Why do we use this "Competitive Ratio" thing anyways?

or

Decision Theories and Optimality

Class Presentation for

15-854: Approximation and Online Algorithms

Patrick Riley and **Elly Winner**

April 5, 2000

Why Competitive Ratio?

The questions you should have been asking, but probably weren't:

- Why do we use the competitive ratio to say an algorithm is good or bad?
- What if I have a grudge against the competitive ratio? Could I reasonably use something else?
- We always consider the worst case adversary. What if I know the adversary probably won't be that bad?

Outline

- Basic Definitions
- Strict Uncertainty
  - Other optimality criteria
  - "Bad" examples
  - Axiomatic characterization of criteria
- Pure Risk
  - Absolute Utility Functions
  - Personalization and Utility Theory
  - Other Approaches
- Conclusion

Simple Decision Theoretic Model

$\mathcal{S}$

$\mathcal{A}$		s_1	s_2	s_3	s_4
	a_1	70	80	90	1
	a_2	1	1	1	10,000

- One-shot problem
- Set $\mathcal{A}$ of actions (assume finite for simplicity)
- Set $\mathcal{S}$ of states (assume finite for simplicity)
- Cost function $C : \mathcal{A} \times \mathcal{S} \to \mathbb{R}$
- We choose an action, nature chooses a state
- We pay the amount specified by the cost function

Preference Relations

In order to be as general as possible, we'll talk about a preference relation $\preccurlyeq$

$a \preccurlyeq b$ means that a is preferred to b.

Notice the similarity to $\leq$, reflecting that smaller cost is usually preferred

Important notes:

- $\preccurlyeq$ must be Complete and Transitive
- $a \prec b$: denotes a is strictly preferred to b (i.e. $b \not\preccurlyeq a$)
- $a \approx b$: denotes a is indifferent to b (i.e. $a \preccurlyeq b$ and $b \preccurlyeq a$)

Certainty and Uncertainty

Several cases based on the amount of information we have:

- Under Certainty: We know what state nature will pick. Classic optimization.
  Not what we'll do here.
- Under Uncertainty:
  - Strict Uncertainty: We have *no* information about likelihood of the states of nature
    We'll do this first
  - Pure Risk: Probabilities of the various states are known
    We'll do this second
  - Partial Information: Somewhere in the middle of the previous two. Most real-world cases.
    This is really hard, so we won't talk about it much

Progress Report

- Basic Definitions
- Strict Uncertainty
  - Other optimality criteria
  - "Bad" examples
  - Axiomatic characterization of criteria
- Pure Risk
  - Absolute Utility Functions
  - Personalization and Utility Theory
  - Other Approaches
- Conclusion

Value Functions

And you thought we were done with definitions!

A Value Function V is just a convenient way to express a preference relation

- V assigns a real number to each action
- Preferred
  $a \preccurlyeq b$ if and only if $V(a) \leq V(b)$
- Strictly Preferred
  $a \prec b$ if and only if $V(a) < V(b)$
- Indifferent
  $a \approx b$ if and only if $V(a) = V(b)$

Some Candidate Decision Criteria

- Competitive Ratio: Minimize multiplicative times worse than OPT.

$$V(a) = \max_s \left\{ \frac{C(a,s)}{\min_x C(x,s)} \right\}$$

- Minimax Cost: Minimize worst-case cost.

$$V(a) = \max_s C(a,s)$$

- Minimax Regret: Minimize additive amount worse than OPT.

$$V(a) = \max_s \left\{ C(a,s) - \min_x C(x,s) \right\}$$

- Pessimism/Optimism Index: Balance worst and best cases. Choose $\alpha \in [0,1]$

$$V(a) = \alpha \max_s C(a,s) + (1-\alpha) \min_s C(a,s)$$

- Principle of Insufficient Reason: Assume uniform probability distribution.

$$V(a) = \sum_s \frac{C(a,s)}{|\mathcal{S}|}$$

Minimax Cost is bad!

	s_1	s_2	s_3	s_4
a_1	1	1	1	10^6+1
$\rightarrow$ a_2	10^6	10^6	10^6	10^6

- Minimax Cost chooses a_2
- However, a_1 is much better in most cases and only slightly worse in one case
- We'll have lots of regret if we use this!
- Minimax Regret and Competitive Ratio do better

Minimax Regret is bad!

	s_1	s_2	s_3	s_4
a_1	1	1	1	$11,000$
$\rightarrow$ a_2	$1,000$	$1,000$	$1,000$	$10,000$

- Minimax Regret chooses a_2
- However, a_1 pays almost nothing in three states and in the other, increases cost by only 10%.
- Diminishing Marginal Utility: $\$10$ matters more to a pauper than to Bill Gates

Pessimism/Optimism Index is bad!

- First, there's the question of how to choose α
- But there is a more serious problem...

	s_1	s_2	s_3	V
a_1	1	0	1	$1/2$
a_2	0	1	1	$1/2$
z	$\frac{1}{2}$	$\frac{1}{2}$	1	$3/4$

Let $\alpha = \frac{1}{2}$.

- actions a_1 and a_2 are both optimal
- action z is just a probability mixture of a_1 and a_2 $z := (1/2)a_1 + (1/2)a_2$
- A probability mixture will eventually choose 1 of the optimal actions, so how could it be suboptimal?

Principle of Insufficient Reason is bad!

- Extremely vague: What are the basic events?
- Is one partition better than another?

An example:

- You have an urn with two balls, both of which are either blue or purple
- What is the probability of drawing two balls of different colors?
- One partition of the outcome space:
  {blue,blue} {blue,purple} {purple,purple}
- Another partition of the outcome space:
  ⟨blue, blue⟩ ⟨blue, purple⟩ ⟨purple, blue⟩ ⟨purple, purple⟩
- Which one is better/right?

Competitive Ratio is bad!

- Can't handle zero or negative costs!

		s_1	s_2	s_3	s_4
$\rightarrow$	a_1	70	80	90	$1/10,000$
	a_2	1	1	1	1

- Competitive ratio choose a_1
- Is that what you would take?

Argh! Everything is bad! Should we give up?

- Nothing is perfect! Welcome to the real world.
- Can we say that some of these are better or worse than others?
- How can we systematically compare decision criteria?
- One possibility is to give axioms which characterize each of the criteria. Depending on which axioms, you like, you can choose your criterion.

Generalized Decision Model

This slide could be filled with lots of symbols and other stuff you would have to stare at for a while.

Instead, I'll just say this:

- Consider the algorithmic decision problem $P := (\mathcal{A}, \mathcal{S}, C)$
- $\mathcal{A}$ is a non-empty set of algorithms

  There could be infinitely many. They can be on-line or offline.
- $\mathcal{S}$ is a non-empty set of finite (but unbounded) length request sequences

  There could be infinitely many.
- $C : \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$

  This is the cost function
- Still compare to OPT, which may not be in $\mathcal{A}$.

Ax 0: Completeness

There is a preference relation $\preccurlyeq_P$ on the algorithms.

Preference relation $\preccurlyeq_P$ over $\mathcal{A}$.

Must be complete and transitive

Ax 1: Algorithmic Labeling

It doesn't matter what you call the algorithms.

$\preccurlyeq_P$ is invariant under algorithm labeling

Let $P' = (\mathcal{A}', \mathcal{S}, C')$.

If there exists an isomorphism $h : \mathcal{A} \to \mathcal{A}'$, then $a \preccurlyeq_P b$ if and only if $h(a) \preccurlyeq_{P'} h(b)$

Ax 2: Request Seq. Labeling

It doesn't matter what you call the request sequences.

- This is a little strange because the request sequences could have different lengths, but this is hard to argue against.

$\preccurlyeq_P$ is invariant under request sequence labeling

Let $P' = (\mathcal{A}, \mathcal{S}', C')$.

If there exists an isomorphism $g : \mathcal{S} \to \mathcal{S}'$, then $a \preccurlyeq_P b$ if and only if $g(a) \preccurlyeq_{P'} g(b)$

Ax 3: Weak Domination

If a is always at least as cheap as b, then $a \preccurlyeq_P b$

- Any rational decision criterion should satisfy this

---

a weakly dominates $b \Rightarrow a \preccurlyeq_P b$

- a weakly dominates b iff $\forall s \in \mathcal{S}, C(a,s) \leq C(b,s)$

Ax 4: Uniform Continuity

Take a sequence of algorithmic decision problems $\{P_k\}$ whose (cost functions) converge (uniformly) to $P = (\mathcal{A}, \mathcal{S}, C)$

If $a \preccurlyeq_{P_k} b$ for all k, then $a \preccurlyeq_P b$

- Note that cost functions are only part of sequence that changes
- We need this technical point of "uniform" convergence because otherwise it's not consistent with the Competitive Ratio

---

Let $\{P_k = (\mathcal{A}, \mathcal{S}, C_k)\}$ be a sequence of decision problems such that $\{P_k\}$ uniformly converges to $P = (\mathcal{A}, \mathcal{S}, C)$.

If $a \preccurlyeq_{P_k} b$ for all k, then $a \preccurlyeq_P b$

Ax 5: Independence of *Weakly Dominated* Alternatives

If we add an algorithm which does not perform better than the previous best at any single point, then the ranking of the original algorithms is unchanged.

- Why this bit about not performing better?
- Because it's not consistent with Competitive Ratio otherwise!

Let $P' = (\mathcal{A} \cup E, \mathcal{S}, C)$ be any algorithmic extension of $P = (\mathcal{A}, \mathcal{S}, C)$ such that $\forall a \in \mathcal{A}$, $\forall s \in \mathcal{S}$, $C'(a, s) = C(a, s)$ and $\forall b \in E$, $\forall s \in \mathcal{S}$, $\exists a \in \mathcal{A}$ such that $C(a, s) \leq C(b, s)$. Then $\preccurlyeq_{P'}$ is equivalent to $\preccurlyeq_P$ over $\mathcal{A}$.

Illustrating Axiom 5

	a	1	3
OPT→	b	2	1
	x	1	100

- x is a valid addition according to Axiom 5.

new OPT→	a	1	3
OPT→	b	2	1
	y	1/2	100

- y is a not a valid addition according to Axiom 5.

Ax 5′: Independence of *Irrelevant* Alternatives

Not Consistent with the Competitive Ratio

If we add an algorithm then the ranking of the original algorithms is unchanged.

- To some people, this seems easier to support than Axiom 5.
- What about the man in the restaurant?
- However, if we require $\mathcal{A}$ to be well-represented at every point (i.e. for every state of nature, there exists an optimal algorithm), Axiom 5 is equivalent to Axiom 5′.

Let $P' = (\mathcal{A} \cup E, \mathcal{S}, C)$ be any algorithmic extension of $P = (\mathcal{A}, \mathcal{S}, C)$ such that $\forall a \in \mathcal{A}$, $\forall s \in \mathcal{S}$, $C'(a, s) = C(a, s)$. Then $\preccurlyeq_{P'}$ is equivalent to $\preccurlyeq_P$ over $\mathcal{A}$.

Ax 6: Convexity

Any probability mixture of indifferent algorithms is no worse than the original algorithms

- Any probability mixture will eventually lead to one of the algorithms, so how could it be worse?

Let $a_1, a_2 \in \mathcal{A}$ with $a_1 \approx a_2$. For any $p \in [0,1]$, let $b = pa_1 + (1-p)a_2$. Use $P' = (\mathcal{A} \cup \{b\}, \mathcal{S}, C)$. Then $b \preccurlyeq_{P'} a_1$ and $b \preccurlyeq_{P'} a_2$

Ax 7: Redundancy

We can eliminate redundant states of nature

in other words

Redundant columns can be removed

$$\begin{array}{c|c|c|c|c|c|} & s_1 & s_2 & s_3 & s_4 & s_5 \\ \hline a & x & w & w & w & w \\ \hline a & y & z & z & z & z \\ \hline \end{array} \quad \Leftrightarrow \quad \begin{array}{c|c|c|} & s_1 & s_2 \\ \hline a & x & w \\ \hline a & y & z \\ \hline \end{array}$$

- Remember we are in strict uncertainty

Let $P' = (\mathcal{A}', \mathcal{S}', C')$, where $S' = S - E$ and C' is identical to C at all $a \in \mathcal{A}, s \in \mathcal{S}'$. If for all $s \in E$, there exists some $s' \in \mathcal{S}'$ such that for all $a \in \mathcal{A}$, $C'(a, s') = C(a, s)$, then $\preccurlyeq_P$ is equivalent to $\preccurlyeq_{P'}$

Ax 8: Proportionality

Scaling any column does not affect the preference relation.

- This is the hardest axiom to justify. Any suggestions?
- If we assume there is a fixed unknown probability distribution, this might make some sense.
- We are only interested in relative performance...

Let $P = (\mathcal{A}, \mathcal{S}, C)$ be a problem ϵ_P-bounded away from zero (i.e. $\inf_{a,s} C(a,s) \geq \epsilon_P > 0$). Let c be any positive real number, and s^* be any element of $\mathcal{S}$. Set $P' = (\mathcal{A}, \mathcal{S}, C')$ such that

$$C'(a,s) = \begin{cases} c \cdot C(a,s) & \text{if } s = s^* \\ C(a,s) & \text{o.w.} \end{cases}$$

Then if P' is ϵ_P-bounded away from zero, $\preccurlyeq_{P'}$ is equivalent to $\preccurlyeq_P$.

Illustrating Ax 8

	s_1	s_2	s_3	s_4
a_1	70	80	90	1
a_2	1	1	1	10,000

	s_1	s_2	s_3	s_4
a_1	70	80	90	1/10,000
a_2	1	1	1	1

- Axiom 8 says that these two decision problems are essentially the same.
- Do you agree?

Ax $8'$: Column Linearity

Not Consistent with the Competitive Ratio

Shifting any column does not affect the preference relation.

- Imagine someone told you they would give you $\$10$ extra if nature happened to pick s_3. Would that change your preferences?
- What about $\$1,000,000$?
- This characterizes another notion of relative performance

Let $P = (\mathcal{A}, \mathcal{S}, C)$ be an algorithmic decision problem. Let c be any real number, and s^* be any element of $\mathcal{S}$. Set $P' = (\mathcal{A}, \mathcal{S}, C')$ such that

$$C'(a,s) = \begin{cases} c + C(a,s) & \text{if } s = s^* \\ C(a,s) & \text{o.w.} \end{cases}$$

Then $\preccurlyeq_{P'}$ is equivalent to $\preccurlyeq_P$.

Ax 9: Linearity

Not Consistent with the Competitive Ratio

If we apply a positive linear transformation to the cost function, our preferences do not change.

- Kind of a combination of Axioms 8 and $8'$, excepts the transformation is applied to the *entire* cost function.
- This is appealing. Should it matter if temperatures are given in Fahrenheit or Celsius?

Let $P' = (\mathcal{A}, \mathcal{S}, C')$, where $C' = \lambda C + \beta$ with $\lambda > 0$. Then $\preccurlyeq_{P'}$ is equivalent to $\preccurlyeq_P$.

Characterization of the Criteria

- Let $+$ denote consistency with an axiom
- Let $-$ denote inconsistency with an axiom
- Let $\oplus$ denote characterization

Criterion	Axiom 1	2	3	4	5	5′	6	7	8	8′	9
Competitive Ratio	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$-$	$\oplus$	$\oplus$	$\oplus$	$-$	$-$
Minimax Cost	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$+$	$\oplus$	$\oplus$	$\oplus$	$-$	$+$	$+$
Minimax Regret	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$-$	$\oplus$	$\oplus$	$-$	$\oplus$	$+$
Pess-Opt	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$+$	$\oplus$	$-$	$\oplus$	$-$	$-$	$\oplus$
Insuff. Reason	$\oplus$	$\oplus$	$\oplus$	$+$	$+$	$\oplus$	$+$	$-$	$-$	$\oplus$	$+$

So What!

This is the point where I try to wake up those who fell asleep when I first said "Axiom" and those who got lost a long time ago.

- We have a several possible optimality criteria (other than the Competitive Ratio on which we focus in this course)
- None of them is perfect! They all have bad cases and weaknesses!
- When you work on your favorite application, you need to pick a criterion.
- One fairly principled way to pick one is to understand what characterizes it; what axioms you have to believe will be true if you want to use this criterion
- While amazingly popular, the competitive ratio may not be the right thing for your application! You need to be able to make this choice!

Progress Report

- Basic Definitions
- Strict Uncertainty
  - Other optimality criteria
  - “Bad” examples
  - Axiomatic characterization of criteria
- Pure Risk
  - Absolute Utility Functions
  - Personalization and Utility Theory
  - Other Approaches
- Conclusion

Pure Risk

Risk: Consequences of decisions are uncertain, but quantifiable with a probability distribution.

Assume:

- We always have a full probability distribution;
- Our distribution is the “true” underlying distribution.

A definition for later usage:
Lottery: Any probability distribution over a set of outcomes (prizes).

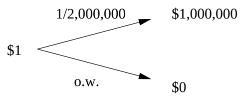
Note: We will now be talking primarily about gains rather than costs.

Fair Value of Lotteries

Fair value of a lottery?
Pascal & Fermat: expected \$ return.

Is expected return really the fair value of a lottery?

What about these two lotteries?

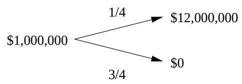


Lottery 1

$\$1 \longrightarrow \$1$

Lottery 2

And these two?



Lottery 3

$\$1{,}000{,}000 \longrightarrow \$1{,}500{,}000$

Lottery 4

Which would most people pick?

St. Petersburg Paradox

Casino in St. Petersburg will run any lottery if we can find an “appropriate” participation price.

What about this one?
Fair coin. Gambler receives $\$2^n$ if the first heads occurs on the nth flip.

Expected return is infinite!

$$1/2*1+1/4*2+1/8*4+... = 1/2+1/2+1/2...$$

Cut off flips at \$1,000,000?
Expected return = \$500,000!

Utility Function for Money

Bernoulli: Value of \$x is $\log x$. Should pay \$4 to enter St. Petersburg lottery.

Cramer: Value of \$x is $\sqrt{x}$. Should pay \$12 to enter St. Petersburg lottery.

Underlying Assumptions:

- More money is better;
- Utility of gain depends on the amount of money you already have;
- Value of gain grows at a diminishing rate.

Paradox Fixed?

What if we rearrange the prizes?

Utility function $U(x)$ like Bernoulli's or Cramer's. Assume that $\forall n \in \mathbb{Z}, \exists x_n$ such that $U(x_n) = 2^n$. Modify lottery so the first head in the nth trial gets a prize of x_n.

Now Bernoulli and Cramer would pay way too much to participate.

What Now?

Two possible ways out:

- Bounded utility—assume there is a maximum possible utility;
- Undervaluing—small probabilities are impossible.

Problems with these:

- Bounded utility removes expressiveness (e.g. can't distinguish between earning \$1,000,000 and \$12,000,000);
- What if all probabilities are small? Can't say everything is impossible!

Personalization

Bernoulli and Cramer solutions are ad hoc.

If we're ad hoc, we should be appropriately ad hoc. Different people could have different utility functions.

Risk Averse: Person not willing to pay expected return of lottery. Utility functions are strictly concave.

Risk Neutral: Person willing to pay exactly expected return. Utility functions strictly linear.

Risk Seeking: Person willing to pay more than expected return. Utility functions strictly convex.

Utility Theory

(von Neumann and Morganstern)

If an individual can consistently specify preferencess, then there is a utility function that represents those preferences.

Assumptions:

- Decision maker can consistently value any set of lotteries.
- Can do so with no intransitivities.

Too much to ask of any person, but could be a useful simplification.

Axioms of Utility Theory

Completeness: Decision maker has a preference relation $\succ$ over the set of all lotteries.

Independence: If lottery 1 $\succ$ lottery 2, then for any probability $\alpha \in (0,1]$ and any lottery 3,

$$\alpha L_1 + (1-\alpha)L_3 \succ \alpha L_2 + (1-\alpha)L_3.$$

Continuity: If lottery 1 $\succ$ lottery 2 $\succ$ lottery 3, then there are probabilities $\alpha, \beta \in (0,1)$ such that

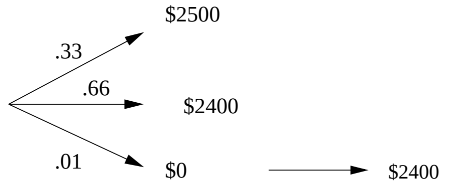
$$\alpha L_1 + (1-\alpha)L_3 \succ L_2 \succ \beta L_1 + (1-\beta)L_3$$

Example: Say lottery 1 is a sure \$100, lottery 2 is a sure \$10, and lottery 3 is sure death. Is continuity reasonable?

Theorem: Binary relation $\succ$ satisfies above axioms iff $\exists U : \mathcal{L} \to \mathbb{R}$ such that $L \succ L' \Leftrightarrow \sum_i p_i U(z_i) > \sum_i {p_i}' U(z_i)$

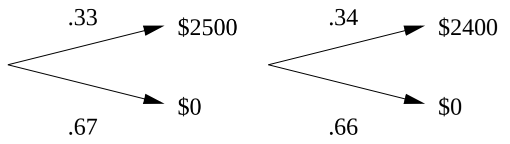
Allais Paradox

Choose between lotteries A and B:



Lottery A Lottery B

Choose between lotteries C and D:



Lottery C Lottery D

Why Allais Is Paradoxical

Most people prefer lotteries B and C. This violates Expected Utility Theory.

Let U be the decision maker's utility function. For convenience, let $U(0) = 0$.

The first preference ($L_B \succ L_A$) specifies:

$$U(2400) > .33U(2500) + .66U(2400)$$

$$\Rightarrow .34U(2400) > .33U(2500).$$

But the second preference ($L_C \succ L_D$) specifies:

$$.33U(2500) + .67U(0) > .34U(2400) + .66U(0)$$

$$\Rightarrow .33U(2500 > .34U(2400).$$

Other Problems with Utility Theory

It's difficult to get and use personal utility functions. Often assume utility function is some "nice" concave (like log) or linear.

Expected Utility Theory widely accepted as descriptive of rational choice under risk, but peoples' choices often violate it.

Another violation: Risk averse in positive domain tend to be risk seekers in negative domain. E.g., prefer 80% risk of losing $4000 to sure loss of $3000.

Other Approaches to Risk

Many non-von Neumann-Morganstern optimality criteria.

One: utility function $U(k_1(L), k_2(L))$ where k_i are *characteristic parameters* of the underlying probability distribution of L.

Usually, $k_1(L)$ is the expected payoff and $k_2(L)$ is a measure of risk.

Most common k_2 is standard deviation. Then this method is called the Mean-Variance criterion and $U(L) = \mathbf{E}(L) - \delta * \sigma(L)$.

Between Uncertainty and Risk

How can we use partial information?

Bayesian approach: try to transform uncertainties into risks (probabilities).

Prior Distribution: Assigned probability model.

When new sample is obtained, update prior using Bayes' rule for conditional probabilities.

Posterior: The result of such an update.

Assigning a Prior

We don't have complete information. How do we come up with a prior? Two main schools of probability theory:

Objectivists: Sample the "true" underlying distribution, then use empirical distribution, maximum likelihood, maximum entropy, ... to recover a prior.

If no samples, start with even or random probability distribution—further sampling will lead to convergence to "true" distribution.

Subjectivists: Set of qualitative beliefs based on personal experience and consistent with a certain set of axioms.

Personal preference relation over set of probability distributions, probability prior and utility function.

Deep Conclusion

The moral of the story:

- There are a lot of ways to measure the quality of choices in strict uncertainty, pure risk, and mixed environments.
- No one way is better in any of the environments. You have to make an educated choice based on the characteristics of the criteria.
- Life is tough!