"통계학의 개념 및 제문제" 제 6 강

- 표본분포의 정의
- 변수변환
- 합과 평균의 확률분포
- 표본분산의 확률분포
- 표준화된 표본평균의 확률분포

통계적 추론의 흐름_개인적인 정리라 틀릴 수 있습니다.

[모집단] $\xrightarrow{\text{확률실험}}$ [표본공간] $\longrightarrow$ [확률변수] $\xrightarrow{\text{확률분포}}$ [확률질량 확률밀도 함수]

1. 관심있는 모집단을 대상으로 확률실험을 하고 그 결과에 따른 표본공간을 정의 한다.
2. 확률변수를 통해 표본공간의 사건을 실수에 매핑한다.
3. 확률변수 값을 관측하여 확률적 분포를 확인하고, 이를 통해 모수를 구하고 모집단의 특성을 파악한다.
4. 그런데, 일반적으로 모집단 전체를 대상으로 확률실험을 할 수가 없고 이때문에 모수를 구할 수 없다.

[모집단] $\xrightarrow{\text{추출}}$ [표본]

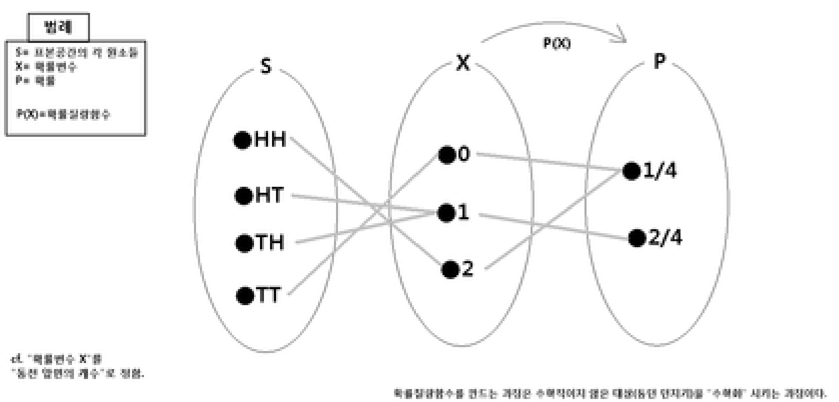
[표본] $\xrightarrow{\text{확률실험}}$ [표본공간] $\longrightarrow$ [확률변수] $\xrightarrow{\text{확률분포}}$ [확률질량 확률밀도 함수]

1. 모집단에 추출한 표본은 모집단과 같은 확률분포를 따른다.
2. 모집단의 분포를 모를 때는 가정(정규분포 등)하거나 표본의 확률변수의 분포를 관측하여 확인한다.
3. 통계적 계산을 통해 모수를 추정할 수 있는 통계량을 구한다.
4. 통계량을 기반으로 통계적 추론 기법에 맞게 모수를 추정한다.
5. 그런데, 모집단의 분포를 가정하지 않아도 평균과 분산등의 모수를 추정할 수 있는 방법이 있다.

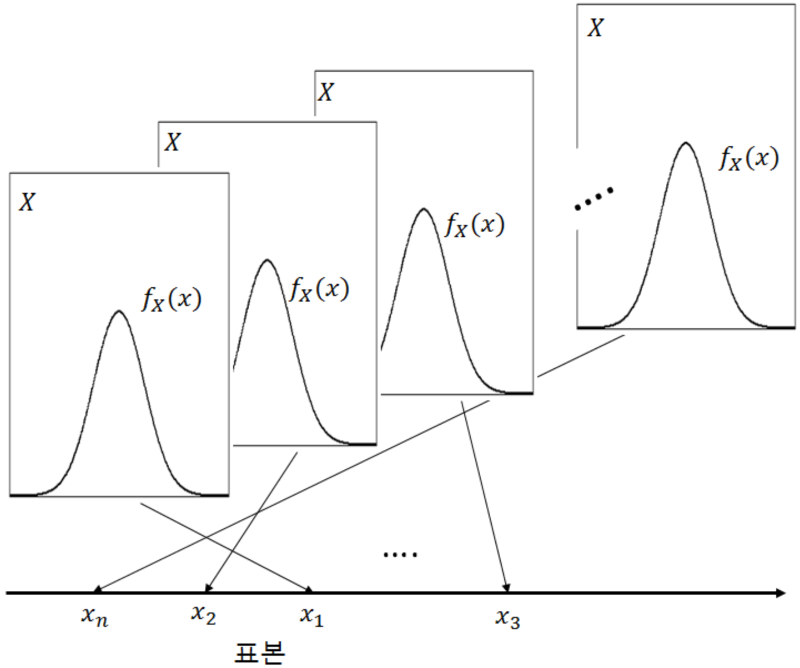
(극한분포, 약대수의 법칙, 중심극한정리 등을 통해)

[확률질량 확률밀도 함수] $\xrightarrow{\text{통계적 계산}}$ [통계량]

[모수추정] $\longleftarrow$ [통계량]



표본분포의 정의



표본조사

- 표본조사
  : 모집단 전체를 조사하는 것이 불필요하거나 불가능할 때
- 표본 : 모집단의 일부
  – 표본을 이용해 모집단의 모수를 추정

통계추론

- 확률변수 X 의 모집단 : 모수 θ 를 바탕으로 한 확률분포
  – 확률질량함수 또는 확률밀도함수 $f\left(x|\theta\right)$
- 표본추출 : 모수 θ 를 추정
  – $\boxed{\text{표본 } X_1, X_2, \cdots, X_n}$ 을 추출
  – 확률표본 : 표본은 서로 독립이고 동일 분포

확률변수!!!!

크기가 n인 표본이란, 모집단에서 n개의 확률변수를 추출한 것이다.

- 통계량 : 모수추정에 적합한 확률표본의 함수
  – 표본분포 : 통계량의 확률분포

통계량의 확률표본의 함수

즉, 확률변수의 함수임으로 통계량을 알기위해서는

확률변수의 변환과 함수계산이 필요하다.

실제 계산에서는 변수의 실제값을 넣어 계산

- 표본평균과 표본분산 : 모집단의 평균과 분산을 추정하는 데 적합
  – 표본평균 : $\overline{X} = \dfrac{1}{n}(X_1 + X_2 + \cdots + X_n) = \dfrac{1}{n}\sum_{i=1}^{n} X_i$
  – 표본분산 : $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
  – 표본 표준편차 : $S = \sqrt{S^2} = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$

변수변환

이산형 확률변수 함수의 분포

- 이산형 확률변수 X : 확률질량함수 $f_X(x)$
- $Y=u(X)$: $u(X)$ 일대일 함수
  - 역함수 $x=u^{-1}(y)$
- Y 의 확률질량함수 : $f_Y(y)$

$$f_Y(y)=P(Y=y)=P\left[X=u^{-1}(y)\right]=f_X\left[u^{-1}(y)\right]$$

X, Y는 확률변수, x, y는 각 변수의 값.

$Y=u(X) \longleftrightarrow X=u^{-1}(Y)$

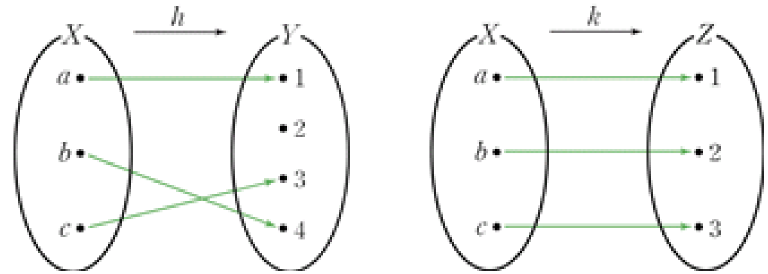
$y=u(x) \longleftrightarrow x=u^{-1}(y)$

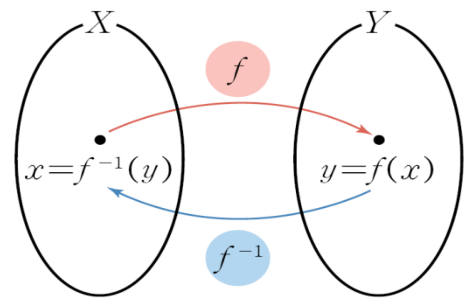
변수변환

$\underline{f_Y(y)} = P(Y=y) = P[X=u^{-1}(y)] = \underline{f_X(u^{-1}(y))}$

해당값 변환.

Y의 함수를 X의 함수로 표현.





예 4.1

$Y=X^2$ 의 확률분포는?

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

$y=x^2 \longleftrightarrow x=\sqrt{y}$ ($\because x=1,2,3 \cdots$)

$\therefore f_Y(y)=f_X(\sqrt{y})=\frac{3}{4}\cdot\left(\frac{1}{4}\right)^{\sqrt{y}-1}, \quad y=1,4,9\cdots$

변수변환

연속형 확률변수 함수의 분포

- 연속형 확률변수 X : 확률밀도함수 $f_X(x)$
- $Y = u(X)$: $u(X)$ 일대일 함수
- Y 의 확률밀도함수 : $f_Y(y)$

$$f_Y(y) = f_X\left[u^{-1}(y)\right]\boxed{\left|\frac{dx}{dy}\right|}$$

연속형에서 필요한 일종의 보정계수,
X, Y의 공간이 다음으로 확률질량함수
의 성질을 만족시키기 위해…

$Y = 2X - 3$ 의 확률밀도함수는?

$$f_X(x) = \begin{cases} \dfrac{x}{12}, & 1 < x < 5 \\ 0, & \text{그 밖에} \end{cases}$$

$y = 2x - 3 \longleftrightarrow x = \dfrac{y+3}{2}, \quad \dfrac{dx}{dy} = \dfrac{1}{2}$

$$f_Y(y) = f_X[u^{-1}(y)]\left|\frac{dx}{dy}\right|$$

$$= f_X\left(\frac{y+3}{2}\right)\cdot\frac{1}{2}$$

$$= \frac{(y+3)/2}{12}\cdot\frac{1}{2}$$

$$= \frac{y+3}{48}, \quad -1 < y < 7.$$

$y = u(x) \longleftrightarrow x = u^{-1}(y)$

$u(\cdot)$ 가 증가함수;

$$F_Y(y) = P(Y \le y) = P[X \le u^{-1}(y)]$$

$$= \int_{-\infty}^{u^{-1}(y)} f_X(x)\,dx \quad —① $$

($f(x) \underset{\text{미분}}{\overset{\text{적분}}{\rightleftarrows}} F(x)$,

아마도, ①은 $F(x)\big|_{x=u^{-1}(y)} - F(x)\big|_{x=-\infty}$)

①를 y로 미분

$$f_Y(y) = f(u^{-1}(y))\frac{d(u^{-1}(y))}{dy} = f(u^{-1}(y))\cdot\frac{dx}{dy} \quad —②$$

$u(\cdot)$ 가 감소함수;

$$F_Y(y) = P(Y \le y) = P[X \ge u^{-1}(y)]$$

$$= \int_{u^{-1}(y)}^{\infty} f_X(x)\,dx \quad —③$$

(아마도, ③는 $\underbrace{F(x)\big|_{x=\infty}}_{\text{미분하면 } 0} - F(x)\big|_{x=u^{-1}(y)}$)

③을 y로 미분

$$f_Y(y) = -f(u^{-1}(y))\frac{dx}{dy} \quad —④$$

부호가 − (감소함수)

②, ④를 함께 쓰면,

$$f_Y(y) = f(u^{-1}(y))\left|\frac{dx}{dy}\right|$$

변수변환

예 4.3

확률변수 X 가 $N(0,1)$ 을 따를 때 $Y=X^2$ 의 확률밀도함수는?

$$P(Y\le a) = P(X^2 \le a)$$
$$= P(-\sqrt{a} \le X \le \sqrt{a})$$
$$= \int_{-\sqrt{a}}^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx$$

여기서,

$$y = x^2 \longleftrightarrow x = \begin{cases} -\sqrt{y}, & x \le 0 \\ \sqrt{y}, & x > 0 \end{cases}$$

$$\frac{dx}{dy} = \begin{cases} -\frac{1}{2\sqrt{y}}, & x \le 0 \\ \frac{1}{2\sqrt{y}}, & x > 0 \end{cases}$$

$$\therefore \int_{-\sqrt{a}}^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx = \int_{-\sqrt{a}}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx + \int_{0}^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx$$

(변수변환)

$$= \int_{-a}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}\left(-\frac{1}{2\sqrt{y}}\right)dy + \int_{0}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}\left(\frac{1}{2\sqrt{y}}\right)dy$$

$x = -\sqrt{y}$

$x = -\sqrt{a} \longleftrightarrow y = -a$. y에 대한 적분이 되어야 함

적분값이 같다

$$= \int_{0}^{a} \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}\,dy$$

χ^2의 확률밀도함수

$$\therefore g(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \le 0 \end{cases}$$

감마함수, 감마분포 관련…

- $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1}\,dy$
- $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$
- $\alpha > 0$: $\Gamma(\alpha) = (\alpha-1)!$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

보조정리 1 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ 이다.

증명 감마함수 $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\,dy$로부터

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty y^{-\frac{1}{2}} e^{-y}\,dy = \int_0^\infty \frac{1}{\sqrt{y}} e^{-y}\,dy$$

이다. 여기서 $\sqrt{y} = \frac{z}{\sqrt{2}}$로 놓으면 $\frac{1}{2\sqrt{y}}dy = \frac{1}{\sqrt{2}}dz$이므로

$$\Gamma(\frac{1}{2}) = \int_0^\infty e^{-\frac{z^2}{2}} \sqrt{2}\,dz = \sqrt{2}\sqrt{2\pi}\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz$$

$$= 2\sqrt{\pi}\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz = \sqrt{\pi}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz$$

$$= \sqrt{\pi}$$

변수변환

함수의 결합확률밀도함수

$Y_1 = u_1(X_1, X_2)$ 와 $Y_2 = u_2(X_1, X_2)$ 의 결합확률밀도함수

$$f(y_1, y_2) = f(x_1, x_2)|J| = f(x_1, x_2)\boxed{\begin{vmatrix} dx_1/dy_1 & dx_1/dy_2 \\ dx_2/dy_1 & dx_2/dy_2 \end{vmatrix}}$$

야코비안(혹은 자코비안) 행렬식,

연속형 변수변환의 확장.

선형대수등에서 변수의 선형변환을 위해 사용한다.

예 4.4

$Y_1 = X_1^2$ 과 $Y_2 = X_1 X_2$ 의 결합확률밀도함수는?

$$f(x_1, x_2) = \begin{cases} 4x_1x_2, & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0, & \text{그 밖에} \end{cases}$$

$y_1 = x_1^2$, $y_2 = x_1 x_2$ $\longleftrightarrow$ $x_1 = \sqrt{y_1}$, $x_2 = \dfrac{y_2}{\sqrt{y_1}}$

y_1 미분: $\dfrac{dx_1}{dy_1} = \dfrac{1}{2\sqrt{y_1}}$, y_2 미분: $\dfrac{dx_1}{dy_2} = 0$

y_1 미분: $\dfrac{dx_2}{dy_1} = -\dfrac{1}{2}\cdot\dfrac{y_2}{y_1^{3/2}}$, y_2 미분: $\dfrac{dx_2}{dy_2} = \dfrac{1}{\sqrt{y_1}}$

$$|J| = \begin{vmatrix} dx_1/dy_1, & dx_1/dy_2 \\ dx_2/dy_1, & dx_2/dy_2 \end{vmatrix} = \begin{vmatrix} 1/2\sqrt{y_1}, & 0 \\ -y_2/2y_1^{3/2}, & \dfrac{1}{\sqrt{y_1}} \end{vmatrix}$$

행렬식 계산.

$$= \frac{1}{2\sqrt{y_1}}\cdot\frac{1}{\sqrt{y_1}} = \frac{1}{2y_1}$$

$$\therefore f(y_1, y_2) = f(x_1, x_2)|J|$$

$$= f(\sqrt{y_1}, y_2/\sqrt{y_1})\cdot\frac{1}{2y_1}$$

$$= 4\cdot\sqrt{y_1}\cdot\frac{y_2}{\sqrt{y_1}}\cdot\frac{1}{2y_1} = \frac{2y_2}{y_1},$$

여기서, $y_2^2 < y_1 < 1$, $0 < y_2 < 1$

합과 평균의 확률분포

적률생성함수의 성질

① $\boxed{M_X(t) = M_Y(t) \Leftrightarrow f_X(x) = f_Y(y)}$ 적률생성함수의 유일성!!!

② $X_1, X_2, \cdots, X_n$ 서로 독립, 적률생성함수 : $M_{X_i}(t)$

– $M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)\cdot \cdots \cdot M_{X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t)$

③ $X_1, X_2, \cdots, X_n$ 서로 독립, 적률생성함수 : $M_X(t)$

– $M_{X_1+X_2+\cdots+X_n}(t) = [M_X(t)]^n$

– $M_{\bar{X}}(t) = [M_X(t/n)]^n$

②
$$M_{X_1+X_2+\cdots+X_n}(t) = E(e^{t(X_1+X_2+\cdots+X_n)})$$
$$= E(e^{tX_1}\cdot e^{tX_2}\cdot \cdots \cdot e^{tX_n})$$
$$= E(e^{tX_1})\cdot E(e^{tX_2})\cdots E(e^{tX_n})$$
$$= \prod_{i=1}^{n} M_{X_i}(t)$$

같은 확률분포, $M_{X_1} = M_{X_2} = \cdots = M_{X_n} = M_X(t)$

$$\therefore M_{X_1+X_2+\cdots+X_n} = [M_X(t)]^n.$$

③
$$M_{\bar{X}}(t) = M_{\left(\frac{X_1+X_2+\cdots X_n}{n}\right)}(t)$$
$$= E(e^{t(\frac{X_1}{n}+\frac{X_2}{n}+\cdots+\frac{X_n}{n})})$$

새로운 t

$$= \underbrace{E(e^{\frac{t}{n}X_1})}_{M_X(\frac{t}{n})}\cdot E(e^{\frac{t}{n}X_2})\cdots E(e^{\frac{t}{n}X_n})$$

$$\therefore M_{\bar{X}}(t) = [M_X(\tfrac{t}{n})]^n$$

합과 평균의 확률분포

예 4.5

X_1, X_2 서로 독립, $B(n_i, p), i = 1, 2$

$X_1 + X_2$ 의 확률분포는?

예 4.6

X_1, X_2 서로 독립, $Poisson(\lambda_i)\ \ i = 1, 2$

$X_1 + X_2$ 의 확률분포는?

예 4.7

X_1, X_2 서로 독립, $Gamma(r_i, \lambda)\ \ i = 1, 2$

$X_1 + X_2$ 의 확률분포는?

$X_1, X_2, \cdots, X_n$ 이 서로 독립

① $X_i \sim B(n_i, p) \Rightarrow X_1 + X_2 + \cdots + X_n \sim B\left(\sum_{i=1}^{n} n_i, p\right)$

② $X_i \sim Poisson(\lambda_i) \Rightarrow X_1 + X_2 + \cdots + X_n \sim Poisson\left(\sum_{i=1}^{n} \lambda_i\right)$

③ $X_i \sim Gamma(r_i, \lambda) \Rightarrow X_1 + X_2 + \cdots + X_n \sim Gamma\left(\sum_{i=1}^{n} r_i, \lambda\right)$

$X \sim B(n, p)$

$M_X(t) = [(1-p) + pe^t]^n$

$M(t) = M_{X_1}(t) \cdot M_{X_2}(t)$

$= [(1-p) + pe^t]^{n_1} \cdot [(1-p) + pe^t]^{n_2}$

$= [(1-p) + pe^t]^{\underline{n_1+n_2}}$ 새로운 n

$\therefore X_1 + X_2 \sim B(n_1 + n_2, p)$

$X \sim Poisson(\lambda)$

$M_X(t) = e^{\lambda(e^t - 1)}$

$M(t) = M_{X_1}(t) \cdot M_{X_2}(t)$

$= \exp[\lambda_1(e^t - 1)] \exp[\lambda_2(e^t - 1)]$

$= \exp[\underbrace{(\lambda_1 + \lambda_2)}_{\text{새로운 } \lambda}(e^t - 1)]$

$\therefore X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2)$

$X \sim Gamma(r, \lambda)$

$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$

$M(t) = M_{X_1}(t) \cdot M_{X_2}(t)$

$= \left(\frac{\lambda}{\lambda - t}\right)^{r_1} \left(\frac{\lambda}{\lambda - t}\right)^{r_2}$

$= \left(\frac{\lambda}{\lambda - t}\right)^{\underline{r_1 + r_2}}$ 새로운 r.

$\therefore X_1 + X_2 \sim Gamma(r_1 + r_2, \lambda)$

합과 평균의 확률분포

예 4.8

X_1, X_2 서로 독립, $N(\mu_i, \sigma_i^2)$ $i = 1, 2$

$X_1 + X_2$ 의 확률분포는?

$X \sim N(\mu, \sigma^2)$

$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

$M(t) = M_{X_1}(t) \cdot M_{X_2}(t)$

$= \exp(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2) \cdot \exp(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2)$

$= \exp(\underbrace{(\mu_1 + \mu_2)}_{\text{새로운 } \mu} + \frac{1}{2}\underbrace{(\sigma_1^2 + \sigma_2^2)}_{\text{새로운 } \sigma^2} t^2)$

$\therefore X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

예 4.9

$X_i \sim N(\mu, \sigma^2)$ 서로 독립, $X_1 + \cdots + X_n$ 의 확률분포는?

$M(t) = E(e^{t(X_1 + X_2 + \cdots + X_n)})$

$= E(e^{tX_1}) E(e^{tX_2}) \cdots E(e^{tX_n})$

$= [\exp(\mu t + \frac{1}{2}\sigma^2 t^2)]^n$

$= \exp\{\underbrace{(\mu t + \frac{1}{2}\sigma^2 t^2) + (\mu t + \frac{1}{2}\sigma^2 t^2) + \cdots + (\mu t + \frac{1}{2}\sigma^2 t^2)}_{n\text{개}}\}$

$= \exp(\underbrace{n\mu}_{\text{새로운 } \mu} t + \frac{1}{2}\underbrace{n\sigma^2}_{\text{새로운 } \sigma^2} t^2)$

$\therefore X_1 + X_2 + \cdots X_n \sim N(n\mu, n\sigma^2)$

정규분포 확률변수 합의 확률분포

① $X_1, X_2, \cdots, X_n$ 독립, $N(\mu_i, \sigma_i^2)$

– $X_1 + \cdots + X_n \sim N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$

② $X_1, X_2, \cdots, X_n$ 독립, 평균과 분산이 μ 와 σ^2 으로 동일

– $X_1 + \cdots + X_n \sim N(n\mu, n\sigma^2)$

합과 평균의 확률분포

정규분포 확률변수 평균의 확률분포

- $X_1, X_2, \cdots, X_n$ 서로 독립, $N(\mu, \sigma^2)$

① $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

② $Z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$$M_{\overline{X}}(t) = \left[M\left(\frac{t}{n}\right)\right]^n$$

$$= \exp\Big\{\underbrace{\left(\mu\frac{t}{n} + \frac{1}{2}\sigma^2\frac{t^2}{n^2}\right) + \left(\mu\frac{t}{n} + \frac{1}{2}\sigma^2\frac{t^2}{n^2}\right) + \cdots + \left(\mu\frac{t}{n} + \frac{1}{2}\sigma^2\frac{t^2}{n^2}\right)}_{n\text{개.}}\Big\}$$

$$= \exp\left(\mu t + \frac{1}{2}\cdot\underbrace{\frac{\sigma^2}{n}}_{\text{새로운 } \sigma^2} t^2\right)$$

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

표본분산의 확률분포

카이제곱(χ^2)분포

- 모분산의 추정, 적합도검정(goodness-of-fit test) 수행, 교차표(contingency table) 관련 검정 등에 이용

카이제곱(χ^2)분포

- 자유도(d.f.) n 인 카이제곱분포 $X \sim \chi^2(n)$
  - $r = n/2,\ \lambda = 1/2$ 인 감마분포
  - 확률밀도함수

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x},\ x > 0$$

* $X \sim N(0,1)$ $Y = X^2$ 의 확률밀도함수.
(예 4.3)

$$f(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} \qquad \boxed{\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}}$$

$$= \frac{1}{\Gamma(\frac{1}{2})} \cdot \frac{1}{\sqrt{2y}} \cdot e^{-\frac{1}{2}y}$$

$$= \underbrace{\frac{1}{\Gamma(\frac{1}{2})} \cdot \left(\frac{1}{2}\right)^{\frac{1}{2}} \cdot y^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}y}}_{n=1 \text{ 인 } \chi^2 \text{ 분포!!!}}$$

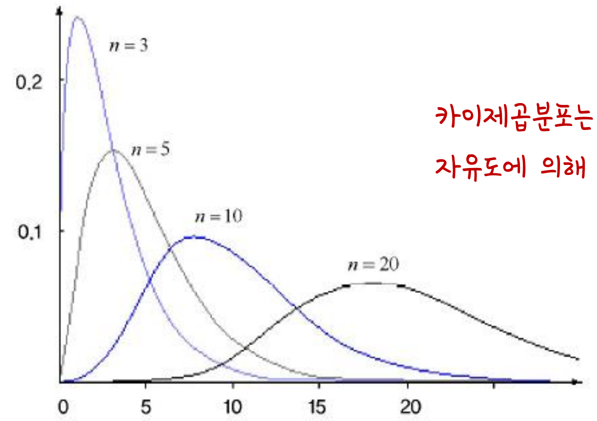
※ 감마분포.

$$f(x) = \frac{\lambda^r}{\underbrace{(r-1)!}_{\text{감마함수}}} x^{r-1} e^{-\lambda x}$$

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$$

$r = \frac{n}{2}$, $\lambda = \frac{1}{2}$ 대입.

$$f(x) = \frac{1}{\Gamma(\frac{n}{2})} \cdot \left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} \rightarrow \chi^2 \text{분포.}$$



카이제곱분포는 비대칭분포이며 비대칭정도는 자유도에 의해 결정된다.

- $\chi^2(n)$ 확률변수 적률생성함수

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

- 기대값 : $E(X) = n$
- 분 산 : $Var(X) = 2n$

$$E(X) = M'(t=0) = -\frac{n}{2} \cdot (-2) \cdot (1-2t)^{-\frac{n}{2}-1}$$

$$= n \cdot (1-2t)^{-\frac{n}{2}-1} \Big|_{t=0}$$

$$= n$$

$$E(X^2) = M''(t=0) = n \cdot \left(-\frac{n}{2}-1\right) \cdot (-2) \cdot (1-2t)^{-\frac{n}{2}-2}$$

$$= (n^2+2n)(1-2t)^{-\frac{n}{2}-2} \Big|_{t=0}$$

$$= n^2+2n.$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$= n^2+2n-n^2 = 2n.$$

표본분산의 확률분포

카이제곱(χ^2)분포의 특성

- $X_1, X_2, \cdots, X_n$ 서로 독립, $N(\mu, \sigma^2)$

① $\left(\dfrac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(1)$, 여기서 $\dfrac{X_i-\mu}{\sigma} \sim N(0,1)$

② $\boxed{\left(\dfrac{X_1-\mu}{\sigma}\right)^2 + \left(\dfrac{X_2-\mu}{\sigma}\right)^2 + \cdots + \left(\dfrac{X_n-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)}$

카이제곱분포의 가법성!!!

*가법성에 대해

$$\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{\sim\chi(n)} = \underbrace{\sum_{i=1}^{n-1}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{\sim\chi(n-1)} + \underbrace{\left(\frac{X_n-\mu}{\sigma}\right)^2}_{\chi(1).}$$

표본분산의 분포

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

표본분산의 확률분포

- $X_1, X_2, \cdots, X_n$ 서로 독립, $N(\mu, \sigma^2)$

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

*표본분산분포를 알기위해.

$X \sim N(\mu, \sigma^2)$ 에서 n개의 표본 추출.

먼저, $\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}[(X_i-\overline{X}) + (\overline{X}-\mu)]^2$

$\overline{X}$를 -, +.

$$\therefore \sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2 + n(\overline{X}-\mu)^2$$

(나머지는 Σ 안에서 0).

양변을 σ^2 으로 나누면,

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \underbrace{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}}_{①} + \underbrace{\frac{n(\overline{X}-\mu)^2}{\sigma^2}}_{②}$$

$$① \rightarrow \frac{(n-1)}{\sigma^2}\underbrace{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{(n-1)}}_{=S^2} = \frac{(n-1)S^2}{\sigma^2}$$

$$② \rightarrow \frac{(\overline{X}-\mu)^2}{\frac{\sigma^2}{n}} = \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

정리하면,

$$\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{W\sim\chi^2(n)} = \underbrace{\frac{(n-1)S^2}{\sigma^2}}_{U} + \underbrace{\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2}_{V\sim\chi^2(1)}$$

χ^2 분포의 가법성에 의해 $\therefore \boxed{\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)}$

*적률생성 함수로 증명.

$$M_W(t) = M_U(t)\cdot M_V(t)$$

$$(1-2t)^{-\frac{n}{2}} = M_U(t)\cdot(1-2t)^{-\frac{1}{2}}$$

$$\therefore M_U(t) = (1-2t)^{-\frac{n-1}{2}}$$

적률생성 함수의 유일성에 의해

$$U \sim \chi^2(n-1)$$

표준화된 표본평균의 확률분포

표본평균의 분포

- X_i 서로 독립, $N(\mu, \sigma^2)$: $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

- 모분산 σ^2 을 모를 경우 : $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$

- 1908년 고셋 : $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t$ 분포

t 분포

- $$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{(\overline{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)S^2/(\sigma^2(n-1))}}$$
$$= \frac{Z}{\sqrt{V/n-1}} \sim t(n-1)$$

t 분포

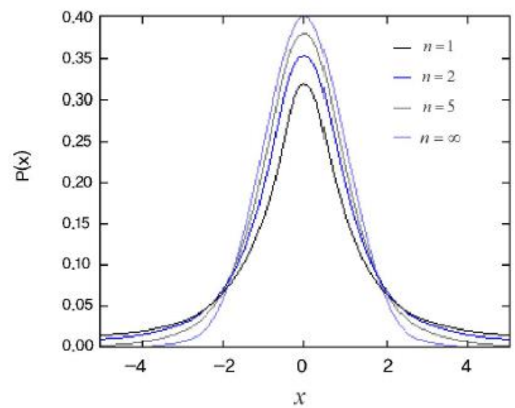
- 자유도(df) $n-1$ 인 t 분포의 확률밀도함수

$$f(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}\left(1+\frac{x^2}{n-1}\right)^{-\frac{n}{2}}\frac{1}{\sqrt{n-1}}, -\infty < x < \infty$$

t 분포

- $X_1, X_2, \cdots, X_n$ 서로 독립, $N(\mu, \sigma^2)$, σ^2 을 알 수 없을 때

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



표준정규분포처럼 평균값이 0에 대하여
좌우대칭이고 종모양이며,
자유도에 의해 그 분포 모양이 결정된다.
자유도가 증가함에 따라 t분포는
표준정규분포에 가까워진다.

$$T = \frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\frac{s}{\sqrt{n}\cdot(\sigma/\sqrt{n})}}$$

$$= \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{s^2(n-1)}{\sigma^2(n-1)}}} = \frac{Z}{\sqrt{V/n-1}} \sim t(n-1).$$

$Z \sim N(0,1)$

$V \sim \chi^2(n-1)$

옆의 식을 변수변환과 계산을 하면
t분포의 확률밀도함수가 구해진다.
t분포는 표준정규분포의 자유도가 (n−1)인
카이제곱 분포의 일종의 비례값으로
나타난다.