SocultPaperV2

2024-05-19

Introduction

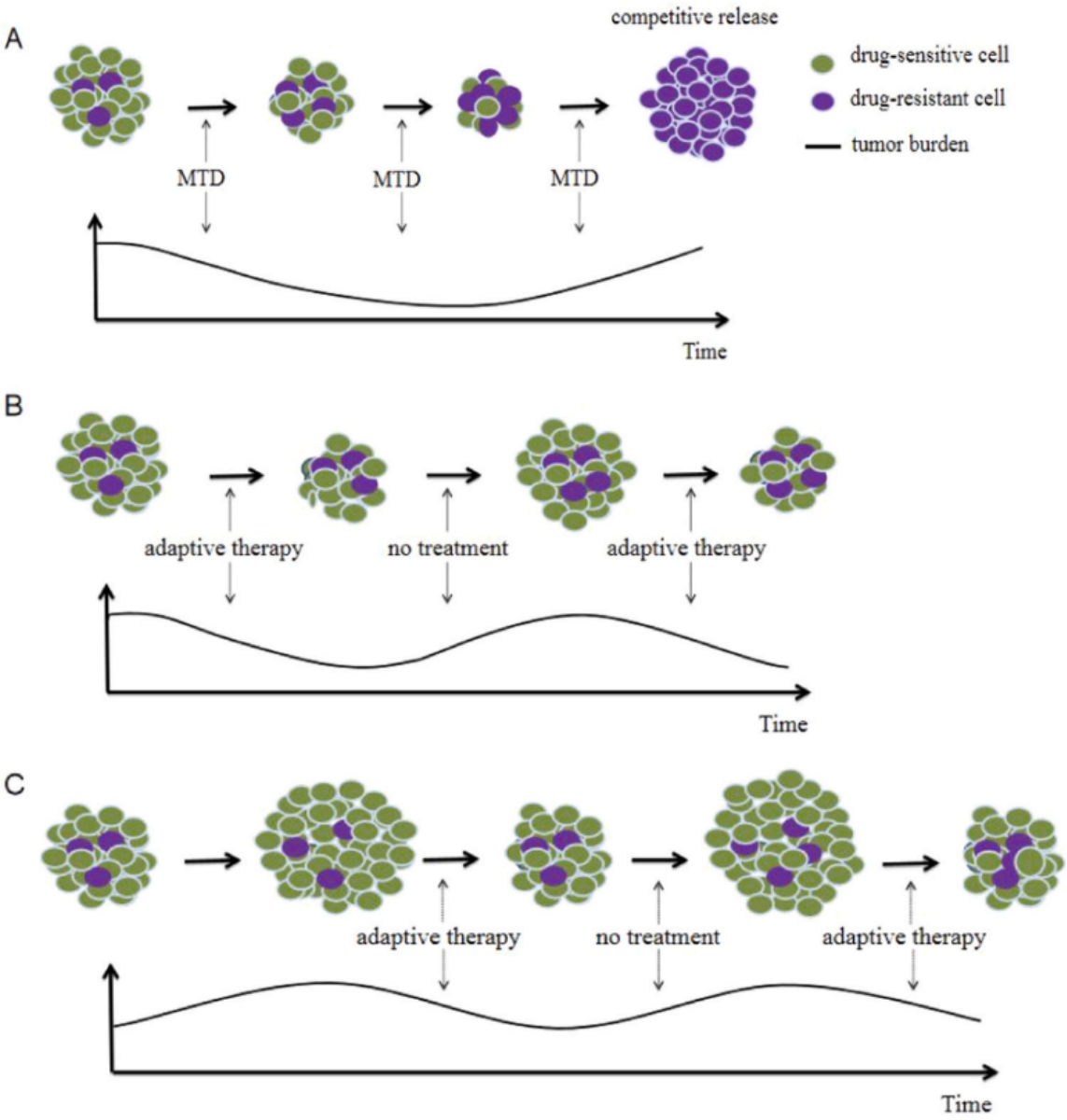
Why we should POMDPs with FEP for Adaptive cancer therapy

What is the adapative cancer therapy

Worldwide, cancer is the cause of 1 out 6 deaths. Of these cancer deaths an estimated 90% are due to development of drug resistance (Bukowski, Kciuk, and Kontek 2020). While intial cancer treatment usually shows positive response in tumor burden, drug resistance develops due. To highlight the inefficincy of traditional approaches, (Staňková et al. 2019) models cancer treatment game theorict contest between a physician and a tumor, where physicians move on each round is to apply a certain treatment, and a tumor makes an adaption. While it is a "Stackelberg game", a game where one player is the leader (the physician) and another player is a follower (the tumor), is assymerty 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 a approach to cancer treatment based on controlling the intra-tumoral evolutionary dynamics. By leveraging that cancer cells can incur a fitness cost to evovling 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. Apative Therapy utilizes this darwinian competition to make cancer fight itself. Thus the dream-sceneario 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. Intitial results showing both less cumuluative dosages and longer survival in comparison similiar group of patients recieving standard care. Firstly, patients were only involved if they showed a substainal 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 Adapative Therapy has produced a review of the use of mathematical modelling in the field, and among other things, indentified 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 minizmize the toxicity. Given that patients likely differ in their ability to tolerate tumor burden and the toxicity of drugs and the evolutioanry dynamics of the cancers that they carry differ, ,odels would ideally therefore have to trade-off each these factors given a specific patient. This will require fitting data on invidual patients. Lotke-volterra models have been fitted frequently.

biological factors Another important aspect of modelling, is illumanitng the actual comppetive disadvantage that resistant cells are. While larger tumor sizes are thought to increase the suppresion 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 authiors argue that adaptive therapy might still be able to delay time-to-progression if this is not the case. Comptettion isn't neccesarily 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 sourrind tissue. For example, prostate cancer cells in bone can utilzied such as the transforming growth factor β can accelerate the proliferation of cancer cells.

** You can keep throwing layers at POMDPs**

The efficay adapative therapy depends strongly on initial resistance rates, and deciding to opt for control strategy such as adpative 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 metastates or simply by the fact that more cells increase the chance total amount of mutations happening. Adapative 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 tother practical constraints.

There is need to include multi-drug treatment protocols in adaptive therapy, but the number of possible permtuations 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 indentical to how conditions of the prior block. In principle only one block would have to be designed then. ** this is a non-steady-state equlibirum **

Constructing real time prediciton Prediciting individual patient repsonses real-time would greatly enhance adaptive therapy since dose modulation could be indivdualized 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 biomarks 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 neither, meaning that decidnig when to treat directly on the basis of a biomarker might not be optimal.

A key issue going forward is rethiking 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 usefull for future decisions too. ** this is EFE** Quantifying the unceartinty in the models belief and the consequences of its suggested actions is also paramount.

What are POMDPs in general

Partially Observable Markov Decision Processes (POMPD) is class of controller models that model and underlying markovian process, that in disctrete time and state enviroment, the next step of the system only depends on the current step. Crucially for partial orbservability is crucial facet of these models, and refers to the fact that these models don't directly observe the actual enviroment or markovian process, but instead only potentially noisy signals emitted by it while trying to manipulate the evivorment (Åström 1969). This allows these models to differiate an orbserved signal from what it "believes" about the enviorment and use a single reward function trade off unceartinty for achieveing 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 pompds

One approximate solutions, of these is born out of the field of neuroscience litterature. The field which has come to be know as active inference suggests that the brain could by using a varitional approach. By minimizing two objective functions. *Free Energy* (EFE) as measure of model and past sensory inputs, and *Expected Free Energy* (EFE) which evaluates future courses of actions against a set preferred observations. Active Inference has been used to model psychopathology (Da Costa et al. 2020) but also applied to control scenearious 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 implemneted in MATLAB and recently in Python with the python package pymdp

How do POMDPs meet these requirements

Active Inference is implemented by minimizing free energy which minimizes surprisal

Methods

Simulation details

Enviroment

The simulated enviroment features discretized of 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 succesfully 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 thiscarries the risk of tumor growing out of control and "killing the patient". In order to succesfully manage the disease, a model will therefore also have to manage the resistance state. However, the resistance state is directly orbserable 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 matricies perfectly reflect the transition probabilties of the enviorment. 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 comparibility 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 POMPD can only observe the tumor level, and its prior preferences are uniform over all tumor states except for the highest. These POMPDS 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 observes 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 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.

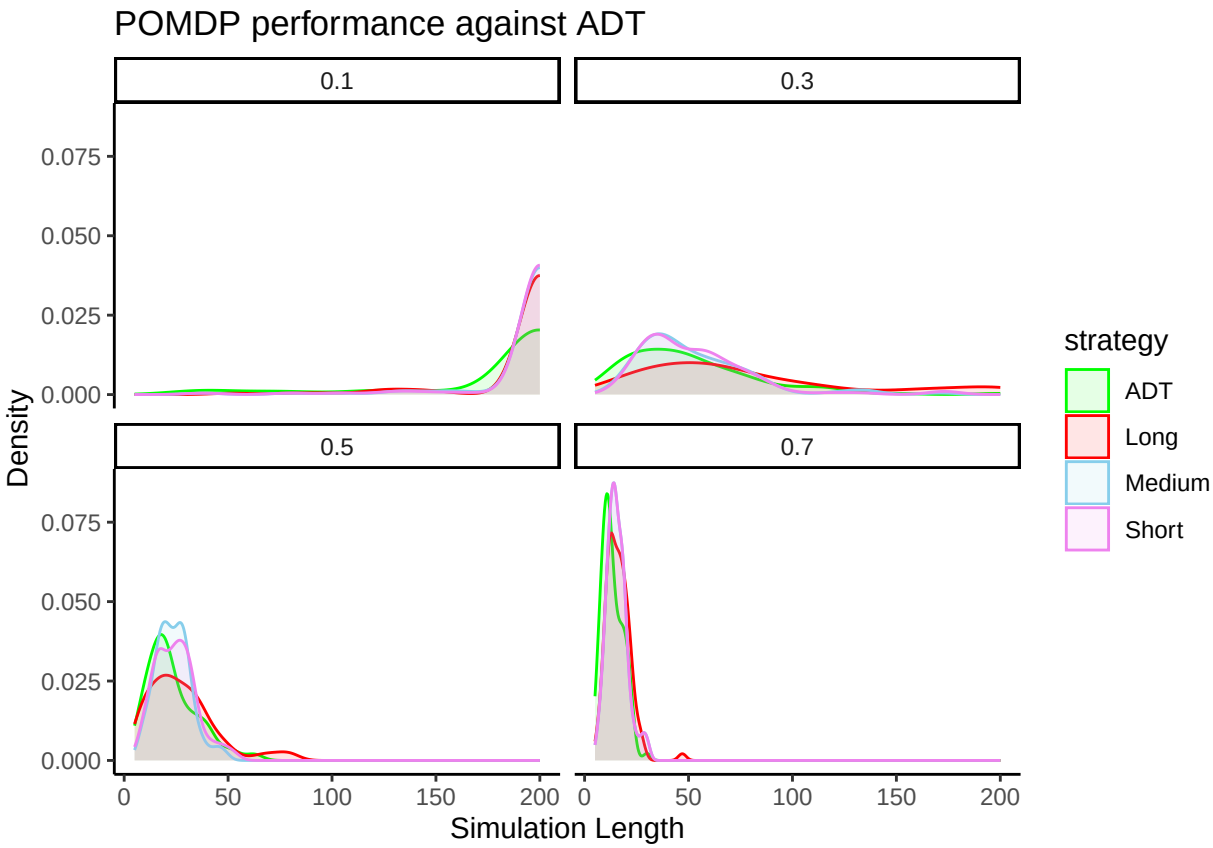
Hidden state factors do typically not affect each other at the state level. Instead their interactions are typically modelled as resulting from observations. This setup was deemed inadequate for the current experiment. It necessitated the that generative model could accurately model how the resistance 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 probabilties in the generative models are constructed using three dimensional "B-tensors", which describe expected transition probabilties 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.

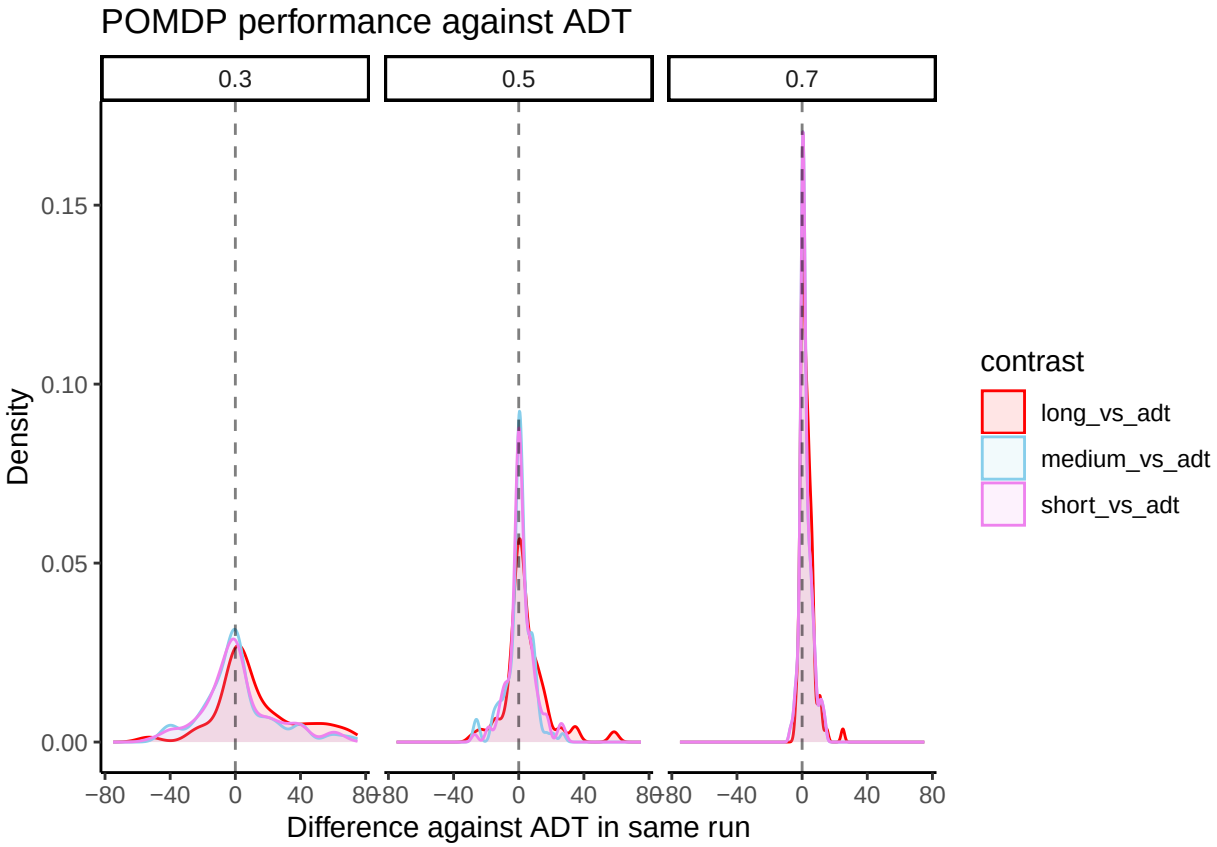
Results

Performance of POMDPs vs ADT heuristic

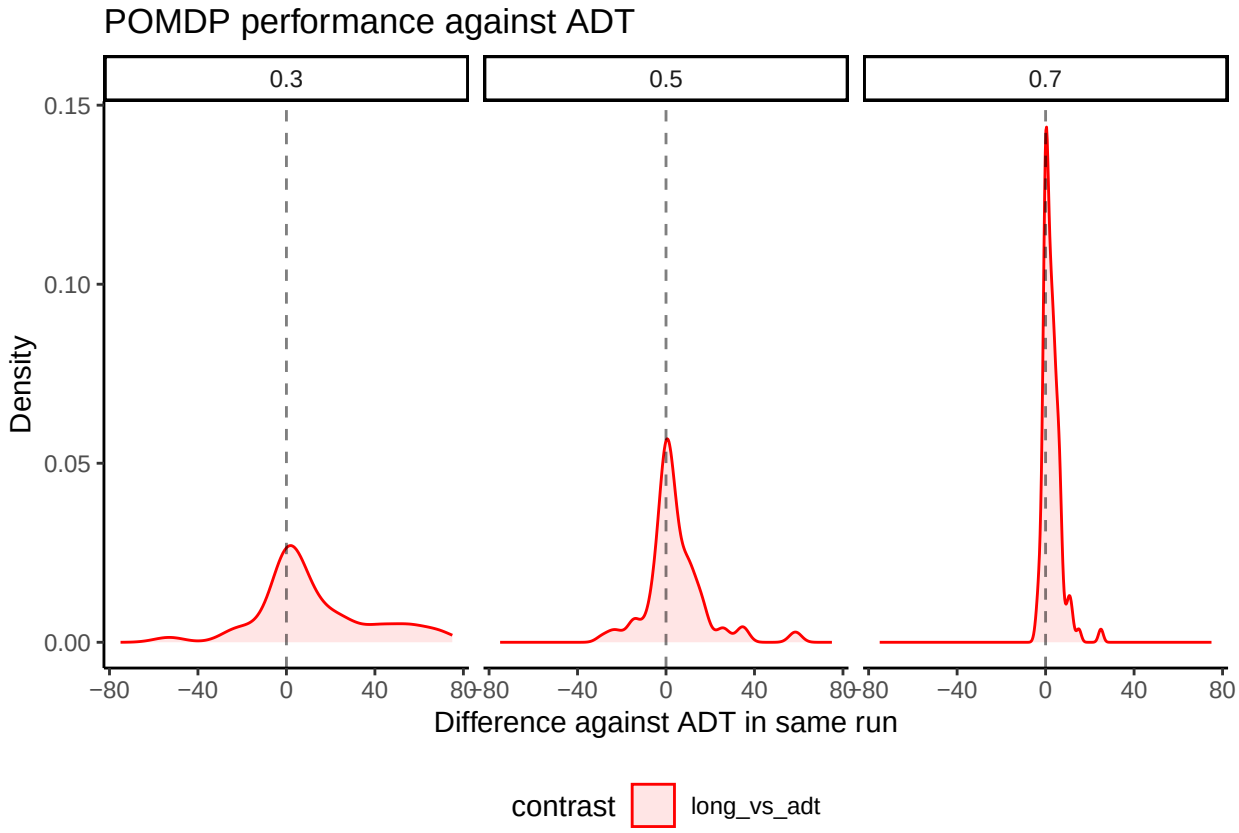


Four different simulation runs each consisting of 100 runs, of maximally 200 timesteps for different rates of tumor growth (respectively 1. .3, .5, and .7 probability of increasing the tumor state) . On each run the the outcomes on each of the potential 200 random variables are generated and each strategy 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 perforamnce 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.

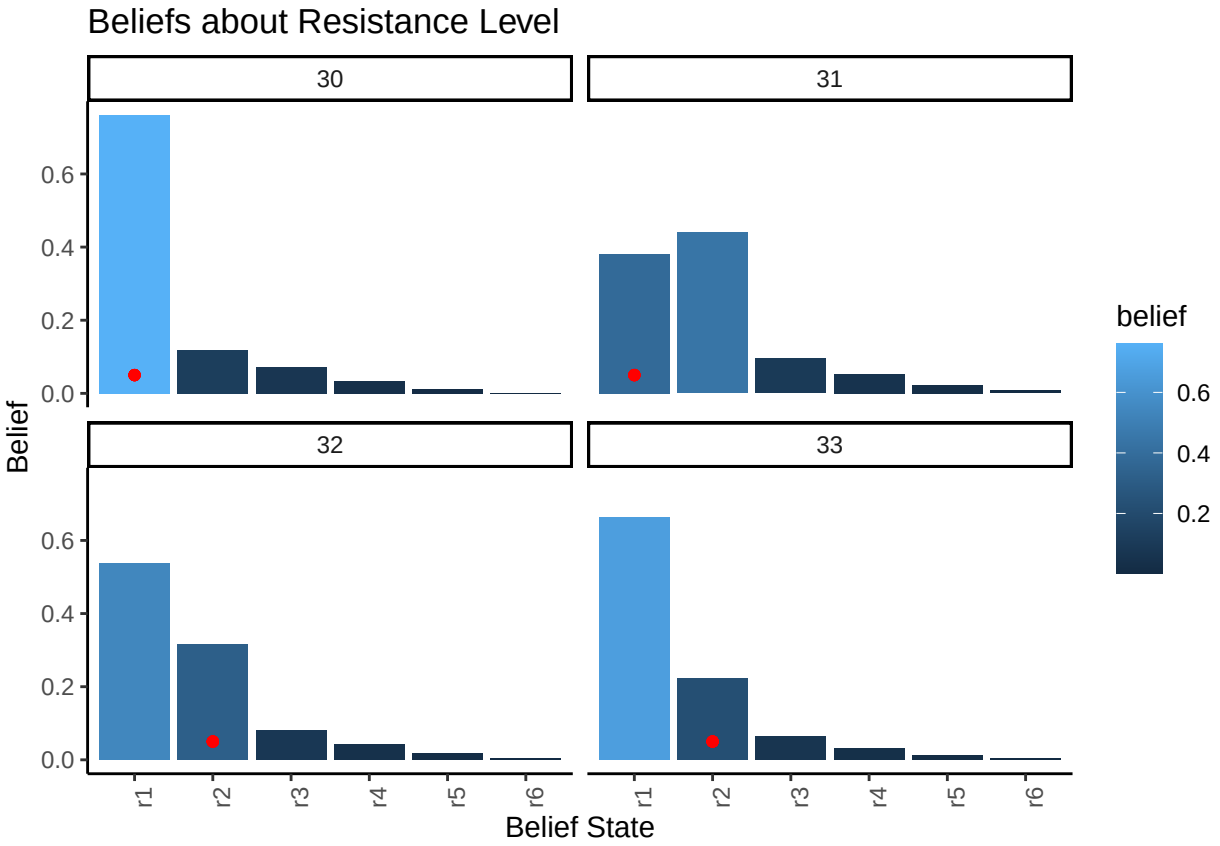


At more aggressive cancer rates, the simulation consistently to outperform the range bounded approach. The absoulute increase in timesteps decreases however.

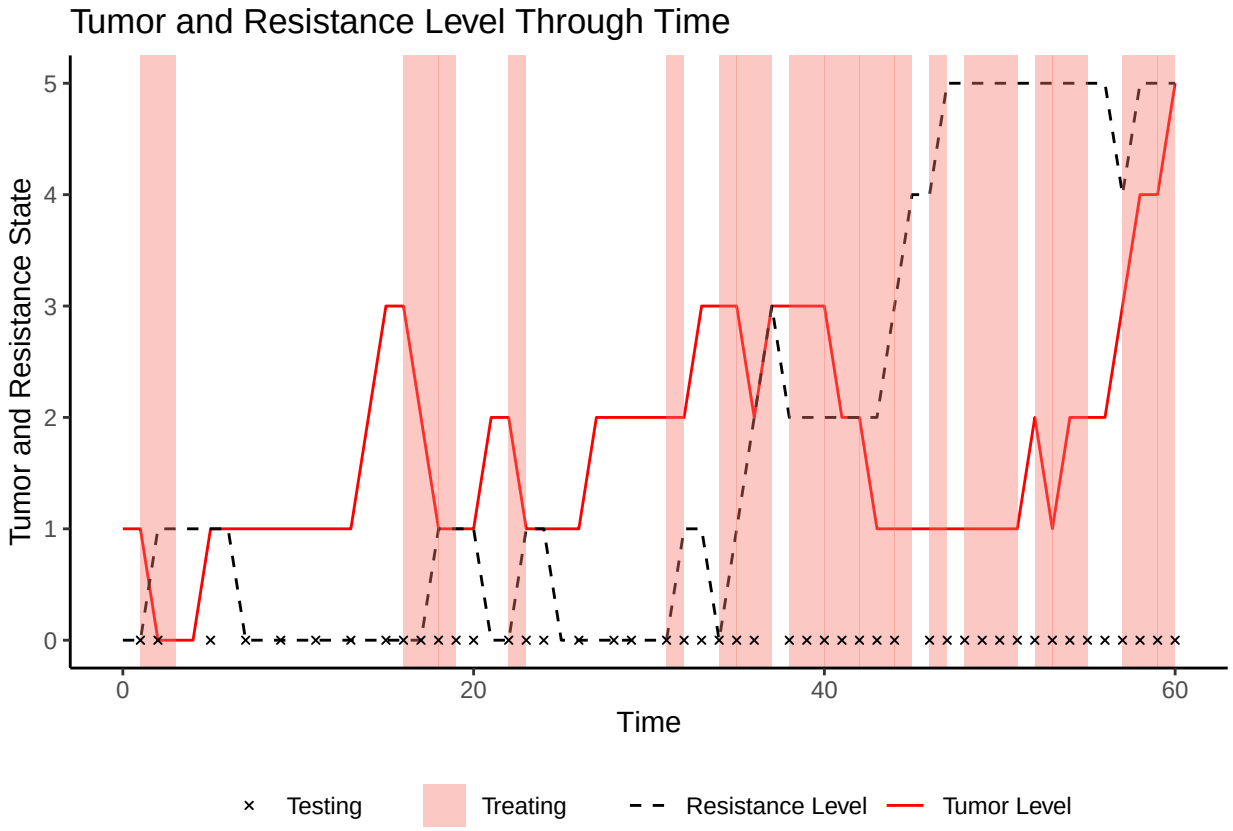
Capabilities

Learning underlying hidden state

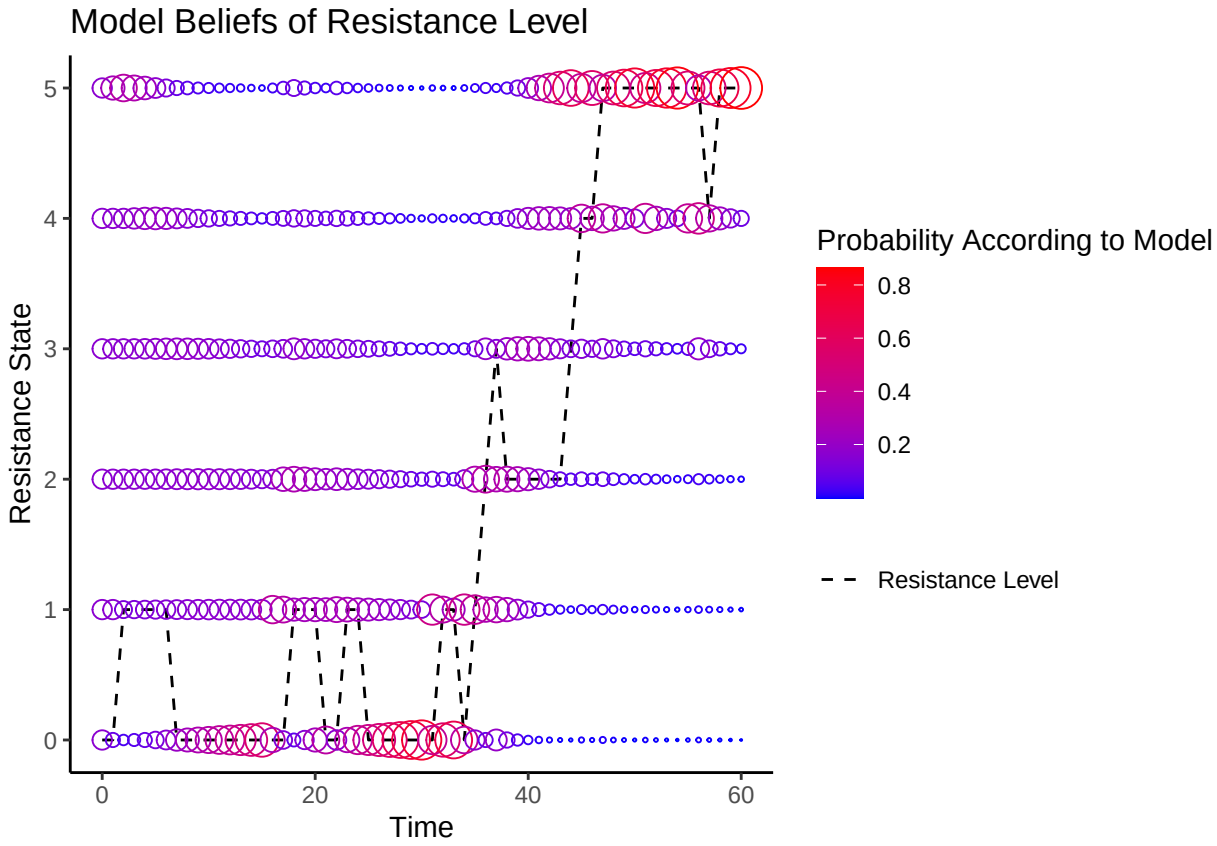
The POMDP structure is capable of inferring an underlying hidden state: the resistance state. Even though this not directly observable, this can be inferred to how the tumor state responds to treatment. Applying treatment for longer time without any beneficial effect would suggest that the resistance 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. A time progresses the resistance level increases, and the model adjusts its beliefs. While the tumor level doesn't increase durings this time_period, this would also signal the model that the resistance level is low. This is the case since the enviroment always has change of increasing the tumor state, but such an increase would be negated by succesful 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 depenping on the the resistance level. While the model far from perfectly knows the resistance level. Model beliefs for the entire run is plotted.



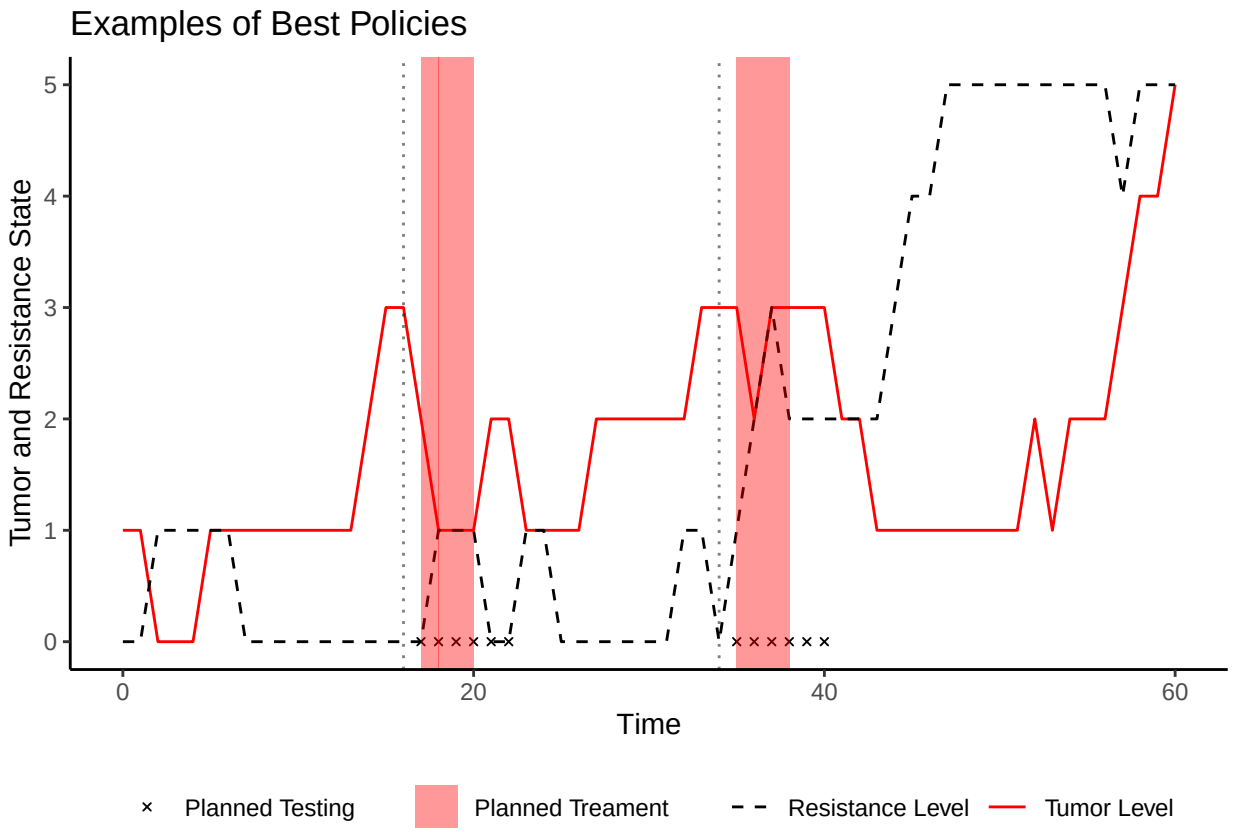
The model begins fairly agnostic. Given that the model was initialized with a uniform prior over resistance states this makes sense. Thorought out the course of the simulation it then finds loweer values of resistance state more likely.

Beliefs, current/future and uncertainty is accessible

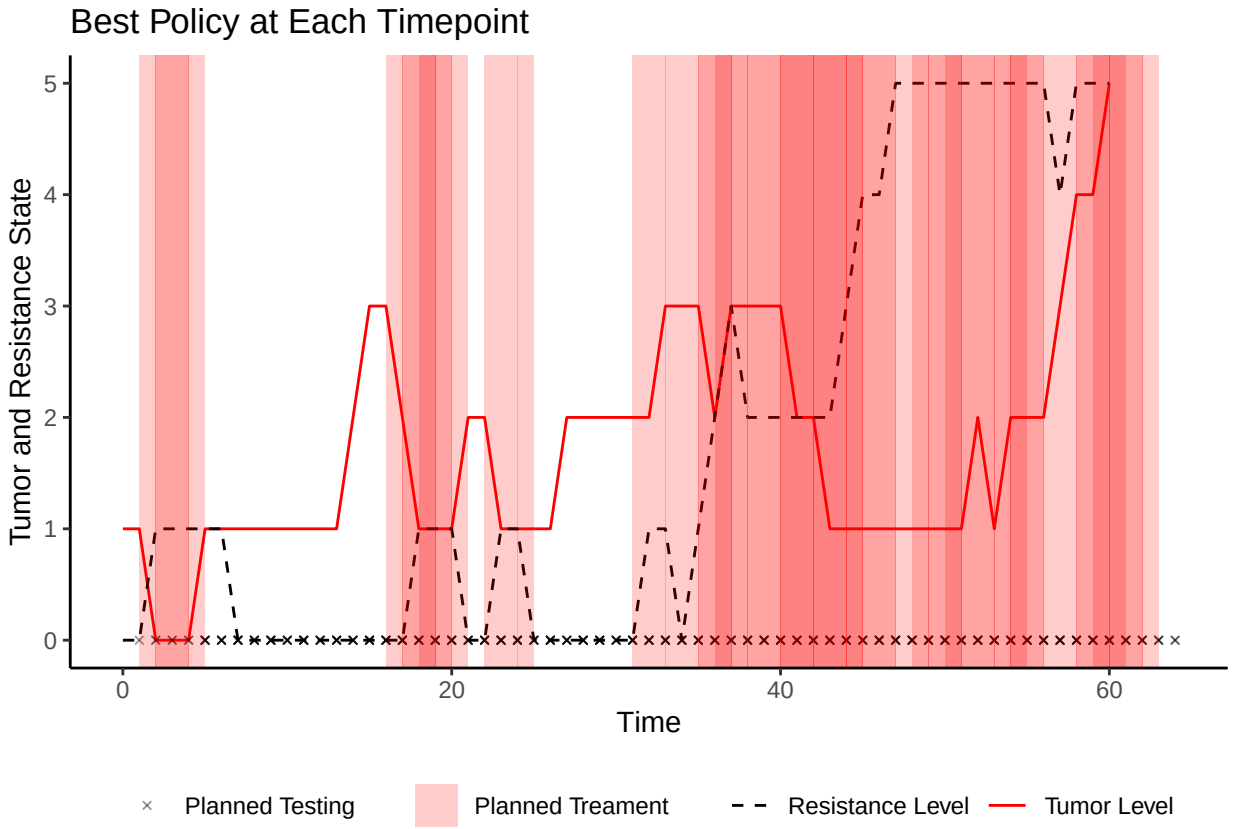
This current simulated model used a slightly different implementation than those compared to performance of rangebounded therapy planning. While the those models had longer policy horizons, they didn't consider the entire space possible actions. A model with a shorter policy horizon that would instead 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 unceartain. 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.

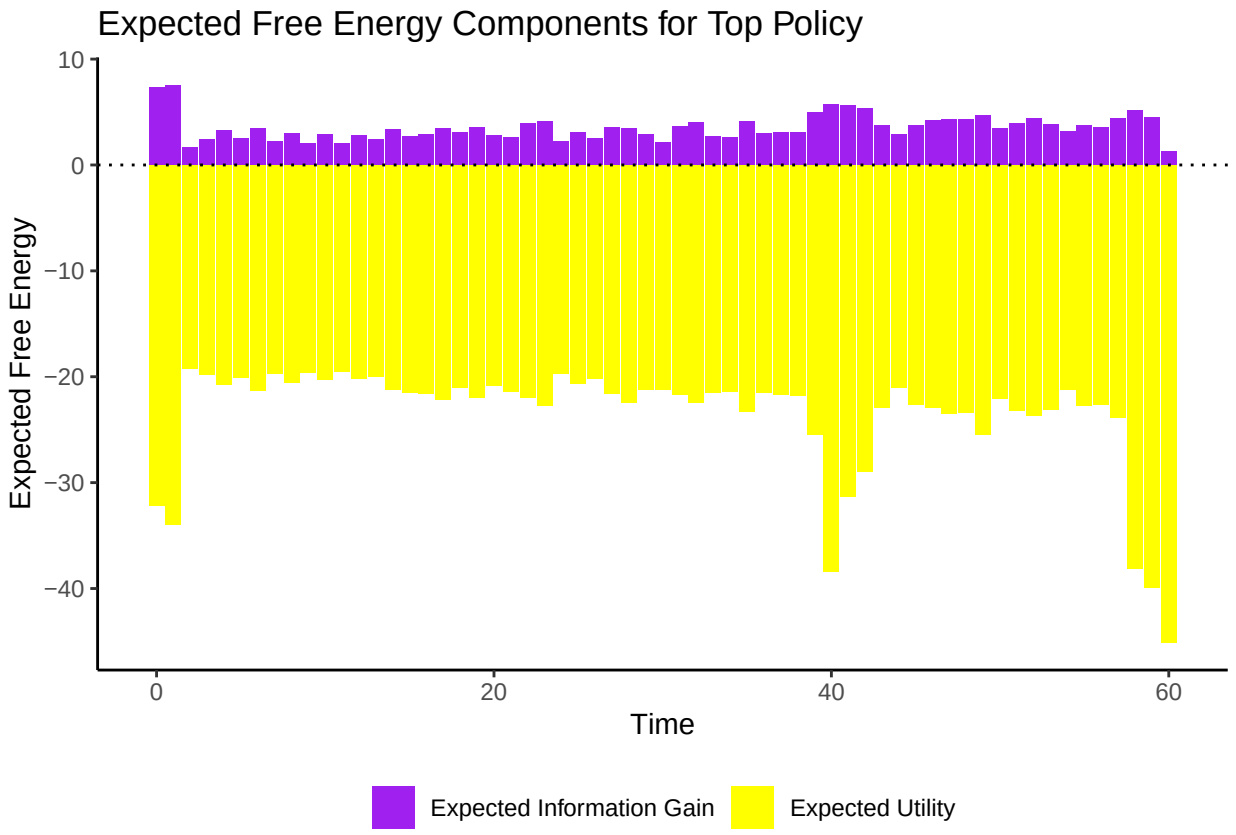
Policy evalutation



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 reaminder 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 doesnt show realized actions but cumultatiove 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 throught time best held withdrawn. A darker shade indicate that on multiple prior timesteps, the model thought acting would be best course of action. Excact 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 enviorment that mimicks particular dynamics of adaptive therapy, POMPD models for which the excact transition probabilties of the enviorment 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 horizion 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 enegery principle. Since free energy can be shown to be the result of factoring in both expected information gain and utility gain, pompds

should be able to plan testing proceudres optimally, given the transitition probabilities, likelihood mappings and prior preferences are correctly specificed. Conceptually this reduces the boundary between testing and treating, since both convey a informational gain. For example, in certain cases, it might be prudent to treat an otherwise mild 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 descprition of the unceartinties 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 the flexiblity of the POMDP structure, computation and our ability to specify the tranisiton 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 suggest 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 POMPDs.

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

Only one range bound

Future Research

NESS Priors over policies

Continuous state space or binning illness

Learning the transition parameters

searching policy space better

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.

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.

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.

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.