Summary

Scale

- ***Ordinal scale*** are invariant up to positive monotone transformations
    - $f(x) \geq f(y)$ if and only if x $\geq$ y
    - Ordinal scale cannot conclude the precise difference between two objects.
- ***Interval scale*** accurately reflects the difference between objects.
    - $f'(x) = k * f(x) + m$
    - k and m are constants
    - It's positive linear transformation
- ***Ratio scale*** accurately reflects the ratios between objects.
    - $f'(x) = k * f(x)$
    - k is constant
    - It's positive multiplication.

Dominance Principle

- Dominated acts must not chosen.
- Dominated acts
    - whose all outcomes under every state are **at least as rational as** other outcomes of other acts under every state
    - **and** there are some outcomes under some states are **less rational** than outcomes of other acts under those states
- Let $v(a_i, s_j)$ be the **value of the outcome** corresponding to the act a_i and the state s_j.

Weak Dominance

- ***Weak Dominance*** : $a_i \succcurlyeq a_j$ if and only if $v(a_i, s) \geq v(a_j, s)$ for every state s.

Strong Dominance

- ***Strong Dominance*** : $a_i \succ a_j$ if and only if v(a_i, s_m) $\geq$ v(a_j, s_m) for every state s_m, and there is some state s_n such that v(a_i, s_n) > v(a_j, s_n).

Maximin Principle

- The maximin principle focuses on the ***worst*** possible outcome of each act.
- One should ***maximize*** the ***minimal*** value obtainable with each act.
- If the **worst possible outcome** of an act is **better than** the **worst possible outcome** of another act, the **first** act should be chosen.
- Formally speaking : $a_i \succeq a_j$ if an only if $\min(a_i) \geq \min(a_j)$.
- If the minimum value corresponding to the outcomes of two or more acts are the same, the maximin principle ranks the acts as equally rational.

Leximin Rule

- Extended version of *Maximin Principle*
- Let $min^1(a_i)$ be the value of the worst outcome of act a_i,
- $min^2(a_i)$ be the value of its second worst outcome,
- In general,$min^n(a_i)$ be the value of its n-th worst outcome
- $a_i \ \ a_j$ if and only if there is some positive integer n such that
    - $min^n(a_i) > min^n(a_j)$
    - regardless $min^m(a_i) = min^m(a_j)$ for all m < n

Maximax principle

- The maximax principle focuses on the best outcomes.
- Rationality requires us to prefer alternatives in which the best possible outcome is as good as possible.
- You should maximize the maximal value obtainable with an act.

Optimism-pessimism principle

- A decision maker's degree of optimism can be represented by a **real number** α between 0 and 1.
- α = 1: maximal optimism
- α = 0: maximal pessimism
- max (a_i) : the best outcome of act a_i
- min (a_i) : the worst outcome of act a_i

- The value of act a_i = [α · max(a_i)] + [(1 - α) · min(a_i)]
    - $\alpha \cdot max(a_i)$ is optimism part on act a_i
    - $(1 - \alpha) \cdot min(a_i)$ is pessimism part on act a_i
- α is a subjective interpretation

- α : an agent's optimism index
- [α · max(a_i)] + [(1 - α) · min(a_i)] : an act's α-index (subjective)

- Principle definition : $a_i > a_j$ if and only if α-index (a_i) > α-index (a_j)

$$[\alpha \cdot (a_i)] + [(1-\alpha) \cdot min(a_i)] > [\alpha \cdot (a_j)] + [(1-\alpha) \cdot min(a_j)]$$

- If $\alpha = 1$, evaluate α-index $\rightarrow max(a_i)$ which means the optimism pessimism rule
  collapses to the maximax rule.
- If $\alpha = 0$, evaluate α-index $\rightarrow min(a_i)$ which means the optimism pessimism rule collapses to the maximin rule.

- Note that, in order for this rule to work, the value of outcomes should be measured on an interval scale.

- Is it rational to focus just on the best and the worst cases?
- Determine a series of α, α_1, α_2, ..., α_n, such that $\alpha_1 + \alpha_2 + ... + \alpha_n = 1$, then define α-index of an act a_i as

$$\alpha_1 \cdot (a_1, s_1) + \alpha_2 \cdot (a_1, s_2) + ... + \alpha_n \cdot (a_1, s_n) = \alpha(a_1)$$

- then choose the act with the **greatest** α-index

- Note that α_1, α_2, ..., α_n are ***not probabilities***
  - similar to (subjective) **probabilities**
  - but not ***equivalent to*** **probabilities**
- They are chosen only according to how **much importance** the decision maker attaches to the best, the second best, ..., and the worst outcome of each act.
- The **relative importance** of an outcome **doesn't necessarily correspond** to its probability.

Minimax Regret

- The best alternative is one that **minimizes** the **maximum** amount of regret.
- $a_i \succ a_j$ if and only if the maximum regret of a_i is less than the maximum regret of a_j, or to put it formally :

$$max\{(v(a_i, s_1) - max(s_1), (v(a_i, s_2) - max(s_2)), ...\} < max\{(a_j, s_1) - max(s_1), (v(a_j, s_2) - max$$

Procedure

- The value of regret for each outcome is calculated by subtracting the value of the **best outcome** of **each state** from the value of the
outcome in question.
    - This obtains the ***regret matrix***
- The act chosen based on Minimax Regret principle is an act whose maximum regret for all states is minimum among other acts for all states.
    - Find the maximum regret for each act.
    - Choose the act with the least maximum regret

The principle Of Insufficient Reason

- If one has no reason to think that one state of the world is **more probable than another**, then all states should be assigned **equal** probability.
- By applying the principle of insufficient reason, an initial decision problem under **ignorance** is transformed into a decision problem under **risk**.
- If one advocates the principle of maximizing expected value as the best rule for decisions under risk, then the principle of insufficient reason can be stated as:
- $a_i \succ a_j$ if and only if $\sum_{x=1}^{n} \frac{1}{n} v(a_i, s_x) > \sum_{x=1}^{n} \frac{1}{n} v(a_j, s_x)$

The Law of Large Number

- Everyone who try to maximize the expected (utility) value will be better off in the long run.
- A random event repeat happen n times. The probability of the corresponding even is p.
- The probability of percentage that differ from p will converges to 0 as $n \to \infty$.

Allais' Paradox

- If one player choose act 1 over act 2, then that player have to choose act 3 over act 4 as well. However, players usually choose act 4 instead

	1/100	1/10	89/100
A1	x	x	x
A2	0	kx	x
A3	x	x	0
A4	0	kx	0

Relationship Properties

- $x \succeq y$ if and only if $x \succ y$ or $x \sim y$
- $x \sim y$ if and only if $x \succeq y$ and $y \succeq x$
- $x \succ y$ if and only if $x \succeq y$ and not $x \sim y$

- For every x and y, it should hold that :
- **Completeness** $x \succ y$ or $x \sim y$ or $y \succ x$
- **Asymmetry** If $x \succ y$, then it is *false* that $y \succ x$.
- **Transitivity** If $x \succ y$ and $y \succ z$, then $x \succ z$.
- **Negative Transitivity** If it is *false* that $x \succ y$ and *false* that $y \succ z$, then it is false that $x \succ z$.
  - It implies that indifference is **transitive** (if $x \sim y$ and $y \sim z$, then $x \sim z$).
  - Indifference does not follow form the assumption in **Transitivity**

- If a binary relation is symmetric and transitive, then it is necessarily reflexive
  - **reflexivity** : xRx for every $x \in X$
- **Irreflexivity**: For every $x,\ xRx$.
- **Seriality**: For every x, there is some y such that xRy.
- **Symmetry**: If xRy, then yRx.
- **Antisymmetry**: If xRy and yRx, then $x = y$.
- **Connectedness**: If xRy and xRz, then either yRz or zRy.
- **Convergence**: If xRy and xRz, then there is some u such that yRu and cRu.

Neumann and Morgenstern Method

- **Z** is a finite set of **basic prizes**
- **L** is a set of lotteries that can be constructed from **Z** by applying the following inductive definition:
  - Every basic prize in **Z** is a lottery
  - If A and B are lotteries, then so is the prospect of getting A with probability p and B with probability (1-p), for every 0 ≤ p ≤ 1.
  - Nothing else is a lottery.

Detailed Example

- Simpson is going to go to a rock concert. Three bands are playing tonight. He thinks A is better than B, which is better than C.
- Two tickets are available :
  - Ticket 1 entitles him to a 70% chance of watching A and a 30% chance of watching C.

    - Ticket 2 entitles him to watch B with 100% certainty.
- Upon reflection, he finds the two tickets equally attractive.
- Suppose we know that Simpson always acts according to the principle of maximizing utility.
- By reasoning backward, we can find the utilities he attaches to A, B, C. That is have to solve the following equation:

$$0.7 \cdot u(A) + 0.3 \cdot u(C) = 1.0 \cdot u(B)$$

- Stipulate that u(A) = 100 and u(C) = 0, then it turns out that u(B) = 70.
    - u(A) and u(C) are picked arbitrary.

The Mathematics of Probability

Kolmogorov Axioms

- **Axiom 1**:
    - $1 \geq p(A) \geq 0$, probability of every event lies between 0 and 1.
- **Axiom 2**:
    - $p(S) = 1$, probability of the entire sample space is 1.
- **Axiom 3**:
    - If $A \cap B = \emptyset$, then $p(A \cup B) = p(A) + p(B)$ (keyword, sample space).
    - $A \cap B = \emptyset$ means A and B are mutually exclusive.
- **Proposition version of Axiom 2:**
    - If A is a logical truth, then p(A) = 1.
- **Proposition version** of **Axiom 3**:
    - A and B are mutually exclusive, then $p(A \vee B) = p(A) + p(B)$.

Theorem

- **Theorem 6.1**: p(A) + p(~A) = 1
    - A v ~A is a logical truth and **Axiom 2**
- **Theorem 6.2**: If A and B are logically equivalent, then $p(A) = p(B)$
    - p(A v ~B) = p(A) + p(~B)
    - p(A) + p(~B) = 1
- **Theorem 6.3**:
    - p(A v B) = p(A) + p(B) - p (A ^ B)
    - p(A $\rightarrow$ B) = p(~A) + p(B) - p(~A ^ B)
- **Definition 6.1**
    - The probability of A given B, p(A|B).
    - p(B) $\neq$ 0.

    - $p(A|B) = \frac{p(A \wedge B)}{p(B)}$
    - p(~A|B) = 1 - p(A|B)
- **Definition 6.2**
    - A is **independent** of B **if and only if** p(A) = p(A|B)
- **Theorem 6.4** (Definition 6.1 and 6.2)
    - If A is **independent** of B, then p(A ^ B) = p(A) * p(B).
- **Theorem 6.5** (*Inverse probability law*)
    - $p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$, given that p(A) $\neq$ 0.
    - **Theorem 6.2** and **Definition 6.1**
- **Theorem 6.6**
    - $$p(B|A) = \frac{p(B) \ \cdot \ p(A|B)}{[p(B) \ \cdot \ p(A|B)] + [p(\neg B) \ \cdot \ p(A|\neg B)]}$$
    - given that p(A) $\neq$ 0
    - eliminate one of the unconditional probability
    - $p(B)$ is known as the ***prior probability***

Unknown Priors

- ***Prior probability*** : p(B)
    - Unconditional probability of B

Case Study

- Playing Roulette with 38 numbered pockets marked (0, 00, 1, 2, 3, ..., 36).
- The house wins whenever the ball falls into the 0 or 00.
- However, the first five times you play, the house wins every single time.
    - Suspect the roulette wheel is manipulated but don't know how many.
    - What is probability that the roulette wheel in front of you is manipulated? (***Given that the house win 5 times in a row).
- p(B), probability that the wheel is rigged.
    - ***Prior Probability***
- p(B|5H), probability that the wheel is rigged
    - given that the house wins five times in a row.
- All five trials are independent of each other
    - p(5H|~B) = (1/19) ^ 5.
- Newspaper article p(5H|B) = (1/2)^5.
    - It's much higher than p(5H|~B).
    - The observation fit better when the hypothesis that the wheel is biased.
- ***Prior Probability***, p(B), is unknown.
- Right now, you have a equation for ***Bayes' Theorem*** but with two unknown

- p(B) and p(B|5H).

- If the ***prior probability***, p(B) is unknown, one can choose whatever value of p(B) he / she wishes.
    - Depend on the situations and personal believe.
    - Case study :
        - Higher prior if there are many manipulated roulette wheels.
        - Lower prior if the owner respects the law.
- By inserting the chosen ***prior probability*** into ***Bayes' Theorem***, one can find
    - p(B|A) and p(~B|A) = 1- p(B|A).
- By applying ***Bayes' Theorem***, one has been able to *update* his / her initial beliefs from the new information.
- The **new probabilities**, p(B|A) and p(~B|A), are ***posterior probability***.
    - The **new probabilities** are yielded by using
    - ***prior probability***, p(B), and
    - some observation on p(A|B) and p(A|~B).
- To choose a (next) prior that is close to "correct" one,
    - use the **first posterior probability** as the **new prior probability**
        - when applying Bayes' theorem in the next time.
    - observe the events several more times,
    - apply Bayes' theorem again to the new data set.
- This strategy "wash out" the incorrect prior probability using new data.
    - Each time Bayes' theorem is applied, one will get somewhat closer to the truth.

Example

- Rolling four fair dice. What is the probability of getting at least one six?
    - A - get at least one six in a roll of four dice.
    - ~A is logically equivalent with getting no six at all when rolling four dice
    - p(~A) = 1- 5/6 * 5/6 * 5/6 * 5/6 = 0.52
- **Theorem 6.2**
    - What is the probability of A and B? Given that the probability of A is 0.2 and the probability of B is 0.1, whereas the probability of A or B is 0.25
    - p(A ^ B) = p(A) + p(B) - p(A v B) = 0.05
- **Definition 6.1**
    - Roll a fair die twice. Given that the first roll is a 5, what is the probability that the total sum will exceed 9?
    - A = "the sum exceeds 9", 10, 11, 12.
    - B = "the first roll is a 5".
    - p(A ^ B) = "The sum exceeds 9 and the first roll is a 5" = 2/36
        - This require you to count.
        - Don't fall into the infinite loop using **Theorem 6.2**.

    - Be careful to each statement and its logical form.
        - They can be very similar but they are completely different in logical sense.
        - Example : $p(A \wedge B)$ and $p(A|B)$.
- **Theorem 6.5**
    - Probability that the gearbox will break down (B) given the appearance of a welding crack (A).
    - 90 % of all broken gearboxes have welding cracks, $p(A|B) = 0.9$.
    - 10% of all gearboxes break down during the lifespan, $p(B) = 0.1$.
    - 20% of all gearboxes have welding cracks, $p(A) = 0.2$.
    - $p(B|A) = 0.1 * 0.9 / 0.2 = 0.45$
    - Be careful to assign variable
    - Be careful reading the statement and identify which variable is the given one / condition.
