Exploring Optimal Risk-Sensitive Behavior in the Balloon Analogue Risk Task (BART)

Xin Sui (xin.sui@tuebingen.mpg.de)
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
University of Tübingen, Tübingen, Germany

Peter Dayan (peter.dayan@tuebingen.mpg.de)
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
University of Tübingen, Tübingen, Germany

Kevin Lloyd (kevin.lloyd@tuebingen.mpg.de)
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Abstract

Attitudes towards risk play a crucial role in everyday decision-making as well as in psychiatric disorders like anxiety. The Balloon Analogue Risk Task (BART) provides a behavioral measure of risk preference that has been widely applied to both clinical and nonclinical populations. However, although most versions of BART involve epistemic as well as aleatoric uncertainty, the implications of this for risk-sensitivity have not been explored. We adopt a prominent theoretical framework for understanding risk, namely conditional value-at-risk (CVaR), to elucidate the effect of risk attitudes on optimal exploration and exploitation in a simple instance of BART. In sequential problems, CVaR comes in two different flavors: pCVaR, which precommits to a level of risk at the very first choice; and nCVaR, which re-applies the same risk level at every step in a nested manner. We show that the structure of stochasticity in the BART is such that pCVaR is more risk-averse than nCVaR in a single trial, for the same nominal risk. We also show that risk preferences and prior expectations interact in risk-sensitive exploration across multiple trials. We hope to provide a normative grounding for a more detailed understanding of behavioral variation in the BART.

Keywords: risk-averse reinforcement learning; exploration; Bayes adaptive Markov decision process (BAMDP); balloon analogue risk task (BART)

Introduction

Risk Sensitivity in Sequential Decisions

Everyday decision making often involves uncertain or probabilistic outcomes, implying a tremendous importance for risk sensitivity. This is also a critical aspect of many psychiatric symptoms, e.g. intolerance to uncertainty (Grupe & Nitschke, 2013; Charpentier et al., 2017), excessive worry (Watkins, 2008), avoidance (Maner & Schmidt, 2006), etc.. Here, we go beyond conventional single-shot and two-step tasks and examine risk sensitive exploration and exploitation (formalized by conditional value-at-risk; CVaR) in longer-range sequential decision-making (exemplified by a simple form of Balloon Analogue Risk Task; BART).

Environment: Balloon Analogue Risk Task (BART)

The Balloon Analogue Risk Task (BART) (Lejuez et al., 2002) is a classic experimental paradigm that generates a behavioral measure of risk attitudes. Human subjects perform a trial of the task by pumping up a virtual balloon. With each pump, the balloon grows larger and the subject accrues a small amount of money, but there is a chance that the balloon bursts, wiping out all the earnings accrued on that trial. At any point before explosion, the subject may choose to stop pumping, cash out, and move to the next trial. Subjects face aleatoric uncertainty from the chance of bursting per step; and epistemic uncertainty if they do not know the distribution governing the chance. We chose the BART in our project as a proof-of-concept environment for its schematic but sequential nature; but simplified it to assume that bursting happens with fixed, but initially unknown, hazard rate or probability θ per pump.

Risk Framework: Conditional Value-at-Risk (CVaR)

Conditional value-at-risk (CVaR) is a common measure in the literature of risk-averse reinforcement learning (Rockafellar & Uryasev, 2002; Gagne & Dayan, 2022). It is defined as the expected value of the outcomes in the lower α-tail of their distribution. When risk preference $\alpha = 1$, the agent is risk-neutral; when $0 < \alpha < 1$, the agent is risk-averse, with a smaller α corresponding to more severe aversion.

Compared to traditional risk frameworks, CVaR offers psychologically attractive qualities by capturing the extremely negative events that seem to motivate risk-avoidant behaviors and anxious thought (Watkins, 2008).

In sequential problems such as the BART, CVaR comes in two different flavors, one which precommits to a level of risk sensitivity at the very first choice ("precommited CVaR"/pCVaR), and the other which re-applies the same level of risk sensitivity at every step in a nested manner ("nested CVaR"/nCVaR). We follow the formalization of pCVaR and nCVaR from Gagne & Dayan (2021; 2022); Chow et al. (2015).

Risk-Sensitive Exploration

The classic trade-off between exploration and exploitation is also inescapable in sequential decisions with epistemic uncertainty and learning. Under the formal framework we establish for risk sensitivity in sequential decisions inherent to the BART, we are especially interested in the implications of risk sensitivity for exploration. What does optimal risk-sensitive exploration look like? And how does it vary from individual to individual, each with their unique risk preferences and priors from past experiences?

Models and Simulation Results

We simulate and analyze exploration behavior of a risk-sensitive agent solving a version of the BART with a fixed hazard rate θ. This assumption is in line with Wallsten et al. (2005), who found that the best model assumes a constant instead of increasing burst probability within a trial, even when the latter is actually true.

For a small-scale simulation as a proof of concept, we cap the number of pumps per trial at 8 and the number of trials per block at 5, i.e. the agent interacts with 5 identical balloons in a row, for each balloon (i.e. trial) choosing between 0 to 8 pumps. We also assume no discounting. The model of exploration behavior in the BART under the CVaR framework is two-fold: within-trial (inner loop) and across-trial (outer loop).

Within-Trial: CVaR Optimization

We model the within-trial behavior as a single-shot optimal policy under CVaR, i.e. no pump-by-pump belief updating and therefore exploration being important only between trials. This assumption is again taken from the best model in Wallsten et al. (2005), in which the number of pumps is already decided at the start of a trial. **Fig 1 (a)** shows that, unless α =

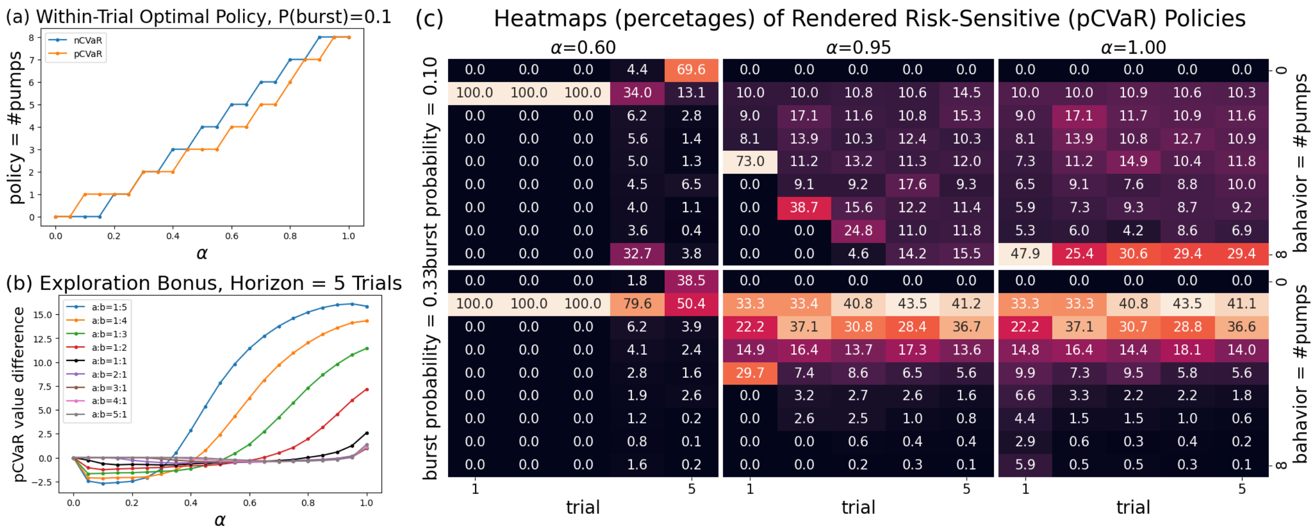

Figure 1: Simulation Results. **(a)** pCVaR vs. nCVaR within-trial optimal policy for all levels of risk-preference α when the balloon's burst probability is known to be 0.1 (as an example; the direction of inequality between pCVaR vs. nCVaR is preserved regardless of this value). **(b)** Exploration bonus for all levels of risk-preference pCVaR$_\alpha$. Each line plot corresponds to a different prior (a and b refer to parameters of Beta distribution); this plot shows the difference between the risk-adjusted value of the initial state for, e.g., $a = 1; b = 1$ versus $a = 30; b = 30$. **(c)** Rendered risk-sensitive (pCVaR) policies. Each of the six heatmaps corresponds to one agent's behavior distribution (in percentages annotated as numbers in the heatmap cells) for every trial of the block. The agent's risk preference α and the balloon's burst probability varies across heatmaps.

1 (risk-neutral), the pCVaR optimal policy given perfect knowledge of θ (which makes the BART a standard Markov decision process or MDP; Sutton & Barto (2018)) is more conservative than the nCVaR optimal policy. This is the opposite of the examples shown in Gagne & Dayan (2022). This is a result of the type of progressive adjustment to risk preferences in pCVaR akin to a justified gambler's fallacy (Chen et al., 2016), in which risk-aversion increases following advantageous stochastic samples. This is because the stochasticity structure of the BART only allows for fortunate state transitions to happen in an unfinished trial (i.e. "surviving a pump" instead of "bursting the balloon"), which keeps adjusting α downwards and making the agent more risk-averse.

Across-Trial: Bayes Adaptive Markov decision process (BAMDP)

We model the across-trial exploration behavior as a risk-sensitive optimal policy in a Bayes adaptive Markov decision process (BAMDP) (Duff, 2002). Bayes adaptivity captures the agent's initial ignorance about the balloon's value of θ and trial-to-trial learning (through belief updating) about this quantity. The agent's belief states are modeled by Beta distributions parameterized by a and b. The BAMDP is again subject to CVaR so as to simulate risk-sensitive exploration across-trial.

Combined: Exploration Bonus and Rendered Policy

Fig 1 (b) visualizes the exploration bonus, which is the difference between the starting belief state value function associated with an uncertain vs. more certain prior of the same mean, e.g. Beta(1,1) vs. Beta(30,30). **Fig 1 (c)** depicts the simulated behavior of agents enjoying various risk preferences ($\alpha = 0.6; 0.9; 1.0$) and interacting with balloons of different explosive properties ($\theta = 0.1; 0.33$). Both figures show rich interactions between risk preferences and priors.

Discussion

Epistemic uncertainty generates risk that needs to be quantified and accommodated to predict and understand behavior in normal and psychiatric populations. We present preliminary results examining this in the context of the BART task and CVaR. Our next steps are to study the effects of nCVaR on exploration, to accommodate risk seeking as well as risk aversion (by focusing on upper rather than lower tails of the outcome distribution) and, by fitting our models to human behavioral data, to probe links between individual differences in risk preference and related psychiatric symptoms.

Acknowledgments

This work was supported by funding from the Max Planck Society and the Humboldt Foundation.

References

Charpentier, C. J., Aylward, J., Roiser, J. P., & Robinson, O. J. (2017). Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. *Biological psychiatry*, *81*(12), 1014–1022.

Chen, D. L., Moskowitz, T. J., & Shue, K. (2016). Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics*, *131*(3), 1181–1242.

Chow, Y., Tamar, A., Mannor, S., & Pavone, M. (2015). Risk-sensitive and robust decision-making: a cvar optimization approach. *Advances in neural information processing systems*, *28*.

Duff, M. O. (2002). *Optimal learning: Computational procedures for bayes-adaptive markov decision processes*. University of Massachusetts Amherst.

Gagne, C., & Dayan, P. (2021). Two steps to risk sensitivity. *Advances in Neural Information Processing Systems*, *34*, 22209–22220.

Gagne, C., & Dayan, P. (2022). Peril, prudence and planning as risk, avoidance and worry. *Journal of Mathematical Psychology*, *106*, 102617.

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, *14*(7), 488–501.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (bart). *Journal of Experimental Psychology: Applied*, *8*(2), 75.

Maner, J. K., & Schmidt, N. B. (2006). The role of risk avoidance in anxiety. *Behavior therapy*, *37*(2), 181–189.

Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, *26*(7), 1443–1471.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Wallsten, T. S., Pleskac, T. J., & Lejuez, C. W. (2005). Modeling behavior in a clinically diagnostic sequential risk-taking task. *Psychological review*, *112*(4), 862.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological bulletin*, *134*(2), 163.