

SocultPaperV2

2024-05-19

Introduction

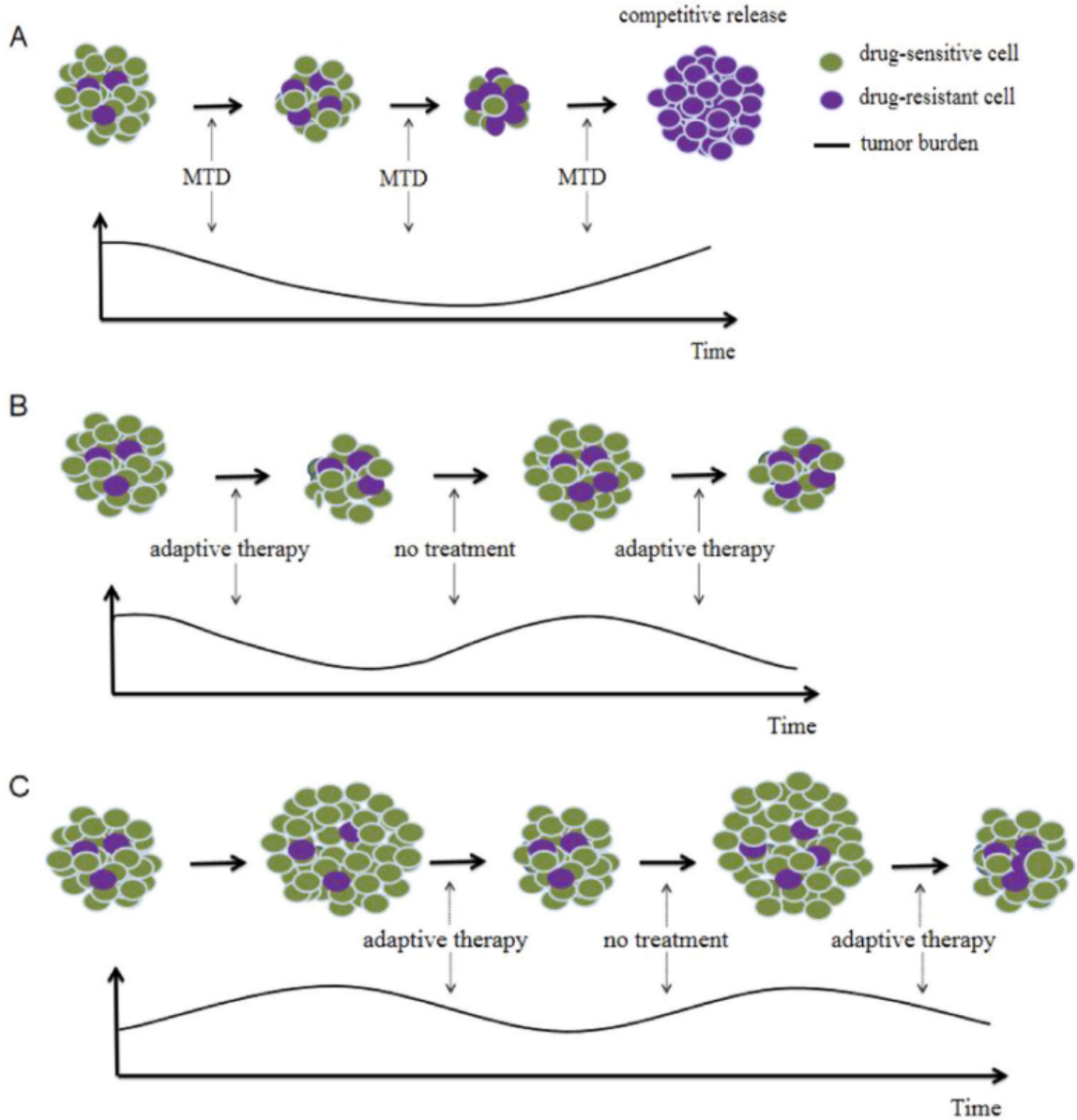
Why we should POMDPs with FEP for Adaptive cancer therapy

What is the adaptive cancer therapy

Worldwide, cancer is the cause of 1 out of 6 deaths. Of these cancer deaths an estimated 90% are due to development of drug resistance (Bukowski, Kciuk, and Kontek 2020). While initial cancer treatment usually shows positive response in tumor burden, drug resistance develops due. To highlight the inefficiency of traditional approaches, (Staňková et al. 2019) models cancer treatment game theoretic contest between a physician and a tumor, where physicians move on each round is to apply a certain treatment, and a tumor makes an adaptation. While it is a “Stackelberg game”, a game where one player is the leader (the physician) and another player is a follower (the tumor), its asymmetry is rarely exploited. Instead of using their advantage to steer the evolutionary pressures placed on tumors, the physician lets the tumor not only adapt to the current round of the game but also to future rounds of the game. The advantage of leading the game is thereby lost. The authors aptly analogize the current practice:

“Consider cancer treatment as a rock-paper-scissors game in which almost all cells within the cancer play, for example, “paper.” It is clearly advantageous for the treating physician to play “scissors.” Yet, if the physician only plays “scissors,” the cancer cells can evolve to the unbeatable resistance strategy of “rock.”” (Staňková et al. 2019)

Adaptive Therapy is an approach to cancer treatment based on controlling the intra-tumoral evolutionary dynamics. By leveraging that cancer cells can incur a fitness cost to evolving mechanisms that yield resistance to drugs. For example it has been shown that tumor cells can mutate to increased expression of PGP membrane pump, which uses ATP to move drugs out of the cell. While this makes cell more resistant to treatment, it also comes at metabolic cost. (Gatenby et al. 2009) found that PGP activity was the culprit of approximately 50% of cell metabolism. As a result, if resistant and non-resistant cells are competing for space and resources, drug-sensitive cells should over time outcompete resistant cells. Adaptive Therapy utilizes this darwinian competition to make cancer fight itself. Thus the dream-scenario of this adaptive therapy is not to eradicate cancer, but instead make it a controllable chronic disease.



Initial results from pilot clinical trial on metastatic castrate-resistant prostate cancer patients are promising. Initial results showing both less cumulative dosages and longer survival in comparison similar group of patients receiving standard care. Firstly, patients were only involved if they showed a substantial positive response to treatment. The trial has then been utilizing a range-bounded treatment rule. If the bloodmarker *Prostate-specific Antigen* (PSA), a proxy of tumor burden, increases back to pre-trial levels, treatment is applied until PSA drops to 50% of pre-trial levels (Zhang et al. 2017). The trial is expected to run until december 2024 (Center and Research Institute 2024).

Desired qualities of models

While the trial reported by (Zhang et al. 2017) only utilized a single drug, key researchers in researchers in Adaptive Therapy has produced a review of the use of mathematical modelling in the field, and among other things, identified that modelling multidrug treatments is a necessity of future models. Additionally, they argue that it is unlikely that any treatment approach can accomplish delaying the emergence of resistant

cells, lower the tumor burden and minimize the toxicity. Given that patients likely differ in their ability to tolerate tumor burden and the toxicity of drugs and the evolutionary dynamics of the cancers that they carry differ, models would ideally therefore have to trade-off each these factors given a specific patient. This will require fitting data on individual patients. Lotke-volterra models have been fitted frequently.

biological factors Another important aspect of modelling, is illuminating the actual competitive disadvantage that resistant cells are. While larger tumor sizes are thought to increase the suppression of resistant cells, the dynamics are likely much more complex. Factors such as the spatial configuration of actual cells and the range of the molecular influence they yield of each other. It is also crucial that these actually incur a fitness cost, however other authors argue that adaptive therapy might still be able to delay time-to-progression if this is not the case. Competition isn't necessarily strong if the tumor isn't at carrying capacity.

**** CONTROLLING THE RESPONSE TO TREATMENT IS PARAMOUNT. MOUNTAIN CAR PROBLEM ****

Gene-expression has been shown to change in cells as a result of treatment, which further complicates modelling the adaptive therapy as case of resistant vs non resistant cells. Phenotypic plasticity should therefore also be accounted for. Another biological factor that should be accounted for is surrounding tissue. For example, prostate cancer cells in bone can utilize such as the transforming growth factor β can accelerate the proliferation of cancer cells.

**** You can keep throwing layers at POMDPs ****

The efficacy adaptive therapy depends strongly on initial resistance rates, and deciding to opt for control strategy such as adaptive therapy would be beneficial if made early. This however necessitates predicting what patients would respond better to adaptive therapy and who benefit better from the other treatment protocols, such as standard maximum tolerable dose protocol.

How to construct dosing protocols High tumor burdens might also come with other costs, such as increasing the risk of new metastases or simply by the fact that more cells increase the chance total amount of mutations happening. Adaptive therapy also depends on frequent monitoring, and could benefit from the use of different testing protocols.

Robustness to changes to in plans due to machine failure or other practical constraints.

There is need to include multi-drug treatment protocols in adaptive therapy, but the number of possible permutations at each point in time grows extremely fast when more treatment options are introduced. A potential solution to this problem is constructing a treatment protocol that steers the tumor in cycle, so that the conditions at the start of one treatment block is identical to how conditions of the prior block. In principle only one block would have to be designed then. **** this is a non-steady-state equilibrium ****

Constructing real time prediction Predicting individual patient responses real-time would greatly enhance adaptive therapy since dose modulation could be individualized further and evolutionary dynamics controllable with more precision. Using relevant biomarkers would be crucial in this regard **** all observable consequences also generate information **** Any chosen model must be able to be calibrated and validated before hand. When to time the collection of biomarkers also seems to be in issue, since the prostate trial found that treatment would at times overshoot, since biomarkers were collected too late, and the PSA levels were dropped well-below 50%. This likely leads to poorer clinical performance. The link between biomarkers and actual tumor progression is not certain either, meaning that deciding when to treat directly on the basis of a biomarker might not be optimal.

A key issue going forward is rethinking how data on patients is collected. It will be crucial to collect data not only to detect progression, but to collect data that will be useful for future decisions too. **** this is EFE **** Quantifying the uncertainty in the models belief and the consequences of its suggested actions is also paramount.

What are POMDPs in general

Partially Observable Markov Decision Processes (POMDP) is a class of controller models that model and underlying markovian process, that in discrete time and state environment, the next step of the system only depends on the current step. Crucially for partial observability is a crucial facet of these models, and refers to the fact that these models don't directly observe the actual environment or markovian process, but instead only potentially noisy signals emitted by it while trying to manipulate the environment (Åström 1969). This allows these models to differentiate an observed signal from what it "believes" about the environment and use a single reward function trade off uncertainty for achieving a certain goal state (Kaelbling, Littman, and Cassandra 1998) while yielding bayes optimal beliefs. For example, if discretized a, POMDP could understand a PSA reading as noised signal of actual tumor state, and thus try to control to tumor state rather than the PSA reading. While POMDPs are typically difficult to solve analytically various approximate approaches exist.

Using active inference to solve pomdp

One approximate solution, of these is born out of the field of neuroscience literature. The field which has come to be known as active inference suggests that the brain could be using a variational approach. By minimizing two objective functions. *Free Energy* (EFE) as a measure of model and past sensory inputs, and *Expected Free Energy* (EFE) which evaluates future courses of actions against a set of preferred observations. Active Inference has been used to model psychopathology (Da Costa et al. 2020) but also applied to control scenarios such as the mountain-car problem (Friston, Daunizeau, and Kiebel 2009) and, albeit augmented with deep-learning, robotics control (Çatal et al. 2020). These have been implemented in MATLAB and recently in Python with the python package pymdp (Heins et al. 2022a) and (Heins et al. 2022b).

Typically POMDPs are modelled for discrete state spaces. Mathematically they are described as a joint probability:

$$p(o, s, u; \phi)$$

where o are observations, s are hidden states, u are "control states" (states that an agent can influence) and ϕ are the hyperparameters of the model, such as α typically used as 'inverse temperature' i.e. how deterministically the model selects actions. By conditioning on certain observations, the POMDP is solvable by various approximations schemes for the optimal posterior beliefs over what the underlying hidden states are and what the optimal course of action is given a set of preferences. For example, a POMDP could model the joint probability of observing a certain amount of tumor burden in a patient, given some hidden states, such as the how resistancy of the underlying tumors, and whether a certain treatment was applied. Disregarding computability this can be done for any time horizon. Due to considering time and states discrete, this joint probability can be made tractable for a time horizon by factorizing the joint probability into the following categorical probability distributions:

- A likelihood model A . Usually modelled as a set of arrays where each array corresponds to an observation modality describes how observations map to a particular state. For example how one modality could be PSA readings, and its likelihood array describes how likely different test results are under different tumor burdens. It can both be modelled as a perfect signal of the tumor burden, i.e. the same tumor burden always gives the same test result, but it could also be implemented as a noised signal. If multiple states are modelled, each likelihood modality is an array with a dimension corresponding to the number of possible observations of that modality, and for each hidden state another dimension is added with the same length as the number of possible hidden states. In this way, every combination of state can be mapped to an observation.
- A transition model B . It describes the probability of state transitioning to another on any particular timestep. This also encodes how actions are expected to influence transition probabilities. It could describe how a tumor is likely to evolve from one timestep to next depending whether treatments

is being applied. The transition model is usually coded as collection of three-dimensional arrays, a first dimension for the next state, a dimension for the current state, and a third dimension with the length of each action that would influence transition probabilities of the state. Modelling the transition probabilities would depend on whether treatment is being applied or not.

- A prior over preferred states C . A particularity of these models is that utility is specified in probabilities. This prior is set over certain observations, meaning that the model artificially expects to see certain observations. As a result, this drives the models behavior to act in ways that bring it close these expectations. At a first pass, describing a goal in terms of probabilities seems odd, but it could be considered a fundamental property of self-organizing systems. For example, to stay alive, I would have to spend time states where my tumor burden is low, even though in all likelihood, given enough time, I would come to develop cancer. Crucially, it allows for inferring what actions would bring about me spending time in states with low tumor burdens.
- A prior over initial states D . These are vectors of probabilities of initial beliefs. In continuation with example above, this component would specify how probable different levels of tumor burden before interacting with the patient.

The POMPD can be solved at any time step by minimizing two measures, *Variational Free Energy* (VFE) and *Expected Free Energy* (EFE). VFE is an upper bound on *surprisal*, and is approximated to infer the most likely current state given an observation, the likelihood model, and the prior over either initial states if no earlier observations have been made. Otherwise the preceding posterior over states is used as the current prior (Smith, Friston, and Whyte 2022). VFE scores how well the models representation aligns with a set of observations, and by using different approximate posteriors over states, choosing the approximate which minimizes VFE will approximate bayesian belief updating (Da Costa et al. 2020). EFE instead score, plans on how to act and their expected outcomes. EFE is used to construct a posterior distribution over what policy is most preferred. Since the model is equipped with C , the set of prior preferences over observations, the distinguishing between probable and preferred is difficult. This could however be considered a benefit, since it EFE is a composite of utility gain and information.

This potential benefit to adaptive therapy could be realized if adaptive therapy is modelled as a system where the tumor burden emits a noised signal. While controlling the tumor burden in short run is imperative, truly successful strategies will have to also control future resistance dynamics. This task seems more challenging since resistance dynamics don't appear directly observable in clinical settings. Two obvious choices emerge, either implement a heuristic such which we expect to control the resistance dynamics well, or explicitly model the resistance dynamics despite paucity of real-time data on resistance. However, by applying a treatment and observing how the changes in tumor burden, the underlying resistance dynamics should be inferable. This paper investigates whether active inference can be used to gain traction on the second strategy.

Consider the following motivations for the approach:

- Real-time data will be sparse in the clinical settings. Extracting the maximum amount of information for each data point could be pivotal. Models with a claim to bayes-optimal beliefs should be favored
- Short-term exploiting tumor vulnerability is essential for keeping the patient alive, but for long term success, but to control resistance dynamics, gaining information about how the resistance dynamics is also essential and this must be done through treating the tumor. Choosing whether to treat would therefore not only be a consideration of the tumor level, but also about how much wiser it would make us about learning the resistance dynamics.

This has the interesting corollary, that two patients could have the exactly the same tumor burden, but if the resistance dynamics is well modelled for one, treating could be poor choice, while for the other patient, despite them having the same tumor burden, treating could be advisable, simply to gain information about the resistance dynamics. Superficially, the Active Inference implementation of POMDPs could be well suited candidate, to optimally make this utility information trade-off. It should be noted that active inference could be implemented in variety of ways.

Methods

Simulation details

Environment

The simulated environment features discretized planning cancer treatments. At each timestep a “cancer state” has a certain risk of increasing. If the tumor ever reaches the state 5, the simulation ends. To avoid this, a model has decide when to apply treatment. Applying treatment can reduce the tumor level - whether the treatment successfully reduces the “cancer state” depends on a underlying *resistance state*. When the resistance is low, the chance of treatment succeeding is high, but probability declines as the resistance state increase. The resistance state has fixed probability of increasing when treatment is applied, and it can decrease when treatment is withdrawn. Whether the resistance state decreases depends on the tumor level. At high tumor levels, the resistance state is more likely to decrease, but this carries the risk of tumor growing out of control and “killing the patient”. In order to successfully manage the disease, a model will therefore also have to manage the resistance state. However, the resistance state is directly observable in the given simulation. It must therefore be inferred from how the tumor state responds to treatment.

Sim Runs:

Both the resistance state, and tumor states start at state 0 out of 5, meaning that a total is six states are possible in each. In both simulations, POMDPs transition matrices perfectly reflect the transition probabilities of the environment. Two slightly different simulations are used:

- Performance simulation, where a range-bounded approach to managing the tumor is used, where the treatment is begun each time the tumor state increases above level 3, and is withdrawn when the tumor state drops below 2. 100 runs at 4 different probabilities of increasing the tumor state is run for maximally 200 timesteps. For each run, vectors of outcomes are pregenerated, meaning that vector of whether the tumor increases of length 200 is predetermined where each entry has the probability specified for the entire simulation. Likewise vectors are generated for treatment outcomes at each resistance level, and outcomes for resistance drops. This is done to ensure comparability between the range-bounded approach and POMDP models that consider different future outcomes at different lengths. In these runs the tumor state is perfectly observable, meaning a tumor state always generates tumor observation that corresponds to hidden factor. The POMDP can only observe the tumor level, and its prior preferences are uniform over all tumor states except for the highest. These POMDPs only consider a combination of policies at each timestep. They can only consider treating or not treating for three timesteps in a row. The shortest horizon model only considers one of these blocks, while the medium considers two blocks of treating or not treating, for total horizon of 6 steps into the future. The longest horizon model considers three blocks for a total of 9 steps into the future. This is done to ease the computational burden, since it greatly reduces the total number of policies to be evaluated.
- In the capacities simulation, a single run of a POMDP with a slightly different model structure is used. This model can observe a perfect signal of whether the patients is alive, whether the model is testing, treating and a noised signal of the tumor state if the model has decided to test the patient on the given run. Its prior preferences are heavily skewed against observing a dead patient, somewhat against treating and little against not treating. This means that it will have to balance the “cost” of all these actions”. It is also handicapped further by the fact that the tumor signal is noised, and the resistance state, which must be inferred through the tumor signal is therefore doubly obfuscated. The likelihood mapping between however perfectly captures the maps the expected noise.

Changes to POMDP specification.

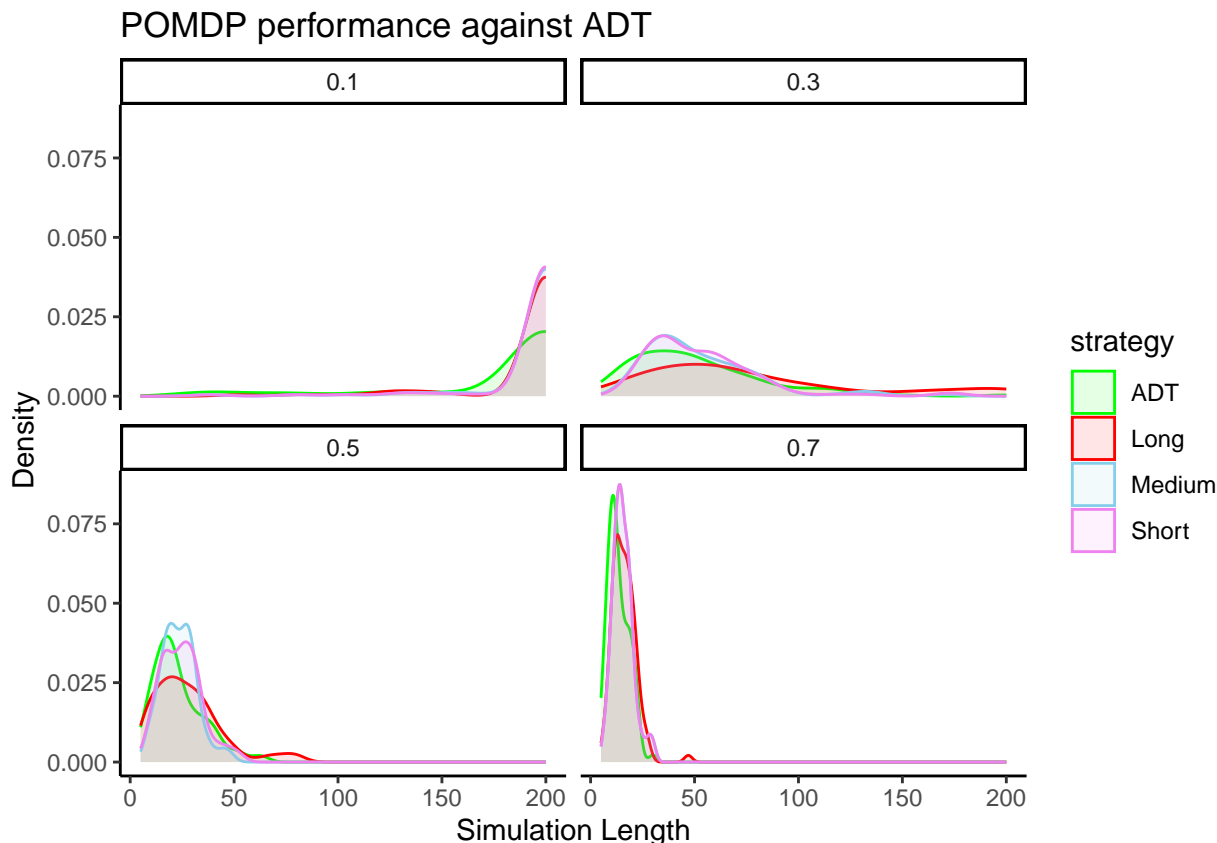
Hidden state factors do typically not affect each other at the state level. Instead their interactions are typically modelled as resulting from observations. This setup was deemed inadequate for the current experiment. It necessitated the that generative model could accurately model how the resistance state factor influences the probability of treatment succeeding. This was done in order to accurately portray issue of resistance levels generally being in accessible to current testing methodology and the evolutionary dynamics that are suspected to be at play in reducing resistance in real-world tumors.

Accurately modelling how higher tumor levels makes decreases in resistance more likely was also crucial. Typically, transition probabilities in the generative models are constructed using three dimensional “B-tensors”, which describe expected transition probabilities within a state factor: one dimension the current state, one dimension for what ever action is chosen and third for the resulting state.

For the present project another dimension was added, this dimension corresponds to the state of another state factor. Concretely, this meant that the tumor state factor had fourth dimension. By matrix multiplication the expectation over this fourth dimension is factored in. One could choose to view it as there are now as many three dimensional b-tensors for the tumor state factor as there are states in the resistance factor. Effectively, matrix multiplying these result in a probability weighted average of expected tumor transition probabilities, i.e. the expectation. However, it should be noted that it is not trivial the order in which states are evaluated and that this must be specified. This also resulted in much pymdp functionality breaking, since it was presumably built only with three dimensional b-tensors in mind.

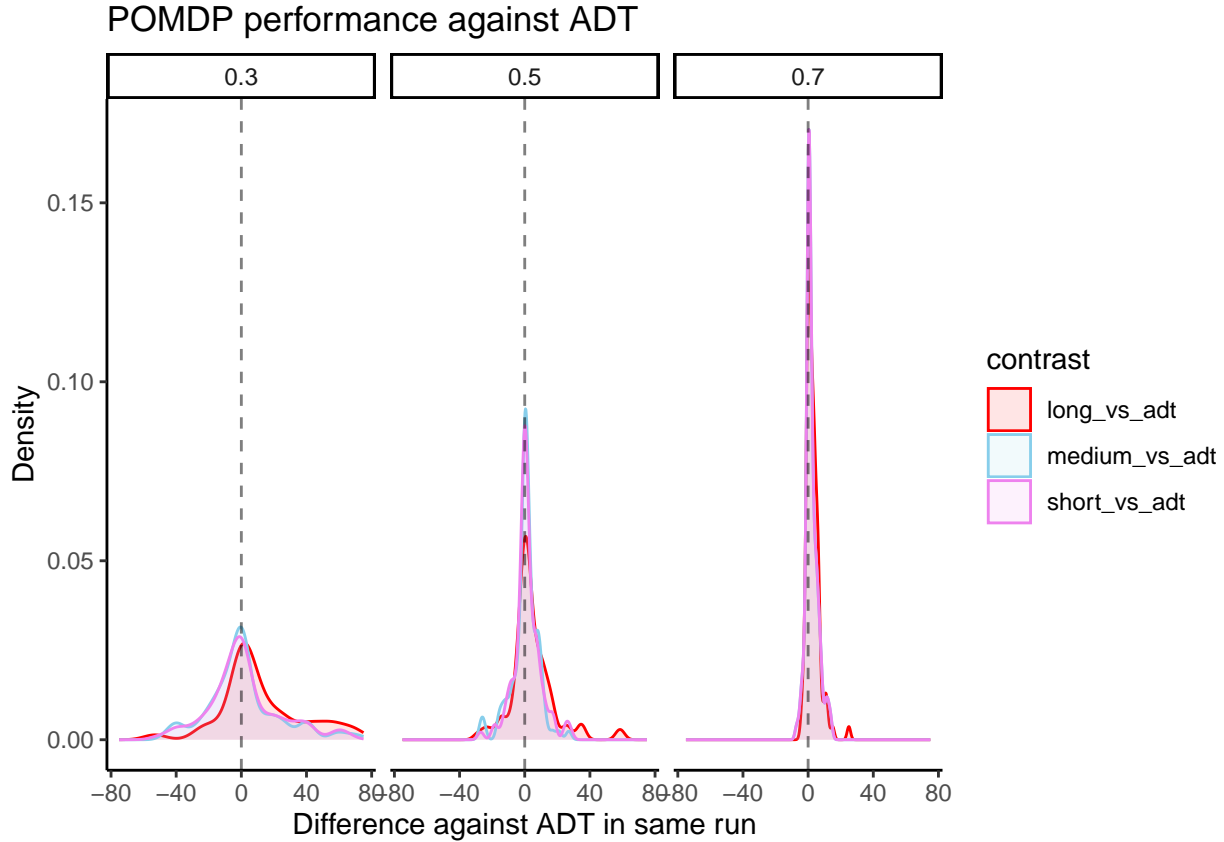
Results

Performance of POMDPs vs ADT heuristic

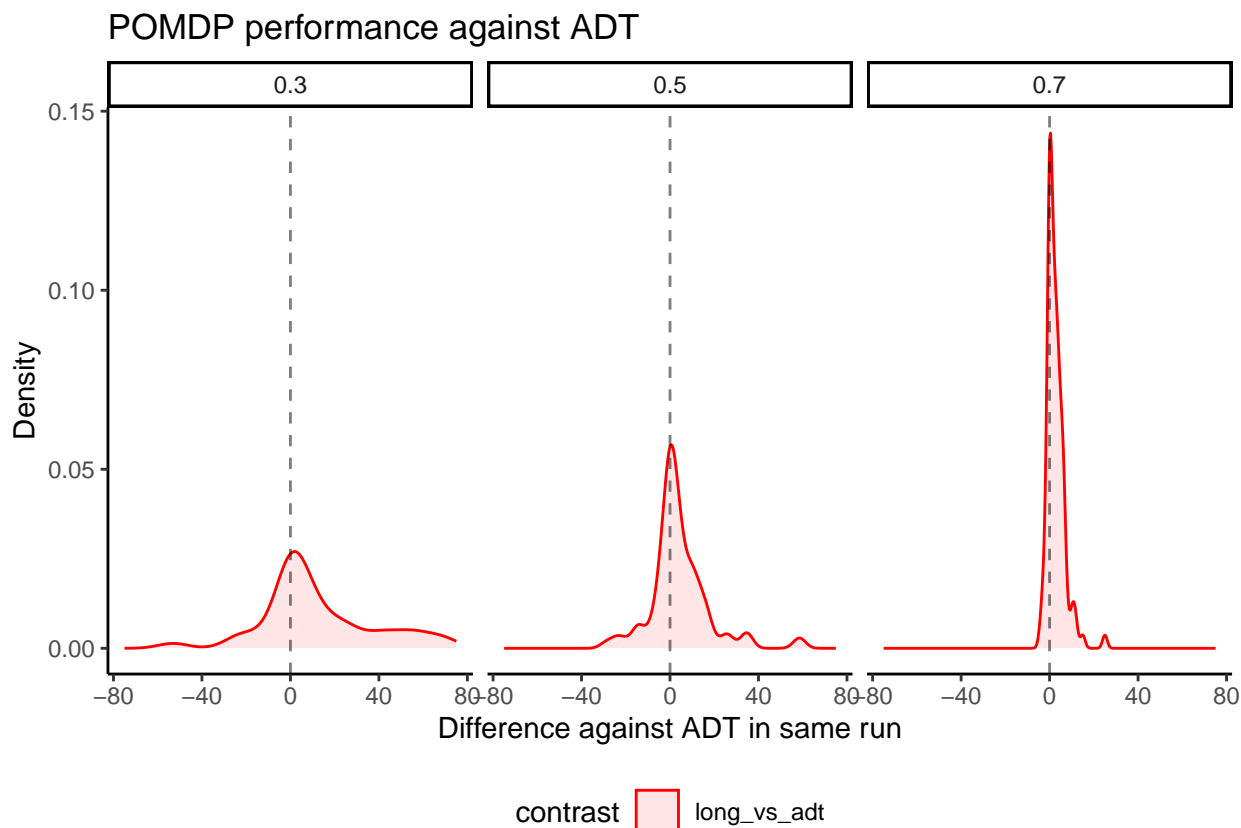


Four different simulation runs each consisting of 100 runs, of maximally 200 timesteps for different rates of tumor growth (respectively 1. .3, .5, and .7 probability of increasing the tumor state) . On each run the the outcomes on each of the potential 200 random variables are generated and each strategy is therefore tested on a the same enviroment. On the lowest tumor growth rate almost all the runs reach maximal length.

Contrasts against ADT-performance for growth rate .3, .5 and .7 are plotted. .1 is omitted since it appears that maximal performamnce was achieved for each policy horizon of the POMPD strategies.



Two interpretations emerge from the contrasted simulation lengths. As the the growth rate increases, the POMDP seems to run for longer than the specific instance of the range bounded approach. It also seems that longer policy horizon seems beneficial. When the tumor state had .7 probability of increasing each timestep, the longest POMPD with the longest policy horizon performed as follows.

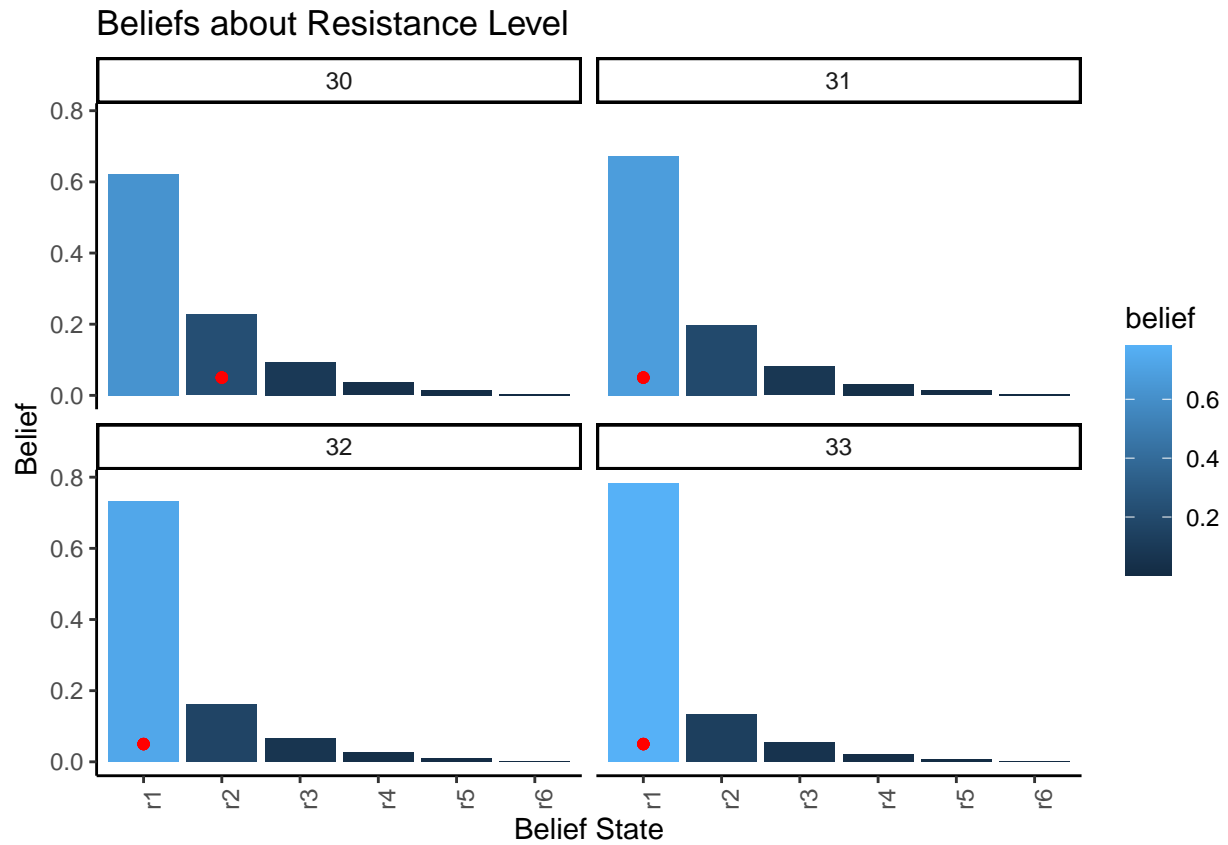


At more aggressive cancer rates, the simulation consistently to outperform the range bounded approach. The absolute increase in timesteps decreases however.

Capabilities

Learning underlying hidden state

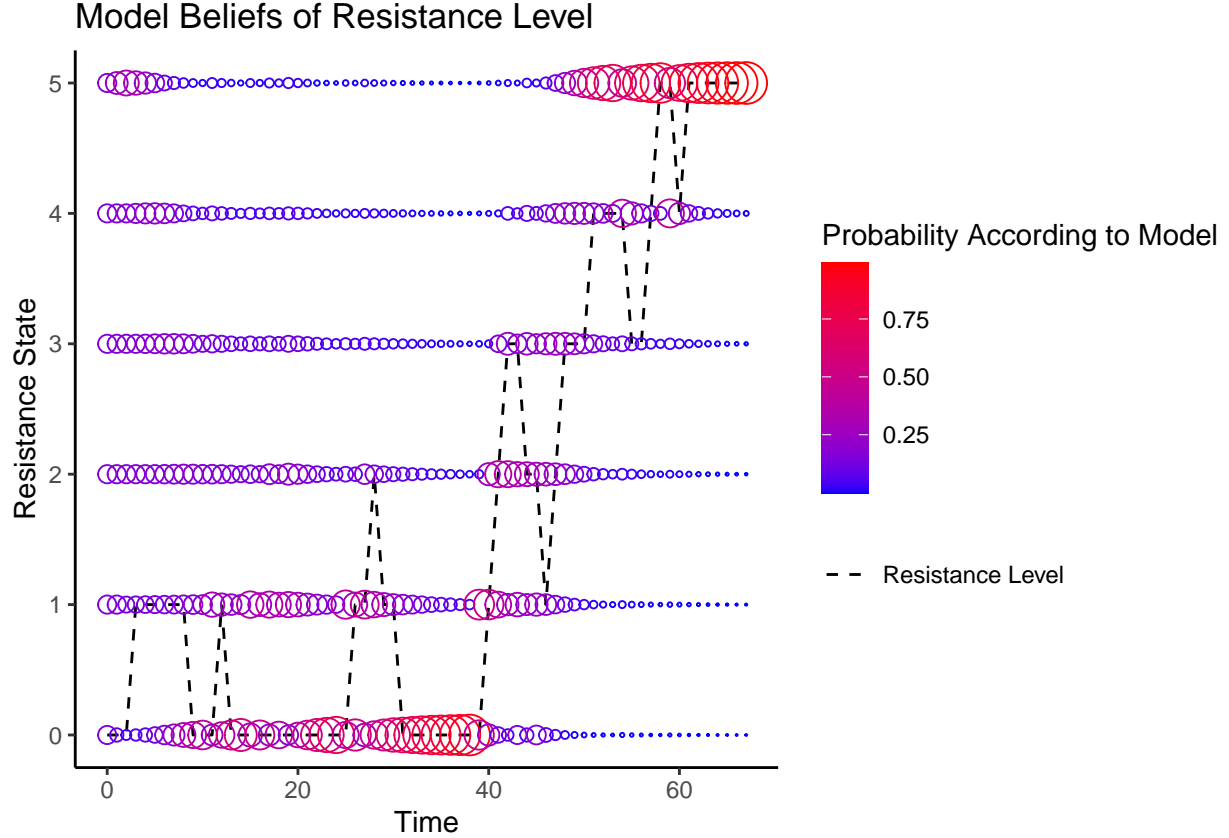
The POMDP structure is capable of inferring an underlying hidden state: the resistance state. Even though this not directly observable, this can be inferred to how the tumor state responds to treatment. Applying treatment for longer time without any beneficial effect would suggest that the resistance level is high, while immediately observing that the tumor level decreases would suggest that the resistance level is low. At each timestep the POMDPs perform infer the most likely state of every state factor, given their current observation and prior beliefs. Through custom changes to the 'get_expected_states' function in PYMDP that allowed the models to consider how one hidden state factor (resistance factor) would influence another (the tumor factor).



The above plot shows the strength of beliefs in t probabilities for a subset of timesteps [30 - 34], and the red dot show the actual the resistance levels. As time progresses the resistance level increases, and the model adjusts its beliefs. While the tumor level doesn't increase during this time period, this would also signal the model that the resistance level is low. This is the case since the environment always has a change of increasing the tumor state, but such an increase would be negated by a successful round of treatment.



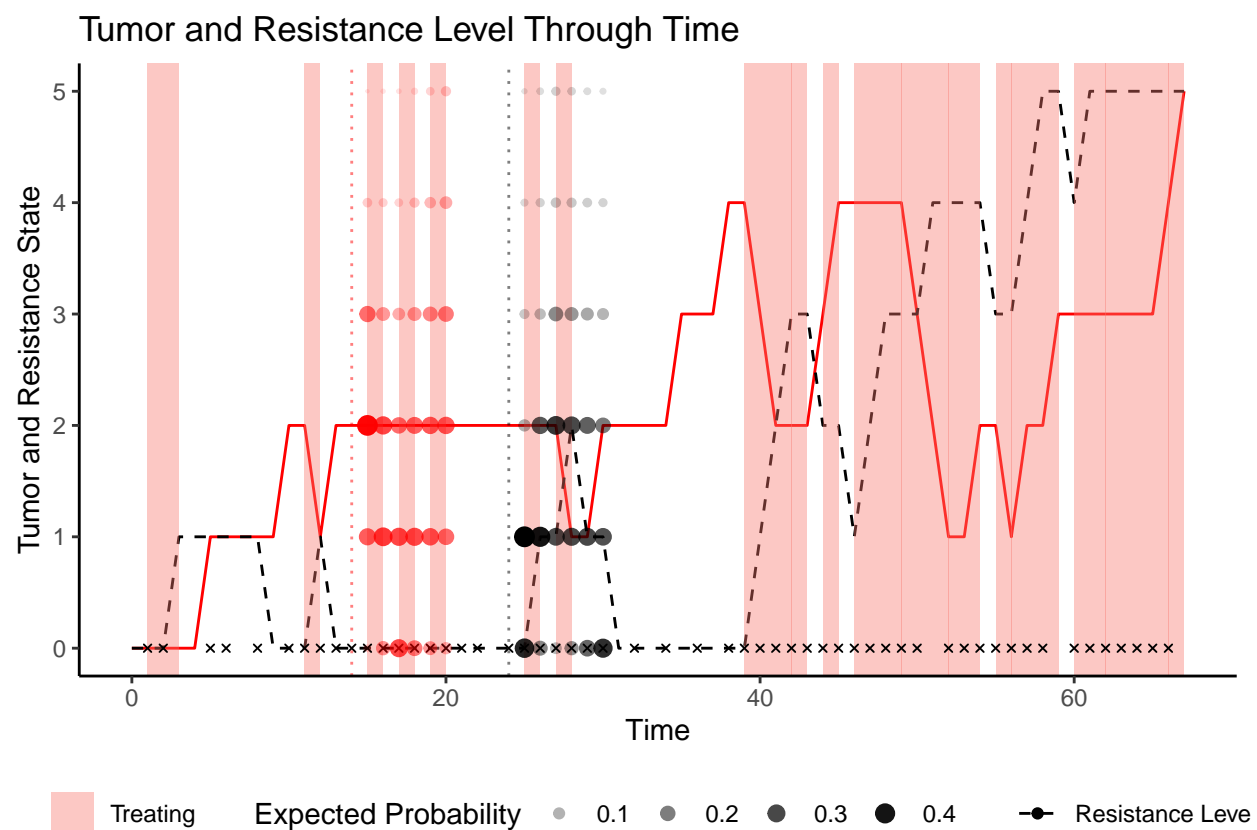
The entire run is plotted for the given. Only the tumor level is observable to the POMDP. It must combine its knowledge about how resistance level likely increases after applying treatment, and how the tumor level responds to treatment depending on the the resistance level. While the model far from perfectly knows the resistance level. Model beliefs for the entire run is plotted.



The model begins fairly agnostic. Given that the model was initialized with a uniform prior over resistance states this makes sense. Throught out the course of the simulation it then finds lower values of resistance state more likely.

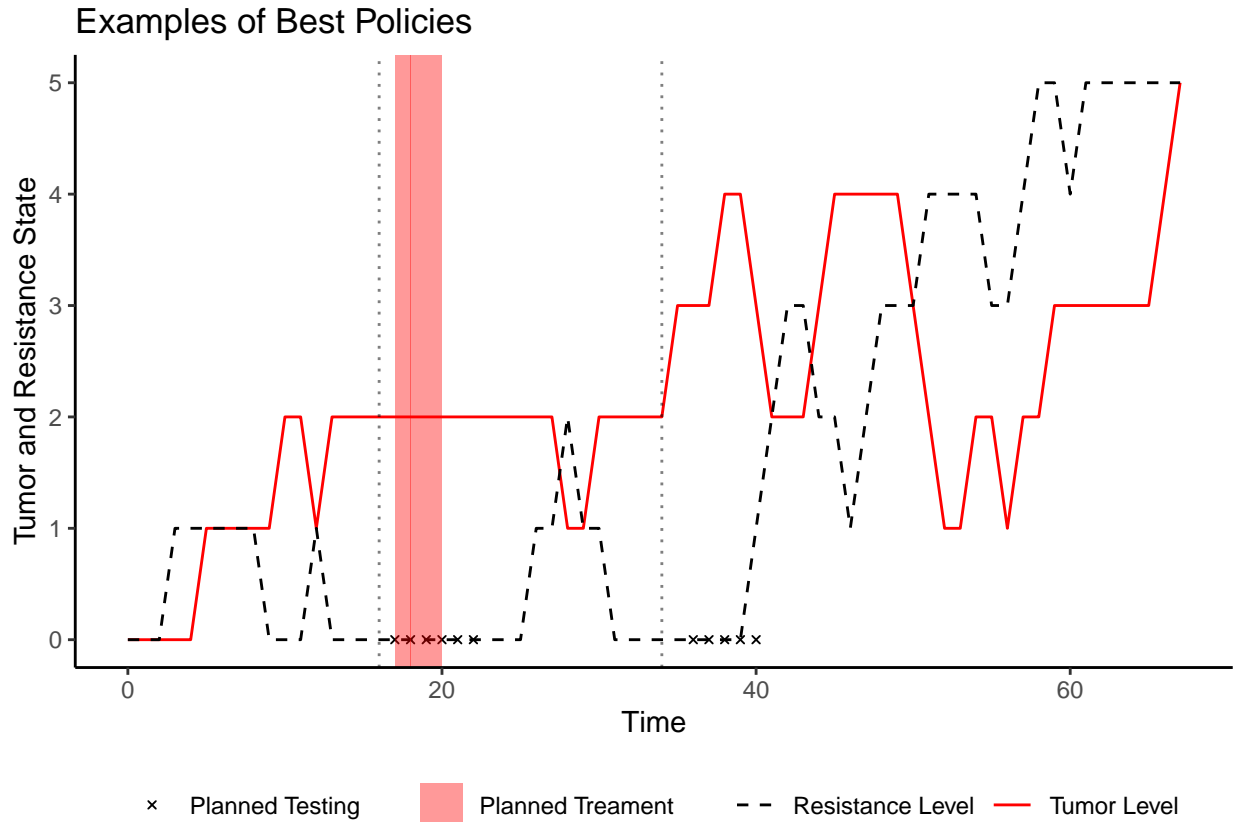
Beliefs, current/future and uncertainty is accessible

This current simulated model used a slightly different implementation than those compared to performance of rangebounded therapy planning. While the those models had longer policy horizons, they didn't consider the entire space possible actions. A model with a shorter policy horizon that would instead search every single possible action four steps is plotted to investigate the structure of decision-making by the POMDP model. The enviroment was also more uncertain. Instead of featuring a 1-to-1 mapping of the observations of the tumor to actual tumor state, it recieved a noised signal. It must therefore. The beliefs about how tumor state will evolve as a consequence of the most promising policy at time step 14 is plotted, and the expectation of resistance states at $t=24$ is plotted too.

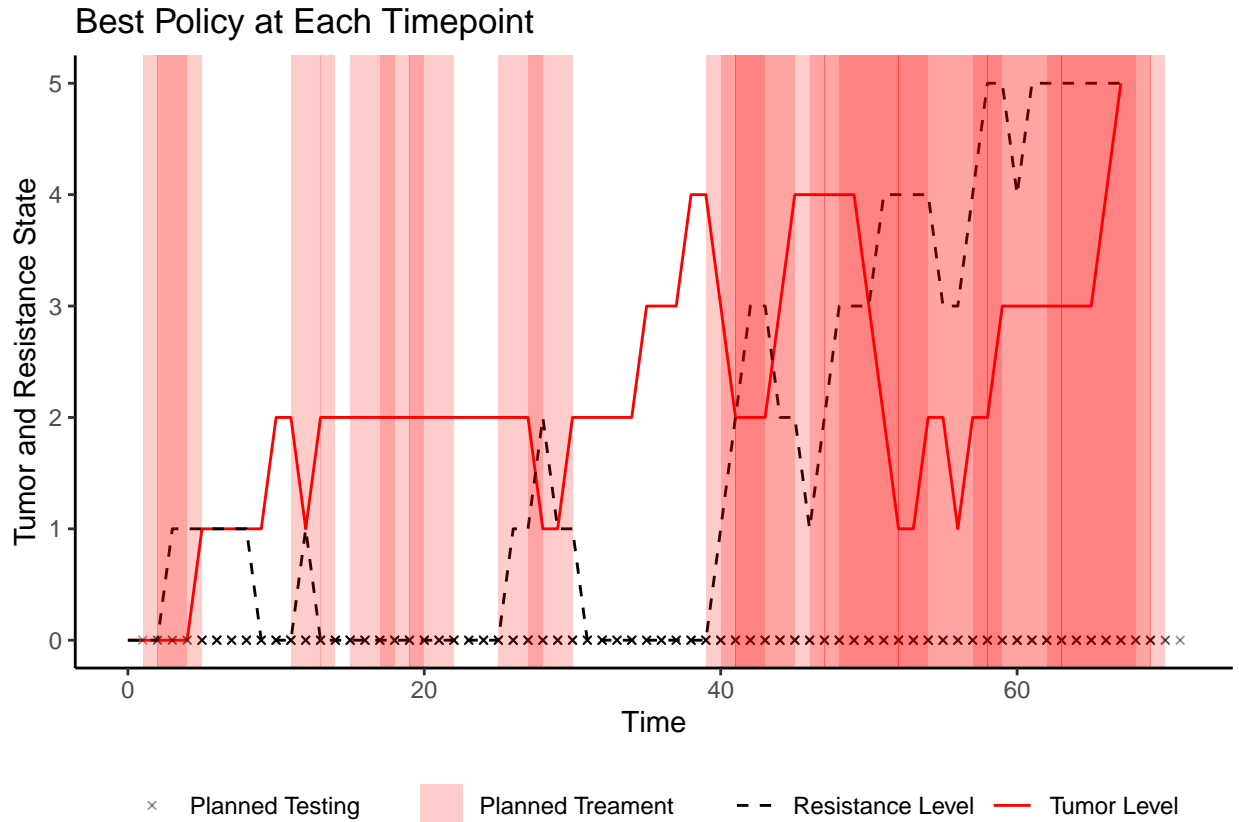


Dotted vertical lines indicated the time-point for which the evaluation of expected states are extracted.

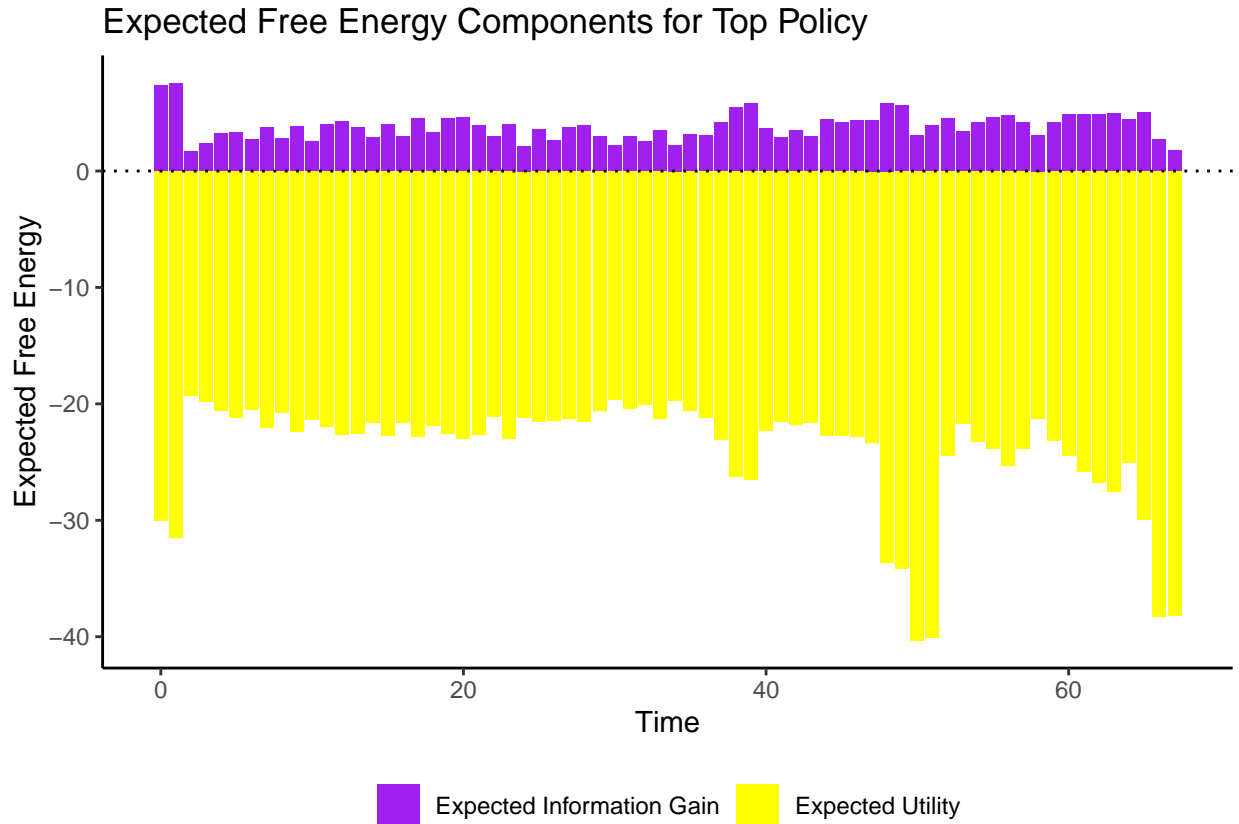
Policy evaluation



The above plot shows the highest evaluated policy at timestep 16 and at timestep 34. At timestep 16 not treatment is planned, and testing is withdrawn for next timestep but picked up again for remainder of the period. For 34 treatment is suggested applied for three rounds. Note that policies are evaluated each timestep, and these plans are likely not finished. Instead a course of action is decided on each turn. The highest rated policy for each timestep is plotted



The above plot doesn't show realized actions but cumulative best action from perspective of each time point. This means that for completely blank spots, there is no point where model thinks treating or testing would be the optimal course of action. This doesn't never happen for treating, but some periods of treatment are thought time best held withdrawn. A darker shade indicate that on multiple prior timesteps, the model thought acting would be best course of action. Exact decision making for each policy can also be extracted. Since each policy is evaluated for its EFE, its expected information and utility gains can be extracted



Discussion

What has this project found

The current simulation demonstrates that in a ableit simplified enviornment that mimicks particular dynamics of adaptive therapy, POMPD models for which the excact transition probabilities of the enviornment has already been specified can plan treatment better than the specific range-bounded treatment heuristic implemented here. It should be noted that only one range bound was tested. The gain in relative time to progression is stronger under more aggressive cancer growth rates, and POMPDs with longer policy horizons perform better than shorter policy horizons. It is quite possible that the celing effect of policy horizon wasn't met.

Additionally, the paper demonstrates that pompd show a number of desiarable qualities especially true for adaptive cancer therapy, but some of these could possibly also be extended to other medical situations. While the POMDP is able to discern the difference between an underlying state and the noised signal it emits. In the present case this would be tumor state and the tumor observatoin, more interestingly, is the ability of the POMDPs to directly control a deeper state (in the present example, the resistance state). This state doesn't produce observations directly, but is mediated through another state (tumor state). It thus has to be inferred through how other states evolve over time. Modelling the underlying states was achieved using a small change to POMPD implementation in pymdp. (See methods section for a thorough explanation). This points to another important quality of POMDPs, namely their flexibility. More complex generative models could straight forwardly by set up and tested. One could for example envision a multi-drug generative model or models that "cancer agressivity" state which would modulate how aggressive fast the tumor changes. Crucially, for each causal node one can dream up, if it can be discretized and describe in terms of transition probabilities, it should be implementable.

Another key feature of the POMDP implementation is the use of the free enegyry principle. Since free energy can be shown to be the result of factoring in both expected information gain and utility gain, pommps

should be able to plan testing procedures optimally, given the transition probabilities, likelihood mappings and prior preferences are correctly specified. Conceptually this reduces the boundary between testing and treating, since both convey an informational gain. For example, in certain cases, it might be prudent to treat an otherwise mild form of cancer simply to gain information about how effective treatment will be in the future.

In a review of modelling approaches for adaptive therapy, (West et al. 2023) highlight desiderata for mathematical models for Adaptive therapy. One of these is need for accurate description of the uncertainties for the mathematical model. As demonstrated, the beliefs about current and future states are completely accessible, and the decision making reduces to Bayesian updating. Due to the flexibility of the POMDP structure, computation and our ability to specify the transition probabilities are the only limit to number of interventions that could be modelled. Planning multidrug approaches should therefore also be a straightforward procedure for POMDPs. The “reasoning” behind the suggested treatment course is also evaluated since each policy is evaluated for its gain in utility, how much the policy will move the system towards desired states, and how accuracy of generative model will change. Only computation and our ability to specify the transition probabilities limit the number of interventions that could be added to POMDPs.

**** more information is present when treating ****

**** less**

Does it work for other sorts of medical planning

FEP components can be used to balance testing and treating

Would longer policy search be better

Perhaps one could build a two-layer model. The bottom layer controls what goes inside the treatment-cycle, like the mountain-car controller. The top layer would have to decide what of the different controllers to use. This

Only one range bound

Future Research

NESS

Especially POMDPs but also other models have been used to model the decision of agents in active inference. Especially in psychopathology, human decision making has been modeled through these. This presents a fairly speculative, but interesting possibility. It could potentially allow to approximate the decision making of otherwise black box models. Perhaps if suitable controller could be designed for example by machine learning or another process leading to an opaque decision making process. POMDP should be fit to its behavior and the model’s decision making would thus become transparent. It would likely come with a cost to performance, but perhaps such a trading some performance for transparency could be worthwhile in a clinical setting.

This approach obviously isn’t any better than our ability to construct such a black-box model and evaluate its performance prior to clinical trials.

Priors over policies

One could also enforce a prior over an adaptive therapy approach, making sure that the model would only deviate from this if its very certain. This could hopefully combine the benefits and robustness of the 50% rule of thumb with the capacities of active inference.

Continuous state space or binning illness

Moving from continuous biomarkers to discrete data will be necessary. either by reformulating problems out for the discrete state space, or by binning, transforming observations into discrete values. Perhaps by dimensionality reduction such as PCA, deep-neural networks, or simply binning observations. For example a tumor state could correspond to a 10% increase in tumor burden. Considering that they grow exponentially, this could perhaps result in the states occupying the same amount of time/width of the underlying continuous distribution.

Learning the transition parameters

Learning the transition parameters, could happen through some combination of learning on the fly and pre-specification by expert/knowledge modelling. Deep-learning on simulation data could probably be used. Perhaps also on already fitted data.

Searching Policy Space better

Demonstrating the usefulness of models in adaptive therapy

An approach to demonstrating the usefulness of POMDP active inference model could be to show that it manages to control tumor progression successfully on more realistic simulations, such as the Zhang et al. 2017 model. Either learning likelihood and transition probabilities on the fly or specifying them in a manner before hand that could be done in clinical setting too would be necessary. Even more impressive would be a model that would given no chance to act, but only observe, accurately predict the actual outcomes of clinical data. For example the trial data reported. If this model would make interesting post-dictions, about better courses of actions, this would be severely interesting. It would still be hard to beat a rct.

Other fep models

- Åström, K. J. 1969. “Optimal Control of Markov Processes with Incomplete State-Information II. The Convexity of the Lossfunction.” *Journal of Mathematical Analysis and Applications* 26 (2): 403–6. [https://doi.org/10.1016/0022-247X\(69\)90163-2](https://doi.org/10.1016/0022-247X(69)90163-2).
- Bukowski, Karol, Mateusz Kciuk, and Renata Kontek. 2020. “Mechanisms of Multidrug Resistance in Cancer Chemotherapy.” *International Journal of Molecular Sciences* 21 (9): 3233. <https://doi.org/10.3390/ijms21093233>.
- Çatal, Ozan, Samuel Wauthier, Cedric De Boom, Tim Verbelen, and Bart Dhoedt. 2020. “Learning Generative State Space Models for Active Inference.” *Frontiers in Computational Neuroscience* 14 (November). <https://doi.org/10.3389/fncom.2020.574372>.
- Center, H. Lee Moffitt Cancer, and Research Institute. 2024. “A Pilot Study of Adaptive Abiraterone Therapy for Metastatic Castration Resistant Prostate Cancer.” <https://clinicaltrials.gov/study/NCT02415621>.
- Da Costa, Lancelot, Thomas Parr, Noor Sajid, Sebastijan Veselic, Victorita Neacsu, and Karl Friston. 2020. “Active Inference on Discrete State-Spaces: A Synthesis.” *Journal of Mathematical Psychology* 99 (December): 102447. <https://doi.org/10.1016/j.jmp.2020.102447>.
- Friston, Karl J., Jean Daunizeau, and Stefan J. Kiebel. 2009. “Reinforcement Learning or Active Inference?” *PLoS ONE* 4 (7): e6421. <https://doi.org/10.1371/journal.pone.0006421>.
- Gatenby, Robert A., Ariosto S. Silva, Robert J. Gillies, and B. Roy Frieden. 2009. “Adaptive Therapy.” *Cancer Research* 69 (11): 4894–4903. <https://doi.org/10.1158/0008-5472.CAN-08-3658>.

- Heins, Conor, Beren Millidge, Daphne Demekas, Brennan Klein, Karl Friston, Iain D. Couzin, and Alexander Tschantz. 2022a. “Pymdp: A Python Library for Active Inference Indiscrete State Spaces.” *Journal of Open Source Software* 7 (73): 4098. <https://doi.org/10.21105/joss.04098>.
- Heins, Conor, Beren Millidge, Daphne Demekas, Brennan Klein, Karl Friston, Iain Couzin, and Alexander Tschantz. 2022b. “Pymdp: A Python Library for Active Inference in Discrete State Spaces.” *Journal of Open Source Software* 7 (73): 4098. <https://doi.org/10.21105/joss.04098>.
- Kaelbling, Leslie Pack, Michael L. Littman, and Anthony R. Cassandra. 1998. “Planning and Acting in Partially Observable Stochastic Domains.” *Artificial Intelligence* 101 (1): 99–134. [https://doi.org/10.1016/S0004-3702\(98\)00023-X](https://doi.org/10.1016/S0004-3702(98)00023-X).
- Smith, Ryan, Karl J. Friston, and Christopher J. Whyte. 2022. “A Step-by-Step Tutorial on Active Inference and Its Application to Empirical Data.” *Journal of Mathematical Psychology* 107 (April): 102632. <https://doi.org/10.1016/j.jmp.2021.102632>.
- Staňková, Kateřina, Joel S. Brown, William S. Dalton, and Robert A. Gatenby. 2019. “Optimizing Cancer Treatment Using Game Theory.” *JAMA Oncology* 5 (1): 96–103. <https://doi.org/10.1001/jamaoncol.2018.3395>.
- West, Jeffrey, Fred Adler, Jill Gallaher, Maximilian Strobl, Renee Brady-Nicholls, Joel Brown, Mark Roberson-Tessi, et al. 2023. “A Survey of Open Questions in Adaptive Therapy: Bridging Mathematics and Clinical Translation.” Edited by Richard M White. *eLife* 12 (March): e84263. <https://doi.org/10.7554/eLife.84263>.
- Zhang, Jingsong, Jessica J. Cunningham, Joel S. Brown, and Robert A. Gatenby. 2017. “Integrating Evolutionary Dynamics into Treatment of Metastatic Castrate-Resistant Prostate Cancer.” *Nature Communications* 8 (1): 1816. <https://doi.org/10.1038/s41467-017-01968-5>.