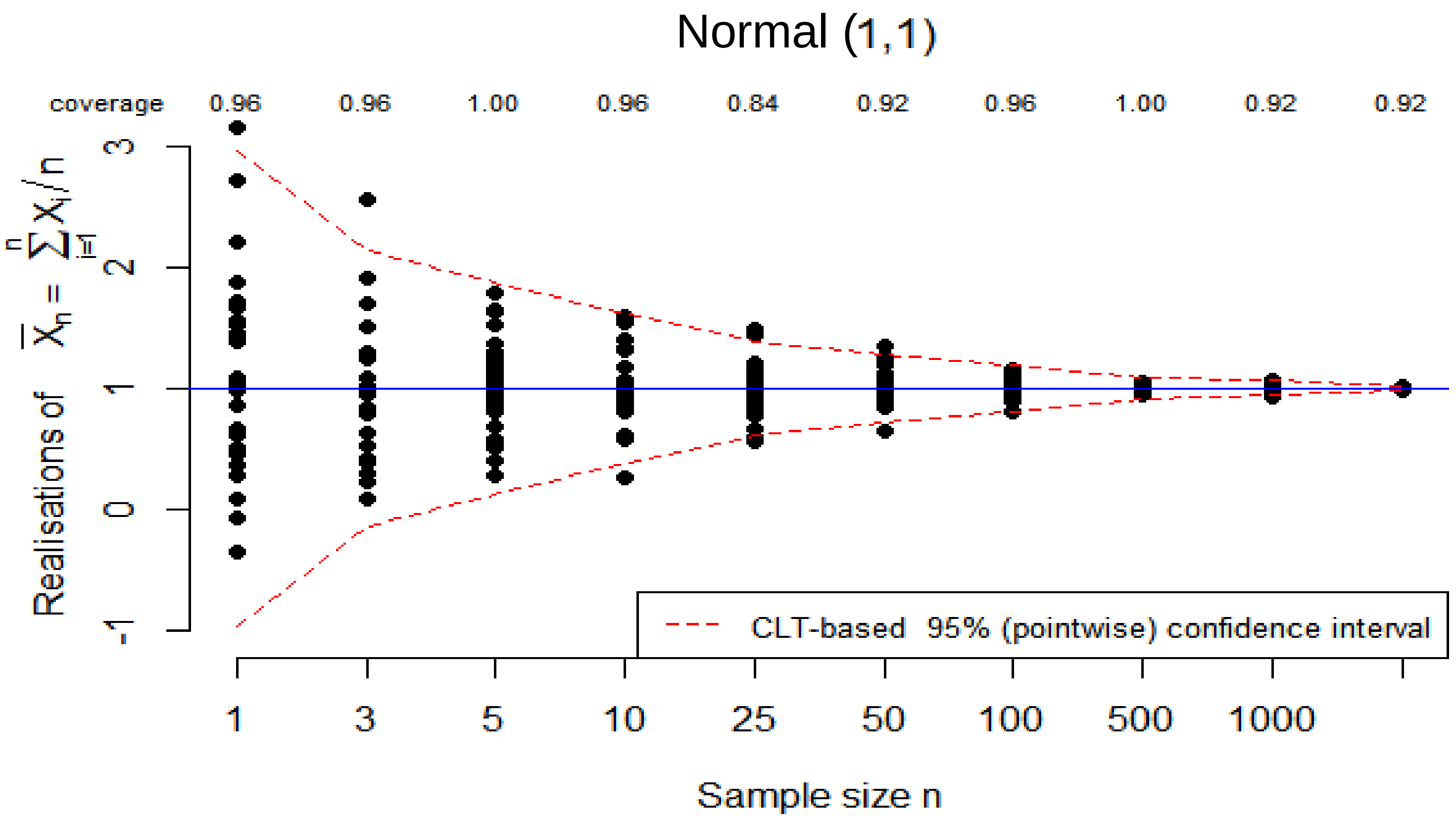
- $X_i \sim N(\mu_1, \sigma_1^2),\ Y_j \sim N(\mu_2, \sigma_2^2),\ X_i, Y_j$ 독립
- $S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2,\ S_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$
- $\frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2(m-1),\ \frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2(n-1)$

2.4 표본분산의 비교

- $$F=\frac{\frac{(m-1)S_1^2}{\sigma_1^2}/(m-1)}{\frac{(n-1)S_2^2}{\sigma_2^2}/(n-1)}=\frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}\sim F(m-1,n-1)$$

3 극한분포

3.1 확률적 수렴


Normal (1,1)
coverage 0.96 0.96 1.00 0.96 0.84 0.92 0.96 1.00 0.92 0.92
Realisations of $\bar{X}_n = \sum_{i=1}^{n} X_i / n$
-1 0 1 2 3
CLT-based 95% (pointwise) confidence interval
1 3 5 10 25 50 100 500 1000
Sample size n

3.1 확률적 수렴

- 모집단 $\sim N(\mu, \sigma^2)$ $\rightarrow$ 표본평균 $\sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- n이 무한히 커지면 표본평균은 μ에 근접
  - $\bar{X}$가 상수값 μ에 확률적 수렴 : $\bar{X} \xrightarrow{p} \mu$
  - 임의 양수 ε, $\lim_{n\to\infty} P(|\bar{X} - \mu| < \varepsilon) = 1$

3.2 마코프 부등식

- $u(X)$: X의 양의 함수

임의의 상수 $\varepsilon > 0,\ P[u(X) \geq \varepsilon] \leq \dfrac{E[u(X)]}{\varepsilon}$

3.2 마코프 부등식

예 4.14

$X_1, \ldots, X_n \sim E(X_i) = \mu,\ Var(X_i) = \sigma^2,$

$\bar{X}_n$가 μ에 확률적으로 수렴함을 증명하라.

3.3 약대수법칙

- $X_1, \dots, X_n \sim E(X_i) = \mu,\ Var(X_i) = \sigma^2$ 확률표본

  $\bar{X}_n$가 μ에 확률적으로 수렴

3.4 확률적 수렴성의 성질

- $Y_n \xrightarrow{p} c, \quad W_n \xrightarrow{p} d$ $(a, b, c, d$ 상수$)$

① $aY_n + b \xrightarrow{p} ac + b$

② $Y_n + W_n \xrightarrow{p} c + d$

③ $Y_n W_n \xrightarrow{p} cd$

④ $\dfrac{1}{Y_n} \xrightarrow{p} \dfrac{1}{c}$

⑤ g가 연속이면 $g(Y_n) \xrightarrow{p} g(c)$

3.4 확률적 수렴성의 성질

예 4.15

$X_1, \cdots, X_n$ 평균 μ, 분산 σ^2의 확률표본.

$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \xrightarrow{p} \sigma^2$임을 보여라.

3.5 분포의 수렴

- $Y_n \xrightarrow{d} Y$
- 누적분포함수나 적률생성함수를 이용

$$\lim_{n\to\infty} F_n(y) = F(y) \quad or \quad \lim_{n\to\infty} M_n(t) = M(t)$$

3.5 분포의 수렴

예 4.17

$X_1, \cdots, X_n \sim Ber(p)$ 확률표본, $Y_n = \sum_{i=1}^{n} X_i$, $np = \mu$, n이 무한히 커짐에 따라 Y_n이 포아송분포를 따름을 보여라.

3.5 분포의 수렴

예 4.17

Y_n이 포아송분포를 따름을 보여라.

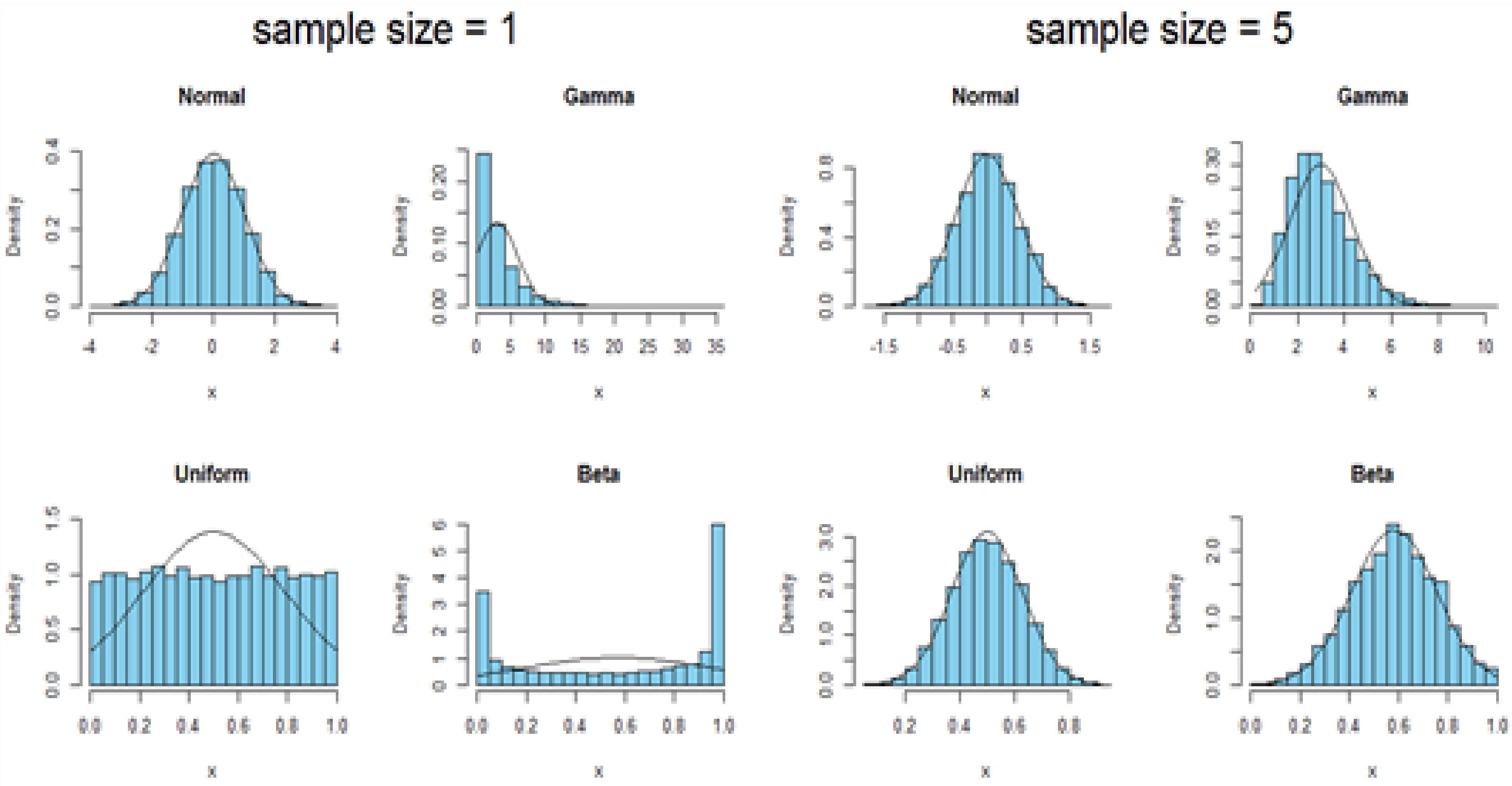
3.6 표본평균의 분포

- $X_1, \cdots, X_n$ 서로 독립 $\sim N(\mu, \sigma^2)$
  - $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

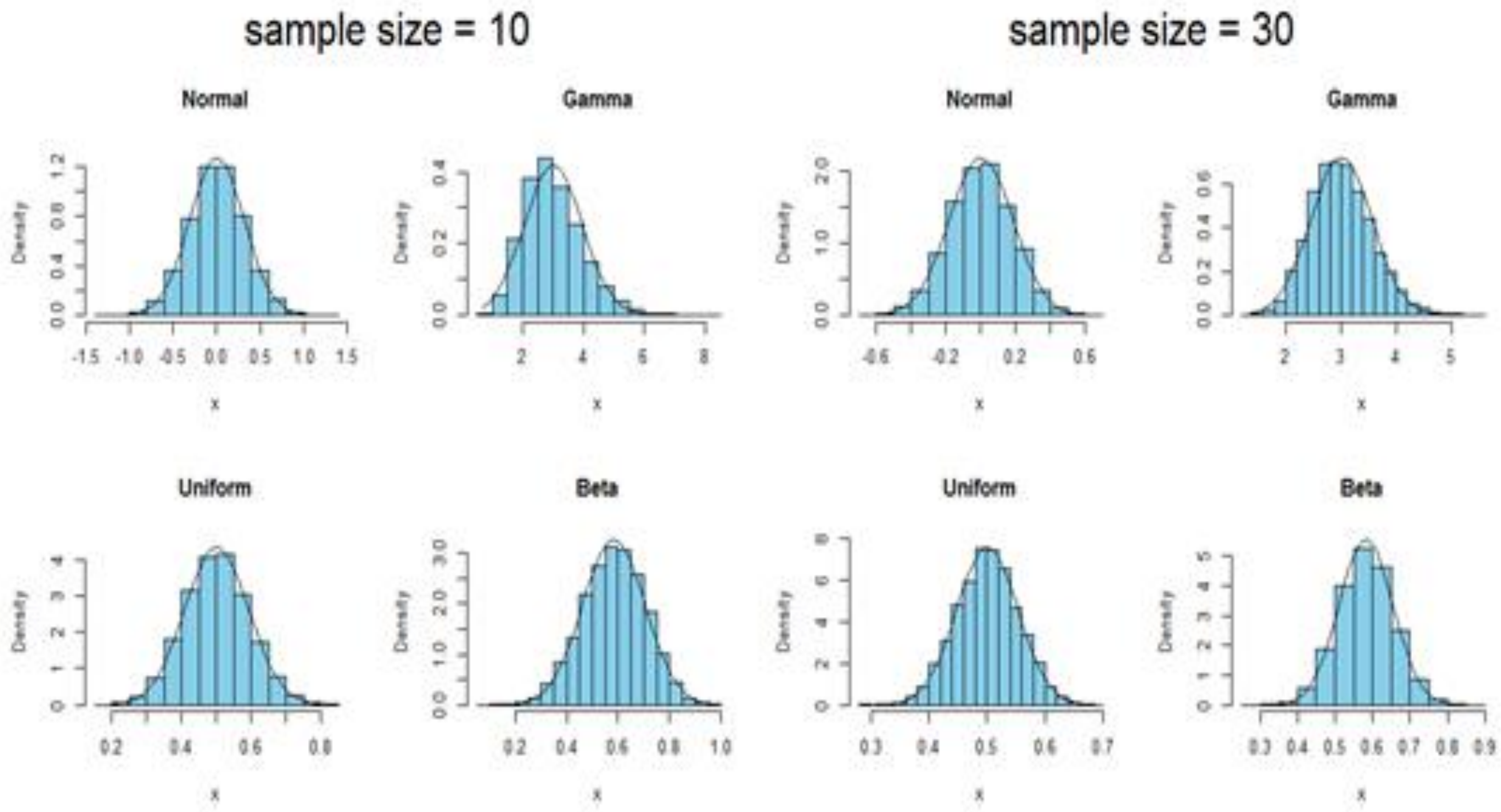
3.7 중심극한정리

- $X_1, \cdots, X_n \sim E(X_i) = \mu,\ Var(X_i) = \sigma^2$ 확률표본
  - $Y_n = \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

3.7 중심극한정리


sample size = 1
Normal
Gamma
Uniform
Beta
sample size = 5
Normal
Gamma
Uniform
Beta
Density
x

3.7 중심극한정리


sample size = 10
sample size = 30
Normal
Gamma
Normal
Gamma
Uniform
Beta
Uniform
Beta
Density
x

3.7 중심극한정리

예 4.18

중심극한정리를 증명하라.

3.7 중심극한정리

예 4.18

중심극한정리를 증명하라.

3.7 중심극한정리

예 4.18

중심극한정리를 증명하라.

3.7 중심극한정리

예 4.19

$Y_n \sim \chi^2(n)$, $(Y_n - n)/\sqrt{2n} \xrightarrow{d} N(0,1)$임을 보여라.

3.8 이항분포의 정규근사

- $Y_n \sim B(n,p)$일 때

$$Z = \frac{Y_n - np}{\sqrt{np(1-p)}}$$의 극한확률분포는 $N(0,1)$

3.8 이항분포의 정규근사

- 이항분포 정규근사의 연속성 수정

$$P(a \leq Y \leq b) \approx \Phi\left[\frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right] - \Phi\left[\frac{a - 0.5 - np}{\sqrt{np(1-p)}}\right]$$

3.8 이항분포의 정규근사

예 4.20

한국인 성인 50%가 최소한 한 개의 신용카드를 가진다. 30명 성인을 표본추출할 때 19명에서 20명 사이의 성인이 최소한 한 개의 신용카드를 소지하고 있을 확률을 정규근사를 사용하여 구하라.

3.9 극한분포의 연산

- $Y_n \xrightarrow{d} Y, W_n \xrightarrow{p} c$일 때(c는 상수)

① $Y_n + W_n \xrightarrow{d} Y + c$

② $Y_n W_n \xrightarrow{d} cY$

③ $\dfrac{Y_n}{W_n} \xrightarrow{d} \dfrac{Y}{c}$

3.9 극한분포의 연산

④ $h(x)$가 미분 가능하고 $h'(x)$가 b에서 연속

$$\sqrt{n}[h(Y_n) - h(b)] \xrightarrow{d} N(0, [h'(b)]^2\sigma^2)$$

3.9 극한분포의 연산

예 4.21

$X_1, \dots, X_n \sim (\mu, \sigma^2)$인 확률표본일 때, $Y_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$

의 극한분포를 구하라.

[7강]

표본분포 (2)

4 정리하기

정리하기

1. $X_i \sim N(\mu_1, \sigma^2)$ m개 확률표본, $Y_j \sim N(\mu_2, \sigma^2)$의 n개 확률표본, σ^2 미지, 변수 X_i와 Y_j가 서로 독립

- $\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p^2\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2)$

정리하기

2. $X_i \sim N(\mu_1, \sigma_1^2)$ m개 확률표본, $Y_j \sim N(\mu_2, \sigma_2^2)$ n개 확률표본, σ^2 미지, 변수 X_i와 Y_j가 서로 독립

- $\frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$

정리하기

3. $X_1, \cdots, X_n$이 $E(X_i) = \mu,\ Var(X_i) = \sigma^2$ 확률표본
   - (약대수법칙) 표본평균 $\bar{X}_n$가 μ에 확률적으로 수렴
   - (중심극한정리) $\bar{Y}_n = \sqrt{n}(\bar{X}_n - \mu)$가 평균이 0이고 분산이 σ^2인 정규분포로 수렴

다음 시간 안내 ▼

8강. 점추정

수고하셨습니다.