Emory University

MATH 361 Mathematical Statistics I

Learning Notes

Jiuru Lyu

September 29, 2023

Contents

1 Prerequisites

Definition 1.0.1 (Geometric Series). A geometric series has the form

$$\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + \cdots$$

If $|r| < 1$, then the series converges to $\dfrac{a}{1-r}$. Otherwise, it diverges.

Example 1.0.2 Does the series $\sum_{n=1}^{\infty} 2^{2n}3^{1-n}$ converge or divers?

Solution 1.
Note that

$$2^{2n}3^{1-n} = \left(2^2\right)^n 3^{1-n} = 4^n\left(\frac{1}{3}\right)^{n-1} = 4 \cdot 4^{n-1}\left(\frac{1}{3}\right)^{n-1} = 4\left(\frac{4}{3}\right)^{n-1}.$$

So,

$$\sum_{n=1}^{\infty} 2^{2n}3^{1-n} = \sum_{n=1}^{\infty} 4\left(\frac{4}{3}\right)^{n-1}$$

is a geometric series, with $a = 4$ and $r = \dfrac{4}{3}$.

Since $|r| = \left|\dfrac{4}{3}\right| = \dfrac{4}{3} > 1$, the series diverges. $\square$

Definition 1.0.3 (Taylor Series). The Taylor series expanded about a of a differentiable function f is

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots .$$

Definition 1.0.4 (Maclaurin Series). The Taylor series expanded about $a = 0$.

Remark. *The Maclurin Series of* e^x *is given by* $e^x = \sum_{n=0}^{\infty} \dfrac{x^n}{n!}$.

Theorem 1.0.5 Binomial Expansion

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},$$

where $\binom{n}{k}$ is read as "n choose k" and can also be written as nCk.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}.$$

Theorem 1.0.6 Integration by Parts

$$\int u \, \mathrm{d}v = uv - \int v \, \mathrm{d}u.$$

Example 1.0.7 Evaluate $\displaystyle\int xe^{-x} \, \mathrm{d}x$.

Solution 2.

Let $u = x, \ \mathrm{d}v = e^{-x} \, \mathrm{d}x$. So, $\mathrm{d}u = \mathrm{d}x$ and $v = \displaystyle\int e^{-x} \, \mathrm{d}x = -e^{-x}$. Then,

$$\int xe^{-x} \, \mathrm{d}x = -xe^{-x} - \int -e^{-x} \, \mathrm{d}x = -xe^{-x} - e^{-x} + C.$$

□

Definition 1.0.8 (Type I Improper Integral). If $\displaystyle\int_a^t f(x) \, \mathrm{d}x$ exists for all $t > 0$, then

$$\int_a^\infty f(x) \, \mathrm{d}x = \lim_{t\to\infty} \int_a^t f(x) \, \mathrm{d}x.$$

Example 1.0.9 Evaluate $\displaystyle\int_0^\infty xe^{-x} \, \mathrm{d}x$.

Solution 3.

$$\begin{aligned}
\int_0^\infty xe^{-x} \, \mathrm{d}x = \lim_{t\to\infty} \int_0^t xe^{-x} \, \mathrm{d}x &= \lim_{t\to\infty} \Big[-xe^{-x} - e^{-x}\Big]_0^t \\
&= \lim_{t\to\infty} \Big(-te^{-t} - e^{-t} + 1\Big) \\
&= -\lim_{t\to\infty} \left(\frac{t}{e^t}\right) - \lim_{t\to\infty} e^{-t} + 1 \\
&= -\lim_{t\to\infty} \left(\frac{1}{e^t}\right) - 0 + 1 = -0 - 0 + 1 = 1.
\end{aligned}$$

□

Example 1.0.10 Double Integrals over Irregular Domains.
Consider

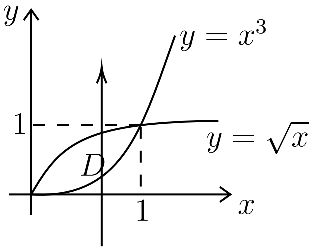
$$\iint_D 4xy - y^4 \, \mathrm{d}A,$$

where D is the region bounded between $y = \sqrt{x}$ and $y = x^3$.

Evaluate this double integral over D.

Solution 4.

Firstly, we draw the diagram representing D as follows:



$$
\begin{aligned}
\iint_D 4xy - y^3 \,\mathrm{d}A = \int_0^1 \int_{x^3}^{\sqrt{x}} 4xy - y^3 \,\mathrm{d}y\mathrm{d}x &= \int_0^1 \left[2xy^2 - \frac{1}{4}y^4\right]_{x^3}^{\sqrt{x}} \mathrm{d}x \\
&= \int_0^1 2x\big(x - x^6\big) - \frac{1}{4}\big(x^2 - x^{12}\big) \,\mathrm{d}x \\
&= \int_0^1 2x^2 - 2x^7 - \frac{1}{4}x^2 + \frac{1}{4}x^{12} \,\mathrm{d}x \\
&= \left[\frac{2}{3}x^3 - \frac{1}{4}x^8 - \frac{1}{12}x^3 + \frac{1}{52}x^{13}\right]_0^1 \\
&= \frac{2}{3} - \frac{1}{4} - \frac{1}{12} + \frac{1}{52} = \frac{55}{156}.
\end{aligned}
$$

□

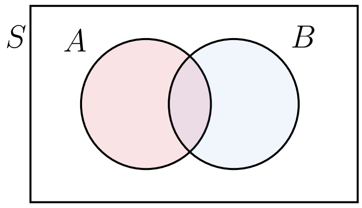
2 Probability

2.1 Sample Space and Probability

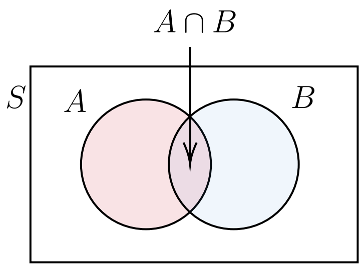
Definition 2.1.1 (Experiment). An *experiment* is a procedure with well-defined outcome.
Definition 2.1.2 (Sample Space/S). The *sample space*, denoted as S is the set of all possible outcomes of an experiment.
Definition 2.1.3 (Event). An *event* is a collection of outcomes.

Example 2.1.4 Consider flipping two coins. Use H to represent heads and T to represent tails. Then, $S = \{HH, HT, TH, TT\}$. Event "one heads"$= \{HT, TH\}$, and the event "at least one heads"$= \{HT, TH, HH\}$.

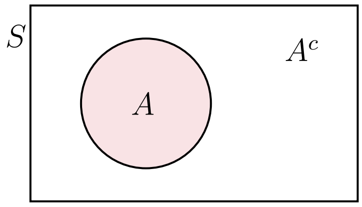
Definition 2.1.5 (Union/$\cup$). $A \cup B$ is the *union* of A and B, meaning everything in A and everything in B.



Definition 2.1.6 (Intersection/$\cap$). $A \cap B$ is the *intersection* of A and B, everything in both A and B



Definition 2.1.7 (Complement/A^c). A^c denotes the *complement* of A, meaning everything in$S$ that is not in A.



Corollary 2.1.8 $A \cap A^c = \{\} = \varnothing$.
Definition 2.1.9 (Mutually Exclusive). Two sets A and B over the same sample space are *mutually exclusive* if they have no outcomes in common. i.e., $A \cap B = \varnothing$.

Remark. *A and A^c are mutually exclusive, but not all sets mutually exclusive are complements of each other.*

Definition 2.1.10 (Probability Function). Let A be an event over a sample space S. Then, $P(A)$ denotes the *probability* of A and P is the *probability function*. The probability function P assigns a number $P(A)$ for each event $A \subseteq S$.

Axiom 2.1.11 Kolmogorov Axioms

1. Let A be an event in S, then $P(A) \geq 0$.
2. $P(S) = 1$.
3. If A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.
4. If $A_1, \ldots, A_n, \ldots$ are mutually exclusive sets, then

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

Proposition 2.1.12 $P(A^c) = 1 - P(A)$.

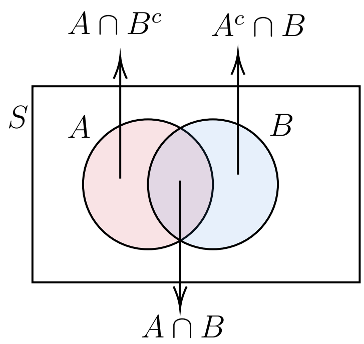
Proof 1. Note that $P(S) = 1$. Since $A^c \cup A = S$, we have $P(A \cap A^c) = 1$. Since A and A^c are mutually exclusive, $P(A \cup A^c) = P(A) + P(A^c) = 1$. So, $P(A^c) = 1 - P(A)$. ■

Proposition 2.1.13 $P(\varnothing) = 0$.

Proof 2. Note that $P(S) = 1$. Then, $P(S^c) = 1 - P(S)$. By definition, we know $S^c = \varnothing$. So, $P(\varnothing) = 1 - P(S) = 1 - 1 = 0$. ■

Proposition 2.1.14 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof 3. Consider the following Venn diagram:



Note that $P(A) = P(A \cap B) + P(A \cap B^c)$ and $P(B) = P(A \cap B) + P(A^c \cap B)$. So, we have

$$P(A) + P(B) = \boxed{P(A \cap B^c) + P(A^c \cap B) + P(A \cap B)} + P(A \cap B). \tag{1}$$

From the Venn diagram, we notice that $P(A \cap B^c) + P(A^c \cap B) + P(A \cap B)$ is exactly $P(A \cup B)$. So, Eq. (1) becomes $P(A) + P(B) = P(A \cup B) + P(A \cap B)$. That is exactly what is required: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. ■

Definition 2.1.15 (Classical Probability). In a discrete and finite case, S is finite and all outcomes are equally likely, and the probability function is defined as

$$P(A) = \frac{|A|}{|S|},$$

where $|A|$ is the cardinality of A and $|S|$ is the cardinality of S.

Example 2.1.16 Despite the definition of classical probability (probability function defined for a discrete and finite case), there are other definitions of probability functions:

1. Discrete and Countably Infinite:

   Let $S = \mathbb{N}$ be the set of natural numbers. Then,

   $$P(k) = \frac{1}{2^k}.$$

   It can also be verified that

   $$P(S) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.$$

2. Continuous and Uncountably Infinite:

   Let $S = [0, 1]$. Suppose E is a subset of $[0, 1]$ such that $\int_E \mathrm{d}x$ is defined. Then,

   $$P(E) = \int_E \mathrm{d}x,$$

   and it can also be verified that $P(S) = 1$.

2.2 Conditional Probability and Independence

Definition 2.2.1 (Conditional Probability). We read $P(A|B)$ as the probability of A given B. Knowing B occurs, we create a new sample space, in which the probability of A occurs changes:

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|}{|B|} \cdot \frac{1/|S|}{1/|S|} = \frac{|A \cap B|/|S|}{|B|/|S|} = \frac{P(A \cap B)}{P(B)}.$$

Corollary 2.2.2 $P(A \cap B) = P(A|B)P(B)$

Example 2.2.3 Find the probability of dealing A first, 2 second, and 3 third.

Solution 1.

$$\begin{aligned} P(\text{dealing } A, 2, 3) &= P(A \text{ first})P(2 \text{ second}|A \text{ first})P(3 \text{ third}|A \text{ first} \cap 2 \text{ second}) \\ &= \frac{4}{52} \cdot \frac{4}{51} \cdot \frac{4}{50} \end{aligned}$$

□

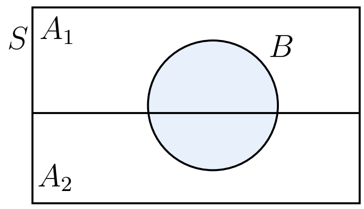
Corollary 2.2.4 $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1)$.

Theorem 2.2.5 The Law of Total Probability
Suppose the sample space $S = A_1 \cup A_2 \cup \cdots \cup A_n$, with $A_i \cap A_j = \varnothing \quad \forall i \neq j$. Then,

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n).$$

Remark. *This theorem gives us a nice way to partition the sample space.*

Example 2.2.6



As represented in the diagram above, $P(B) = P(B \cap A_1) + P(B \cap A_2)$.

Theorem 2.2.7 Bayes Theorem

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}.$$

Example 2.2.8 Coronary Artery Disease (CAD)

The probability of someone having CAD is 60%. In a study of 101 patients, 37 of them are known to NOT have CAD and 64 are known to have CAD. Of the 37 patients without CAD, 34 had negative tests while 3 had positive tests. Of the 64 with CAD, 54 had positive tests and 10 had negative tests. Find the probability of a patient has CAD given positive test.

Solution 2.

Let $T+$ be positive test, $T-$ be negative test, $D+$ be presence of CAD, and $D-$ be absence of CAD. Then, from the problem, we have

$$P(D+) = 0.6; \quad P(D-) = 1 - P(D+) = 0.4$$

and

$$P(T+|D+) = \frac{54}{64} \approx 0.84; \quad P(T-|D-) = \frac{34}{37} \approx 0.92; \quad P(T+|D-) = \frac{3}{37} \approx 0.08.$$

Then, by Bayes Theorem,

$$\begin{aligned} P(D+|T+) &= \frac{P(T+|D+)P(D+)}{P(T+|D+)P(D+) + P(T+|D-)P(D-)} \\ &= \frac{0.84 \times 0.6}{0.84 \times 0.6 + 0.08 \times 0.4} \approx \boxed{0.94}. \end{aligned}$$

□

Definition 2.2.9 (Independence). Events A and B are *independent* if $P(A|B) = P(A)$, meaning the occurrence of B does note affect the occurrence of A.

Corollary 2.2.10 If A and B are independent, then $P(A \cap B) = P(A|B)P(B) = P(A)P(B)$.

Example 2.2.11 Draw a card from 52 card deck

Let A: The card is an Ace and H: The card is a hearts. Then,

$$P(A \cap H) = P(\text{The card is an Ace of hearts}) = \frac{1}{52} = \frac{1}{4} \cdot \frac{1}{13} = P(H)P(A).$$

So, ranks and suits are independent.

Example 2.2.12 Mutually Exclusive v.s. Independence

A coin is flipped twice: $S = \{HH, TH, HT, TT\}$. Let $A =$ The first flip is $H = \{HH, HT\}$ and $B =$ The second flip is $T = \{HT, TT\}$.

- A and B are independent: $A \cap B = \{HT\}$. So, $P(A \cap B) = \frac{1}{4}$. Since $P(A \cap B)\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B)$, we know A and B are independent.

- A and B are not mutually exclusive because $P(A \cap B) = \frac{1}{4} \neq 0$.

Definition 2.2.13 (Repeated Trials). A sequence of events $A_1, \ldots, A_n$ is called independent if for any combination

$$P(A_{i1} \cap A_{i2} \cap \cdots \cap A_{ik}) = P(A_{i1})P(A_{i2}) \cdots P(A_{ik}).$$

In this case, each individual event is called a *trial.*

Example 2.2.14 Roll a fair die repeatedly. What is the probability that the first 6 appears on the roll k? If I win when 6 is rolled, what is the probability that I win?

Solution 3.

Let $A_j =$ the first 6 is rolled on roll j.

$$\begin{aligned}
j = 1 \quad & P(A_1) = \frac{1}{6} \\
j = 2 \quad & P(A_2) = \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) \\
j = 3 \quad & P(A_3) = \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) \\
j = 4 \quad & P(A_4) = \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) \\
\vdots \\
j = k \quad & P(A_k) = \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right)
\end{aligned}$$

So,

$$\begin{aligned}
P(\text{I win}) &= P(A_1) + P(A_2) + \cdots + P(A_k) + \cdots \\
&= \frac{1}{6} + \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \cdots + \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right) + \cdots \\
&= \left(\frac{1}{6}\right)\left(1 + \left(\frac{5}{6}\right) + \cdots + \left(\frac{5}{6}\right)^{k-1} + \cdots\right) \\
&= \left(\frac{1}{6}\right)\sum_{i=0}^{\infty}\left(\frac{5}{6}\right)^i = \frac{1}{6} \cdot \frac{1}{1 - \frac{5}{6}} = \frac{1}{6} \cdot 6 = \boxed{1}.
\end{aligned}$$

□

Example 2.2.15 Three people A, B and C take turn to flip a coin. Whoever gets a heads wins. Find the probability of each individual winning.

Solution 4.
First consider the case when Player A wins. Let $A_j =$ Player A wins on the j-th turn.

$$\begin{aligned}
j = 1 &\quad P(A_1) = \frac{1}{2} \\
j = 2 &\quad P(A_2) = \left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)\left(\frac{1}{2}\right) \\
j = 3 &\quad P(A_3) = \left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)\left(\frac{1}{2}\right) = \left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)^2\left(\frac{1}{6}\right) \\
j = 4 &\quad P(A_4) = \left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)^3\left(\frac{1}{2}\right) \\
&\vdots \\
j = k &\quad P(A_k) = \left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)^{k-1}\left(\frac{1}{2}\right) = \left(\frac{1}{8}\right)^{k-1}\left(\frac{1}{2}\right)
\end{aligned}$$

So,

$$\begin{aligned}
P(A\text{ wins}) = \sum_{j=1}^{\infty} P(A_j) &= \frac{1}{2} + \left(\frac{1}{8}\right)\left(\frac{1}{2}\right) + \cdots + \left(\frac{1}{8}\right)^{k-1}\left(\frac{1}{2}\right) + \cdots \\
&= \frac{1}{2}\sum_{i=0}^{\infty}\left(\frac{1}{8}\right)^i \\
&= \frac{1}{2}\cdot\frac{1}{1-\frac{1}{8}} = \frac{1}{2}\cdot\frac{8}{7} = \boxed{\frac{4}{7}}.
\end{aligned}$$

Similarly, we can get the probability of player B wins to be $P(B\text{ wins}) = \frac{2}{7}$. Finally, we can compute the probability of player C wins by

$$P(C\text{ wins}) = 1 - P(A\text{ wins}) - P(B\text{ wins}) = 1 - \frac{4}{7} - \frac{2}{7} = \frac{1}{7}.$$

□

2.3 Combinatorics

Theorem 2.3.1 Multiplication Rule
If operation A can be performed in n ways and operation B in m ways, then the sequence (operation A, operation B) can be performed in $n \times m$ ways.

Corollary 2.3.2 Ordered Sequence Consider a set A and $|A| = n$. Then, an *ordered sequence* of A, $(x_1, x_2, \ldots, x_k)$ $s.t.$ $x_i \in A$, is picked with replacement of elements. Then,

$$|(x_1, x_2, \ldots, x_k)| = n^k.$$

Remark. *In this situation, repetition is allowed.*

Definition 2.3.3 (Permutation). *Permutation* is an ordered sequence without replacement of elements. That is, $(x_1, x_2, \ldots, x_k)\ s.t.\ x_i \in A$ and $x_i \neq x_j \forall i \neq j$. Then,

$$|(x_1, x_2, \ldots, x_k)| = n(n-1)\cdots(n-k+1).$$

It is also written as ${}_nP_k = \dfrac{n!}{(n-k)!}$.

Definition 2.3.4 (Combination). *Combination* is an unordered permutation (no order, no replacement of elements). So, we have

$$\begin{aligned} \text{permutation} &= \text{combination} \times \text{orderings} \\ {}_nP_k &= {}_nC_k \times k! \\ {}_nC_k &= \frac{{}_nP_k}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k} \end{aligned}$$

Remark. *People are always distinct. Letter or coins are not usually distinct.*

Example 2.3.5 How many ways can we scramble the letters in STATISTICS?

Solution 1.

If the letter are distinct, then $10!$ ways to scramble the word. However, they are not distinct:

Non-distinct Letters	Ways to Scramble
$S-3$	$3!$
$T-3$	$3!$
$I-2$	$2!$

So, ways to scramble the word N satisfies

$$\begin{aligned} 10! &= N \cdot 3! \cdot 3! \cdot 2! \\ N &= \frac{10!}{3! \cdot 3! \cdot 2!} \quad \text{Multinomial Coefficient} \end{aligned}$$

□

Definition 2.3.6 (Multinomial Coefficient). The *multinomial coefficient* is the number of ways that n objects with n_j of type j, where $j = 1, \ldots, r$, can be distinctly ordered. So,

$$\sum_{j=1}^{r} n_j = n$$

and

$$\text{Multinomial Coefficient} = \frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot n_r!}$$

Remark. *Tips for Counting:*

1. *Draw a picture of the structure*
2. *Construct a smaller problem when there are large numbers or variables.*
3. *If the structure of the problem falls into different categories, then add instead of multiple.*

2.4 Combinatorial Probabilities

Remark. *Probability Tips*

1. *Avoid multiplying probabilities. Always set up quotient.*
2. *Keep track of order. If we have order in the sample space, we will need order in the event.*
3. *Know some basic sample spaces:*
   - *Rolling n fair die:* $|S| = 6^n$ *(ordered).*
   - *Flipping n coins:* $|S| = 2^n$
   - *Dealing a hand of n cards:* $\binom{52}{n}$

Example 2.4.1 Roll 5 Fair Die. What is the size of the sample space? What is the probability that the first three have one face and the last two another? What is the probability that two faces show up exactly twice?

Solution 1.

Size of the sample space: $|S| = 6^5$.

Let $A =$ the probability that the first three have one face and the last two another. $|A| = (6 \times 1 \times 1) \times (5 \times 1) = 30$. So, $P(A) = \frac{30}{6^5} = \frac{5}{6^4}$.

Let $B =$ the probability that two faces show up exactly twice. Note that we use $\binom{6}{2}$ to give the faces of the pairs. $\binom{4}{1}$ to the last one. $\frac{5!}{2! \cdot 2! \cdot 1!}$ ways to order the faces. So, $|B| = \binom{6}{2}\binom{4}{1}\frac{5!}{2! \cdot 2! \cdot 1!}$. Then, $P(B) = \frac{\binom{6}{2}\binom{4}{1}\frac{5!}{2!\cdot 2!\cdot 1!}}{6^5}$. □

3 Random Variables

3.1 Discrete RV: Binomial & Hypergeometric

Definition 3.1.1 (Random Variable). A *random variable* is a number determined by the outcome of an experiment, $X : S \to \mathbb{R}$. Usually, we are not particularly interested in S but in the distribution of the outcomes for X. We want to describe the probability associated with different values of X.

Example 3.1.2 Flip three coins. Count the number of heads (H):

$$S = \{HHH, HHT, HTH, TTHTTHTHT, HTT, TTT\}$$

$$\widetilde{S} = \{3H, 2H, 1H, 0H\}.$$

Note in $\widetilde{S}$, not every outcome is equally likely. We would need to define a function for or a list of the values for each outcome in $\widetilde{S}$. This is an example of a *discrete random variable.*

Example 3.1.3 Pick a student. Let the random variable Y = height of the students in cm. Then, Y is an example of a *continuous random variable.* Continuous random variables can take on an interval of values.

Notation 3.1.4. X is a random variable and has a distribution (it is still abstract and unrealized). x is a number and a realized random variable X.

Definition 3.1.5 (Discrete Random Variable). A *discrete random variable* is a random variable whose range is finite of countable.

Definition 3.1.6 (Probability Density (Mass) Function / pdf).

$$P_X(x) = P(X = x) = P(\{s \in S \mid X(s) = x\}).$$

Definition 3.1.7 (Cumulative Density Function / cdf).

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

Remark. *CDFs and PDFs can be represented by functions, graphs, or tables.*

Example 3.1.8 Roll three fair die. Let X be the largest value of the three die. Find the pdf.

Solution 1.

Note the pdf

$$P_X(x) = P(X = x) = P(X \leq x) - P(X \leq x - 1).$$

Find the cdf of x. The die that take on at most the value x, so each die have x possible outcomes, and considering order, we know

$$F_X(x) = \frac{x^3}{6^3}.$$

Therefore,

$$P_X(x) = P(X \le x) - P(X \le x-1) = F_X(x) - F_X(x-1) = \frac{x^3}{6^3} - \frac{(x-1)^3}{6^3}.$$

□

Definition 3.1.9 (Bernoulli Distribution). The *Bernoulli distribution* is the classic "flip one coin," where X is the number of heads. Let $X \sim$ Bernoulli(p), where p stands for the probability of success. $x = 1$ for success and $x = 0$ for failure. The pdf of Bernoulli distribution is

$$P_X(x) = p^x(1-p)^{1-x}$$

So,

$$P_X(1) = p; \qquad P_X(0) = (1-p).$$

Definition 3.1.10 (Binomial Distribution). The *binomial distribution* is adding Bernoulli trials together. Let $Y = X_1 + \cdots + X_n$ be the number of success with $X \sim$ Bernoulli(p) and $Y \sim$ Binomial(n, p). n is the number of trials and p is the probability of success. The pdf of binomial distribution is

$$P_Y(y) = \binom{n}{y} p^y (1-p)^y.$$

Definition 3.1.11 (Hypergeometric Distribution). Suppose we have a bag of red (r) and white (w) chips and $r + w = N$. Let $X =$ the number of red chips when choosing n chips *without replacement*. Then, $X \sim$ Hypergeometric(r, w, n), and the pdf is given by

$$P_X(x) = P(X = x) = \frac{\binom{r}{x}\binom{w}{n-x}}{\binom{r+w}{n}},$$

where $\binom{r}{x}$ is the number of ways to get x red, $\binom{w}{n-x}$ is the number of ways to get $n - x$ white, and $\binom{r+w}{n}$ is the total number of picking a size of x.

Remark. *If we choose sequence (ordered choose without replacement), we should get the exact same answer:* ${}_rP_x$ *is the number of ways the red chips can be selected in order,* ${}_wP_{n-x}$ *is the ways to choose the white chips,* $\binom{n}{x}$ *is the locations of the red chips, and* ${}_NP_n$ *is the* *orders of n chips. Then,*

$$P_X(x) = P(X = x) = \frac{({}_rP_x)\cdot({}_wP_{n-x})\cdot\binom{n}{x}}{{}_NP_n} = \frac{\dfrac{r!}{(r-x)!}\cdot\dfrac{w!}{(w-(n-x))!}\cdot\dfrac{n!}{(n-x)!x!}}{\dfrac{N!}{(N-n)!}}$$

$$= \frac{\dfrac{r!}{(r-x)!x!}\cdot\dfrac{w!}{(w-(n-x))!(n-x)!}}{\dfrac{N!}{(N-n)!n!}}$$

$$= \frac{\binom{r}{x}\binom{w}{n-x}}{\binom{N}{n}}.$$

3.2 Continuous Random Variables

Definition 3.2.1 (Random Variable). A *random variable* is an outcome of an experiment mapped to a number: $X : S \to \mathbb{R}$. We are most interested describing the probability associated with different values of X.

> **Example 3.2.2** In a chemistry experiment that depends on temperature. Let $Y =$ the temperature measured for experiment, then the unit of temperature will/may lead to different "looking" results.

Definition 3.2.3 (Continuous Random Variables). A *continuous random variable* is a function from a sample space S to the real numbers $s.t.$ for any values a, b with $a < b$, there exists a function $f_Y(y)$ $s.t.$

$$P(a < Y < b) = \int_a^b f_Y(y)\,\mathrm{d}y,$$

where $f_Y(y)$ is called the *pdf (probability density function)*, and (1) $f_Y(y) \geq 0 \quad \forall y \in \mathbb{R}$; (2) $\int_{-\infty}^{\infty} f_Y(y)\,\mathrm{d}y = 1$. Meanwhile, the *cdf (cumulative distribution function)* is defined as

$$P(Y \leq y) = F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t.$$

The *q-th quantile* c can be defined as $F(c) = q \quad \forall q \in (0, 1)$.

Remark. *The probability for $a < Y < b$ can be regarded as the area under $f_Y(y)$ over the interval (a, b). Therefore, $P(Y = a) = 0$.*

Definition 3.2.4 (Uniform Distribution). Suppose Y is a continuous random variable, and $Y \sim \text{Uniform}(a, b)$, then

$$f_Y(y) = \begin{cases} \dfrac{1}{b-a} & y \in [a, b]; \\ 0 & \text{otherwise}. \end{cases}$$

Remark. *Y is continuous does not imply $f_Y(y)$ is also continuous.*

Definition 3.2.5 (Exponential Distribution). Suppose Y is a continuous random variable, and $Y \sim$ Exponential(λ) with $\lambda > 0$ is defined as

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y \geq 0.$$

Theorem 3.2.6 Fundamental Theorem of Calculus

- $\dfrac{\mathrm{d}}{\mathrm{d}x}\displaystyle\int_a^x f(t)\,\mathrm{d}t = F(x)$; and
- $\displaystyle\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)\ s.t.\ \frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x)$.

Example 3.2.7 The Temperature Example - Cont'd.

Let X = temperature in °F and Y = temperature in °C. Given $f_X(x)$, $F_X(x)$, and $Y = \frac{5}{9}(X - 32)$. Use the cdf→pdf method to find $f_Y(y)$.

Solution 1.

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P\left(Y = \frac{5}{9}(X-32) \leq y\right) \\ &= P\left(X \leq \frac{9}{5}y + 32\right) \\ &= F_X\left(\frac{9}{5}y + 32\right). \end{aligned}$$

The derivative of cdf gives pdf:

$$\begin{aligned} f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}F_X\left(\frac{9}{5}y + 32\right) \\ &= f_X\left(\frac{9}{5}y + 32\right)\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{9}{5}y + 32\right) && \left[\mathit{Chain\ Rule}\right] \\ &= \frac{9}{5}f_X\left(\frac{9}{5} + 32\right). \end{aligned}$$

□

3.3 Expected Values and Variances

Definition 3.3.1 (Expected Values). For discrete random variables,

$$E(X) = \sum_{\text{all values of } x} xP_X(x) = \sum_{\text{all values of } x} xP(X = x).$$

For continuous random variables,

$$E(Y) = \int_{-\infty}^{\infty} y f(y)\, \mathrm{d}y$$

Remark. $E(X) = \mu_X = \mu$ *is the balancing point of the distribution, also known as the* ***first moment****. Another center of the distribution is the* ***median****,* m*, such that*

$$P(X \geq m) = P(X < m) = \frac{1}{2}.$$

Theorem 3.3.2 Properties of Expected Values

- Let $Y = g(X)$. Then,

$$E(Y) = E(g(X)) = \begin{cases} \sum_{\text{all } x} g(x) P_X(x), & \text{discrete} \\ \\ \int_{-\infty}^{\infty} g(x) f(x)\, \mathrm{d}x, & \text{continuous} \end{cases}$$

- Special Case: $Y = aX + b$.

$$E(Y) = E(aX + b) = aE(X) + b.$$

  This special case also indicates that $E(X)$ is linear.

Definition 3.3.3 (Variance). The width of a distribution can be described by the variance. The *variance* is the *second centered moment*:

$$Var(X) = E\big((X - \mu)^2\big)$$

Another way to write variance is

$$Var(X) = E\big(X^2\big) - E(X)^2$$

Theorem 3.3.4 Properties of Variance

- Variance is not linear.
- Variance is translation invariant:

$$Var(X + b) = Var(X).$$

Remark. *Variance of a line:*

$$Var(aX + b) = a^2 Var(X)$$

Proof 1.

$$
\begin{aligned}
Var(aX+b) &= E\big((aX+b)^2\big) - E(aX+b)^2 \\
&= E\big(a^2X^2 + b^2 + 2abX\big) - \big(aE(X)+b\big)^2 \\
&= a^2E\big(X^2\big) + b^2 + 2abE(X) - a^2E(X)^2 - b^2 - 2abE(X) \\
&= a^2\big(E\big(X^2\big) - E(X)^2\big) \\
&= a^2Var(X)
\end{aligned}
$$

■