# 6장. 확률변수의 극한(Limiting Distribution)

**Large sample size properties**

Sometimes, the probability distribution for a statistic, $\overline{X}$, is hard to describe for small size $n$. But we can say something for "large" $n$.

**정의. 확률수렴 ( Convergence in Probability )**

A sequence of r.v. $X_1, X_2, \cdots$ converges in probability to a r.v. $X$ if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0 \text{ or } \lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1.$$

Notation: $X_n \xrightarrow{p} X$

## 6.1 대수의 법칙

**Thm 6.1 ( *Law of large numbers* : 대수의 법칙 )**

Let $X_1, X_2, \cdots$ be i.i.d random variables with mean $\mu$ and variance $\sigma^2 < \infty$.

Let $\overline{X}_n = \sum_{i=1}^{n} X_i / n$ then for every $\varepsilon > 0$, $\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0$.

i.e., $\overline{X}_n$ converges in probability to $\mu$.

Proof) Recall. Chebyshev's inequality (Thm 4.6)

Use Chebyshev's inequality.

i.e., $P(|\overline{X}_n - \mu| \geq \varepsilon) = P((\overline{X}_n - \mu)^2 \geq \varepsilon^2) \leq \dfrac{E(\overline{X}_n - \mu)^2}{\varepsilon^2} = \dfrac{Var(\overline{X}_n)}{\varepsilon^2} = \dfrac{\sigma^2}{n\varepsilon^2} \to 0$

as $n \to \infty$.

Example ( Consistency of $S^2$ )

$E(S^2) = \sigma^2$

Does $S^2 \to \sigma^2$ in probability ??

Same proof works if $Var(S^2) \to 0$.

## 6.2 분포수렴

### Def 6.1 분포수렴 ( Converges in Distribution )

$X_1, X_2, \cdots$ converges in distn to a r.v. $X$ if $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$ at all $x$ where $F_X(x)$ is continuous.

(i.e., $F_{X_n}(x) = P(X_n \leq x) \quad \to \quad P(X \leq x) = F_X(x)$ )

Notation: $X_n \xrightarrow{d} X$, or $X_n \to X$ in dist.

Example

Let $Y_n \sim \text{Beta}(1+\frac{1}{n}, 1)$, $Y \sim \text{Beta}(1,1)$ : i.e., Uniform r.v.

Then,

$$F_{Y_n}(y) = \int_0^y \frac{\Gamma(2+\frac{1}{n})}{\Gamma(1+\frac{1}{n})\Gamma(1)} u^{(1+\frac{1}{n}-1)} du = \int_0^y (1+\frac{1}{n}) u^{\frac{1}{n}} du = y^{1+\frac{1}{n}}$$

$$\Longrightarrow F_{Y_n}(y) = \begin{cases} 0, & y \leq 0 \\ y^{1+\frac{1}{n}}, & 0 < y < 1 \\ 1, & 1 \leq y \end{cases}$$

$$\& \qquad F_Y(y) = \begin{cases} 0, & y \leq 0 \\ y, & 0 < y < 1 \\ 1, & 1 \leq y \end{cases}$$

Therefore $F_{Y_n}(y) \to F_Y(y)$ at all $y$ where $F_Y(y)$ is continuous.

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

**Ex 6.1** Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} U(0,\theta)$. What is the limiting distribution of $X_{(n)}$?

*Sol.*

$$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(X_1 \le x, X_2 \le x, \cdots, X_n \le x)$$

$$= \left(\frac{x}{\theta}\right)^n,\ 0 \le x \le \theta.$$

Then, $\lim_{n\to\infty} F_{X_{(n)}}(x) = \begin{cases} 0 & x < \theta \\ 1 & x \ge \theta \end{cases}$

Let $X = \theta$. Then, $F_X(x) = \begin{cases} 0 & x < \theta \\ 1 & x \ge \theta \end{cases}$ i.e., $F_{X_{(n)}}(x) \to F_X(x)$.

$\therefore\ X_{(n)} \xrightarrow{d} X$, *i.e.*, $X_{(n)} \xrightarrow{d} \theta$. & $X_{(n)} \xrightarrow{p} \theta$.

**Ex 6.2** Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} U(0,\theta)$.

What is the limiting distribution of $Y_n = n(\theta - X_{(n)})$?

*Sol.*

*Fill out by yourself*

**Thm 6.2** ( *Convergence of mgf implies convergence in distribution* )

Let the mgfs of $X_1, X_2, \cdots$ be $M_1(t), M_2(t), \cdots$, respectively, and $M(x)$ be the mgf of $X$.

If $\lim_{n\to\infty} M_n(t) \to M(t)$ for all $t$, then $X_n \xrightarrow{d} X$.

**Ex 6.3** Let $X_n \sim \text{Poisson}(\lambda_n)$ and $\lambda_n \to \infty$.

What is the limiting distribution of $Z_n = \dfrac{X_n - E(X_n)}{\sqrt{Var(X_n)}} = \dfrac{X_n - \lambda_n}{\sqrt{\lambda_n}}$?

*Sol.* Use convergence of mgf

$$
\begin{aligned}
M_{Z_n}(t) = E(e^{tZ_n}) &= E\left[e^{t\left(\frac{X_n-\lambda_n}{\sqrt{\lambda_n}}\right)}\right] \\
&= E\left(e^{-t\sqrt{\lambda_n}+\frac{t}{\sqrt{\lambda_n}}X_n}\right) \\
&= e^{-t\sqrt{\lambda_n}}E\left(e^{\frac{t}{\sqrt{\lambda_n}}X_n}\right) \\
&= e^{-t\sqrt{\lambda_n}}M_{X_n}\left(\frac{t}{\sqrt{\lambda_n}}\right) \\
&= e^{-t\sqrt{\lambda_n}}e^{\lambda_n\left(e^{\frac{t}{\sqrt{\lambda_n}}}-1\right)} \\
&= e^{-t\sqrt{\lambda_n}+\lambda_n\left(e^{\frac{t}{\sqrt{\lambda_n}}}-1\right)} \\
&= e^{-t\sqrt{\lambda_n}+\lambda_n\left(1+\frac{t}{\sqrt{\lambda_n}}+\frac{1}{2}\left(\frac{t}{\sqrt{\lambda_n}}\right)^2+\sum_{k=3}^{\infty}\frac{1}{k!}\left(\frac{t}{\sqrt{\lambda_n}}\right)^k-1\right)} \\
&= e^{\frac{t^2}{2}+\sum_{k=3}^{\infty}\frac{1}{k!}\frac{t^k}{\lambda_n^{k/2-1}}} \\
&\to e^{\frac{t^2}{2}} \text{ as } n\to\infty. \ :\ \text{mgf of } N(0,1)
\end{aligned}
$$

$$\therefore\ Z_n \xrightarrow{d} X,\ \ X \sim N(0,1)$$

**Ex 6.4** *Read*

## 6.3 중심극한정리 (Central Limit Theorem)

**Thm 6.3 ( Central Limit Theorem : CLT )**

$X_1, X_2, \cdots, X_n \overset{iid}{\sim}$ mean $\mu$ and variance $\sigma^2$, $0 < \sigma^2 < \infty$.

Let $S_n = \sum_{i=1}^{n} X_i$, $\overline{X}_n = \sum_{i=1}^{n} X_i / n$. Then,

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma} = \frac{(\overline{X}_n - \mu)}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

That is, for any $x$, $-\infty < x < \infty$,

$$\lim_{n\to\infty} P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leqq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = P(Z \leqq x)$$

**Proof.**

**Note** : m.g.f. of $N(\mu, \sigma^2)$ is $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.

**Recall** : If $M_{Y_n}(t) \to M_Y(t)$, then $Y_n \xrightarrow{d} Y$.

We will assume m.g.f of $X_i$ exists and show that

the m.g.f of $\dfrac{(\overline{X}_n - \mu)}{\sigma/\sqrt{n}}$ converges to $e^{t^2/2}$ as $n \to \infty$.

Define $Y_i = \dfrac{X_i - \mu}{\sigma}$ and $M_Y(t)$ denoted m.g.f of $Y$

$$==> \; Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}\sum_{i=1}^{n}(X_i - \mu)}{\sigma n} = \frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}}$$

$$\text{Then, } M_{Z_n}(t) = M_{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i}(t) = M_{\sum_{i=1}^{n} Y_i}\left(\frac{t}{\sqrt{n}}\right) = \left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Expand $M_Y\left(\dfrac{t}{\sqrt{n}}\right)$ in a Taylor's series about "0"

Note (Taylor's expansion) :

Taylor's expansion of function $f(x)$ at $x=a$

$$\Rightarrow f(x)=f(a)+\frac{f^{(1)}(a)}{1!}(x-a)+\frac{f^{(2)}(a)}{2!}(x-a)^2+\cdots$$

When $a=0$,

$$f(x)=f(0)+\frac{f^{(1)}(0)}{1!}(x-0)+\frac{f^{(2)}(0)}{2!}(x-0)^2+\cdots$$

$$M_Y\left(\frac{t}{\sqrt{n}}\right)=\sum_{k=0}^{\infty}M_Y^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!}$$

$$M_Y(0)=E(e^{0Y})=1$$

$$M_Y^{(1)}(0)=E(Y)=E\left(\frac{X-\mu}{\sigma}\right)=0$$

$$M_Y^{(2)}(0)=E(Y^2)=Var(Y)=\frac{1}{\sigma^2}Var(X)=\frac{1}{\sigma^2}\sigma^2=1$$

$$==>\ M_Y\left(\frac{t}{\sqrt{n}}\right)=1+\frac{(t/\sqrt{n})^2}{2}+R_Y(t/\sqrt{n})$$

Note that

$$\lim_{n\to\infty}\frac{R_Y(t/\sqrt{n})}{(t/\sqrt{n})^2}=0. \quad ==>\ \lim_{n\to\infty}nR_Y\left(\frac{t}{\sqrt{n}}\right)=0.$$

$$\therefore \lim_{n\to\infty}(M_Y(t/\sqrt{n}))^n=\lim_{n\to\infty}\left[1+\frac{(t/\sqrt{n})^2}{2}+R_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

$$=\lim_{n\to\infty}\left[1+\frac{1}{n}\left(\frac{t^2}{2}+nR_Y\left(\frac{t}{\sqrt{n}}\right)\right)\right]^n$$

$$=e^{t^2/2}\quad : mgf\ \ of\ \ N(0,1)$$

Here, we use the fact that

$$\lim_{n\to\infty}a_n=a \text{ 일 때, } \lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n=e^a.$$

Meaning of the CLT (Important) :

*Read text p.153*

Example : $Y_1, Y_2, \cdots, Y_n \overset{iid}{\sim} Bernoulli(p)$ and $S_n = \sum_{i=1}^{n} Y_i$

$$\text{i.e., } Y_i = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$$

$$\Longrightarrow \frac{S_n}{n} = \overline{X}_n, \quad \frac{\overline{X}_n - p}{\sqrt{p(1-p)/n}} \xrightarrow{d} N(0,1)$$

Normal approximation to binomial

$Y \sim Bin(n,p)$ is approximated by $N(np, np(1-p))$

Approximation is good if $n$ is large, $p$ is not close to "0" or "1".

( $np \geqq 5$ & $n(1-p) \geqq 5$ )

Ref. Continuity correction (연속성 수정)

When we use CLT, a continuity correction is needed for discrete distn.

If $Y \sim Bin(n,p)$ & $n$ is large enough, $\Phi$ : cdf of $N(0,1)$

$$P(Y \leq k) \approx \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right)$$

$$P(Y < k) \approx \Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right)$$

Ex 6.5 : $X \sim Bin(100, 0.5)$일 때, $P(X \geq 60)$의 근사값?

*Sol.*

$$P(X \geq 60) = 1 - P(X < 60) = 1 - P\left(\frac{X - 0.5 - 50}{\sqrt{25}} \leq \frac{60 - 0.5 - 50}{\sqrt{25}}\right)$$

$$= \Phi\left(\frac{60 - 0.5 - 50}{\sqrt{25}}\right)$$

Example : $X_1, X_2, \cdots, X_{25} \overset{iid}{\sim} Bin(5, 0.4)$일 때,

$$P(1 \leqq \overline{X} \leqq 2.5) = P(\frac{1-2}{\sqrt{6/5/}\ \sqrt{25}} \leqq \frac{\overline{X}-2}{\sqrt{6/5/}\ \sqrt{25}} \leqq \frac{2.5-2}{\sqrt{6/5/}\ \sqrt{25}})$$

$$\approx P(-4.56 \leqq Z \leqq 2.28) = 0.9887$$