ST102 - 2023 - Group 14

Yutong Wang

Nov 2023

Contents

1 Ex2 Q2 - De Morgan's Law

1.1 Identify a Set

What do we actually mean by two sets are equal to each other? Intuitively, they should contain exactly the same elements. Formally:

Definition 1. *Sets A and B are equal (denote as $A = B$) if and only if:* $(A \subset B) \wedge (B \subset A)$.

The symbol $A \subset B$ is read as: A is a subset of B, and define by:

Definition 2. *$A \subset B$ if and only if:*

$$\forall a \in A, a \in B.$$

1.2 Proof general De Morgan's Law

An amazing fact is that, De Morgan's Law holds not only for 2 sets, but for any index set including finite union (/intersection) of sets, countable union (/intersection) of sets, and uncountable union (/intersection) of sets.

Theorem 1. *For any index set I_α, we have:*

$$\begin{aligned}
(\bigcup_{\alpha \in I_\alpha} A_\alpha)^c &= \bigcap_{\alpha \in I_\alpha} (A_\alpha^c) \\
(\bigcap_{\alpha \in I_\alpha} A_\alpha)^c &= \bigcup_{\alpha \in I_\alpha} (A_\alpha^c)
\end{aligned}$$

The conclusion is that two sets are equal, which is just defined in the previous section. We use the definition and the 'prove by contradiction' to prove the De Morgan's Law.

2 Ex3 Q1 - Combinations

2.1 Why it is always an integer?

For a combination, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, since it is a quotient, you might want to ask, why this is an integer for any n and k satisfying $n \geq k$?

We can see that the number of each prime number as its factor in the denominator is always less than that in the numerator.

2.2 Useful formulas

Combinations have clear meaning in real life, intuitions could lead to lots of useful equations.

Proposition 1. *For any $n \in \mathbb{Z}_+$,*

$$2^n = \sum_{k=0}^{n} \binom{n}{k}.$$

Proof. Consider a set $A = \{a_1, \dots, a_n\}$ with n number of elements in it (note that the definition of a set implies that any two elements in it is not identical).

- 2^n is the number of subsets of A. Each '2' represents the choice of 'whether a_i is in this subset'. For any two subsets, if they have different answers for any of these questions, then they are different subsets of A.
- We could also count the number of subsets by restricting the number of elements to $\{0, 1, \dots, n\}$. The number of elements with k elements is $\binom{n}{k}$, from the definition of the combinations.

□

3 Ex4 Q1 - Bayesian's View/Philosophy

3.1 The Original Question

Question: Consider a modified version of the Monty Hall problem. In this version, there are n boxes ($n \geq 3$), of which 1 box contains the prize and the other $n-1$ boxes are empty. You again select one box at first. Monty, who knows where the prize is, then opens $n-2$ of the remaining $n-1$ boxes, all of which are shown to be empty. If Monty has a choice of which boxes to open (*i.e.* , if the prize is in the box you chose at first), he will choose at random which one of the boxes to leave unopened.

Suppose that you have chosen Box 1, and then Monty opens Boxes 3 to n, leaving Box 2 unopened. After we have observed this, what is the probability that the prize is in Box 1, and what is the probability that it is in Box 2?

Solution. The standard Bayesian procedure would provide the answer. Our goal is:

$$\mathrm{P}(\text{prize in 1} \mid \text{Monty does not open 2}),$$

To solve for this, we denote events $A_i = \{\text{prize in } i\}$, for $i = 1, \cdots, n$, and $B = \{\text{Monty does not open 2}\}$. Apply the Bayes' Theorem and the Law of total probability:

$$\begin{aligned}
\mathrm{P}(A_1|B) &= \frac{\mathrm{P}(B \cap A_1)}{\mathrm{P}(B)} \\
&= \frac{\mathrm{P}(B|A_1)\mathrm{P}(A_1)}{\mathrm{P}(B)} \\
&= \frac{\mathrm{P}(B|A_1)\mathrm{P}(A_1)}{\sum_{i=1}^{n} \mathrm{P}(B|A_i)\mathrm{P}(A_i)},
\end{aligned} \tag{3.1}$$

Clearly, $\mathrm{P}(A_i) = \frac{1}{n}$ for all i, so the problem becomes solving $\mathrm{P}(B|A_i)$ for all $i = 1, \cdots, n$. For $i = 1$, we have:

$$\begin{aligned}
\mathrm{P}(B|A_1) &= \mathrm{P}(\text{does not open 2} \mid \text{prize in 1}) \\
&= \mathrm{P}(\text{randomly select } n-2 \text{ objects from } \{2, 3, \cdots, n\}, \text{ does not contain 2}) \\
&= \frac{\#\ \{\text{select } n-2 \text{ objects from } \{3, 4, \cdots, n\}\ \}}{\#\{\text{select } n-2 \text{ objects from } \{2, 3, 4, \cdots, n\}\ \}} \\
&= \frac{1}{\binom{n-1}{n-2}} \\
&= \frac{1}{n-1}
\end{aligned}$$

For $i = 2$, then:

$$\mathrm{P}(B|A_2) = \mathrm{P}(\text{does not open 2} \mid \text{prize in 2}) = 1$$

This is because Monty knows where the prize is: if the prize is in 2, he will leave 2 unopened. For $i \geq 2$, we have:

$$\begin{aligned}
\mathrm{P}(B|A_i) &= \mathrm{P}(\text{does not open 2} \mid \text{prize in } i) \\
&= \frac{\#\ \{\text{select } n-2 \text{ objects from } \{3, 4, \cdots, i-1, i+1, \cdots, n\}\ \}}{\#\{\text{select } n-2 \text{ objects from } \{2, 3, 4, \cdots, n\}\ \}} \\
&= 0.
\end{aligned}$$

The last step is because the set $\{3, 4, \cdots, i-1, i+1, \cdots, n\}$ only contains $n-3$ elements, so there is no way to choose $n-2$ without replacement from it, thus, the numerator is 0.

Putting these values back into 3.1, we get:

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} = \frac{\frac{1}{n-1}}{\frac{1}{n-1}+1} = \frac{1}{n}. \tag{3.2}$$

Similarly, we can show that $P(A_2|B) = \frac{n-1}{n}$.

3.2 Change the Setting a little bit

Now we change the question a little bit.

Question: Now we assume that Monty does not know where the prize is. What we see is that, he opened boxes $3, 4, \cdots, n$, and these boxes are empty.

You are still asked to solve for the conditional probability that the prize is in 1 and that the prize is in 2.

Solution: Apply Bayes' Theorem:

$$\begin{aligned}
P(\text{prize in 1} \mid 3, \cdots, n \text{ are empty}) &= \frac{P(3, \cdots, n \text{ are empty} \mid \text{prize in 1}) \cdot P(\text{prize in 1})}{P(3, \cdots, n \text{ are empty})} \\
&= \frac{1 \cdot \frac{1}{n}}{P(\text{the prize is in 1 or 2})} \\
&= \frac{1 \cdot \frac{1}{n}}{\frac{2}{n}} \\
&= \frac{1}{2}.
\end{aligned}$$

The same applies to $P(\text{prize in 2} \mid 3, \cdots, n \text{ are empty}) = \frac{1}{2}$.

3.3 Comparison and Comments

In both cases, despite the condition is written in different ways (in the original version, the condition is 'Monty does not open 2', while in the new version, the condition is '$3, \cdots, n$ are empty'), what we saw is the same: we saw that Monty choose boxes $\{3, 4, \cdots, n\}$, opened them, and it turns out that all these boxes are empty.

A key question is that why we have different result for these two cases? From the basic principle in science, the difference must lie in the difference of the condition we have: Monty knows where the prize is or not.

Another view is rather more interesting: In both cases, by the law of total probability, we have:

$$P(\text{prize in 1} \mid \text{condition}) + P(\text{prize in 2} \mid \text{condition}) = 1.$$

since $P(\text{prize in } i \mid \text{condition}) = 0$ for $i \geq 3$.

Thus, we could think of it as the unconditional probability $P(\text{prize in } i)$ 'flows' from box i into box 1 or 2, when the condition happens. With these in mind, we see that, if Monty knows where the prize is, the probability only flows into box 2; however, if he does not know it, the probability evenly flows into box 1 and box 2.

More concepts (such as 'entropy' in information theory, for those interested) are required to interpret it in more depth. I guess the result is quite amazing so far.

3.4 Bayesian's View on Conditional Probability

Conditional probability could be interpreted from a Bayesian's point of view. Basically, Bayesianist treat probabilities as subjective object - different persons could have different view on the probability of a certain events.

information could

4 Ex5 Q4 - Logics/Convergence of Sequences

4.1 The Original Question

In Ex5, question 4, given X's pmf $p(x) = \frac{1}{2^x}$ for $x = 1, 2, 3, \cdots$, you are asked to prove that $\mathbb{E}[2^X]$ does not exist.

$$\mathbb{E}[2^X] = \sum_{x=1}^{\infty} 2^x p(x) = \sum_{x=1}^{\infty} 1 := \lim_{n\to\infty} \sum_{x=1}^{n} 1 = \lim_{n\to\infty} n,$$

thus, it does not exist.

4.2 Existence of a Limit

However, this is not the end of the story if one wants to be mathematically rigorous. To be rigorous, first, recall the definition of convergence for a sequence $\{a_n\}$, say, '$a_n \to a$'.

Definition 3. *If for any $\varepsilon > 0$, there exists an $n_0 \in \mathbb{N}$, such that, for any $n > n_0$, $|a_n - a| < \varepsilon$, then we say sequence $\{a_n\}$ converge to a (denote as $a_n \to a$).*

Then, one should know how to take the negation of a proposition. Take the definition 1 as an example, we show how to take its negation. Basically, we want to express 'a_n does not converge to a'. Let us write the definition of '$a_n \to a$' using mathematical notations.

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n > n_0, |a_n - a| < \varepsilon. \tag{4.3}$$

4.3 Negation of a Proposition

Then, following the procedure:

1. Exchange $\forall$ and $\exists$;
2. take negation for the last short proposition;

We get:

$$\exists \varepsilon_0 > 0, \forall n \in \mathbb{N}, \exists n_1 > n, |a_{n_1} - a| \geq \varepsilon_0, \tag{4.4}$$

and this is the correct expression of 'a_n does not converge to a'.

Why this procedure works could be shown theoretically, but at this stage, knowing it and using it would be enough.

However, this is only 'a_n does not converge to a', for a specific a, which is not enough. Think of it intuitively, what do we actually mean by 'a_n does not converge'?

Bingo! 'a_n does not converge' means, 'for any $a \in \mathbb{R}$, a_n does not converge to a'. In logic notation:

$$\forall a > 0, \exists \varepsilon_0 > 0, \forall n \in \mathbb{N}, \exists n_1 > n, |a_{n_1} - a| \geq \varepsilon_0, \tag{4.5}$$

This is the sufficient and necessary condition for 'a_n does not converge'.

4.4 Two Useful Results with Examples

To prove 'a_n does not converge', by definition, we could directly verify 4.5 holds. Another method is to use some necessary conditions which has been found useful, and show that those necessary conditions are not satisfied, as a result, the original proposition does not hold.

Simple examples are:

Proposition 2. *For the sequence $\{a_n\}_{n=1}^{\infty}$, if there exists a sub-sequence $\{a_{n_k}\}_{k=1}^{\infty}$ s.t. $a_{n_k} \to a_0$ as $k \to \infty$, and another sub-sequence $\{a_{n_m}\}_{m=1}^{\infty}$ s.t. $a_{n_m} \to a_1$ as $m \to \infty$, where $a_0 \neq a_1$, then $\{a_n\}$ does not converge.*

Proposition 3. *For the sequence $\{a_n\}_{n=1}^{\infty}$, if the sequence is unbounded, then $\{a_n\}$ does not converge.*

Definition 4. *We say a sequence $\{a_n\}_{n=1}^{\infty}$ to be unbounded*[1] *if,*

$$\forall C \in \mathbb{R}_+, \forall n \in \mathbb{N}, \exists n_0 > n, a_{n_0} > C.$$

Sometimes it is denoted as $a_n \to \infty$ as $n \to \infty$, but I would say this is a misuse of the notation technically.

To prove Proposition 1 and Proposition 2, one should use the assumption of the sequence to show 4.5 holds. This could be found in a standard mathematical analysis textbook.

To gain some intuition, let's look at two examples, that correspond to two propositions. We will use the definition to verify these sequence does not converge:

1. $a_n = (-1)^n$, for $n = 0, 1, 2, \cdots$.

   For $a \in \mathbb{R}$ such that $a \neq 1$ or -1, take $\varepsilon_0 = \frac{\min(|1-a|,|1+a|)}{2}$, then for any $n \in \mathbb{N}$, take $n_1 = n+1$, we see that $|a_{n_1} - a| \geq \min(|1-a|, |-1-a|) = \min(|1-a|, |1+a|) > \frac{\min(|1-a|,|1+a|)}{2} = \varepsilon_0$;

   For $a = 1$, take $\varepsilon_0 = 1.5$ (lots of value could be taken typically), then for any $n \in \mathbb{N}$, we take $n_1 = 2n+1$ to be an odd number, then $|a_{n_1} - a| = |(-1) - 1| = 2 > 1.5 = \varepsilon_0$.

   For $a = -1$, take $\varepsilon_0 = 1.5$, then for any $n \in \mathbb{N}$, we take $n_1 = 2n$ to be an even number, then $|a_{n_1} - a| = |1 - (-1)| = 2 > 1.5 = \varepsilon_0$.

   To summarise, we have proved, for any $a \in \mathbb{R}$, there exists a $\varepsilon_0 > 0$, for any $n \in \mathbb{N}$, exists a $n_1 > n$, such that $|a_{n_1} - a| \geq \varepsilon_0$. It is complicated mainly because we need to design different values of ε_0 and n_1, for different values of a. If we use proposition 1, then no proof is needed any more.

2. $a_n = n$, $n = 0, 1, 2, \cdots$.

   For any $a \in \mathbb{R}$, we take $\varepsilon_0 = 1$, for any $n \in \mathbb{N}$, we take $n_1 = \max(n+1, \lceil a \rceil + 1)$[2], then:

   $$|a_{n_1} - a| \geq (\lceil a \rceil + 1) - a = 1 + (\lceil a \rceil - a) \geq 1.$$

5 Ex5 Q4 - St. Petersburg Paradox

It is called a 'paradox' since it seems to contradict our intuition that we cannot do anything to make sure we can win in a fair gambling. However, it does not contradict anything. If you do not agree, write down your opinion and check how it is explained in theory.

I am sure doing this will help foster one's intuition.

See this link for details.

[1] It is equivalent to 'there exists a monotonic sub-sequence of a_n goes to infinity'. Check!

[2] $f(x) = \lceil x \rceil$ is called ceiling function, it takes the smallest integer k satisfying $k \geq x$.

6 Ex6 Q1 - Cauchy-Schwartz Inequality

Basically, the motivation to introduce Cauchy-Schwartz inequality is to show why we have $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$ holds for any distribution.

6.1 Discrete Case

Proposition 4. *For any real values $a_1, \cdots, a_n, b_1, \cdots, b_n$, we have:*

$$\big(\sum_{i=1}^n a_i^2\big)\big(\sum_{i=1}^n b_i^2\big) \geq \big(\sum_{i=1}^n a_i b_i\big)^2$$

Proof. This is the classic version and is a nice practice of dealing with more than 1 summation notation.

$$\begin{aligned}
\text{LHS - RHS} &= \sum_{i=1}^n \sum_{j=1}^n a_i^2 b_j^2 - \sum_{i=1}^n \sum_{j=1}^n a_i b_i a_j b_j \\
&= \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n a_i^2 b_j^2 + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n a_i^2 b_j^2 - \sum_{i=1}^n \sum_{j=1}^n a_i b_i a_j b_j \\
&= \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n a_i^2 b_j^2 + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n a_j^2 b_i^2 - \sum_{i=1}^n \sum_{j=1}^n a_i b_i a_j b_j \\
&= \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n (a_i^2 b_j^2 - 2a_i b_i a_j b_j + a_j^2 b_i^2) \\
&= \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n (a_i b_j - a_j b_i)^2 \\
&\geq 0.
\end{aligned}$$

□

6.2 General Case

Proposition 5. *Suppose squared-integrable functions $f(x)$ and $g(x)$ ($\int_a^b f^2(x)dx < \infty$) are defined on $[a, b]$. We have:*

$$\Big(\int_a^b f^2(x)dx\Big)\Big(\int_a^b g^2(x)dx\Big) \geq \Big(\int_a^b f(x)g(x)dx\Big)^2$$

The proof can be found throughout the internet, and interested students could also check out 'Hölder's inequality'.

With this proposition, we have:

$$\begin{aligned}
\mathbb{E}[X^2]\mathbb{E}[1] &= \Big(\int x^2 f_X(x)dx\Big)\Big(\int 1 \cdot f_X(x)dx\Big) \\
&\geq \Big(\int |x| f_X(x)dx\Big)^2 \\
&\geq \Big(\int x f_X(x)dx\Big)^2 \\
&= \big(\mathbb{E}[X]\big)^2
\end{aligned}$$

7 Ex6 Q3 - Uniqueness of Moment Generating Function

In Q3, we are given a moment-generating function of r.v. X: $M_X(t) = \frac{2}{11}e^t + \frac{4}{11}e^{2t} + \frac{4}{11}e^{4t} + \frac{1}{11}e^{8t}$ for $t \in (-\infty, \infty)$, and we are asked to solve for the probability distribution of X. It is easy to obtain the result by simply observing: $\mathrm{P}(X = 1) = \frac{2}{11}, \mathrm{P}(X = 2) = \mathrm{P}(X = 4) = \frac{4}{11}, \mathrm{P}(X = 8) = \frac{1}{11}$, and for $x \notin \{1, 2, 4, 8\}, \mathrm{P}(X = x) = 0$.

However, a natural problem is proposed: Is such a X unique? Is there another different distribution with the same moment-generating function?

I collected several results.

7.1 Discrete Case

Proposition 6. [3] *For $n \geq 0$, we have:*

$$\frac{d^n}{ds^n}\mathbb{E}[s^X]\bigg|_{s=0} = n! \cdot \mathrm{P}(X = n),$$

Proof. In fact,

$$\frac{\mathrm{d}^n}{\mathrm{d}s^n}\mathbb{E}[s^X] = \mathbb{E}[\frac{\mathrm{d}^n}{\mathrm{d}s^n}s^X] = \mathbb{E}[X(X-1)\cdots(X-n+1)s^{X-n}],$$

when $s = 0$,

$$X(X-1)\cdots(X-n+1)s^{X-n} = \begin{cases} 0 & \text{if } X \neq n \\ n! & \text{if } X = n \end{cases},$$

thus, $\mathbb{E}[X(X-1)\cdots(X-n+1)s^{X-n}]\big|_{s=0} = n! \cdot \mathrm{P}(X = n)$. □

With this result, note that $\mathbb{E}[s^X] = \mathbb{E}[e^{X \log s}] = M_X(\log s)$. If r.v.s' X and Y only takes value in, say, $\mathbb{N} = \{0, 1, \cdots\}$, and they have the same MGF, *i.e.* , $M_X(t) = M_Y(t)$ for $t \in (-\delta, \delta)$ for some $\delta > 0$.

From this proposition, we see that, for any $n \in \mathbb{N}$, $\mathrm{P}(X = n) = \mathrm{P}(Y = n)$.

7.2 General Case - Fourier Transformation

In probability theory, $\psi_X(t) := \mathbb{E}[e^{itX}]$ is called the characteristic function of random variable X. It is obvious that $\psi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[e^{(it)X}] = M_X(it)$.

Fourier inversion theorem implies the uniqueness of the distribution. I would suggest reading this lecture notes provided by Stanford University.

7.3 Warning - Same Distribution

We have seen that if two random variables X, Y have the same moment-generating function $M_X(t) = M_Y(t)$, then they have the same distribution. I want to emphasize the difference between '' and '$X = Y$'.

In theory, X, Y could have the same distribution when they are defined on different sample spaces.

As an example, suppose James has a fair coin in hand, and Yutong also has a fair coin in hand. Denote $X = \{$the number of heads, tossing James' coin n times$\}$, and $Y = \{$the number of heads, tossing Yutong's coin n times$\}$. According to the definition of binomial distribution, $X \sim \text{Binomial}(n, \frac{1}{2})$, and $Y \sim \text{Binomial}(n, \frac{1}{2})$.

However, if we want to talk about '$X = Y$', we should bear in mind that the sample space S_X of r.v. X is the result of random experiments for the coin in James' hand, and the sample space S_Y of r.v. Y is the result of random experiments for the coin in Yutong's hand. Since X and Y are defined on different

[3]The proposition is collected from this link.

sample spaces, the event $\{X = Y\}$ is actually defined on the compound sample space, mathematically, the Cartesian product $S_X \times S_Y$.

In this example, one can see that 'X, Y have the same distribution' does not imply '$X = Y$'. One shall remember that, 'identical distributed' is a weaker proposition to say the two random variables are quite similar to each other.

8 Ex6 Q4 - The Integral does not Exist

We would like to talk about the relationship between the following quantities:

$$\sum_{k=1}^{\infty} \frac{1}{k}, \quad \int_0^1 \frac{1}{x} \mathrm{d}x, \quad \int_1^{\infty} \frac{1}{x} \mathrm{d}x$$

8.1 Harmonic Series

$\sum_{k=1}^{\infty} \frac{1}{k}$ is called the harmonic series, and it is quite famous. First of all, it does not converge.

Proposition 7. $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$. *(to be precise,* $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k}$ *does not exist.)*

Proof. A simple argument could give the result:

$$\sum_{k=2}^{2^m} \frac{1}{k} = \sum_{k=2^0+1}^{2^m} \frac{1}{k} = \sum_{j=0}^{m} \Big(\sum_{k=2^j+1}^{2^{j+1}} \frac{1}{k} \Big) > \sum_{j=0}^{m} \Big(\sum_{k=2^j+1}^{2^{j+1}} \frac{1}{2^{j+1}} \Big) = \sum_{j=0}^{m} \Big(\frac{1}{2^{j+1}} \cdot 2^j \Big) = \frac{1}{2} m.$$

□

8.2 Relationships

Proposition 8. *For any* $N \in \mathbb{N}$*, it holds that:*

$$\int_1^{N+1} \frac{1}{x} dx < \sum_{i=1}^{N} \frac{1}{i} < 1 + \int_1^{N} \frac{1}{x} dx.$$

Proof. The only argument used in the proof is that $f(x) = \frac{1}{x}$ is a decreasing function. For the left-hand side:

$$\int_1^{N+1} \frac{1}{x} \mathrm{d}x = \sum_{i=1}^{N} \int_i^{i+1} \frac{1}{x} \mathrm{d}x < \sum_{i=1}^{N} \int_i^{i+1} \frac{1}{i} \mathrm{d}x = \sum_{i=1}^{N} \frac{1}{i}.$$

For the right-hand side:

$$1 + \int_1^{N} \frac{1}{x} \mathrm{d}x = 1 + \sum_{i=1}^{N-1} \int_i^{i+1} \frac{1}{x} \mathrm{d}x > 1 + \sum_{i=1}^{N-1} \int_i^{i+1} \frac{1}{i+1} \mathrm{d}x = 1 + \sum_{i=2}^{N} \frac{1}{i} = \sum_{i=1}^{N} \frac{1}{i}.$$

□

Set $y = \frac{1}{x}$, then $y = 1$ when $x = 1$, and $y \to 0$ when $x \to \infty$, thus, $\int_1^{\infty} \frac{1}{x} \mathrm{d}x = \int_1^0 y \mathrm{d}(\frac{1}{y}) = \int_0^1 \frac{1}{y} \mathrm{d}y$ (if the limit exists. Here we use the proof towards contradiction).

8.3 Riemann Zeta Function

It is relevant to the Riemann Zeta function:

$$\zeta(x) := \sum_{n=1}^{\infty} \frac{1}{n^x}.$$

We can see that $\zeta(1) = \sum_{n=1}^{\infty} \frac{1}{n}$ is the harmonic series which does not converge. Another famous example is that:

Proposition 9.

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

I spent days thinking about why a summation of lots (infinitely many) of rational numbers could be an irrational number when I was a high school student. For finite many summations, we could prove by induction that it is still a rational number. It reveals the truth that there are essential differences between finite summation and infinite summations.

Proof. Expand $f(x) = x$ in terms of the complete orthogonal basis $\{e^{-ix}, e^{-2ix}, e^{-3ix}, \cdots\}$. Calculate $\|x\|^2 = \langle x, x\rangle$. $\square$

It is also clear for which value of x, does $\zeta(x)$ converge.

Proposition 10. *For $x > 1$, $\zeta(x)$ converges.*

Proof. For any $x > 1$, $f(t) = \frac{1}{t^x}$ is a decreasing function, and there must be some $\varepsilon > 0$ such that $x > 1 + \varepsilon$. Then:

$$\begin{aligned}
\sum_{k=1}^{N} \frac{1}{k^x} &< 1 + \int_1^N \frac{1}{t^x} \mathrm{d}t \\
&< 1 + \int_1^N \frac{1}{t^{1+\varepsilon}} \mathrm{d}t \\
&= 1 + \Big(-\frac{1}{\varepsilon}\frac{1}{t^{\varepsilon}}\Big)\Big|_{t=1}^{t=N} \\
&= 1 + \frac{1}{\varepsilon}(1 - \frac{1}{N^{\varepsilon}})
\end{aligned}$$

Thus, for any $N \in \mathbb{N}$, $\sum_{k=1}^{N} \frac{1}{k^x} < 1 + \frac{1}{\varepsilon}(1 - \frac{1}{N^{\varepsilon}}) < 1 + \frac{1}{\varepsilon}$, which does not depend on N. $\square$

9 Ex11 Q3 - Normal Approximation

From the Central Limit Theorem, we know that, for any distribution P such that its second-order moment exists, if X_i's are i.i.d. copies following P, then:

The cdf of $\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)$ - denote as $F_n(x)$, converges to $\Phi(x)$.

$$F_n(x) := \mathrm{P}\Big(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \le x\Big),$$

$$\sup_{x\in\mathbb{R}} \Big| F_n(x) - \Phi(x) \Big| \to 0.$$

Here I would like to talk about two things: 1. To foster one's intuition, it is helpful to do simulations to verify it does converge to normal; and 2. the quantitative central limit theorem (Berry-Esseen Theorem), which gives the order of convergence.

9.1 Intuition

The best part of the central limit theorem is that, we do not need X to follow any particular distribution. As long as the variance exists, the sample mean will eventually converge.

The thoughts to do simulation:

1. Choose a distribution P (with finite second moment);
2. Choose several 'n' to illustrate, for example, $n = 1, 5, 15, 25, 50$;
3. Choose a m. For each n chosen in the last step, (i.i.d.) generate $m \times n$ numbers of $X_i \sim P$, and calculate m numbers of entity of $\overline{X_n}$;
4. draw the histogram for these m numbers of $\overline{X_n}$.

9.2 Berry-Esseen Theorem

Basically, it says: all distribution with finite third moment satisfies:

$$\sup_{x \in \mathbb{R}} \left| F_n(x) - \Phi(x) \right| \leq \frac{C\rho}{\sigma^3 \sqrt{n}},$$

where C is a constant greater than 0.4, and $\rho = \mathbb{E}[|X_1|^3]$, and $\sigma = (\mathbb{E}[|X_1|^2])^{1/2}$ is the standard deviation of X_1.

The implication of this theorem is that: the normal approximation is not that accurate, it is of order $O(\frac{1}{\sqrt{n}})$. In many applications, or for some particularly important distributions like Bernoulli etc., there are better ways to approximate the distribution of $\sum_{i=1}^{n} X_i$. For those interested, check Bernstein's inequality, Hoeffding's inequality, etc.