표본분포 : 표본통계량(평균, 분산, 중앙값 등)의 분포.

Random sample : 표본의 크기가 같고 표본끼리 독립이라면 Random sample이다.

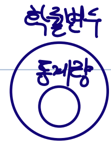
E.3.1

성공확률 $= p$ 인 베르누이 분포.

$f(x) = p^x(1-p)^{1-x}$ $(x=0,1)$ 서로 독립인 베르누이 시행을 10번 진행할 때, 표본을 $X_1, \cdots, X_{10}$이라고 할때,

결합 확률밀도함수는 $f_{X_1 \cdots X_{10}}(x_1, \cdots, x_{10}) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_{10}}(x_{10}) = p^{\sum_{i=1}^{10} x_i}(1-p)^{\sum_{i=1}^{10} 1-x_i}$ $(x_i = 0, 1)$

$X_1, \cdots, X_{10}$은 확률밀도함수가 $f(x) = p^x(1-p)^{1-x}$ $(x=0,1)$인 크기 10인 랜덤표본이다.



확률변수 : 표본공간에서 관찰가능한 결과를 실수로 대응하는 것

↳ 일종의 함수이다.

정의 3.3

미지의 모수를 포함하지 않는 랜덤표본의 함수 $T = T(X_1, \cdots, X_n)$를 통계량이라고 한다.

↳ 통계량도 확률변수이다.

↳ 통계량도 표본들에 함수를 가해서 만들어지므로 특수한 경우의 확률변수인 것이다.

E.3.2

$f(x;\theta)$로부터 얻은 Random sample $X_1, \cdots, X_n$. (θ는 미지의 모수)

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ 와 $\max\{X_1, \cdots, X_n\}$은 통계량이다.

$\bar{X}_n - \theta$ 와 $\max\{X_1/\theta, \cdots, X_n/\theta\}$은 미지의 모수 θ에 의존하므로 통계량이 아니다.

정의 3.4

$f(x;\theta)$으로부터 크기 n인 Random Sample $X_1, \cdots, X_n$의 r차 표본적률(r-th sample moment)는

$$m_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r \text{ 이다.}$$

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_n$ 에 대한 r차 표본중심적률(r-th sample central moment)는

$$m_r = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^r \text{ 으로 정의된다.}$$

위의 표본적률은 모수 θ를 포함 X. 확률표본의 함수 즉, 통계량이고

m'_r의 기댓값을 이용하여 모적률 추정이 가능하다.

$E(m'_r) = \mu'_r$

→ 모적률 추정을 위해선 표본평균, 표본분산의 성격을 알아본다.

$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^r\right) = E(X^r)$

정리 3.1

$f(x : \mu, \sigma^2)$으로부터 Random Sample $X_1, \cdots, X_n$.

→ Random Sample 이기에 모든 X_i에서 평균은 μ, 분산은 σ^2으로 동일하다.

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

① $E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\cdot\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot n\cdot\mu = \mu$

② $Var(\bar{X}_n) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\cdot n\cdot\sigma^2 = \frac{\sigma^2}{n}$

③ $E(S_n^2) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right) = \sigma^2$의 증명

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n + \bar{X}_n - \mu)^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{n}(\bar{X}_n - \mu)^2 + \cancel{\sum_{i=1}^{n} 2(X_i - \bar{X}_n)(\bar{X}_n - \mu)}$$

$$\Rightarrow \sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\bar{X}_n)^2 + \sum_{i=1}^{n}(\bar{X}_n-\mu)^2$$

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

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

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

정규분포와 관련된 분포들.

1. 카이제곱분포 (chi-squared distribution)

$X \sim GAM(\frac{n}{2}, 2)$는 $\chi^2_{(n)}$이다.

$f_X(x) = \frac{1}{\Gamma(\frac{n}{2}) \cdot 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}$, $x>0$ 이다. $M_X(t) = \frac{1}{(1-2t)^{\frac{n}{2}}}$ $(t<\frac{1}{2})$

$\frac{1}{(1-\beta t)^{\alpha}}$ $(GAM(\alpha, \beta))$

$E(X)=n$, $Var(X)=2n$

정리 3.3

X_i들이 각각 $\chi^2_{(k_i)}$을 가질 때, $Y=\sum_{i=1}^n$는 $\chi^2_{(\sum_{i=1}^n k_i)}$을 가진다.

<증명>

$M_X(t) = \frac{1}{(1-2t)^{\frac{k}{2}}}$ $M_Y(t) = \underbrace{(1-2t)^{-\frac{k_1}{2}} \times \cdots \times (1-2t)^{-\frac{k_n}{2}} = (1-2t)^{\sum_{i=1}^n \frac{k_i}{2}}}_{\text{지수가 합이 되어야하기에 곱으로 표현.}}$ 이므로 $\chi^2(\sum_{i=1}^n k_i)$ 이다.

- χ^2의 자유도가 커지면 정규분포와 비슷해진다.

정리. 3.4

확률변수 Z가 $N(0,1)$을 따를 때, $Z^2 \sim \chi^2_{(1)}$을 따른다.

<증명>

$$M_{Z^2}(t) = E(e^{tZ^2}) = \int_{-\infty}^{\infty} e^{tz^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz^2 - \frac{z^2}{2}} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{z^2(\frac{2t-1}{2})} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2 \cdot \frac{1-2t}{2}} dz$$

$$= \underbrace{\int_{-\infty}^{\infty} \frac{\sqrt{1-2t}}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot z^2 \cdot (1-2t)} dz}_{N(0, \frac{1}{1-2t}) \text{ 이므로 } 1} \cdot \frac{1}{\sqrt{1-2t}} = \frac{1}{\sqrt{1-2t}} = (1-2t)^{-\frac{1}{2}}, \ t < \frac{1}{2}$$

이것이 $\chi^2_{(1)}$의 mgf이다.

그리고 $Z \sim N(0,1)$일 때, $\sum_{i=1}^k Z_i^2$는 $\chi^2_{(k)}$를 따른다.

<증명>

$X \sim N(\mu, \sigma^2)$일 때, $X_1, \cdots, X_i$들은 모두 독립이므로 그들의 함수인 $\frac{X_i - \mu}{\sigma^2}$도 모두 독립이다.

$Z = \frac{X-\mu}{\sigma^2}$라고 할 때, $Z^2 \sim \chi^2_{(1)}$을 따르므로 Z^2들의 합은 자유도가 $\sum_{i=1}^k Z_i^2$인 카이제곱분포를 따른다.

2. F-분포 (F-distribution)

· 정규분포로부터 구한 <u>독립인 두 표본의 분산비</u>에 대한 분포를 설명할 때 중요.

U, V는 서로 독립인 카이제곱분포이다. $U \sim \chi^2(n)$, $V \sim \chi^2(m)$

$$X = \frac{\chi^2(n)/n}{\chi^2(m)/m} \sim F(n,m), \quad f_X(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{n}{m}\right)^{\frac{n}{2}} \frac{x^{\frac{n-2}{2}}}{\left(1+\frac{n}{m}x\right)^{\frac{n+m}{2}}}, \quad x>0$$

정리 3.6

$$E(X) = \frac{m}{m-2} \ (m>2) \qquad Var(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}, \ (m>4)$$

<증명>

$X \sim F(n,m)$ $U \sim \chi^2(n)$, $V \sim \chi^2(m)$ $\qquad X = \frac{\chi^2(n)/n}{\chi^2(m)/m} = \frac{U/n}{V/m}$

$$E(X) = E\left(\frac{U/n}{V/m}\right) = \frac{m}{n}E\left(\frac{U}{V}\right) = \frac{m}{n} \times E(U) \times E\left(\frac{1}{V}\right) = \frac{m}{n} \times n \times E\left(\frac{1}{V}\right) = m \cdot E\left(\frac{1}{V}\right)$$

$$E\left(\frac{1}{V}\right) = \int_{-\infty}^{\infty} \frac{1}{v} \cdot \frac{1}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} v^{\frac{m}{2}-1} \cdot e^{-\frac{v}{2}} I(v>0)\, dv$$

$$= \frac{1}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \int_0^\infty \frac{1}{v} \cdot v^{\frac{m-2}{2}} \cdot e^{-\frac{v}{2}} dv = \frac{1}{\Gamma\left(\frac{m}{2}\right)} \cdot \left(\frac{1}{2}\right)^{\frac{m}{2}} \underline{\int_0^\infty v^{\frac{m-4}{2}} \cdot e^{-\frac{v}{2}} dv}$$

$$= \frac{1}{\Gamma\left(\frac{m}{2}\right)} \cdot \left(\frac{1}{2}\right)^{\frac{m}{2}} \int_0^\infty \frac{1}{2^{\frac{m-2}{2}}\Gamma\left(\frac{m-2}{2}\right)} v^{\frac{m-2}{2}-1} \cdot e^{-\frac{v}{2}} \cdot 2^{\frac{m-2}{2}} \cdot \Gamma\left(\frac{m-2}{2}\right) dv$$

$$= \frac{\Gamma\left(\frac{m-2}{2}\right) \cdot 2^{\frac{m-2}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \underbrace{\int_0^\infty \frac{1}{2^{\frac{m-2}{2}} \cdot \Gamma\left(\frac{m-2}{2}\right)} v^{\frac{m-2}{2}-1} e^{-\frac{v}{2}} dv}_{GAM\left(\frac{m-2}{2},\, 2\right) = 1.} = \frac{\Gamma\left(\frac{m}{2}-1\right)}{\Gamma\left(\frac{m}{2}\right)} \cdot \frac{2^{\frac{m-2}{2}}}{2^{\frac{m}{2}}}$$

$$= \frac{\cancel{\Gamma\left(\frac{m}{2}-1\right)}}{\left(\frac{m}{2}-1\right)\cancel{\Gamma\left(\frac{m}{2}-1\right)}} \cdot \frac{2^{\frac{m-2}{2}}}{2^{\frac{m}{2}}} = \frac{1}{\frac{m-2}{2}} \cdot \frac{\cancel{2^{\frac{m}{2}}} \cdot 2^{-1}}{\cancel{2^{\frac{m}{2}}}} = \frac{2}{m-2} \cdot \frac{1}{2} = \frac{1}{m-2}, \ m>2.$$

$X \sim F(n,m)$ 일때, $\frac{1}{X} \sim F(m,n)$을 따른다. $\left(X = \frac{U/n}{V/m} \quad \frac{1}{X} = \frac{V/m}{U/n} = F(m,n) \right)$

$F_\alpha(n,m)$을 $F(n,m)$의 $100(1-\alpha)$ 백분위수라고 하면

$$\alpha = P[X \geq F_\alpha(n,m)] = \underline{P\left[\frac{1}{X} \leq \frac{1}{F_\alpha(n,m)}\right]} = 1 - P\left[\frac{1}{X} \geq \frac{1}{F_\alpha(n,m)}\right]$$

① 부등식 부호가 바뀐다.

② $\frac{1}{F_\alpha(n,m)}$은 $F(m,n)$의 누적확률과 관련이 있다.

③ 이제 확률의 대칭성 때문에 $\frac{1}{F_\alpha(n,m)} = F_\alpha(m,n) = F_{1-\alpha}(m,n)$ 이 성립.

④ 즉 $\frac{1}{F_\alpha(n,m)} = F_{1-\alpha}(m,n)$ 이다.

E.3.3

$X \sim F(5,10)$ $P(X \leq a) = 0.01$을 만족하는 a는 $F_{0.99}(5,10)$ 이다.

$$a = F_{0.99}(5,10) = \frac{1}{F_{0.99}(10,5)} = \frac{1}{F_{0.01}(10,5)} = \frac{1}{10.1} = 0.099$$

더 나아가 $X \sim F(6,12)$, $\frac{1}{X} \sim F(12,6)$ 이므로.

$$P(X \geq 0.25) = P\left[\frac{1}{X} \leq \frac{1}{0.25}\right] = P\left[\frac{1}{X} \leq 4.00\right] = 0.95$$

3. t-분포 (student's t-distribution)

· 정규분포의 모평균 추정, 가설검정 등 검정이론에서 중요하다.

$X = \frac{Z}{\sqrt{U/k}} \sim t(k)$ $\quad Z \sim N(0,1)$, $U \sim \chi^2(k)$ $\quad X = \frac{Z}{\sqrt{U/k}}$, $Y = U$로 변수변환.

Z와 U는 서로 독립이므로 $f_{Z,U}(z,u) = f_Z(z) f_U(u)$ $\quad z = X \cdot \sqrt{\frac{Y}{k}}$, $u = Y$로 표현

$$f_{Z,U}(z,u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \cdot \frac{1}{\Gamma(\frac{k}{2}) 2^{\frac{k}{2}}} y^{\frac{k}{2}-1} \cdot e^{-\frac{u}{2}} I(z>0) I(u>0)$$

$$|J| = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{y}{k}} & 0 \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{y}{k}}$$

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\Gamma(\frac{k}{2}) 2^{\frac{k}{2}}} y^{\frac{k}{2}-1} \cdot e^{-\frac{1}{2}(x^2 \cdot \frac{y}{k})} \cdot e^{-\frac{y}{2}} \cdot \sqrt{\frac{y}{k}} \, I(x>0) I(y>0)$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \frac{1}{\sqrt{2\pi}\Gamma(\frac{k}{2}) 2^{\frac{k}{2}}} \cdot \sqrt{\frac{1}{k}} \cdot \int_0^\infty e^{(-\frac{x^2}{2k} - \frac{1}{2})y} \cdot y^{\frac{k-1}{2}} dy =$$

$$= \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})\sqrt{k\pi}} \cdot \frac{1}{(1+\frac{x^2}{k})^{\frac{k+1}{2}}} I(-\infty < x < \infty)$$

정리 3.7

$X \sim t(n)$ 일때, $E(X)=0\ (n>1)$, $Var(X)=\frac{n}{n-2}$, $n>2$

또한 $T \sim t(n)$이라면 $T^2 \sim F(1,n)$을 따른다.

정규분포로부터의 표본추출.

정리 3.6

서로 독립인 확률변수 $X_1, \cdots, X_i$들이 정규분포 $N(\mu_i, \sigma_i^2)$를 따르면

$\sum_{i=1}^n X_i$는 $N(\sum_{i=1}^n X_i, \sum_{i=1}^n \sigma_i^2)$를 따른다.

<증명>

먼저 정규분포의 적률생성함수는 $e^{\mu t + \frac{\sigma^2 t^2}{2}}$

$\sum_{i=1}^n X_i$의 mgf는 $M(t)=E(e^{t \cdot \sum_{i=1}^n X_i}) = E(e^{tX_1} \cdot e^{tX_2} \cdot \cdots \cdot e^{tX_n}) = E(e^{tX_1}) \cdots E(e^{tX_n})$

$= M_{X_1}(t) \times \cdots \times M_{X_n}(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \times \cdots \times e^{\mu_n t + \frac{\sigma_n^2 t^2}{2}}$

$= e^{\sum_{i=1}^n \mu_i t + \frac{\sigma_i^2 t^2}{2}}$는 $N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$의 mgf.

즉, $\sum_{i=1}^n X_i$는 $N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$를 따른다.

E. 3.5

$X \sim N(6, 2^2)$ $\sum_{i=1}^{10} X_i \sim N(\sum_{i=1}^{10} 6, \sum_{i=1}^{10} 2^2) = N(60, 40)$.

$P(X \geq 70) = P\left(\frac{X-\mu}{\sigma} \geq \frac{70-60}{\sqrt{40}}\right) = P\left(Z \geq \frac{\sqrt{40}}{4}\right) = 1 - P\left(Z < \frac{\sqrt{40}}{4}\right) = 1 - \Phi\left(\frac{\sqrt{40}}{4}\right) = 0.057$

- $X_i - \bar{X}_n$은 서로 독립이 아니다.

$\sum_{i=1}^n (X_i - \bar{X}_n) = 0$ 이다. 즉 $X_1 + \cdots + X_n = n\bar{X}_n$이므로 $X_i - \bar{X}_n$들은 서로 독립이 아니다.

그러므로 $X_i - \bar{X}_n$이 정규분포를 따름에도 불구하고 S_n^2의 분포를 구할때

정리 3.5를 사용할 수 없다.

↳ 서로 독립인 확률변수가 각각 정규분포를 따를때, $\sum_{i=1}^k \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2(k)$를 만족한다.

정리 3.9

$X_1, \cdots, X_n$을 $N(\mu, \sigma^2)$으로부터의 크기가 n인 Random Sample이라고 하면

① $\bar{X}_n \left(= \sum_{i=1}^{n} \frac{X_i}{n}\right)$과 $S_n^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$은 서로 독립이다.

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

E.3.6

$X \sim N(6, 2^2)$ $\quad P(S_n^2 > 5) = P\left(\frac{9 \cdot S_n^2}{\sigma^2} > \frac{9 \times 5}{\sigma^2}\right) = P\left(\frac{9S_n^2}{4} > \frac{45}{4}\right) = P\left(\chi^2_{(9)} > \frac{45}{4}\right) = 0.259$

정리 3.9

$N(\mu_X, \sigma_X^2)$인 크기가 n인 Random sample $X_1, \cdots, X_n$

$N(\mu_Y, \sigma_Y^2)$인 크기가 m인 Random sample $Y_1, \cdots, Y_m$

두 표본이 독립이라면 $F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(n-1, m-1)$.

<증명>

$$\frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi^2(n-1), \quad \frac{(m-1)S_Y^2}{\sigma_Y^2} \sim \chi^2(m-1) \quad \frac{\frac{(n-1)S_X^2}{\sigma_X^2} \cdot \frac{1}{n-1}}{\frac{(m-1)S_Y^2}{\sigma_Y^2} \cdot \frac{1}{m-1}} = \frac{\chi^2(n-1)/n-1}{\chi^2(m-1)/m-1} = F(n-1, m-1).$$

정리 3.11

$N(\mu, \sigma^2)$으로부터의 Random Sample $X_1, \cdots, X_n$. $\quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$T = \frac{Z}{\sqrt{V/k}} = \frac{(\bar{X}_n - \mu)/\frac{\sigma}{\sqrt{n}}}{\sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma^2} \Big/ n-1}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 / n-1}} \sim t(n-1).$$

• 확률변량 T는 μ가 알려져 있을 때, 통계량이다. 이를 스튜던트화 t통계량이라고 한다.

물론 μ를 몰라도 T는 t분포를 따르지만 unknown parameter를 포함하므로 통계량이 아니다.

[illegible]

그리고 t분포는 n이 커짐에 따라 표준정규분포로 가까워지는데 수식으로 증명 가능하다.

$\underline{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2/n-1}\ \text{이}\ n\text{이 커짐에 따라}\ \sigma\text{로 수렴한다.}}$ 따라서 T는 표준정규분포 확률로 근사된다.

슬러츠키 정리.

• 대수의 법칙과 중심극한정리.

지금까지는 모분포의 정규성을 가정하고 통계량의 표본분포를 다루었다.

하지만 때로는 모분포의 정규성을 가정하지 못할 수 있다.

그러므로 n이 클 때, 표본분포의 근사를 다루기로 한다.

• 실수의 수렴 $\neq$ 확률변수의 수렴

1. 실수의 수렴 : $n \to \infty$, $|x_n - x| \to 0$을 의미한다.
2. 확률변수의 수렴 : 확률변수 X_n은 값이 실현값이 아니라 확률공간에서 정의된 함수이다.

따라서 X_n의 수렴은 확률공간에서의 분포적, 확률적 관점에서 정의되어야 한다.

$\boxed{1}$ 대수의 법칙 (Law of large numbers)

확률변수의 열 $X_1, \cdots, X_n, \cdots$ 과 확률변수 X가 같은 확률공간에 정의된다.

만약 임의의 $\varepsilon > 0$에 대하여 $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$ 또는 $\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$ 이면

X_n이 X로 확률적으로 수렴한다고 하고 $X_n \xrightarrow{P} X$으로 표기한다.

평균이 $\mu < \infty$인 분포 $f(x)$로부터 Random sample $X_1, \cdots, X_n$을 얻었다면, $\bar{X}_n \xrightarrow{p} \mu$가 성립.

↳ n이 커짐에 따라 표본평균이 모평균으로 확률적으로 수렴한다는 것을 의미한다.

〈증명〉

$f(x)$가 $\sigma^2 < \infty$라고 가정.

$$P[|\bar{X}_n - \mu| < \varepsilon] = \underline{P[|\bar{X}_n - \mu|^2 < \varepsilon^2] \geq 1 - \frac{E[\bar{X}_n - \mu]^2}{\varepsilon^2}} = 1 - \frac{\sigma^2/n}{\varepsilon^2} \to 1$$

체비셰프 부등식.

E.3.7

$X_1, \cdots, X_n$ 서로 독립인 베르누이(p) 확률변수라고 하자. $E(X_1)=p$ 이므로

표본비율 $\hat{p}_n = \sum_{i=1}^{n} \frac{X_i}{n}$ 은 모비율 p로 확률적으로 수렴한다.

↳ $p[|\hat{p}_n - p| > \varepsilon] = 0$ 이 성립한다는 의미.

$$p[|\hat{p}_n - p|^2 > \varepsilon^2] = \frac{E(\hat{p}_n - p)^2}{\varepsilon^2} = \frac{Var(\hat{p}_n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0 \text{ 이므로 확률적으로 수렴.}$$

$$Var(\hat{p}_n) = Var\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \times n \times p(1-p) = \frac{p(1-p)}{n}$$

중심극한정리.

확률변수열 $X_1, \cdots, X_n, \cdots$ 이 cdf $F_{X_1}, \cdots, F_{X_n}, \cdots$ 을 각각 갖고

확률변수 X가 cdf F_X를 갖는다고 하자. 만약 함수 F_X가 연속인 모든점 x에서

$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ 가 만족된다면, X_n이 X로 분포수렴한다고 말하고, $X_n \xrightarrow{d} X$로 표기한다.

확률변수 $X_1, X_2, \cdots$ 의 누적분포함수가 $F_1, F_2, \cdots$ 이고 적률생성함수가 $M_1(t), M_2(t), \cdots$

라고 하자. 이때 어떤 t의 저근한 $-h<t<h$에 대하여 $\lim_{n\to\infty} M_n(t) = M(t)$가 성립하며

$M(t)$가 누적분포함수 $F(x)$를 갖는 어떤 확률분포의 적률생성함수라고 하면

$F(x)$가 연속인 모든 점에서 $\lim_{n\to\infty} F_n(x) = F(x)$가 성립한다.

정리 3.14

(중심극한정리) $\mu, \sigma^2 < \infty$ 인 확률밀도함수 $f(x)$로부터 랜덤표본 $X_1, \cdots, X_n$을 얻었다면.

이때 확률변량 $$Z_n = \frac{\sum_{i=1}^{n} X_i - E[\sum_{i=1}^{n} X_i]}{\sqrt{Var(\sum_{i=1}^{n} X_i)}} = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n}\,\sigma}$$ 는 표본크기가 무한대에 가까워지면

$N(0,1)$으로 분포수렴한다.

↳ ① 분자 : $\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n}(X_i - E(X_i)) = \sum_{i=1}^{n}(X_i - \mu)$

② 분모 : $Var(\sum_{i=1}^{n} X_i) = Var(n \cdot \frac{1}{n}\sum_{i=1}^{n} X_i) = Var(n \times \bar{X}) = n^2 \cdot Var(\bar{X}) = n^2 \cdot \frac{\sigma^2}{n} = n\sigma^2$

중심극한정리 증명

$X_i-\mu$의 mgf는 $m(t)=E(e^{t(X-\mu)})$ 라고 하면 $m'(0)=E(X_i-\mu)=0$, $m''(0)=E((X_i-\mu)^2)=\sigma^2$ 이다.

테일러 급수에 의해 $m(t)=\frac{m(0)}{0!}(t-0)^0+\frac{m'(0)}{1!}(t-0)^1+\frac{m''(0)}{2!}(t-0)^2+\frac{m'''(0)}{3!}(t-0)^3+\cdots$

$$=m(0)+m'(0)t+\frac{m''(0)}{2}t^2+\cdots$$

1. 테일러 전개를 $t=0$에서 진행했기에 ξ는 $0<\xi<t$에 있다고 정의한다.

$$=m(0)+m'(0)t+\frac{m''(\xi)}{2}t^2 \quad \underline{(0<\xi<t)}$$

2. ξ의 값은 알 수 없고, 계산할 필요가 없다.

1. 중심극한 정리는 t는 충분히 작다고 가정하므로 2차항(까지만) 전개한다.
2. $m''(\xi)$는 테일러전개의 나머지항을 의미한다.
3. 정확도를 보장하기위해 $0<\xi<t$인 값을 사용한다.

$$=1+\frac{m''(\xi)}{2}t^2=1+\frac{\sigma^2t^2}{2}+\frac{(m''(\xi)-\sigma^2)t^2}{2}$$ 이 된다.

편의상 $Z_n=\frac{\sum_{i=1}^{n}(X_i-\mu)}{\sqrt{n}\sigma}$ 으로 표현하면 Z_n의 mgf는

$$M_{Z_n}(t)=E(e^{tZ_n})=E\left(e^{t\frac{\sum_{i=1}^{n}(X_i-\mu)}{\sqrt{n}\sigma}}\right)=M_{\sum_{i=1}^{n}(X_i-\mu)}\left(\frac{t}{\sqrt{n}\sigma}\right)$$

$$=\prod_{i=1}^{n}M_{(X-\mu)}\left(\frac{t}{\sqrt{n}\sigma}\right)=\left[M_{(X-\mu)}\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n=\left[m\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n$$

$$m\left(\frac{t}{\sqrt{n}\sigma}\right)=m(0)+m'(0)\left(\frac{t}{\sqrt{n}\sigma}-0\right)+m''(0)\left(\frac{t}{\sqrt{n}\sigma}-0\right)^2+\cdots$$

$$=m(0)+m'(0)\left(\frac{t}{\sqrt{n}\sigma}\right)+\frac{m''(\xi)}{2!}\left(\frac{t}{\sqrt{n}\sigma}\right)^2 \quad \left(0<\xi<\frac{t}{\sqrt{n}\sigma}\right)$$

$$=1+\frac{m''(\xi)}{2n\sigma^2}t^2=1+\frac{\sigma^2t^2}{2n\sigma^2}+\frac{(m''(\xi)-\sigma^2)t^2}{2n\sigma^2}$$

즉 $M_{Z_n}(t)=\left[1+\frac{\sigma^2t^2}{2n\sigma^2}+\frac{(m''(\xi)-\sigma^2)t^2}{2n\sigma^2}\right]^n$ $\left(0<\xi<\frac{t}{\sqrt{n}\sigma}\right)$ 이다

그런데 $n\to\infty$ 일때, $\frac{t}{\sqrt{n}\sigma}\to 0$, $\xi\to 0$

$m''(\xi)\to m''(0)=\sigma^2$ 이다. 이 논리는 $m''(t)$이 연속성이 전제 되었을 때 가능하다.

적률생성함수는 $-h<t<h$에서 무한히 미분 가능하므로 $m''(t)$는 연속성이 전제될 수 있다.

즉 $\lim_{n\to\infty}M_{Z_n}(t)=\lim_{n\to\infty}\left[1+\frac{t^2}{2n}\right]^n=\lim_{n\to\infty}\left(1+\frac{t^2}{2n}\right)^{\frac{2n}{t^2}\times t^2\times\frac{1}{2}}=e^{\frac{t^2}{2}}$ 이다.

$e^{\frac{t^2}{2}}$은 $N(0,1)$의 mgf이다.

(정리 3.13) $\lim_{n\to\infty}M_n(t)=M(t)$가 성립하며 $M(t)$가 $F(x)$를 가지는 어떤 pdf의 mgf라고 한다면

$F(x)$는 연속인 모든 점에서 $\lim_{n\to\infty} F_n(x) = F(x)$가 성립하므로

(정리 2.26) 확률변수 X, Y가 같은 적률생성함수를 가지면 두 확률변수는 같은 확률밀도함수를 가진다.

E.3.8

$X_1, \dots, X_n$이 $U(0,1)$로부터 얻은 랜덤표본이라고 하자. $E(X_i)=\frac{1}{2}$ $Var(X_i)=\frac{1}{12}$이므로

$\sum_{i=1}^n X_i$의 분포를 $N(\frac{n}{2}, \frac{n}{12})$으로 근사할 수 있다.

$\dfrac{\sum_{i=1}^n X_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$은 근사적으로 표준정규분포를 따르므로 $Z \sim N(0,1)$이라고 하면

$$P(a \le \sum_{i=1}^n X_i \le b) = P\left(\frac{a-\frac{n}{2}}{\sqrt{\frac{n}{12}}} \le \frac{\sum_{i=1}^n X_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \le \frac{b-\frac{n}{2}}{\sqrt{\frac{n}{12}}}\right) \doteq P\left(\frac{a-\frac{n}{2}}{\sqrt{\frac{n}{12}}} \le Z \le \frac{b-\frac{n}{2}}{\sqrt{\frac{n}{12}}}\right)$$

$= \Phi\left(\frac{b-\frac{n}{2}}{\sqrt{\frac{n}{12}}}\right) - \Phi\left(\frac{a-\frac{n}{2}}{\sqrt{\frac{n}{12}}}\right)$가 성립한다.

즉 $\sum_{i=1}^n X_i$과 관련된 확률의 계산에서 표준정규분포를 사용하여 계산할 수 있다.

E.3.9

$X \sim B(n,p)$는 독립이고 성공확률이 p인 베르누이 분포를 n개 합한 것이다.

중심극한 정리에 따라 $P(a \le X \le b) = P(a \le \sum_{i=1}^n X_i \le b) \doteq P\left(\frac{a-np}{\sqrt{npq}} \le Z \le \frac{b-np}{\sqrt{npq}}\right)$

$= \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right)$이 된다.

이항분포가 n이 커짐에 따라 정규분포에 가까워지고 만약 p가 너무 작아서 np가 작다면 포아송근사를 사용하면 된다.

한편 이항분포는 이산형, 정규분포는 연속형이므로 연속성수정(continuity correction)을 하여 근사의 정확성을 더 높일 수 있다

이항분포는 이산형이므로 $P(X \le k) = P(X < k+1)$이지만 연속형은 만족하지 않는다.
그래서 k나 $k+1$을 사용하지 않고 $k+\frac{1}{2}$을 사용한다.

즉, $P(a \le X \le b)$가 아니라 $P(a-\frac{1}{2} \le X \le b+\frac{1}{2})$을 이용해서 근사하는 것이다.

(추가설명)

정수 k를 구간으로 $(k-0.5, k+0.5)$로 확장해서 계산한다.

a는 $[a-0.5, a+0.5]$로 처리.
b는 $[b-0.5, b+0.5]$로 처리
$\therefore a-0.5 \le X \le b+0.5$.

실제로 연속성 수정 후 더 정확하게 값을 근사할 수 있다.

• 슬러츠키(slutsky) 정리.

확률변수의 열 $X_1, \cdots, X_n$이 상수 c로 확률적으로 수렴 $(X_n \xrightarrow{p} c)$

" $Y_1, \cdots, Y_n$은 상수 Z로 " $(Y_n \xrightarrow{p} Z)$이라고 한다면

$Y_n + X_n \xrightarrow{d} Z + c$, $X_n Y_n \xrightarrow{d} c \cdot Z$가 성립한다.

E.3.10

$\mu, \sigma^2 < \infty$인 모분포로부터 Random Sample $X_1, \cdots, X_n$

$S_n \xrightarrow{p} \sigma$, 중심극한정리에 의해 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$ 이므로 슬러츠키 정리를 사용하여

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \xrightarrow{d} N(0,1)$이 성립함을 보일 수 있다.

• 델타 방법. - 점근적으로 정규분포를 따르는 확률변량의 함수의 극한분포를 근사하기 위한 방법.

$X_1, \cdots, X_n$에 대하여, $\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2)$이라고 하자

이때 함수 $g(\theta)$의 연속인 도함수 $g'(\theta)$가 존재하고 0이 아니면 $\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2 \cdot (g'(\theta))^2)$이 성립한다.

〈증명〉

평균값정리에 의하여 X_n과 θ 사이에 있는 ξ에 대하여 $g'(\xi) = \frac{g(X_n) - g(\theta)}{X_n - \theta} \Rightarrow g(X_n) = g(\theta) + g'(\xi)(X_n - \theta)$

이 성립한다. $X_n \xrightarrow{p} \theta$이므로 $\xi \xrightarrow{p} \theta$이고, $g'(\xi) \xrightarrow{p} g'(\theta)$가 성립한다.

그러므로 $g(\bar{X}_n) = g(\theta) + g'(\theta)(\bar{X}_n - \theta)$이 성립하고 $\sqrt{n}\,(g(\bar{X}_n) - g(\theta)) = g'(\theta)\sqrt{n}\,(\bar{X}_n - \theta)$이다.

$\sqrt{n}\,(\bar{X}_n - \theta) \xrightarrow{d} N(0,\sigma^2)$을 가정했으므로 $\sqrt{n}\,(g(\bar{X}_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2 \cdot [g'(\theta)]^2)$이 성립한다.

E.3.11

포아송분포로부터 Random Sample $X_1, \dots, X_n$이 주어졌을 때, 중심극한정리에 의해

$$Z_n = \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}/\sqrt{n}} \xrightarrow{d} N(0,1) \Rightarrow \sqrt{n}\,(\bar{X}_n - \lambda) \xrightarrow{d} N(0,\lambda)$$이 성립한다.

$g(\lambda) = \sqrt{\lambda}$, $g'(\lambda) = \frac{1}{2\sqrt{\lambda}}$, $(g'(\lambda))^2 = \frac{1}{4\lambda}$ $\quad$ $\sqrt{n}\,(g(\bar{X}_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2[g'(\theta)]^2)$인 델타 방법에 따라

$$\sqrt{n}\,(\sqrt{\bar{X}_n} - \sqrt{\lambda}) \xrightarrow{d} N\left(0, \lambda \cdot \frac{1}{4\lambda}\right) = \underline{N\left(0, \frac{1}{4}\right)}$$

↳ 극한분포의 분산이 모수에 의존하는 것을 방지하는 이런 방법을 분산 안정화 변환 (Variance stabilizing transformation)이라고 한다.

E.3.12

두 개의 표본 $X_1, \dots, X_n$, $Y_1, \dots, Y_m$으로부터 구한 표본평균 $\bar{X}_n$, $\bar{Y}_m$의 기댓값이 μ_X, μ_Y일 때,

두 모평균의 비 $g(\mu_X, \mu_Y) = \frac{\mu_X}{\mu_Y}$를 고려하자. $\frac{\partial}{\partial \mu_X} g(\mu_X, \mu_Y) = \frac{1}{\mu_Y}$이고 $\frac{\partial}{\partial \mu_Y} g(\mu_X, \mu_Y) = -\frac{\mu_X}{\mu_Y^2}$

$g(\bar{X}_n, \bar{Y}_m) = \frac{\bar{X}_n}{\bar{Y}_m}$의 기댓값, 분산을 구해보자.

$g(\bar{X}_n, \bar{Y}_m)$을 (μ_X, μ_Y) 근처에서 테일러 급수를 전개하여 1차 근사를 사용한다.

$$g(\bar{X}_n, \bar{Y}_m) \approx g(\mu_X, \mu_Y) + (\bar{X}_n - \mu_X)\frac{\partial g}{\partial \mu_X} + (\bar{Y}_m - \mu_Y)\frac{\partial g}{\partial \mu_Y} = \frac{\mu_X}{\mu_Y} + \frac{1}{\mu_Y}(\bar{X}_n - \mu_X) - \frac{\mu_X}{\mu_Y^2}(\bar{Y}_m - \mu_Y)$$

$$E[g(\bar{X}_n, \bar{Y}_m)] = E\left(\frac{\bar{X}_n}{\bar{Y}_m}\right) \approx E\left(\frac{\mu_X}{\mu_Y}\right) + E\left(\frac{1}{\mu_Y}(\bar{X}_n - \mu_X)\right) - E\left(\frac{\mu_X}{\mu_Y^2}(\bar{Y}_m - \mu_Y)\right)$$

$E(\bar{X}_n - \mu_X) = 0$, $E(\bar{Y}_m - \mu_Y) = 0$ 이므로 $E\left(\frac{\bar{X}_n}{\bar{Y}_m}\right) \approx \frac{\mu_X}{\mu_Y}$이다.

$$Var\left(\frac{\bar{X}_n}{\bar{Y}_m}\right) = Var\left(\frac{\mu_X}{\mu_Y} + \frac{1}{\mu_Y}(\bar{X}_n - \mu_X) - \frac{\mu_X}{\mu_Y^2}(\bar{Y}_m - \mu_Y)\right) = \frac{1}{\mu_Y^2}Var(\bar{X}_n) + \frac{\mu_X^2}{\mu_Y^4}Var(\bar{Y}_m) - 2\cdot\frac{\mu_X}{\mu_Y^3}Cov(\bar{X}_n, \bar{Y}_m)$$

• 순서통계량. (Order statistics)

Random Sample $X_1, \cdots, X_n$ 이면 이 Random Sample을 크기순으로 나열하여

순서통계량을 구할 수 있다.

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

이산형 확률변수는 확률변수와 순서통계량이 1:1 대응이다.

연속형 확률변수는 1:1 대응 X. 변환의 대상영역을 각 영역에서 1:1 변환이 되도록 나눈다.

예를 들어 $n=3$인 경우.

$A_1 = \{(x_1, x_2, x_3) : \overset{X_{(1)}}{x_1} < \overset{X_{(2)}}{x_2} < \overset{X_{(3)}}{x_3}\}$ 각 영역에서는 1:1 대응이 된다.

A_2 $\vdots$ 자코비안의 절댓값은 1, 순서통계량의 jcdf는 $f(x_{(1)})f(x_{(2)})f(x_{(3)})$

A_3 $\vdots$ 따라서 가능한 3!개를 합하면 $3!f(x_{(1)})f(x_{(2)})f(x_{(3)})$이 된다.

$\vdots$ $\vdots$

$A_6 = \{(x_1, x_2, x_3) : x_3 < x_2 < x_1\}$

일반화 : $g(x_{(1)}, \cdots, x_{(n)}) = \begin{cases} n!\, f(x_{(1)}) f(x_{(2)}) \cdots f(x_{(n)}), & x_{(1)} < \cdots < x_{(n)} \text{인 경우} \\ 0 & \text{otherwise} \end{cases}$

→ n!을 곱하는 이유는 n!가지 순열을 하나로 축약했기에 곱해준다.

• 자코비안 : 다변수 함수에서 변수 간 관계를 나타내는 편미분 행렬이다.
행렬식의 절댓값은 좌표변환에서 부피 변화율을 나타낸다.

E. 3.13

X_1, X_2, X_3을 pdf $f(x) = 3x^2\ (0<x<1)$으로부터 얻은 Random sample 이라고 하자.

이때, 순서통계량 $X_{(1)}, X_{(2)}, X_{(3)}$의 결합확률밀도함수를 구하면

$g(x_{(1)}, x_{(2)}, x_{(3)}) = 3!(3x_{(1)}^2)(3x_{(2)}^2)(3x_{(3)}^2) = 162\, x_{(1)}^2 x_{(2)}^2 x_{(3)}^2\, I(0<x_{(1)}<x_{(2)}<x_{(3)}<1)$ 이고

그 외의 경우는 0으로 주어진다.

일반적인 k번째 순서통계량 $X_{(k)}$의 pdf는 다음과 같다.

$X_1, \cdots, X_n$ Random Sample. $a<x<b$ 에 대하여 $f(x)>0$ 이라고 하자.

$$f(x_{(k)}) = \begin{cases} \frac{k!}{(k-1)!(n-k)!}[F(x_{(k)})]^{k-1}[1-F(x_{(k)})]^{n-k} \cdot f(x_{(k)}) & (a<x_{(k)}<b) \\ 0 & \text{otherwise.} \end{cases}$$

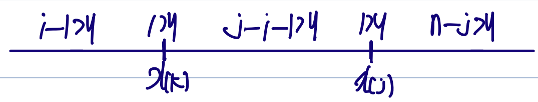
<증명>

$X_{(k)} = x_{(k)}$이므로 k-1개는 $x_{(k)}$보다 작고, n-k개는 $x_{(k)}$보다 크다.

여기서 $P(X \le x_{(k)}) = F(x_{(k)})$이고 $P(X > x_{(k)}) = 1 - P(X \le x_{(k)}) = 1 - F(x_{(k)})$이다.

가능한 경우의 수는 $\frac{k!}{(k-1)!(n-k)!}$이므로 확률밀도함수는 위와 같다.

i번째, j번째 순서통계량의 결합확률밀도함수는 $P[X \le x_{(k)}]$, $P[x_{(k)} < X \le x_{(j)}]$, $P[X > x_{(j)}]$이다.



$$f_{X_{(i)},X_{(j)}}(x_{(i)},x_{(j)})=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x_{(i)})]^{i-1}f(x_{(i)})\times[F(x_{(j)})-F(x_{(i)})]^{j-i-1}f(x_{(j)})\cdot[1-F(x_{(j)})]^{n-j}$$

1-1 표본최솟값 $X_{(1)}$의 cdf

$$G(x_{(1)})=P[X_{(1)}\le x_{(1)}]=1-P[X_{(1)}>x_{(1)}]=1-[1-F(x_{(1)})]^n$$

2-1 표본최댓값 $X_{(n)}$의 cdf

$$G(x_{(n)})=P[X_{(n)}\le x_{(n)}]=[F(x_{(n)})]^n$$

1-2 표본최솟값의 pdf

$$f_{X_{(1)}}(x_{(1)})=\frac{d}{dx_{(1)}}G(x_{(1)})=-n[1-F(x_{(1)})]^{n-1}(-f(x_{(1)}))=nf(x_{(1)})\cdot[1-F(x_{(1)})]^{n-1}$$

2-2 표본최댓값의 pdf

$$f_{X_{(n)}}(x_{(n)})=\frac{d}{dx_{(n)}}G(x_{(n)})=n\cdot F(x_{(n)})^{n-1}\cdot f(x_{(n)})$$

E. 3. 14.

지수분포로부터의 $X_1,\dots,X_n$ Random Sample $f(x)=\frac{1}{\lambda}e^{-\frac{x}{\lambda}}I(x>0)$

$X_{(1)}$의 pdf

$$n\cdot f(x_{(1)})\cdot[1-F(x_{(1)})]^{n-1}=n\cdot\frac{1}{\lambda}e^{-\frac{x_{(1)}}{\lambda}}\cdot[1-(1-e^{-\frac{x_{(1)}}{\lambda}})]^{n-1}=\frac{n}{\lambda}e^{-\frac{nx_{(1)}}{\lambda}}\quad(x_{(1)}>0)$$

즉, 평균이 $\frac{\lambda}{n}$인 지수분포의 pdf가 된다

$X_{(n)}-X_{(1)}$ = 표본 범위 (Sample Range).