Information-Theoretic Bounded Rationality

Pedro A. Ortega

Penn
University of Pennsylvania

September 26, 2013

Outline

Perfect Rationality

Motivation

The mathematical foundation of

- economics,
- artificial intelligence,
- and control

is the **theory of subjective expected utility** (SEU), leading to the **maximum subjective expected utility principle** [Savage 1954].

The Maximum SEU Principle

Simply stated, SEU theory says that, given:

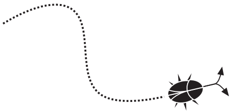
- a set of **policies** Π,
- a set of **outcomes** $\mathcal{X}$,
- a **utility** function $U : \mathcal{X} \to \mathbb{R}$,
- and **beliefs** $P(x|\pi)$ of outcomes x given a policies π,

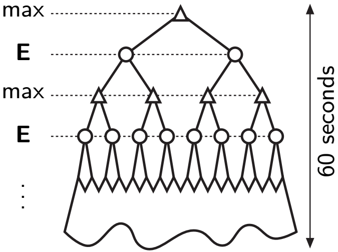
a **rational decision maker** chooses π^* as

$$\pi^* = \arg\max_{\pi \in \Pi} \sum_x P(x|\pi) U(x).$$

Designing a Rational Robobug

Objective: construct policy of simple robot using maximum expected utility principle (anything else is irrational!)





- 1 minute at 2 interactions/second, 2 actions, 2 observations
- nodes of decision tree:

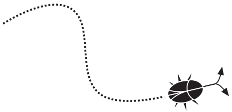
$$\frac{4^{120}-1}{4-1} \approx 5.89 \cdot 10^{71}$$

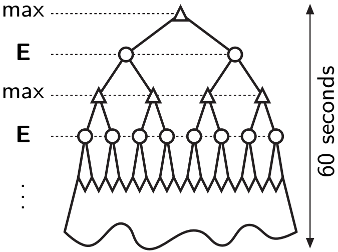
- computation time at 10^{16} FLOPS:

$$1.12 \cdot 10^{50} \text{ years}$$

Designing a Rational Robobug

Objective: construct policy of simple robot using maximum expected utility principle (anything else is irrational!)





- 1 minute at 2 interactions/second, 2 actions, 2 observations
- nodes of decision tree:

$$\frac{4^{120}-1}{4-1} \approx 5.89 \cdot 10^{71}$$

- **atoms in the world:** 10^{50}
- computation time at 10^{16} FLOPS:

$$1.12 \cdot 10^{50} \text{ years}$$

- **age of the universe:** $1.37 \cdot 10^{10}$

Problems of the Maximum Expected Utility Principle

3 Central Problems:

1. **Intractability** of policy search
   or "no approximations allowed"
2. **Causal precedence** of policy choice
   or "no delay of policy choice"
3. **Self-Contradictory** as a design principle
   or "no unrationalized design choices"

Problems of the Maximum Expected Utility Principle

3 Central Problems:

1. **Intractability** of policy search
   or "no approximations allowed"
2. **Causal precedence** of policy choice
   or "no delay of policy choice"
3. **Self-Contradictory** as a design principle
   or "no unrationalized design choices"

$\Rightarrow$ The very theory of rationality has a **intrinsic bottleneck**!

Shortcomings of Subjective Expected Utility Theory

1. Behavioral inconsistencies

2. Problems as a normative theory

Shortcomings of Subjective Expected Utility Theory

1. Behavioral inconsistencies
    - 1954: Allais' showed systematic violation of EU.
    - 1961: Ellsberg showed systematic violation of SEU.
        - Revives Knight's distinction between **risk** and **ambiguities**.
    - '60s & '70s: more inconsistencies appeared (see Machina) . . .
    - 1979 & 1992: Kahnemann & Tversky propose **prospect theory** and **cumulative prospect theory**), solving most of the behavioral inconsistencies.
2. Problems as a normative theory

Shortcomings of Subjective Expected Utility Theory

1. Behavioral inconsistencies

2. Problems as a normative theory
   - 1957: Simon proposes **bounded rationality**.
   - 1989: Russell (& Wefald) — Metareasoning.
     - Points out that bounded rationality is one of the most important open problems in AI.
   - 1998: Rubinstein — Bounded rationality.
   - 2006: Gigerenzer — Heuristics.
   - 2007: Hansen & Sargent — Robustness.
   - –today: AI & control based on **approximations** of SEU.
   - –today: **No widely-accepted theory** of bounded rationality.

Bounded Rationality

Caveat: Metareasoning does not work!

Straightforward solution: penalize choice costs [e.g. Russell]

- desired behavior:

$$U(x)$$

- reasoning about costs:

$$U'(x, \pi) := U(x) - C(\pi)$$

Caveat: Metareasoning does not work!

Straightforward solution: penalize choice costs [e.g. Russell]

- desired behavior:

$$U(x)$$

- reasoning about costs:

$$U'(x, \pi) := U(x) - C(\pi)$$

- reasoning about costs of costs:

$$U''(x, \pi, \pi') := U'(x, \pi) - C'(\pi')$$

Caveat: Metareasoning does not work!

Straightforward solution: penalize choice costs [e.g. Russell]

- desired behavior:

$$U(x)$$

- reasoning about costs:

$$U'(x,\pi) := U(x) - C(\pi)$$

- reasoning about costs of costs:

$$U''(x,\pi,\pi') := U'(x,\pi) - C'(\pi')$$

- and so on...

Caveat: Metareasoning does not work!

Problem of metareasoning:

- Unbounded metalevels + growing solution spaces:

$$\mathcal{X} \to (\mathcal{X} \times \Pi) \to (\mathcal{X} \times \Pi \times \Pi') \to \cdots$$

- We have to accept that metareasoning is **not permitted**.

Caveat: Metareasoning does not work!

Problem of metareasoning:

- Unbounded metalevels + growing solution spaces:

$$\mathcal{X} \to (\mathcal{X} \times \Pi) \to (\mathcal{X} \times \Pi \times \Pi') \to \cdots$$

- We have to accept that metareasoning is **not permitted**.
- Solution: **Interrupted Decisions**.

Interrupted Decisions

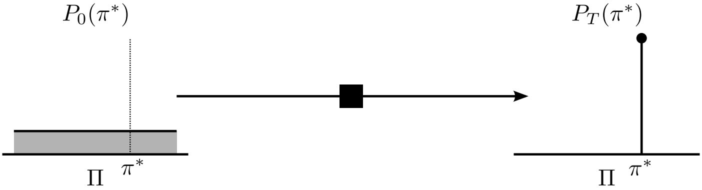
Q: Given that the decision maker is not allowed to reason about his own resources, how do we capture the notion of **boundedness**?

A: Assumption: Computation transforms uncertainty into certainty.

Interrupted Decisions

Q: Given that the decision maker is not allowed to reason about his own resources, how do we capture the notion of **boundedness**?

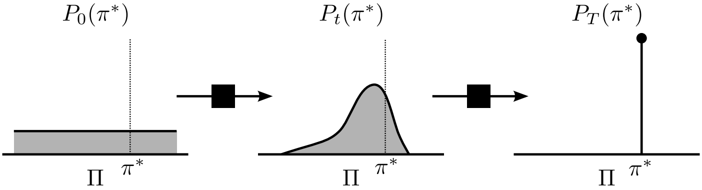
A: Assumption: Computation transforms uncertainty into certainty.



Interrupted Decisions

Q: Given that the decision maker is not allowed to reason about his own resources, how do we capture the notion of **boundedness**?

A: Assumption: Computation transforms uncertainty into certainty.



Computation is **interrupted** $\rightarrow$ uncertainty is **reduced**, but not eliminated!

Information-Theoretic Bounded Rationality

Question: How do we **characterize** behavior when the decision maker is **bounded rational**, i.e. when his **processing resources are limited**?

Information-Theoretic Bounded Rationality

Question: How do we **characterize** behavior when the decision maker is **bounded rational**, i.e. when his **processing resources are limited**?

Our Answer: A **bounded rational** decision maker **can be thought of** as maximizing the functional

$$\sum_x p(x) \left\{ U(x) - \frac{1}{\alpha} \log \frac{p(x)}{q(x)} \right\}.$$

Information-Theoretic Bounded Rationality

Question: How do we **characterize** behavior when the decision maker is **bounded rational**, i.e. when his **processing resources are limited**?

Our Answer: A **bounded rational** decision maker **can be thought of** as maximizing the functional

$$\sum_x p(x) \left\{ U(x) - \frac{1}{\alpha} \log \frac{p(x)}{q(x)} \right\}.$$

Why? Result is based on an information-theoretic assumption about transformation costs, i.e. the cost of "changing".

Transformation Costs

The Cost of Transformations

Our Fundamental Assumption: The difficulty of producing an event determines its probability ("probabilities encode costs").

The Cost of Transformations

Our Fundamental Assumption: The difficulty of producing an event determines its probability ("probabilities encode costs").

Examples:

- Biologists infer behavior from anatomy. Energy-efficient behavior is more frequent than energy-inefficient behavior.
- Conversely, engineers design systems such that desirable behavior is cheaper than undesirable behavior.
- Every action/observation/interaction of a system **necessarily** transforms its **information state**, simply because "before" and "after" are distinguishable!

The Cost of Transformations

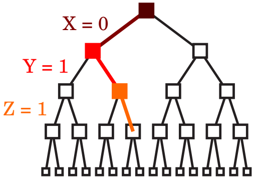
Our Fundamental Assumption: The difficulty of producing an event determines its probability ("probabilities encode costs").

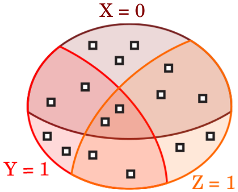
Examples:

- Biologists infer behavior from anatomy. Energy-efficient behavior is more frequent than energy-inefficient behavior.
- Conversely, engineers design systems such that desirable behavior is cheaper than undesirable behavior.
- Every action/observation/interaction of a system **necessarily** transforms its **information state**, simply because "before" and "after" are distinguishable!

What is a Transformation? Chemical reaction, memory update, consulting a random number generator, changing location...

Measure-Theoretic Formalization of Transformations





- **Sequential realizations** are modeled as **filtrations**.
- An **information state** is a measurable set.
- A transformation is a **condition** on the information state:

$$\begin{array}{lccc} \text{State:} & A & & (A \cap B) \\ & & \longrightarrow \text{“}B \text{ is true”} \longrightarrow & \\ \text{Measure:} & P(S|A) & & P(S|A \cap B) \end{array}$$

Axioms of Transformation Costs

Let

- (Ω, Σ) measurable space;
- $P(\cdot|\cdot) : (\Omega \times \Omega) \to [0, 1]$ conditional probability measure.

Then $\rho(\cdot|\cdot) : (\Sigma \times \Sigma) \to \mathbb{R}^+$ is a **transformation cost function** iff

A1. $\rho(A|B) = f(P(A|B))$ for some real-valued, continuous f,

A2. $\rho(A \cap B|C) = \rho(B|C) + \rho(A|B \cap C)$ (additive),

A3. $\rho(A|B) > \rho(C|D) \Leftrightarrow P(A|B) < P(C|D)$ (monotonic),

for all $A, B, C, D \in \Sigma$.

Axioms of Transformation Costs

Let

- (Ω, Σ) measurable space;
- $P(\cdot|\cdot) : (\Omega \times \Omega) \to [0, 1]$ conditional probability measure.

Then $\rho(\cdot|\cdot) : (\Sigma \times \Sigma) \to \mathbb{R}^+$ is a **transformation cost function** iff

A1. $\rho(A|B) = f(P(A|B))$ for some real-valued, continuous f,

A2. $\rho(A \cap B|C) = \rho(B|C) + \rho(A|B \cap C)$ (additive),

A3. $\rho(A|B) > \rho(C|D) \Leftrightarrow P(A|B) < P(C|D)$ (monotonic),

for all $A, B, C, D \in \Sigma$.

Solution: ρ must be

$$\rho(A|B) = -\frac{1}{\alpha} \log P(A|B), \qquad \alpha > 0.$$

What have we achieved?

The relation

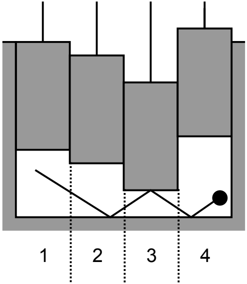
$$\rho(A|B) = -\frac{1}{\alpha} \log P(A|B)$$

establishes a conversion between the cost of changing an information state and conditional probabilities.

More importantly, this will serve as a **primitive** to establish a relation between utility and information.

Variational Principle

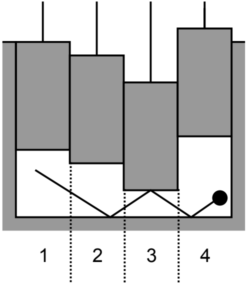
Thermodynamic Example



Knowledge Determines Work:

- Particle bounces **uniformly**.
- Move each piston to control **probabilities**.

Thermodynamic Example



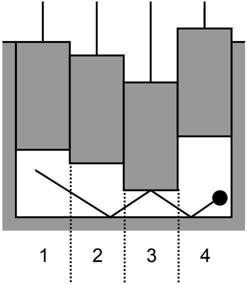
Knowledge Determines Work:

- Particle bounces **uniformly**.
- Move each piston to control **probabilities**.
- The system does isothermal **work**:

$$W = -\gamma \ln \frac{V}{V'}, \qquad \gamma > 0.$$

because **we don't know** where the particle is!

Thermodynamic Example



Knowledge Determines Work:

- Particle bounces **uniformly**.
- Move each piston to control **probabilities**.
- The system does isothermal **work**:

$$W = -\gamma \ln \frac{V}{V'}, \qquad \gamma > 0.$$

  because **we don't know** where the particle is!
- **If we knew** the location x of the particle, then the work would be

$$w(x) = -\gamma \ln \frac{v(x)}{v'(x)}.$$

Thermodynamic Example II

Define:

$$W = -\gamma \ln \frac{V}{V'} \qquad w(x) = -\gamma \ln \frac{v(x)}{v'(x)} \qquad p(x) = \frac{v'(x)}{V'} \qquad q(x) = \frac{v(x)}{V}$$

Total Work:

$$\begin{aligned} W &= -\gamma \ln \frac{V}{V'} \\ &= -\gamma \sum_x \frac{v'(x)}{V'} \ln\left\{ \frac{V}{V'} \cdot \frac{v(x)}{v(x)} \cdot \frac{v'(x)}{v'(x)} \right\} \\ &= -\sum_x \frac{v'(x)}{V'} \left\{ \gamma \ln \frac{v(x)}{v'(x)} \right\} - \gamma \sum_x \frac{v'(x)}{V'} \ln\left\{ \frac{v'(x)}{V'} \Big/ \frac{v(x)}{V} \right\} \\ &= \mathbf{E}_p[w(x)] - \gamma D(p \| q) \end{aligned}$$

Thermodynamic Example III

The First Law:

$$\begin{array}{ccccc} \mathbf{E}_p[w(x)] & = & W & + & \gamma D(p\|q) \\ \{\text{Expected Work}\} & = & \{\text{Work}\} & + & \{\text{Waste}\} \\ \Delta U & = & W & + & Q \end{array}$$

Interpretation:

1. **Knowledge determines** the amount of work.
2. Every change in expected work comes with a **loss**.
3. Total work is the **negative free energy difference** (NFED).

Thermodynamic Example IV

The Variational Principle:

- The transformation from $q(x)$ to $p(x)$ can be **thought of as** maximizing the work with fixed local work $w(x)$:

$$\forall \tilde{p}, \qquad \mathbf{E}_{\tilde{p}}[w(x)] - \gamma D(\tilde{p} \| q) \quad \leq \quad \mathbf{E}_{p}[w(x)] - \gamma D(p \| q)$$

which is concave in $\tilde{p}$.

Thermodynamic Example IV

The Variational Principle:

- The transformation from $q(x)$ to $p(x)$ can be **thought of as** maximizing the work with fixed local work $w(x)$:

$$\forall \tilde{p}, \qquad \mathbf{E}_{\tilde{p}}[w(x)] - \gamma D(\tilde{p} \| q) \quad \leq \quad \mathbf{E}_{p}[w(x)] - \gamma D(p \| q)$$

  which is concave in $\tilde{p}$.

- Equivalently,

$$\forall \tilde{p}, \qquad \mathbf{E}_{\tilde{p}}[U(x)] - \gamma D(\tilde{p} \| q) \quad \leq \quad \mathbf{E}_{p}[U(x)] - \gamma D(p \| q)$$

  where $U(x) := w(x) + C$.

Thermodynamic Example IV

The Variational Principle:

- The transformation from $q(x)$ to $p(x)$ can be **thought of as** maximizing the work with fixed local work $w(x)$:

$$\forall \tilde{p}, \qquad \mathbf{E}_{\tilde{p}}[w(x)] - \gamma D(\tilde{p} \| q) \quad \leq \quad \mathbf{E}_{p}[w(x)] - \gamma D(p \| q)$$

  which is concave in $\tilde{p}$.

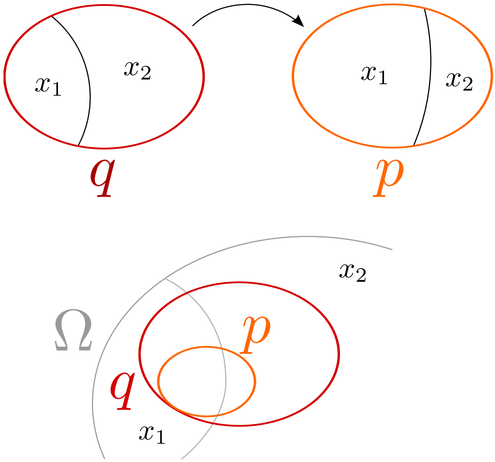
- Equivalently,

$$\forall \tilde{p}, \qquad \mathbf{E}_{\tilde{p}}[U(x)] - \gamma D(\tilde{p} \| q) \quad \leq \quad \mathbf{E}_{p}[U(x)] - \gamma D(p \| q)$$

  where $U(x) := w(x) + C$.

- Only the differences in the $U(x)$'s matter!

The Abstract Formulation



The Abstract Formulation II

Transformation Cost $q \rightarrow p$:

$$\rho(p|q) = \sum_x P(x|p \cap q)\rho(x \cap p \cap q|x \cap q) + \frac{1}{\alpha}\sum_i P(x|p \cap q) \log \frac{P(x|p \cap q)}{P(x|q)}$$
$$= \sum_x p(x)\varrho(x) + \frac{1}{\alpha}\sum_x p(x) \log \frac{p(x)}{q(x)},$$

where

$$p(x) := P(x|p \cap q) \qquad q(x) := P(x|q) \qquad \varrho(x) := \rho(x \cap p \cap q|x \cap q).$$

The Abstract Formulation II

Transformation Cost $q \to p$:

$$\rho(p|q) = \sum_x P(x|p \cap q)\rho(x \cap p \cap q|x \cap q) + \frac{1}{\alpha}\sum_i P(x|p \cap q) \log \frac{P(x|p \cap q)}{P(x|q)}$$
$$= \sum_x p(x)\varrho(x) + \frac{1}{\alpha}\sum_x p(x) \log \frac{p(x)}{q(x)},$$

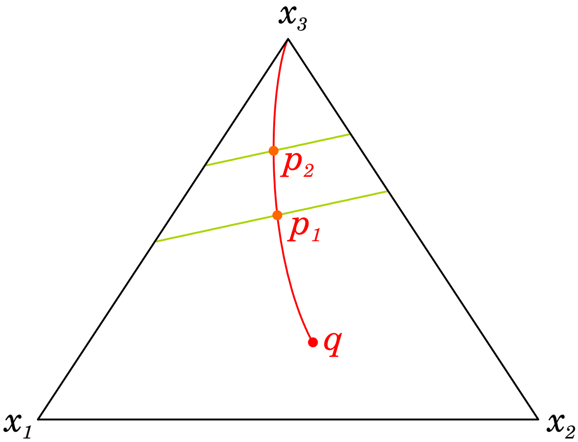
where

$$p(x) := P(x|p \cap q) \qquad q(x) := P(x|q) \qquad \varrho(x) := \rho(x \cap p \cap q|x \cap q).$$

Variational Principle:

$$\forall \tilde{p}, \quad \sum_x \tilde{p}(x)U(x) - \frac{1}{\alpha}\sum_x \tilde{p}(x) \log \frac{\tilde{p}(x)}{q(x)} \leq \sum_x p(x)U(x) - \frac{1}{\alpha}\sum_x p(x) \log \frac{p(x)}{q(x)}$$

Information-Geometric Picture



Sequential Decisions

Free Energy Principle

Let q be a probability distribution and U be a real-valued utility over $\mathcal{X}$. Given $\alpha \in \mathbb{R}$, the **negative free energy difference (NFED)** is given by

$$-\Delta F_{\alpha}[p] := \sum_{x} p(x) U(x) - \frac{1}{\alpha} \sum_{x} P(x) \log \frac{p(x)}{q(x)}.$$

Interpretation

- NFED $=$ expected utility - transformation costs
- models **net utility gain** obtained in transforming q into p
- relative entropy models information content of transformation
- inverse temperature α models (transformation-) bits per utile
- higher inverse temperature $\longrightarrow$ higher net utility gain

Equilibrium Distribution

The solution to the NFED is the **equilibrium distribution**

$$p(x) = \frac{1}{Z(\alpha)} q(x) \exp\{\alpha U(x)\},$$

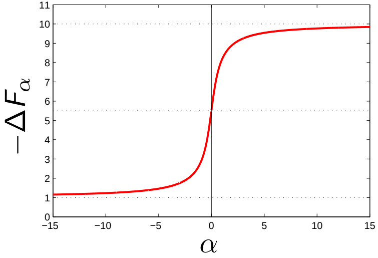
where $Z(\alpha)$ is the **partition function**

$$Z(\alpha) = \sum_x q(x) \exp\{\alpha U(x)\}.$$

The NFED extremum is the **certainty equivalent**

$$\frac{1}{\alpha} \log Z(\alpha) = \frac{1}{\alpha} \log \left(\sum_x q(x) \exp \left\{ \alpha U(x) \right\} \right).$$

Certainty Equivalent



The inverse temperature α parameterizes the **degree of control**:

$$\alpha \to \infty \quad : \quad \frac{1}{\alpha}\log Z \longrightarrow \max U(x) \quad \text{(maximum)}$$

$$\alpha \to 0 \quad : \quad \frac{1}{\alpha}\log Z \longrightarrow \mathbf{E}_x[U(x)] \quad \text{(expectation)}$$

$$\alpha \to -\infty \quad : \quad \frac{1}{\alpha}\log Z \longrightarrow \min U(x) \quad \text{(minimum)}$$

Operational Interpretation of Free Energy

What does it mean to solve the NFED?

- *Given:* prior $q(x)$, utility $U(x)$, and inverse temperature α.
- *Problem:* Obtain a sample from $p(x)$. (This is what "solving" means!)
- Dramatically different from classical decision making: we do not have to check all the outcomes!

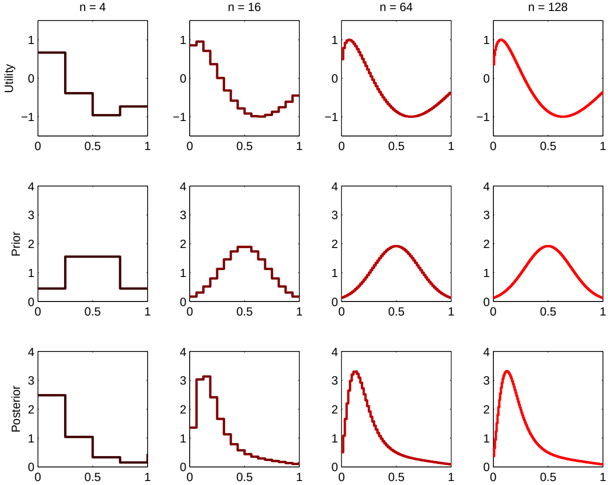
Algorithm:

1. Obtain sample $x' \sim q(x)$.
2. Accept x' with probability

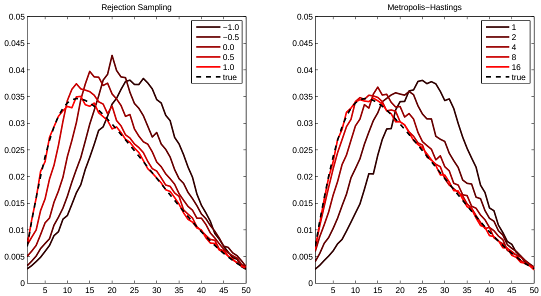
$$A(x'|x) = \exp\{\alpha(U(x') - U(x))\}.$$

3. Rejection sampling: Compare against target utility U^*.
4. Metropolis-Hastings: Compare against sample from last iteration.

Equilibrium Distribution

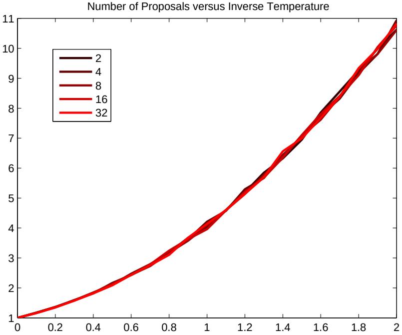


Monte Carlo Simulation of Equilibrium Distribution

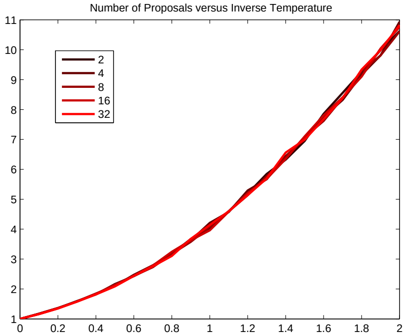


- *Rejection Sampling:* If target value is larger or equal than maximum utility, then samples come from equilibrium distribution.
- *Metropolis-Hastings:* If chain is "long enough", then samples come from equilibrium distribution.

Number of Proposals for Different Number of Outcomes



Number of Proposals for Different Number of Outcomes



The number of proposals depends on the inverse temperature, not on the number of outcomes!

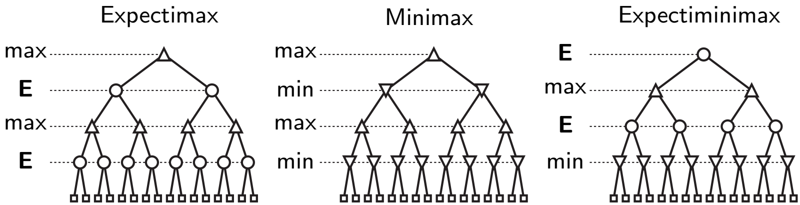
Number of Proposals

Theorem:

Let $\delta > 0$ be a precision. The number of proposals in rejection sampling needed for an acceptance probability $(1 - \delta)$ is

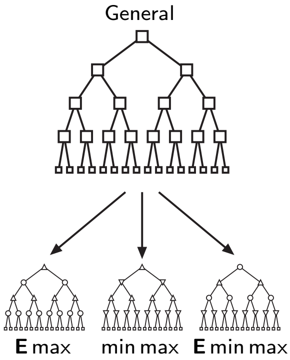
$$n = \frac{\log \delta}{\log(1 - p_\alpha)}, \qquad \text{where} \qquad p_\alpha = \frac{Z_\alpha}{\exp \alpha U^*}.$$

Decision Trees



- Sequential decision problems are stated as decision trees and solved using backward induction.
- Decision rules depend on system: stochastic, cooperative, competitive, hybrid, . . .
- This intuitive distinction between "types of systems" is formally unsatisfactory.
- Decision rules can be reexpressed in a unified way using the free energy functional.

Goal: Generalized Decision Trees



- Different operators express different degrees of control (DoCs):
  - max ⇔ full control
  - **E** ⇔ no control
  - min ⇔ full anti-control
- Goal: Find a generalized operator □ that expresses
  - the 3 classical DoCs,
  - \+ all the other DoCs in between.

Equivalent Lotteries

Definition

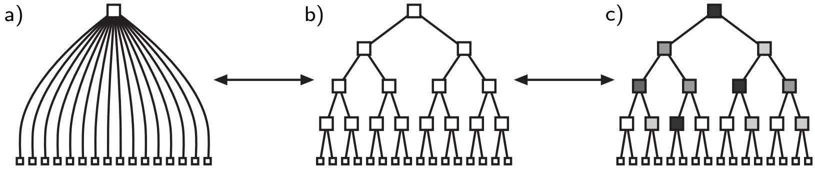
Two lotteries are said to be equivalent iff they have the same prior q, posterior p, and the same certainty equivalent.

Theorem

Let p be the equilibrium distribution given α, U and q.
If *α changes to β with fixed p and q,* **then** *U changes to V:*

$$\beta\Big(V(x) - \tfrac{1}{\beta}\log Z_\beta\Big) = \alpha\Big(U(x) - \tfrac{1}{\alpha}\log Z_\alpha\Big).$$

Construction of Generalized Decision Trees



a) $q(x), U(x), \alpha$

$$\sum_x p(x)U(x) + \frac{1}{\alpha}\sum_x p(x)\log\frac{p(x)}{q(x)}$$

b) $q(x_t|x_{1:t-1}), S(x_t|x_{1:t}), \alpha$

$$\sum_{x_{\leq T}} p(x_{\leq T})\sum_{t=1}^{T}\left\{S(x_t|x_{<t}) + \frac{1}{\alpha}\log\frac{p(x_t|x_{<t})}{q(x_t|x_{<t})}\right\}$$

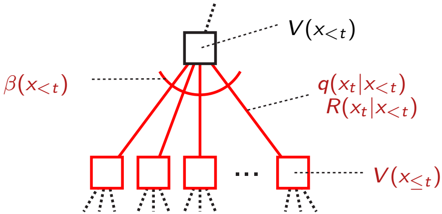
c) $q(x_t|x_{<t}), R(x_t|x_{<t}), \beta(x_{<t})$

$$\sum_{x_{\leq T}} p(x_{\leq T})\sum_{t=1}^{T}\left\{R(x_t|x_{<t}) + \frac{1}{\beta(x_{<t})}\log\frac{p(x_t|x_{<t})}{q(x_t|x_{<t})}\right\}$$

Generalized Optimality Equations

Given

Generalized decision problem $q(x_t|x_{<t})$, $R(x_t|x_{<t})$ and $\beta(x_{<t})$.



Generalized Value/Utility

$$V(x_{<t}) = \frac{1}{\beta(x_{<t})} \log\left\{ \sum_{x_t} q(x_t|x_{<t}) \exp\left\{ \beta(x_{<t}) \left[R(x_t|x_{<t}) + V(x_t) \right] \right\} \right\}$$

Conclusions

Conclusions

1. The free energy principle serves as an **axiomatic foundation** for bounded rational decision-making.
2. It formalizes a **trade-off** between the gains of maximizing the utility and the losses of transformation costs.
3. It establishes clear **links to** information theory and thermodynamics.
4. Inverse temperature **parameterizes** the resource limitations/degree of control.
5. It allows **generalizing** decision trees.

Open Questions

1. What are the **exact** relations to:
   - game theory,
   - search theory,
   - and computational complexity?
2. What are the implications for search algorithms?
3. What are the causal implications?

References

1. Ortega, P.A. and Braun, D.A. *Thermodynamics as a theory of decision-making with information processing costs*. Proceedings of the Royal Society A (20120683), 2013.
2. Ortega, P.A. *Free Energy and the Generalized Optimality Equations*. European Workshop on Reinforcement Learning, 2012.
3. Braun, D.A., Ortega, P.A., Theodorou, E. and Schaal, S. *Path Integral Control and Bounded Rationality*, IEEE Symposium on adaptive dynamic programming and reinforcement learning, pp. 202–209, 2011.
4. Ortega, P.A. *A Unified Framework for Resource-Bounded Agents Interacting with Unknown Environments*, PhD Thesis, Department of Engineering, University of Cambridge, 2011.

References—Related Views

1. Kappen, H.J. *A Linear Theory for Control of Non-Linear Stochastic Systems*, Physical Review Letters 95:200201, 2005.
2. Todorov, E. *Linearly Solvable Markov Decision Problems.* NIPS, 2006.
3. Friston, K. *The Free-Energy Principle: A Unified Brain Theory?* Nat Rev Neurosci, 11:127–38, 2010.
4. Tishby, N. & Polani, D. *Information Theory of Decisions and Actions*, in Perception-Action Cyle, 601–636, Springer 2011.