“통계학의 개념 및 제문제” 제 7 강

- 표본분포
- 표본평균차의 확률분포
- 표본분산비의 확률분포
- 극한분포(약대수 법칙, 중심극한정리)

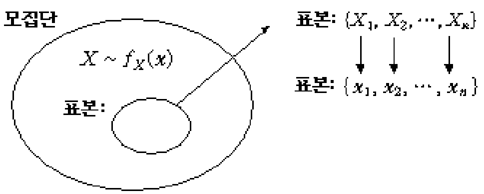
표본의 개념

표본 1 : 모집단의 일부로서 모집단으로부터 실제로 얻어낸 구체적인 자료

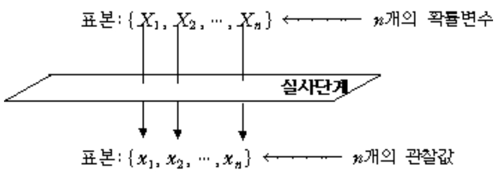
$$\{x_1, x_2, \cdot \cdot \cdot, x_n\}$$

표본 2 : 통계학의 이론을 전개하기 위한 "변수로서의 표본"

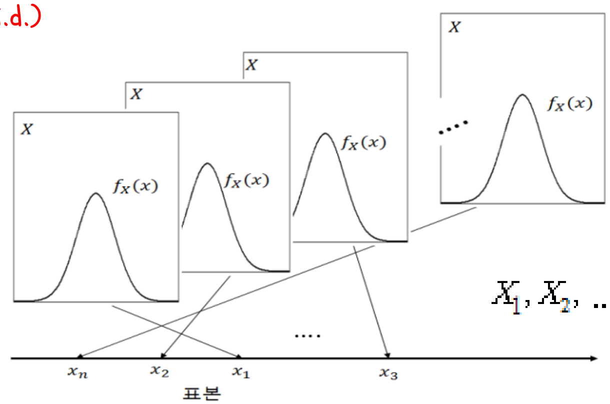
$$\{X_1, X_2, \cdot \cdot \cdot, X_n\}$$



표본 1 $\{x_1, x_2, \cdot \cdot \cdot, x_{1000}\}$ 은 표본 2 $\{X_1, X_2, \cdot \cdot \cdot, X_n\}$의 구현된 값!



만약 수능응시자 50만명의 성적(X:모집단의 확률변수)이 정규분포를 따른다고 하면, 모집단을 구성하는 X값이 50만개 존재하고 몇 점인지는 모르는 상태, n개의 표본을 얻을 경우 $\{X_1, X_2, \cdot \cdot \cdot, X_n\}$로 표현하는데 50만개의 X중 n개의 X를 얻는다는 의미이고, 당연히 $\{X_1, X_2, \cdot \cdot \cdot, X_n\}$도 정규분포를 따르며, 서로 영향을 주지 않는다. 똑같이 독립적으로 분포한다.(i.i.d.)



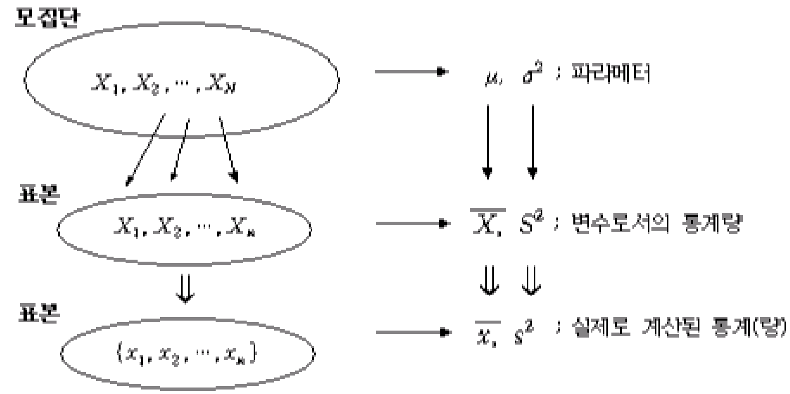
$$X_1, X_2, \ldots, X_r \sim i.i.d. f_x(x)$$

통계량 : 표본으로부터 얻어진 어떤 것.

즉, $\{x_1, x_2, \cdot \cdot \cdot, x_n\}$ 또는 $\{X_1, X_2, \cdot \cdot \cdot, X_n\}$으로부터 얻어진 결과

$\overline{X}, \overline{x}, S^2, s^2, X_1, (x_1 + x_2)/2, ..$ 등 무수히 많다.

여기서 $\{X_1, X_2, \cdot \cdot \cdot, X_n\}$로부터 얻어진 통계량($\overline{X}, S^2, X_1$ 등)은 하나의 또다른 확률변수로 어떤 분포를 한다.



예를들어 표본평균($\overline{X}$)의 값은 표본이 어떻게 얻어졌느냐에 따라 다르게 된다. 1000개의 개체로 구성된 모집단에서 100개의 표본을 뽑는 경우 ${}_{1000}C_{100}$ 의 경우의 수가 생기고 이에 따라 ${}_{1000}C_{100}$ 개의 표본평균값들이 있을 수 있다. 때문에 표본평균값들의 분포가 있게 되고

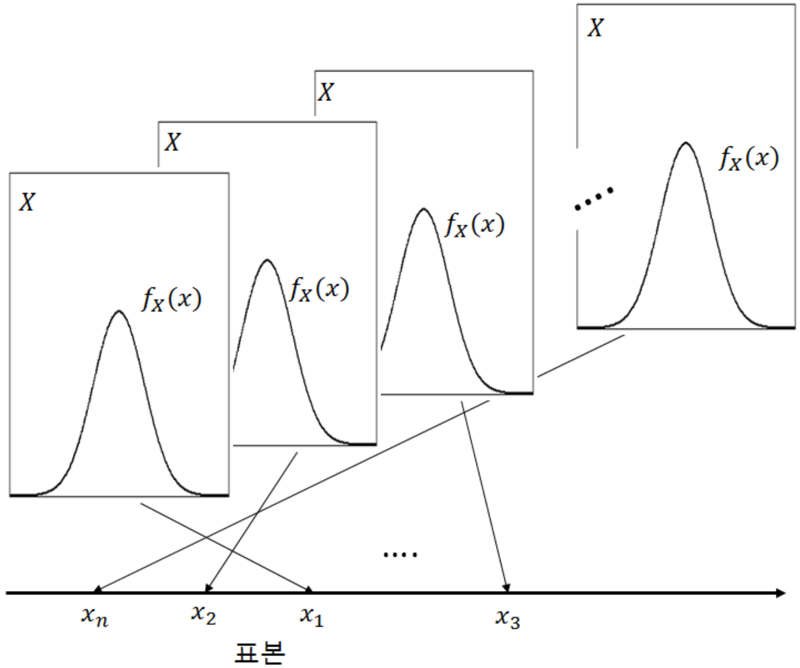
이를 "표본평균의 분포" (정확하게 표현하면 "표본평균의 표본분포-sampling distribution of sample means-")라 하는 것이다.

물론, 실제로는 하나의 표본 $\{x_1, x_2, \cdot \cdot \cdot, x_n\}$에서 그에 대응하는 하나의 표본평균값 $\overline{x}$ 을 얻게 되는데, 이는 분포하는 $\overline{X}$ 의 특정 실제값이다.

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
  – 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 : $\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$
  – 표본분산 : $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
  – 표본 표준편차 : $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$

표본평균차의 확률분포

표본평균의 차이에 대한 분포

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

* 사전지식.

$$Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y).$$

$$X_1 + X_2 + \cdots + X_n \sim N\left(\sum_{i=1}^{n} \mu_i , \sum_{i=1}^{n} \sigma_i^2\right)$$

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

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

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

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

$\chi(n-1)$ 을 따르는 확률변수.

* 표본평균차의 확률분포.

$$\overline{X} \sim N\left(\mu_1 , \frac{\sigma_1^2}{m}\right) \qquad \overline{Y} \sim N\left(\mu_2 , \frac{\sigma_2^2}{n}\right)$$

$$\therefore \overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2 , \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)$$

표본평균차의 확률분포

표본평균의 차이에 대한 분포

- m 개 X_i 독립, $N(\mu_1, \sigma_1^2)$, n개 Y_j 독립, $N(\mu_2, \sigma_2^2)$, X_i 와 Y_j 가 독립, 분산 동일
  - $\overline{X}-\overline{Y} \sim N(\mu_1-\mu_2\,, \sigma^2[1/m+1/n])$
  - $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}} \sim N(0,1)$

합동표본분산

$$S_p^2=\frac{1}{m+n-2}\left\{\sum_{i=1}^{m}(X_i-\overline{X})^2+\sum_{j=1}^{n}(Y_j-\overline{Y})^2\right\}$$

$$\frac{(m+n-2)S_p^2}{\sigma^2}\sim\chi^2(m+n-2)$$

표본평균의 차이에 대한 분포

- 두 집단의 모집단 분산을 모르지만 같다고 가정

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

(1) $\sigma_1^2=\sigma_2^2=\sigma^2$, σ^2을 알 때,

표준화 $\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim N(0,1)$ — ①

※ 합동표본분산(S_p^2)은 σ^2의 추정량. 통계량.

$$S_p^2=\frac{1}{m+n-2}\left\{\sum_{i=1}^{m}(X_i-\bar{X})^2+\sum_{j=1}^{n}(Y_i-\bar{Y})^2\right\}$$

$$S_p^2=\frac{1}{m+n-2}\left\{(m-1)S_1^2+(n-1)S_2^2\right\}$$

양변을 σ^2으로 나누고 정리.

$$\frac{S_p^2(m+n-2)}{\sigma^2}=\underbrace{\frac{(m-1)S_1^2}{\sigma^2}}_{\sim\chi^2(m-1)}+\underbrace{\frac{(n-1)S_2^2}{\sigma^2}}_{\chi^2(n-1)}$$

∴ χ^2 분포의 가법성에 의해 $\dfrac{S_p^2(m+n-2)}{\sigma^2}\sim\chi^2(m+n-2)$ — ②

(2) σ^2을 모를 때,

①, ②에 의해.

$$T=\frac{Z}{\sqrt{V/n-1}}=\frac{\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\cancel{\sigma}\sqrt{\frac{1}{m}+\frac{1}{n}}}}{\sqrt{\dfrac{S_p^2\cancel{(m+n-2)}}{\cancel{\sigma^2}\cancel{(m+n-2)}}}}$$

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

표본평균차의 확률분포

예 4.12 $X_1, \cdots, X_9 \sim N(\mu_1, \sigma^2)$, $Y_1, \cdots, Y_{16} \sim N(\mu_2, \sigma^2)$

$$P\left[-\frac{S_p}{2} \le (\overline{X}-\overline{Y})-(\mu_1-\mu_2) \le \frac{S_p}{2}\right]$$

$m=9$, $n=16$

$$\therefore \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{9}+\frac{1}{16}}} \sim t(23)$$

$$P\left[-\frac{S_p}{2} \le (\overline{X}-\overline{Y})-(\mu_1-\mu_2) \le \frac{S_p}{2}\right]$$

$$= P\left[-\frac{1}{2\sqrt{\frac{1}{9}+\frac{1}{16}}} \le \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{9}+\frac{1}{16}}} \le \frac{1}{2\sqrt{\frac{1}{9}+\frac{1}{16}}}\right]$$

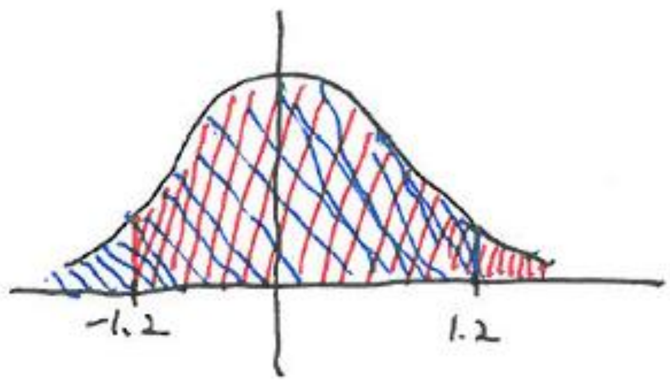
$$= P\left[-1.2 \le t(23) \le 1.2\right]$$

$$= 2P(T \le 1.2) - 1$$

여기서 $P(T \le 1.2) = 0.87882$
(R에서)

$$= 0.7576$$

예제에서 확인할 수 있는 것 처럼 두 집단의 평균을 비교할 때 이 분포를 사용할 수 있다.
두 집단의 평균이 같다고 귀무가설을 세우면 모평균의 차가 0이 됨으로 표본평균의 차에 따른 통계량의 확률값을 기준으로 통계적 검정을 실시한다.



표본분산비의 확률분포

두 집단 분산의 비교

- 표본분산비 : 두 개의 독립인 카이제곱분포 통계량의 비
  - F 분포로 표현

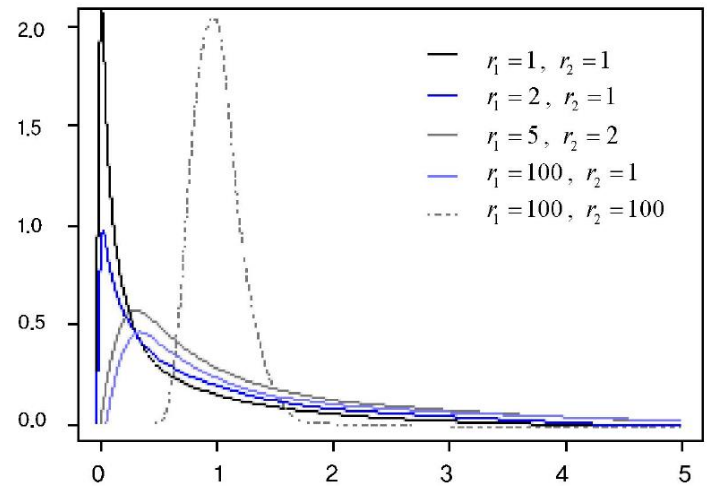
F 분포

- $V_1 \sim \chi^2(r_1), V_2 \sim \chi^2(r_2)$, 서로 독립
- $F = \dfrac{V_1/r_1}{V_2/r_2} \sim F(r_1, r_2)$
- 확률밀도함수

$$f(x) = \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \frac{r_1}{r_2}\left(\frac{r_1}{r_2}x\right)^{\frac{r_1}{2}-1}\left(1+\frac{r_1}{r_2}x\right)^{-\frac{r_1+r_2}{2}}, \quad x>0$$

- 분산분석(analysis of variance)에서 쓰이는 분포
- 피셔–스네데커(Fisher–Snedecor) 분포

F 분포의 모양



F 분포의 특성

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

$T \sim t(n)$ 이면

$$T = \frac{Z}{\sqrt{V/n}}$$

$$T^2 = \frac{Z^2/1}{V/n} \sim F(1,n).$$

$\sim \chi^2(1)$

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

$$\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(1)$$

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

표본분산비의 확률분포

표본분산의 비교

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

예 4.13

$X_1, \cdots, X_9 \sim N(\mu, \sigma_1^2)$, $Y_1, \cdots, Y_{16} \sim N(\mu, \sigma_2^2)$ 의 확률표본

$$P\left[\frac{S_1^2}{S_2^2} \Big/ \frac{\sigma_1^2}{\sigma_2^2} \le 2.64\right]$$

$$\frac{S_1^2}{S_2^2} \Big/ \frac{\sigma_1^2}{\sigma_2^2} \sim F(8, 15)$$

$$P\left[\frac{S_1^2}{S_2^2} \Big/ \frac{\sigma_1^2}{\sigma_2^2} \le 2.64\right] = 0.95$$

$X_1, X_2, \cdots X_m \sim N(\mu_1, \sigma_1^2)$

$Y_1, Y_2 \cdots Y_n \sim N(\mu_2, \sigma_2^2)$

$S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$

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

$\sim \chi^2(m-1)$

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

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

예제에서 확인할 수 있는 것 처럼 두 집단의 분산을 비교할 때 이 분포를 사용할 수 있다.
두 집단의 분산이 같다고 귀무가설을 세우면 모분산의 비가 1이 됨으로 표본분산의 비에 따른 통계량의 확률값을 기준으로 통계적 검정을 실시한다.

극한분포_확률적 수렴

확률적 수렴

- 모집단 $\sim N(\mu, \sigma^2)$ $\rightarrow$ 표본평균 $\sim N(\mu, \frac{\sigma^2}{n})$
- n 이 무한히 커지면 표본평균은 μ 에 근접
  - 표본평균 $\overline{X}$ 가 상수값 μ 에 확률적으로 수렴
  - $\overline{X} \xrightarrow{p} \mu$

마코프 부등식

- $u(X)$: X 의 양의 함수, 임의의 상수 $\varepsilon > 0$

$$P[u(X) \geq \varepsilon] \leq \frac{E[u(X)]}{\varepsilon}$$

*마코프 부등식.

여기서 $f(x)$ 는 $u(x)$ 의 확률밀도함수.

$$E(u(X)) = \int_0^\infty u(x) f(x)\,dx$$

$$= \underbrace{\int_0^{\varepsilon} u(x) f(x)\,dx}_{\text{0보다 크거나 같다.}} + \int_{\varepsilon}^\infty u(x) f(x)\,dx .$$

$$\therefore E(u(X)) \geq \underbrace{\int_{\varepsilon}^\infty u(x) f(x)\,dx \geq \int_{\varepsilon}^\infty \varepsilon f(x)\,dx}_{\varepsilon\text{은 상수, } u(x)\text{의 범위는 } \varepsilon\text{부터 이므로.}}$$

$$E(u(X)) \geq \varepsilon \int_{\varepsilon}^\infty f(x)\,dx = \varepsilon P(u(X) \geq \varepsilon)$$

$$\therefore P(u(X) \geq \varepsilon) \leq \frac{E(u(X))}{\varepsilon}$$

여사건의 확률.

또는 $P(u(X) > \varepsilon) \geq 1 - \frac{E(u(X))}{\varepsilon}$

극한분포_확률적 수렴

약대수법칙

임의의 양수 ε에 대하여 $\lim_{n\to\infty} P(|\overline{X}-\mu| < \varepsilon) = 1$

약대수법칙

$X_1, X_2, \cdots, X_n \sim E(X_i) = \mu$, $Var(X_i) = \sigma^2$ 확률표본일 때
$\overline{X}_n$ 가 상수값 μ 에 확률적으로 수렴

예 4.14

$X_1, X_2, \cdots, X_n \sim E(X_i) = \mu$, $Var(X_i) = \sigma^2$ 일 때
$\overline{X}_n$ 가 μ 에 확률적으로 수렴

*약대수 법칙.

$X_1, X_2, - X_n \quad E(X_i) = \mu, \quad Var(X_i) = \sigma^2$

이므로,

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad E(\overline{X}_n) = \mu \quad Var(\overline{X}_n) = \frac{\sigma^2}{n}$

$P(|\overline{X}_n - \mu| < \varepsilon)$

$= P(|\overline{X}_n - \mu|^2 < \varepsilon^2) \geq 1 - \frac{E(|\overline{X}_n - \mu|^2)}{\varepsilon^2}$

마코프부등식에서 $u(X)$

$= 1 - \frac{Var(\overline{X}_n)}{\varepsilon^2} = 1 - \frac{\sigma^2}{\varepsilon n}$

n이 무한히 커지면,

$\therefore \lim_{n\to\infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1,$

즉, $\overline{X} \xrightarrow{P} \mu.$

극한분포_확률적 수렴

확률적 수렴성의 성질

확률변수 $Y_n \to c, W_n \to d$ 일 때 (단, a,b,c,d 는 상수임)

① $aY_n + b \xrightarrow{p} ac+b$ ② $Y_n + W_n \xrightarrow{p} c+b$

③ $Y_n W_n \xrightarrow{p} cd$ ④ $\frac{1}{Y_n} \xrightarrow{p} \frac{1}{c}$

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

예 4.16

$X_1, X_2, \cdots, X_n \sim Ber(p)$ 의 확률표본,

$\overline{X}_n(1-\overline{X}_n) \xrightarrow{p} p(1-p)$

$E(\bar{X}_n) = \frac{1}{n}E(X_1 + \cdots + X_n) = \frac{1}{n}np = p$

$Var(\bar{X}_n) = Var(\frac{1}{n}(X_1 + \cdots + X_n))$

$= \frac{1}{n^2}Var(X_1 + \cdots + X_n)$

$= \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$

$P(|\bar{X}_n - p| < \varepsilon)$

$= P(|\bar{X}_n - p|^2 < \varepsilon^2) \geq 1 - \frac{E(|\bar{X}_n - p|^2)}{\varepsilon^2} = 1 - \frac{Var(\bar{X}_n)}{\varepsilon^2}$

$= 1 - \frac{p(1-p)}{n\varepsilon^2}$

$\therefore \bar{X}_n \xrightarrow{p} p$

$\therefore \bar{X}_n(1-\bar{X}_n) \xrightarrow{p} p(1-p)$

예 4.15

$X_1, X_2, \cdots, X_n$ 이 평균 μ , 분산 σ^2 의 확률표본일 때

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

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

$= \frac{n}{n-1} \cdot \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

$= \frac{n}{n-1} \cdot \{\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \frac{1}{n} \cdot n(\bar{X}_n)^2\}$

여기서 $\frac{n}{n-1} = \frac{n-1}{n-1} + \frac{1}{n-1} \longrightarrow 1$

$\underline{\frac{1}{n}\sum_{i=1}^{n}X_i^2} \xrightarrow{P} E(X_i^2)$

X_i^2의 평균

$(\bar{X}_n)^2 \xrightarrow{P} \mu^2$ (정리 4.13 ⑤)

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

$\xrightarrow{P} E(X_i^2) - \mu^2 = \sigma^2$

즉, $\underline{S_n^2 \xrightarrow{P} \sigma^2}$ //

극한분포_분포의 수렴

분포의 수렴

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

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

예 4.17

$X_1, X_2, \cdots, X_n \sim Ber(p)$ 의 확률표본, $Y_n = \sum_{i=1}^{n} X_i$.

n 이 무한히 커짐에 따라 p 가 0의 값에 가깝고 $np = \mu$ 일 때

Y_n 이 포아송분포를 따름을 보여라.

$X_1, X_2, \cdots X_n \sim Ber(p)$

$Y_n = \sum_{i=1}^{n} X_i$, $np = \mu$, p는 0에 가깝다.

$Y_n \sim B(n, p)$ $p = \frac{\mu}{n}$

이항분포 적률생성함수 $M(t) = [(1-p) + pe^t]^n$

포아송분포 〃 $M(t) = e^{\lambda(e^t-1)}$

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

$$= \left(1 - \frac{\mu}{n} + \frac{\mu e^t}{n}\right)^n = \left(1 + \frac{\mu(e^t-1)}{n}\right)^n$$

$$\lim_{n\to\infty} M_n(t) = \lim_{n\to\infty} \left(\underbrace{\left(1 + \frac{\mu(e^t-1)}{n}\right)^{\frac{n}{\mu(e^t-1)}}}_{e}\right)^{\mu(e^t-1)}$$

$$= \underbrace{e^{\mu(e^t-1)}}_{\text{포아송분포의 적률생성함수.}}$$

$\therefore Y_n \xrightarrow{d}$ 포아송분포.

극한분포_분포의 수렴

표본평균의 분포

- $X_1, X_2, \cdots, X_n$ 서로 독립 $\sim N(\mu, \sigma^2)$
  - $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$
- 중심극한정리

중심극한정리

- 드 무아르(A Braham de Moivre, 1667~1754)
  : 1733년 동전 던지기 앞면의 수의 분포
- 라플라스
  : 1812년 드 무아르의 결과를 확장, 이항분포를 정규분포로 근사
- 리아프노프(A. Lyapunov, 1857~1918)
  : 수학적으로 정리

많은 경우, 표준화 하여,

$$\boxed{Y_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)}$$

으로 쓰기도 한다.

중심극한정리

- $X_1, X_2, \cdots, X_n \sim E(X_i) = \mu\,, Var(X_i) = \sigma^2$ 의 확률표본
  - $Y_n = \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

정규분포 가정이 없다!!!
어떤 분포라도 표본의 크기가 충분히 크다면
표본평균의 분포는 정규분포를 따른다!!

Y_n의 적률생성함수

$M_n(t) = E[e^{t\sqrt{n}(\bar{X}_n-\mu)}] = E[e^{t\sqrt{n}\bar{X}} e^{-t\sqrt{n}\mu}]$

$= e^{-\mu\sqrt{n}t} \cdot E[e^{t\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i)}] = e^{-\mu\sqrt{n}t} \cdot E[e^{\frac{1}{\sqrt{n}}(\sum_{i=1}^{n} X_i)t}]$

X_i는 독립이므로

$= e^{-\mu\sqrt{n}t} \underbrace{E(e^{X_1\frac{t}{\sqrt{n}}}) E(e^{X_2\frac{t}{\sqrt{n}}}) \cdots E(e^{X_n\frac{t}{\sqrt{n}}})}_{\text{적률생성함수가 모두 같다.}}$

X_i의 적률생성함수를 $M(t)$라 하면,

$= e^{-\mu\sqrt{n}t}[M(\frac{t}{\sqrt{n}})]^n = [e^{-\frac{\mu}{\sqrt{n}}t} M(\frac{t}{\sqrt{n}})]^n$

여기서 $m(t) = e^{-\mu t} M(t)$ 로 놓으면,

$= [m(\frac{t}{\sqrt{n}})]^n$

* 테일러급수 $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$

$m(\frac{t}{\sqrt{n}})$를 0에서 테일러전개,

$= [m(0) + m'(0)\frac{t}{\sqrt{n}} + \frac{1}{2!}m''(0)\frac{t^2}{n} + \underbrace{\frac{1}{3!}m'''(0)\frac{t^3}{n\sqrt{n}} + \cdots}_{R(\frac{t}{\sqrt{n}})\text{로 표기.}}]^n$

* $m(0) = 1$

$m'(t) = e^{-\mu t}M(t) = -\mu e^{-\mu t}M(t) + e^{-\mu t}M'(t)$

$\therefore M'(0) = -\mu + \mu = 0$

$m''(0) = \sigma^2$

정리하면,

$= [1 + \frac{\sigma^2}{2}\cdot\frac{t^2}{n} + R(\frac{t}{\sqrt{n}})]^n$

여기서, $\lim_{n\to\infty} nR(\frac{t}{\sqrt{n}}) = 0$ (???)

$\therefore \lim_{n\to\infty} M_n(t) = \lim_{n\to\infty} \{1 + \frac{1}{n}[\frac{\sigma^2t^2}{2} + \cancel{nR(\frac{t}{\sqrt{n}})}^{\,0}]\}^n$

$= \lim_{n\to\infty} [\underbrace{\{1 + \frac{(\frac{\sigma^2t^2}{2})}{n}\}^{\frac{+n}{\sigma^2t^2}}}_{e}]^{\frac{\sigma^2t^2}{2}}$

$= e^{\frac{\sigma^2t^2}{2}} = \boxed{e^{\frac{1}{2}\sigma^2t^2}}$ ⇒ $N(0, \sigma^2)$인 적률생성함수

* $X \sim N(\mu, \sigma^2)$의 적률생성함수

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

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

극한분포_분포의 수렴

예 4.19

$Y_n \sim \chi^2(n)$ 따를 때 $(Y_n - n)/\sqrt{2n} \xrightarrow{d} N(0, \sigma^2)$

$Y_n \sim \chi^2(n)$ 이므로,

χ^2분포의 가법성에 따라,

$X_i \sim \chi^2(1)$ 이라고 하면

$Y_n = \sum_{i=1}^{n} X_i$.

* $X \sim \chi^2(n)$ 의 $E(X) = n$, $Var(X) = 2n$

$E(X_i) = 1$, $Var(X_i) = 2$

$$\frac{Y_n - n}{\sqrt{2n}} = \frac{\sum_{i=1}^{n} X_i - n}{\sqrt{2n}} = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - 1\right)}{\sqrt{2}}$$

$1 = \mu$, $2 = \sigma^2$ 이므로

$$= \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)}{\sigma}$$ 로 쓸 수 있다.

$$= \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

∴ 중심극한정리에 의해 $\frac{Y_n - n}{\sqrt{2n}} \xrightarrow{d} Z \sim N(0,1)$ 이다.

이항분포의 정규근사

$Y_n \sim B(n, p)$ 일 때 $Z_n = \dfrac{Y_n - np}{\sqrt{np(1-p)}}$ 의 극한확률분포는 $N(0,1)$

$Y_n \sim B(n, p)$ 이므로,

$X_i \sim Ber(p)$ 라고 하면,

$Y_n = \sum_{i=1}^{n} X_i$, $\bar{X}_n = \frac{Y_n}{n}$, $E(X_i) = p$ $Var(X_i) = p(1-p)$

중심극한정리에 의해,

$$\frac{(\bar{X}_n - p) \times n}{\left(\sqrt{\frac{p(1-p)}{n}}\right) \times n} = \frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} Z \sim N(0,1)$$

이항분포 정규근사의 연속성 수정

$$P(a \le Y \le b) = P\left[\frac{a-np}{\sqrt{np(1-p)}} \le \frac{Y-np}{\sqrt{np(1-p)}} \le \frac{b-np}{\sqrt{np(1-p)}}\right]$$

$$\approx P\left[\frac{a-np}{\sqrt{np(1-p)}} \le Z \le \frac{b-np}{\sqrt{np(1-p)}}\right]$$

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

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

극한분포_분포의 수렴

극한분포의 연산

확률변수 $Y_n \xrightarrow{d} Y,\ W_n \xrightarrow{p} c$ 일 때 (단, c 는 상수임)

① $Y_n + W_n \xrightarrow{d} Y + c$

② $Y_n W_n \xrightarrow{d} cY$

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

④ $h(x)$ 가 미분 가능하고 $h'(x)$ 가 b 에서 연속

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

예 4.21

$X_1, X_2, \cdots, X_n \sim (\mu, \sigma^2)$ 인 확률표본일 때

$Y_n = \dfrac{\bar{X}_n - \mu}{S_n / \sqrt{n}}$ 의 극한분포는?

$X_1, X_2, \cdots, X_n$

$E(X_i) = \mu, \quad Var(X_i) = \sigma^2$

$$Y_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} = \frac{\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\dfrac{S_n}{\sigma}}$$

분자 분모를 σ로 나눠서 정리

여기서 중심극한정리에 의해 $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$

약대수 법칙에 의해 $S_n^2 \xrightarrow{p} \sigma^2 \Rightarrow \dfrac{S_n}{\sigma} \xrightarrow{p} 1.$

슬러츠키 정리 (정리 4.16 ③) 에 의해,

$$Y_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} = \frac{\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\dfrac{S_n}{\sigma}} \xrightarrow{d} \frac{Z}{1} = Z \sim N(0,1)$$