alex-levesque.com

PHIL 266 — Probability and Inductive Logic

Winter 2026

Based on lectures by J. Mozersky – Queen's University

Notes written by Alex Lévesque

These notes are my own interpretations of the course material and they are not endorsed by the lecturers.
Feel free to reach out if you point out any errors.

Contents

1 Deductive and Inductive Logic **3**
1.1 Deductive Logic . . . 3
1.1.1 Truth Functions . . . 4
1.1.2 Soundness . . . 4
1.2 Inductive Logic . . . 5

2 Probability **6**
2.1 Logical Notation and Language . . . 6
2.2 Axioms of Probability . . . 6
2.3 Compound Events . . . 6
2.4 Conditional Probability and Bayes' Rule . . . 7

3 Decision Theory **9**
3.1 Expected Value . . . 9
3.2 Insurance Puzzle and Risk Averseness . . . 10
3.3 Paradoxes . . . 10
3.4 Decision Tables and Dominance . . . 11
3.5 The Prisoner's Dilemma . . . 11

4 Four Main Theories of Probability **13**

5 Logical Theory **14**
5.1 Bertrand's Paradox . . . 14

6 Personal (Bayesian) Theory **16**
6.1 Payoff Matrix . . . 16
6.2 Sure Loss Contracts (Disjunction) . . . 16
6.3 Sure Loss Contracts (Conditional) . . . 17

7 Limiting Frequency Theory **19**

7.1 Limiting Frequency and Convergence 19
7.2 Bernoulli Trials and Normal Approximations 19
7.3 Significance and Hypothesis Testing 20
7.4 Confidence Intervals and Sampling 21

8 Propensity Theory **22**
8.1 Propensity 22

9 Exam Essay Prep **23**
9.1 What are the four main interpretations of probability? Which do you prefer and why? 23

1 Deductive and Inductive Logic

1.1 Deductive Logic

Definition: In **deductive logic**, when you have true premises and a valid argument, the conclusion must be true too. Deductive logic does not take risks, i.e., there is no probability involved.

Inductive logic takes risks.

Definition: Propositions are statements that can be either true or false

Definition: An **argument** is a point or series of reasons presented to support a proposition which is the conclusion of the argument

Consider the arguments below:

Argument 1:
1. Either John is home or Sue is home.
2. John is not home.

So,
3. Sue is home

Argument 2:
1. Either Bill is happy or Kim is happy
2. Bill is not happy

So,
3. Kim is happy

These are different arguments but they have the same logical structure. Therefore, to get a sense of logical form, contrast it with grammatical form.

Definition: In logic, propositions are **true or false**. Arguments are **valid or invalid**. Strictly speaking, there is no such thing as a valid or invalid proposition, and no such thing as a true or false argument.

Definitions: A **logically necessary** argument cannot be false. A **logically impossible** argument cannot be true. A **logically contingent** is neither logically necessary nor impossible.

Definition: A **truth-preserving** argument is an argument where if all of its premises are true, then the conclusion must be true. A truth-preserving argument is valid.

Example: All humans are mortal. Socrates is a human. Therefore, Socrates is mortal. This is a truth-preserving argument. If the premises are true, the conclusion cannot be false.

1.1.1 Truth Functions

Truth functions

Inclusive 'or'

P	Not-*P*
T	F
F	T

P	*Q*	*P* or *Q*
T	T	T
T	F	T
F	T	T
F	F	F

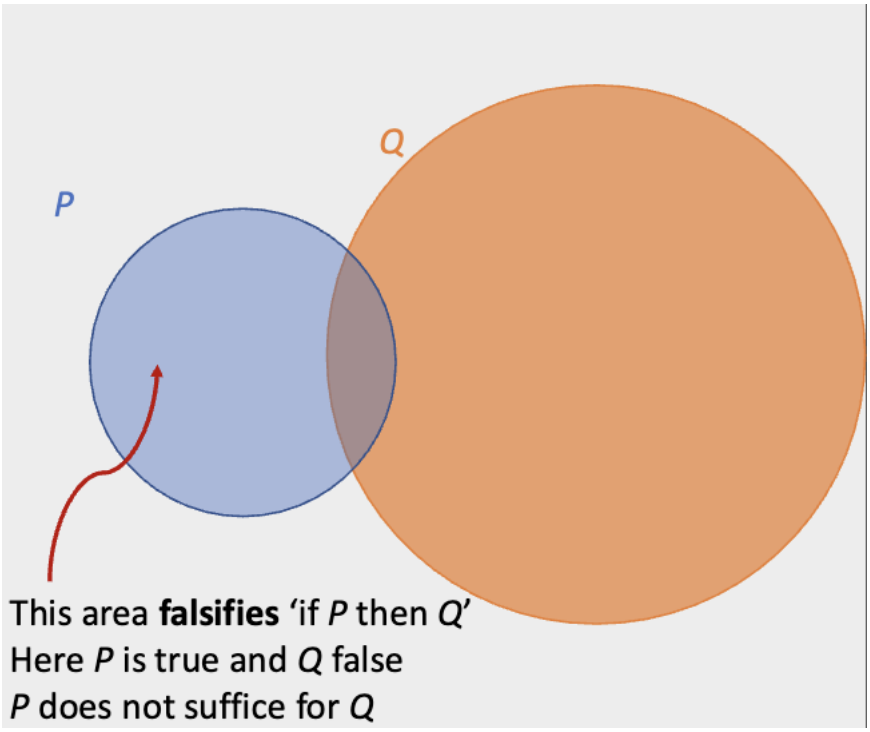
P	*Q*	*P* and *Q*
T	T	T
T	F	F
F	T	F
F	F	F

P	*Q*	If *P* then *Q*
T	T	T
T	F	F
F	T	T
F	F	T

P	*Q*	*P* if and only if *Q*
T	T	T
T	F	F
F	T	F
F	F	T

Considering *If P then Q*, we note that P is a subset of Q.

If, *if P then Q* is false, then there must be some region within P but outside Q.



1.1.2 Soundness

Definition: An argument is **sound** if and only if (1) it is valid and (2) all its premises are true. Sound arguments must have true conclusions.

Logic cannot tell us whether an argument is sound, but can tell us whether an argument is valid.

1.2 Inductive Logic

This is the introduction to uncertainty and risk in logic. We evaluate risky arguments using inductive logic.

Example: All math majors I have met went on to get good jobs. If I major in math, then I will get a good job. This outcome is actually uncertain and we must reason about risk to answer the question.

Definition: Probability is a fundamental tool for inductive logic. Say 95% of math majors get good jobs, so **probably** if I major in math, then I will get a good job.

2 Probability

It is important to understand what is meant by 'fair'. A biased coin is unfair. If is a pattern to an outcome, then this would be unfair. A fair set-up is (1) random, (2) has no memory of previous outcomes (it's irregular) and (3) is too complex for the next outcome to be predicted.

Definition: Trials on a chance set-up are **independent** if and only if the probabilities of the outcomes of the trial are not influenced by the outcomes of previous trials

Definition: A chance set-up is **fair** if and only if it is unbiased and outcomes are independent of each other

Gambler's Fallacy is the assumption that the set-up is fair, but also assumes that the set-up is not independent, so not fair.

Definition: Examples like tossed coins, dice, roulette wheels, etc. are simple and artificial **models** of real-life situations that can be much more complex.

Definition: An inference to a **plausible explanation** is a form of inductive reasoning in which you infer that a particular hypothesis best explains a set of observed facts.

Definition: An inference based on **testimony** is an inductive inference in which you accept a claim as likely true primarily because someone else asserts it.

2.1 Logical Notation and Language

In probability, we can discuss either **propositions** (statements that are t/f) or **events** (occurrences). These are equivalent in this context.

Concept	Logic Symbol	Set Theory	Meaning
Negation	$\sim A$	A'	Not A
Conjunction	$A\&B$	$A \cap B$	A and B (Both must occur)
Disjunction	$A \vee B$	$A \cup B$	A or B (At least one occurs)
Probability	$Pr(A)$	—	The probability that A occurs

2.2 Axioms of Probability

1. Range: For any outcome A, $0 \leq Pr(A) \leq 1$
2. Certainty: $Pr(\Omega) = 1$, where Ω is a guaranteed event
3. Addition Rule: If A and B are mutually exclusive (cannot happen at the same time), then $Pr(A \vee B) = Pr(A) + Pr(B)$
4. Multiplication Rule: If A and B are independent, then $Pr(A \wedge B) = Pr(A) \times Pr(B)$

2.3 Compound Events

Deck 1: 2 red cards, 3 black cards. H = die lands 5 or 6. R1 = draw red from first deck.

Deck 2: 4 red cards, 1 black card. L = die lands 1, 2, 3, or 4. R2 = draw red from second deck.

First, we throw the die. If 5-6, we pick from deck 1; If 1-4, we pick from deck 2.

What is the probability of drawing a red card?

$Pr(H\&R_1) = \frac{2}{6} \times \frac{2}{5} = \frac{2}{15}$

$Pr(L\&R_2) = \frac{4}{6} \times \frac{4}{5} = \frac{8}{15}$

$Pr(\text{Red}) = Pr[(H\&R_1) \vee (L\&R_2)] = \frac{2}{15} + \frac{8}{15} = \frac{2}{3}$

2.4 Conditional Probability and Bayes' Rule

Definition: $Pr(A|B) = \frac{Pr(A\&B)}{Pr(B)}$, provided $Pr(B) > 0$

$Pr(A|B)$ is read as the probability of event A given that event B has occurred

Example: The conditional probability of rolling an even number of a six sided fair die is $\frac{Pr(A\&B)}{Pr(B)} = \frac{1/6}{1/2} = \frac{1}{3}$

Definition: Categorical statements are absolute (e.g., The chance of winning is 0.4), while **conditional** statements include a premise (e.g. The chance of winning if the opponent is tired).

Definition: Bayes' Rule: $Pr(H|E) = \frac{Pr(H)Pr(E|H)}{Pr(H)Pr(E|H)+Pr(\neg H)Pr(E|\neg H)}$, where H is the hypothesis and E is the evidence i.e. new data observed

For several mutually exclusive hypothesis, we have:

- $\Pr(H_j/E) = \frac{\Pr(H_j) \times \Pr(E/H_j)}{\Sigma[\Pr(H_i) \times \Pr(E/H_i)]}$
  - i = 1, 2, 3, ..., i, j, ... , k.

Bayes' Rule

1.	$\Pr(A\&B) = \Pr(A/B) \times \Pr(B)$	Rule 5
2.	$\Pr(B\&A) = \Pr(B/A) \times \Pr(A)$	Rule 5
3.	$(A\&B) = (B\&A)$	Logical equiv.
4.	$\Pr(A/B) \times \Pr(B) = \Pr(B/A) \times \Pr(A)$	1, 2
5.	$\Pr(A/B) = \frac{\Pr(A) \times \Pr(B/A)}{\Pr(B)}$	4, rearranging
6.	$\Pr(B) = \Pr(B/A) \times \Pr(A) + \Pr(B/\sim A) \times \Pr(\sim A)$	total prob.
7.	$\Pr(A/B) = \frac{\Pr(A) \times \Pr(B/A)}{\Pr(A) \times \Pr(B/A) + \Pr(\sim A) \times \Pr(B/\sim A)}$	5, 6

Summary: Rules of probability

1. $0 \leq Pr(A) \leq 1$

2. $Pr(\Omega) = 1$

4. $\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A\&B)$

6. $\Pr(A \,\&\, B) = \Pr(\mathrm{A/B}) \times \Pr(B)$

7. $\Pr(A) = \Pr(\mathrm{A/B}) \times \Pr(B) + \Pr(\mathrm{A/\sim B}) \times \Pr(\sim B)$

- $\Pr(H/E) = \frac{\Pr(H) \times \Pr(E/H)}{\Pr(H) \times \Pr(E/H) + \Pr(\sim H) \times \Pr(E/\sim H)}$ BAYES' RULE

3 Decision Theory

What we decide to do depends on what we believe and what we want. What we believe is represented by probabilities and what we desire is represented by utilities (what we want).

To assess a decision, we must consider both the consequences of an
action and the probability of those consequences occurring

Definition: Decision theory combines the probabilities of consequences of an act with the utility (benefit or value) of those consequences

3.1 Expected Value

To evaluate an act, we:

1. Determine the probability of each possible consequence of the act
2. Determine the utility of each consequence
3. Multiply the probability of each consequence by its utility
4. Add up the results (this is called the expected value of the act)

$U(C)$ is Utility of a consequence A, B, C. $E(A)$ is expected value, where:

$$E(A) = \sum_{i=0}^{n} [Pr(C_i) \times U(C_i)]$$

where C_i are the consequences of A

2.2 Example: The Coin Toss Games

- **Game 1:** Toss a coin. Heads: win \$2. Tails: lose \$2.

  $$Exp(G1) = (0.5 \times 2) + (0.5 \times -2) = 0$$

  Since the value is 0, one should be indifferent to playing.

- **Game 2:** Toss coin twice. Two Heads: win \$10. Otherwise: lose \$1.

  $$Exp(G2) = (0.25 \times 10) + (0.75 \times -1) = 2.50 - 0.75 = \$1.75$$

  Since $Exp(G2) > 0$, it is advantageous to play.

2.3 Fair Price

A **fair price** is the cost to enter a game such that the expected value becomes 0. For Game 2 above, the fair price is \$1.75. If charged \$1.75 to play:

$$Exp(P) = (0.25 \times (10 - 1.75)) + (0.75 \times (-1 - 1.75)) = 0$$

In decision theory, "fair" simply means neither side has a mathematical advantage.

A fair price is the cost to enter a game such that the expected value becomes 0. For Game 2 above, the fair price is \$1.75. If charged \$1.75 to play:

$$Exp(P) = (0.25 \times (10 - 1.75)) + (0.75 \times (-1 - 1.75)) = 0$$

3.2 Insurance Puzzle and Risk Averseness

The Expected Value Rule states that one should act to maximize expected value. However, this rule sometimes conflicts with intuition regarding insurance.

Suppose there is a 1% chance your 750 iPad is stolen ($Pr(S) = 0.01$). An insurance plan costs \$10.

**Not Buying Insurance ($\sim B$): $Exp(\sim B) = (0.01 \times -\$750) + (0.99 \times 0) = -\$7.50$

**Buying Insurance (B): $Exp(B) = -\$10$

Therefore, the expected value rule suggests not buying insurance, yet people buy insurance because they are **risk averse**

3.3 Paradoxes

Allais Paradox: This paradox shows that people's choices often violate the EVR, showing irrational inconsistencies when choosing between gambles with certain vs. risky outcomes. Most people prefer a guaranteed moderate gain in one scenario (such as stocks) but prefer a high-risk, higher-potential reward in a second scenario where the certain outcome is removed

The Classic Example (Two Choices):

- **Choice 1:**
  - **Option A:** 100% chance of winning \$1 million. (Most choose this)
  - **Option B:** 89% chance of \$1 million, 10% chance of \$5 million, 1% chance of nothing.
- **Choice 2:**
  - **Option C:** 11% chance of \$1 million, 89% chance of nothing.
  - **Option D:** 10% chance of \$5 million, 90% chance of nothing. (Most choose this)

St. Petersburg Paradox: This paradox arises when a game has an infinite expected value but seems worth very little in reality

The Game: Toss a coin until it comes up tails.

- If heads on 1st toss, win \$2.
- If heads on 2nd toss, win \$4.
- If heads on nth toss, win 2^n.

The Calculation:

$$Exp(S) = \left(\frac{1}{2} \times 2\right) + \left(\frac{1}{4} \times 4\right) + \left(\frac{1}{8} \times 8\right) + \cdots = 1 + 1 + 1 + \cdots = \infty$$

Since the expected value is infinite, a rational agent should theoretically be willing to pay any finite amount to play. However, no rational person would pay millions to play this.

Pascal's Wager: Pascal argued one should believe in God even if the probability of

God's existence is low, because the utility of belief in God exists in infinite. Whether God exists or not, you gain more and lose less by believing.

- If you believe and he doesn't exist, all you lose is a few Sunday mornings
- If you believe and he does exist, you get infinite bliss after death
- If you don't believe and he doesn't exist, you only gain a few Sunday mornings
- If you don't believe and he does exist, you get infinite suffering after death

3.4 Decision Tables and Dominance

Definition: If one act yields a better (or equal) outcome than all other acts in every possible state of affairs, that act **dominates**. The rule is: If one act dominates, do it.

The Dominance Rule only works if acts don't causally influence which states of affairs occur. If one state is likely to cause another state (like if one act is going to raise the probability of a low utility state), dominance is not a good guideline here.

Improved Dominance Rule: If no act in the set of acts has any causal influence on the states of affairs in the partition, then if one act dominates, do it.

5.2 The Conflict: Causation

Dominance fails when acts **causally influence** the state of affairs. Consider the decision to exercise (E) vs. not exercise ($\sim E$) for health (H):

	Healthy (H)	**Unhealthy (∼H)**
Exercise (E)	$U = 50$	$U = 1$
Don't Exercise (∼E)	$U = 100$	$U = 2$

Here, $\sim E$ strictly dominates E (100 ¿ 50 and 2 ¿ 1). However, not exercising *causes* the probability of being unhealthy to rise.

- If $Pr(H/E) = 0.8$ and $Pr(H/\sim E) = 0.3$:
- $Exp(E) = (0.8 \times 50) + (0.2 \times 1) = 40.2$
- $Exp(\sim E) = (0.3 \times 100) + (0.7 \times 2) = 31.4$

Conclusion: When dominance and expected value conflict due to causation, **prioritize expected value**.

3.5 The Prisoner's Dilemma

When more than one person is involved, we enter the world of game theory.

- Two criminals, Bill and Sue, are given a choice by the prosecutor:

1. If one confesses, I will have enough evidence to put the other away for 20 years.
2. I will reward the confessor with freedom.
3. If you both confess, I will reduce each sentence to 10 years, for cooperation.
4. If you both stay silent, the best I can get is 1-year sentences each.

Bill/Sue	Sue confesses	Sue does not confess
Bill confesses	-10/-10	0/-20
Bill does not confess	-20/0	-1/-1

The principle of dominance recommends confessing, but the best for the group is to each stay silent.

In repeated interactions (iterated games), rational self-interest can lead to cooperation via strategies like tit-for-tat (start cooperative, then copy the opponent's previous move), creating stable patterns of mutual benefit

4 Four Main Theories of Probability

There are two types:

- Belief Type: Probability is a measure of the strength of belief. Frequencies are **evidence** for probabilities
- Frequency Type: Probability is a measure of the physical world. Frequencies/dispositions **are** probabilities

I. BELIEF-TYPE THEORIES		II. FREQUENCY-TYPE THEORIES	
A. Logical Theory	**B. Personal (Bayesian) Theory**	**A. Limiting Frequency Theory**	**B. Propensity Theory**
Probability is a measure of how strong one's beliefs **should be**.	Probability is a measure of how strong one's beliefs **are**.	Probability is a measure of **long-run frequencies** in nature.	Probability is a measure of a physical system's **disposition to produce** an outcome.
Key feature: • *What the evidence entails, 'objective'.*	*Key feature:* • *Individual response, 'subjective'.*	*Key feature:* • *Focus on outcomes in the world.*	*Key Feature:* • *Focus on the cause of outcomes.*

Frequency Principle: The frequency principle in philosophy, often attributed to thinkers like Ian Hacking, is a rule of thumb for assigning probability to a single event based on known frequencies. It states that if you know the long-run frequency of an outcome in similar trials, you should assign that same probability to a single trial, provided you know nothing else about it.

5 Logical Theory

Deductive validity is the relation of truth-preservation, i.e., when the truth of the premises **guarantees** the truth of the conclusion.

In logical theory, probability is **always** relative to evidence, and no proposition has a probability of its own

The principle of insufficient reason (indifference): if there is no reason (evidence) to favour one alternative over nay other, they should each be treated as equally probable

This concerns what we **should** believe, not what we **in fact** believe

Example:

There is a car under the tarp, it could be pink, or not pink. Indifference means that the probability of it being pink is 1/2, which is also the probability of it being not pink.

Keynes says that none of the alternative possibilities should be 'divisible' into anything that has the **same form** as another possibility. Like the car example, non-pink means red or blue or ..., which is the same form as pink, so non-pink is divisible into the same form as 'pink', therefore inadmissible

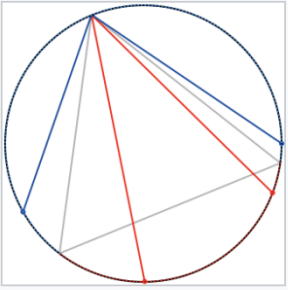
The following problem is paradoxical. It's failure does not come from the logic but from the specification. The core argument is that logical theory is **evidence-relative through-and-through**. This means probability is a property of the information you have about them, thus if we have under-specified evidence, logical theory cannot produce a single answer.

One could argue this is a strength, but some argue that one needs some beliefs before evidence can even be understood.

5.1 Bertrand's Paradox

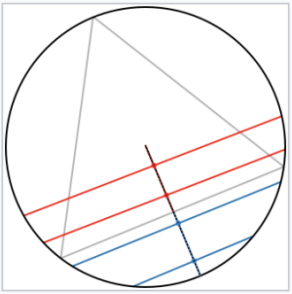
Consider an equilateral triangle that is inscribed in a circle. Supposed a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?

1. The "random endpoints" method: Choose two random points on the circumference of the circle and draw the chord joining them. To calculate the probability in question imagine the triangle rotated so its vertex coincides with one of the chord endpoints. Observe that only if the other chord endpoint lies on the arc between the endpoints of the triangle side opposite the first point, the chord is longer than a side of the triangle. The length of that arc is one third of the circumference of the circle, therefore the probability that a random chord is longer than a side of the inscribed triangle is $\frac{1}{3}$.



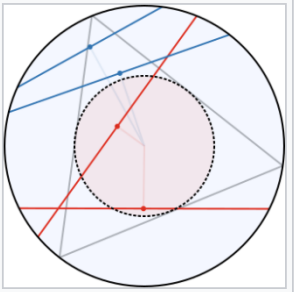
An equilateral triangle (grey) inscribed in a circle. Four chords of the circle are shown, two being longer than the triangle's side (red) and two being shorter (blue).

2. The "random radial point" method: Choose a radius of the circle, choose a point on the radius and construct the chord through this point and perpendicular to the radius. To calculate the probability in question imagine the triangle rotated so a side is perpendicular to the radius. The chord is longer than a side of the triangle if the chosen point is nearer the center of the circle than the point where the side of the triangle intersects the radius. The side of the triangle bisects the radius, therefore the probability a random chord is longer than a side of the inscribed triangle is $\frac{1}{2}$.



Random chords, selection method 2

3. The "random midpoint" method: Choose a point anywhere within the circle and construct a chord with the chosen point as its midpoint. The chord is longer than a side of the inscribed triangle if the chosen point falls within a concentric circle of radius $\frac{1}{2}$ the radius of the larger circle. The area of the smaller circle is one fourth the area of the larger circle, therefore the probability a random chord is longer than a side of the inscribed triangle is $\frac{1}{4}$.



Random chords, selection method 3

6 Personal (Bayesian) Theory

This theory demonstrates that belief, while personal, is not entirely subjective or arbitrary. We can believe whatever we want, but there are constraints on how we should **update** our beliefs.

We can see how numbers can be assigned on the basis of subjective preferences:

- Mathematicians use **betting rates** to represent degree of belief.
- E.g., you say, 'the Leafs will win the Cup'. Consider the following:
  1. You win $10 if the Leafs win the Cup.
  2. You win $10 if a fair coin falls heads (Pr = ½)
  3. You win $10 if a fair die comes up six (1/6)
  4. You win $10 if the coin falls 3 heads in a row (1/8)

- It seems clear that (2) is preferable to (3), which is preferable to (4).
  - So, we pair **each** off against (1), ask which you prefer, and thereby zero in on your degree of belief.

We can get quite precise in representing beliefs with a number, but we cannot go on forever, so no absolute final precision.

6.1 Payoff Matrix

A = the Leafs win the Cup. $\$S$ is the stake. We can represent this in a payoff matrix.

	Payoff for bet on A	Payoff for bet against A
A	$\$(1-p)S$	$\$-(1-p)S$
$\neg A$	$\$-pS$	$\$pS$

p is the rate at which you are indifferent between betting on A at rate p or $\neg A$ at rate $1-p$

6.2 Sure Loss Contracts (Disjunction)

If rates are A and $\neg A$ don't add up to 1, you are breaking the rules of logic. Then, you are subject to a sure-loss contract.

Coherence: A set of betting rates is coherent if and only if it is no open to a sure-loss contract

- **Example**: Your betting rate on the Leafs winning, W, is ¾; your betting rate **against** them winning is 5/8.

- I accept these rates:
  - You bet $6 that the Leafs win; I bet $2 that they lose.
  - Also, you bet $5 that they lose; I bet $3 that they win.
  - These are **your** rates, but…

	Bet on W	Bet against W	Payoff
W	2	-5	-3
~W	-6	3	-3

- This is the problem with inconsistency, **sure loss**! So:
  - A set of betting rates is **coherent** if and only if it is **not** open to a sure-loss contract.
- In this case, you violate the rule of certainty:
  - Pr(W v ~W) = 1, so Pr(~W) = 1 - Pr(W)

This explains why personal degrees of belief must satisfy the basic rules of probability (normality, certainty, and additivity).

Example:

1 The Disjunction Rule ($r = p + q$)

If an individual P assigns rates p to D, q to G, and r to $D \vee G$, and $r < p + q$, a sure-loss contract is constructed using a stake S.

Payoff Table (Case: $r < p + q$)

In this case, the bookie offers three bets: a bet on D, a bet on G, and a bet against $D \vee G$.

Outcome	Bet on D	Bet on G	Bet vs $D \vee G$	Total Payoff
$D \wedge \neg G$	$(1-p)S$	$-qS$	$-(1-r)S$	$(r-p-q)S$
$\neg D \wedge G$	$-pS$	$(1-q)S$	$-(1-r)S$	$(r-p-q)S$
$\neg D \wedge \neg G$	$-pS$	$-qS$	rS	$(r-p-q)S$

If $r \neq p + q$, the player is subject to exploitation

6.3 Sure Loss Contracts (Conditional)

For a conditional example, we must accept that $Pr(P\&Q) = Pr(P/Q) \times Pr(Q)$. Else, we are subject to a sure-loss contract.

2 The General Conjunction Rule ($p = q/r$)

For conditional probability, P assigns p to $R|F$, q to $R \wedge F$, and r to F. Coherence requires $p = q/r$. If $p < q/r$, a conditional sure-loss is created where the loss occurs only if F is true.

Payoff Table (Case: $p < q/r$)

Here, the bookie sets stakes based on r and q to ensure the individual P cannot win.

Outcome	**Bet on $R \wedge F$**	**Bet vs F**	**Bet vs $R\|F$**	**Total Payoff**
$R \wedge F$	$(1-q)S_r$	$-(1-r)S_q$	$-(1-p)S_r$	$pr - q$
$\neg R \wedge F$	$-qS_r$	$-(1-r)S_q$	pS_r	$pr - q$
$\neg F$	$-qS_r$	rS_q	0 (Bet Off)	0

If $p \neq \frac{q}{r}$, the player is subject to exploitation

7 Limiting Frequency Theory

Bayes' Rule helps us to use new evidence to update our beliefs, but how do we gain new evidence? Most of the time we must engage in empirical investigation, therefore rely on samples and partial information.

7.1 Limiting Frequency and Convergence

Let C represent a collective of some repetitive events or mass phenomena, e.g., 100 tosses of a coin, and k represent an attribute of that collective, e.g., heads:

If, in n members of C, k occurs $m(k)$ times, then the **relative frequency** of $k = \frac{m(k)}{n}$

If k is an attribute of a collective C, then:

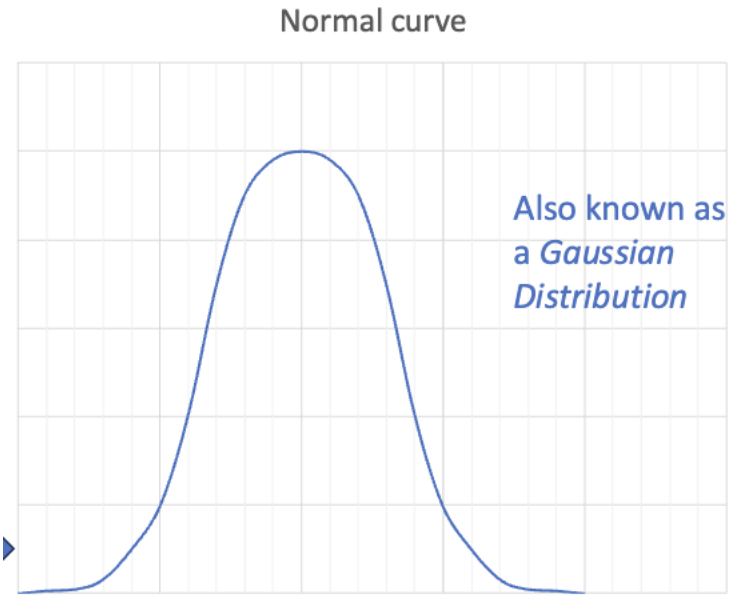
$$\lim_{n\to\infty} \frac{m(k)}{n} \text{ exists}$$

We define $Pr\left(\frac{k}{C}\right) = \lim_{n\to\infty} \frac{m(k)}{n}$. This is the limiting frequency concept. We assume that, as n increases, $\frac{m(k)}{n}$ approaches a fixed value, then, we define probability as this value

Example: for a coin, $\lim_{n\to\infty} \frac{m(heads)}{n} = \frac{1}{2}$

7.2 Bernoulli Trials and Normal Approximations

Bernoulli trials consist of repeated, independent events that share a constant probability p. As the number of trials increases, the binomial distribution of these events approximates a normal, or Gaussian, distribution curve



The theoretical mean, which represents the most probable number of outcomes, is calculated as $\mu = pn$.

The standard deviation determines the width of the normal curve and is calculated using $\sigma = \sqrt{(1-p)pn}$

The properties of an ideal normal curve dictate that sample results will fall within 1σ of the mean 68% of the time, within 2σ 95% of the time, and within 3σ 99% of the time.

Example 1

- Assume that 1/5000 car rides results in an accident. Assume, further, that in Toronto, there are 250,000 car rides per day. What is the most probable number of accidents today?
  - pn = 1/5000 x 250,000 = 50
- Assuming a normal distribution, what is the standard deviation?
  - $\sigma = \sqrt{(1-p)pn} = \sqrt{(0.9998)0.0002 \times 250{,}000} = \sqrt{49.99} = 7.07$

Example 2

- Amazing Cars Inc. claims that 99% of its cars can drive 100,000 km before requiring an oil change. The government tests 1125 cars by driving them for 100,000 km. If Amazing is telling the truth, what is the probability that the government will find that between 1104 and 1124 of Amazing's cars **don't** require an oil change?

- Assume a normal distribution:
  - μ = 0.99 x 1125 = 1113.75 = 1114
  - $\sigma = \sqrt{(1-p)pn}$
  - $= \sqrt{(0.01)0.99 \times 1125}$
  - $= \sqrt{11.14}$
  - $= 3.34$
  - So, $3\sigma = 10.02 \cong 10$
- So, the probability of finding between 1104 and 1124 cars not needing an oil change is 0.99

7.3 Significance and Hypothesis Testing

Frequency defined probability provides the mathematical foundation for significance tests, which evaluate observed data against a null hypothesis.

A null hypothesis is an assumption that the object is fair, or has no effect. A significant result occurs when the data you collect is very unlikely to happen if the null hypothesis were true, typically using a threshold of 5% or 1%.

Examples:

1. The Amazing Cars Case

Amazing Cars Inc. claims 99% of their cars go 100,000 km without an oil change ($p = 0.99$). +1

- **The Test**: The government tests 1125 cars. +1
- **The Calculation**: Based on the normal distribution, 99% of all samples should fall between 1104 and 1124 cars. +1
- **The Result**: If the government finds only 1100 cars passed, that specific result happens less than 1% of the time if the company is telling the truth.
- **The Conclusion**: The result is **significant at the 1% level**. Either a 1-in-100 fluke happened, or the company's claim is false. +1

2. The Seven Heads Coin Toss

- **Null Hypothesis**: The coin is fair ($Pr = 0.5$).
- **The Observation**: You flip the coin 7 times and get 7 heads. +1
- **The Probability**: The chance of this happening with a fair coin is $1/128$, which is approximately **0.0078 (less than 1%)**. +2
- **The Conclusion**: Since the p-value (0.01) is so low, you conclude that either the coin is biased or you just witnessed a very rare event. +1

7.4 Confidence Intervals and Sampling

Confidence intervals provide an estimated range in which an unknown population quantity is expected to lie. A confidence interval, such as 95%, indicates that the statistical procedure used to generate the interval is correct 95% of the time.

For a 95% confidence interval, the standard margin of error is calculated as $\frac{1}{\sqrt{n}}$

- Recall that $\sigma = \sqrt{(1-p)pn}$. The margin of error is largest when p = ½, i.e., when
  - $\sigma = \sqrt{\frac{1}{2} \times \frac{1}{2} \times n} = \sqrt{\frac{1}{4} \times n} = \frac{\sqrt{n}}{2}$ So: $\frac{\sigma}{n} = \frac{1}{2\sqrt{n}}$

- Therefore, we can conclude that:
  - The probability is **at least** 0.68 that the proportion, s, of Gs in the sample is within $\frac{1}{2\sqrt{n}}$ of p.
  - The probability is **at least** 0.95 that the proportion, s, of Gs in the sample is within $\frac{1}{\sqrt{n}}$ of p.
  - The probability is **at least** 0.99 that the proportion, s, of Gs in the sample is within $\frac{3}{2\sqrt{n}}$ of p.

To ensure our sample is representative, we use a stratified sample, so we break a group into MECE sub-groups and sample from each.

8 Propensity Theory

von Mises: Probability deals with problem sin which either the same event repeats itself again and again, or a great number of uniform elements are involved at the same time.

One primary objection of limiting frequency theory is that it relies on an idealization of convergence at infinity that can never be actualized in reality. Because frequency is defined on the long run, it fails to provide meaningful information for single, particular cases.

8.1 Propensity

Supposed everyone if your family but you are prone to food poisoning. On the limiting frequency view, the probability that a member of your family gets sick is basically $\frac{1}{4}$, but this is not **your** probability.

Propensity Theory: On this view, probability is a measure of the underlying structure of a system that leads to observed frequencies. Even if a system has never been tested and never will be, it can still have a tendency or disposition to give a certain frequency in the long run. **The point** is to have a theory that can make sense of single cases in an objective, scientific manner.

Simpson's Paradox:

The interpretation of probability becomes critical when dealing with statistical puzzles like Simpson's Paradox. This paradox occurs when a trend appears in several different groups of data but disappears or reverses when these groups are combined. For example, a drug might show a higher recovery rate for men and a higher recovery rate for women individually, yet show a lower recovery rate when the data is aggregated. Such cases highlight the necessity of choosing the correct theoretical framework—whether frequency, propensity, or Bayesian—to move beyond raw numbers and understand the "mathematics of the mathematics".

9 Exam Essay Prep

9.1 What are the four main interpretations of probability? Which do you prefer and why?

Logical theory treats probabilities by the **logical relationships** between propositions. The degree of belief in a statement **depends on the logic connecting** evidence to the claim. In this theory, two rational agents with identical information should **assign the same probability** to any given proposition. The **strength** is that it is **independent** from subjective preference. The **weakness** is that the theory struggles to handle cases in real-world **situations with no logical relationship**.

Bayesian theory **allows for subjectivity** by assigning initial probabilities based on prior experience of an agent. Two Bayesian agents with identical evidence may assign different probabilities **if their prior beliefs differ**. The **strength** is that it is **flexible and practical** for updating beliefs and learning. The **weakness** is that it permits arbitrary assignments of initial probabilities.

Limiting frequency theory defines probability **from the outcome of long-term repeated trials**. The **strength** is that it is **accurate for empirical data** in science and engineering. The **weakness** is that it is not applicable for **one-off events or where infinite repetition is impossible**.

Propensity theory assigns **objective probabilities to real systems** even before any trials. The **strength** is that it allows for objectivity and handling singular events, for example in fundamental physics. The **weakness** is that propensities are unobservable theoretical entities.

I prefer Bayesian theory because it **aligns with how I actually think about uncertainty in problem-solving**, as a structured process of refining belief from new evidence. The theory's inclusion of prior knowledge is applied to when I work with **machine learning models**. Bayesian reasoning is also **flexible** because it can handle repeatable experiments and also singular decisions.