Information, Utility & Bounded Rationality

Pedro A. Ortega and Daniel A. Braun

Department of Engineering, University of Cambridge
Trumpington Street, Cambridge, CB2 1PZ, UK
{dab54,pao32}@cam.ac.uk

Abstract. Perfectly rational decision-makers maximize expected utility, but crucially ignore the resource costs incurred when determining optimal actions. Here we employ an axiomatic framework for bounded rational decision-making based on a thermodynamic interpretation of resource costs as information costs. This leads to a variational "free utility" principle akin to thermodynamical free energy that trades off utility and information costs. We show that bounded optimal control solutions can be derived from this variational principle, which leads in general to stochastic policies. Furthermore, we show that risk-sensitive and robust (minimax) control schemes fall out naturally from this framework if the environment is considered as a bounded rational and perfectly rational opponent, respectively. When resource costs are ignored, the maximum expected utility principle is recovered.

Keywords: bounded rationality, expected utility, risk-sensitivity

1 Introduction

According to the principle of maximum expected utility (MEU), a *perfectly rational* decision-maker chooses its action so as to maximize its expected utility, given a probabilistic model of the environment [18]. In contrast, a *bounded rational* decision-maker trades off the action's expected utility against the computational cost of finding the optimal action [12]. In this paper we employ a previously published axiomatic conversion between utility and information [11] as a basis for a framework for bounded rationality that leads to such a trade-off based on a thermodynamic interpretation of resource costs [5]. The intuition behind this interpretation is that ultimately any real decision-maker has to be incarnated in a thermodynamical system, since any process of information processing must always be accompanied by a pertinent physical process [16]. In the following we conceive of information processing as changes in information states represented by probability distributions in statistical physical systems, where states with different energy correspond to states with different utility [4]. Changing an information state therefore implies changes in physical states, such as flipping gates in a transistor, changing voltage on a microchip, or even changing location of a gas particle. Changing such states is costly and requires thermodynamical work [5]. We will interpret this work as a proxy for resource costs of information processing.

2 Bounded Rationality

Since bounded rational decision-makers need to trade off utility and information costs, the first question is how to translate between information and utility. In canonical systems of statistical mechanics this relationship is given by the Boltzmann distribution that relates the probability $\mathbf{P}$ of a state to its energy $\mathbf{U}$ (utility), thus forming a *conjugate* pair $(\mathbf{P}, \mathbf{U})$. As shown previously, the same relationship can be derived axiomatically in a choice-theoretic context [11], and both formulations satisfy a variational principle [4]:

Theorem 1. *Let X be a random variable with values in $\mathcal{X}$. Let $\mathbf{P}$ and $\mathbf{U}$ be a conjugate pair of probability measure and utility function over X. Define the **free utility** functional as $\mathbf{J}(\mathbf{Pr}; \mathbf{U}) := \sum_{x \in \mathcal{X}} \mathbf{Pr}(x)\mathbf{U}(x) - \alpha \sum_{x \in \mathcal{X}} \mathbf{Pr}(x) \log \mathbf{Pr}(x)$, where $\mathbf{Pr}$ is an arbitrary probability measure over X. Then, $\mathbf{J}(\mathbf{Pr}; \mathbf{U}) \leq \mathbf{J}(\mathbf{P}; \mathbf{U})$ with $\mathbf{P}(X) = \frac{1}{Z} e^{\frac{1}{\alpha}\mathbf{U}(X)}$ and $Z = \sum_{X' \in \mathcal{X}} e^{\frac{1}{\alpha}\mathbf{U}(X')}$.*

A proof can be found in [8]. The constant $\alpha \in \mathbb{R}$ is usually strictly positive, unless one deals with an adversarial agent and it is strictly negative.

The variational principle of the free utility also allows measuring the cost of transforming the state of a stochastic system required for information processing. Consider an initial system described by the conjugate pair $\mathbf{P}_i$ and $\mathbf{U}_i$ and free utility $\mathbf{J}_i(\mathbf{P}_i, \mathbf{U}_i)$. We now want to transform this initial system into another system by adding new constraints represented by the utility function $\mathbf{U}_*$. Then, the resulting utility function $\mathbf{U}_f$ is given by the sum $\mathbf{U}_f = \mathbf{U}_i + \mathbf{U}_*$ and the resulting system has the free utility $\mathbf{J}_f(\mathbf{P}_f, \mathbf{U}_f)$. The difference in free utility is

$$\mathbf{J}_f - \mathbf{J}_i = \sum_{x \in \mathcal{X}} \mathbf{P}_f(x)\mathbf{U}_*(x) - \alpha \sum_{x \in \mathcal{X}} \mathbf{P}_f(x) \log \frac{\mathbf{P}_f(x)}{\mathbf{P}_i(x)}. \tag{1}$$

These two terms can be interpreted as determinants of bounded rational decision-making in that they formalize a trade-off between an expected utility $\mathbf{U}_*$ (first term) and the information cost of transforming $\mathbf{P}_i$ into $\mathbf{P}_f$ (second term). In this interpretation $\mathbf{P}_i$ represents an initial probability or policy, which includes the special case of the uniform distribution where the decision-maker has initially no preferences. Deviations from this initial probability incur an information cost measured by the KL divergence. If this deviation is bounded by a non-zero value, we have a bounded rational agent. This allows formulating a variational principle both for control and estimation:

1. **Control.** Given an initial policy represented by the probability measure $\mathbf{P}_i$ and the constraint utilities $\mathbf{U}_*$, we are looking for the final system $\mathbf{P}_f$ that optimizes the trade-off between utility and resource costs. That is,

$$\mathbf{P}_f = \arg\max_{\mathbf{Pr}} \sum_{x \in \mathcal{X}} \mathbf{Pr}(x)\mathbf{U}_*(x) - \alpha \sum_{x \in \mathcal{X}} \mathbf{Pr}(x) \log \frac{\mathbf{Pr}(x)}{\mathbf{P}_i(x)}. \tag{2}$$

   The solution is given by $\mathbf{P}_f(x) \propto \mathbf{P}_i(x) \exp\left(\frac{1}{\alpha}\mathbf{U}_*(x)\right)$. In particular, at very low temperature $\alpha \approx 0$ we get $\mathbf{J}_f - \mathbf{J}_i \approx \sum_{x \in \mathcal{X}} \mathbf{P}_f(x)\mathbf{U}_*(x)$, and hence

resource costs are ignored in the choice of $\mathbf{P}_f$, leading to $\mathbf{P}_f \approx \delta_{x^*}(x)$, where $x^* = \max_x \mathbf{U}_*(x)$. Similarly, at a high temperature, the difference is $\mathbf{J}_f - \mathbf{J}_i \approx -\alpha \sum_{x \in \mathcal{X}} \mathbf{P}_f(x) \log \frac{\mathbf{P}_f(x)}{\mathbf{P}_i(x)}$, and hence only resource costs matter, leading to $\mathbf{P}_f \approx \mathbf{P}_i$.

2. **Estimation.** Given a final probability measure $\mathbf{P}_f$ that represents the environment and the constraint utilities $\mathbf{U}_*$, we are looking for the initial system $\mathbf{P}_i$ that satisfies

$$\mathbf{P}_i = \arg\max_{\mathbf{Pr}} \sum_{x \in \mathcal{X}} \mathbf{P}_f(x) \mathbf{U}_*(x) - \alpha \sum_{x \in \mathcal{X}} \mathbf{P}_f(x) \log \frac{\mathbf{P}_f(x)}{\mathbf{Pr}(x)} \tag{3}$$

which translates into $\mathbf{P}_i = \arg\min_{\mathbf{Pr}} \sum_{x \in \mathcal{X}} \mathbf{P}_f(x) \log \frac{\mathbf{P}_f(x)}{\mathbf{Pr}(x)}$ and thus we have recovered the minimum relative entropy principle for estimation, having the solution $\mathbf{P}_i = \mathbf{P}_f$. The minimum relative entropy principle for estimation is well-known in the literature as it underlies Bayesian inference [6], but the same principle can also be applied to problems of adaptive control [9, 10, 2].

3 Applications

Consider a system that first emits an action symbol x_1 with probability $P_0(x_1)$ and then expects a subsequent input signal x_2 with probability $P_0(x_2|x_1)$. Now we impose a utility on this decision-maker that is given by $U(x_1)$ for the first symbol and $U(x_2|x_1)$ for the second symbol. How should this system adjust its action probability $P(x_1)$ and expectation $P(x_2|x_1)$? Given the boundedness constraints, the variational problem can be formulated as a nested expression

$$\max_{p(x_1,x_2)} \sum_{x_1} p(x_1) \left[U(x_1) - \alpha \log \frac{p(x_1)}{p_0(x_1)} + \sum_{x_2} p(x_2|x_1) \left[U(x_2|x_1) - \beta \log \frac{p(x_2|x_1)}{p_0(x_2|x_1)} \right] \right].$$

with α and β as Lagrange multipliers. We have then an inner variational problem:

$$\max_{p(x_2|x_1)} \sum_{x_2} p(x_2|x_1) \left[-\beta \log \frac{p(x_2|x_1)}{p_0(x_2|x_1)} + U(x_2|x_1) \right] \tag{4}$$

with the solution

$$p(x_2|x_1) = \frac{1}{Z_2} p_0(x_2|x_1) \exp\left(\frac{1}{\beta} U(x_2|x_1) \right) \tag{5}$$

and the normalization constant $Z_2(x_1) = \sum_{x_2} p_0(x_2|x_1) \exp\left(\frac{1}{\beta} U(x_2|x_1) \right)$ and an outer variational problem

$$\max_{p(x_1)} \sum_{x_1} p(x_1) \left[-\alpha \log \frac{p(x_1)}{p_0(x_1)} + U(x_1) + \beta \log Z_2 \right] \tag{6}$$

with the solution

$$p(x_1) = \frac{1}{Z_1} p_0(x_1) \exp\left(\frac{1}{\alpha} \left(U(x_1) + \beta \log Z_2 \right) \right) \tag{7}$$

and the normalization constant $Z_1 = \sum_{x_1} p_0(x_1) \exp\left(\frac{1}{\alpha}\left(U(x_1) + \beta \log Z_2\right)\right)$. For notational convenience we introduce $\lambda = \frac{1}{\alpha}$ and $\mu = \frac{1}{\beta}$. Depending on the values of λ and μ we can discern the following cases:

1. **Risk-seeking bounded rational agent:** $\lambda > 0$ and $\mu > 0$
   When $\lambda > 0$ the agent is bounded and acts in general stochastically. When $\mu > 0$ the agent considers the move of the environment as if it was his own move (hence "risk-seeking" due to the overtly optimistic view). We can see this from the relationship between Z_1 and Z_2 in (7), if we assume $\mu = \lambda$ and introduce the value function $V_t = \frac{1}{\lambda} \log Z_t$, which results in the recursion

   $$V_{t-1} = \frac{1}{\lambda} \log \sum_{x_{t-1}} P_0(x_{t-1}|\cdot) \exp\left(\lambda \left(U(x_{t-1}|\cdot) + V_t\right)\right).$$

   Similar recursions based on the log-transform have been previously exploited for efficient approximations of optimal control solutions both in the discrete and the continuous domain [3, 7, 15]. In the perfectly rational limit $\lambda \to +\infty$, this recursion becomes the well-known Bellman recursion

   $$V^*_{t-1} = \max_{x_{t-1}} \left(U(x_{t-1}|\cdot) + V^*_t\right)$$

   with $V^*_t = \lim_{\lambda \to +\infty} V_t$.
2. **Risk-neutral perfectly rational agent:** $\lambda \to +\infty$ and $\mu \to 0$
   This is the limit for the standard optimal controller. We can see this from (7) by noting that

   $$\lim_{\mu \to 0} \frac{1}{\mu} \log \sum_{x_2} p_0(x_2|x_1) \exp\left(\mu U(x_2|x_1)\right) = \sum_{x_2} p_0(x_2|x_1) U(x_2|x_1),$$

   which is simply the expected utility. By setting $U(x_1) \equiv 0$, and taking the limit $\lambda \to +\infty$ in (7), we therefore obtain an expected utility maximizer

   $$p(x_1) = \delta(x_1 - x^*_1)$$

   with

   $$x^*_1 = \arg\max_{x_1} \sum_{x_2} p_0(x_2|x_1) U(x_2|x_1).$$

   As discussed previously, action selection becomes deterministic in the perfectly rational limit.
3. **Risk-averse perfectly rational agent:** $\lambda \to +\infty$ and $\mu < 0$
   When $\mu < 0$ the decision-maker assumes a pessimistic view with respect to the environment, as if the environment was an adversarial or malevolent agent. This attitude is sometimes called risk-aversion, because such agents act particularly cautiously to avoid high uncertainty. We can see this from (7) by writing a Taylor series expansion for small μ

   $$\frac{1}{\mu} \log \sum_{x_2} p_0(x_2|x_1) \exp\left(\mu U(x_2|x_1)\right) \approx \mathbb{E}[U] - \frac{\mu}{2} \mathbb{VAR}[U],$$

where higher than second order cumulants have been neglected. The name risk-sensitivity then stems from the fact that variability or uncertainty in the utility of the Taylor series is subtracted from the expected utility. This utility function is typically *assumed* in risk-sensitive control schemes in the literature [19], whereas here it falls out naturally. The perfectly rational actor with risk-sensitivity μ picks the action

$$p(x_1) = \delta(x_1 - x_1^*)$$

with

$$x_1^* = \arg\max_{x_1} \frac{1}{\mu} \log \sum_{x_2} p_0(x_2|x_1) \exp\left(\mu U(x_2|x_1)\right),$$

which can be derived from (7) by setting $U(x_1) \equiv 0$ and by taking the limit $\lambda \to +\infty$. Within the framework proposed in this paper we might also interpret the equations such that the decision-maker considers the environment as an adversarial opponent with bounded rationality μ.

4. **Robust perfectly rational agent:** $\lambda \to +\infty$ and $\mu \to -\infty$
When $\mu \to -\infty$ the decision-maker makes a worst case assumption about the adversarial environment, namely that it is also perfectly rational. This leads to the well-known game-theoretic minimax problem with the solution

$$x_1^* = \arg\max_{x_1} \arg\min_{x_2} U(x_2|x_1),$$

which can be derived from (7) by setting $U(x_1) \equiv 0$, taking the limits $\lambda \to +\infty$ and $\mu \to -\infty$ and by noting that $p(x_1) = \delta(x_1 - x_1^*)$. Minimax problems have been used to reformulate robust control problems that allow controllers to cope with model uncertainties [1]. Robust control problems are also known to be related to risk-sensitive control [1]. Here we derived both control types from the same variational principle.

4 Conclusion

In this paper we have proposed a thermodynamic interpretation of bounded rationality based on a free utility principle. Accordingly, bounded rational agents trade off utility maximization against resource costs measured by the KL divergence with respect to an initial policy. The use of the KL divergence as a cost function for control has been previously proposed to measure deviations from passive dynamics in Markov systems [14, 15]. Other methods of statistical physics have been previously proposed as an information-theoretic approach to interactive learning [13] and to game theory with bounded rational players [20]. The contribution of our study is to devise a single axiomatic framework that allows for the treatment of control problems, game-theoretic problems and estimation and learning problems for perfectly rational and bounded rational agents. In the future it will be interesting to relate the thermodynamic resource costs of bounded rational agents to more traditional notions of resource costs in computer science like space and time requirements when computing optimal actions [17].

References

1. T. Basar and P. Bernhard. *H-Infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Birkhauser Boston, 1995.
2. D. A. Braun and P. A. Ortega. A minimum relative entropy principle for adaptive control in linear quadratic regulators. In *The 7th conference on informatics in control, automation and robotics*, volume 3, pages 103–108, 2010.
3. D.A. Braun, P.A. Ortega, E. Theodorou, and S. Schaal. Path integral control and bounded rationality. In *IEEE symposium on adaptive dynamic programming and reinforcement learning*, 2011.
4. H.B. Callen. *Thermodynamics and an Introduction to Themostatistics*. John Wiley & Sons, 2nd edition, 1985.
5. R. P. Feynman. *The Feynman Lectures on Computation*. Addison-Wesley, 1996.
6. D. Haussler and M. Opper. Mutual information, metric entropy and cumulative relative entropy risk. *The Annals of Statistics*, 25:2451–2492, 1997.
7. B. Kappen. A linear theory for control of non-linear stochastic systems. *Physical Review Letters*, 95:200201, 2005.
8. G. Keller. *Equilibrium States in Ergodic Theory*. London Mathematical Society Student Texts. Cambridge University Press, 1998.
9. P. A. Ortega and D. A. Braun. A minimum relative entropy principle for learning and acting. *Journal of Artificial Intelligence Research*, 38:475–511, 2010.
10. P.A. Ortega and D.A. Braun. A bayesian rule for adaptive control based on causal interventions. In *The third conference on artificial general intelligence*, pages 121–126, Paris, 2010. Atlantis Press.
11. P.A. Ortega and D.A. Braun. A conversion between utility and information. In *The third conference on artificial general intelligence*, pages 115–120, Paris, 2010. Atlantis Press.
12. H Simon. *Models of Bounded Rationality*. MIT Press, 1982.
13. S. Still. An information-theoretic approach to interactive learning. *Europhysics Letters*, 85:28005, 2009.
14. E. Todorov. Linearly solvable markov decision problems. In *Advances in Neural Information Processing Systems*, volume 19, pages 1369–1376, 2006.
15. E. Todorov. Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences U.S.A.*, 106:11478–11483, 2009.
16. M. Tribus and E.C. McIrvine. Energy and information. *Scientific American*, 225:179–188, 1971.
17. P.M.B. Vitanyi. Time, space, and energy in reversible computing. In *Proceedings of the 2nd ACM conference on Computing frontiers*, page 435444, 2005.
18. J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
19. P. Whittle. *Risk-sensitive optimal control*. John Wiley and Sons, 1990.
20. D.H. Wolpert. *Information theory - the bridge connecting bounded rational game theory and statistical physics. In: Complex Engineering Systems. Braha, D. and Bar-Yam, Y. (Eds.)*. Perseus Books, 2004.