07 강

통계적 추론

표본분포 2

한국방송통신대학교 통계·데이터과학과
이긍희 교수

학습내용

1. 표본평균의 차의 분포를 이해한다.
2. 표본분산의 비의 분포를 이해한다.
3. 확률적 수렴과 분포의 수렴을 이해한다.
4. 중심극한정리를 이해한다.

01

표본평균과

표본분산 관련 분포

1 표본평균의 차이에 대한 분포

표본평균과 표본분산의 분포

- $X_i \sim N(\mu, \sigma^2), i = 1,2,\cdots,n,$ 확률표본
  - $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(\mathrm{n}-1)$
  - $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

1 표본평균의 차이에 대한 분포

표본평균의 차이 분포 : 분산이 다를 때

- $X_i \sim N(\mu_1, \sigma_1^2),\ i = 1,2,\cdots,m$
- $Y_j \sim N(\mu_2, \sigma_2^2),\ j = 1,2,\cdots,n$
- 확률표본, X_i, Y_j 독립
- $\bar{X}\text{-}\bar{Y} \sim N(\mu_1 - \mu_2, \sigma_1^2/m + \sigma_2^2/n)$

1 표본평균의 차이에 대한 분포

표본평균의 차이 분포 : 분산이 같을 때

- $X_i \sim N(\mu_1, \sigma^2),\ i = 1,2,\cdots,m$
- $Y_j \sim N(\mu_2, \sigma^2),\ j = 1,2,\cdots,n$
- 확률표본, X_i, Y_j 독립
- $\bar{X}$-$\bar{Y} \sim N(\mu_1 - \mu_2, \sigma^2\left[\frac{1}{m} + \frac{1}{n}\right])$

1 표본평균의 차이에 대한 분포

공동 표본분산의 분포

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

1 표본평균의 차이에 대한 분포

표본평균의 차이 분포 : 모분산 미지, 동일

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

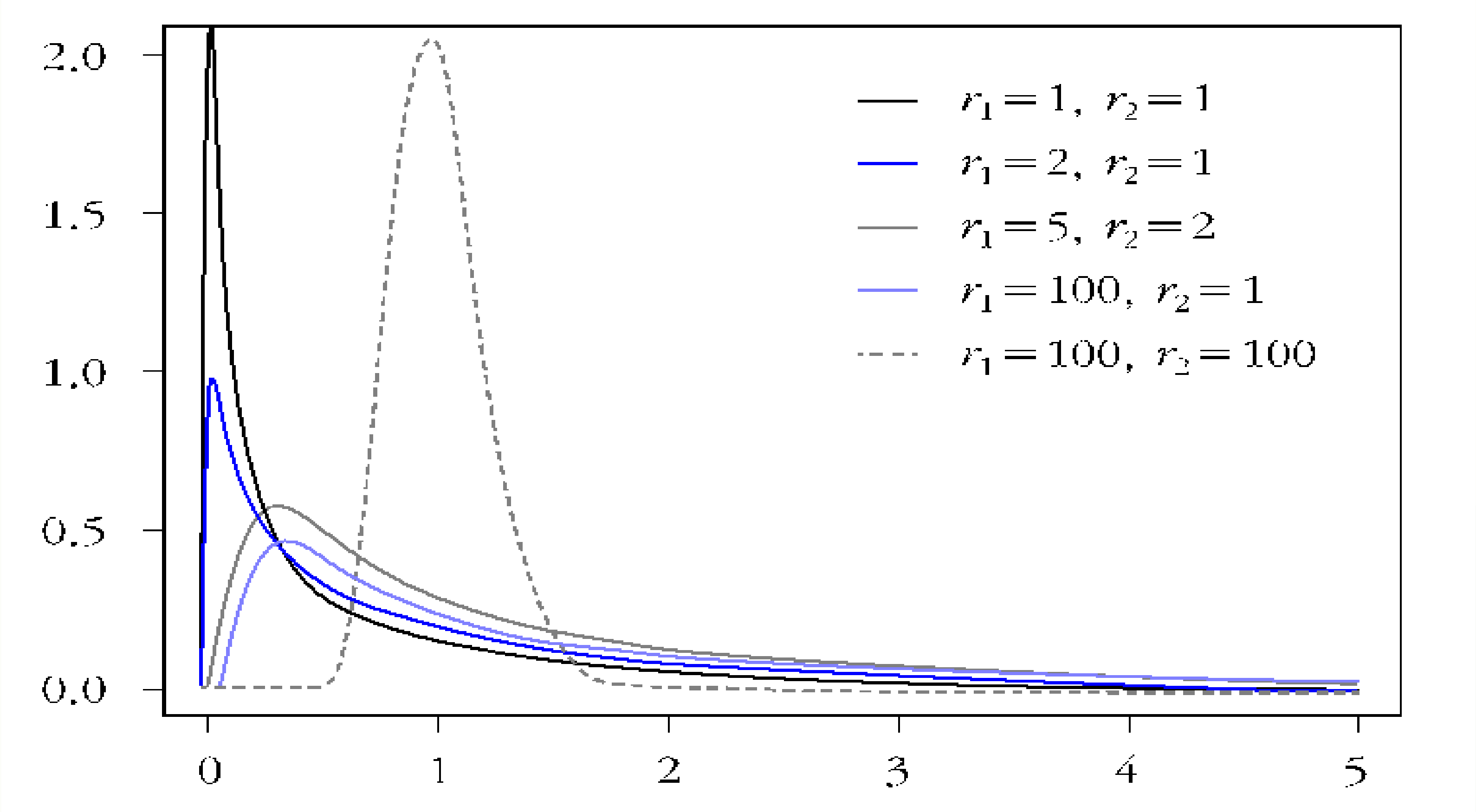
2 표본분산의 비에 대한 분포

F분포

- $V_1 \sim \chi^2(r_1), V_2 \sim \chi^2(r_2)$, 독립
- $F = \dfrac{V_1/r_1}{V_2/r_2} \sim F(r_1, r_2)$
- $f(x) = \dfrac{\Gamma\left(\frac{r_1+r_2}{2}\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \dfrac{r_1}{r_2} \left[\dfrac{r_1}{r_2} x\right]^{\frac{r_1}{2}-1} \left[1 + \dfrac{r_1}{r_2} x\right]^{\frac{r_1+r_2}{2}}, x > 0$

2 표본분산의 비에 대한 분포

F분포



2 표본분산의 비에 대한 분포

F분포의 성질

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

2 표본분산의 비에 대한 분포

표본분산의 비교

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

2 표본분산의 비에 대한 분포

표본분산의 비교

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

02

극한 분포

1 확률적 수렴

확률적 수렴의 정의

- $Y_n \xrightarrow{p} c$
- $\lim_{n \to \infty} P(|Y_n - c| \geq \varepsilon) = 0$, 임의의 상수 $\varepsilon > 0$

  $\Leftrightarrow \lim_{n \to \infty} P(|Y_n - c| < \varepsilon) = 1$, 임의의 상수 $\varepsilon > 0$

1 확률적 수렴

확률적 수렴

- $X_i \sim N(\mu, \sigma^2), i = 1,2,\cdots,n,$ 확률표본 $\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- $n \to \infty \ \Rightarrow \bar{X}$는 μ에 근접

1 확률적 수렴

마코프 부등식

- $u(X)$: X의 양의 함수, 임**의의 상수** $\varepsilon > 0$
- $P[u(X) \geq \varepsilon] \leq \frac{E[u(X)]}{\varepsilon}$
- $P[u(X) < \varepsilon] \geq 1 - \frac{E[u(X)]}{\varepsilon}$

1 확률적 수렴

약대수 법칙

- $X_i \sim (\mu, \sigma^2), i = 1,2,\cdots,n,$ 확률표본일 때 $\bar{X}_n$이 μ에 확률적으로 수렴

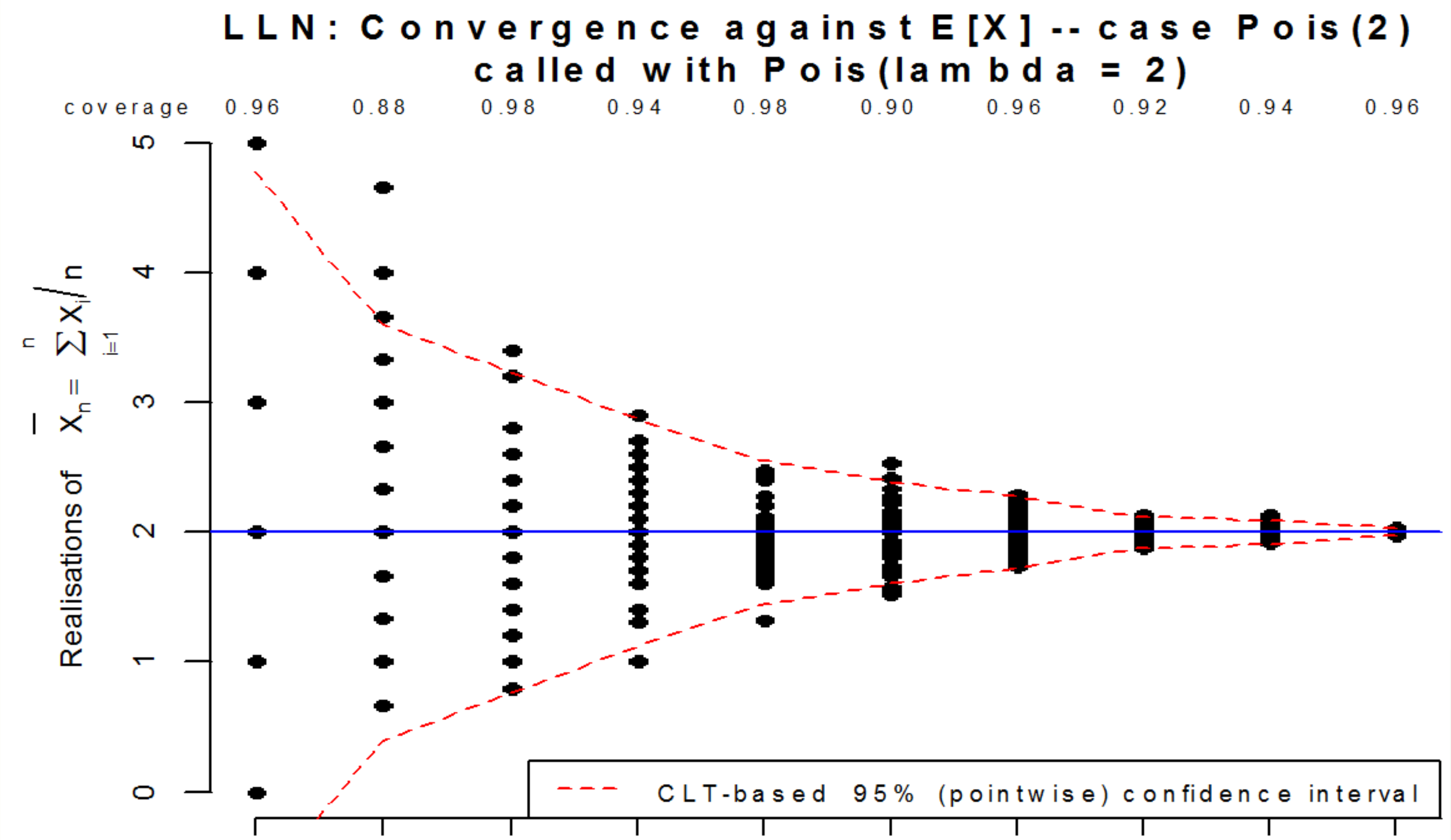
1 확률적 수렴

약대수 법칙

- $X_i \sim (\mu, \sigma^2), i = 1,2,\cdots,n,$ 확률표본일 때 $\bar{X}_n$이 μ에 확률적으로 수렴

1 확률적 수렴

약대수 법칙



1 확률적 수렴

확률적 수렴의 성질

- $Y_n \xrightarrow{p} c, W_n \xrightarrow{p} d$, ($a, b, c, d$ 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 $\frac{1}{Y_n} \xrightarrow{p} \frac{1}{c}$
  - ⑤ g 연속 $\Rightarrow g(Y_n) \xrightarrow{p} g(c)$

1 확률적 수렴

예 $X_1, X_2, \cdots, X_n$: $(\mu,\ \sigma^2)$ 의 확률표본일 때

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

1 확률적 수렴

예 $X_1, X_2, \cdots, X_n$: $(\mu,\ \sigma^2)$ 의 확률표본일 때

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

2 분포의 수렴

분포의 수렴의 정의

- $Y_n \xrightarrow{d} Y$
- $\lim_{n\to\infty} F_n(y) = F(y) \ \ or \ \ \lim_{n\to\infty} M_n(t) = M(t)$

2 분포의 수렴

예 $X_1, X_2, \cdots, X_n$:$Ber(p)$의 확률표본, p가 0에 가깝고, $np = \mu$일 때 $Y_n = \sum_{i=1}^{n} X_i$이 n이 무한히 커짐에 따라 포아송분포를 따름을 보여라.

2 분포의 수렴

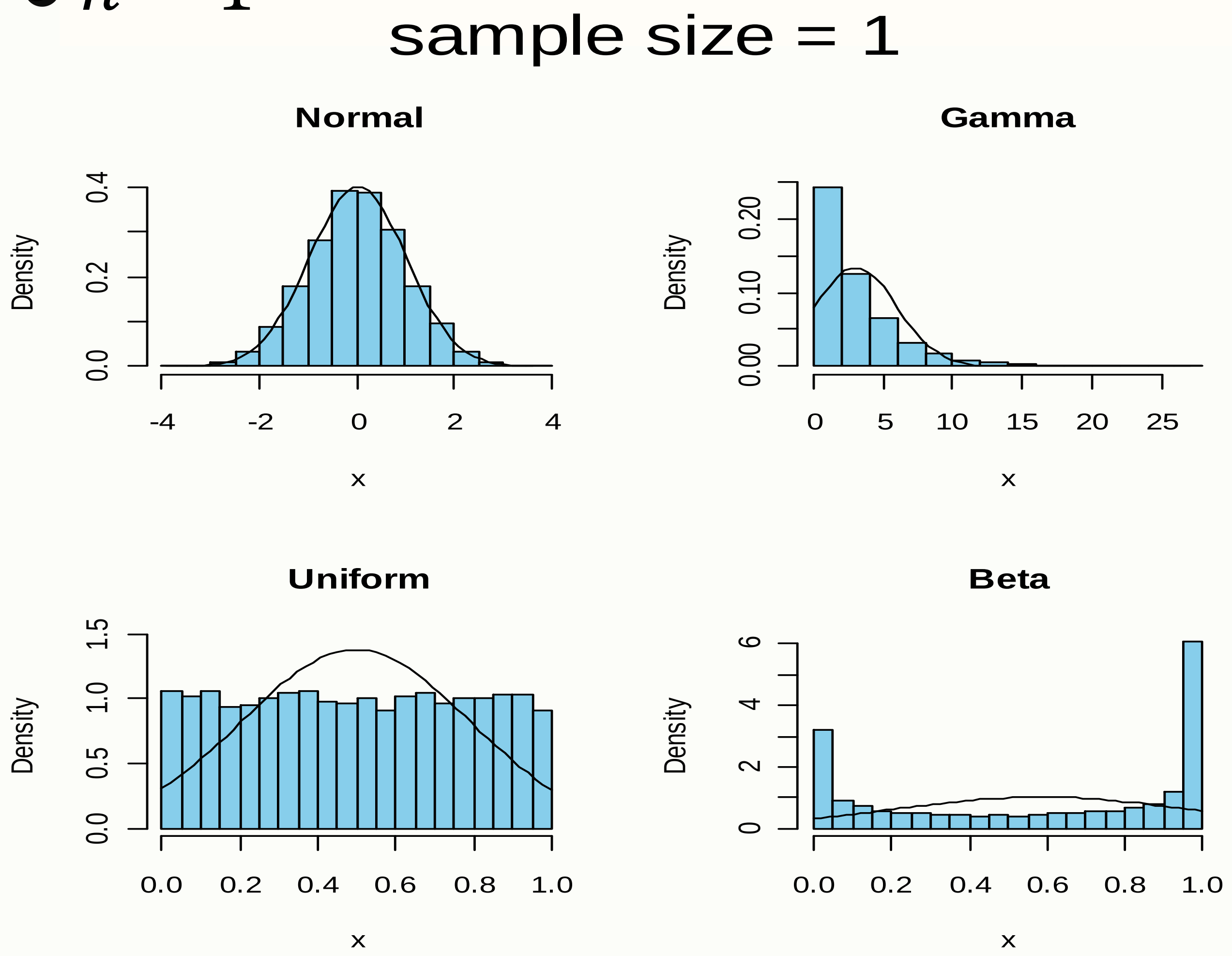
예 $X_1, X_2, \cdots, X_n$:$Ber(p)$의 확률표본, p가 0에 가깝고, $np = \mu$일 때 $Y_n = \sum_{i=1}^{n} X_i$이 n이 무한히 커짐에 따라 포아송분포를 따름을 보여라.

03

중심극한정리

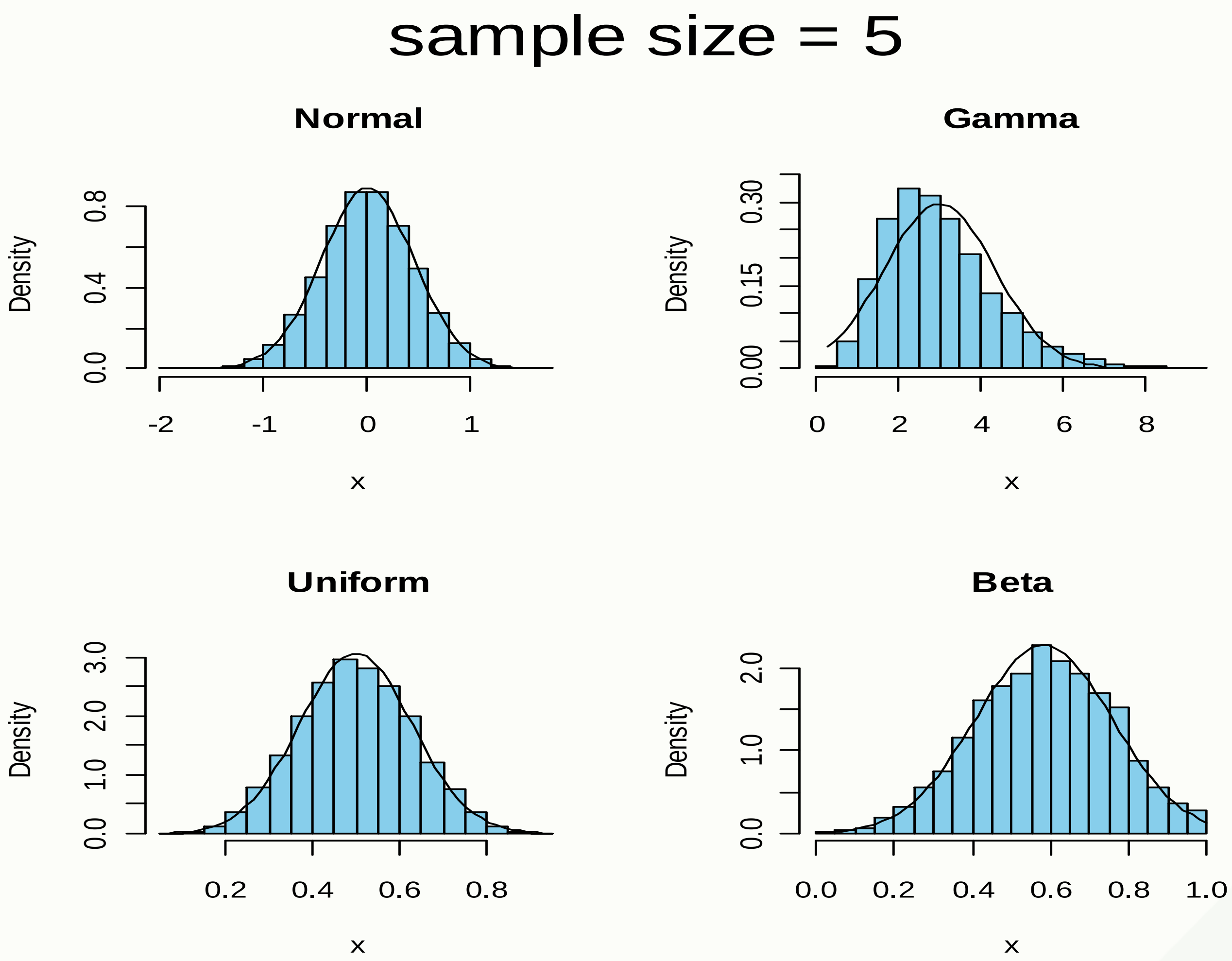
1 분포의 극한

- $n = 1$


sample size = 1
Normal
Gamma
Uniform
Beta
Density
x
-4 -2 0 2 4
0.0 0.2 0.4
0 5 10 15 20 25
0.00 0.10 0.20
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.5 1.0 1.5
0 2 4 6

1 분포의 극한

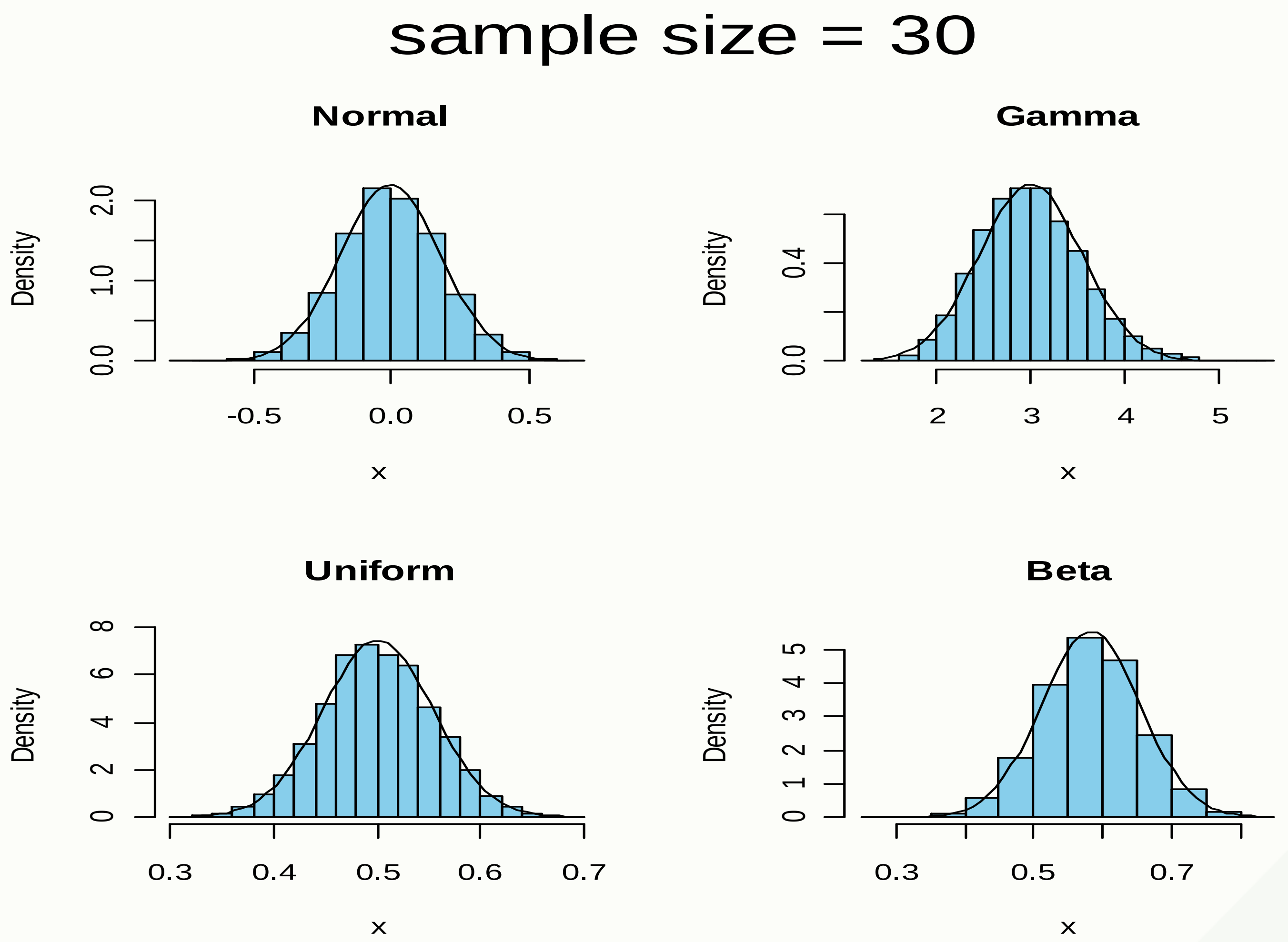
- $n = 5$



sample size = 5
Normal
Density
0.0
0.4
0.8
-2
-1
0
1
x
Gamma
Density
0.00
0.15
0.30
0
2
4
6
8
x
Uniform
Density
0.0
1.0
2.0
3.0
0.2
0.4
0.6
0.8
x
Beta
Density
0.0
1.0
2.0
0.0
0.2
0.4
0.6
0.8
1.0
x

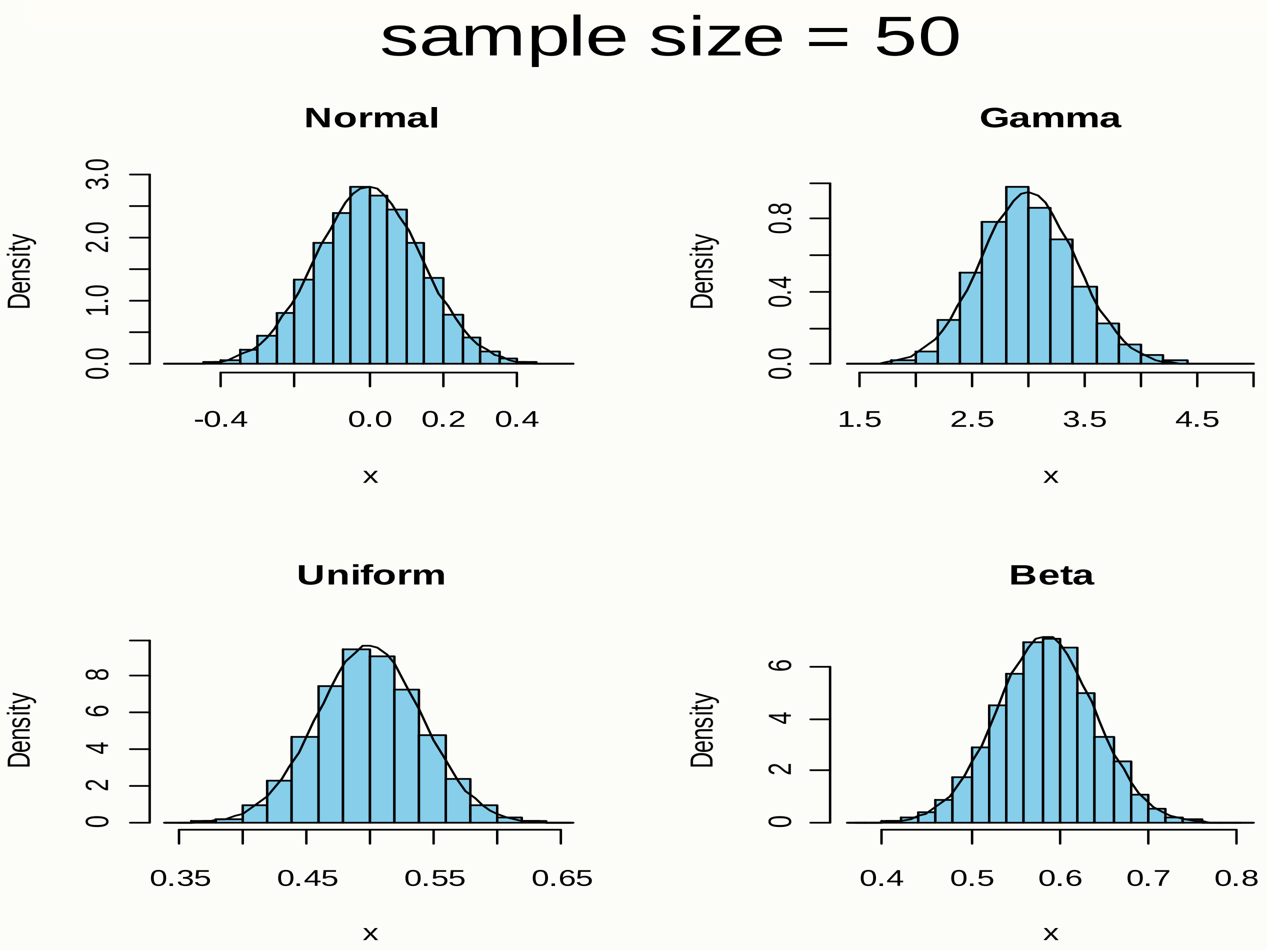
1 분포의 극한

- $n = 30$


sample size = 30
Normal
Density
0.0
1.0
2.0
-0.5
0.0
0.5
x
Gamma
Density
0.0
0.4
2
3
4
5
x
Uniform
Density
0
2
4
6
8
0.3
0.4
0.5
0.6
0.7
x
Beta
Density
0
1
2
3
4
5
0.3
0.5
0.7
x

1 분포의 극한

- $n = 50$


sample size = 50
Normal
Gamma
Uniform
Beta
Density
x
-0.4
0.0
0.2
0.4
1.5
2.5
3.5
4.5
0.35
0.45
0.55
0.65
0.4
0.5
0.6
0.7
0.8
0.0
1.0
2.0
3.0
0.4
0.8
0
2
4
6
8

2 중심극한정리

중심극한정리에 필요한 정리

- $X_1, X_2, \cdots, X_n : M(t) \;\Rightarrow M_{X_1+X_2+\cdots+X_n}(t) = [M(t)]^n$
- $\lim_{n\to\infty}(1+\frac{c}{n})^n = e^c$
- Taylor 급수 : $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$
  - $f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_n$

2 중심극한정리

중심극한정리

- $X_i \sim (\mu, \sigma^2), i = 1,2,\cdots,n,$ 확률표본일 때

$$Y_n = \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

2 중심극한정리

중심극한정리

- $X_i \sim (\mu, \sigma^2), i = 1,2,\cdots,n,$ 확률표본일 때

$$Y_n = \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

2 중심극한정리

중심극한정리

- $X_i \sim (\mu, \sigma^2), i = 1,2,\cdots,n,$ 확률표본일 때

$$Y_n = \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

2 중심극한정리

중심극한정리

- $X_i \sim (\mu, \sigma^2), i = 1,2,\cdots,n,$ 확률표본일 때

$$Y_n = \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

2 중심극한정리

예 $Y_n \sim \chi^2(n)$, $(Y_n - n)/\sqrt{2n}$ 의 극한 분포는?

3 이항분포의 정규근사

이항분포의 정규근사

- $Y_n \sim B(n,p), Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$

3 이항분포의 정규근사

이항분포의 정규근사

- $Y_n \sim B(n,p), Z_n = \dfrac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$

4 극한분포의 연산

극한분포의 연산 (Slutsky 정리)

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

  $\Rightarrow\ Y_n + W_n \xrightarrow{d} Y + c,\ Y_n W_n \xrightarrow{d} cY$

4 극한분포의 연산

확률변수 함수의 극한분포

- $\sqrt{n}(Y_n - \mu) \xrightarrow{d} N(0, \sigma^2)$
- $h(x)$가 μ에서 미분가능, $h'(\mu) \neq 0$

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

4 극한분포의 연산

예 $Y_n \sim B(n,p), \hat{p}_n = \frac{Y_n}{n}, Z_n = \frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1-\hat{p}_n)/n}}$의 극한 분포는?

4 극한분포의 연산

예 $Y_n \sim B(n,p), \hat{p}_n = \frac{Y_n}{n}, Z_n = \frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1-\hat{p}_n)/n}}$의 극한 분포는?

정리하기

- $X_1, X_2, \cdots X_m \sim N(\mu_1, \sigma^2)$ 확률표본, $Y_1, Y_2, \cdots Y_n \sim N(\mu_2, \sigma^2)$ 확률표본, σ^2 는 미지일 때 다음이 성립한다.

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

정리하기

- $X_1, X_2, \cdots X_m \sim N(\mu_1, \sigma_1^2)$ 확률표본, $Y_1, Y_2, \cdots Y_n \sim N(\mu_2, \sigma_2^2)$ 확률표본, X_i 와 Y_j 는 독립일 때 다음이 성립한다.

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

정리하기

- $X_1, X_2, \cdots, X_n \sim (\mu, \sigma^2)$ 의 확률표본일 때 $\bar{X}_n$는 μ에 확률적으로 수렴한다.
- $X_1, X_2, \cdots, X_n \sim (\mu, \sigma^2)$ 의 확률표본일 때 다음이 성립한다.

$$Y_n = \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

09 강

다음시간 안내

점추정 1