An Adversarial Interpretation of Information-Theoretic Bounded Rationality

Pedro A. Ortega
GRASP Laboratory
University of Pennsylvania
ope@seas.upenn.edu

Daniel D. Lee
GRASP Laboratory
University of Pennsylvania
ddlee@seas.upenn.edu

Abstract

Recently, there has been a growing interest in modelling planning with information constraints. Accordingly, a decision maker maximizes a regularized expected utility known as the free energy, where the regularizer is given by the information divergence from a prior to a posterior choice probability distribution. While this approach can be justified in various ways, most importantly from statistical mechanics and information theory, it is still unclear how it relates to game theory. This connection has been suggested previously in work relating the free energy to risk-sensitive control and to extensive form games. In this work, we present an adversarial interpretation that is equivalent to the free energy optimization problem. The adversary can, by paying an exponential penalty, generate costs that diminish the decision maker's payoffs. It turns out that the optimal strategy of the adversary consists in choosing costs so as to render the decision maker indifferent among its choices, which is a definining property of a Nash equilibrium, thus tightening the connection between free energy optimization and game theory.

1 Introduction

There has been a growing interest in modelling planning and decision making with information constraints. In this paradigm, a bounded-rational decision maker is thought of as extremizing the functional

$$F_\beta[p] := \sum_{x\in\mathcal{X}} p(x)U(x) - \frac{1}{\beta}\sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{p_0(x)}, \tag{1}$$

where $p_0, p \in \Delta(\mathcal{X})$ are a prior and posterior distribution respectively in the simplex $\Delta(X)$ over strategies $\mathcal{X}$, U is a utility function, and where $\beta \in \mathbb{R}$ is a parameter that controls the trade-off between maximizing the expected utility and minimizing the information divergence from the prior to the posterior. The functional (1) is known under various names, the most commonly adopted ones being *KL-control cost* and *free energy*. The particular regularization embodied by (1) has been justified in many ways. The first ones [2, 5] took their inspiration from the maximum entropy principle of statistical mechanics as a way to model systems with noisy control mechanisms. In [12], the information divergence was used as a way to formulate a convex control problem having linear solutions. [6, 7] proposed equation (1) as a way to formalize bounded rationality by deriving it from an axiomatic formalization of transformation costs that render utilities and information commensurable. [10] use (1) as a way to control the loss of information resulting from policy updates to avoid overfitting. [11] gave an information-theoretic justification of (1) that parallels rate-distortion theory. In their interpretation, the optimization problem $\min_p -\beta F_\beta[p]$ represents the minimal expected amount of bits necessary to specify a choice that yields a given expected utility. The utility plays the role of a linear constraint, thereby pointing out the information-geometric nature of (1). Other related justifications arise in the context of computational neuroscience [3, 4] and Quantal Response Equilibrium [14, 15].

Previous results have also suggested that (1) has a relation to game theory. This is seen follows: the extremum of (1) given by

$$\frac{1}{\beta}\log Z_\beta, \qquad \text{where} \quad Z_\beta := \sum_x p_0(x)e^{\beta U(x)}, \tag{2}$$

but now seen *as a function of* β, can be thought of as an interpolation of the maximum, expectation and minimum operator, since

$$\begin{aligned}
\tfrac{1}{\beta}\log Z_\beta &= \max_x\{U(x)\} && \beta \to +\infty \\
\tfrac{1}{\beta}\log Z_\beta &= \quad \mathbf{E}_q[U] && \beta \to \quad 0 \\
\tfrac{1}{\beta}\log Z_\beta &= \min_x\{U(x)\} && \beta \to -\infty.
\end{aligned}$$

This has two important consequences. First, (2) corresponds, in economic jargon, to the *certainty-equivalent*, i.e. the value a risk-sensitive decision-maker is thought to assign to an uncertain choice ([6], also see [13]). Second, in [8], it has been pointed out that the certainty-equivalent generalizes the aggregation operators of *decision trees*, also known as *extensive form games* in the game-theoretic literature. This is important, as decision making in the face of an adversarial, an indifferent, and a cooperative environment, which have previously been treated as unrelated modeling assumptions, can now be treated as particular instances of a single decision rule; the different decision attitudes just correspond to different choices of the parameter β. Despite the identification with the certainty-equivalent, it is as yet unknown how to *derive* (1) *directly* from a game-theoretic scenario, which would elucidate the conditions under which (2) is exact.

2 Results

The main result of our present work is very simple: the maximization of equation (1) is equivalent to

$$\min_C \max_p \sum_x p(x)[U(x) - C(x)] + \sum_x p_0(x)e^{\beta C(x)}. \tag{3}$$

This equation can be interpreted as follows. The decision maker is playing against an adversarial environment. *After* the decision maker chooses its probabilities $p(x)$, the adversarial enviroment can generate costs $C(x)$ reacting to the choice. However, the adversary is not allowed to generate these costs arbitrarily; instead, it pays an exponential penalty to generate them.

Our second result characterizes the solution to this adversarial setup. It turns out that the adversary's best strategy is to choose costs such that

$$U(x) - C(x) = \text{constant}, \tag{4}$$

that is, costs are chosen such that the decision-maker's eventual payoffs are *uniform* over the options.

3 Derivation

3.1 First Claim

Define

$$f[p, C] := \sum_x p(x)[U(x) - C(x)] + \sum_x p_0(x)e^{\beta C(x)}.$$

Since f is continuous, concave in the $p(x)$ and convex in the $C(x)$, then

$$\min_C \max_p f[p, C] = \max_p \min_C f[p, C].$$

Then, minimizing over the costs yields the worst case

$$\frac{\partial f}{\partial C(x)} = 0 \quad \Longrightarrow \quad p(x) = \beta p_0(x)e^{\beta C^*(x)}.$$

This implies that the costs C^* must be chosen as

$$C^*(x) = \frac{1}{\beta} \log \frac{p(x)}{p_0(x)} + \frac{1}{\beta} \log \frac{1}{\beta}.$$

Substituting the worst case costs back into f yields

$$\begin{aligned} f[p, C^*] &= \sum_x p(x)U(x) - \frac{1}{\beta} \sum_x p(x) \Big\{ \log \frac{p(x)}{\beta p_0(x)} \Big\} + \sum_x p_0(x) \Big\{ \frac{p(x)}{\beta p_0(x)} \Big\} \\ &= \sum_x p(x)U(x) - \frac{1}{\beta} \sum_x p(x) \log \frac{p(x)}{p_0(x)} + K \end{aligned}$$

with constant $K = -\frac{1}{\beta}(\log \beta + 1)$. Obviously, the constant can be dropped without changing the maximizing distribution. That is,

$$\arg\max_p f[p, C] - K = \arg\max_p F_\beta[p].$$

This proves that (1) is equivalent to (3). Mathematically, the adversarial term is generated from the entropy term via a generalized Legendre transformation. The convex conjugate of the logarithmic term is related to the exponential penalty for the adversary.

3.2 Second Claim

Define $\mathcal{X}^* \subset \mathcal{X}$ as the subset of elements maximizing the penalized utility, that is for all $x^* \in \mathcal{X}$ and $x \in \mathcal{X}$,

$$U(x^*) - C(x^*) \geq U(x) - C(x). \tag{5}$$

If we now maximize f with respect to the choice probabilities, then the optimal probabilities $p^*(x)$ are given by

$$\frac{\partial f}{\partial p(x)} = 0 \quad \Longrightarrow \quad p^*(x) = \begin{cases} q(x) & \text{if } x \in \mathcal{X}^*, \\ 0 & \text{otherwise,} \end{cases}$$

where q is any distribution over $\mathcal{X}^*$. In other words, the optimal choice probabilities. Given this, the worst case costs $C^*(x)$ are

$$\frac{\partial f[C, p^*]}{\partial C(x)} = 0 \quad \Longrightarrow \quad C^*(x) = \begin{cases} \frac{1}{\beta} \log \frac{q(x)}{\beta p_0(x)} & \text{if } x \in \mathcal{X}^*, \\ -\infty & \text{otherwise.} \end{cases} \tag{6}$$

However, if $\mathcal{X}^* \neq \mathcal{X}$, then we get a contradiction, since

$$U(x^*) - C^*(x^*) \not\geq U(x) - C^*(x)$$

for all $x \notin \mathcal{X}^*$, violating (5). Hence, it must be that for all $x \in \mathcal{X}$,

$$U(x) - C(x) = \text{constant},$$

concluding the proof of our second claim.

4 Discussion

Indifference and Nash Equilibrium. Our results establish an interesting relation to game theory. Equations (4) and (6), immediately yield the indifference relations

$$U(x) - C(x) = \text{constant} \quad \text{and} \quad U(x) - \frac{1}{\beta} \log \frac{p(x)}{p_0(x)} = \text{constant}$$

for all $x \in \mathcal{X}$, where we have used (6) and the fact that it must be that $p^*(x) = q(x)$ to avoid the contradiction pointed out in the previous section. This is a characterization of the solution to the free energy functional, i.e. the equilibrium distribution

$$p^*(x) = \frac{p_0(x) e^{\beta U(x)}}{\sum_{x'} p_0(x') e^{\beta U(x')}}.$$

This is interesting because it is well-known in game theory that a Nash equilibrium is a strategy profile such that each player chooses a (mixed) strategy that renders the others players indifferent to their choices [9]. In our current setup, the adversary chooses costs such that the decision-maker is indifferent between his choices and vice-versa.

Equivalence to General Regularizers. One could argue that the exponential regularizer

$$\sum_x p_0(x) e^{\beta C(x)} \tag{7}$$

is arbitrary and unmotivated. In particular, consider a more general regularizer of the form

$$\sum_x g_x\big(C(x)\big), \tag{8}$$

where each option $x \in \mathcal{X}$ has its own penalization function $g_x : \mathbb{R} \to \mathbb{R}^+$ for the cost $C(x)$. For the functions $g_x(z)$ to intuitively correspond to penalizations, let us assume that they are continuously differentiable, strictly convex and monotonically increasing. Then, the worst-case costs $C^*(x)$ the adversary can choose are given by

$$\frac{\partial f}{\partial C(x)} = 0 \quad \Longrightarrow \quad C^*(x) = \begin{cases} \gamma(x) & \text{if } x \in \mathcal{X}^*, \\ -\infty & \text{otherwise.} \end{cases} \tag{9}$$

where $\gamma(x)$ is the solution to $dg_x/dz = q(x)$. It is immediately seen that this enforces the indifference relation (4). Now, eliminating the contradictions by enforcing $\mathcal{X}^* = \mathcal{X}$, equations (6) and (9) say that

$$p(x) = \beta p_0(x) e^{\beta C^*(x)} \quad \text{and} \quad p(x) = \frac{dg_x}{dz}\big(C^*(x)\big).$$

Hence, equating the latter two and assuming that they have the same optimal costs $C^*(x)$ yields

$$p_0(x) = \frac{1}{\beta} e^{-\beta C^*(x)} \frac{dg_x}{dz}\big(C^*(x)\big); \tag{10}$$

which is to say that the general regularizer (8) is equivalent to the exponential regularizer (7) where the prior weights $p_0(x)$ have been chosen as in (10).

5 Conclusions

In this report, we have presented an adversarial interpretation of the free energy functional, which is readily obtained via a generalized Legendre transformation. In this representation, an adversarial player can choose costs that reduce the decision maker's payoffs, where the adversary is subject to an exponential penalty for its choice of costs. We have shown that the worst-case adversary will select costs that render the decision maker's payoff uniform, in accordance to the Nash equilibrium.

This adversarial interpretation readily suggests novel algorithms for the calculation of the equilibrium distribution, such as the ones inspired by differential game theory [1] and convex programming [16].

References

[1] E. Dockner, S. Jorgensen, N.V Long, and G. Sorger. *Differential Games in Economics and Management Science*. Cambridge University Press, 2001.

[2] W.H. Fleming. Exit probabilities and optimal stochastic control. *Applied Mathematics and Optimization*, 4:329–346, 1977/78.

[3] K. Friston. The free-energy principle: a rough guide to the brain? *Trends in Cognitive Science*, 13:293–301, 2009.

[4] K. Friston. The free-energy principle: a unified brain theory? *Nature Review Neuroscience*, 11:127–138, 2010.

[5] H.J. Kappen. A linear theory for control of non-linear stochastic systems. *Physical Review Letters*, 95:200201, 2005.

[6] P.A. Ortega and D.A. Braun. A conversion between utility and information. In *Proceedings of the third conference on artificial general intelligence*, pages 115–120. Atlantis Press, 2010.

[7] P.A. Ortega and D.A. Braun. Information, utility and bounded rationality. In *Lecture notes on artificial intelligence*, volume 6830, pages 269–274, 2011.

[8] P.A. Ortega and D.A. Braun. Free energy and the generalized optimality equations for sequential decision making. In *European Workshop on Reinforcement Learning (EWRL10)*, 2012.

[9] M.J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1999.

[10] J. Peters, K. Mülling, and Y. Altün. Relative entropy policy search. In *AAAI*, 2010.

[11] N. Tishby and D. Polani. *Perception-Action Cycle*, chapter Information Theory of Decisions and Actions, pages 601–636. Springer New York, 2011.

[12] E. Todorov. Linearly solvable markov decision problems. In *Advances in Neural Information Processing Systems*, volume 19, pages 1369–1376, 2006.

[13] B. van den Broek, W. Wiegerinck, and H.J. Kappen. Risk sensitive path integral control. In *UAI*, pages 615–622, 2010.

[14] D.H. Wolpert. *Complex Engineering Systems*, chapter Information theory - the bridge connecting bounded rational game theory and statistical physics. Perseus Books, 2004.

[15] D.H. Wolpert, M. Harré, E. Olbrich, N. Bertschinger, and J. Jost. Hysteresis effects of changing the parameters of noncooperative games. *Physical Review E*, 85(036102), 2012.

[16] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*, pages 928–936, 2003.