Ch.5 Approximation of Sample Distribution

5.1. CENTRAL LIMIT THEOREM

The approximation of the cumulative probability of a binomial distribution was introduced in Chapter 3 Section 6:

$$\sum_{x:\, a\le \frac{x-np}{\sqrt{np(1-p)}} \le b} \binom{n}{x} p^x (1-p)^{n-x} \sim \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \,, \quad n\to\infty$$

Meanwhile, if $X_1, \cdots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, the number of defects in the sample, $X_1+\cdots+X_n$, follows in binomial distribution $B(n,p)$. Hence, for a random variable $Z \sim N(0,1)$,

$$\lim_{n\to\infty} P\left(a \le \frac{X_1+\cdots+X_n - np}{\sqrt{np(1-p)}} \le b \right) = P(a \le z \le b)$$

Meaning, for the sample proportion $\hat{p} = (X_1+\cdots+X_n)/n$, we can approximate its probability as:

$$\lim_{n\to\infty} P\left(a \le \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \le b \right) = P(a \le z \le b)$$

As such, oftentimes the approximation of the sample probability are given as the probability of the standard normal distribution. The most fundamental and useful statistic that can be approximated in this manner is the sample mean. The following "Central limit theorem" introduces the precedent approximation and takes a pivotal role in the approximation of all probabilities.

〈Theorem 5.1.1. The Central Limit Theorem〉

When the random variables $X_1, \cdots, X_n$ are i.i.d and $\text{Var}(X_i) > 0$,

Let $\quad E[X] = \mu \,,\ \text{Var}(X) = \sigma^2 \quad (0 < \sigma^2 < +\infty)$

the following holds for $Z \sim N(0,1)$

$$\lim_{n\to\infty} P\left(\frac{(X_1+\cdots+X_n)/n - \mu}{\sigma/\sqrt{n}} \le x \right) = P(z \le x) \quad \forall x : -\infty < x < +\infty$$

(proof)

The theorem can be proved for the general case but as it is beyond the topics covered by this book, we will prove the theorem given that the mgf of X_i exists. We will show that the mgf of

$$\sqrt{n}(\bar{X}_i - \mu)/\sigma = \left\{ (X_1+\cdots+X_n)/n - \mu \right\} / (\sigma/\sqrt{n})$$

converges to the mgf of Z : $mgf_Z(t) = \exp(t^2/2)$. [67] In short, we will prove the following:

$$\lim_{n\to\infty} mgf_{\sqrt{n}(\bar{X}-\mu)/\sigma}(t) = mgf_Z(t) = \exp(t^2/2)$$

From the properties of mgf in 〈Theorem 2.5.11〉,

$$\begin{aligned} mgf_{\sqrt{n}(\bar{X}-\mu)/\sigma}(t) &= E\left[\exp\left\{t \cdot \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right\}\right] \\ &= E\left[\exp\left\{t \frac{\sqrt{n}\cdot\left(\frac{1}{n}(X_1+\cdots+X_n)-\mu\right)}{\sigma}\right\}\right] \\ &= E\left[\exp\left\{t \frac{\sqrt{n}\cdot\frac{1}{n}(X_1-\mu)}{\sigma}\right\}\right] \cdots E\left[\exp\left\{t \frac{\sqrt{n}\cdot\frac{1}{n}(X_n-\mu)}{\sigma}\right\}\right] \\ &= E\left[\exp\left\{\frac{t}{\sqrt{n}}\left(\frac{X_1-\mu}{\sigma}\right)\right\}\right]^n \\ &= mgf_{(X_1-\mu)/\sigma}(t/\sqrt{n})^n \end{aligned}$$

Let $m(s) = mgf_{(X_1-\mu)/\sigma}(s)$. From the properties of mgf and Taylor series expansion, $m(t/\sqrt{n}) = m(0) + \frac{m'(0)}{1!}(t/\sqrt{n}) + \frac{m''(0)}{2!}(t/\sqrt{n})^2 + R_{n,t}$

$$= 1 + E\left[\frac{X_1-\mu}{\sigma}\right] t/\sqrt{n} + E\left[\left(\frac{X_1-\mu}{\sigma}\right)^2\right] \frac{t^2}{2n} + R_{n,t}$$

* $\lim_{n\to\infty} R_{n,t} / (t/\sqrt{n})^2 = 0$

$$\therefore\ m(t/\sqrt{n}) = 1 + \frac{1}{2n}t^2 + R_{n,t} \,, \quad \lim_{n\to\infty} n R_{n,t} = 0$$

Hence, [68]

$$\log mgf_{\sqrt{n}(\bar{X}_n-\mu)/\sigma}(t) = n \log\left(1 + \underbrace{\frac{1}{2}\frac{t^2}{n} + R_{n,t}}_{x}\right), \quad \lim_{n\to\infty} n R_{n,t} = 0$$

[68] $\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots = x + R(x)$, $\lim_{x\to 0} R(x) = 0$

$$= n\left\{ \frac{1}{2}\frac{t^2}{n} + \underbrace{R_{n,t} - \frac{1}{2}\left(\frac{1}{2}\frac{t^2}{n} + R_{n,t}\right)^2 + \cdots}_{r_{n,t}} \right\}$$

$$= \frac{1}{2}t^2 + n \cdot r_{n,t} \,, \quad \lim_{n\to\infty} n r_{n,t} = 0$$

$$\therefore \lim_{n\to\infty} mgf_{\sqrt{n}(\bar{X}-\mu)/\sigma}(t) = \exp\left\{\log mgf_{\sqrt{n}(\bar{X}-\mu)/\sigma}(t)\right\} = \exp\left\{\frac{1}{2}t^2\right\} = mgf_Z(t) \,, \quad Z \sim N(0,1)$$

* 증명순서
① $mgf_{\frac{z}{n}}$ n개의 mgf 곱으로 나타내기
② Taylor series 전개하기
③ $m'(0) = E[\cdot]$, $m''(0) = E[\cdot^2]$ 꼴로 치환하기
④ log 씌우기
⑤ $\log(1+A) = A - \frac{1}{2}A^2 + \frac{1}{3}A^3 + \cdots$ 이용해서 전개하기
⑥ e 씌우기 (④ 단계 되돌리기)

[67] Generally, it is known that following is true if the cdf of Z is continuous,

$$\lim_{n\to\infty} mgf_{Z_n}(t) = mgf_Z(t) \quad \forall t : |t| < h \ (\exists h > 0) \Rightarrow \lim_{n\to\infty} P(Z_n \le x) = P(Z \le x) \quad \forall x$$

Ex. 5.1.1

(a) Normal approximation of Poisson distribution.

For random variables $X_1, \cdots, X_n \overset{i.i.d}{\sim} \text{Poisson}(\lambda)$

$$E[X] = \text{Var}(X) = \lambda$$

Thus, $\lim_{n\to\infty} P\left\{\frac{(X_1+\cdots+X_n)/n - \lambda}{\sqrt{\lambda/n}} \le x\right\} = P(z \le x)$, $z \sim N(0,1)$

Meanwhile, using the properties of a Poisson distribution, $X_1+\cdots+X_n \sim \text{Poisson}(n\lambda)$

$$\lim_{n\to\infty} \sum_{k:\ \frac{k-n\lambda}{\sqrt{n\lambda}} \le x} \frac{e^{-n\lambda}(n\lambda)^k}{k!} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Hence, if $Y_n \sim \text{Poisson}(n)$ and n is large enough, we can use the cdf of the Standard normal distribution Φ to make the following approximation: [69]

$$P(a < Y_n \le b) \simeq \Phi\left(\frac{b-n}{\sqrt{n}}\right) - \Phi\left(\frac{a-n}{\sqrt{n}}\right)$$

[69] Such approximations are possible due to the uniform convergence of the mean to the normal distribution in CLT.

이러한 근사계산은 중심극한정리에서 정규분포의 수렴이 균등수렴 (연습문제 (5.9) 참조) 이기 때문에 가능한 것이다.

(b) Normal Approximation of Gamma · Chi-Square distribution.

For $X_1, \cdots, X_n \overset{iid}{\sim} \text{Gamma}(\alpha,\beta)$, $E[X_i] = \alpha\beta$, $\text{Var}(X_i) = \alpha\beta^2$ $(\alpha>0, \beta>0)$

Hence, $\lim_{n\to\infty} P\left(\frac{(X_1+\cdots+X_n)/n - \alpha\beta}{\sqrt{\alpha\beta^2/n}} \le x\right) = P(z \le x)$, $Z \sim N(0,1)$

Meanwhile if $Y_n \sim \chi^2(n)$, $Y_n \overset{d}{\equiv} X_1+\cdots+X_n$, $X_i \overset{iid}{\sim} \chi^2(1) = \text{Gamma}(1/2, 2)$ $(i=1,\cdots,n)$

Hence, $\lim_{n\to\infty} P\left(\frac{Y_n - n}{\sqrt{2n}} \le x\right) = P(z \le x)$, $Z \sim N(0,1)$

Therefore, $Y_n \sim \chi^2(n)$ and for large enough n, similar to (a),

$$P(a \le Y_n \le b) \simeq \Phi\left(\frac{b-n}{\sqrt{2n}}\right) - \Phi\left(\frac{a-n}{\sqrt{2n}}\right)$$

The Central Limit Theorem is true for the multivariate case and holds a pivotal role in approximating Sample distribution of many Statistics. The proof is similar to the Single variable case.

(Theorem 5.1.2. The Central Limit Theorem in the Multivariate Case.

If the multivariate random variable $X_i = (X_{i1}, \cdots, X_{iK})^t, \cdots, X_n = (X_{n1}, \cdots, X_{nK})^t$ is i.i.d., and $\text{var}(X_i)$ exists,

Let $E[X_i] = \mu$, $\text{Var}(X_i) = \Sigma$

for $Z = (Z_1, \cdots, Z_K)^t \sim N_K(0, \Sigma)$, the following is true:

$$\lim_{n\to\infty} P\left\{\sqrt{n}\,(X_{1j} + \cdots + X_{nj})/n - \mu_j) \le x_j,\ j = 1, \cdots, K\right\}$$

$$= P(Z_1 \le x_1, \cdots, Z_K \le x_K) \quad \forall x_j$$

Ex. 5.1.2. Multivariate Normal Approximation of Multinomial Distribution

From the representational definition of multinomial distribution,

$$\text{If } X = (X_{n1}, \cdots, X_{nK})^t \sim \text{Multi}(n, (p_1, \cdots, p_K)^t)$$

$$X_n \overset{d}{\equiv} Z_1 + \cdots + Z_n,\ Z_i = (Z_{i1}, Z_{i2}, \cdots, Z_{iK})^t \overset{iid}{\sim} \text{Multi}(1, (p_1, p_2, \cdots, p_K)^t)$$
$$i = 1, \cdots, n$$

Meanwhile, let $D(p_j)$ be the matrix whose diagonal entries are $p_1, \cdots, p_K$. From (Theorem 3.2.2),

$$E[X] = p = (p_1, \cdots, p_K)^t, \quad \text{Var}(X) = n(D(p_j) - pp^t)$$

Thus, from CLT,

$$\lim_{n\to\infty} P\left\{(X_n - np)/\sqrt{n} \le x\right\} = P(z \le x),\ Z \sim N_K(0,\ n(D(p_j) - pp^t))$$

* $D(p_j) - pp^t$ is a Singular matrix and, hence, does not have an inverse matrix.

$$D(p_j) - pp^t = \begin{bmatrix} p_1 & 0 & \cdots & 0 & 0 \\ 0 & p_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & p_{K-1} & 0 \\ 0 & 0 & \cdots & 0 & p_K \end{bmatrix} - \begin{bmatrix} p_1^2 & p_1p_2 & \cdots & p_1p_{K-1} & p_1p_K \\ p_2p_1 & p_2^2 & \cdots & p_2p_{K-1} & p_2p_K \\ \vdots & \vdots & & \vdots & \vdots \\ p_{K-1}p_1 & p_{K-1}p_2 & \cdots & p_{K-1}^2 & p_{K-1}p_K \\ p_Kp_1 & p_Kp_2 & \cdots & p_Kp_{K-1} & p_K^2 \end{bmatrix}$$

$$\text{Var}(X) = n\left\{D(p_j) - pp^t\right\}$$

5.2. ASYMPTOTIC DISTRIBUTION AND CONVERGENCE IN PROBABILITY

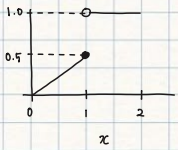
In the CLT, the distribution that was used in approximation was normal, whose cdf was continuous. But the cdf of the distribution used in approximation, such as the Poisson approximation of the Binomial distribution, may not be continuous. We will now consider the distribution approximation in the general case.

Ex. 5.2.1

Consider a test in which you flip a coin. If it is heads, show a random number between 0 and $1+n^{-1}$. If it is tails, show the number $1+n^{-1}$. Let X_n denote the number we observe in each test ($n=1,2,\cdots$). If the cdf of X_n ($n=1,2,\cdots$) is as follows, think about the approximation of the distribution.

$$cdf_{X_n}(x) = P(X_n \le x) = \begin{cases} 0, & x<0 \\ \frac{1}{2}\frac{1}{1+n^{-1}}x, & 0 \le x \le 1+n^{-1} \\ 1, & x \ge 1+n^{-1} \end{cases}$$

$$G(x) = \lim_{n\to\infty} cdf_{X_n}(x) = \begin{cases} 0, & x<0 \\ \frac{1}{2}x, & 0 \le x \le 1 \\ 1, & x>1 \end{cases}$$



$G(x)$ is not right continuous at $x=1$. Thus, from <Theorem 1.5.1>, it is not a cdf.

$F(x)$ is the modified function of $G(x)$ s.t. it is right continuous at $x=1$.

$$F(x) = \begin{cases} 0, & x<0 \\ \frac{1}{2}x, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

$F(x)$ is a cdf and for random variables X_n ($n=1,2,\cdots$) in Ex. 5.2.1,

$$\lim_{n\to\infty} cdf_{X_n}(x) = F(x) \quad \forall x : x \ne 1$$

Let X be the number observed in a test in which a coin is flipped and if it is head, show a random number between 0 and 1, if it is tails, show 1. $F(x)$ is the cdf of X. Thus,

$$\lim_{n\to\infty} cdf_{X_n}(x) = cdf_X(x) \quad \forall x : x \ne 1$$

and we can say that the distribution of X_n ($n=1,2,\cdots$) approximates to the distribution of X.

We now formally define the "limiting distribution" that covers cases, such as Ex. 5.2.1, where the limiting distribution's cdf is not continuous.

Definition: Limiting Distribution

For the random variable's columns X_n and Z, if $\lim_{n\to\infty} P(X_n \le x) = P(Z \le x)$ is true for all x that are continuous in cdf_Z - Cumulative distribution function of Z, then Z is called limiting distribution or the asymptotic distribution of X_n. It is denoted as: $X_n \xrightarrow[n\to\infty]{d} Z$

Let $Conti(cdf_Z)$ denote the set of all points continuous in cdf_Z.

$$X_n \xrightarrow[n\to\infty]{d} Z \Leftrightarrow \lim_{n\to\infty} cdf_{X_n}(x) = cdf_Z(x) \quad \forall x : x \in Conti(cdf_Z)$$

Ex. 5.2.2. Poisson approximation of the Binomial Distribution.

It was introduced in chapter 3 section 4 that the following is true for $X_n \sim Bin(n, \lambda/n)$ $(0<\lambda/n<1)$ and $X \sim Poisson(\lambda)$.

$$P(X_n=k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \xrightarrow[n\to\infty]{} \frac{1}{k!}\lambda^k e^{-\lambda} = P(X=x), \quad k=1,2,\cdots$$

Meanwhile, the random variables take 0 or $n \in \mathbb{N}$, thus for all points in x,

$$cdf_{X_n}(x) = \sum_{k:\,0\le k\le x} P(X_n=k) \xrightarrow[n\to\infty]{} \sum_{0\le k\le x} P(X=k) = cdf_X(x)$$

Hence, $Poisson(\lambda)$ is the limiting distribution of $B(n,\lambda/n)$ and this is denoted as:

$$B(n,\lambda/n) \approx Poisson(\lambda), \quad n\to\infty$$

$$* \lim_{n\to\infty} P(X_n=k) = \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^k = \lim_{n\to\infty}\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n\to\infty}\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \lim_{n\to\infty}\underbrace{\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)}_{\to 1}\frac{\lambda^k}{k!}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1^{-k}=1} = 1\cdot\frac{\lambda^k}{k!}e^{-\lambda} = \frac{\lambda^k e^{-\lambda}}{k!}$$

Ex. 5.2.3.

Let $U_{(1)} < \cdots < U_{(n)}$ denote the order statistics of random samples from $U(0,1)$.

$$P\{n(1-U_{(n)}) \le x\} = P\{1-U_{(n)} \le \frac{x}{n}\} = P\{1-\frac{x}{n} \le U_{(n)}\}$$

$$= 1 - P\{U_{(n)} \le 1-\frac{x}{n}\}$$

$$P\{U_{(n)} \le 1-\frac{x}{n}\} = \begin{cases} 0, & 1-x/n<0 \Leftrightarrow x>n \\ (1-\frac{x}{n})^n, & 0 \le 1-x/n < 1 \Leftrightarrow 0<x\le n \\ 1, & 1 \le x/n \Leftrightarrow x \le 0 \end{cases}$$

$$\lim_{n\to\infty} P\{U_{(n)} \le 1-\frac{x}{n}\} = \begin{cases} 0 & x>+\infty \ (\text{doesn't exist}) \\ e^{-x} & 0<x\le+\infty \\ 1 & x\le 0 \end{cases}$$

Hence the limiting distribution of $n(1-U_{(n)})$ is,

$$\lim_{n\to\infty} P\{n(1-U_{(n)}) \le x\} = \begin{cases} 1-e^{-x}, & x \ge 0 \\ 0, & x<0 \end{cases}$$

← 누적분포함수의 성질 (3) 오른쪽 연속성을 위해서 변경

$$\therefore n(1-U_{(n)}) \xrightarrow[n\to\infty]{d} Z, \quad Z \sim Exp(1)$$

asympotic distribution 찾는법:
① X_n의 cdf 찾기
② $\lim_{n\to\infty} cdf_{X_n}(x)$ 찾기
③ 혹은 $\lim_{n\to\infty} cdf_{X_n}(x)$ 계산하기

⟨Theorem 5.2.1. Limiting distribution is degenerative (i.e., 상수의 분포)⟩

The condition for the random variable's column $X_n (n=1,2,\cdots)$ to have a limiting distribution that is a constant c distribution is:

$$X_n \xrightarrow[n\to\infty]{d} X,\ P(X=c)=1 \Leftrightarrow \lim_{n\to\infty} P(|X_n-c|\geq\epsilon)=0 \quad \forall \epsilon>0.$$

(proof)

If $P(X=c)=1$, the cdf of X is:

$$cdf_X(x)=\begin{cases}0, & x<c\\ 1, & x\geq c\end{cases}$$

Hence $cdf_X(x)$ is continuous on $\forall x: x\neq c$.

WTS:

$(\Leftarrow)$ $\lim_{n\to\infty} P(|X_n-c|\geq\epsilon)=0 \Leftrightarrow \lim_{n\to\infty} P(|X_n-c|<\epsilon)=1 \Leftrightarrow \lim_{n\to\infty} P(-\epsilon<X_n-c<\epsilon)=1$

$\Leftrightarrow \lim_{n\to\infty} P(c-\epsilon<X_n<c+\epsilon)=1$

$\Leftrightarrow \lim_{n\to\infty} P(X_n<c+\epsilon)-P(X_n<c-\epsilon)=1$

$\Leftrightarrow \lim_{n\to\infty} P(X_n<c+\epsilon)-P(X_n<c-\epsilon)=cdf_X(c+\epsilon)-cdf_X(c-\epsilon)=1-0=1$ ✓✓

$(\Rightarrow)$ (go backwards in the precedent proof)

$(\Rightarrow)$ 극한분포의 정의에 따라

$$\lim_{n\to\infty} cdf_{X_n}(x)=cdf_X(x) \quad \forall x: x\neq c$$

따라서 임의의 양수 ϵ에 대하여 다음이 성립하는 것을 알 수 있다.

$$\lim_{n\to\infty} P(X_n\leq c-\epsilon)=cdf_X(c-\epsilon)=0$$

$$0\leq \lim_{n\to\infty} P(X_n\geq c+\epsilon)\leq \lim_{n\to\infty} P(X_n>c+\epsilon/2)=1-cdf_{X_n}(c+\epsilon/2)=0$$

$$\therefore \lim_{n\to\infty} P(|X_n-c|\geq\epsilon)=0$$

$(\Leftarrow)$ 주어진 조건이 성립한다면, 임의의 양수 ϵ에 대하여 다음이 성립하게 된다.

$$0\leq cdf_{X_n}(c-\epsilon)=P(X_n\leq c-\epsilon)\leq P(|X_n-c|\geq\epsilon)\xrightarrow[n\to\infty]{} 0$$

$$0\leq 1-cdf_{X_n}(c+\epsilon)=P(X_n>c+\epsilon)\leq P(|X_n-c|\geq\epsilon)\xrightarrow[n\to\infty]{} 0$$

$$\therefore \lim_{n\to\infty} cdf_{X_n}(x)=cdf_X(x) \quad \forall x: x\neq c.$$

Definition: Convergence in Probability

If the conditions in Theorem 5.2.1 are satisfied, we say that the random variable X_n converges in probability to constant c and denote it as $X_n \xrightarrow[n\to\infty]{p} c$ or $\operatorname{plim}_{n\to\infty} X_n=c$. Meaning,

$$\operatorname{plim}_{n\to\infty} X_n=c \Leftrightarrow \lim_{n\to\infty} P(|X_n-c|\geq\epsilon)=0 \quad \forall\epsilon>0$$

⟨Theorem 5.2.2. Law of Large Numbers⟩

If the random variables $X_1,\cdots,X_n$ are iid, and $E[|X|]<+\infty$,

$$p\lim_{n\to\infty} \frac{1}{n}\sum X_i = E[X_i]$$

(proof)

The theorem is true generally but its proof is beyond the scope of the book hence we will show the proof under the condition that $Var(x)<+\infty$

Apply the Chebyshev's inequality on $\bar{X}_n=(X_1+\cdots+X_n)/n$. Then for an arbitrary ϵ,

$$0\leq P(|\bar{X}_n-E(\bar{X}_n)|\geq\epsilon) \leq Var(\bar{X}_n)/\epsilon^2 = Var(X_i)/n\epsilon^2$$

$$\therefore \lim_{n\to\infty} P(|\bar{X}_n-E(X_i)|\geq\epsilon)=0$$

When we observe random samples $X_1,\cdots,X_n$, $\frac{1}{n}\sum_{i=1}^{n} I_A(X_i)$ is the relative frequency (상대도수) of observation that belongs to A. Apply the Law of Large Numbers to the relative frequency and we get

$$p\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I_A(X_i)=E[I_A(X_i)]=P(X_i\in A)$$

Meaning, as the number of trials increase, relative frequency approaches the probability indefinitely.

Think about the Law of Large Numbers in the context of statistical inference. Law of Large Numbers mean that as the sample size increases, the sample mean will go indefinitely closer to the population mean. In order to apply the concept of "convergence in probability" to more statistics, we must expand its definition to the multivariate case.

Definition: Convergence in Probability in the Multivariate Case.

For the multivariate random variable $X_n=(X_{n1},\cdots,X_{nK})^t$ and constant vector $C=(c_1,\cdots,c_K)^t$, let

$$\|X_n-C\| = \sqrt{(X_{n1}-c_1)^2+\cdots+(X_{nK}-c_K)^2}$$

If the following is true, we say that $X_n=(X_{n1},\cdots,X_{nK})^t$ converges to $C=(c_1,\cdots,c_K)^t$ and Similar to the one dimensional case, we denote it as $X_n \xrightarrow[n\to\infty]{p} C$ or $\operatorname{plim}_{n\to\infty} X_n=c$.

$$\lim_{n\to\infty} P(\|X_n-C\|\geq\epsilon)=0 \quad \forall\epsilon>0$$

〈Theorem 5.2.3〉

The following statements are equivalent to $\operatorname{plim}_{n\to\infty} X_n = C$, $X_n = (X_{n1}, \cdots, X_{nK})^t$ converges in probability to $C = (C_1, \cdots, C_K)^t$.

(a) $\lim_{n\to\infty} P(\max_{1\le i\le K} |X_{ni} - C_i| \ge \epsilon) = 0$

(b) $\operatorname{plim}_{n\to\infty} X_{n1} = C_1, \cdots, \operatorname{plim}_{n\to\infty} X_{nK} = C_K$

(proof)

(a) In the multivariate case, from the definition of $\|X_n - C\|$,

$$\max_{1\le i\le K} |X_{ni} - C_i| \le \|X_n - C\| \le K \max_{1\le i\le K} |X_{ni} - C|$$

$$\therefore P(\max_{1\le i\le K} |X_{ni} - C_i| \ge \epsilon) \le P(\|X_n - C\| \ge \epsilon) \le P(K \max_{1\le i\le K} |X_{ni} - C| \ge \epsilon)$$

$$= P(\max_{1\le i\le K} |X_{ni} - C| \ge \epsilon/K)$$

Hence, the convergence in probability and (a) are equivalent.

(b) Inclusion Relationship (포함관계)

$$(|X_{ni} - C_i| \ge \epsilon) \subseteq (\max_{1\le i\le K} |X_{ni} - C_i| \ge \epsilon) = \bigcup_{i=1}^{K} (|X_{ni} - C_i| \ge \epsilon) \quad (i = 1, \cdots, K)$$

From the precedent relationship, the following inequality can be derived.

$$\max_{1\le i\le K} P(|X_{ni} - C_i| \ge \epsilon) \le P(\max_{1\le i\le K} |X_{ni} - C_i| \ge \epsilon) \le \sum_{i=1}^{K} P(|X_{ni} - C_i| \ge \epsilon)$$

Hence, (a) and (b) are equivalent.

〈Theorem 5.2.4〉

If iid multivariate random variables $X_i = (X_{i1}, \cdots, X_{iK})^t, \cdots, X_n = (X_{n1}, \cdots, X_{nK})^t$ and $E(X_i) = (E(X_{i1}), \cdots, E(X_{iK}))^t$ can be defined,[7a]

$$\operatorname{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = E(X_i)$$

(proof)

The Theorem is evident from the Law of Large Numbers in the one-dimensional case and Theorem 5.2.3.

* 7a) From the definition of Expectation, this means $E(|X_{i1}|) < +\infty, \cdots, E(|X_{iK}|) < +\infty$

Ex. 5.2.4. Convergence in Probability of sample moment vector.

$$\hat{m}_r = \frac{1}{n} \sum_{i}^{n} X_i^r$$

The rth sample moment is a function of $X_1, \cdots, X_n$ and is used in the estimation of $m_r = E[X_i^r]$. Meanwhile from 〈Theorem 1.6.2〉 we have learned that if $E(|X^k|) < +\infty$, any lower moments can be defined. Hence from the Law of Large Numbers if $E(|X^k|) < +\infty$,

$$\operatorname{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (X_{i1}, \cdots, X_i^k)^t = (E(X_i), \cdots, E(X_i^k))^t$$

gets indefinitely closer to the population moment as the sample size grows.

〈Theorem 5.2.5. Continuous Function and Convergence in Probability〉

If the multivariate random variable $X_n = (X_{n1}, \cdots, X_{nK})^t$ converges in probability to the constant vector $C = (C_1, \cdots, C_K)^t$ and the real value function g is continuous on C, $g(X_n)$ converges in probability to $g(c)$. Meaning,

$$\text{if } \operatorname{plim}_{n\to\infty} X_n = C \Rightarrow \operatorname{plim}_{n\to\infty} g(X_n) = g(c)$$

(proof)

Since g is continuous on C,

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. if } |x - c| < \delta, |g(x) - g(c)| < \epsilon$$

for an arbitrary ϵ, choose δ satisfying the precedent inequality.

$$(\|X_n - C\| < \delta) \subseteq (|g(X_n) - g(c)| < \epsilon)$$

$$\therefore (|g(X_n) - g(c)| \ge \epsilon) \subseteq (\|X_n - C\| \ge \delta)$$

$$\therefore P(|g(X_n) - g(c)| \ge \epsilon) \le P(\|X_n - C\| \ge \delta)$$

$$\therefore 0 \le \lim_{n\to\infty} P(|g(X_n) - g(c)| \ge \epsilon) \le \lim_{n\to\infty} P(\|X_n - C\| \ge \delta) = 0$$

$$\therefore \lim_{n\to\infty} P(|g(X_n) - g(c)| \ge \epsilon) = 0 \quad \forall \epsilon > 0$$

〈Theorem 5.2.6. Convergence in Probability and operations $(+, -, \times, \div)$〉

The following are true given that X_n and Y_n converges in probability to a and b respectively.

(a) $\operatorname{plim}_{n\to\infty} (X_n + Y_n) = \operatorname{plim}_{n\to\infty} X_n + \operatorname{plim}_{n\to\infty} Y_n = a + b$

(b) $\operatorname{plim}_{n\to\infty} (X_n - Y_n) = \operatorname{plim}_{n\to\infty} X_n - \operatorname{plim}_{n\to\infty} Y_n = a - b$

(c) $\operatorname{plim}_{n\to\infty} (X_n \cdot Y_n) = \operatorname{plim}_{n\to\infty} X_n \cdot \operatorname{plim}_{n\to\infty} Y_n = a \cdot b$

(d) $\operatorname{plim}_{n\to\infty} (X_n / Y_n) = \operatorname{plim}_{n\to\infty} X_n / \operatorname{plim}_{n\to\infty} Y_n = a/b \quad (b \ne 0)$

(proof)

From 〈Theorem 5.2.3〉, the bivariate random variable $(X_n, Y_n)^t$ converges in probability to $(a, b)^t$. i.e.,

$$\operatorname{plim}_{n\to\infty} (X_n, Y_n)^t = (a, b)^t$$

Hence, from 〈Theorem 5.2.5〉, the convergence in probability is conserved if we apply a continuous bivariate function.

Meanwhile, the four operations are bivariate functions continuous in its range and therefore the theorem is applicable. Specifically, we can apply the following functions:

$$g_1(x, y) = x + y, \; g_2(x, y) = x - y, \; g_3(x, y) = x \cdot y, \; g_4(x, y) = x/y$$

Ex. 5.2.5. Convergence in Probability of Sample Variance and Standard deviation.

We can express $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ as a function of two sample moments $\left\{\sum_i^n X_i/n,\ \sum_i^n X_i^2/n\right\}^t$.

$$S_n^2 = \frac{1}{n-1}\sum_i^n\left(X_i^2 - 2X_i\bar{X}+\bar{X}^2\right)$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_i^n X_i^2 - \left(\frac{1}{n}\sum_i^n X_i\right)^2\right)$$

Meanwhile, from the Law of Large Numbers, if $E[X_i^2] < +\infty$,

$$\operatorname{plim}_{n\to\infty}\frac{1}{n}\sum_i^n X_i^2 = E[X_i^2],\quad \operatorname{plim}_{n\to\infty}\frac{1}{n}\sum_i^n X_i = E[X]$$

Thus, from ⟨Theorem 5.2.6⟩,

$$\operatorname{plim}_{n\to\infty} S_n^2 = \operatorname{plim}_{n\to\infty}\frac{n}{n-1}\left\{\frac{1}{n}\sum_i^n X_i^2 - \left(\frac{1}{n}\sum_i^n X_i\right)^2\right\} = 1\times\left\{E(X_i^2)-E(X_i)^2\right\}$$

Hence if $\sigma^2 = \mathrm{Var}(X_i) < +\infty$

$$\operatorname{plim}_{n\to\infty} S_n^2 = \sigma^2$$

In addition, standard deviation $S_n = \sqrt{S_n^2}$ and the square root function is a continuous function. From ⟨Theorem 5.2.5⟩ it follows that

$$\operatorname{plim}_{n\to\infty} S_n = \sqrt{\operatorname{plim}_{n\to\infty} S_n^2} = \sqrt{\sigma^2} = \sigma$$

The following Theorem is useful when it is difficult to prove the convergence in probability using the Law of Large Numbers.

⟨Theorem 5.2.7⟩

For random variables $X_n\ (n=1,2,\cdots)$ whose variance can be defined in $\mathbb{R}$, the following holds:

If $\lim_{n\to\infty}\mathrm{Var}(X_n)=0$, $\lim_{n\to\infty}E[X_n]=a$, $\Rightarrow \operatorname{plim}_{n\to\infty} X_n = a$

(proof)

For an arbitrary ϵ, the following is true due to Markov's Inequality in ⟨Theorem 1.6.4⟩.

$$P(|X_n-a|\geq\epsilon) \leq E[(X_n-a)^2]/\epsilon^2$$

Meanwhile, $E[(X_n-a)^2] = \mathrm{Var}(X_n-a) + E[X_n-a]^2$

$$= \mathrm{Var}(X_n) + (E[X_n]-a)^2$$

$$= 0+0 \quad \because \mathrm{Var}(X_n)=0 \text{ and } E[X_n]=a$$

$$\therefore\ 0 \leq \lim_{n\to\infty} P(|X_n-a|\geq\epsilon) \leq E[(X_n-a)^2]/\epsilon^2 = 0$$

Ex. 5.2.6.

Let $U_{(1)} < \cdots < U_{(n)}$ denote the order statistics of n random samples from $U(0,1)$. The pdf of $U_{(n)}$ is:

$$pdf_{U_{(n)}}(x) = n x^{n-1} I_{(0,1)}(x)$$

$$E[U_{(n)}] = \frac{n}{n+1},\quad \mathrm{Var}(U_{(n)}) = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2$$

$$\lim_{n\to\infty}\mathrm{Var}(U_{(n)}) = 0,\quad \lim_{n\to\infty}E[U_{(n)}] = 1$$

$$\therefore \operatorname{plim}_{n\to\infty} U_{(n)} = 1$$

Ex. 5.2.7. Convergence in Probability of Sample Quantile.

If the population distribution is continuous and it cdf $F(x)$ is strictly increasing[74], the value at which the cdf equals a, or simply $F^{-1}(a)$, is called the a quantile of the population. Meanwhile, let $X_{(1)},\cdots,X_{(n)}$ denote the order statistics of n iid random samples.

$$r_n \sim an,\ \text{i.e.,}\ \lim_{n\to\infty} r_n/n = a\quad (0<a<1)$$

for $r_n\in\mathbb{N}$, $X_{(r_n)}$ is called the "sample quantile"[75]. From ⟨Theorem 4.3.4⟩, the distribution of sample quantile is:

$$X_{(r_n)} \overset{d}{\equiv} h\left(\frac{1}{n}Z_1+\cdots+\frac{1}{n-r_n+1}Z_{r_n}\right),\quad Z_i \overset{iid}{\sim} Exp(1)\quad (i=1,\cdots,n)$$

Here, $h(y) = F^{-1}(1-e^{-y})\quad (y>0)$

Meanwhile, the expectation and variance of standard exponential distribution is 1. Let

$$Y_n = \frac{1}{n}Z_1+\cdots+\frac{1}{n-r_n+1}Z_{r_n}$$

$$E[Y_n] = \frac{1}{n}+\cdots+\frac{1}{n-r_n+1},\quad \mathrm{Var}(Y_n) = \frac{1}{n^2}+\cdots+\left(\frac{1}{n-r_n+1}\right)^2$$

Approximate the Sum:

$$\frac{1}{n}+\cdots+\frac{1}{n-r_n+1} = \frac{1}{n}\sum_{k=0}^{r_n-1}\frac{1}{1-k/n} \sim \int_0^a \frac{1}{1-x}\,dx = -\log(1-a)$$

$$\frac{1}{n^2}+\cdots+\frac{1}{(n-r_n+1)^2} = \frac{1}{n}\frac{1}{n}\sum_{k=0}^{r_n-1}\frac{1}{(1-k/n)^2} \sim \frac{1}{n}\int_0^a \frac{1}{(1-x)^2}\,dx = \frac{1}{n}\frac{a}{1-a}$$

Hence,

$$\lim_{n\to\infty}\mathrm{Var}(Y_n) = \lim_{n\to\infty}\frac{1}{n}\frac{a}{1-a} = 0,\quad E[Y_n] = -\log(1-a)$$

$$\therefore \operatorname{plim}_{n\to\infty} Y_n = \operatorname{plim}_{n\to\infty}\left(\frac{1}{n}Z_1+\cdots+\frac{1}{n-r_n+1}Z_{r_n}\right) = -\log(1-a)$$

Therefore, if the function $h(y) = F^{-1}(1-e^{-y})\ (y>0)$ is continuous, from ⟨Theorem 5.2.5⟩,

$$\operatorname{plim}_{n\to\infty} h\left(\frac{1}{n}Z_1+\cdots+\frac{1}{n-r_n+1}Z_{r_n}\right) = h(-\log(1-a)) = F^{-1}(a)$$

and from ⟨Theorem 5.2.1⟩'s necessary conditions,

$$h\left(\frac{1}{n}Z_1+\cdots+\frac{1}{n-r_n+1}Z_{r_n}\right) \xrightarrow[n\to\infty]{d} F^{-1}(a)$$

Since $X_{[r_n]} \overset{d}{\equiv} h(\frac{1}{n}Z_1 + \cdots + \frac{1}{n-r_n+1}Z_{r_n})$

$X_{[r_n]} \xrightarrow[n\to\infty]{d} F^{-1}(a)$

Hence if F^{-1} is continuous, from the necessary conditions of 〈Theorem 5.2.1〉,

$$p\lim_{n\to\infty} X_{[r_n]} = F^{-1}(a)$$

7-4) 확률적분변환에 관한 〈정리 4.3.3〉에서와 마찬가지로 누적분포함수가 순증가가 아닌 경우에는 F의 역변환을

$$F^{-1}(u) = \inf\{x : F(x) \geq u\} \quad (0 \leq u \leq 1)$$

로 정의하고, 일반적으로는 이 역변환을 이용하여 분위수를 정의한다.

7-5) 표본분위수에 대한 이러한 정의는 표본크기가 충분히 큰 경우에 사용하는 정의이며 이는 〈예 4.1.2〉의 표본 중앙값의 정의와 다른 것을 알 수 있다. 이러한 정의는 확률수렴과 같은 극한 성질을 간략히 밝히기 위한 것이며, 일반적인 정의에 따른 표본분위수의 경우에도 오일단의 분위수로 확률수렴하는 것이 알려져 있다.

Slutsky's Theorem : Shinkook's proof

WTS: $P(X_n+Y_n<x) \xrightarrow[n\to\infty]{d} P(Z+c<x)$ — introduce Y_n in (a)·(b), introduce X_n in (c)·(d)

$P(X_n+Y_n<x) = P(X_n+Y_n<x,\ Y_n<c-\epsilon) + P(X_n+Y_n<x,\ Y_n>c-\epsilon)$

$\leq P(Y_n<c-\epsilon) + P(X_n<x-c+\epsilon)$

$\therefore \lim_{n\to\infty} P(X_n+Y_n<x) \leq \lim_{n\to\infty} P(Y_n<c-\epsilon) + P(X_n<x-c+\epsilon) = 0 + P(Z<x-c+\epsilon)$

$= P(Z+c<x+\epsilon)$

$P(X_n+Y_n>x) = P(X_n+Y_n>x,\ Y_n<c+\epsilon) + P(X_n+Y_n>x,\ Y_n>c+\epsilon)$

$\leq P(X_n>x-c+\epsilon) + P(Y_n>c+\epsilon)$

$\therefore \lim_{n\to\infty} P(X_n+Y_n>x) \leq \lim_{n\to\infty} P(X_n>x-c-\epsilon) + P(Y_n>c+\epsilon)$

$= P(Z>x-c-\epsilon) + 0$

$= P(Z+c>x-\epsilon)$

$\therefore \lim_{n\to\infty} P(X_n+Y_n<x) \geq P(Z+c<x-\epsilon)$

Conclusion: $P(Z+c<x-\epsilon) \leq \lim_{n\to\infty} P(X_n+Y_n<x) \leq P(Z+c<x+\epsilon)$

$\epsilon>0$ is arbitrary. $\therefore \lim_{n\to\infty} X_n+Y_n = Z+c$

(c) $\lim_{n\to\infty} X_nY_n = c\cdot Z$

* $(c=0)$ and $(c\neq 0)$ case.

if $c=0$, $P(|X_nY_n|>x) = P(|X_nY_n|>x, |X_n|>K) + P(|X_nY_n|>x, |X_n|<K)$

$\leq P(|X_n|>K) + P(|Y_n|>x/K)$

$\lim_{n\to\infty} P(|X_nY_n|>x) \leq \lim_{n\to\infty} P(|Y_n|>x/K) + P(|X_n|>K) = 0 + P(|X_n|>K) \xrightarrow[K\to+\infty]{} 0+0=0$

if $c\neq 0$, $\lim_{n\to\infty} X_n(Y_n-c) = 0 \ \therefore \lim_{n\to\infty} X_nY_n - cX_n = 0 \Leftrightarrow \lim_{n\to\infty} X_nY_n = \lim_{n\to\infty} cX_n = cZ$

5.3 CALCULATION OF LIMITING DISTRIBUTIONS.

☆☆ 〈Theorem 5.3.1. Slutsky's Theorem〉

For random variables X_n, Y_n $(n=1,2,\cdots)$, Z and constant c, if

$$X_n \xrightarrow[n\to\infty]{d} Z, \quad p\lim_{n\to\infty} Y_n = c$$

the following are true:

(a) $X_n+Y_n \xrightarrow[n\to\infty]{d} Z+c$ (b) $X_n-Y_n \xrightarrow[n\to\infty]{d} Z+c$

(c) $Y_nX_n \xrightarrow[n\to\infty]{d} cZ$ (d) $X_n/Y_n \xrightarrow[n\to\infty]{d} Z/c \quad (c\neq 0)$

(proof)

The theorem is true generally but here we will prove the theorem given that the cdf of Z is continuous.

$P(X_n+Y_n\leq z) = P(X_n+Y_n\leq z,\ |Y_n-c|<\epsilon) + P(X_n+Y_n\leq z,\ |Y_n-c|\geq\epsilon)$

$P(X_n+Y_n\leq z,\ -\epsilon<Y_n-c<\epsilon)$
$= P(X_n+Y_n\leq z,\ c-\epsilon<Y_n<c+\epsilon)$
$\leq P(X_n+Y_n\leq z,\ c-\epsilon<Y_n)$

$P(X_n+Y_n\leq z,\ |Y_n-c|\geq\epsilon)$
$\leq P(|Y_n-c|\geq\epsilon)$

$\leq P(X_n+Y_n\leq z,\ c-\epsilon<Y_n) + P(|Y_n-c|\geq\epsilon)$

$\overline{\lim} P(X_n+Y_n\leq z) = \lim_{n\to\infty} P(X_n+Y_n\leq z,\ c-\epsilon<Y_n) + P(|Y_n-c|\geq\epsilon)$

$= \lim_{n\to\infty} P(X_n\leq z-Y_n,\ c-\epsilon<Y_n) + P(|Y_n-c|\geq\epsilon)$

$\leq \lim_{n\to\infty} P(X_n\leq z-(c-\epsilon)) = \lim_{n\to\infty} P(X_n\leq z-c+\epsilon) = P(Z\leq z-c+\epsilon)$

$\therefore \overline{\lim} P(X_n+Y_n\leq z) \leq P(Z\leq z-c+\epsilon) \xrightarrow[\epsilon\downarrow 0]{} P(Z+c\leq z)$

Similarly, $P(X_n+Y_n\geq z) = P(X_n+Y_n\geq z,\ |Y_n-c|<\epsilon) + P(X_n+Y_n\geq z,\ |Y_n-c|\geq\epsilon)$

$\leq P(Z\geq z-c+\epsilon) + P(|Y_n-c|\geq\epsilon)$

$\therefore \underline{\lim} P(X_n+Y_n\geq z) \leq P(X_n+Y_n\geq z+\epsilon) \xrightarrow[\epsilon\downarrow 0]{} P(Z+c\geq z)$

$\therefore \underline{\lim} P(X_n+Y_n\leq z) \geq P(Z+c\leq z)$

Hence, under the condition that the cdf of Z is continuous,

$\lim_{n\to\infty} P(X_n+Y_n\leq z) = P(Z+c\leq z)$ (∵ Squeeze theorem)

(b) if $p\lim_{n\to\infty} Y_n = c$, $p\lim_{n\to\infty}(-Y_n) = -c$ and using (a) we can prove (b)

Steps of proof
① Introduce $|Y_n-c|\geq\epsilon$ and $|Y_n-c|<\epsilon$ to the probability
② Eliminate $X_n+Y_n\geq Z$ in the 2nd probability.
③ use $(c-\epsilon<Y_n<c+\epsilon)\subset(c-\epsilon<Y_n)$
④ substitute Y_n with $c-\epsilon$
⑤ Find upper/lower bound and use squeeze theorem

→ when finding lower bound, start with $X_n+Y_n>Z$ then later switch the

(c) If $C=0$, for an arbitrary $\epsilon>0$,

$$P(|Y_nX_n|\geq\epsilon) = P(|Y_nX_n|\geq\epsilon, |X_n|>K) + P(|Y_nX_n|\geq\epsilon, |X_n|\leq K)$$
$$\leq P(|X_n|>K) + P(|Y_n||X_n|\geq\epsilon, |X_n|\leq K)$$
$$\leq P(X_n>K) + P(X_n<-K) + P(K|Y_n|\geq\epsilon)$$
$$= P(X_n>K) + P(X_n<-K) + P(|Y_n|\geq\epsilon/K)$$

$\lim_{n\to\infty} X_n \xrightarrow{d} Z$ and $p\lim_{n\to\infty} Y_n = C$

$$\therefore \Rightarrow P(Z>K) + P(Z<-K) + 0 \xrightarrow{K\to\infty} 0+0$$

$$\therefore\ 0 \leq \lim_{n\to\infty} P(|X_nY_n|>\epsilon) \leq P(Z>K) + P(Z<-K) \xrightarrow[K\to\infty]{} 0$$

textbook에는 ≤ ?

$$\therefore\ p\lim_{n\to\infty} Y_nX_n = 0$$

Key Ideas
① Divide the problem to 2 cases : $C=0$ vs $C\neq 0$
② Introduce $|X|>K$ and $|X|<K$ to the probability
③ Control $K\to\infty$
④ If $C\neq 0$, $Y_n - C \to 0$ let $C^* = Y_n - C$

($C\neq 0$)

$p\lim_{n\to\infty}(Y_n-C)=0$ and $X_n \xrightarrow[n\to\infty]{d} Z$

$$\therefore\ p\lim_{n\to\infty}(Y_n-C)X_n = 0 \quad \text{from the precedent proof.}$$

Also, from the conditions given for the cdf of Z, $cX_n \xrightarrow[n\to\infty]{d} cZ$ hence from (a),

$$Y_nX_n = (Y_n-C)X_n + cX_n \xrightarrow[n\to\infty]{d} 0 + cZ = cZ$$

(d) If $p\lim_{n\to\infty} Y_n = C\ (C\neq 0)$, $p\lim_{n\to\infty} 1/Y_n = 1/C$ and since (c) is true, (d) holds.

Ex 5.3.1 The asymptotic distribution of Studentized Sample mean.

Let $\bar{X}_n$ and S_n respectively denote sample mean and sample standard deviation of n random samples sampled from a distribution with mean μ and standard deviation σ.

$$(\bar{X}_n-\mu)/(S_n/\sqrt{n})$$

is called the "Studentized" sample mean. Meanwhile from the CLT and <Ex. 5.2.5>

$$\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{d} Z,\ Z\sim N(0,1) \quad \text{and} \quad p\lim_{n\to\infty} S_n = \sigma$$

Hence, from Slutsky's Theorem,

$$\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \Big/ \frac{S_n}{\sigma} \xrightarrow[n\to\infty]{d} Z/1 = Z\sim N(0,1)$$

Especially, if the population distribution is normal, the studentized sample mean follows the t-distribution with degrees of freedom $n-1$. Thus, results above imply that the t-distribution approximates to the standard normal (for large dfs) distribution and it is denoted as:

$$t(n-1) \approx N(0,1),\ n\to\infty$$

In addition, regardless of the population distribution, if the variance can be defined ($\sigma^2>0$),

$$\lim_{n\to\infty}\left(-z_{\alpha/2} < \frac{\bar{X}_n-\mu}{S_n/\sqrt{n}} \leq z_{\alpha/2}\right) = 1-\alpha$$

$$\therefore P\left\{\bar{X}_n - z_{\alpha/2}S_n/\sqrt{n} \leq \mu \leq \bar{X}_n + z_{\alpha/2}S_n/\sqrt{n}\right\} \simeq 1-\alpha,\ n\to\infty$$

Hence, the $100(1-\alpha)\%$ asymptotic confidence interval of the population mean μ is given as:

$$\mu \in \left[\bar{X}_n - z_{\alpha/2}S_n/\sqrt{n},\ \bar{X}_n + z_{\alpha/2}S_n/\sqrt{n}\right)$$

Ex. 5.3.2. Asymptotic Distribution of Sample Variance.

Consider the sample distribution of sample variance, S_n^2, of a distribution with mean μ and standard deviation σ ($0<\sigma<+\infty$). Let $\bar{X}_n$ denote the sample mean.

$$S_n^2 = \frac{1}{n-1}\sum_i^n\left[(X_i-\mu)-(\bar{X}_n-\mu)\right]^2 = \frac{n}{n-1}\left\{\frac{1}{n}\sum_i^n(X_i-\mu)^2 - (\bar{X}_n-\mu)^2\right\}$$

Meanwhile, let $Y_i = (X_i-\mu)^2$ $(i=1,2,\cdots,n)$. Since Y_i are iid, from the CLT, if $\mathrm{Var}(Y_i)<+\infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_i^n Y_i - E(Y_i)\right) \xrightarrow[n\to\infty]{d} W,\ W\sim N(0,\mathrm{Var}(Y_i))$$

$E[Y_i]=\sigma^2$, $\mathrm{Var}(Y_i) = E[Y_i^2]-E[Y_i]^2 = E[(X_i-\mu)^4]-\sigma^4$

$$\sqrt{n}\left(\frac{1}{n}\sum_i^n Y_i - \sigma^2\right) \xrightarrow[n\to\infty]{d} W,\ W\sim N(0, E[(X_i-\mu)^4]-\sigma^4)$$

From the CLT and LLN (Law of Large Numbers),

$$\sqrt{n}(\bar{X}_n-\mu) \xrightarrow[n\to\infty]{d} Z,\ Z\sim N(0,\sigma^2),\quad p\lim_{n\to\infty}(\bar{X}_n-\mu)=0$$

Thus from Slutsky's theorem and <Theorem 5.2.1>,

$$\sqrt{n}(\bar{X}_n-\mu)(\bar{X}_n-\mu) \xrightarrow[n\to 0]{d} 0\times Z,\ Z\sim N(0,\sigma^2)$$

$$\therefore\ p\lim_{n\to\infty}\sqrt{n}(\bar{X}_n-\mu)^2 = 0$$

Hence, if $E[(X_i-\mu)^4]<+\infty$, let $W\sim N(0, E[(X_i-\mu)^4]-\sigma^4)$

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X}_n)^2-\sigma^2\right) = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^2-\sigma^2\right\} - \sqrt{n}(\bar{X}_n-\mu)^2 \xrightarrow[n\to\infty]{d} W-0=W$$

Thus if $E[(X_i-\mu)^4]<+\infty$, let $\rho_4 = E\left[\left(\frac{X_i-\mu}{\sigma}\right)^4\right]-3$ 76)

$$\sqrt{n}(S_n^2-\sigma^2)$$
$$= \sqrt{n}\left\{\frac{1}{n}\sum_i^n(X_i-\bar{X})^2-\sigma^2\right\} + \frac{1}{\sqrt{n}}S_n^2 \xrightarrow[n\to\infty]{d} W,\ W\sim N(0,(\rho_4+2)\sigma^4)$$

$$\sqrt{n}(S_n^2-\sigma^2) = \sqrt{n}\left(\frac{1}{n-1}\sum(X_i-\bar{X})^2-\sigma^2\right)$$
$$= \sqrt{n}\left(\frac{1}{n-1}\sum(X_i-\mu)^2 - \frac{n}{n-1}(\bar{X}_i-\mu)^2 - \sigma^2\right)$$
$$\approx \sqrt{n}\left(\frac{1}{n}\sum(X_i-\mu)^2-\sigma^2\right) - \underbrace{\sqrt{n}(\bar{X}_i-\mu)^2}_{\to W\sim N(0,(\rho_4+2)\sigma^4)}$$

$\mathrm{Var}([X_i-\mu]^2) = E[(X_i-\mu)^4]-E[(X_i-\mu)^2]^2$

$\sqrt{n}(\bar{X}_i-\mu)\times(\bar{X}_i-\mu) \to Z\times 0 = 0$, $Z\sim N(0,\sigma^2)$

Many times in finding the asymptotic distribution of statistics, we utilize the CLT as ⟨Ex.5.3.1⟩ and ⟨Ex.5.3.2⟩. In these cases, oftentimes the multivariate CLT is used, given in ⟨Theorem 5.1.2⟩. The definition of asymptotic distribution for multivariate random variables are given below.

Definition: Multivariate Asymptotic Distribution.

For columns of the multivariate random variables $X_n=(X_{n1},\cdots,X_{nK})^t$ $(n=1,2\cdots)$ and $Z=(Z_1,\cdots,Z_K)^t$, if

$$\lim_{n\to\infty} P(X_{n1}\le x_1,\cdots X_{nK}\le x_K)=P(Z_1\le x_1,\cdots,Z_K\le x_K) \quad {}^{77)}$$

is true for all continuous points $x=(x_1,\cdots,x_K)^t$ in $cdf_Z(x_1,\cdots,x_K)$, we say that Z is the limiting distribution or the asymptotic distribution of $X_n=(X_{n1},\cdots,X_{nK})^t$. It is denoted as:

$$X_n \xrightarrow[n\to\infty]{d} Z \quad \text{or} \quad (X_{n1},\cdots,X_{nK})^t \xrightarrow[n\to\infty]{d} (Z_1,\cdots,Z_K)^t$$

76) 여기에서 정의된 β_4를 모집단의 첨예도(kurtosis)라고 하며 이는 모집단 분포가 평균 부근에 밀집되어 있는 정도를 나타내는 척도이고, 모집단의 분포가 정규분포인 경우에 첨예도는 0 이다.

77) 이러한 결합확률을 간략히 $P(Z_1\le x_1,\cdots,Z_K\le x_K)=P(Z\le x)$로 나타내기로 한다.

⟨Theorem 5.3.2. Continuous Function and Limiting Distribution⟩

For multivariate random variables $X_n=(X_{n1},\cdots,X_{nK})^t$ $(n=1,2,\cdots)$ and $Z=(Z_1,\cdots,Z_K)^t$, if

$$(X_{n1},\cdots,X_{nK})^t \xrightarrow[n\to\infty]{d} (Z_1,\cdots,Z_K)^t$$

for continuous function g,

$$g(X_n) \xrightarrow[n\to\infty]{d} g(Z)$$

(proof) The proof of this theorem is beyond the scope of this book.

Ex.5.3.3. Binomial Distribution and Chi-squared Approximation.

From the normal approximation of Binomial distribution we know that if $X_n\sim B(n,p)$,

$$\frac{X_n-np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{d} Z,\ Z\sim N(0,1)$$

In addition the square function is continuous. Hence from ⟨Theorem 5.3.2⟩,

$$\left(\frac{X_n-np}{\sqrt{np(1-p)}}\right)^2 \xrightarrow[n\to\infty]{d} Z^2,\ Z\sim N(0,1)$$

$$\therefore \left(\frac{X_n-np}{\sqrt{np(1-p)}}\right)^2 \xrightarrow[n\to\infty]{d} W,\ W\sim\chi^2(1)$$

Meanwhile, let $X_{n1}=X_n$ and $X_{n2}=n-X_n$, $p_1=p$ and $p_2=1-p$ then,

$$(X_{n1},X_{n2})^t\sim \text{Multi}(n,(p_1,p_2)^t)$$

and we know that the following equality holds:

$$\frac{(X_{n1}-np_1)^2}{np_1}+\frac{(X_{n2}-np_2)^2}{np_2}=\frac{(X_n-np)^2}{np}+\frac{(n-X_n-n(1-p))^2}{n(1-p)}$$

$$=\frac{(X_n-np)^2q+p[n-X_n-nq]^2}{npq}=\frac{(X_n-np)^2q+p(X_n+nq-n)^2}{npq}$$

$$=\frac{(X_n-np)^2q+p[X_n-n(1-q)]^2}{npq}=\frac{(X_n-np)^2q+(X_n-np)^2p}{npq}$$

$$=\frac{(X_n-np)^2}{npq}$$

Hence, $(X_{n1},X_{n2})^t\sim \text{Multi}(n,(p_1,p_2)^t)$ and

$$\sum_1^2\frac{(X_{nj}-np_j)^2}{np_j}\xrightarrow[n\to\infty]{d} W,\ W\sim\chi^2(1)$$

Ex.5.3.4. Multinomial Distribution and Chi-squared Approximation.

Let's generalize ⟨Ex.5.3.3⟩

$$(X_{n1},\cdots,X_{nK})^t\sim \text{Multi}(n,(p_1,\cdots,p_K)^t)\quad (p_1+\cdots+p_K=1,\ p_i>0,\ i=1,\cdots,K)$$

We will show: $\sum_{i=1}^{n}\frac{(X_{nj}-np_j)^2}{np_j}\xrightarrow[n\to\infty]{d} W,\ W\sim\chi^2(K-1)$

Here, let $X_n=(X_{n1},\cdots,X_{nK})^t$, $p=(p_1,\cdots,p_K)^t$. From ⟨Ex.5.1.2⟩ and the multivariate CLT,

$$(X_n-np)/\sqrt{n}\xrightarrow[n\to\infty]{d} X,\ X\sim N_k(0,D(p)-pp^t)$$

$* \ Var(X)=n\{D(p)-pp^t\}$

$$\therefore \frac{\bar{X_n}-\bar{X}}{S_n/\sqrt{n}}\sim N(0,I)\Rightarrow\frac{\bar{X_n}-\bar{X}}{1/\sqrt{n}}\sim N(0,D(p)-pp^t)\Rightarrow\frac{\bar{X_n}-\bar{X}}{\sqrt{n}}\sim N(0,D(p)-pp^t)$$

Let $Y_{nj}=X_{nj}$, $\mu_j=p_j$ $(j=1,\cdots,r)$, $Y_n=(Y_{n1},\cdots,Y_{nr})^t$, $\mu=(\mu_1,\cdots,\mu_r)^t$, $r=K-1$

$D(\mu_j)$: Matrix with i^{th} diagonal entry μ_i

$$(Y_n-n\mu)/\sqrt{n}\xrightarrow[n\to\infty]{d} Z,\ Z\sim N_r(0,\Sigma),\ \Sigma=D(\mu_j)-\mu\mu^t$$

In addition, $z^t\Sigma^{-1}z$ is a continuous function hence from ⟨Theorem 5.3.2⟩,

$$\{(Y_n-n\mu)/\sqrt{n}\}^t\Sigma^{-1}\{(Y_n-n\mu)/\sqrt{n}\}\xrightarrow[n\to\infty]{d} z^t\Sigma^{-1}z,\ z\sim N_r(0,\Sigma)$$

From ⟨Theorem 4.4.5⟩, if $z\sim N_r(0,\Sigma)$, $z^t\Sigma^{-1}z\sim\chi^2(r)$, $r=K-1$

Therefore:

$$\{(Y_n-n\mu)/\sqrt{n}\}^t\Sigma^{-1}\{(Y_n-n\mu)/\sqrt{n}\}\xrightarrow[n\to\infty]{d} W,\ W\sim\chi^2(K-1)$$

Meanwhile, if we utilize the operations on special types of matrices similar to (Ex. 4.4.5)[78],

$$\Sigma^{-1} = (D(\mu_j) - \mu\mu^t)^{-1} = (I - \underbrace{D(\mu_j)^{-1}\mu}_{a}\underbrace{\mu^t}_{b^t})D^{-1}(\mu_j) = (I + \frac{\overbrace{D(\mu_j)^{-1}\mu\mu^t}^{ab^t}}{1-\underbrace{\mu^t D^{-1}(\mu_j)\mu}_{b^t a}})D^{-1}(\mu_j)$$

* $D^{-1}(\mu_j)\mu = \begin{bmatrix}\mu_1^{-1} & & \\ & \ddots & \\ & & \mu_r^{-1}\end{bmatrix}\begin{bmatrix}\mu_1 \\ \vdots \\ \mu_r\end{bmatrix} = \begin{bmatrix}1 \\ \vdots \\ 1\end{bmatrix} = \mathbb{1}$,

$\mu^t D^{-1}(\mu_j) = [1 \cdots 1] = \mathbb{1}^t$

$b^t a = 1 + [\mu_1 \cdots \mu_r]\begin{bmatrix}\mu_1^{-1} & & \\ & \ddots & \\ & & \mu_r^{-1}\end{bmatrix}\begin{bmatrix}\mu_1 \\ \vdots \\ \mu_r\end{bmatrix} = \mu_1 + \cdots + \mu_r = 1 - p_K$

$$\therefore \Sigma^{-1} = (I + \frac{D(\mu_j)^{-1}\mu\mu^t}{1-\mu^t D^{-1}(\mu_j)\mu})D^{-1}(\mu_j) = D^{-1}(\mu_j) + \frac{\mathbb{1}\mathbb{1}^t}{1-(1-p_K)} = D^{-1}(\mu_j) + \frac{1}{p_K}\mathbb{1}\mathbb{1}^t$$

[78] (Refer to Appendix II) for vector a,b if $1 + b^t a \neq 0$

$$(I + ab^t)^{-1} = I + cab^t,\ c = -\frac{1}{1+b^t a}$$

$$\{(Y_n - n\mu)/\sqrt{n}\}^t \Sigma^{-1}\{(Y_n - n\mu)/\sqrt{n}\}$$, Let $Y_{nj}^* = (Y_{nj} - n\mu_j)$

$$= \frac{1}{n}\begin{bmatrix}Y_{n1}^* & \cdots & Y_{nr}^*\end{bmatrix}\begin{bmatrix}p_1^{-1}+p_K^{-1} & p_K^{-1} & \cdots & p_K^{-1} \\ p_K^{-1} & p_2^{-1}+p_K^{-1} & \cdots & p_K^{-1} \\ \vdots & p_K^{-1} & \cdots & p_K^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_K^{-1} & p_K^{-1} & \cdots & p_{K-1}^{-1}+p_K^{-1}\end{bmatrix}\begin{bmatrix}Y_{n1}^* \\ \vdots \\ Y_{nr}^*\end{bmatrix}$$

$$= \frac{1}{n}\left[p_1^{-1}Y_{n1}^* + p_K^{-1}(Y_{n1}^* + \cdots + Y_{nr}^*),\ p_2^{-1}Y_{n2}^* + p_K^{-1}(Y_{n1}^* + \cdots + Y_{nr}^*), \cdots,\ p_{K-1}^{-1}Y_{nr}^* + p_K^{-1}(Y_{n1}^* + \cdots + Y_{nr}^*)\right]\begin{bmatrix}Y_{n1}^* \\ \vdots \\ Y_{nr}^*\end{bmatrix}$$

$$= \frac{1}{n}\left[\sum_{j=1}^{K-1} p_j^{-1}Y_{nj}^{*2} + (Y_{n1}^* + \cdots + Y_{nr}^*)^2 p_K^{-1}\right]$$

$$Y_{n1}^* + \cdots + Y_{nr}^* = (Y_{n1} - n\mu_1) + \cdots + (Y_{nr} - n\mu_r)$$
$$= (Y_{n1} + \cdots + Y_{nr}) - n(\mu_1 + \cdots + \mu_r)$$
$$= (n - X_{nK}) - n(1 - p_K)$$
$$= n - X_{nK} - n + np_K$$

$$\therefore (Y_{n1}^* + \cdots + Y_{nr}^*)^2 = (-X_{nK} + np_K)^2 = (X_{nK} - np_K)^2$$

$$\therefore \Rightarrow \frac{1}{n}\left[\sum_{j=1}^{K-1}\frac{(Y_{nj} - n\mu_j)^2}{p_j} + \frac{(X_{nK} - np_K)^2}{p_K}\right] = \sum_{j=1}^{K-1}\frac{(Y_{nj} - n\mu_j)^2}{np_j} + \frac{(X_{nK} - np_K)^2}{np_K}$$

$$\therefore \sum_{j=1}^{K-1}\frac{(Y_{nj} - n\mu_j)^2}{np_j} + \frac{(X_{nK} - np_K)^2}{np_K} \xrightarrow[n\to\infty]{d} W,\ W \sim \chi^2(K-1)$$

⟨Theorem 5.3.3. Calculation of the limiting distribution using first order approximation⟩

For multivariate random variables $X_n = (X_{n1}, \cdots, X_{nK})^t$ $(n = 1, 2, \cdots)$, Z and vector $\theta = (\theta_1, \cdots, \theta_K)^t$, if

$$\sqrt{n}(X_n - \theta) \xrightarrow[n\to\infty]{d} Z$$

and the first order partial derivative $\dot{g}(\theta)$ is continuous, the following holds:

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow[n\to\infty]{d} (\dot{g}(\theta))^t Z$$

79) In the one-dimensional case, $\dot{g}(\theta) = \partial g(\theta)/\partial\theta$, in the multivariate case, $\dot{g}(\theta) = (\partial g(\theta)/\partial\theta_1, \cdots, \partial g(\theta)/\partial\theta_K)^t$.

(proof)

We will show the proof in one-dimensional case since the proof is similar in the multivariate case.

Since the function $g(\theta)$ is differentiable, there exists an $r(x)\,(= r_\theta(x))$ which satisfies:

$$g(x) = g(\theta) + \{\dot{g}(\theta) + r(x)\}(x - \theta),\quad r(x) \xrightarrow[x\to\theta]{} 0$$

↳ Taylor series expansion.

We know that the following is true for $r(x)$:

$$\forall\epsilon > 0,\ \exists\delta \text{ such that if } |x-\theta| < \delta \Rightarrow |r(x)| < \epsilon$$

Hence, for an arbitrary ϵ, select δ which satisfies the above

$$(|X_n - \theta| < \delta) \subseteq (|r(X_n)| < \epsilon) \Rightarrow (|r(X_n)| \geq \epsilon) \subseteq (|X_n - \theta| \geq \delta)$$

$$\therefore P(|r(X_n)| \geq \epsilon) \leq P(|X_n - \theta| \geq \delta)$$

Meanwhile, it is given that $\sqrt{n}(X_n - \theta) \xrightarrow[n\to\infty]{d} Z$. Thus from Slutsky Theorem an ⟨Theorem 5.2.1⟩,

$$X_n - \theta = \frac{1}{\sqrt{n}} \times \sqrt{n}(X_n - \theta) \xrightarrow[n\to\infty]{d} 0 \times Z = 0$$

$$\therefore \operatorname*{plim}_{n\to\infty} X_n = \theta \quad \text{i.e.,} \quad \lim_{n\to\infty} P(|X_n - \theta| \geq \delta) = 0$$

$$\therefore 0 \leq \lim_{n\to\infty} P(|r(X_n)| \geq \epsilon) \leq \lim_{n\to\infty} P(|X_n - \theta| \geq \delta) = 0$$

$$\therefore \operatorname*{plim}_{n\to\infty} r(X_n) = 0$$

Therefore, $\sqrt{n}(g(X_n) - g(\theta)) = (\dot{g}(\theta) + r(X_n))\sqrt{n}(X - \theta)$, $\operatorname*{plim}_{n\to\infty}(\dot{g}(\theta) + r(X_n)) = \dot{g}(\theta)$

Finally, from Slutsky's Theorem,

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow[n\to\infty]{d} \dot{g}(\theta)Z$$

· Key Takeaways

① expand $g(X_n)$ into taylor series $g(X_n) = g(\theta) + \dot{g}(\theta)(X-\theta) + r(x)(X-\theta)^2$

② $r(x)$ is continuous. Use the δ-ϵ definition to show that $r(X_n) \to 0$ as $X_n \to \theta$

③ Use slutsky's Theorem to show: $\sqrt{n}(g(x) - g(\theta)) \xrightarrow[n\to\infty]{d} \dot{g}(\theta)Z$

Ex. 5.3.5. Limiting Distribution of Sample Standard deviation.

Let S_n^2 denote the sample variance of random samples from a distribution with mean μ and standard deviation σ $(0<\sigma<+\infty)$.
From ⟨Ex. 5.3.2⟩ we have learned that the distribution of sample variance is as the following, given that $E[(X_i-\mu)^4]<+\infty$:

$$\sqrt{n}(S_n^2-\sigma^2) \xrightarrow[n\to\infty]{d} W,\quad W\sim N(0,(\rho_4+2)\sigma^4),\quad \rho_4 = E\left[\left(\frac{X_i-\mu}{\sigma}\right)^4\right]-3$$

Sample standard deviation S_n is the square root of sample variance, $\sqrt{S_n^2}$, and the square root function $g(x)=\sqrt{x}$ has the first-order derivative $\dot{g}(x)=1/(2\sqrt{x})$. Thus, given that $E[(X_i-\mu)^4]<+\infty$, from ⟨Theorem 5.3.3⟩,

$$\sqrt{n}(\sqrt{S_n^2}-\sqrt{\sigma^2}) \xrightarrow[n\to\infty]{d} \frac{1}{2\sqrt{\sigma^2}}W,\quad W\sim N(0,(\rho_4+2)\sigma^4)$$

$$\therefore \sqrt{n}(S_n-\sigma) \xrightarrow[n\to\infty]{d} Z,\quad Z\sim N(0,(\rho_4+2)\sigma^2/4)$$

Ex. 5.3.6 Limiting Distribution of Sample Correlation coefficient.

Consider a bivariate population in which mean: $(\mu_1,\mu_2)^t$, variance: σ_1^2, σ_2^2 $(0<\sigma_1^2<+\infty,\ 0<\sigma_2^2<+\infty)$ and correlation coefficient: ρ $(0<\rho<1)$. Let $(X_1,Y_1)^t,\cdots,(X_n,Y_n)^t$ denote the random samples from the bivariate distribution. The Sample correlation coefficient is:

$$\hat{\rho}_n = \frac{\sum_1^n (X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_1^n (X_i-\bar{X})^2}\sqrt{\sum_1^n (Y_i-\bar{Y})^2}}$$

We know that the value of sample correlation coefficient does not change even if we substitute X_i and Y_i with $(X_i-\mu_1)/\sigma_1$ and $(Y_i-\mu_2)/\sigma_2$

Thus, when calculating the sample distribution of ρ, we can suppose that $\mu_1=0$, $\mu_2=0$, $\sigma_1^2=1$ and $\sigma_2^2=1$.

For convenience, let:

$$\bar{X}_n=\frac{1}{n}\sum_1^n X_i,\quad \bar{Y}_n=\frac{1}{n}\sum_1^n Y_i,\quad \overline{(XY)}_n=\frac{1}{n}\sum_1^n X_iY_i$$

$$\overline{(X^2)}_n=\frac{1}{n}\sum_1^n X_i^2,\quad \overline{(Y^2)}_n=\frac{1}{n}\sum_1^n Y_i^2$$

for function $g(t_1,t_2,t_3,t_4,t_5)=(t_3-t_1t_2)(t_4-t_1^2)^{-1/2}(t_5-t_2^2)^{-1/2}$

$$\hat{\rho}_n=\frac{\overline{(XY)}_n-\bar{X}_n\bar{Y}_n}{\sqrt{\overline{(X^2)}_n-(\bar{X}_n)^2}\sqrt{\overline{(Y^2)}_n-(\bar{Y}_n)^2}} = g(\bar{X}_n,\bar{Y}_n,\overline{(XY)}_n,\overline{(X^2)}_n,\overline{(Y^2)}_n)$$

as such, we can express $\hat{\rho}_n$ as a function of mean of $(X_i,Y_i,X_iY_i,X_i^2,Y_i^2)^t$ $(i=1,2,\cdots,n)$

Also, let $Z_i=(X_i,Y_i,X_iY_i,X_i^2,Y_i^2)^t$ $(i=1,2,\cdots,n)$. Then, Z_i are iid random variables in $\mathbb{R}^5$. Thus, the following holds given that $Var(Z_i)$ exists.

$$\sqrt{n}\left(\frac{1}{n}\sum_1^n Z_i - E(Z_1)\right) \xrightarrow[n\to\infty]{d} V,\quad V\sim N(0,Var(Z_1))$$

Meanwhile, from our supposition, $E[Z_1]=(0,0,\rho,1,1)^t$
$\rho = g(0,0,\rho,1,1) = g(E[Z_i])$

and the derivative of the function $g(t_1,t_2,t_3,t_4,t_5)$ $=(t_3-t_1t_2)(t_4-t_1^2)^{-1/2}(t_5-t_2^2)^{-1/2}$ is continuous. Hence, from ⟨Theorem 5.3.3⟩, if $Var(Z_i)$ exists[80], and if $\theta=E[Z_1]=(0,0,\rho,1,1)^t$,

$$\sqrt{n}(\hat{\rho}_n-\rho)=\sqrt{n}\left(g\left(\frac{1}{n}\sum_1^n Z_i\right)-g(\theta)\right) \xrightarrow[n\to\infty]{d} (\dot{g}(\theta))^tV,\quad V\sim N(0,Var(Z_i))$$

$\dot{g}(t_1,t_2,t_3,t_4,t_5)=$

$$\partial g/\partial t_1 = -t_2(t_4-t_1^2)^{-1/2}(t_5-t_2^2)^{-1/2} - t_1t_2\cdot -\tfrac{1}{2}\cdot -2t_1(t_4-t_1^2)^{-3/2}(t_5-t_2)^{-1/2}$$

$$= -t_2(t_4-t_1^2)^{-1/2}(t_5-t_2^2)^{-1/2} - t_1^2t_2(t_4-t_1^2)^{-3/2}(t_5-t_2)^{-1/2}$$

$$\partial g/\partial t_2 = -t_1(t_4-t_1^2)^{-1/2}(t_5-t_2^2)^{-1/2} - t_1t_2^2(t_4-t_1^2)^{-1/2}(t_5-t_2)^{-3/2}$$

$$\partial g/\partial t_3 = (t_4-t_1^2)^{-1/2}(t_5-t_2^2)^{-1/2}$$

$$\partial g/\partial t_4 = -\tfrac{1}{2}(t_3-t_1t_2)(t_4-t_1^2)^{-3/2}(t_5-t_2^2)^{-1/2}$$

$$\partial g/\partial t_5 = -\tfrac{1}{2}(t_3-t_1t_2)(t_4-t_1^2)^{-1/2}(t_5-t_2^2)^{-3/2}$$

$$\dot{g}(0,0,\rho,1,1)=[0,0,1,-\rho/2,-\rho/2]^t$$

$$(\dot{g}(\theta))^tV\sim N(0,(\dot{g}(\theta))^tVar(Z_1)(\dot{g}(\theta)))=N(0,Var[(\dot{g}(\theta))^tZ_1])$$

$$\dot{g}(\theta)^tZ = X_iY_i-\frac{\rho}{2}X_i^2-\frac{\rho}{2}Y_i^2$$

Therefore, if $E[X^4]<+\infty$ and $E[Y^4]<+\infty$,

$$\sqrt{n}(\hat{\rho}_n-\rho) \xrightarrow[n\to\infty]{d} W,\quad W\sim N(0,Var(X_iY_i-\tfrac{\rho}{2}X_i^2-\tfrac{\rho}{2}Y_i^2))$$ □

80) $Z_i=(X_i,Y_i,X_iY_i,X_i^2,Y_i^2)$ 이므로 ⟨정리 1.6.2⟩로부터 $E[X_i^4]<+\infty$, $E[Y_i^4]<+\infty$ 이면 Z_i의 분산 행렬이 존재하는 것을 알 수 있다.

Steps of proof
① Define correlation coefficient $\hat{\rho}_n$ as the mean of $Z_i=(X_i,Y_i,X_iY_i,X_i^2,Y_i^2)^t$
② Find a function $g(z)$ s.t. $g(\bar{z})=\hat{\rho}_n$
③ Find the Jacobian matrix $\dot{g}(z)$
④ Find the variance of the limiting distribution: $Var(\dot{g}(\theta)^tZ)$
* the value of correlation coefficient doesn't change even after standardization. Use this property.

Ex. 5.3.7. Sample correlation coefficient and variance stabilizing transformation.

In Ex. 5.3.6, let's consider the sample correlation coefficient, given that the bivariate normal distribution $N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2, \rho)$ ($\sigma_1>0$, $\sigma_2>0$, $-1<\rho<1$) is the population distribution.

In this case, from ⟨Ex. 4.4.3⟩,

$$Y_1 - \rho X_1 \big|_{X_1=x_1} \sim N(0, 1-\rho^2)$$

* ⟨Ex. 4.4.3⟩ : 이변량 정규분포 $N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2, \rho)$ 에서의 조건부분포

$$X_2\big|_{X_1=x_1} \sim N\left(\mu_2 + \rho\sigma_2(x_1-\mu_1)/\sigma_1,\ \sigma_2^2(1-\rho^2)\right)$$

$$\left(Y_1 - \rho X_1\right)\big|_{X_1=x_1} \sim N(a, b)$$

$$Y_1\big|_{X_1=x_1} - \rho x_1 \sim N\left(\mu_2 + \rho\sigma_2(x_1-\mu_1)/\sigma_1 - \rho x_1,\ \sigma_2^2(1-\rho^2)\right)$$

$$\sim N(\rho x_1 - \rho x_1,\ 1-\rho^2) \qquad (\because \mu_1=\mu_2=0,\ \sigma_1^2=\sigma_2^2=1)$$

$$\sim N(0, 1-\rho^2)$$

Thus the conditional distribution is independent of $X_1=x_1$,
$\Rightarrow$ $Y_1-\rho X_1$ and X_1 are independent. If $\mu_1=\mu_2=0$, $\sigma_1=\sigma_2=1$, and $T=(Y_1-\rho X_1)/\sqrt{1-\rho^2}$, X_1 and T are independent and both follow the standard normal distribution $N(0,1)$.

Substitute $Y_1 = \rho X_1 + \sqrt{1-\rho^2}\,T$

$$X_1Y_1 - \frac{\rho}{2}X_1^2 - \frac{\rho}{2}Y_1^2 = X_1(\rho X_1 + \sqrt{1-\rho^2}\,T) - \frac{\rho}{2}X_1^2 - \frac{\rho}{2}(\rho X_1 + \sqrt{1-\rho^2}\,T)^2$$

$$= \rho X_1^2 + \sqrt{1-\rho^2}\,TX_1 - \frac{\rho}{2}X_1^2 - \frac{\rho}{2}(\rho^2X_1^2 + 2\rho\sqrt{1-\rho^2}X_1T + (1-\rho^2)T^2)$$

$$= \rho X_1^2 + \sqrt{1-\rho^2}\,TX_1 - \frac{\rho}{2}X_1^2 - \frac{\rho^3}{2}X_1^2 - \rho^2\sqrt{1-\rho^2}X_1T - \frac{\rho}{2}(1-\rho^2)T^2$$

$$= X_1^2\left(\rho - \frac{\rho}{2} - \frac{\rho^3}{2}\right) + X_1T\left(\sqrt{1-\rho^2} - \rho^2\sqrt{1-\rho^2}\right) - T^2\cdot\frac{\rho}{2}(1-\rho^2)$$

$$= \frac{\rho}{2}(1-\rho^2)X_1^2 + (1-\rho^2)^{3/2}X_1T - \frac{\rho}{2}(1-\rho^2)T^2$$

$$\mathrm{Var}\left(X_1Y_1 - \frac{\rho}{2}X_1 - \frac{\rho}{2}Y_1\right) = \mathrm{Var}\left(\frac{\rho}{2}(1-\rho^2)X_1^2 + (1-\rho^2)^{3/2}X_1T - \frac{\rho}{2}(1-\rho^2)T^2\right)$$

$$= \frac{\rho^2}{4}(1-\rho^2)^2\mathrm{Var}(X_1^2) + (1-\rho^2)^3\mathrm{Var}(X_1T) + \frac{\rho^2}{4}(1-\rho^2)^2\mathrm{Var}(T^2)$$

$\rho_4 = E[X^4] - 3 = 0$ $(X\sim N(0,1))$
$\therefore E[X^4]=3$

$$\mathrm{Var}(X_1^2) = E[X_1^4] - E[X_1^2]^2 = 3-1 = 2$$
$$\therefore \mathrm{Var}(X_1^2) = \mathrm{Var}(T^2) = 2$$

$$\mathrm{Var}(X_1T) = E[X_1^2T^2] - E[X_1T]^2 = E[X_1^2]E[T^2] - (E[X_1]E[T])^2$$
$$= 1\cdot 1 - (0\cdot 0)^2 = 1$$

$$\therefore \mathrm{Var}\left(X_1Y_1 - \frac{\rho}{2}X_1 - \frac{\rho}{2}Y_1\right) = \frac{\rho^2}{4}(1-\rho^2)^2\cdot 2 + (1-\rho^2)^3\cdot 1 + \frac{\rho^2}{4}(1-\rho^2)^2\cdot 2$$
$$= (1-\rho^2+\rho^2)(1-\rho^2)^2 = (1-\rho^2)^2$$

$$\therefore \sqrt{n}(\hat\rho - \rho) \xrightarrow[n\to\infty]{d} W, \quad W\sim N(0, (1-\rho^2)^2)$$

Here, if we find the limiting distribution of $g(\hat\rho_n)$ using ⟨Theorem 5.3.3⟩,

$$\sqrt{n}(g(\hat\rho) - g(\rho)) \xrightarrow[n\to\infty]{d} \dot g(\rho)W, \quad \dot g(\rho)W \sim N(0, [\dot g(\rho)]^2(1-\rho^2)^2)$$

In particular, the transformation $g(\hat\rho)$ which causes the variance of $\hat\rho_n$ to be independent of ρ is called the "variance stabilizing transformation". Especially,

$$g(\hat\rho) = \frac{1}{2}\log\frac{1+\hat\rho}{1-\hat\rho}$$

is called the "Fischer's Transformation" and causes $[\dot g(\hat\rho)]^2(1-\rho^2)^2=1$ allowing the limiting distribution to be $N(0,1)$ i.e.,

$$\sqrt{n}[g(\hat\rho) - g(\rho)] \xrightarrow[n\to\infty]{d} Z, \quad Z\sim N(0,1)$$

* $\dot g(\hat\rho) = \frac{1}{2}\left\{\frac{1}{1-\hat\rho} + \frac{1+\hat\rho}{(1-\hat\rho)^2}\right\} \Big/ \left\{\frac{1+\hat\rho}{1-\rho}\right\}$

$$= \frac{1}{2}\left\{\frac{1-\hat\rho+1+\hat\rho}{(1-\hat\rho)^2}\right\} \Big/ \left\{\frac{1+\hat\rho}{1-\hat\rho}\right\}$$

$$= \frac{1}{2}\left\{\frac{2}{(1-\hat\rho)^2}\right\} \Big/ \left\{\frac{1+\hat\rho}{1-\hat\rho}\right\}$$

$$= \frac{1}{(1-\hat\rho)(1-\hat\rho)} = \frac{1}{1-\hat\rho^2}$$

$$\therefore [\dot g(\hat\rho)]^2 = (1-\hat\rho^2)^{-2}$$

Steps.

① Find a random variable that is independent of X_1
② Use CLT to find a r.v. T that is independent of X_1 and $\sim N(0,1)$
③ Substitute Y with T
④ Find the variance of $\hat\rho$ using X and T (use iid property)
⑤ Use CLT to find the distribution of $\sqrt{n}(\hat\rho-\rho)\sim N(0, \mathrm{Var}(\hat\rho))$
⑥ What is the variance stabilizing transformation?

Ex. 5.3.8. Limiting Distribution of Quantile Number.

In ⟨Ex. 5.2.7⟩, we learned about the convergence in probability of the quantile $X_{(r_n)}$ $(r_n \underset{n\to\infty}{\sim} an,\ 0<a<1)$. Here, we will think about the limiting distribution of sample quantile given that the conditions in ⟨Ex. 5.2.7⟩ are satisfied.

Let F^{-1} be the inverse of the population distribution's cdf and let $h(y) = F^{-1}(1-e^{-y})$ $(y>0)$. From ⟨Theorem 4.3.4⟩, the distribution of sample quantile is:

$$X_{(r_n)} \overset{d}{=} \left(\frac{1}{n} Z_1 + \cdots + \frac{1}{n-r_n+1} Z_{r_n} \right), \quad Z_i \overset{iid}{\sim} \mathrm{Exp}(1)\ (i=1,\cdots,n)$$

Meanwhile, if

$$Y_n = \frac{1}{n} Z_1 + \cdots + \frac{1}{n-r_n+1} Z_{r_n}$$

we know that the mean and variance of Y_n approximates to:

$$E[Y_n] = \frac{1}{n} + \cdots + \frac{1}{n-r_n+1} \underset{n\to\infty}{\sim} -\log(1-a)$$

$$\mathrm{Var}(Y_n) = \frac{1}{n^2} + \cdots + \frac{1}{(n-r_n+1)^2} \underset{n\to\infty}{\sim} \frac{1}{n} \frac{a}{1-a}$$

Let

$$W_n = \sqrt{n}\, \frac{Y_n - (-\log(1-a))}{\sqrt{a/(1-a)}}$$

* This is NOT CLT.

If we approximate the mgf of W_n, [81] we find that the following holds:

$$\lim_{n\to\infty} \mathrm{mgf}_{W_n}(t) = \exp\left\{\tfrac{1}{2}t^2\right\} = \mathrm{mgf}_W(t),\ W \sim N(0,1)$$

$$\therefore W_n \xrightarrow[n\to\infty]{d} W,\ W \sim N(0,1)$$

$$\therefore \sqrt{n}(Y_n - (-\log(1-a))) \xrightarrow[n\to\infty]{d} \sqrt{a/(1-a)}\, W$$

Thus, from ⟨Theorem 5.3.3⟩ if the function f is differentiable, [82]

$$\sqrt{n}\left\{ h(Y_n) - h(-\log(1-a)) \right\} \xrightarrow[n\to\infty]{d} h'(-\log(1-a)) \sqrt{a/(1-a)}\, W$$

However, under these conditions, $h'(-\log(1-a)) = (1-a)/f(F^{-1}(a))$, * $f = F'$

$$\therefore \sqrt{n}\left\{ X_{(r_n)} - F^{-1}(a) \right\} \xrightarrow[n\to\infty]{d} Z,\ Z \sim N\left(0,\ a(1-a)/[f(F^{-1}(a))]^2\right)$$

[81] 연습문제 5.15 에 이러한 정규생성함수를 근사하는 과정이 소개되어 있다.

[82] 확률밀도함수가 $f(x) = \partial F(x)/\partial x$ 로 주어지고 $f(F^{-1}(a)) > 0$ 이면 이 조건이 만족된다.

* $h(y) = F^{-1}(1-e^{-y})$ let $z = F^{-1}(1-e^{-y}) \Rightarrow F(z) = (1-e^{-y})$

$\partial z \cdot f(z) = e^{-y}\, dy$

$$\therefore \frac{\partial z}{\partial y} = e^{-y} f(z)^{-1} = e^{-y} / f(F^{-1}(1-e^{-y}))$$

$$\therefore h'(-\log(1-a)) = (1-a)/f(F^{-1}(a))$$ □

5.4. APPROXIMATION USING SIMULATION

In this section, we will explore methods of approximating the distribution of statistics or researching its characteristics using the observation of random samples. The observation of random samples that follow a specific distribution is called "random number" and recently many statistics packages allow us to generate these so called "random numbers".

The "uniform random number", which refers to the observations from the $U(0,1)$ distribution, is the foundation of generation of random numbers.

Ex. 5.4.1. Random Number Generation in Logistic Distribution.

The pdf and cdf of the Logistic distribution introduced in ⟨Ex. 4.1.5⟩ are

$$f(z) = \frac{e^z}{(1+e^z)^2},\quad F(z) = 1 - \frac{1}{(1+e^z)},\quad -\infty < z < \infty$$

From the above, we know that the inverse of the cdf is:

$$F^{-1}(u) = \log \frac{u}{1-u}$$

Hence, from ⟨Theorem 4.3.3⟩, for $U \sim U(0,1)$,

$$Z = \log \frac{U}{1-U} \sim L(0,1)$$

using the above, we can find the random number from the Logistic Distribution $L(0,1)$. Also, by using:

$$\sigma Z + \mu = \sigma \log \frac{U}{1-U} + \mu \sim L(\mu,\sigma)$$

we can generate random numbers from the Logistic distribution $L(\mu,\sigma)$.

Similarly to ⟨Ex. 5.4.1⟩, for the random variable $U \sim U(0,1)$,

$$-\log(1-U) \sim \mathrm{Exp}(1),\quad \sigma\{-\log(1-U)\} \sim \mathrm{Exp}(\sigma)$$

using the precedent property, we can generate random numbers from $\mathrm{Exp}(\sigma)$, and for the inverse of the cdf of standard normal distribution, $\Phi^{-1}(u)$,

$$\Phi^{-1}(U) \sim N(0,1),\quad \sigma\Phi^{-1}(U) + \mu \sim N(\mu,\sigma^2)$$

we can use this to generate random samples from $N(\mu,\sigma^2)$. Other methods of random number generation are introduced in Appendix Ⅲ.

〈Theorem 5.4.1. Definite integral approximation using random numbers〉

For $U_1,\cdots,U_n \overset{iid}{\sim} U(0,1)$ and function $g(x)$ that is continuous on $[a,b]$, the following holds:

$$X_i = (b-a)U_i + a \overset{iid}{\sim} U(a,b) \quad (i=1,\cdots,n)$$

$$\operatorname{plim}_{n\to\infty} \frac{b-a}{n}\sum_{i=1}^{n} g(X_i) = \int_a^b g(x)\,dx$$

(proof)

$X_1,\cdots,X_n \overset{iid}{\sim} U(a,b)$ and its pdf is $pdf_{X_i}(x) = \frac{1}{b-a} I_{[a,b]}(x)$

$$E[g(X_i)] = \frac{1}{b-a}\int_a^b g(x_i)\,dx$$

Hence, from the Law of Large Numbers,

$$\operatorname{plim}_{n\to\infty} \frac{b-a}{n}\sum_{i=1}^{n} g(X_i) = (b-a)E[g(X_i)] = \int_a^b g(x)\,dx$$

* $\lim_{n\to\infty} \frac{1}{n}\sum g(x_i) = E[g(X)]$

Using the convergence in probability to approximate the definite integral as 〈Theorem 5.4.1〉 is called "Monte Carlo Integration". In other words, using independently generated random numbers $u_1,\cdots,u_n$ to find the approximate of definite integral:

$$y_i = (b-a)\,g((b-a)u_i + a) \quad (i=1,\cdots,n)$$

$$\int_a^b g(x)\,dx \simeq \frac{1}{n}\sum_{i=1}^{n} y_i \,, \quad n\to\infty$$

is called "Monte Carlo Integration". This utilizes the observations of Y_i:

$$Y_i = (b-a)\,g((b-a)U_i + a) \quad (i=1,\cdots,n)$$

to estimate $E[Y_i]$ and its asymptotic Confidence Interval can be used to describe the precision(정밀도) of estimation.

Ex. 5.4.2.

If we use the integral formula, we know that $\int_1^3 x^2\,dx = 26/3$. Meanwhile for independently generated random samples $u_1,\cdots,u_n \overset{iid}{\sim} U(0,1)$, let

$$x_i = 2u_i + 1, \quad y_i = 2x_i^2 \quad (i=1,\cdots,n)$$

$$\int_1^3 x^2\,dx \simeq \frac{2}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The approximation and 95% asymptotic confidence interval generated from random numbers generated using R package is as follows, and as one can see, the precision of approximation improves as n increases.

n	100	1000	10,000	100,000
$\bar{y}$	8.7612	8.5633	8.6515	8.6473
$\bar{y} - 1.96 S_y/\sqrt{n}$	7.8241	8.2728	8.5604	8.6185
$\bar{y} + 1.96 S_y/\sqrt{n}$	9.6983	8.8538	8.7426	8.6762

Ex. 5.4.3. Variance of Logistic Distribution.

It is known that the variance of Logistic Distribution L(0,1), which is symmetric about $x=0$, is : 83)

$$\int_{-\infty}^{+\infty} x^2 \frac{e^x}{(1+e^x)^2}\,dx = \frac{\pi^2}{3}$$

Meanwhile, from 〈Ex. 5.4.1〉, for $U \sim U(0,1)$

$$X = \log\frac{U}{1-U} \sim L(0,1)$$

Hence we can approximate the variance of Logistic Distribution L(0,1) using the random numbers $u_1,\cdots,u_n$ independently generated from $U(0,1)$

$$\therefore \int_{-\infty}^{+\infty} x^2 \frac{e^x}{(1+e^x)^2}\,dx = E[x^2] \simeq \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The approximation and 95% asymptotic confidence interval generated from random numbers generated using R package is as follows

n	100	1,000	10,000	100,000
$\bar{y}$	2.9775	3.3375	3.3201	3.2949
$\bar{y} - 1.96 S_y/\sqrt{n}$	2.0043	2.9799	3.2034	3.2584
$\bar{y} + 1.96 S_y/\sqrt{n}$	3.9506	3.6952	3.4369	3.3313

84) 부분적분을 이용하고 무한급수로 나타내어 적분하면

$$\int_{-\infty}^{+\infty} x^2 \frac{e^x}{(1+e^x)^2}\,dx = 4\int_{-\infty}^{+\infty} \frac{e^x}{(1+e^x)^2} x\,dx = 4\int_{-\infty}^{+\infty} \sum_{n=1}^{\infty} (-e^{-x})^{n-1} e^{-x}\,dx$$

$$= 4\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^2}$$

한편 푸리에 급수를 이용한 다음의 항등식에서 $x=0$을 대입하면 위의 무한급수의 값을 얻을 수 있다.

$$x^2 = \frac{1}{3}\pi^2 + 4\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}\cos(nx) \quad (-\pi \le x \le \pi)$$

Ex. 5.4.4. Limiting Distribution of Sample proportion (표본비율).

Let $X_1,\cdots,X_n$ denote the observation of n Bernoulli trials with population proportion(모비율) p. We know that the following is true for $\hat{p}_n = (X_1+\cdots+X_n)/n$ from LLN and CLT:

$$\operatorname{plim}_{n\to\infty} \hat{p}_n = p, \quad \frac{\hat{p}_n - p}{\sqrt{p(1-p)/n}} \xrightarrow[n\to\infty]{d} Z, \; Z\sim N(0,1)$$

Hence,

$$\lim_{n\to\infty} P\left(-z_{\alpha/2} \le \frac{\hat{p}_n - p_n}{\sqrt{\hat{p}_n(1-\hat{p}_n)/n}} \le z_{\alpha/2}\right) = 1-\alpha$$

$$\therefore \lim_{n\to\infty} p\left\{\hat{p}_n - z_{\alpha/2}\sqrt{\hat{p}_n(1-\hat{p}_n)/n} \le p \le \hat{p}_n + z_{\alpha/2}\sqrt{\hat{p}_n(1-\hat{p}_n)/n}\right\} = 1-\alpha$$

From above, we can denote the $100(1-\alpha)\%$ asymptotic confidence interval p as:

$$p \in \left[\hat{p}_n - z_{\alpha/2}\sqrt{\hat{p}_n(1-\hat{p}_n)/n},\; \hat{p}_n + z_{\alpha/2}\sqrt{\hat{p}_n(1-\hat{p}_n)/n}\right]$$

Ex. 5.4.5 Approximation of π using the asymptotic distribution of sample proportion.

For $(U_1, U_2) \overset{iid}{\sim} U(0,1)$, observing (U_1, U_2) can be thought of as dropping random points on a square with length/width of 1. The probability that the dot will fall inside a quarter of a circle inside the square is:

$$P(U_1^2 + U_2^2 \le 1) = \pi/4$$

The Law of Large Numbers imply that the relative proportion of dots falling inside the quarter of a circle approximates to $\pi/4$ in a repeated and independent test. Thus for independently generated random numbers $u_{11}, u_{12}, \cdots, u_{n1}, u_{n2} \sim U(0,1)$, let

$$x_i = u_{i1}^2 + u_{i2}^2 \quad (i = 1, 2, \cdots, n)$$

Then,

$$\hat{p}_n = \frac{1}{n} \sum_{1}^{n} I_{(x_i \le 1)}$$

is the approximation of $p = P(U_1^2 + U_2^2 \le 1) = \pi/4$ and we can use $4\hat{p}$ as the approximation of $4p = \pi$. Also from (Ex. 5.4.4) we can express the precision of approximation using the asymptotic confidence interval of $4p$.

The approximation and 95% asymptotic confidence interval generated from random numbers generated using R package is as follows.

n	100	1,000	10,000	100,000
$4\hat{p}$	3.3600	3.1320	3.1344	3.1452
$4\hat{p}_n - 4 \times 1.96\sqrt{\hat{p}_n(1-\hat{p}_n)/n}$	3.0755	3.0308	3.1024	3.1351
$4\hat{p}_n + 4 \times 1.96\sqrt{\hat{p}_n(1-\hat{p}_n)/n}$	3.6445	3.2331	3.1664	3.1552