06 강

통계적 추론

표본분포 1

한국방송통신대학교 통계·데이터과학과
이긍희 교수

학습내용

1. 표본분포의 개념을 이해한다.
2. 확률변수 함수의 확률분포를 구한다.
3. 확률변수 합의 분포를 구한다.
4. 카이제곱 분포와 t분포를 이해한다.

01

표본분포의 개요

1 확률표본

- 확률표본(random sample)
  - **표본** $X_1, X_2, \cdots, X_n$ **을 모집단(동일한 분포)에서 독립적으로 추출**

2 통계적 추론

- 통계량 : 확률표본의 함수
  - **표본평균** : $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - **표본분산** : $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- 표본분포 : 통계량의 분포

02

변수 변환

1 이산형 확률변수 함수의 분포

- 이산형 확률변수 : $X \sim f_X(x)$
  - $Y = u(X)$: **1-1 함수**

    → **역함수** : $x = u^{-1}(y)$
  - $f_Y(y) = f_X[u^{-1}(y)]$

1 이산형 확률변수 함수의 분포

예 $Y = X^2$ 의 확률분포는?

$$f_X(x) = \frac{3}{4}\left(\frac{1}{4}\right)^{x-1}, x = 1,2,\cdots$$

2 연속형 확률변수 함수의 분포

- 연속형 확률변수 : $X \sim f_X(x)$
  - $Y = u(X)$: **1-1 함수 → 역함수** : $x = u^{-1}(y)$
  - $f_Y(y) = f_X[u^{-1}(y)] \left|\frac{dx}{dy}\right|$

2 연속형 확률변수 함수의 분포

- 연속형 확률변수 : $X \sim f_X(x)$
  - $Y = u(X)$: **1-1 함수 → 역함수 :** $x = u^{-1}(y)$
  - $f_Y(y) = f_X[u^{-1}(y)]\left|\frac{dx}{dy}\right|$

2 연속형 확률변수 함수의 분포

예 $X \sim N(0,1)$ 일 때 $Y = X^2$ 의 확률밀도함수는?

2 연속형 확률변수 함수의 분포

예 $X \sim N(0,1)$ 일 때 $Y = X^2$ 의 확률밀도함수는?

2 연속형 확률변수 함수의 분포

예 $X \sim U(0,1)$ 일 때 $Y = -log(1-X)$ 의 확률밀도함수는?

2 연속형 확률변수 함수의 분포

예 $X \sim U(0,1)$ 일 때 $Y = -log(1 - X)$ 의 확률밀도함수는?

2 연속형 확률변수 함수의 분포

- 연속형 확률변수의 누적분포함수 : F
  - $F(X) \sim U(0,1)$

3 결합확률분포

이산형 결합확률분포

- $Y_1 = u_1(X_1, X_2), Y_2 = u_2(X_1, X_2), S \to T$: **1-1 대응변환**
- **결합확률질량함수 :**

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}\left(u_1(x_1, x_2), u_2(x_1, x_2)\right), (y_1, y_2) \in T$$

3 결합확률분포

예 $X_1 \sim B(n_1, p)$, $X_2 \sim B(n_2, p)$, 서로 독립, $X_1 + X_2$의 확률질량함수는?

3 결합확률분포

예 $X_1 \sim B(n_1, p)$, $X_2 \sim B(n_2, p)$, 서로 독립, $X_1 + X_2$의 확률질량함수는?

3 결합확률분포

결합확률밀도함수

- $Y_1 = u_1(X_1, X_2), Y_2 = u_2(X_1, X_2), , S \to T$: **1-1 대응변환**
- **결합확률밀도함수 :**

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}|J|, \quad |J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

3 결합확률분포

예 $X_1,\ X_2 \sim N(0,1)$ 의 확률표본, $Y_1 = X_1 + X_2$과 $Y_2 = X_1 - X_2$의 결합확률밀도함수는?

3 결합확률분포

예 $X_1,\ X_2 \sim N(0,1)$ 의 확률표본, $Y_1 = X_1 + X_2$과 $Y_2 = X_1 - X_2$의 결합확률밀도함수는?

3 결합확률분포

예 $X_1,\ X_2 \sim N(0,1)$ 의 확률표본, $Y_1 = X_1 + X_2$과
$Y_2 = X_1 - X_2$는 독립인가?

3 결합확률분포

예 $X_1,\ X_2 \sim N(0,1)$ 의 확률표본, $Y_1 = X_1 + X_2$과 $Y_2 = X_1 - X_2$는 독립인가?

03

합과 평균의 확률분포

1 확률변수 합의 분포

- 합성곱 (convolution) 공식
  - $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$

1 확률변수 합의 분포

예 $X_1,\ X_2 \sim N(0,1)$, 서로 독립, $X_1 + X_2$의 확률분포는?

1 확률변수 합의 분포

예 X_1, $X_2 \sim N(0,1)$, 서로 독립, $X_1 + X_2$의 확률분포는?

1 확률변수 합의 분포

예 X_1, $X_2 \sim N(0,1)$, 서로 독립, $X_1 + X_2$의 확률분포는?

1 확률변수 합의 분포

기댓값과 분산

- $X_1, X_2, \cdots, X_n \sim (\mu_i, \sigma_i^2)$
- $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mu_i$
- $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \sigma_i^2$

1 확률변수 합의 분포

적률생성함수

- $X_1, X_2, \cdots, X_n\ : M_X(t)$
- $M_{X_1+X_2+\cdots+X_n}(t) = [M_X(t)]^n$
- $M_{\bar{X}}(t) = [M_X(t/n)]^n$

1 확률변수 합의 분포

예 $X_i \sim B(n,p), i = 1,2,$ 서로 독립, $X_1 + X_2$의 확률분포는?

1 확률변수 합의 분포

- 분포별 확률변수들 합의 분포
  - $X_1, X_2, \cdots, X_n$ 서로 독립
  - $X_i \sim B(n_i, p) \Rightarrow \sum_{i=1}^{n} X_i \sim B(\sum_{i=1}^{n} n_i, p)$
  - $X_i \sim Poisson(\lambda_i) \Rightarrow \sum_{i=1}^{n} X_i \sim Poisson(\sum_{i=1}^{n} \lambda_i)$
  - $X_i \sim Gamma(r_i, \lambda) \Rightarrow \sum_{i=1}^{n} X_i \sim Gamma(\sum_{i=1}^{n} r_i, \lambda)$

1 확률변수 합의 분포

정규분포 확률변수 합의 분포

- $X_i \sim N(\mu_i, \sigma_i^2), i = 1,2,\cdots,n,$ 서로 독립
  $\Rightarrow \sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$
- $X_i \sim N(\mu, \sigma^2), i = 1,2,\cdots,n,$ 서로 독립
  $\Rightarrow \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

1 확률변수 합의 분포

예 $X_i \sim N(\mu_i,\ \sigma_i^2), i = 1,2,\cdots,n,$ 서로 독립
$\sum_{i=1}^{n} X_i$의 확률분포는?

2 확률변수 평균의 분포

정규분포 확률변수 평균의 분포

- $X_i \sim N(\mu, \sigma^2), i = 1,2,\cdots,n,$ 서로 독립
  $\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

04

표본분산과 표본평균의 분포

1 카이제곱 분포

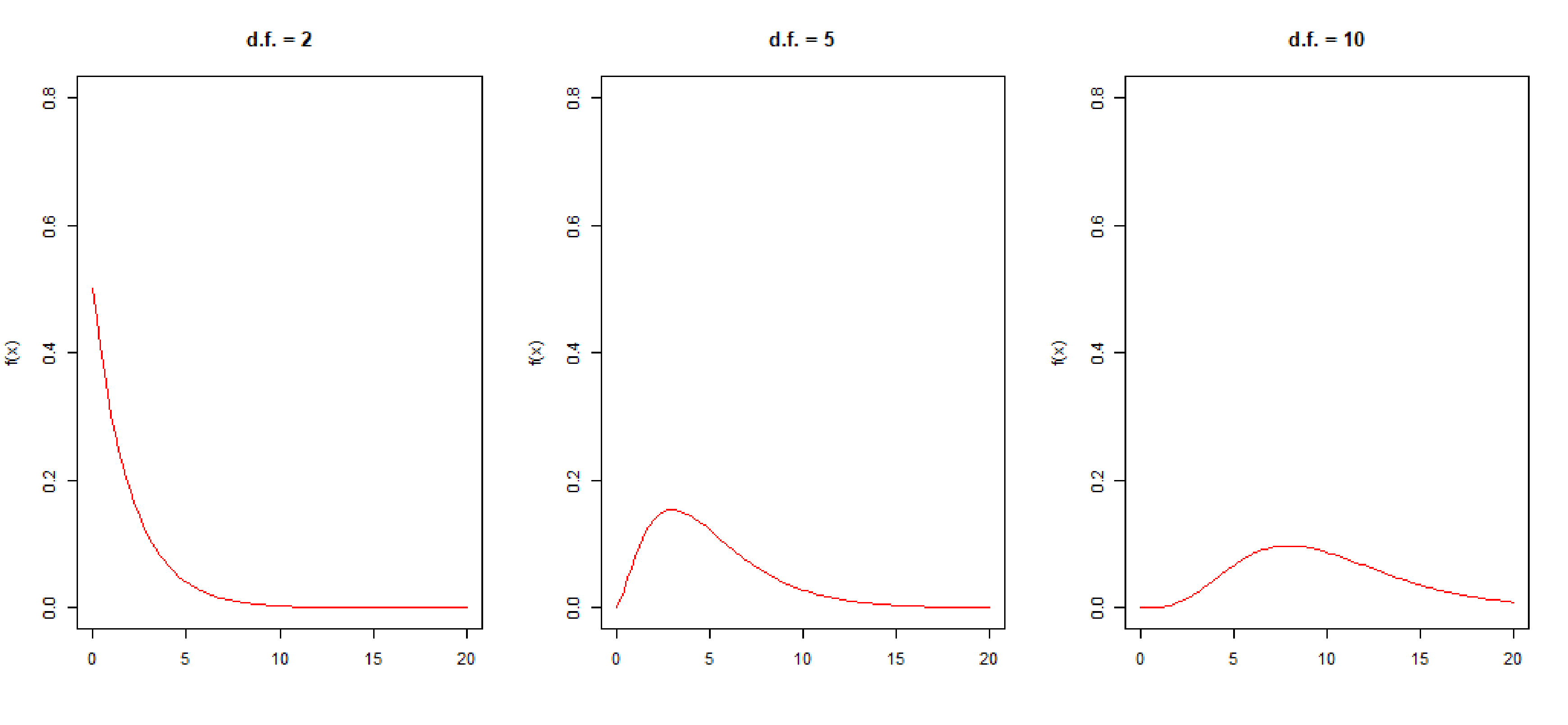
카이제곱분포의 개요

- $X \sim \chi^2(n)$
- $Gamma(\frac{n}{2}, \frac{1}{2}),\ f(x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2}$
- $E(X) = n, Var(X) = 2n$

1 카이제곱 분포

카이제곱분포의 모습

- $X \sim \chi^2(n)$



1 카이제곱 분포

- $\chi^2(n)$ 의 적률생성함수
  - $M(t) = (1-2t)^{-\frac{n}{2}}, \quad t < \frac{1}{2}$

1 카이제곱 분포

정규분포와 카이제곱 분포

- $X_i \sim N(\mu, \sigma^2), i = 1,2,\cdots,n,$ 서로 독립
- $\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(1), \frac{X_i-\mu}{\sigma} \sim N(0,1)$
- $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$

2 표본분산의 분포

표본분산의 분포

- $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$

2 표본분산의 분포

표본분산의 분포

- $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2+\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$

2 표본분산의 분포

표본분산의 분포

- $X_i \sim N(\mu, \sigma^2), i = 1,2,\cdots,n,$ 서로 독립
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

3 표본평균의 분포

표본평균의 분포

- $X_i \sim N(\mu, \sigma^2), i = 1,2,\cdots,n,$ 서로 독립, σ^2 기지
- $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

3 표본평균의 분포

표본평균의 분포

- $X_i \sim N(\mu, \sigma^2), i = 1,2,\cdots,n,$ 서로 독립, σ^2 미지
- $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

3 표본평균의 분포

t 분포

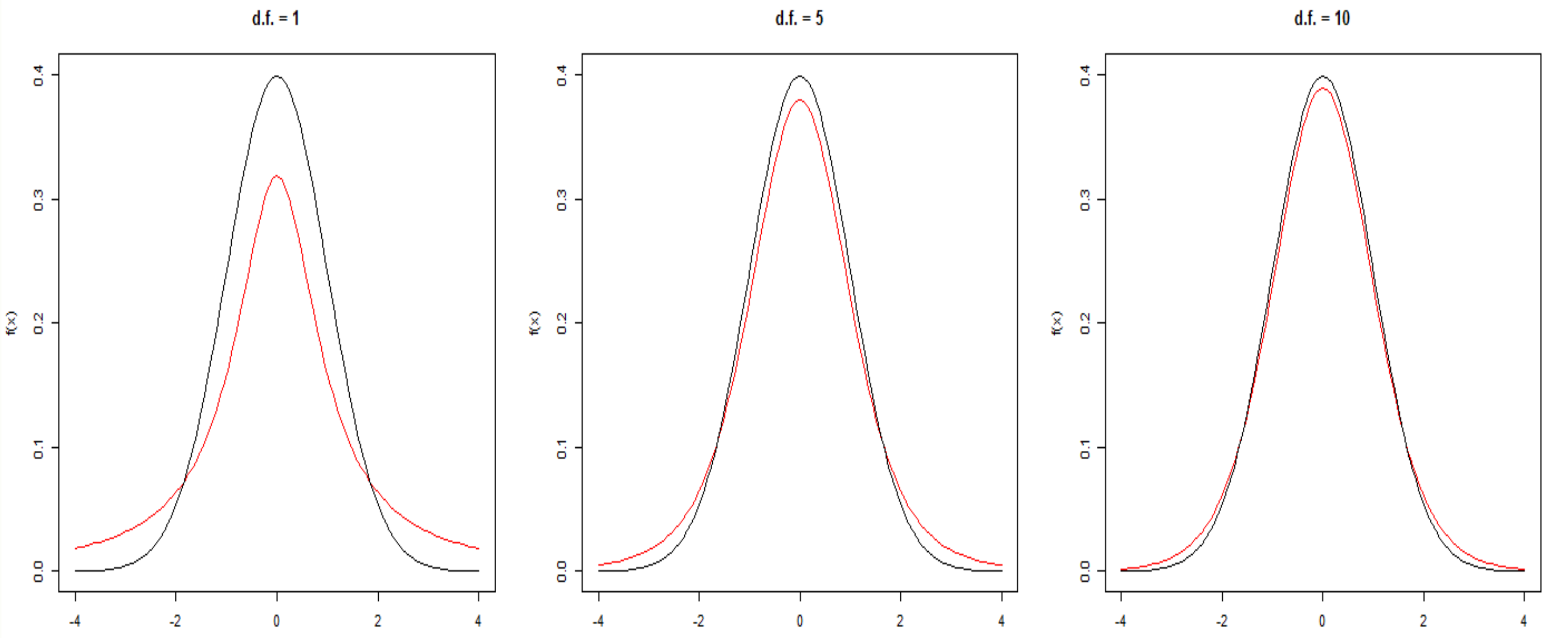
- $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)/(\frac{\sigma}{\sqrt{n}})}{\sqrt{(n-1)S^2/[\sigma^2(n-1)]}} = \frac{Z}{\sqrt{\frac{V}{n-1}}}$

3 표본평균의 분포

- 자유도 $n-1$인 t분포의 확률밀도함수
  - $f(x) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})}\left[1+\frac{x^2}{n-1}\right]^{-n/2}\frac{1}{\sqrt{n-1}}, -\infty < x < \infty$

3 표본평균의 분포

t 분포의 모습



3 표본평균의 분포

t (r) 분포의 기대값

- $E(X) = 0, \qquad r > 1$
- $Var(X) = \frac{r}{r-2}, \; r > 2$

정리하기

- 표본분포는 확률표본의 함수인 통계량의 분포이다.
- 확률변수의 함수에 대한 확률밀도함수는 변수변환법을 통해 구할 수 있다.
- $X_1, X_2, \cdots X_n$ 이 $N(\mu, \sigma^2)$ 를 따르는 확률표본일 때

  표본평균 $\bar{X}$ 은 $N\left(\mu, \frac{\sigma^2}{n}\right)$ 을 따른다.

정리하기

- $X_1, X_2, \cdots X_n$ 이 $N(\mu, \sigma^2)$ 를 따르는 확률표본일 때

- 표본평균 $\bar{X}$ 과 표본분산 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 은 서로 독립

- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(\mathrm{n}-1)$

- $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

07 강

다음시간 안내

표본 분포 2