Law of Large Numbers / Central Limit Theorem

Tadao Hoshino （星野匡郎）

ver. 2018 Fall Semester

The Law of Large Numbers: Introduction

The Law of Large Numbers: Introduction

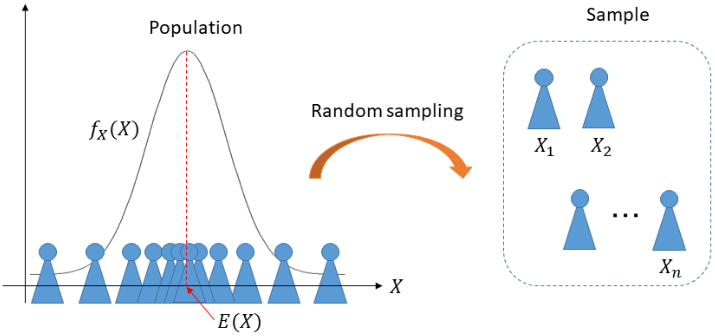
LLN: Informal statement

Suppose that we have data $\{X_1, ..., X_n\}$ of sample size n randomly drawn from the same population. Let

Population mean of X: $\mu = E(X)$, Sample average of X: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Then, as n increases to infinity, $\bar{X}_n$ approaches to μ.

The Law of Large Numbers: Introduction



	Population mean		Sample average
If n is sufficiently large. . .	$E(X)$	$\approx$	$\frac{1}{n}\sum_{i=1}^{n} X_i$

A simulation of dice rolling

- Click on the Windows Start button at the bottom left of your computer screen, and start **R** by selecting "R" (or R XXX, where XXX gives the version of **R**, e.g., R x64 3.4.3) – the **R** console pops up.
- To simulate dice rolls in **R**, we can use a function called `sample`. The usage of this function is:

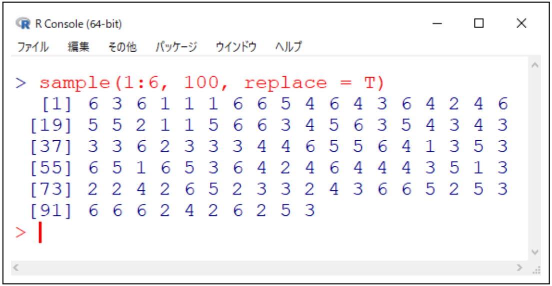
```
sample(1:6, n, replace = T)
```

  - This function draws n random numbers from the vector specified in the first argument.
  - The notation "`1:6`" means the vector of integers $(1,2,...,6)$.
  - "`replace = T`" is required when the size of sample is larger than the length of the first argument.

A simulation of dice rolling

- For example, if you want to simulate 100 dice rolls, type the following in the **R** console:

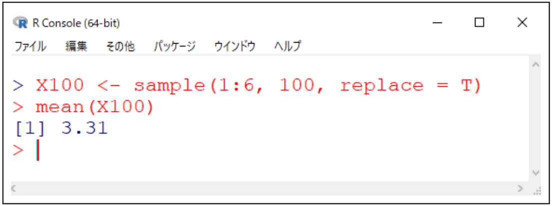
```
sample(1:6, 100, replace = T)
```



NOTE: Your results may be different from mine. To fix the simulation results, you need to set the "random seed" before you start generating random numbers.

A simulation of dice rolling

- Save this simulation result as `X100` (you can use any name here), and compute the sample average by `mean(X100)`:



- Similarly, create `X10000` (10,000 dice rolls) and `X10` (10 dice rolls), and compute their averages:

A simulation of dice rolling

R Console (64-bit)

ファイル 編集 その他 パッケージ ウインドウ ヘルプ

```
> X100 <- sample(1:6, 100, replace = T)
> mean(X100)
[1] 3.31
> X10000 <- sample(1:6, 10000, replace = T)
> mean(X10000)
[1] 3.513
> X10 <- sample(1:6, 10, replace = T)
> mean(X10)
[1] 3.2
>
```

- Recall that $E(X) = 3.5$.
- From above results, we can see that the sample average with $n = 10000$ ($n = 10$) is the most (least) accurate approximation of $E(X)$, and the result of $n = 100$ is in between.

The Law of Large Numbers: Introduction

- The result of LLN is intuitively understandable, but in what sense does the sample average approaches to the population mean formally?
  $\Rightarrow$ Convergence in probability
- Under what conditions does LLN hold true?
  $\Rightarrow$ A sufficient (but not necessary) condition is that X_i's are drawn independently from the same population and the variance of X_i is finite.

Convergence in Probability

Convergence of real numbers

- Let $(a_n)_{n=1}^{\infty}$ denote a sequence of real numbers:

$$(a_n)_{n=1}^{\infty} = \{a_1,\ a_2,\ a_3, ...\}$$

- Example (i):

$$a_1 = 1,\ a_2 = 1.4,\ a_3 = 1.41\ \cdots a_n = 1.4142... \cdots$$

  a_n converges to $\alpha = \sqrt{2}$ as n increases to infinity.

- Example (ii):

$$a_1 = 3,\ a_2 = 3.1,\ a_3 = 3.14\ \cdots a_n = 3.1415... \cdots$$

  a_n converges to $\alpha = \pi$ as n increases to infinity.

- The number α is called the limit of $(a_n)_{n=1}^{\infty}$, and we write

$$\lim_{n\to\infty} a_n = \alpha \text{ or } a_n \to \alpha\ (n \to \infty)$$

  Equivalently, they can be denoted as $\lim_{n\to\infty} |a_n - \alpha| = 0$ or $|a_n - \alpha| \to 0\ (n \to \infty)$.

Some more examples of sequences

- $a_1 = 1,\ a_2 = 1/2,\ a_3 = 1/3, \cdots, a_n = 1/n, \cdots$

$$\lim_{n\to\infty} a_n = 0$$

- $a_1 = (1+(1/1))^1,\ a_2 = (1+(1/2))^2, \cdots, a_n = (1+(1/n))^n, \cdots$

$$\lim_{n\to\infty} a_n = e \text{ (Napier's constant: the base of natural log)}$$

- $a_1 = 1,\ a_2 = 2,\ a_3 = 3, \cdots, a_n = n, \cdots$

$$\lim_{n\to\infty} a_n = \infty$$

  When the limit is infinity, we say the sequence $(a_n)_{n=1}^{\infty}$ is divergent.

- $a_1 = -1,\ a_2 = 1,\ a_3 = -1, \cdots a_n = (-1)^n, \cdots$
  The sequence $(a_n)_{n=1}^{\infty}$ alternates between 1 and -1 and is neither convergent nor divergent.

Exercise 1

1. Show that the infinite decimal $0.999...$ is equal to 1.

Convergence in probability

- Now, suppose $(a_n)_{n=1}^{\infty}$ is a sequence of random variables.
- Note that "the sequence $(a_n)_{n=1}^{\infty}$ converges to α" is a stochastic event.

Convergence in probability

A sequence of random variables $(a_n)_{n=1}^{\infty}$ is said to converge in probability to α if

$$\lim_{n\to\infty} \Pr\left(|a_n - \alpha| < \epsilon\right) = 1$$

or equivalently

$$\lim_{n\to\infty} \Pr\left(|a_n - \alpha| \geq \epsilon\right) = 0$$

for <u>any</u> positive $\epsilon > 0$.

When $(a_n)_{n=1}^{\infty}$ converges to α in probability, we write

$$\operatorname*{plim}_{n\to\infty} a_n = \alpha \text{ or } a_n \xrightarrow{P} \alpha \ (n \to \infty)$$

Here, plim stands for "probability limit".

Convergence in probability

- Convergence in probability:

$$\lim_{n\to\infty} \Pr\left(|a_n - \alpha| < \epsilon\right) = 1 \text{ for any } \epsilon > 0.$$

- An important point to note in this definition is that ϵ can be arbitrarily small, say $\epsilon = 0.001$, $\epsilon = 0.00...001$, or even smaller.
- The definition says that, however small ϵ is, the probability of the event $\{|a_n - \alpha| < \epsilon\}$ (note that this is a real number) converges to 1 as n increases.

LLN: Formal statement

Suppose we have data $\{X_1, ..., X_n\}$ of sample size n independently drawn from the same population. Let

$$\text{Population mean of } X\text{: } \mu = E(X), \quad \text{Sample average of } X\text{: } \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then, we have

$$\bar{X}_n \xrightarrow{P} \mu \ (n \to \infty)$$

This result is called the weak law of large numbers.

NOTE : In the definition of convergence in probability, each a_n corresponds to the sample average of n observations:

$$a_1 = \frac{1}{1}\sum_{i=1}^{1} X_i,\ a_2 = \frac{1}{2}\sum_{i=1}^{2} X_i,\ ..., a_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sketch of the Proof of Weak LLN

Proof of Weak LLN

Expectation of $\bar{X}_n$

- Using linearity of expectation, we have

$$E(\bar{X}_n) = \frac{1}{n} \underbrace{\sum_{i=1}^{n} E(X_i)}_{=n\mu} = \mu$$

Variance of $\bar{X}_n$

- Recall that when X and Y are independent,

$$V(X+Y) = V(X) + V(Y)$$

  (see the previous lecture note).
- Similarly, if X_i's are independent, we have $V(\sum_{i=1}^{n} X_i) = nV(X)$. Thus,

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

Proof of Weak LLN

Markov's inequality

Let X be a non-negative random variable, and $\epsilon > 0$ be a positive constant. Then,

$$\Pr(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$$

This result is called Markov's inequality.

Proof. Note that $\Pr(X \geq \epsilon) = E(\mathbf{1}\{X \geq \epsilon\})$, where

$$\mathbf{1}\{X \geq \epsilon\} = \begin{cases} 1 & \text{if } X \geq \epsilon \\ 0 & \text{if } X < \epsilon \end{cases}$$

We have

$$\mathbf{1}\{X \geq \epsilon\} \leq \frac{X}{\epsilon}$$

for any $\epsilon > 0$. Finally, taking the expectation of both sides yields the desired result. ■

Proof of Weak LLN

Chebyshev's inequality

Let X be a random variable, and $\epsilon > 0$ be a positive constant. Then,

$$\Pr\left(|X-\mu| \geq \epsilon\right) \leq \frac{V(X)}{\epsilon^2}$$

This result is called Chebyshev's inequality.

Proof. Noting the equivalence of the two events

$$|X-\mu| \geq \epsilon \iff |X-\mu|^2 \geq \epsilon^2,$$

it holds that

$$\Pr\left(|X-\mu| \geq \epsilon\right) = \Pr\left(|X-\mu|^2 \geq \epsilon^2\right).$$

Then, by Markov's inequality, we have

$$\Pr\left(|X-\mu|^2 \geq \epsilon^2\right) \leq \frac{E(|X-\mu|^2)}{\epsilon^2} = \frac{V(X)}{\epsilon^2}$$

■

Proof of Weak LLN

Now we are ready to prove Weak LLN by showing

$$\lim_{n\to\infty} \Pr\left(|\bar{X}_n - \mu| \geq \epsilon\right) = 0 \text{ for any } \epsilon > 0.$$

Proof of Weak LLN

Step 1. By Chebyshev's inequality, for any $\epsilon > 0$,

$$\Pr\left(|\bar{X}_n - \mu| \geq \epsilon\right) \leq \frac{V(\bar{X}_n)}{\epsilon^2}$$

Step 2. By the independence of $\{X_1, ..., X_n\}$,

$$V(\bar{X}_n) = \frac{V(X)}{n}.$$

Step 3. Combining these results, we have

$$\Pr\left(|\bar{X}_n - \mu| \geq \epsilon\right) \leq \frac{V(X)}{n\epsilon^2}$$

Taking the limit of both sides as $n \to \infty$,

$$\lim_{n\to\infty} \Pr\left(|\bar{X}_n - \mu| \geq \epsilon\right) \leq 0.$$

Since any probability cannot be less than zero, the above inequality holds with equality. ■

The Central Limit Theorem

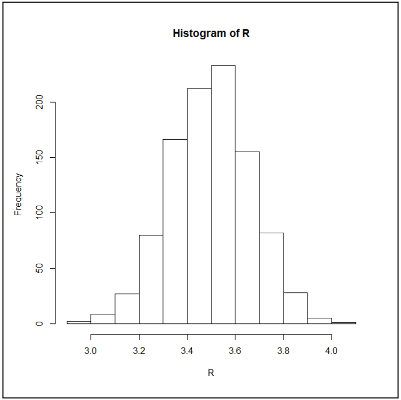
A simulation of dice rolling

- Again, we conduct a simulation of dice rolling with **R**. This time, we repeat the computation of the average of 100 dice rolls 1,000 times, and plot the result in a histogram.
- Type the following in the **R** console:

```
> R <- numeric(1000)
> for(i in 1:1000) R[i] <- mean(sample(1:6, 100, replace = T))
> hist(R)
```

  - **First line**: we create a vector of empty values of length 1,000 by `numeric(1000)`, and name it as `R`.
  - **Second line**: we calculate the average of 100 dice rolls, and store the result in the i-th element of `R`. Repeat this from $i = 1$ to $i = 1000$.
  - **Third line**: plot the histogram of `R` by `hist(R)`.

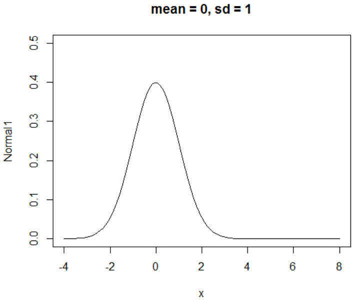
A simulation of dice rolling

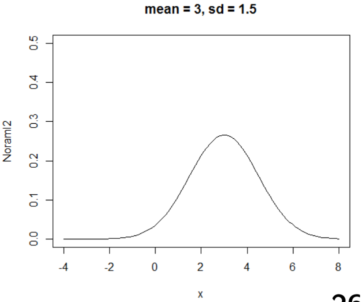


- This histogram looks almost symmetric and unimodal, and peaks at $3.5 = E(X)$.

Normal distribution

- The most important probability distribution in the entire field of statistics is the normal distribution.
- The normal distribution is symmetric and unimodal.
- The shape of the normal distribution is fully characterized by two parameters: the mean μ and the standard deviation σ.
- The mean μ determines the center of the distribution, and the standard deviation σ determines the shape of the curve.





The Central Limit Theorem

- The normal distribution with mean μ and standard deviation σ is denoted as $N(\mu, \sigma^2)$, and its PDF is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

  (You do not need to memorize the definition.)
  Thus, if X follows the normal distribution $N(\mu, \sigma^2)$, the probability that X is less than or equal to a is

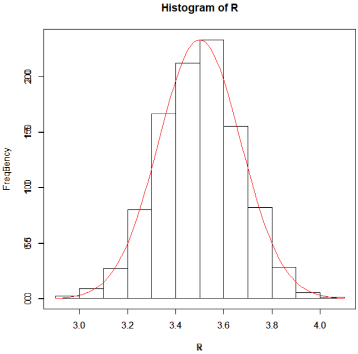
$$\Pr(X \leq a) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{a} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

- In particular, the normal distirubution $N(0, 1)$ is called the standard normal distribution. The PDF of the standard normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

The Central Limit Theorem

It seems possible that we can approximate the shape of the histogram of the averages of dice rolls by normal distribution with appropriately chosen mean and standard deviation...



NOTE : The red curve is the PDF of normal distribution $N(3.5, 0.171^2)$.

The Central Limit Theorem

This approximation is not a coincidence, but the result of the central limit theorem.

CLT: Informal statement

Suppose we have data $\{X_1, ..., X_n\}$ of sample size n randomly drawn from the same population. Let

$$\text{Pop mean of } X\text{: } \mu = E(X), \quad \text{Pop variance of } X\text{: } \sigma^2 = V(X),$$

$$\text{Sample average of } X\text{: } \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then, as n increases to infinity, the probability distribution of $\bar{X}_n$ can be approximated by the normal distribution $N(\mu, \sigma^2/n)$.

Convergence in distribution

A sequence of random variables $(a_n)_{n=1}^{\infty}$ is said to converge in distribution to $N(\mu, \sigma^2)$ if

$$\lim_{n\to\infty} \underbrace{\Pr\left(a_n \leq x\right)}_{\text{CDF of } a_n} = \Phi(x; \mu, \sigma)$$

for any x, where $\Phi(\cdot; \mu, \sigma)$ is the CDF of $N(\mu, \sigma^2)$.

When $(a_n)_{n=1}^{\infty}$ converges to $N(\mu, \sigma^2)$ in distribution, we write

$$a_n \xrightarrow{d} N(\mu, \sigma^2)$$

- The formal statement of the CLT is: under the aforementioned assumptions,

$$\underbrace{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}}_{\text{standardized sample average}} \xrightarrow{d} N(0,1)$$

- The proof of the CLT is much more complicated than that of the LLN, and thus is omitted.

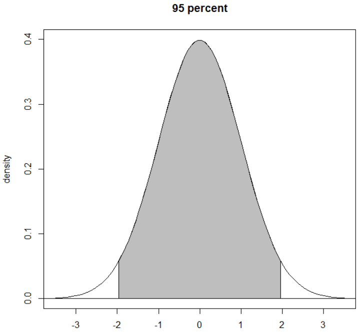
Confidence Interval

Confidence Interval

Let Z be distributed as the standard normal $N(0,1)$. Then, it holds that

$$\Pr(-1.9599 \le Z \le 1.9599) = 0.95$$

That is, the area of grayed part in the following figure is equal to 0.95.

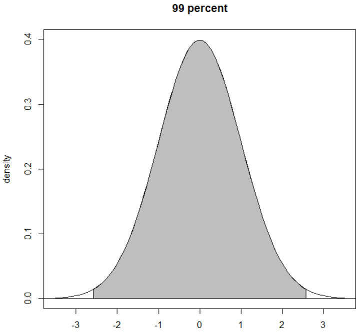


* When drawing a random number from $N(0,1)$, the value will be included in $[-1.9599, 1.9599]$ about 95 times out of 100.

Similarly, it holds that

$$\Pr(-2.5758 \leq Z \leq 2.5758) = 0.99$$

That is, the area of grayed part in the following figure is equal to 0.99.



* When drawing a random number from $N(0, 1)$, the value will be included in $[-2.5758, 2.5758]$ about 99 times out of 100.

Confidence Interval

- Let $\{X_1, ..., X_n\}$ be a random sample of n observations from a population. According to the CLT, the distribution function of the standardized sample average

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

can be approximated by the distribution function of $N(0, 1)$ as n increases.

- Thus, for sufficiently large n, we have

$$\Pr(-1.9599 \leq Z_n \leq 1.9599) \approx 0.95$$

* Taking the limit of the left hand side as $n \to \infty$, the approximation holds with equality.

- $0.95 \approx \Pr(-1.9599 \leq Z_n \leq 1.9599)$

$$
\begin{aligned}
&= \Pr\left(-1.9599 \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq 1.9599\right) \\
&= \Pr\left(-1.9599\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq 1.9599\frac{\sigma}{\sqrt{n}}\right) \\
&= \Pr\left(\bar{X}_n - 1.9599\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.9599\frac{\sigma}{\sqrt{n}}\right)
\end{aligned}
$$

- This implies that the population mean μ is included in the interval

$$
\left[\bar{X}_n - 1.9599\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.9599\frac{\sigma}{\sqrt{n}}\right]
$$

with 95% probability. This interval is called μ's 95% confidence interval.

Confidene Interval

- What is important about this fact is

even though the true value of μ is unknown, the interval μ falls into can be computed.

(This is the power of the central limit theorem.)

- Similarly, the 99% confidene interval of μ can be computed by

$$\left[\bar{X}_n - 2.5758\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 2.5758\frac{\sigma}{\sqrt{n}}\right]$$

- Note that the length of the confidence interval is inversely proportional to $\sqrt{n}$; **the larger the sample size, the more precise the confidence interval**.
- If, for example, the lower bound of the 95% confidence interval is positive, we can infer that μ is positive with at least 95% probability $=>$ this is the basic framework of statistical hypothesis testing.

Glossary I