중심극한정리

정의

- 개요

1. $X_1, \dots, X_n$을 평균 μ, 분산 σ^2인 정규분포에서 추출한 확률표본이라 할 때

1) $\frac{\sum_{1=1}^{n} x_i - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma}$ ~ N(0,1) 이다.

2) 이 때, 어떤 특수한 경우에서 $X_1, \dots, X_n$은 정규분포가 아니더라도 극한에선 표준정규분포로 수렴한다.

정의

- 증명

1. $X_1, \ldots, X_n$을 평균 μ, 분산 σ^2인 어떤 분포에서 추출한 확률표본이라 할 때

$\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \xrightarrow{D}$ N(0,1) 이다.

1) $\mathrm{M(t)} = \mathrm{E}(\mathrm{e}^{t(x-\mu)}) = \mathrm{e}^{-\mu t}\mathrm{m(t)}$에 대해서 $\mathrm{M(t)}$는 $(x-\mu)$의 mgf이므로

(1) $\mathrm{M}(0) = \mathrm{E}(\mathrm{e}^0)=1$

(2) $\mathrm{M}'(0) = \mathrm{E}[(x-\mu)\mathrm{e}^0] = \mathrm{E}[(x-\mu)] = 0$

(3) $\mathrm{M}''(0) = \mathrm{E}[(x-\mu)^2\mathrm{e}^0] = \mathrm{E}[(x-\mu)^2] = \sigma^2$

(4) $\mathrm{M(t)}$를 2차까지 테일러 전개하면

- $\mathrm{M(t)} = 1 + \mathrm{M}'(0)\mathrm{t} + \frac{1}{2}\mathrm{M}''(\xi)t^2 = 1 + \frac{1}{2}\mathrm{M}''(\xi)t^2$

(5) $\frac{\sigma^2 t^2}{2}$를 더하고 빼면

- $\mathrm{M(t)} = 1 + \frac{\sigma^2 t^2}{2} + \frac{[\mathrm{M}''(\xi) - \sigma^2]t^2}{2}$

정의

- 증명

2)이제, 다변량 결합분포의 mgf를 생각하자. $\mathrm{M}(\mathrm{t};n)$을

(1) $\mathrm{M}(\mathrm{t};n) = \mathrm{E}\left[\exp\left(\mathrm{t}\frac{\sum_{1=1}^{n} x_i - n\mu}{\sigma\sqrt{n}}\right)\right]$ 인 확률변수의 mgf를 정의하면

(2) $\mathrm{E}\left[\exp\left(\mathrm{t}\frac{\sum_{1=1}^{n} x_i - n\mu}{\sigma\sqrt{n}}\right)\right] = \mathrm{E}\left[\exp\left(\mathrm{t}\frac{X_1-\mu}{\sigma\sqrt{n}}\right)\cdots\exp\left(\mathrm{t}\frac{X_n-\mu}{\sigma\sqrt{n}}\right)\right]$

$$= \mathrm{E}\left[\exp\left(\mathrm{t}\frac{X_1-\mu}{\sigma\sqrt{n}}\right)\right]\cdots E\left[\exp\left(\mathrm{t}\frac{X_n-\mu}{\sigma\sqrt{n}}\right)\right]$$

(3) $X_1, \ldots, X_n$은 동일한 분포에서 추출된 확률표본이므로

$$\left\{\mathrm{E}\left[\exp\left(\mathrm{t}\frac{X-\mu}{\sigma\sqrt{n}}\right)\right]\right\}^n = \left\{\mathrm{M}\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n$$

정의

- 증명

3) 이 때, M(t) = $1+\frac{\sigma^2 t^2}{2}+\frac{[M''(\xi)-\sigma^2]t^2}{2}$ 에서 t를 $\frac{t}{\sigma\sqrt{n}}$로 교체하면

(1) $M(\frac{t}{\sigma\sqrt{n}}) = 1+\frac{t^2}{2n}+\frac{[M''(\xi)-\sigma^2]t^2}{2\sigma^2 n}$

(2) 따라서 $\left\{M\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n = \left\{1+\frac{t^2}{2n}+\frac{[M''(\xi)-\sigma^2]t^2}{2\sigma^2 n}\right\}^n$ 이다.

(3) $\lim_{n\to\infty}\left[1+\frac{b}{n}+\frac{\psi(n)}{n}\right]^{cn}$ 에서 $\lim_{n\to\infty}\frac{\psi(n)}{n}=0$ 이면 $\lim_{n\to\infty}\left[1+\frac{b}{n}\right]^{cn}=e^{bc}$인 정리를 이용

- $\lim_{n\to\infty}\left\{1+\frac{t^2}{2n}+\frac{[M''(\xi)-\sigma^2]t^2}{2\sigma^2 n}\right\}^n$ 에서 $\frac{[M''(\xi)-\sigma^2]t^2}{2\sigma^2 n}\to 0$ 이므로,
- 이 극한 mgf는 $\lim_{n\to\infty}\left\{1+\frac{t^2}{2n}\right\}^n = e^{\frac{1}{2}t^2}$

(4) 이 극한 mgf는 N(0,1)인 정규분포와 같다.

정의

- 다변량 분포로 확장

1. $\{X_n\}$이 p차원 벡터의 열일 경우, $\lim_{n\to\infty} P[||X_n - X|| \geq \varepsilon]$ = 0일때 $X_n \xrightarrow{p} X$ 이다.

2. 또한, $\{X_n\}$에서 X_n의 성분 $X_{nj}(\mathrm{j} = 1 \cdots p) \xrightarrow{p} X_j$ 일 때 한해 $X_n \xrightarrow{p} X$ 이다.

1) $$\underset{X_1}{\begin{bmatrix} X_{11} \\ \cdots \\ X_{1j} \end{bmatrix}} \underset{X_2}{\begin{bmatrix} X_{21} \\ \cdots \\ X_{2j} \end{bmatrix}} \cdots \underset{X_n}{\begin{bmatrix} X_{n1} \\ \cdots \\ X_{nj} \end{bmatrix}} \xrightarrow{p} \begin{bmatrix} X_1 \\ \cdots \\ X_j \end{bmatrix} \xrightarrow{p} X$$

2) $X_n \xrightarrow{p} X$ 라고 할 때

(1) $\varepsilon < |X_{nj}\text{- } X_j| \leq ||X_{nj}\text{- } X_j||$ 임이 선형대수학에서 증명되어있다. 이를 이용하면

(2) $\overline{\lim_{n\to\infty}}\,\mathrm{P}[(X_{nj}\text{-}X_j) \geq \varepsilon] \leq \overline{\lim_{n\to\infty}}\,\mathrm{P}[||\, X_{nj}\text{-}X_j \,|| \geq \varepsilon]$ = 0 이다.

정의

- 다변량 분포로 확장

2. 위 정리를 이용하면

1) $\{X_n\}$을 공통인 평균벡터 μ와 공분산행렬 Σ을 갖는 iid인 확률벡터 열이라고 한다면

(1) $\overline{x_n} = \frac{\sum x_i}{n}$를 표본평균들의 벡터라고 가정한다면

(2) $\begin{bmatrix} \overline{X_1} \\ \cdots \\ \overline{X_n} \end{bmatrix} \overset{p}{\to} \begin{bmatrix} \mu_1 \\ \cdots \\ \mu_n \end{bmatrix} \overset{p}{\to} \mu$ 일 때, $\bar{x} \overset{p}{\to} \mu$ 이다.

정의

- 다변량 분포로 확장

3. 한편 $\{X_n\}$이 분포함수 $F_n(X)$을 갖는 확률벡터 X_n의 열이고, 확률변수 X가 분포함수 $F(X)$를 가질 때

1) $\lim_{n\to\infty} F_n(X) = F(X)$ 이면 $\{X_n\}$은 X에 분포수렴 한다고 하고, $X_n \xrightarrow{D} X$ 이다.

4. 마찬가지로, $\{X_n\}$이 분포함수 $F_n(X)$과 mgf $M_n(t)$ 를 갖는 확률벡터 X_n의 열이고,
   X가 분포함수 $F(X)$와 mgf $M(t)$ 를 가질 때

1) $\lim\limits_{n\to\infty} M_n(t) = \mathrm{M}(t)$ 일때 한해 $\{X_n\}$은 X에 분포수렴한다.

정의

- 다변량 중심극한정리

1. $\{X_n\}$을 공통인 평균벡터 μ와 공분산행렬 Σ을 갖는 iid인 확률벡터 열이라고 한다면

$$Y_n = \frac{1}{\sqrt{n}} \sum_{1=1}^{n} (X_i - \mu) = \sqrt{n}(\bar{X} - \mu) \text{ 는}$$

$$Y_n \xrightarrow{D} N_p(0, \Sigma) \text{ 이다.}$$

1) $M_n(t)$가 Y_n의 적률생성함수 일 때

(1) $M_n(t) = \mathrm{E}\left[\exp\left\{t^T \frac{1}{\sqrt{n}} \sum_{1=1}^{n} (X_i - \mu)\right\}\right] = \mathrm{E}\left[\exp\left\{\frac{1}{\sqrt{n}} \sum_{1=1}^{n} t^T (X_i - \mu)\right\}\right]$

(2) 이 때, $W_i = t^T (X_i - \mu)$ 로 놓으면, W_i는 평균 0, 분산 $t^T \Sigma t$ 인 iid이다.

(3) 따라서, 중심극한정리에 따라

$$\frac{1}{\sqrt{n}} \sum_{1=1}^{n} W_i \xrightarrow{D} \mathrm{N}(0, t^T \Sigma t)$$

2) 이 때,

정의

- 다변량 중심극한정리

2) 이 때,

(1) $\mathrm{E}\left[\exp\left\{\frac{1}{\sqrt{n}}\sum_{1=1}^{n} t^T(X_i-\mu)\right\}\right] = \mathrm{E}\left[\exp\left\{\frac{1}{\sqrt{n}}\sum_{1=1}^{n} W_i\right\}\right]$ 에서

(2) $\mathrm{M}(1) = \mathrm{E}\left[\exp\left\{\frac{1}{\sqrt{n}}1\cdot\sum_{1=1}^{n} W_i\right\}\right]$ 이므로 이 극한은

$\mathrm{E}\left[\exp\left\{\frac{1}{\sqrt{n}}1\cdot\sum_{1=1}^{n} W_i\right\}\right] \rightarrow e^{1^2 t^T\Sigma t} = e^{t^T\Sigma t}$

(3) 이는 $N_p(0, \Sigma)$ 의 적률생성함수이다.

예제

- μ에 대한 대표본 추론

1. $X_1, \ldots, X_n$이 X~N(μ, σ^2)인 확률변수에서 추출한 확률표본이라고 할 때

1) $\bar{x}$, S를 표본평균과 표본표준편차라고 한다면

2) $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ = $(\frac{\sigma}{s})\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ 인데, 이 때 $\frac{\sigma}{s} \xrightarrow{p}$ 1 이다. 따라서

(1) $(\frac{\sigma}{s})\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \xrightarrow{p} \mathrm{N}(0,1)$ 이다. 그렇기 때문에 $\frac{\bar{x}-\mu}{s/\sqrt{n}} \xrightarrow{p} \mathrm{N}(0,1)$ 이다.

예제

- 이항분포에 대한 정규화

1. $X_1, \ldots, X_n$을 b(1,p)에서 추출한 확률표본이라고 하자.

1) 이 때, $Y_1 = \sum_{1=1}^{n} X_i$ 이면 Y_1~b(n,p) 이다.

2) 앞서 구한대로 Y_1의 극한분포는 푸아송분포를 따르지만, 한편으로는 극한 정규분포이기도 하다.

(1) $\frac{Y_n - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\overline{X} - p)}{\sqrt{p(1-p)}}$ ~ N(0,1) 이고

(2) N = 100, p = 1/2 , P(Y={48,49,50,51,52})라고 할 때

- P($47.5 < Y < 52.5$) = p($\frac{(47.5-50)}{5} \le \frac{(\mathrm{Y}-50)}{5} \le \frac{(52.5-50)}{5}$) = p(-0.5≤ $\frac{(\mathrm{Y}-50)}{5} \le 0.5$) 이다.

예제

- P를 모를때 극한 정규분포의 추정

1. Y_n=b(n,p) 라고 할 때, $\frac{Y}{n} = p$는 N(p, $\frac{p(1-p)}{n}$)을 따른다.

1) 이 때, 분산을 p에 의존하지 않고 구하기 위해서 이를 테일러 전개하면

(1) U($\frac{Y}{n}$) = u(p) + u′(p) ($\frac{Y}{n}$-p)

2) $\Delta -$ 방법으로 이 분포를 추측하면, 분포는

(1) N(u(p), $[u'(p)]^2$ $\frac{p(1-p)}{n}$) 을 따른다.

(2) 이를 미분방정식으로 바꾸면

- $u'(p)$ = $\frac{C}{\sqrt{p(1-p)}}$ 에서, 해는 u(p) = (2c)arcsin($\sqrt{p}$) 이다.
- C = 1/2이면, u($\frac{Y}{n}$) = arcsin($\sqrt{\frac{Y}{n}}$) ~N(arcsin(p), $\frac{1}{4}$n) 을 따른다.