Summary: Probability Theory for EOR

Lingwei Kong

University Groningen

Course outline used for summary and reviewing

Some general information

Why study probability and the course structure

- Econometrics = uncertainty
- We never know the exact outcome (economy one year from now?)
- Maybe we can determine the probability of different outcomes/events.

What does this mean: "*probability*"?

Some general information

Why study probability and the course structure

- Tutorial exercises (Bonuses)
- Assignments (20%, in pairs *within* tutorial groups, week 3/6).
- Closed-book on-campus exam (80%, individual). **One exercise, 3/10, you would have seen either from assignments, weekly tutorials or selected exercises.**

- Practice, practice, practice. We have other selected exercises. Understand instead of knowing by heart. Earn bonus by working hard and participating the tutorials.
- Keep up. **One chapter each week!**

Some general information

Some useful info about R, Python and GitHub

- All this coin flipping and dice throwing is a lot of work.
- It helps to have a computer.
- We will use R (Python is also quite similar), a freely available programming tool that is used a lot for statistical purposes.
- I recommend using RStudio from rstudio.com (or Google Colab where you can also find interesting materials on Python coding). You may find useful information on R coding from this website.
- Let's do a short tour on R and Python. Pick one you like. The textbook uses R though.

Week 1

Week 1

I. Layman's talk: what is probability

Probability is a function with special properties.

Week 1

A. Sample space and naive definition of probability

Sets of possible outcomes are elements in the domain of one probability function

Definition (Event, sample space)

The **sample space** S is the set of all possible **outcomes** of the **experiment**. An **event** A is a **subset** of the sample space S, and we say that A **occurred** if the actual outcome is in A.

Note that: (1) $S \subseteq S$, and we always include S as one event (the largest one as it contains all possible outcomes); (2) if A and B are two events, then their intersections, unions are also included as events, so are their complements; (3) $\emptyset$ is also an event (zero probability though); (4) if A_i's are events, so is $\cup_{i=1}^{\infty} A_i$.

Week 1

A. Sample space and naive definition of probability

- Events are sets of possible outcomes.
- Let's consider a special case: **finite outcomes, all the outcomes are equally likely to happen.**
- The more outcomes one event contains, the more likely it happens.

Definition (Naive definition of probability)

Denote A be an event for an experiment with a **finite*** sample spaces S and each outcome is **equally likely*** to happen:

$$P_{\text{Naive}}(A) = \frac{\text{number of outcomes favorable to } A}{\text{number of outcomes in } S} = \frac{|A|}{|S|}.$$

Week 1

B. How to count

Number of all outcomes $|S|$: Multiplication rule (sampling without replacement)

- We have n cards.
- We select k cards *one at a time*.
- We keep the selected card (without replacement).

How many outcomes (N) when order matters?

- Step 1: n possible outcomes. Keep the card!
- Step 2: $n-1$ possible outcomes. Keep the card!

$\vdots$

- Step k: $n-k+1$ possible outcomes.

MR: $N = n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$ *ordered* outcomes.

Week 1

B. How to count

Definition

Factorial

We define $n!$ (say: n factorial) as

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$$
$$0! = 1$$

$$N = n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

Alternative interpretation of how we arrive at this number of ordered outcomes:

- **We can order the cards in $n!$ ways.** Permutation of n items.
- Don't care about the ordering of the $(n-k)$ not selected cards. The not-selected cards can be ordered in $(n-k)!$ ways.

But to calculate $|S|$, we still need to adjust for over-counting! **We don't care for the orders of selected cards either!** Divide by $k!$.

Week 1

B. How to count

Number of all outcomes $|S|$: we adjust for overcounting:
$$|S| = N/k! = \frac{n!}{k!(n-k)!} = \binom{n}{k}.$$

The **binomial coefficient** gives the number of possible outcomes of picking k items out of n items (where orders do not matter).

In our case: $n = 52, k = 5$.

Also, we know these outcome are equally likely, which is why we can use the naive probability.

Week 1

C. Binomial coefficient (story proof)

- We know that the binomial coefficient is defined as

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

  We set it to be zero if $n < k$.
- Many mathematical identities involve the binomial coefficient.
- Sometimes, these identities can be easily proven algebraically.

Story proof of $\binom{n}{k} = \binom{n}{n-k}$.

Week 1

D. General definition of probability and properties

Definition (General definition of probability)

A **probability function (measure)** P maps an event , a (well-constructed) subset A of the sample space S ($A \subseteq S$), to the probability of the event $P(A)$, a real number within $[0, 1]$. It should satisfy the *Axioms of Probability*

1. $P(\emptyset) = 0$, and $P(S) = 1$.
2. If $A_1, A_2, \ldots$ are disjoint (i.e. mutually exclusive, $A_i \cap A_j = \emptyset, i \neq j$) events, then

$$P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- **Note:** reporting probabilities outside $[0, 1]$ without the disclaimer that this cannot be the right answer will be considered a capital blunder and renders your complete answer invalid (even if that answer is 99% correct).
- The second axiom (countable additivity) involves a summation of a infinite series. It is more intuitive if you only choose finitely many non-empty A_i's: $P\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.

Week 1

D. General definition of probability and properties

General definition embeddes the naive one

- $P_{\text{Naive}}(\emptyset) = |\emptyset|/|S| = 0.$
- $P_{\text{Naive}}(S) = |S|/|S| = 1.$
- $P_{\text{Naive}}\left(\cup_{i=1}^{m} A_i\right) = \left|\cup_{i=1}^{m} A_i\right|/|S| = \sum_{i=1}^{m} |A_i|/|S| = \sum_{i=1}^{m} P_{\text{Naive}}(A_i)$

Now we can directly work with the general definition, and the naive one is simply a special case of the general definition.

The properties derived based on the general definition also holds for the naive one.

Week 1

D. General definition of probability and properties

Theorem

For any events A and B, we would have

- $P(A^c) = 1 - P(A)$.
- *If* $A \subseteq B$, $P(A) \leq P(B)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Week 1

D. General definition of probability and properties

Theorem

(Inclusion-exclusion principle: n events)
For n events $A_i, i = 1, \cdots, n$, *we would have* $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(\cap_{i=1}^n A_i)$.

Week 1

Summary with a quiz

A small quiz: formalize a fair coin tossing game with the probability notation.

Week 1

Summary with a quiz

- **Domain**: $\{\{Head, Bottom\}, \{Head\}, \{Bottom\}, \emptyset\} = \{S, A, A^c, \emptyset\}$ with $S = \{Head, Bottom\}, A = \{Head\}$.

  When an event only contains one outcome, sometimes we drop the bracket "$\{\}$" so you can also write $A = Head$.
- **Codomain**: $[0, 1]$.
- According to the naive definition, $P(A) = 1/2, P(A^c) = 1/2, P(S) = 1, P(\emptyset) = 0$.

Check whether it satisfies the general definition?

Week 1

Summary with a quiz

Right or wrong:

- $P(A) = 0$;
- $P(B) < 0$;
- $P(B) > 1$;
- $P(A) + P(B) > 1$;
- $P(A) = 1 - P(B)$ so $A = B^c$.

Week 1

Summary with a quiz

Right or wrong:

- $P(A) = 0$; sometimes correct sometimes wrong. Note there are events also having zero probability other than the empty set.
- $P(B) < 0$; wrong for sure.
- $P(B) > 1$; wrong for sure.
- $P(A) + P(B) > 1$; sometimes correct sometimes wrong. Wrong for sure if $B \subseteq A^c$.
- $P(A) = 1 - P(B)$ so $A^c = B$. sometimes correct sometimes wrong. For example, let $B = A^c \cup C$ where $P(C) = \emptyset$. The converse is correct though.

Week 2

Week 2

- Assignment 1 will be published today 11 AM, need to enroll first to gain access. (The enrollment is temporarily re-opened and will be closed today 3 PM).
- Enroll within your own tutorial groups! Find available assignment groups on the Enrollment page (they would show groups that are not full yet). Use discussion board. Or email me as the last try. Or work alone.
- You can finish all the exercises now, except the last two may require some additional concepts from the next week material.
- Deadline on Dec 3rd. Hand in PDF report via Nestor.

Week 2

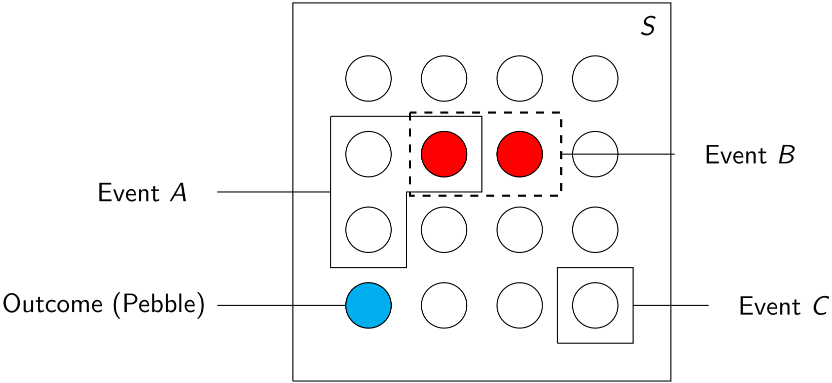
II. Layman's talk: what is conditional probability

$$\begin{array}{llcr} \mathbb{P}(\cdot|B): & \{A_i \subseteq S, i \in I\} & \mapsto & [0,1] \\ & A & \mapsto & P(A \cap B)/P(B)\ (*) \end{array}$$

Measure how likely A occurred given the B already occurred ($P(B) > 0$).
(Informally, the percentage of outcomes in B that also in A.)

Week 2

II. Layman's talk: what is conditional probability



If we know B occured, what can you say about event A and C?

$P(A|B) = 1/2$, $P(C|B) = 0$, $P(B|B) = 1$ (B is like our new sample space.)

Week 2

A. conditional probability

Definition

Conditional probability of A given B ($P(B) > 0$):
if A and B are events with $P(B) > 0$, then the conditional probability of A given B, denoted by $P(A|B)$, is defined as $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

$$\begin{array}{llll} \mathbb{P}(\cdot|B): & \{A_i \subseteq S, i \in I\} & \mapsto & [0,1] \\ & A & \mapsto & P(A \cap B)/P(B) \ (*) \end{array}$$

Week 2

B. Bayes' Rule and LOTP

- **Bayes'rule**
  Suppose $P(A), P(B) > 0$.
  $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$
- **Law of total probability**
  If we have a partition of S: disjoint A_i's such that $\cup_i A_i = S$ and $P(A_i) > 0$, then
  $$P(B) = \sum_i P(B|A_i)P(A_i).$$

Week 2

C. Independence/conditional independence concerning two random events

- A and B are independent if

$$P(A \cap B) = P(A)P(B).$$

- Conditional independence!: A and B are independent **conditional on/given** C if

$$P(A \cap B|C) = P(A|C)P(B|C)$$

- Knowing that B (or A) also occured does not provide any extra information on the probability of A (or B) once we already learn that C occured (we assume $P(C \cap B) > 0$):

$$P(A|C,B) = \frac{P(A \cap B|C)}{P(B|C)} = \frac{P(A|C)P(B|C)}{P(B|C)} = P(A|C)$$
$$P(B|C,A) = \frac{P(A \cap B|C)}{P(A|C)} = \frac{P(A|C)P(B|C)}{P(A|C)} = P(B|C)$$

$P(A|C,B) := P(A|C \cap B) = \frac{P(A \cap (C \cap B))}{P(C \cap B)} = \frac{P((A \cap B) \cap C))}{P(B \cap C)} = \frac{P((A \cap B) \cap C))/P(C)}{P(B \cap C)/P(C)} = \frac{P(A \cap B|C))}{P(B|C)}$.

Week 2

D. show the usefulness of conditioning via four exercises

Exercise: three fair die.

- At the start of the game, you have one dollar. Options
  - Lose the dollar, then D occurs
  - Keep the dollar, then the exact same game repeats
  - Win a dollar, take a moment!
- The complement of a **Never-Ending Game** is the probability that you lose all your money (Game Ends).
- Define D as the event **Game Ends**.

$$
\begin{aligned}
&P(D)\\
=&P(D|\{L1\$\})P(\{L1\$\})+P(D|\{K1\$\})P(\{K1\$\})+P(D|\{W1\$\})P(\{W1\$\})\\
=&\frac{1}{3}+\frac{1}{3}P(D)+\frac{1}{3}P(D)^2.
\end{aligned}
$$

Solving for $P(D)$ yields $P(D)=1$, so $P(NEG)=0$. **Sometiems, it is easier to work with conditional probability.**

Week 2

D. show the usefulness of conditioning via four exercises

Exercise: I

- High income: $P(H) = \frac{8}{100}$.
- High income | E: $P(H|E) = \frac{3}{5}$
- High income | M: $P(H|M) = \frac{3}{4}$.
- High income | E and M $P(H|M, E) = \frac{3}{4}$.
- High income | not M: $P(H|M^C) = \frac{2}{92}$
- High income | E, but not M, $P(H|M^C, E) = 0$

Lessons:

- H is not independent of E.
- **H is independent of E given M. Conditional independence!**
- H is not independent of E given M^C.
- *What fits is best.*
  - *If M, it doesn't matter whether you E.*
  - *If M^C, you are better of E^C.*
- But we are still not sure about the actual case, honestly.

Week 2

D. show the usefulness of conditioning via four exercises

Exercise: II

- The expert witness says

$$P(D_1|I) \approx \frac{1}{8500}, \quad P(D_1 \cap D_2|I) \approx \frac{1}{8500^2}$$

- Then he states that *therefore*

$$P(I|D_1 \cap D_2) \approx \frac{1}{8500^2}$$

- Bayes says no!

$$P(I|D_1 \cap D_2) = \frac{P(D_1 \cap D_2|I)P(I)}{P(D_1 \cap D_2|I)P(I) + P(D_1 \cap D_2|I^C)P(I^C)}$$

If *Presumption of innocence* (namely, $P(I^C) \approx 0$), then $P(I|D_1 \cap D_2) \approx 1$.

Week 2

D. show the usefulness of conditioning via four exercises

Exercise: III (Simpson's paradox)

Job Performance	You (Bert)		Jim	
	Task 1	Task 2	Task 1	Task 2
Fails	40	1	10	18
Successes	80	22	15	100

Aggregating across the two tasks however

$$P(S|B) \approx \frac{102}{143} < \frac{115}{143} \approx P(S|B^C)$$

Using the LOTP

$$P(S|B) = P(S|T_1, B)P(T_1|B) + P(S|T_2, B)P(T_2|B)$$
$$P(S|B^C) = P(S|T_1, B^C)P(T_1|B^C) + P(S|T_2, B^C)P(T_2|B^C)$$

Week 2

Summary by quiz

Suppose now we have

$$P(A \cap B) = P(A)P(B)$$
$$P(A \cap C) = P(A)P(C)$$
$$P(C \cap B) = P(C)P(B)$$

Is it true that if $P(B \cap C) > 0$

$$P(A|B, C) = P(A)?$$

Week 2

Summary by quiz

Note that, $P(A|B,C) = \frac{P(A \cap B \cap C)}{P(B \cap C)}$, we would require $P(A \cap B \cap C) = P(A)P(B)P(C)$.

- It could be the cases such that: A={First get head}; B={Second get head}; C={One head one bottom}.
- We only discuss independence concerning two events, and with more events, we require more *products* (i.e., the probability of the joint event of any arbitry combination of the events needs to be equal to the products of the probabilties of each event present in the joint event) in order to achieve *independence*
- Read textbook 2.5.

Week 2

Summary by quiz

- Conditional probabilities are not always intuitive!
- Read the textbook. For example, DIY study 2.2.5 – 2.2.7, 2.4.

Week 2

Summary by quiz

Right or wrong?

- $P(A|B) = 1.1$.
- $P(A|B) = P(B|A)$.
- If A and B are independent, A and C are independent, B and C are independent, then $P(A|B, C) = P(A)$.
- There are cases that $A_1 \cup A_2 \subsetneq S$ but $P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)$.

Week 2

Summary by quiz

Right or wrong?

- $P(A|B) = 1.1.$ wrong
- $P(A|B) = P(B|A).$ wrong (unless additionally $P(A) = P(B) > 0.$)
- If A and B are independent, A and C are independent, B and C are independent, then $P(A|B, C) = P(A)$.worng (unless additionally $P(A \cap B \cap C) = P(A)P(B)P(C)$.)
- There are cases that $A_1 \cup A_2 \subsetneq S$ but $P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2).$ For example: $B = \emptyset$ or $A_1 \cup A_2 \cup A_3 = S$ with $P(A_3) = 0.$

Week 3

Week 3

III. Layman's talk: what is random variable

X maps elements from S to numbers! Now we can work with functions/numbers!

$$\begin{tikzcd}
s \in S \arrow[r, "X"] \arrow[dr, dashed, "P : A \subseteq S"'] & Numbers \in \mathcal{C} \arrow[d, dashed, "P : \{X \in C \subseteq \mathcal{C}\} \subseteq S"] \\
& {[0,1]}
\end{tikzcd}$$

X essentially is a function, and when we manipulate random variables, it would be similar to manipulating functions and there are so many transformations we can work with: summation, division, max, min...

- We can now use $\{X \in C\}$ with $C \subseteq \mathcal{C}$ to represent different random events, and properties derived for random events would be inherited.
- Given the introduction of these additional structures, we have a larger space to explore: conditional probability/independence based on random variables, random variables with specific probability functions...

Random variables are functions, there are selected functions with nice properties, and so are some selected random variables.

Week 3

A. random variables (real-valued and discrete)

A simple example:

Throw a coin n times. Define X to be the number of heads.

- X takes an outcome s in the sample space S (the domain of function X), and maps that to a number on the real line (the co-domain of real-valued function X).

$$X(HHTHTHHTHT) = 6,$$
$$X(HHTHTHHTTH) = 6,$$
$$X(TTTTTTTTTT) = 0,$$
$$\vdots$$

- X is a *function* that maps each $s \in S \to \mathbb{R}$.
- Because the outcome is random, X is called a *random variable* (r.v.). But it is simply a function.

Week 3

B. CDF and PMF

Definition (CDF of real-valued r.v.)

The **cumulative distribution function (CDF)** of an **real-valued** r.v., $X : S \to \mathbb{R}$, is the function, usually denoted by F_X, which maps numbers to numbers within [0,1].

$$F_X(x) = P(X \leq x).$$

- We only need to know probability values of the following types of events: $\{X \leq x\}, x \in \mathbb{R}$.

Properties:

- **Increasing function from zero to one:** $0 \leq F_X(x_1) \leq F_X(x_2) \leq 1, \forall -\infty < x_1 \leq x_2 < \infty$.

  (a): $\lim_{x \to -\infty} F_X(x) = 0$; think about $\emptyset$!

  (b): $\lim_{x \to \infty} F_X(x) = 1$; think about S!

- **Right-continuous:** $F_X(a) = \lim_{x \downarrow a} F(x)$.

Week 3

B. CDF and PMF

Definition (PMF of discrete r.v.)

The **probability mass function (PMF)** of a **discrete** r.v., $X : S \mapsto \{x_1, x_2, \cdots\}$, is the function:

$$p_X(x) = P(X = x).$$

- We only need to know probability values of the following types of events: $\{X = x\}, x \in supp(X)$.

Properties:

- **Non-negativity**: $p_X(x) > 0$ if $x \in \text{Supp}(X)$; $p_X(x) = 0$ otherwise.
- **Sum to 1.** Suppose X has support $x_1, x_2, \ldots$: $\sum_{j=1}^{\infty} p_X(x_j) = 1$.

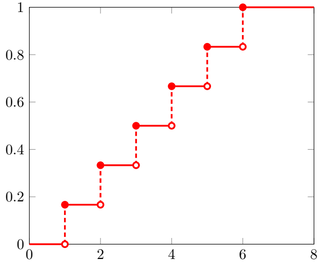
Week 3

B. CDF and PMF

Simple example: Throw a fair six-sided die. What is the PMF for the random variable X denotes the number of eyes of the outcome if we throw a six-sided fair die for one time?

$$P(X = i) = 1/6, i = 1, 2, 3, 4, 5, 6.$$

From the PMF, we can also derive CDF, let's draw a graph:



Week 3

C. Binomial; Uniform discrete; and Hypergeometric

- Throw a special coin n times (with probability p getting head in each flip), let X denotes the number of heads in n throws.
- $X \sim \text{Bin}(n, p)$.
- The PMF of X is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; k = 0, 1, \cdots, n.$$

- $\text{Bin}(1, p)$ is also called Bernolli distribution $\text{Bern}(p)$.

Week 3

C. Binomial; Uniform discrete; and Hypergeometric

- C contains n different numbers ($|C| = n$), Throw a n-sided fair die and each side is marked with a different number in C. Denote X the number of one die roll, and $\{X = x\}$ for any $x \in C$ are equally likely to occur.
- $X \sim$Dunif(C).
- The PMF of X is

$$P(X = x) = 1/n; x \in C.$$

- What is $P(X \in \{x_1, x_2\}), x_1 \neq x_2, x_1, x_2 \in C$?
  $2/n$.
- Bin$(1, 0.5)$ also a Dunif$(\{0, 1\})$.

Week 3

C. Binomial; Uniform discrete; and Hypergeometric

- Mark n different balls with $*$'s from a urn of w white balls and b **black** balls (each ball is equally likely to be marked), and X is the number of balls being both white and marked with $*$'s.
- $X \sim$HGeom(w, b, n).
- The PMF of X is

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}; \max\{n-b, 0\} \le k \le \min\{n, w\}.$$

- $X \sim$HGeom(w, b, n) and $Y \sim$HGeom$(n, w+b-n, w)$ have the identical distribution:

$$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = \frac{\binom{n}{k}\binom{w+b-n}{w-k}}{\binom{w+b}{w}}$$

Mark w different balls with white from a urn of n balls with $*$'s and $w+b-n$ balls without $*$'s (each ball is equally likely to be marked), and Y is the number of balls being both white and marked with $*$'s.

Week 3

D. Independence

Real-valued r.v.'s X, Y are independent if

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y); \forall x, y \in \mathbb{R}.$$

Discrete r.v.'s X, Y are independent if

$$P(X = x, Y = y) = P(X = x)P(Y = y); \forall x \in \text{supp}(X), y \in \text{supp}(Y).$$

Real-valued r.v.'s X, Y are **conditionally** independent given a discrete r.v. Z if

$$P(X \leq x, Y \leq y | Z = z) = P(X \leq x | Z = z)P(Y \leq y | Z = z);$$
$$\forall x, y \in \mathbb{R}, z \in \text{supp}(Z).$$

Discrete r.v.'s X, Y are are **conditionally** independent given a discrete r.v. Z if

$$P(X = x, Y = y | Z = z) = P(X = x | Z = z)P(Y = y | Z = z);$$
$$\forall x \in \text{supp}(X), y \in \text{supp}(Y), z \in \text{supp}(Z).$$

Week 3

E. Functions of random variables

Functions of a real-valued r.v.'s

For an r.v. $X : S \mapsto \mathbb{R}$ with sample space S, and a proper function $g : \mathbb{R} \to \mathbb{R}$, $g(X)$ is the r.v. that maps s to $g(X(s)) \in \mathbb{R}$ for all $s \in S : S \mapsto \mathbb{R}$.

Example:

$X \sim$Bern(p), then $Y = 2(X + 1)$ is also an r.v. with PMF:

$$P(Y = 2) = 1 - p; P(Y = 4) = p.$$

For a discrete r.v. X, the PMF of $Y = g(X)$:

$$P(Y = y) = \sum_{x:g(x)=y} P(X = x); y \in \{g(x) : x \in \text{supp}(X)\}.$$

Week 3

E. Functions of random variables

Functions of multi real-valued r.v.'s

For multiple r.v.'s $X_i : S \mapsto \mathbb{R}, i = 1, 2, 3 \cdots, n$ with sample space S, and a proper function $g : \mathbb{R}^n \mapsto \mathbb{R}$, $g(X_1, X_2, \cdots, X_n)$ is the r.v. that maps s to $g(X_1(s), X_2(s), \cdots, X_n(s)) \in \mathbb{R}$ for all $s \in S : S \mapsto \mathbb{R}$.

Example:

$X_1, X_2 \sim$Bern(0.5) and they are independent, then $Y = X_1 + X_2$ is also an r.v. with PMF:

$$P(Y = 0) = 1/4; P(Y = 1) = 1/2; P(Y = 2) = 1/4.$$

Week 3

Some additional results

- $X \sim$Bin(n, p), $Y \sim$Bin(m, p). X and Y are independent:
  (a) $Z = X + Y$, $Z \sim$Bin$(n + m, p)$.
  (b) The conditional distribution of X given $Z = r$ is HGeom(n, m, r):

$$P(X = k|Z = r) = \frac{\binom{n}{k}\binom{m}{r-k}}{\binom{n+m}{r}}$$

- $X \sim$HGeom(w, b, n), if $N = w + b \to \infty$ and $p = \frac{w}{w+b}$:
  The PMF of X converges to the Bin(n, p) PMF:

$$P\left(X = k\right) = \binom{n}{k} \frac{\prod_{i=0}^{k-1}(p - \frac{i}{N}) \prod_{j=0}^{n-k-1}(q - \frac{j}{N})}{\prod_{l=1}^{n-1}(1 - \frac{l}{N})} \to \binom{n}{k} p^k q^{n-k}$$

Week 3

Summary with Quizzes

- A particle moves n steps on a number line. The particle starts at 0, and at each step it moves 1 unit to the right or to the left, with equal probabilities. Assume all steps are independent.

  Let Y be the particle's position after n steps.

  Find the PMF of Y.

Week 3

Summary with Quizzes

- n is the total number of steps. Let X be the number of steps to the right. Then $X \sim \text{Bin}(n, 1/2)$.
- Suppose $X = j$, then we have set $n - j$ steps to the left.
- Denote by Y the position after n steps.
- $Y \sim 2X - n$.

Week 3

Summary with Quizzes

- $X \sim \text{Bin}(n, 1/2)$.
- $Y = 2X - n$
- Find the PMF.

$$P(Y = k) = P(2X - n = k) = P(X = \frac{1}{2}(k + n)) = \binom{n}{\frac{k+n}{2}} \frac{1}{2^n}.$$

Week 3

Summary with Quizzes

Right or wrong

- X,Y are conditional independent given $Z = 1$, then X,Y are conditional independent given Z.
- $P(X + Y = 2|Z = 1) = -0.1$.
- Conditional independence implies independence.

Week 3

Summary with Quizzes

Right or wrong

- X,Y are conditional independent given $Z = 1$, then X,Y are conditional independent given Z. (Not correct, the converse is correct though.)
- $P(X + Y = 2|Z = 1) = -0.1$. ([0,1])
- Conditional independence implies independence. (Not correct, try to find a counter example)

Week 4

Week 4

IV. Layman's talk: what is expectation

- Distributions assign (probability values) numbers to random events associated with random variables.

  Many values associated with many events...

- Sometimes, one would prefer to know what happens "on average".

  E.g., if heads for 1 bottoms for 0, we know from our daily experience, the average would be 0.5.

- **Expectation** formalizes the above idea and assigns a number (if exists) for one random variable.

Week 4

A. Definition of expectation of discrete random variables

Definition (Expectation of discrete r.v.'s)

The expectation/expected value/first moment/mean/average of a **discrete** random variable X for $x_1, x_2, x_3, \ldots$ (if exists) is defined by

$$E[X] = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i) \quad \left(= \sum_{i=1}^{\infty} x_i \Delta F_X(x_i)\right)$$

(we order all values from the support of X by the increasing order: $x_1 \leq x_2 \leq x_3, \cdots$.
We may choose $\Delta F_X(x_i) = F_X(x_i) - F_X(x_{i-1})$, let $x_0 < x_1, x_0 \notin supp(X)$ and thus $F(x_0) = 0$.
We may choose $\Delta F_X(x) = F(x) - \lim_{y \uparrow x} F(y)$, as another possible expression.
Essentially both expressions give: $\Delta F_X(x_i) = P(X = x_i)$.)

If $|\mathrm{supp}(X)|$ is finite with elements $x_1, x_2, x_3, \ldots, x_n$, then

$$E[X] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i) \quad \left(= \sum_{i=1}^{n} x_i \Delta F_X(x_i)\right)$$

- Expectation may not exist for some random variables, e.g., the summation $\sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i)$ may diverge.
- Same distributions, same expectations (if exist).

Week 4

B. Properties of expectation

Linearity. Assume all expectations are well defined.
For any **real-valued** r.v.'s X, Y and any constant c,

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$
$$\mathbb{E}(cX) = c\mathbb{E}(X)$$

Monotonicity. Assume all expectations are well defined.

For any **real-valued** r.v.'s X, Y such that $X \geq Y$ with probability one ($P(X \geq Y) = 1$), then

$$\mathbb{E}(X) \geq \mathbb{E}(Y)$$

with equality holding iif $\mathbb{P}(X = Y) = 1$.

Week 4

B. Properties of expectation

Proof I linearity with discrete r.v.'s:

$$\mathbb{E}(X+Y) = \sum_{c} cP(X+Y=c) = \sum_{c} c\left(\sum_{a\in \mathrm{supp}(X)} P(X=a, Y=c-a)\right)$$

$$= \sum_{a\in \mathrm{supp}(X)} \sum_{c} cP(X=a, Y=c-a) = \sum_{a\in \mathrm{supp}(X)} \sum_{c} ((a)+(c-a))P(X=a, Y=c-a)$$

$$= \sum_{a\in \mathrm{supp}(X)} \sum_{b\in \mathrm{supp}(Y)} (a+b)P(X=a, Y=b)$$

$$= \sum_{a\in \mathrm{supp}(X)} \sum_{b\in \mathrm{supp}(Y)} aP(X=a, Y=b) + \sum_{a\in \mathrm{supp}(X)} \sum_{b\in \mathrm{supp}(Y)} bP(X=a, Y=b)$$

$$= \sum_{a\in \mathrm{supp}(X)} a \sum_{b\in \mathrm{supp}(Y)} P(X=a, Y=b) + \sum_{b\in \mathrm{supp}(Y)} b \sum_{a\in \mathrm{supp}(X)} P(X=a, Y=b)$$

$$= \sum_{a\in \mathrm{supp}(X)} aP(X=a) + \sum_{b\in \mathrm{supp}(Y)} bP(Y=b)$$

$$= E(X) + E(Y)$$

Week 4

B. Properties of expectation

Proof II linearity with discrete r.v.'s and finite outcomes $|S| < \infty$:

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_c cP(X+Y=c) \\
&= \sum_c c \sum_{s\in S:(X+Y)(s)=c} P(\{s\}) \\
&= \sum_c \sum_{s\in S:(X+Y)(s)=c} cP(\{s\}) \\
&= \sum_c \sum_{s\in S:(X+Y)(s)=c} (X(s)+Y(s))P(\{s\}) \\
&= \sum_{s\in S} (X(s)+Y(s))P(\{s\}) \\
&= \left(\sum_{s\in S} X(s)P(\{s\})\right) + \left(\sum_{s\in S} Y(s)P(\{s\})\right) \\
&= \left(\sum_{a\in \mathrm{supp}(X)} \sum_{s\in S:X(s)=a} X(s)P(\{s\})\right) + \left(\sum_{b\in \mathrm{supp}(Y)} \sum_{s\in S:Y(s)=b} Y(s)P(\{s\})\right) \\
&= \left(\sum_{a\in \mathrm{supp}(X)} aP(X=a)\right) + \left(\sum_{b\in \mathrm{supp}(Y)} bP(Y=b)\right) \\
&= E(X)+E(Y)
\end{aligned}$$

Week 4

B. Properties of expectation

Proof monotonicity with discrete r.v.'s: Note that $Z = X - Y$ would be non-negative with probability one, such that $Z(s) \geq 0$ for all $s \in S$, and thus $E(Z)$ is a weighted sum of non-negative values (≥ 0), and by linearity

$$E(X) - E(Y) = E(Z) \geq 0.$$

If $E(X) = E(Y)$, then $E(Z) = \sum_{z_i \in \mathrm{supp}(Z)} z_i P(Z = z_i) = 0$. Since $z_i \geq 0, P(Z = z_i) > 0$, we know $z_i = 0$ (otherwise $E(Z) > 0$).

Week 4

C. Expectations of functions of random variables I: LOTUS and variane

Distributions determine expectations.

The **distribution** of X determines the **distribution** of $g(X)$.

LOTUS:
The **distribution** of X determines the **expectation** of $g(X)$.

If X is a discrete random variable with support $\text{supp}(X) = \{x_1, x_2, x_3. \cdots\}$, and $g(x)$ is a function from $R \to R$ such that $g(X)$ is a discrete r.v., then

$$\text{E}[g(X)] = \sum_{x \in \text{supp}(X)} g(x) P(X = x) \quad (= \sum_{x \in \text{supp}(X)} g(x) \Delta F_X(x))$$

with $\Delta F_X(x) = F(x) - \lim_{y \uparrow x} F(y)$.

Week 4

C. Expectations of functions of random variables I: LOTUS and variane

Proof with discrete r.v.'s:

$$E(g(X)) = \sum_{c} cP(g(X) = c)$$

$$= \sum_{c} c \left(\sum_{a \in \text{supp}(X): g(a) = c} P(X = a) \right)$$

$$= \sum_{c} \sum_{a \in \text{supp}(X): g(a) = c} cP(X = a)$$

$$= \sum_{c} \sum_{a \in \text{supp}(X): g(a) = c} g(a)P(X = a)$$

$$= \sum_{a \in \text{supp}(X)} g(a)P(X = a)$$

Week 4

C. Expectations of functions of random variables I: LOTUS and variane

- The **variance** of an **real-valued** r.v. X (if exists) is

$$\mathrm{Var}(X) = E(X - EX)^2$$

and the **standard deviation** is

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

- Variance of X is essentially the expectation of a function of X, $g(X)$, with $g(x) = (x - \mu)^2, \mu = E(X)$.

- The distribution of X determines the variance of X.

Week 4

C. Expectations of functions of random variables I: LOTUS and variane

Some properties of variance:

- For scalars $a, c \in \mathbb{R}$, for **any** r.v. X (if variance well-defined):

$$\mathrm{Var}(a + cX) = \mathrm{Var}(cX) = c^2\mathrm{Var}(X).$$

- For **independent** r.v.'s, $X_1, X_2, \cdots, X_n$ (if exists):

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right).$$

Different from the expectation, as for expectation for **any** r.v.'s $X_1, X_2, \cdots, X_n$ (if exists):

$$E(a + cX_1) = a + cE(X_1).$$

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

Week 4

D. Expectations of functions of random variables II: indicator function

$$I_A : S \mapsto \{0, 1\}; A \subset S.$$

- $$I_A(s) = \begin{cases} 1, & s \in A \\ 0, & s \in A^c (= S \setminus A) \end{cases}$$

- $(I_A)^k = I_A$ for any $k \in \mathbb{N}_+$.
- $I_{A^c} = 1 - I_A$.
- $I_{A \cap B} = I_A I_B$.
- $I_{A \cup B} = I_A + I_B - I_A I_B \leq I_A + I_B$.

  $I_{A \cup B} = 1 - I_{(A \cup B)^c} = 1 - I_{A^c \cap B^c} = 1 - I_{A^c} I_{B^c} = 1 - (1 - I_A)(1 - I_B) = I_A + I_B - I_A I_B$.
  Similar to inclusion-exclusion principle with two events
  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Week 4

D. Expectations of functions of random variables II: indicator function

- Fundamental bridge:
  - Inclusion-exclusion principle
  - Let X be a **non-negative integer-valued** r.v. then if its expectation exists

$$\mathbb{E}X = \sum_{i=0}^{\infty} G_X(i)$$

    where G_X is the survival function of X such that $G_X(x) = 1 - F_X(x)$.
  - III. Indicators are a useful tool to simplify questions.

Week 4

E. Negative Binomial and Negative Hypergeometric

$X \sim$Bin(n, p) if X denotes the number of heads in n throws of a special coin (with probability p getting head in each flip).

Now we consider another case:

- If we keep throwing a special coin (with probability p getting head in each flip) until we get r heads, let X denote the number of tails/bottomes/failures.
- $X \sim$NBin(r, p).
- The PMF of X is

$$P(X = n) = \binom{n+r-1}{r-1} p^r (1-p)^n; n = 0, 1, \cdots .$$

- NBin$(1, p)$ is also called Geometric distribution Geom(p) (the number of bottoms before the first head).

Week 4

E. Negative Binomial and Negative Hypergeometric

$X \sim$HGeom(w, b, n). Mark n different balls with $*$'s from a urn of w white balls and b **black** *balls (each ball is equally likely to be marked), and X is the number of balls being both white and marked with $*$'s.*

Now we consider another case:

- Keep marking balls with $*$'s from a urn of w white balls and b **black** balls (each ball is equally likely to be marked) until there are r white balls with marks, and let X denote the number of balls being both **black** and marked with $*$'s.
- $X \sim$NHGeom(w, b, r).
- The PMF of X is

$$P(X=k) = \frac{\binom{w}{r-1}\binom{b}{k}}{\binom{w+b}{r+k-1}} \frac{w-r+1}{w+b-r-k+1}; k = 0, 1, \cdots$$

Week 4

F. Poisson

The probability of more than one hit in a very short interval is negligible.
The probability of a single arrival in a very short interval is proportional to the length of the interval.

- The arrival of one request within a very small interval $(s, s+\Delta t]$ follows $\mathrm{Bern}(\lambda \Delta t)$, denoted as $I_{(s,s+\Delta t]}$;
- The number of total requests within interval $(s, s+t]$

$$X_{\Delta t} = \sum_{i=1}^{t/(\Delta t)} I_{(s+(i-1)\Delta t, s+i\Delta t]},$$

where for simplicity we assume that $t/(\Delta t)$ is a positive integer.

The number of hits in non-overlapping time intervals are independent.

- $X_{\Delta t}$ follows $\mathrm{Bin}(t/\Delta t, \lambda \Delta t)$.
- $P(X_{\Delta t} = k) = \binom{t/(\Delta t)}{k} (\lambda \Delta t)^k (1 - \lambda \Delta t)^{t/(\Delta t) - k}$.
  Let's consider a special case such that $\Delta t = t/n$, what happens to the PMF of the random variable $X_{\Delta t}$ if we let $n \to \infty$ ($\Delta t \to 0$, we are splitting into smaller and smaller intervals).
  Denote the limit random variable as X.
- $P(X = k) = \lim_{n\to\infty} \binom{n}{k} (\lambda t/n)^k (1 - \lambda t/n)^{n-k} =$
  $\lim_{n\to\infty} (\lambda t)^k / k! \left(\prod_{i=1}^{k} (n-i+1)/n \right) (1 - (\lambda t)/n)^n (1 - (\lambda t)/n)^{-k} = e^{-\lambda t} (\lambda t)^k / k!$.
  We use the fact that $(1 + x/s)^s \to_{s\to\infty} e^x$.

Week 4

F. Poisson

Denote X the number of random requests within a time interval of **unit length** from the aforementioned scenario (t=1).

- The PMF of X is

$$P(X = k) = e^{-\lambda}\lambda^k/k!; k = 0, 1, \cdots .$$

where $\lambda > 0$.

- $X \sim$Pois(λ).
- $E(X) = V(X) = \lambda$.

Week 4

F. Poisson

A web server would receive requests from computer A randomly, but also requests from computer B. A and B are independent.

How to properly model the random variable Y,
the number of requests from either computer A or B within a time interval of unit lenghth.

- Simply sum their arrival rate, and all the rest follow the same logic.
- $A \sim \text{Pois}(\lambda_A)$, $B \sim \text{Pois}(\lambda_B)$ and A,B are independent. Then
$$Y = A + B \sim \text{Pois}(\lambda_A + \lambda_B)$$
- Verify the result by looking at the distribution (PMF).

$$P(A+B=k) = \sum_{i=0}^{k} P(A=k-i|B=i)P(B=i) = \sum_{i=0}^{k} P(A=k-i)P(B=i)$$
$$= \sum_{i=0}^{k} \frac{1}{(k-i)!}\lambda_A^{k-i}e^{-\lambda_A} \cdot \frac{1}{i!}\lambda_B^{i}e^{-\lambda_B} = e^{-(\lambda_A+\lambda_B)}\frac{1}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\lambda_A^{k-i}\lambda_B^{i}$$
$$= e^{-(\lambda_A+\lambda_B)}\frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda_A^{k-i}\lambda_B^{i} = e^{-(\lambda_A+\lambda_B)}\frac{1}{k!}(\lambda_A+\lambda_B)^k$$

- $(a+b)^k = (a+b)(a+b)\cdots(a+b) = \sum_{i=0}^{k}\binom{k}{i}a^{k-i}b^{i}$.

Week 4

F. Poisson

- **Poisson conditional on the sum is Binomials** .
- If a web server receive total n requests from either computer A or B.
- Given these two computers are independent,
  for each signal it could either be from A or B:
  $I_{iA} \sim \text{Bern}(\frac{\lambda_A}{\lambda_A+\lambda_B})$.
- If $A \sim \text{Pois}(\lambda_A)$, $B \sim \text{Pois}(\lambda_B)$, A is independent from B, then the number of requests from computer A given $A+B=n$:

$$A \text{ given } \{A+B=n\} \text{ follows } \text{Bin}(n, \frac{\lambda_A}{\lambda_A+\lambda_B})$$

- Verify the result by looking atthe distributio (PMF).

$$P\left(A=k|A+B=n\right)$$
$$=\frac{P\left(A=k,B=n-k\right)}{P(A+B=n)}=\frac{P\left(A=k\right)P\left(B=n-k\right)}{P(A+B=n)}$$
$$=\frac{\left(e^{-\lambda_A}\frac{\lambda_A{}^k}{k!}\right)\left(e^{-\lambda_B}\frac{\lambda_B{}^{(n-k)}}{(n-k)!}\right)}{\left(e^{-(\lambda_A+\lambda_B)}\frac{(\lambda_A+\lambda_B)^n}{n!}\right)}=\frac{n!}{k!(n-k)!}\left(\frac{\lambda_A}{\lambda_A+\lambda_B}\right)^k\left(\frac{\lambda_B}{\lambda_A+\lambda_B}\right)^{n-k}$$
$$=\binom{n}{k}\left(\frac{\lambda_A}{\lambda_A+\lambda_B}\right)^k\left(1-\frac{\lambda_A}{\lambda_A+\lambda_B}\right)^{n-k}.$$

Week 4

F. Poisson

Previously, The arrival of one request within a very small interval $(s, s+\Delta t]$ follows Bern$(\lambda \Delta t)$, denoted as $I_{(s,s+\Delta t]}$; and the number of total requests within interval $(s, s+t]$

$$X_{\Delta t} = \sum_{i=1}^{t/(\Delta t)} I_{(s+(i-1)\Delta t, s+i\Delta t]}.$$

The limit once we let Δt go to zero follows

$$\text{Pois}(\lambda), \lambda = \sum_{i=1}^{t/(\Delta t)} \lambda \Delta t.$$

Poisson paradigm Let $A_i, i = 1, \cdots, n$ be **independent (or weakly dependent)** events with probability p_i, **n is large and p_i are very small.** Let

$$X = \sum_{j=1}^{n} I_{A_j}$$

count how many of the A_i occur (how many j visit one specific island within one day for example). Then **X is approximated distributed** as

$$\text{Pois}(\lambda); \lambda = \sum_j p_j.$$

- λ as the *rate* (expected number within a certain time period) of occurrence of *rare* events.

Week 4

Summary with Quizzes

Poisson serves as a good approximation in practice

Week 4

Summary with Quizzes

Recorded number of deaths by a horse kick in the Prussian army (1875-1894)

Year	GC	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C14	C15
1875	0	0	0	0	0	0	0	1	1	0	0	0	1	0
1876	2	0	0	0	1	0	0	0	0	0	0	0	1	1
1877	2	0	0	0	0	0	1	1	0	0	1	0	2	0
1878	1	2	2	1	1	0	0	0	0	0	1	0	1	0
1879	0	0	0	1	1	2	2	0	1	0	0	2	1	0
1880	0	3	2	1	1	1	0	0	0	2	1	4	3	0
1881	1	0	0	2	1	0	0	1	0	1	0	0	0	0
1882	1	2	0	0	0	0	1	0	1	1	2	1	4	1
1883	0	0	1	2	0	1	2	1	0	1	0	3	0	0
1884	3	0	1	0	0	0	0	1	0	0	2	0	1	1
1885	0	0	0	0	0	0	1	0	0	2	0	1	0	1
1886	2	1	0	0	1	1	1	0	0	1	0	1	3	0
1887	1	1	2	1	0	0	3	2	1	1	0	1	2	0
1888	0	1	1	0	0	1	1	0	0	0	0	1	1	0
1889	0	0	1	1	0	1	1	0	0	1	2	2	0	2
1890	1	2	0	2	0	1	1	2	0	2	1	1	2	2
1891	0	0	0	1	1	1	0	1	1	0	3	3	1	0
1892	1	3	2	0	1	1	3	0	1	1	0	1	1	0
1893	0	1	0	0	0	1	0	2	0	0	1	3	0	0
1894	1	0	0	0	0	0	0	0	1	0	1	1	0	0

Week 4

Summary with Quizzes

Fitting the data with the Poisson p.m.f.

$$P(X = k) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Total number of observations: 280

$X = k$	Amount	Probability	Poisson ($\lambda = 0.7$)
$X = 0$	144	0.5143	0.4966
$X = 1$	91	0.3250	0.3476
$X = 2$	32	0.1143	0.1217
$X \geq 3$	13	0.0464	0.0341

Week 4

Summary with Quizzes

- Let X be the number of hits in a time interval of length t. Let λ be the arrival rate in an interval of length t.

  Then for $x = 0, 1, 2, \ldots,$ we have

$$P(X = x) = f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- Let Y be the number of hits in a time interval of length $a \cdot t$

Week 4

Summary with Quizzes

- Let X be the number of hits in a time interval of length t. Let λ be the arrival rate in an interval of length t.

  Then for $x = 0, 1, 2, \ldots,$ we have

$$P(X = x) = f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- Let Y be the number of hits in a time interval of length $a \cdot t$

$$P(Y = y) = f_Y(y) = \frac{(a\lambda)^y e^{-a\lambda}}{y!}$$

- Per 8 hour working day on average 8 people arrive at a customer service desk. What is probability that 2 arrive in the next hour?

Week 4

Summary with Quizzes

- Let X be the number of hits in a time interval of length t. Let λ be the arrival rate in an interval of length t.

  Then for $x = 0, 1, 2, \ldots$, we have

$$P(X = x) = f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- Let Y be the number of hits in a time interval of length $a \cdot t$

$$P(Y = y) = f_Y(y) = \frac{(a\lambda)^y e^{-a\lambda}}{y!}$$

- Per 8 hour working day on average 8 people arrive at a customer service desk. What is probability that 2 arrive in the next hour?
- Denote Z the number of arrivals in the next hour:

$$P(Z = 2) = \frac{(\frac{1}{8} \cdot 8)^2 e^{-\frac{1}{8} \cdot 8}}{2!} = 0.0948$$

Week 4

Summary with Quizzes

Completely right or possibly wrong

- The sum of Poisson distributed r.v.'s is still Poisson distributed r.v.
- Variance also has the linearity property.
- Variance of a given random variable is simply a non-negative number if exists.
- Expectation of an indicator r.v. could be negative.

Week 4

Summary with Quizzes

Completely right or possibly wrong

- The sum of Poisson distributed r.v.'s is still Poisson distributed r.v. (*independence is necessary, possibly wrong*)
- Variance also has the linearity property. (*wrong*)
- Variance of a given random variable is simply a non-negative number if exists. (Correct.)
- Expectation of an indicator r.v. could be negative. (*Wrong, recall the fundamental bridge, plus average of a 0-1 r.v. can not be negative.*)

Week 5

Week 5

- Assignment 1 feedbck:
  - Use proper notations: when using X, explain what X represents.

Week 5

- Assignment 1 feedbck:
  - Use proper notations: when using X, explain what X represents.
    One bad example from myself:
    4.3.a By definition of expectation of discrete r.v.'s: $E(X_i) = \sum_{i=1}^{6} \frac{i}{6} = 3.5$. What should I add?

Week 5

- Assignment 1 feedbck:
    - Always motivate your answers, only writing down a number/conclusion is not of full marks (actually, may lose most points).

Week 5

- Assignment 1 feedbck:
  - Always motivate your answers, only writing down a number/conclusion is not of full marks (actually, may lose most points).
    E.g., 2.I.b, why is it beneficial to have one additional card?

Week 5

- Assignment 1 feedbck:
    - Link with what we have learned.

Week 5

- Assignment 1 feedbck:
    - Link with what we have learned.
      E.g., 2.II.b, inclusion-exclusion principle. Draw Venn diagram to get some ideas.

Week 5

- Assignment 1 feedbck:
    - Solutions are not unique.

Week 5

- Assignment 1 feedbck:
  - Learning the probability theory is a challenging task. It is a different way of thinking though all concepts root from things you have learned from math I: functions.

Week 5

- Assignment 1 feedbck:

  - Learning the probability theory is a challenging task. It is a different way of thinking though all concepts root from things you have learned from math I: functions. Most time we are working with functions/function values of the pre-image/event generated by another functions, e.g., $P(X \in C)$'s (distributions!), which are like a *Matryoshka* of multi-layers.

Week 5

- Assignment 1 feedbck:
  - Use proper notations: when using X, explain what X represents.
    One bad example from myself:
    4.3.a By definition of expectation of discrete r.v.'s: $E(X_i) = \sum_{i=1}^{6} \frac{i}{6} = 3.5$. What should I add?
  - Always motivate your answers, only writing down a number/conclusion is not of full marks (actually, may lose most points).
    E.g., 2.I.b, why is it beneficial to have one additional card?
  - Link with what we have learned.
    E.g., 2.II.b, inclusion-exclusion principle. Draw Venn diagram to get some ideas.
  - Solutions are not unique.
  - Learning the probability theory is a challenging task. It is a different way of thinking though all concepts root from things you have learned from math I: functions.
    Most time we are working with functions/function values of the pre-image/event generated by another functions, e.g., $P(X \in C)$'s (distributions!), which are like a *Matryoshka* of multi-layers.

Week 5

Comments on week 5 materials:

- We have see a lot of different **discrete** r.v.'s in previous weeks:

Week 5

Comments on week 5 materials:

- We have see a lot of different **discrete** r.v.'s in previous weeks:
  If X can take values of all non-negative integers, it could be Geom or Pois (or other distributions not covered in this course). How do we know what X follows?

Week 5

Comments on week 5 materials:

- We have see a lot of different **discrete** r.v.'s in previous weeks:
  If X can take values of all non-negative integers, it could be Geom or Pois (or other distributions not covered in this course). How do we know what X follows?
  Check distributions! E.g., look at the CDF/PMF.

Week 5

Comments on week 5 materials:

- We have see a lot of different **discrete** r.v.'s in previous weeks:
  If X can take values of all non-negative integers, it could be Geom or Pois (or other distributions not covered in this course). How do we know what X follows?
  Check distributions! E.g., look at the CDF/PMF.
- We see one new type of random variables: the continuous random variables, some of which can take any values from the whole real line instead of a list of numbers that can be indexed by natural numbers.

Week 5

Comments on week 5 materials:

- We have see a lot of different **discrete** r.v.'s in previous weeks:
  If X can take values of all non-negative integers, it could be Geom or Pois (or other distributions not covered in this course). How do we know what X follows?
  Check distributions! E.g., look at the CDF/PMF.
- We see one new type of random variables: the continuous random variables, some of which can take any values from the whole real line instead of a list of numbers that can be indexed by natural numbers.
- How do we classify different continuous random varaibles?

Week 5

Comments on week 5 materials:

- We have see a lot of different **discrete** r.v.'s in previous weeks:
  If X can take values of all non-negative integers, it could be Geom or Pois (or other distributions not covered in this course). How do we know what X follows?
  Check distributions! E.g., look at the CDF/PMF.
- We see one new type of random variables: the continuous random variables, some of which can take any values from the whole real line instead of a list of numbers that can be indexed by natural numbers.
- How do we classify different continuous random varaibles?
  Again, we can look at their distriutions: CDF/PDF.

Week 5

Comments on week 5 materials:

- We have see a lot of different **discrete** r.v.'s in previous weeks:
  If X can take values of all non-negative integers, it could be Geom or Pois (or other distributions not covered in this course). How do we know what X follows?
  Check distributions! E.g., look at the CDF/PMF.
- We see one new type of random variables: the continuous random variables, some of which can take any value from the whole real line instead of a list of numbers that can be indexed by natural numbers.
- How do we classify different continuous random varaibles?
  Again, we can look at their distriutions: CDF/PDF.

Week 5

V. Layman's talk: what is continuous random variable

- By throwing a 10-sided fair die infinite (countably many) times, we get Unif[0,1]: 0.1234565346562341 7545....!
-
  - Let's cut $[0,1]$ into 10^n grids such that each grid can be expressed as

    $$I_i = [x_{i-1,n}, x_{i,n}], i \in \{1, \cdots, 10^n\}$$

    with $x_j = j10^{-n}, j \in \{0, 1, \cdots, 10^n\}$. E.g., $i = 2$, $[10^{-n}, 2 * 10^{-n}]$.
    It is easy to check that for each grids $P(X \in I_j) = 10^{-n}$.
    Let's approximate the probability value of $\{X \leq a\}$ with

    $$\sum_{i=1}^{j_a} \Delta x_i, \Delta x_i = x_i - x_{i-1}, j_a = \min\{i \in \{1, \cdots, 10^n\} : a \in I_i\}.$$

  - $\left| P(X \leq a) - \sum_{i=1}^{j_a} \Delta x_i \right| \leq 10^{-n}$.
  - $\sum_{i=1}^{j_a} \Delta x_i \to \int_0^a 1dx = a$. If we let n go to infinity, then the difference goes to zero. Then we recover the value $P(X \leq a)$.

Week 5

V. Layman's talk: what is continuous random variable

We have a probability density function f_X (integrate the (probability) density via Riemann integral, we have the (probability) value). If an interval has positive density function values, then there are uncountably many possible values in the interval taken by X.

$$\int_{-\infty}^{a} f_X(x)dx = P(X \leq a); a \in \mathbb{R}.$$

Any real-valued random variable that has a non-negative function f_X satisfying the above Riemann integral condition is a continuous real-valued random variable.

Let's compare.

- In the example, we see that we can choose $f_X(x) = F'_X(x)$ and $F'_X(x) = 1, x \in [0, 1]$.

  There are **uncountably infinite** possible values in [0,1] taken by X.

- For any discrete random variable Y, almost everywhere F'_Y is simply **zero**, there is no f_X satisfying the above identity (via Riemann integral).

  There are **at most countably many** possible values taken by Y.

Week 5

A. Continuous r.v.'s v.s. Discrete r.v.'s

- A **real-valued** random variable X is a **discrete** random variable if there exists a set C with at most countably many numbers (so you can index all the elements within C using a set of natural numbers):

$$P(\{X \in C\}) = 1, \text{ or equivalently, } P(X \in C) = 1.$$

The **support** of a **discrete real-valued** random variable is the set of all numbers, x's, such that $P(X = x) > 0$ **(PMF)**:

$$supp(X) = \{x \in C : P(X = x) > 0\}.$$

The support of a discrete random variable has **empty interior** (it contains no open intervals at all): e.g., $\{0, 1, 2, 3, \cdots\}$.

- A **real-valued** random variable X is a **continuous** random variable if there exists a non-negative function f_X such that:

$$P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(s)ds$$

If F_X is differentiable, we can choose $f_X(x) = F'_X(x)$.

The **support** of a **continuous** random variable with differentialble CDF F_X, conventionally, is the set of all numbers, x's, such that $f_X(x) = F'_X(x) > 0$ **(PDF)**:

$$supp(X) = \{x \in C : F'_X(x) > 0\}.$$

The support of a continuous random variable has **non-empty interior** (it contains open intervals): e.g., $(0, 1)$.

Week 5

A. Continuous r.v.'s v.s. Discrete r.v.'s

Definition (PDF of continuous real-valued r.v.)

The **probability density function (PDF)** of a **continuous** real-valued r.v. is a non-negative function f_X on the real line such that via the Riemann integral:

$$\int_{-\infty}^{x} f_X(s) ds = P(X \leq x).$$

For a **continuous** random variable with differentiable CDF F_X, conventionally, $f_X(x) = F_X'(x)$.

- f_X can be used to calculate the CDF function values, and thus f_X is also a way describing the distribution of **continuous real-valued** random variables.
  Properties:
- **Non-negativity**: $f_X(x) > 0$ if $x \in \text{Supp}(X)$; $f_X(x) = 0$ otherwise.
- **Integrate to 1.** $\int_{-\infty}^{\infty} f_X(x) dx = 1.$

Week 5

A. Continuous r.v.'s v.s. Discrete r.v.'s

Real-valued random variables

	CDF $\mathbb{P}(Z \leq z)$	PMF/PDF	Expectation $\mathbb{E}Z$
Discrete			
$R(X) = supp(X) =$	Step function $F(x)$	$p_X(x_i) = F(x_i) - F(x_{i-1})$	$\sum_{x_i \in R(x);\, x_i < x_{i+1}} x_i \Delta F(x_i)$
$\{x_1 < . < x_n < .\}$	increase from 0 to 1	$= F(x_i) - \lim\limits_{s \uparrow x_i} F(s)$	$\sum_{x_i \in R(x);\, x_i < x_{i+1}} x_i p_X(x_i)$
$\mathring{R}(X) = \emptyset$.	**Jumps**		
		$F(x) = \sum\limits_{x_i \in R(x);\, x_i \leq x} p_X(x_i)$	
		$p_X(x) \geq 0; \sum\limits_{x_i \in R(X)} p_X(x_i) = 1$	$(\sum\limits_{n=0}^{\infty} G_X(n))$
Continuous			
	Continuous $F(y)$	$f_Y(y) = F'(y)$	$\int_{R(Y)} y dF(y)$
$R(Y) = supp(Y)$		$= \lim\limits_{s \to y} \frac{F(y) - F(s)}{y - s}$	
$\mathring{R}(Y) \neq \emptyset$, e.g. (a, b)	increase from 0 to 1		$\int_{R(Y)} y f(y) dy$
	No jumps	$F(y) = \int_{s<y} f_Y(s) ds$	
	differentiable		
		$f_Y(y) \geq 0; \int_{R(Y)} f_Y(y) dy = 1$	$\left(\int_0^{\infty} G_Y(y) dy\right)$

Week 5

A. Continuous r.v.'s v.s. Discrete r.v.'s

- **I. uncountably many possible values, so it is of zero probability that you get a specific number. True for all continuous r.v.'s.**

- **II. X_i i.i.d. from a continuous distribution, then**

$$P(X_{a_1} < X_{a_2} < X_{a_3} < \cdots < X_{a_n}) = \frac{1}{n!}.$$

for any permutations $a_1, a_2, a_3, \cdots, a_n$ of $1, 2, 3, \cdots, n$.

Week 5

B. Some continuous random varaibles

Exponential.

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

and zero for $t \leq 0$.

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

  and zero for $t \leq 0$.
- CDF: $F_X(t) = 1 - e^{-\lambda t}$.

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

  and zero for $t \leq 0$.
- CDF: $F_X(t) = 1 - e^{-\lambda t}$.
- **Properties**:

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

  and zero for $t \leq 0$.
- CDF: $F_X(t) = 1 - e^{-\lambda t}$.
- **Properties**:

1. If X ∼Expo(1), then $Y = \frac{X}{\lambda}$ ∼Expo(λ).

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

and zero for $t \leq 0$.
- CDF: $F_X(t) = 1 - e^{-\lambda t}$.
- **Properties**:

1. If X ~Expo(1), then $Y = \frac{X}{\lambda}$ ~Expo(λ).
2. $EY = E\left(\frac{X}{\lambda}\right) = 1/\lambda, VY = V\left(\frac{X}{\lambda}\right) = 1/\lambda^2$.

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

and zero for $t \leq 0$.
- CDF: $F_X(t) = 1 - e^{-\lambda t}$.
- **Properties**:

1. If X ∼Expo(1), then $Y = \frac{X}{\lambda}$ ∼Expo(λ).
2. $EY = E\left(\frac{X}{\lambda}\right) = 1/\lambda, VY = V\left(\frac{X}{\lambda}\right) = 1/\lambda^2$.
3. Waiting time between the first and the second arrivals is also Expo.
   Suppose at time t there is one arrival, we can check the survival funcion (in order to derive the CDF $F_{T_2-T_1}(s) = P(T_2 - T_1 \leq s)$) $P(T_2 - T_1 > s | T_1 = t) =$ $P(N_{t+s} - N_t = 0 | N_t = 1, N_s = 0, s < t) = P(N_{t+s} - N_t = 0) = P(N_s = 0)$ (Hard, you will learn in the future again: the second last equality roots from the fact that arrival numbers are independent for disjoint time intervals (so conditioning on the past interval does not provide any extra info) while the last equality roots from the fact that $N_{t+s} - N_t$ is essential the number of arrivals within a time interval of length s).

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

and zero for $t \leq 0$.
- CDF: $F_X(t) = 1 - e^{-\lambda t}$.
- **Properties**:

1. If $X \sim$Expo(1), then $Y = \frac{X}{\lambda} \sim$Expo($\lambda$).
2. $EY = E\left(\frac{X}{\lambda}\right) = 1/\lambda, VY = V\left(\frac{X}{\lambda}\right) = 1/\lambda^2$.
3. Waiting time between the first and the second arrivals is also Expo.
Suppose at time t there is one arrival, we can check the survival funcion (in order to derive the CDF $F_{T_2-T_1}(s) = P(T_2 - T_1 \leq s)$) $P\left(T_2 - T_1 > s | T_1 = t\right) = P\left(N_{t+s} - N_t = 0 | N_t = 1, N_s = 0, s < t\right) = P\left(N_{t+s} - N_t = 0\right) = P\left(N_s = 0\right)$ (Hard, you will learn in the future again: the second last equality roots from the fact that arrival numbers are independent for disjoint time intervals (so conditioning on the past interval does not provide any extra info) while the last equality roots from the fact that $N_{t+s} - N_t$ is essential the number of arrivals within a time interval of length s).
4. Memoryless property: $P(X > t + s | X > t) = P(X > s)$. (Some second-hand items are as if they are new. They may last for the same life duration.)

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is
$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$
and zero for $t \leq 0$.
- CDF: $F_X(t) = 1 - e^{-\lambda t}$.
- **Properties**:

1. If X ~Expo(1), then $Y = \frac{X}{\lambda}$ ~Expo(λ).
2. $EY = E\left(\frac{X}{\lambda}\right) = 1/\lambda, VY = V\left(\frac{X}{\lambda}\right) = 1/\lambda^2$.
3. Waiting time between the first and the second arrivals is also Expo.
Suppose at time t there is one arrival, we can check the survival funcion (in order to derive the CDF $F_{T_2-T_1}(s) = P(T_2 - T_1 \leq s)$) $P\left(T_2 - T_1 > s | T_1 = t\right) =$
$P\left(N_{t+s} - N_t = 0 | N_t = 1, N_s = 0, s < t\right) = P\left(N_{t+s} - N_t = 0\right) = P\left(N_s = 0\right)$ (Hard, you will learn in the future again: the second last equality roots from the fact that arrival numbers are independent for disjoint time intervals (so conditioning on the past interval does not provide any extra info) while the last equality roots from the fact that $N_{t+s} - N_t$ is essential the number of arrivals within a time interval of length s).
4. Memoryless property: $P(X > t + s | X > t) = P(X > s)$. (Some second-hand items are as if they are new. They may last for the same life duration.)
5. Minimum of independent Expo's is still Expo. Max is not.
Solutions are not unique, we can simply directly calculate it $P\left(\min_i T_i \leq t\right) =$
$1 - P\left(\min_i T_i > t\right) = 1 - P\left(T_1 > t, T_2 > t, \cdots, T_n > t\right) =_{independent} 1 - \prod_i P\left(T_i > t\right)$.

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

and zero for $t \leq 0$.
- CDF: $F_X(t) = 1 - e^{-\lambda t}$.
- **Properties**:

1. If $X \sim$Expo(1), then $Y = \frac{X}{\lambda} \sim$Expo($\lambda$).
2. $EY = E\left(\frac{X}{\lambda}\right) = 1/\lambda, VY = V\left(\frac{X}{\lambda}\right) = 1/\lambda^2$.
3. Waiting time between the first and the second arrivals is also Expo.
   Suppose at time t there is one arrival, we can check the survival funcion (in order to derive the CDF $F_{T_2-T_1}(s) = P(T_2 - T_1 \leq s)$) $P\left(T_2 - T_1 > s | T_1 = t\right) = P\left(N_{t+s} - N_t = 0 | N_t = 1, N_s = 0, s < t\right) = P\left(N_{t+s} - N_t = 0\right) = P\left(N_s = 0\right)$ (Hard, you will learn in the future again: the second last equality roots from the fact that arrival numbers are independent for disjoint time intervals (so conditioning on the past interval does not provide any extra info) while the last equality roots from the fact that $N_{t+s} - N_t$ is essential the number of arrivals within a time interval of length s).
4. Memoryless property: $P(X > t + s | X > t) = P(X > s)$. (Some second-hand items are as if they are new. They may last for the same life duration.)
5. Minimum of independent Expo's is still Expo. Max is not.
   Solutions are not unique, we can simply directly calculate it $P\left(\min_i T_i \leq t\right) = 1 - P\left(\min_i T_i > t\right) = 1 - P\left(T_1 > t, T_2 > t, \cdots, T_n > t\right) =_{independent} 1 - \prod_i P\left(T_i > t\right)$.
6. From Exponential to Poisson: **The waiting time between two Poisson process arrivals follows independent exponential distributions, while if the waiting time between two successive arrivals follows i.i.d. exponential distributions then the number of arrivals follows a Poisson process.**

Week 5

B. Some continuous random varaibles

Exponential.

- A continuous real-valued r.v. X follows Expo(λ) if its PDF $f_X(x)$ is

$$f_X(x) = \lambda e^{-\lambda t}, t > 0$$

and zero for $t \leq 0$.

- X takes values in $(0, +\infty)$. CDF: $F_X(t) = 1 - e^{-\lambda t}$.
- **Properties**:

1. If $X \sim$Expo(1), then $Y = \frac{X}{\lambda} \sim$Expo($\lambda$).
2. $EY = E\left(\frac{X}{\lambda}\right) = 1/\lambda, VY = V\left(\frac{X}{\lambda}\right) = 1/\lambda^2$.
3. Waiting time between the first and the second arrivals is also Expo.
   Suppose at time t there is one arrival, we can check the survival funcion (in order to derive the CDF $F_{T_2-T_1}(s) = P(T_2 - T_1 \leq s)$) $P\left(T_2 - T_1 > s | T_1 = t\right) = P\left(N_{t+s} - N_t = 0 | N_t = 1, N_s = 0, s < t\right) = P\left(N_{t+s} - N_t = 0\right) = P\left(N_s = 0\right)$ (Hard, you will learn in the future again: the second last equality roots from the fact that arrival numbers are independent for disjoint time intervals (so conditioning on the past interval does not provide any extra info) while the last equality roots from the fact that $N_{t+s} - N_t$ is essential the number of arrivals within a time interval of length s).
4. Memoryless property: $P(X > t + s | X > t) = P(X > s)$. (Some second-hand items are as if they are new. They may last for the same life duration.)
5. Minimum of independent Expo's is still Expo. Max is not.
   Solutions are not unique, we can simply directly calculate it $P\left(\min_i T_i \leq t\right) = 1 - P\left(\min_i T_i > t\right) = 1 - P\left(T_1 > t, T_2 > t, \cdots, T_n > t\right) =_{independent} 1 - \prod_i P\left(T_i > t\right)$.
6. From Exponential to Poisson: **The waiting time between two Poisson process arrivals follows independent exponential distributions, while if the waiting time between two successive arrivals follows i.i.d. exponential distributions then the number of arrivals follows a Poisson process.**

Week 5

B. Some continuous random varaibles

Uniform.

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif((a, b))), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif((a, b))), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

U takes values in (a,b). Probability is proportional to length (e.g., $(c, d) \subseteq (a, b)$):
$P(c \leq U \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF
$(x \in (a, b))$:$F_X(x) = P(a + (b-a)U \leq x) = P(U \leq (x-a)/(b-a)) = (x-a)/(b-a)$.

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif$((a, b))$), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

U takes values in (a,b). Probability is proportional to length (e.g., $(c, d) \subseteq (a, b)$): $P(c \leq U \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF $(x \in (a, b))$:$F_X(x) = P(a + (b-a)U \leq x) = P(U \leq (x-a)/(b-a)) = (x-a)/(b-a)$.

- **Properties**:

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif$((a, b))$), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

U takes values in (a,b). Probability is proportional to length (e.g., $(c, d) \subseteq (a, b)$): $P(c \leq U \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF $(x \in (a, b))$:$F_X(x) = P(a + (b-a)U \leq x) = P(U \leq (x-a)/(b-a)) = (x-a)/(b-a)$.

- **Properties**:

1. $Y \sim$Unif(a,b), $X \sim$Unif(0,1).
   $EY = E\left(a + (b-a)X\right) = a + \frac{b-a}{2} = \frac{a+b}{2}$, $VY = V\left(a + (b-a)X\right) = (b-a)^2/12$.

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif$((a,b))$), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{otherwise} \end{cases}$$

U takes values in (a,b). Probability is proportional to length (e.g., $(c,d) \subseteq (a,b)$): $P(c \le U \le d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF $(x \in (a,b))$:$F_X(x) = P(a+(b-a)U \le x) = P(U \le (x-a)/(b-a)) = (x-a)/(b-a)$.

- **Properties**:

1. $Y \sim$Unif(a,b), $X \sim$Unif(0,1).
   $EY = E\left(a+(b-a)X\right) = a + \frac{b-a}{2} = \frac{a+b}{2}$, $VY = V\left(a+(b-a)X\right) = (b-a)^2/12$.
2. Conditional on a unif is a unif. Let's draw the PDF.

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif((a, b))), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

  U takes values in (a,b). Probability is proportional to length (e.g., $(c, d) \subseteq (a, b)$): $P(c \leq U \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF $(x \in (a, b))$:$F_X(x) = P(a + (b-a)U \leq x) = P(U \leq (x-a)/(b-a)) = (x-a)/(b-a)$.
- **Properties**:

1. Y ~Unif(a,b), X ~Unif(0,1).
   $EY = E\left(a + (b-a)X\right) = a + \frac{b-a}{2} = \frac{a+b}{2}$, $VY = V\left(a + (b-a)X\right) = (b-a)^2/12$.
2. Conditional on a unif is a unif. Let's draw the PDF.
3. Universality of the Unif.

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif((a,b))), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{otherwise} \end{cases}$$

U takes values in (a,b). Probability is proportional to length (e.g., $(c,d) \subseteq (a,b)$): $P(c \leq U \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF ($x \in (a,b)$):$F_X(x) = P(a + (b-a)U \leq x) = P(U \leq (x-a)/(b-a)) = (x-a)/(b-a)$.

- **Properties**:

1. $Y \sim$Unif(a,b), $X \sim$Unif(0,1).
   $EY = E\left(a + (b-a)X\right) = a + \frac{b-a}{2} = \frac{a+b}{2}$, $VY = V\left(a + (b-a)X\right) = (b-a)^2/12$.
2. Conditional on a unif is a unif. Let's draw the PDF.
3. Universality of the Unif.

Let F be a strictly increasing (so F^{-1} the inverse function is well defined) CDF of a continuous r.v.:

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif((a, b))), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

U takes values in (a,b). Probability is proportional to length (e.g., $(c, d) \subseteq (a, b)$): $P(c \leq U \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF ($x \in (a, b)$):$F_X(x) = P(a + (b-a)U \leq x) = P(U \leq (x-a)/(b-a)) = (x-a)/(b-a)$.

- **Properties**:

1. Y $\sim$Unif(a,b), X $\sim$Unif(0,1).
   $EY = E\left(a + (b-a)X\right) = a + \frac{b-a}{2} = \frac{a+b}{2}$, $VY = V\left(a + (b-a)X\right) = (b-a)^2/12$.
2. Conditional on a unif is a unif. Let's draw the PDF.
3. Universality of the Unif.

   Let F be a strictly increasing (so F^{-1} the inverse function is well defined) CDF of a continuous r.v.:

   - **Let U $\sim$Unif(0,1), $F^{-1}(U)$ is an r.v. with CDF F.**

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif((a, b))), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

U takes values in (a,b). Probability is proportional to length (e.g., $(c, d) \subseteq (a, b)$): $P(c \leq U \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF $(x \in (a, b))$:$F_X(x) = P(a + (b-a)U \leq x) = P(U \leq (x-a)/(b-a)) = (x-a)/(b-a)$.

- **Properties**:

1. Y $\sim$Unif(a,b), X $\sim$Unif(0,1).
   $EY = E\left(a + (b-a)X\right) = a + \frac{b-a}{2} = \frac{a+b}{2}$, $VY = V\left(a + (b-a)X\right) = (b-a)^2/12$.
2. Conditional on a unif is a unif. Let's draw the PDF.
3. Universality of the Unif.

Let F be a strictly increasing (so F^{-1} the inverse function is well defined) CDF of a continuous r.v.:

- **Let U $\sim$Unif(0,1), $F^{-1}(U)$ is an r.v. with CDF F.**
- **Let X be an r.v. with CDF F, then $F(X) \sim Unif(0, 1)$.**

Week 5

B. Some continuous random varaibles

Uniform.

- A continuous real-valued r.v. U is said to have the uniforma distribution on (a, b) (or follows Unif((a, b))), if its PDF is

$$f_U(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

U takes values in (a,b). Probability is proportional to length (e.g., $(c, d) \subseteq (a, b)$): $P(c \leq U \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{c-d}{b-a}$. CDF ($x \in (a, b)$):$F_X(x) = P(a + (b-a)U \leq x) = P(U \leq (x-a)/(b-a)) = (x-a)/(b-a)$.

- **Properties**:

1. Y $\sim$Unif(a,b), X $\sim$Unif(0,1).
   $EY = E\left(a + (b-a)X\right) = a + \frac{b-a}{2} = \frac{a+b}{2}$, $VY = V\left(a + (b-a)X\right) = (b-a)^2/12$.
2. Conditional on a unif is a unif. Let's draw the PDF.
3. Universality of the Unif.

Let F be a strictly increasing (so F^{-1} the inverse function is well defined) CDF of a continuous r.v.:

- **Let U $\sim$Unif(0,1), $F^{-1}(U)$ is an r.v. with CDF F.**
- **Let X be an r.v. with CDF F, then $F(X) \sim Unif(0, 1)$.**

Week 5

B. Some continuous random varaibles

Normal/Gaussian.

Week 5

B. Some continuous random varaibles

Normal/Gaussian.

- A continuous r.v. Z is said to have the *standard normal distribution* $N(0,1)$ (mean zero and varaince one) if its PDF ψ is given by

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

with corresponding CDF $\Psi(z) = \int_{-\infty}^{z} \psi(z)dz$.

Week 5

B. Some continuous random varaibles

Normal/Gaussian.

- A continuous r.v. Z is said to have the *standard normal distribution* $N(0,1)$ (mean zero and varaince one) if its PDF ψ is given by

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

with corresponding CDF $\Psi(z) = \int_{-\infty}^{z} \psi(z) dz$.

- As for $X = \mu + \sigma Z$ $(\sigma > 0)$, it is said to have the *normal distribution* $N(\mu, \sigma^2)$ with mean μ and variance σ^2.

Week 5

B. Some continuous random varaibles

Normal/Gaussian.

- A continuous r.v. Z is said to have the *standard normal distribution* $N(0,1)$ (mean zero and varaince one) if its PDF ψ is given by

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

with corresponding CDF $\Psi(z) = \int_{-\infty}^{z} \psi(z)dz$.

- As for $X = \mu + \sigma Z$ $(\sigma > 0)$, it is said to have the *normal distribution* $N(\mu, \sigma^2)$ with mean μ and variance σ^2.

  The PDF of X:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Week 5

B. Some continuous random varaibles

Normal/Gaussian.

- A continuous r.v. Z is said to have the *standard normal distribution* $N(0,1)$ (mean zero and varaince one) if its PDF ψ is given by

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

with corresponding CDF $\Psi(z) = \int_{-\infty}^{z} \psi(z)dz$.

- As for $X = \mu + \sigma Z$ ($\sigma > 0$), it is said to have the *normal distribution* $N(\mu, \sigma^2)$ with mean μ and variance σ^2.

  The PDF of X:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **Properties**:

Week 5

B. Some continuous random varaibles

Normal/Gaussian.

- A continuous r.v. Z is said to have the *standard normal distribution* $N(0, 1)$ (mean zero and varaince one) if its PDF ψ is given by

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

with corresponding CDF $\Psi(z) = \int_{-\infty}^{z} \psi(z)dz$.

- As for $X = \mu + \sigma Z$ $(\sigma > 0)$, it is said to have the *normal distribution* $N(\mu, \sigma^2)$ with mean μ and variance σ^2.

  The PDF of X:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **Properties**:

1. Very special PDF here, bell-shape.

Week 5

B. Some continuous random varaibles

Normal/Gaussian.

- A continuous r.v. Z is said to have the *standard normal distribution* $N(0,1)$ (mean zero and varaince one) if its PDF ψ is given by

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

with corresponding CDF $\Psi(z) = \int_{-\infty}^{z} \psi(z) dz$.

- As for $X = \mu + \sigma Z$ ($\sigma > 0$), it is said to have the *normal distribution* $N(\mu, \sigma^2)$ with mean μ and variance σ^2.

  The PDF of X:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **Properties**:

1. Very special PDF here, bell-shape.
2. Symmetry of PDF ($\psi(z) = \psi(-z)$); of CDF ($\Psi(z) = 1 - \Psi(-z)$); of Z and -Z.

Week 5

B. Some continuous random varaibles

Normal/Gaussian.

- A continuous r.v. Z is said to have the *standard normal distribution* $N(0,1)$ (mean zero and varaince one) if its PDF ψ is given by

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

with corresponding CDF $\Psi(z) = \int_{-\infty}^{z} \psi(z) dz$.

- As for $X = \mu + \sigma Z$ $(\sigma > 0)$, it is said to have the *normal distribution* $N(\mu, \sigma^2)$ with mean μ and variance σ^2.

  The PDF of X:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **Properties**:

1. Very special PDF here, bell-shape.
2. Symmetry of PDF ($\psi(z) = \psi(-z)$); of CDF ($\Psi(z) = 1 - \Psi(-z)$); of Z and -Z.
3. $EX = E(\mu + \sigma Z) = \mu$, $VX = V(\mu + \sigma Z) = \sigma^2$

Week 5

B. Some continuous random varaibles

Normal/Gaussian. **Most important continuous distribution!**

- A continuous r.v. Z is said to have the *standard normal distribution* $N(0, 1)$ (mean zero and varaince one) if its PDF ψ is given by

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

with corresponding CDF $\Psi(z) = \int_{-\infty}^{z} \psi(z)dz$.

- As for $X = \mu + \sigma Z$ ($\sigma > 0$), it is said to have the *normal distribution* $N(\mu, \sigma^2)$ with mean μ and variance σ^2.

  The PDF of X:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **Properties**:

1. Very special PDF here, bell-shape.
2. Symmetry of PDF ($\psi(z) = \psi(-z)$); of CDF ($\Psi(z) = 1 - \Psi(-z)$); of Z and -Z.
3. $EX = E(\mu + \sigma Z) = \mu$, $VX = V(\mu + \sigma Z) = \sigma^2$

Week 5

Summary with Quizzes

Can you show that $\max_{i=1}^{2} T_i$ where T_i follows i.i.d. Expo(1) is not exponentially distributed?

Week 5

Summary with Quizzes

- Look at the CDF of $\max_{i=1}^{2} T_i$ for $t > 0$.
- $$\begin{aligned} P\left(\max_{i=1}^{2} T_i \leq t\right) &= P\left(T_1 \leq t, T_2 \leq t\right) \\ &= P\left(T_1 \leq t\right) P\left(T_2 \leq t\right) = \left(1 - e^{-t}\right)^2 \end{aligned}$$
- Its CDF can not be written as the exponential distribution CDF format.

Week 5

Summary with Quizzes

Completely right or possibly wrong

- The waiting time till the second Poisson process arrivals follows Exponential distribution.
- The sum of i.i.d. exponentially distributied r.v.'s does not follow exponential distribution.
- Z and $-Z$ have identical distributions if $Z \sim N(0, 1)$.
- EX and $E(-X)$ could be the same number in some cases.

Week 5

Summary with Quizzes

Completely right or possibly wrong

- The waiting time till the second Poisson process arrivals follows Exponential distribution.
  Wrong. The waiting time between successive Poisson process arrivals follows Exponential distribution. E.g., between the zero and the first arrivals, second and third arrivals, third and forth arrivals, $\cdots$
- The sum of i.i.d. exponentially distributied r.v.'s does not follow exponential distribution. Correct.
- Z and $-Z$ have identical distributions if $Z \sim N(0,1)$. Correct
- EX and $E(-X)$ could be the same number in some cases. Correct.

Week 6

Week 6

VI. Layman's talk: what is MGF

$$M_X(t) = \sum_{i=0}^{\infty} E(X^i)t^i/i! = E\sum_{i=0}^{\infty}(X^i)t^i/i! = Ee^{tX}.$$

Week 6

A. Moments and sample moments

For a real-valued r.v. X (assume the following expectation values exist):

- ***n*th Moment**
$$\mathbb{E}X^n.$$
- ***n*th Central Moment**
$$\mathbb{E}(X-\mu)^n.$$
- ***n*th Standardized Moment**
$$\mathbb{E}((X-\mu)/\sigma)^n.$$
- **Distributions determine moments** (moments are numbers, or you may regard them as parameters of fixed numbers for given random varaibles) since they are defined based on the concept: expectation.

Week 6

A. Moments and sample moments

Let $X_i, i = 1, \cdots, n$ be i.i.d. r.v.'s;

The kth sample moment:

$$M_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

.

- E.g., the sample mean is M_1 (the fist sample moment)

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- **Sample moments serve as approximations (observed from realized data) to the moments (theoretical values)**.

Week 6

A. Moments and sample moments

Sample moments serve as approximations (observed from realized data) to the moments (theoretical values).

Let $X_i, i = 1, \cdots, n$ be i.i.d. r.v.'s with μ, σ^2.

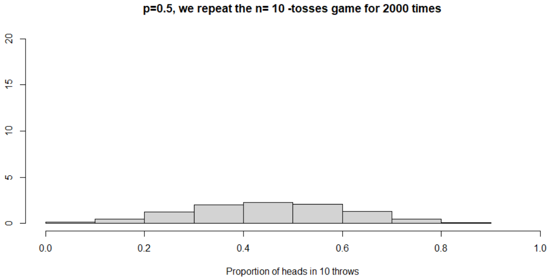
- **Sample mean** $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

  By the linearity of expectation, $\mathbb{E}\bar{X}_n = \mu$;

  due to the independence, $\text{Var}\bar{X}_n = \sigma^2/n \to 0$. (The only random variable type with zero variacne is the degenerated random varaible: constants.)
- $X_i^k, i = 1, \cdots, n$ **are also i.i.d. r.v.'s for any non-negative integer** k**,** so the above results also hold true if X_i^k has first and second moments.
- **Sample variance** $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$
- From $\sum_i^n (X_i - c)^2 = \sum_i^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - c)^2$, we know $\mathbb{E}S_n^2 = \sigma^2$.

Week 6

A. Moments and sample moments

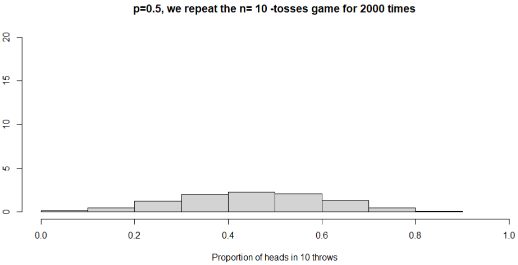


For each time, we throw a fair coin for n tosses, and calculate one number: the proportion of heads (one realisation of the sample average), we repeat this for 2000 times. Then we have 2000 numbers, we draw the histogram (density) plot of these 2000 numbers.

As n increases, we see the realised sample averages (random realisations) are centering towards $p = 1/2$.

Week 6

A. Moments and sample moments



For each time, we throw a fair coin for n tosses, and calculate one number: the proportion of heads (one realisation of the sample average), we repeat this for 2000 times. Then we have 2000 numbers, we draw the histogram (density) plot of these 2000 numbers.

As n increases, we see the realised sample averages (random realisations) are centering towards $p = 1/2$.

Week 6

B. MGF

Moment generating function (MGF)

$$M_X(t) = Ee^{tX}.$$

We say it is well-defined if we can find $a > 0$ **such that**

$$M_X(t) : (-a, a) \mapsto \mathbb{R}.$$

Namely, $M_X(t)$ **is only well defined if the summation below is finite for any** t **from some given open interval containing 0:**

$$\sum_{i=0}^{\infty} E(X^i)t^i/i! = E\sum_{i=0}^{\infty}(X^i)t^i/i! = Ee^{tX}.$$

- **Example** (MGF for $X \sim$Expo(1)): $M_X(t) = Ee^{tX} = \int_0^{\infty} e^{tx}e^{-x}dx = \frac{1}{1-t}, \forall t \neq 1$, $t = 0, M_X(t) = Ee^{tX} = 1$, **note that** $\frac{1}{1-t}$ **is finite for, e.g.,** $t \in (-0.5, 0.5)$ **so we have a well-defined MGF for** X.
- **Question:** If$M_Y(t) = M_X(t), \forall t$ in some open interval containing zero, what do we know about $(\mathbb{E}(X^n))$ and $(\mathbb{E}(Y^n))$?

Week 6

B. MGF

How do we judge two MGFs are equal?

- We know for X ~Expo(1), $MGF_X(t) = \frac{1}{1-t}$, which is well defined for $t < 1$ (see Example 6.4.13)
- We also know $\frac{1}{1-t} = \sum_{i=0}^{\infty} t^n$ for $|t| < 1$ (see Example 6.5.1)
- Is $\sum_{i=0}^{\infty} t^n$ also one expression of the moment generating function of Expo(1)?
- Short answer: yes as $\frac{1}{1-t} = \sum_{i=0}^{\infty} t^n$ satisfies the requirements of moment generating functions (see Definition 6.4.1). Now the main confusion part lies in the difference between two different intervals mentioned above: one is $t < 1$ and another one is $|t| < 1$. **But as long as they coincide in one small interval containing zero, we do not care for their extensions outside that small interval.**

Week 6

B. MGF

- **I. We can derive moments from the MGF (if exists).**
  Not a surprise, as all moments are encoded in the MGF function by design.

  Taking nth derivative w.r.t. t and evaluate the nth derivative function at $t = 0$:

$$\mathbb{E}X^n = M_X^{(n)}(0)$$

- **II. MGFs (if exist) determine the distributions.**
  If two r.v.s have the same MGF, they have the same distribution!
  CDF, PMF/PDF, MGF (if exists).

  Very useful when dealing with a sum of independent r.v.'s.

Week 6

B. MGF

Show that a sum of n independent $N(\mu_i, \sigma_i^2)$, $i \leq n$ is still normal.

- **MGF of $X \sim N(\mu, \sigma^2)$.**

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2},$$

which is finite for $t \in \mathbb{R}$, so the MGF, M_X, is well defined.

- **MGF of $\sum_{i=1}^n X_i$** with $X_i \sim$ *independent* $N(\mu_i, \sigma_i^2)$, $i \leq n$:

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2} = e^{\left(\sum_{i=1}^n \mu_i\right)t + \frac{1}{2}\left(\sum_{i=1}^n \sigma_i^2\right)t^2}.$$

which is finite for, e.g., $t \in (-1, 1)$ (so the MGF is well-defined) and is the MGF of a $N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$-dsitributed r.v.

- The MGF of $\sum_{i=1}^n X_i$ is the MGF of $N\left(\sum_{i=1}^n \mu_i, \frac{1}{2}\left(\sum_{i=1}^n \sigma_i^2\right)\right)$, then we know a sum of n independent $N(\mu_i, \sigma_i^2)$, $i \leq n$ is still Normal/Gaussian.

Week 6

C. Characterization of distributions.

Median.

- c **(may not be unique for a given r.v.)** is a median if $\mathbb{P}(X \leq c) \geq 1/2$ and $\mathbb{P}(X \geq c) \geq 1/2$.

Mode.

- c **(may not be unique for a given r.v.)** is a mode if the PMF/PDF takes its maximum value at c.

Symmetry.

- We say that an r.v. X has a symmetric distribution about μ (or X is symmetric) if $X - \mu$ has the same/identical distribution as $\mu - X$.

Skewness. measure of asymmetry: Negative (left) skewed (e.g., longer left tail, so more probabilities are assigned to values on the left side of the middle, and skewness is negative); positive (right) skewed (e.g., longer right tail, so more probabilities are assigned to values on the right side of the middle, and skewness is positive).

- The skewness of an r.v. X with μ, σ^2 is the third standarized moment:

$$\text{Skew}(X) = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right)^3$$

Kurtosis. measure of heavy tail.

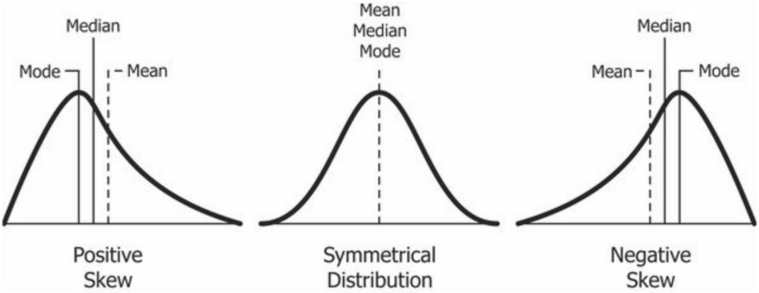
- The kurtosis of an r.v. X with μ, σ^2 is (**fourth standardized moment shifted by 3**):

$$\text{Kurt}(X) = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right)^4 - 3$$

Week 6

C. Characterization of distributions.

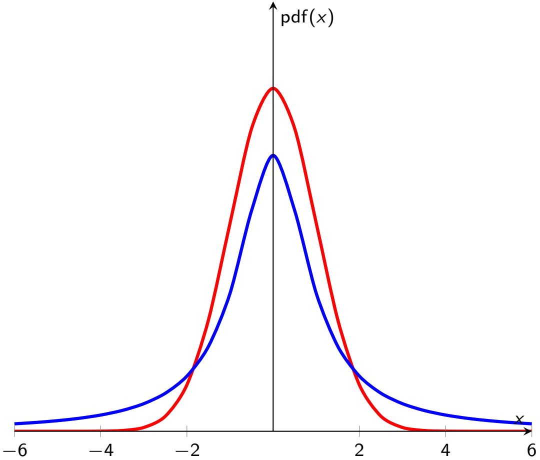
Figures from wikipedia (https://en.wikipedia.org/wiki/Skewness) on negative (left) skewed and positive (right) skewed **unimodal** distributions.



Week 6

C. Characterization of distributions.

Student-t(n) PDF (n=5) (Kurtosis is 5) and standard normal bell-shape PDF (Kurtosis is 0)



Week 6

Summary with Quizzes

Completely right or possibly wrong

- Every random varaible has its distribution.
- Every random varaible has its MGF.
- Two MGF function candidates for a given random variable may take different values when evaluated at some non-zero points.
- Skew(X)=0 does not necessarily imply that $X - \mu$ and $\mu - X$ have identical distributions.

Week 6

Summary with Quizzes

Completely right or possibly wrong

- Every random varaible has its distribution. Correct, actually for any function satisfying the CDF properties, we can find a random variable with CDF being the given function.
- Every random varaible has its MGF. Wrong, for example, Student-t distributed random variable does not even have all moments let alone moment generating functions.
- Two MGF function candidates for a given random variable may take different values when evaluated at some non-zero points. Correct, think about the example we just discussed.
- Skew(X)=0 does not necessarily imply that $X - \mu$ and $\mu - X$ have identical distributions. Correct, this is not surprising as distributions are normally more complicated than a single skewness number.

Week 7

Week 7

- Assignment 2 feedbck:

Week 7

- Assignment 2 feedbck:
  - The grading is strict this time.

Week 7

- Assignment 2 feedbck:
  - The grading is strict this time.
  - Always motivate your answers.

Week 7

- Assignment 2 feedbck:
  - The grading is strict this time.
  - Always motivate your answers.
  - Do not mix different concepts: e.g., distributions are not r.v.'s. When we say: sum of independent Pois is Pois. Techniqually and preceisely speaking, we are saying that the sum of independent Pois-distributed r.v.'s follows Poisson distribution. When you use sloppy terms (sometimes, our textbook uses as the meanings of these terms are clear based on the context), make sure it makes sense.

Week 7

- Assignment 2 feedbck:
  - The grading is strict this time.
  - Always motivate your answers.
  - Do not mix different concepts: e.g., distributions are not r.v.'s. When we say: sum of independent Pois is Pois. Techniqually and preceisely speaking, we are saying that the sum of independent Pois-distributed r.v.'s follows Poisson distribution. When you use sloppy terms (sometimes, our textbook uses as the meanings of these terms are clear based on the context), make sure it makes sense.
- Exam set-up (for now, hopefully, no more changes): two hour on-site closed book exam.

Week 7

- Assignment 2 feedbck:
  - The grading is strict this time.
  - Always motivate your answers.
  - Do not mix different concepts: e.g., distributions are not r.v.'s. When we say: sum of independent Pois is Pois. Techniqually and preceisely speaking, we are saying that the sum of independent Pois-distributed r.v.'s follows Poisson distribution. When you use sloppy terms (sometimes, our textbook uses as the meanings of these terms are clear based on the context), make sure it makes sense.
- Exam set-up (for now, hopefully, no more changes): two hour on-site closed book exam.
- Additional sessions for you: Jan 20th 8AM-10AM, additional review sessions (by me); and Jan 24th 8AM-10AM additional walk-in sessions (by me or could be our TAs) (ask last minute questions you have when preparing for the exam).

Week 7

- Assignment 2 feedbck:
  - The grading is strict this time.
  - Always motivate your answers.
  - Do not mix different concepts: e.g., distributions are not r.v.'s. When we say: sum of independent Pois is Pois. Techniqually and preceisely speaking, we are saying that the sum of independent Pois-distributed r.v.'s follows Poisson distribution. When you use sloppy terms (sometimes, our textbook uses as the meanings of these terms are clear based on the context), make sure it makes sense.
- Exam set-up (for now, hopefully, no more changes): two hour on-site closed book exam.
- Additional sessions for you: Jan 20th 8AM-10AM, additional review sessions (by me); and Jan 24th 8AM-10AM additional walk-in sessions (by me or could be our TAs) (ask last minute questions you have when preparing for the exam).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)
- discrete random variables (definition, CDF, PMF)

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
    - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
    - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} X_n$).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$).
- Expectations (definitions and calculations of discrete/continuous r.v.'s, linearity, monotonicity, **LOTUS**, **fundamental bridge (indicator r.v.'s)**, variance (sum of independent r.v.'s), calculations for r.v.'s and functions of r.v.'s).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$).
- Expectations (definitions and calculations of discrete/continuous r.v.'s, linearity, monotonicity, **LOTUS**, **fundamental bridge (indicator r.v.'s)**, variance (sum of independent r.v.'s), calculations for r.v.'s and functions of r.v.'s).

- moments

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$).
- Expectations (definitions and calculations of discrete/continuous r.v.'s, linearity, monotonicity, **LOTUS**, **fundamental bridge (indicator r.v.'s)**, variance (sum of independent r.v.'s), calculations for r.v.'s and functions of r.v.'s).

- moments
  - moments, central moments, standardized moments.

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$).
- Expectations (definitions and calculations of discrete/continuous r.v.'s, linearity, monotonicity, **LOTUS**, **fundamental bridge (indicator r.v.'s)**, variance (sum of independent r.v.'s), calculations for r.v.'s and functions of r.v.'s).

- moments
  - moments, central moments, standardized moments.
  - **sample moments** (approximations/estimators for moments)

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$).
- Expectations (definitions and calculations of discrete/continuous r.v.'s, linearity, monotonicity, **LOTUS**, **fundamental bridge (indicator r.v.'s)**, variance (sum of independent r.v.'s), calculations for r.v.'s and functions of r.v.'s).

- moments
  - moments, central moments, standardized moments.
  - **sample moments** (approximations/estimators for moments)
  - **MGF** (definitions and derivations (MGFs of sum of independent r.v.'s, of location-scale transformed r.v.'s)). Use MGFs to calculate moments, to determine distributions (for proofs).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$).
- Expectations (definitions and calculations of discrete/continuous r.v.'s, linearity, monotonicity, **LOTUS**, **fundamental bridge (indicator r.v.'s)**, variance (sum of independent r.v.'s), calculations for r.v.'s and functions of r.v.'s).

- moments
  - moments, central moments, standardized moments.
  - **sample moments** (approximations/estimators for moments)
  - **MGF** (definitions and derivations (MGFs of sum of independent r.v.'s, of location-scale transformed r.v.'s)). Use MGFs to calculate moments, to determine distributions (for proofs).
- Describe distributions (mean, mode, median, symmetry, skewness (right or left), kurtosis (heaviness of tail)).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$).

- Expectations (definitions and calculations of discrete/continuous r.v.'s, linearity, monotonicity, **LOTUS**, **fundamental bridge (indicator r.v.'s)**, variance (sum of independent r.v.'s), calculations for r.v.'s and functions of r.v.'s).

- moments
  - moments, central moments, standardized moments.
  - **sample moments** (approximations/estimators for moments)
  - **MGF** (definitions and derivations (MGFs of sum of independent r.v.'s, of location-scale transformed r.v.'s)). Use MGFs to calculate moments, to determine distributions (for proofs).
- Describe distributions (mean, mode, median, symmetry, skewness (right or left), kurtosis (heaviness of tail)).

Week 7

- probability (a function) (sample space, random events, definitions (two, the naive one involves **how to count** and the general one nests the naive one), properties (e.g., inclusion-exclusion principle, probabilities of the joint/unions/complement of events))
- how to count (binomial coefficient, and story proof of binomial coefficient)
- conditional probability (also probability, a function): Bayes' rule and LOTP, independence concerning two events, apply the conditioning to solve problems.
- real-valued random variables (functions from sample space to numbers) (CDF, PMF/PDF, definitions, properties, derivations, functions of random variables (min, max, sum, indicator...), differences and connections between discrete and continuous ones...)
- **independence concerning random variables** (definitions), and its properties/implications (recall MGFs, expectations (sum, product), variances (sum), sum of independent/i.i.d. r.v.'s)

- discrete random variables (definition, CDF, PMF)
  - Binomial (Bern, **indicator r.v.**), Uniform discrete, Hypergeometric, Negative Binomial (Geom) and Negative Hypergeometric: their PMFs, properties (e.g., HGeom(w, b, n) and HGeom(n,w + b n,w) have the identical distribution), background stories.
  - Poisson: PMF, properties (background story: from Bin to Poisson, Poisson Paradigm (approximate the distribution of the sum of independent/weakly dependent and small-success-rate indicator r.v.'s), from Pois to Bin (conditional on the sum), **relation between Pois and Expo**, sum).

- continuous random variables
  - Expo: PDF, CDF, properties (**Memoryless property, min**, distributions of X and X/a, **relation between Pois and Expo**...).
  - Unif: PDF, CDF, properties (**Universality**, Conditional on a unif is a unif, distributions of X and $aX + b$, from binomial to Unif).
  - Normal: PDF, CDF, properties (symmetry, bell-shape PDF, sum, approximation for the distribution of $\sqrt{N}\bar{X} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$).

- Expectations (definitions and calculations of discrete/continuous r.v.'s, linearity, monotonicity, **LOTUS**, **fundamental bridge (indicator r.v.'s)**, variance (sum of independent r.v.'s), calculations for r.v.'s and functions of r.v.'s).

- moments
  - moments, central moments, standardized moments.
  - **sample moments** (approximations/estimators for moments)
  - **MGF** (definitions and derivations (MGFs of sum of independent r.v.'s, of location-scale transformed r.v.'s)). Use MGFs to calculate moments, to determine distributions (for proofs).
- Describe distributions (mean, mode, median, symmetry, skewness (right or left), kurtosis (heaviness of tail)).
- **In short, all materials from the first 6 chapters of the textbook except the stared sections (sections whose titles have little stars "*").**

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

   conditional probability, LOTP, Bayes' rule, independence.

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

   conditional probability, LOTP, Bayes' rule, independence.

3. Understand the concept of a random variable and derive properties of well-known discrete and continuous random variables.

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

   conditional probability, LOTP, Bayes' rule, independence.

3. Understand the concept of a random variable and derive properties of well-known discrete and continuous random variables.

   random variables, and distributions (CDF, PMF/PDF), and some specific random varaibles and their properties (**Bin, Pois, Expo, Normal, Unif**).

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

   conditional probability, LOTP, Bayes' rule, independence.

3. Understand the concept of a random variable and derive properties of well-known discrete and continuous random variables.

   random variables, and distributions (CDF, PMF/PDF), and some specific random varaibles and their properties (**Bin, Pois, Expo, Normal, Unif**).

4. Calculate the expected value of discrete and continuous random variables.

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

   conditional probability, LOTP, Bayes' rule, independence.

3. Understand the concept of a random variable and derive properties of well-known discrete and continuous random variables.

   random variables, and distributions (CDF, PMF/PDF), and some specific random varaibles and their properties (**Bin, Pois, Expo, Normal, Unif**).

4. Calculate the expected value of discrete and continuous random variables.

   From expectations (determined by distributions) we have mean, variance, moments and MGFs. The MGFs(if exist) determine distributions.

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

   conditional probability, LOTP, Bayes' rule, independence.

3. Understand the concept of a random variable and derive properties of well-known discrete and continuous random variables.

   random variables, and distributions (CDF, PMF/PDF), and some specific random varaibles and their properties (**Bin, Pois, Expo, Normal, Unif**).

4. Calculate the expected value of discrete and continuous random variables.

   From expectations (determined by distributions) we have mean, variance, moments and MGFs. The MGFs(if exist) determine distributions.

5. Simulate and visualize the outcomes of chance experiments on your computer.

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

   conditional probability, LOTP, Bayes' rule, independence.

3. Understand the concept of a random variable and derive properties of well-known discrete and continuous random variables.

   random variables, and distributions (CDF, PMF/PDF), and some specific random varaibles and their properties (**Bin, Pois, Expo, Normal, Unif**).

4. Calculate the expected value of discrete and continuous random variables.

   From expectations (determined by distributions) we have mean, variance, moments and MGFs. The MGFs(if exist) determine distributions.

5. Simulate and visualize the outcomes of chance experiments on your computer.

   Simulations are experiments, which help us to uncover some interesting results.

Week 7

Be able to make use of the following concepts:

1. Quantify uncertainty using a coherent framework for probability.

   probability, sample space, outcomes, random events, random varaibles (indicator random varaibles)...

2. Understand the concept of a conditional probability and use conditional probability as a problem solving tool.

   conditional probability, LOTP, Bayes' rule, independence.

3. Understand the concept of a random variable and derive properties of well-known discrete and continuous random variables.

   random variables, and distributions (CDF, PMF/PDF), and some specific random varaibles and their properties (**Bin, Pois, Expo, Normal, Unif**).

4. Calculate the expected value of discrete and continuous random variables.

   From expectations (determined by distributions) we have mean, variance, moments and MGFs. The MGFs(if exist) determine distributions.

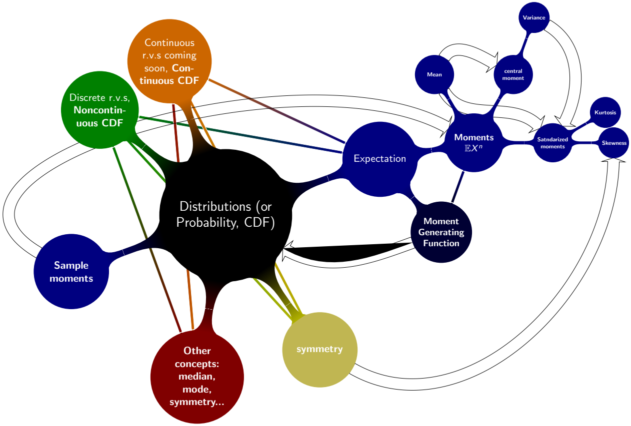
5. Simulate and visualize the outcomes of chance experiments on your computer.

   Simulations are experiments, which help us to uncover some interesting results.

Week 7

Once we have r.v.'s and their distributions, there are many derivative concepts:

Week 7

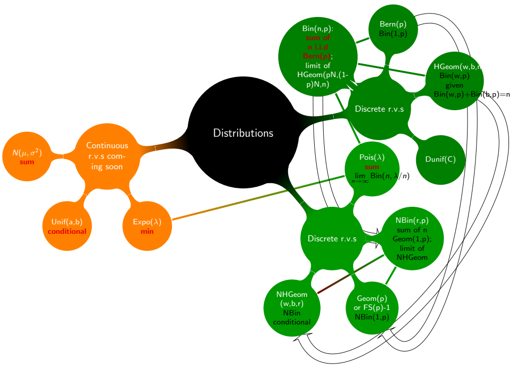


Continuous r.v.s coming soon, **Continuous CDF**
Discrete r.v.s, **Noncontinuous CDF**
Variance
Mean
central moment
Kurtosis
Moments $\mathbb{E}X^n$
Satndarized moments
Skewness
Expectation
Distributions (or Probability, CDF)
Moment Generating Function
Sample moments
symmetry
Other concepts: median, mode, symmetry...

Week 7

We have learned many commonly used distributions with background stories for intuitions. (1) We look at their CDFs, PDF/PMFs, MGFs (means, variances, modes, medians);

(2) Some of their transformations (sum, min, conditional...) and relations (e.g., from Pois to Expo).



Week 7

No more quiz.

Week 7

Thank you, and I wish you find this course useful for your future study and career.