Hyperbolic discounting in reinforcement learning

Report submitted in fulfillment of the requirements
for the M.Tech. Project of

Fourth Year IDD

by

Khush Chopra

16074008

Under the guidance of

Dr. Lakshmanan Kailasam

Department of Computer Science and Engineering
INDIAN INSTITUTE OF TECHNOLOGY (BHU) VARANASI
Varanasi 221005, India
June 2020

Dedicated to

My parents, teachers,.....

Declaration

I certify that

1. The work contained in this report is original and has been done by myself and the general supervision of my supervisor.

2. The work has been submitted for any project.

3. Whenever I have used materials (data, theoretical analysis, results) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

4. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Place: IIT (BHU) Varanasi
Date: June 24, 2020

Khush Chopra
IDD Student
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi,
Varanasi, INDIA 221005.

Certificate

*This is to certify that the work contained in this report entitled "**Hyperbolic discounting in reinforcement learning**" being submitted by **Khush Chopra** (**Roll No. 16074008**) carried out in the Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi, is a bona fide work of our supervision.*

Place: IIT (BHU) Varanasi
Date: June 24, 2020

Dr. Lakshmanan Kailasam
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi,
Varanasi, INDIA 221005.

Acknowledgments

I would like to express my gratitude to my supervisor Dr. Lakshmanan Kailasam who gave me the opportunity to work under his guidance and learn while doing the project on the topic of reinforcement learning, which had me doing a lot of research, I got to know about a lot of new things, I are really thankful. Secondly I would also like to thank my parents for their constant love and support and my friends who helped me whenever I were stuck and couldn't get to my mentor.

Place: IIT (BHU) Varanasi

Date: June 24, 2020 Khush Chopra

Abstract

The amount of a future reward should be discounted when there is a risk that the reward will not be realized. If the risk is manifested at a known, fixed hazard rate, a risk-neutral recipient would discount the reward using an exponential time-preference function. Experimental studies using humans and animals, however, exhibit short-term time preferences that differ from the exponential in a manner consistent with a hazard rate that drops with increasing delay. This trend can be explained by an underlying hazard that is uncertainty. Reinforcement learning (RL) generally defines a discount factor (γ) as part of the Markov Decision Process. The value of a future reward is decreased by an exponential scheme. This exponential scheme leads to theoretical convergence guarantees of the Bellman equation because of its properties. We revisit the basics of discounting in RL and demonstrate that a simple approach approximates hyperbolic discount functions while still using familiar temporal-difference learning techniques in RL.

Contents

List of Figures

Chapter 1

Hyperbolic Discounting

Introduction

The standard solution of the reinforcement learning (RL) problem using the Markov Decision Process (MDP) includes a discount factor $0 \leq \gamma \leq 1$ which reduces the present value of future rewards exponentially (Bellman, 1957; Sutton Barto, 1998). A reward r_t received in t-time steps is reduced in value to $\gamma^t r_t$, a discounted utility model introduced by Samuelson (1937). This puts a time preference for rewards to be realized sooner rather than later. The exponential discounting of future rewards by leads to value functions that satisfy theoretical convergence properties (Bertsekas, 1995). The magnitude of γ also plays a role in stabilizing learning dynamics of RL algorithms (Prokhorov Wunsch, 1997; Bertsekas Tsitsiklis, 1996) and has recently been treated as a hyperparameter of the optimization (OpenAI, 2018; Xu et al., 2018).

The magnitude of the γ and the functional use of this parameter implicitly establishes priors over the solution learned. The magnitude of γ chosen sets an effective horizon for the agent to be looking forward to. Rewards that are far beyond this fixed horizon are neglected (Kearns Singh, 2002). This effectively sets a time scale limit to any reward that is fixed in all decision, which might not be an accurate scheme to act on. The exponential discounting of potential future reward is consistent with the

a prior belief that there exists a known constant risk to the agent in the environment (Sozou (1998)). This is a strong assumption that may not be supported in more complex environments.

There is also the problem of inherent cognitive biases, discounting future values exponentially and according to a single discount factor γ does not fit with the measured value preferences in humans and animals (Mazur, 1985; 1997; Ainslie, 1992; Green Myerson, 2004; Maia, 2009). Agents in such experiments were humans, monkeys, rats and pigeons whose discounting patterns match that of a hyperbolic function, where $d_k(t) = \frac{1}{1+kt}$, for some positive $k > 0$ (Ainslie, 1975; 1992; Mazur, 1985; 1997; Frederick et al., 2002; Green et al., 1981; Green Myerson, 2004).

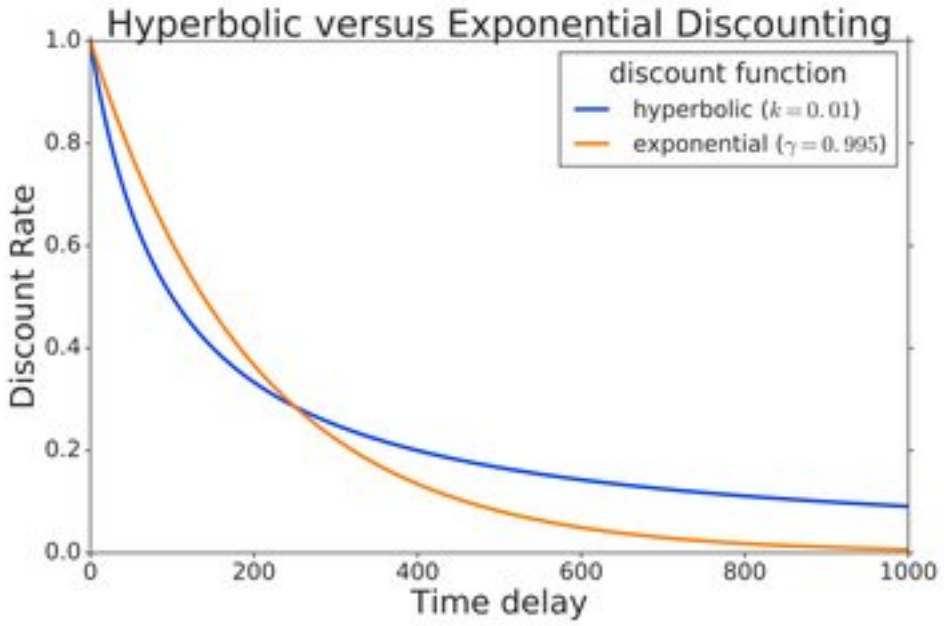

Figure 1.1: Hyperbolic versus exponential discounting. Humans and animals often exhibit hyperbolic discounts (blue curve) which have shallower discount declines for large horizons. In contrast, RL agents often optimize exponential discounts (orange curve) which drop at a constant rate regardless of how distant the return.

Consider this example as an experiment on change in time preference: You are faced with a simple proposition. A person offers you $1M right now with no risk,

but if you choose to wait till tomorrow, you will receive \$1.1M dollars. With no other information, many would be skeptical of this would-be benefactor and choose to get the immediate \$1M reward. With the rationale behind the decision being the belief that the future promise holds risk. However, in a different proposition, you are instead promised \$1M in 365 days or \$1.1M in 366 days. With these new terms many will instead choose the \$1.1M offer. Effectively, the discount rate has decreased as the time to fruition of the reward increased, indicating the belief that it is less likely for the promise to be reneged on the 366^{th} day if it were not already broken on the 365^{th} day. Note that discount rates in humans have been demonstrated to vary with the size of the reward so this time-reversal might not surface for a scenario between \$1 versus \$1.1 (Myerson Green, 1995; Green et al., 1997).

Hyperbolic discounting is consistent with these reversals of time-preferences (Green et al., 1994). Exponential discounting, on the other hand, always remains consistent between these choices and was shown in Strotz (1955) to be the only time-consistent sliding discount function. This discrepancy between the time-preferences of animals from the exponential discounted measure of value might be presumed irrational. However, Sozou (1998) demonstrates that this behavior is mathematically consistent with the agent maintaining some uncertainty over the hazard rate in the environment. In this formulation, rewards are discounted based on the possibility the agent will succumb to a risk and will thus not survive to collect them. Hazard rate, defined later, measures the per-time-step risk the agent incurs as it acts in the environment.

Application in Reinforcement Learning

In deterministic environments like the Arcade Learning Environment (ALE) (Bellemare et al., 2013) stochasticity is introduced using techniques like no-ops (Mnih et al., 2015) and sticky actions (Machado et al., 2018) where the execution of the action is noisy. Physics simulators may also have some noise and randomness which itself induces risk. But even after these stochastic injections, the risk to reward emerges in a more restricted sense. Episode-to-episode risk may vary as the value function and resulting policy evolve. States once safely navigable may become dangerous through catastrophic forgetting (McCloskey Cohen, 1989; French, 1999) or through exploration the agent may venture to new dangerous areas of the state space. However, this is still a narrow manifestation of risk as the environment is generally stable and repetitive. A prior distribution reflecting the uncertainty over the hazard rate, has an associated discount function in the way that an MDP with either this hazard distribution or the discount function, has the same value function for all policies. This equivalence implies that learning policies with a discount function can be interpreted as making them robust to the associated hazard distribution. Thus, discounting serves as a tool to ensure that policies deployed in the real world perform well even under risks they were not trained under.

Chapter 2

Hazards in MDP

Introduction

Sozou (1998) formalizes time preferences in which future rewards are discounted based on the probability that the agent will not *survive* to collect them due to an encountered risk or *hazard*.

Definition 1. Survival s(t) is the probability of the agent surviving until time t.

$$s(t) = P(agent\ is\ alive \mid at\ time\ t) \tag{2.1}$$

A future reward r_t is less valuable presently if the agent is unlikely to survive to collect it. If the agent is risk-neutral, the present value of a future reward r_t received at time-t should be discounted by the probability that the agent will survive until time t to collect it, $s(t)$.

$$v(r_t) = s(t)r_t \tag{2.2}$$

Consequently, if the agent is certain to survive, $s(t) = 1$, then the reward is not discounted per Equation 2.2. From this it is then convenient to define the hazard rate.

Definition 2. Hazard rate h(t) is the negative rate of change of the log-survival at time t.

$$h(t) = -\frac{ds(t)}{dt}\frac{1}{s(t)} \tag{2.3}$$

Therefore the environment is considered hazardous at time t if the log survival is decreasing sharply. Sozou (1998) demonstrates that the prior belief of the risk in the environment implies a specific discounting function. When the risk occurs at a known constant rate than the agent should discount future rewards exponentially. However, when the agent holds *uncertainty* over the hazard rate then hyperbolic and alternative discounting rates arise.

Exponential discounts with known hazards

We recover the familiar exponential discount function in RL based on a prior assumption that the environment has a *knownconstant* hazard. Consider a known hazard rate of $h(t) = \lambda \geq 0$. Definition 2 sets a first order differential equation $\lambda = -\frac{d}{dt} lns(t) = -\frac{ds(t)}{dt}\frac{1}{s(t)}$. The solution for the survival rate is $s(t) = e^{-\lambda t}$ which can be related to the RL discount factor γ

$$s(t) = s^{-\lambda t} = \gamma^t \tag{2.4}$$

This interprets γ as the per-time-step probability of the episode continuing. This also allows us to connect the hazard rate $\lambda \in [0, \infty]$ to the discount factor $\lambda \in [0, 1)$.

$$\gamma = e^{-\lambda} \tag{2.5}$$

As the hazard increases $\lambda \to \infty$, then the corresponding discount factor becomes increasingly myopic $\gamma \to 0$. Conversely, as the environment hazard vanishes, $\lambda \to 0$, the corresponding agent becomes increasingly far-sighted $\gamma \to 1$.

In RL we commonly choose a single γ which is consistent with the prior belief that there exists a known constant hazard rate $\lambda = -ln(\gamma)$. We now relax the assumption that the agent holds this strong prior that it exactly knows the true hazard rate. From a Bayesian perspective, a looser prior allows for some uncertainty in the underlying hazard rate of the environment which we will see in the following section.

Non-Exponential discounts with unknown hazards

We may not always be so confident of the true risk in the environment and instead reflect this underlying uncertainty in the hazard rate through a hazard prior $p(\lambda)$. Our survival rate is then computed by weighting specific exponential survival rates defined by a given λ over our prior $p(\lambda)$.

$$s(t) = \int_{\lambda=0}^{\infty} p(\lambda) s^{-\lambda t} d\lambda \tag{2.6}$$

Sozou (1998) shows that under an exponential prior of hazard $p(\lambda) = \frac{1}{k} exp(-\lambda/k)$ the expected survival rate for the agent is *hyperbolic*

$$s(t) = \frac{1}{1+kt} \equiv \Gamma_k(t) \tag{2.7}$$

We denote the hyperbolic discount by $\Gamma_k(t)$ to make the connection to γ in reinforcement learning explicit. Further, Sozou (1998) shows that different priors over hazard correspond to different discount functions. Figure 2.1 shows the correspondence between different hazard rate priors and the resultant discount functions. The common approach in RL is to maintain a delta-hazard (black line) which leads to exponential discounting of future rewards. Different priors lead to non-exponential discount functions.

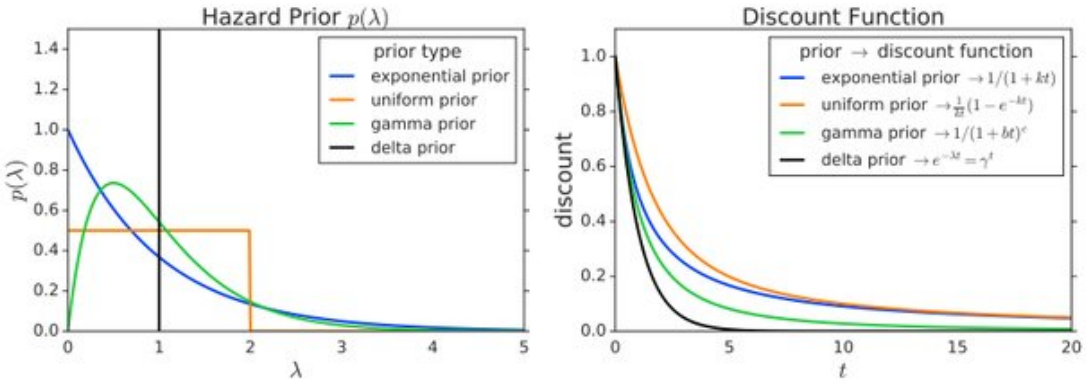

Figure 2.1: Two figures from Sozou (1998). There is a correspondence between hazard rate priors and the resulting discount function. In RL, we typically discount future rewards exponentially which is consistent with a Dirac delta prior (black line) on the hazard rate indicating no uncertainty of hazard rate. However, this is a special case and priors with uncertainty over the hazard rate imply new discount functions. All priors have the same mean hazard rate $\mathbb{E}p(\lambda) = 1$.

Reformulating MDP to include risks

To study MDPs with hazard distributions and general discount functions we introduce two modifications. The hazardous MDP now is defined by the tuple $< S, A, R, P, H, d >$. In standard form, the state space S and the action space A may be discrete or continuous. The learner observes samples from the environment transition probability $P(s_{t+1}|s_t, a_t)$ for going from $s_t \in S$ to $s_{t+1} \in S$ given $a_t \in A$. Considering the case where P is a sub-stochastic transition function, which defines an episodic MDP. The environment emits a bounded reward $r : S \times A \to [r_{min}, r_{max}]$ on each transition. Only considering non-infinite episodic MDPs.

The first difference is that at the beginning of each episode, a hazard $\lambda \in [0, \infty)$ is sampled from the hazard distribution H. This is equivalent to sampling a continuing probability $\gamma = e^{-\lambda}$. During the episode, the hazard modified transition function will be P_λ, in that $P_\lambda(s_{new}|s, a) = e^{-\lambda} P(s_{new}|s, a)$. The second difference is that we now consider a general discount function $d(t)$. This differs from the standard approach of exponential discounting in RL with γ according to $d(t) = \gamma^t$, which is a special case.

A policy $\pi : S \to A$ is a mapping from states to actions. The state action value function $Q_\pi^{H,d}(s, a)$ is the expected discounted rewards after taking action a in state s and then following policy π until termination.

$$Q_\pi^{H,d}(s, a) = \mathbb{E}_\lambda \mathbb{E}_{\pi, P_\lambda} \left[\sum_{t=0}^{\infty} d(t) R(s_t, a_t) | s_0 = s, a_0 = a \right] \tag{2.8}$$

where $\lambda \sim H$ and $\mathbb{E}_{\pi, P_\lambda}$ implies that $s_{t+1} \sim P_\lambda(\cdot|s_t, a_t)$ and $a_t \sim \pi(\cdot|s_t)$.

Chapter 3

Computing Hyperbolic Q-values from Exponential Q-values

Introduction

We can re-purpose exponentially-discounted Q-values to compute hyperbolic (and other-non-exponential) discounted Q-values. The central challenge with using non-exponential discount strategies is that most RL algorithms use some form of TD learning (Sutton, 1988). This family of algorithms exploits the Bellman equation (Bellman, 1958) which, when using exponential discounting, relates the value function at one state with the value at the following state.

$$Q_{\pi}^{\gamma^t}(s,a) = \mathbb{E}_{\pi,P}\left[R(s,a) + \gamma Q_{\pi}(s_{next}, a_{next})\right] \tag{3.1}$$

where expectation $\mathbb{E}_{i,P}$ denotes sampling $a \sim \pi(\cdot|s)$, $s_{next} \sim P(\cdot|s,a)$, and $a_{next} \sim \pi(\cdot|s)$ Being able to reuse the literature on TD methods without being constrained to exponential discounting is thus an important challenge.

Computing Hyperbolic Q-values

Let's start with the case where we would like to estimate the value function where rewards are discounted hyperbolically instead of the common exponential scheme. We refer to the hyperbolic Q-values as Q_π^Γ below in equation 3.3.

$$Q_\pi^{\Gamma_k}(s,a) = \mathbb{E}_\pi \left[\Gamma_k(1)R(s_1,a_1) + \Gamma_k(2)R(s_2,a_2) + ...|s,a\right] \tag{3.2}$$

$$Q_\pi^{\Gamma_k}(s,a) = \mathbb{E}_\pi \left[\sum_t \Gamma_k(t)R(s_t,a_t)|s,a\right] \tag{3.3}$$

We may relate the hyperbolic Q_π^Γ-value to the values learned through standard Q-learning. To do so, notice that the hyperbolic discount Γ_t can be expressed as the integral of a certain function $f(\gamma,t)$ for $\gamma = [0,1)$ in Equation 3.4.

$$\int_{\gamma=0}^{1} \gamma^{kt} d\gamma = \frac{1}{1+kt} = \Gamma_k(t) \tag{3.4}$$

The integral over this specific function $f(\gamma,t) = \gamma^{kt}$ yields the desired hyperbolic discount factor $\Gamma_k(t)$ by considering an infinite set of exponential discount factors γ over its domain $\gamma \in [0,1)$. We visualize the hyperbolic discount factors $\frac{1}{1+t}$(consider $k=1$) for the first few time-steps t in Figure 3.1.

Recognize that the integrand γ^{kt} is the standard exponential discount factor which suggests a connection to standard Q-learning (Watkins Dayan, 1992). This suggests that if we could consider an infinite set of γ then we can combine them to yield hyperbolic discounts for the corresponding time-step t. We build on this idea of modeling many γ throughout this work.

We employ Equation 3.4 and return to the task of computing hyperbolic Q-values.

$$Q_\pi^{\Gamma}(s,a) = \mathbb{E}_\pi \left[\sum_t \Gamma_k(t)R(s_t,a_t)|s,a\right] \tag{3.5}$$

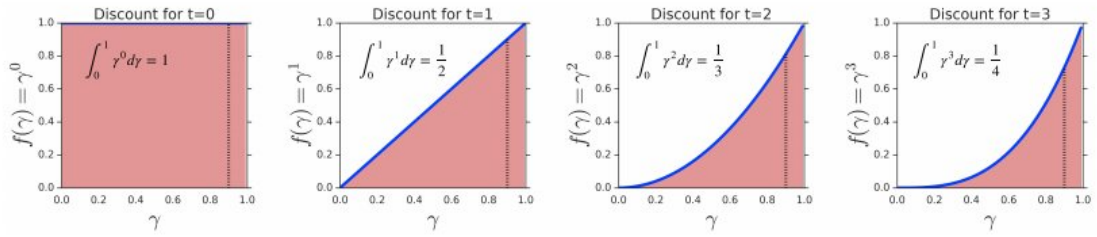

Figure 3.1: From left to right we consider the first four time-steps ($t = 0, 1, 2, 3$) of the function γ^t (shown in blue) over the valid range. The integral (red) of γ^t at time t equals the hyperbolic discount function $\frac{1}{1+t}$ shown in each subplot. Time $t = 0$ is not discounted since the integral of $\gamma^0 = 1$ from 0 to 1 is 1. Then $t = 1$ is discounted by $\frac{1}{2}$, $t = 2$ is discounted by $t = \frac{1}{3}$ and so on. For illustration, the black dotted vertical line indicates the discount that we would use for each time-step if we considered only a single discount factor $\gamma = 0.9$.

$$Q_\pi^\Gamma(s,a) = \mathbb{E}_\pi \left[\sum_t \left(\int_{\gamma=0}^1 \gamma^{kt} d\gamma \right) R(s_t, a_t) | s, a \right] \tag{3.6}$$

$$Q_\pi^\Gamma(s,a) = \int_{\gamma=0}^1 \mathbb{E}_\pi \left[\sum_t R(s_t, a_t)(\gamma^k)^t | s, a \right] d\gamma \tag{3.7}$$

$$Q_\pi^\Gamma(s,a) = \int_{\gamma=0}^1 Q_\pi^{(\gamma^k)^t}(s,a) d\gamma \tag{3.8}$$

where $\Gamma_k(t)$ has been replaced on the first line by $\left(\int_{\gamma=0}^1 \gamma^{kt} d\gamma \right)$ and the exchange is valid if $\sum_{t=0}^\infty \gamma^{kt} r_t < \infty$. This shows us that we can compute the Q_π^Γ-value according to hyperbolic discount factor by considering an infinite set of $Q_\pi^{\gamma^k}$-values computed through standard Q-learning. Examining further, each $\gamma \in [0, 1)$ results in TD-errors learned for a new γ^k. For values of $k < 1$, which extends the horizon of the hyperbolic discounting, this would result in larger γ.

Approximating Hyperbolic Q-values

An equivalence between hyperbolically-discounted Q-values and integrals of exponentially-discounted Q-values requiring evaluating an infinite set of value functions has been discussed in last section. We now present a practical approach to approximate discounting $\Gamma(t) = \frac{1}{1+kt}$ using standard Q-learning.

To avoid estimating an infinite number of Q_π^γ-values we introduce a free hyperparameter (n_γ) which is the total number of Q_π^γ-values to consider, each with their own γ. We use a practically-minded approach to choose G that emphasizes evaluating larger values of γ rather than uniformly choosing points.

$$G = [\gamma_0, \gamma_1, ...\gamma_{n\gamma}] \tag{3.9}$$

Each $Q_\pi^{\gamma^i}$ computes the discounted sum of returns according to that specific discount factor $Q_\pi^{\gamma_i}(s,a) = \mathbb{E}_\pi\left[\sum_t^\pi (\gamma_i)^t r_t | s_0 = s, a_0 = a\right]$,

The set of Q-values permits us to estimate the integral through a Riemann sum (Equation 3.11).

$$Q_\pi^\Gamma(s,a) = \int_0^1 w(\gamma) Q_\pi^\gamma(s,a) d\gamma \tag{3.10}$$

$$Q_\pi^\Gamma(s,a) \approx \sum_{\gamma_i \in G} (\gamma_{i+1} - \gamma_i) w(\gamma_i) Q_\pi^{\gamma_i}(s,a) \tag{3.11}$$

where we estimate the integral through a lower bound. We consolidate this entire process in Figure 3.2 where we show the full process of rewriting the hyperbolic discount rate, hyperbolically-discounted Q-value, the approximation and the instantiated agent. This approach is similar to that of KurthNelson Redish (2009) where each μAgent models a specific discount factor γ. However, this differs in that our final agent computes a weighted average over each Q-value rather than a sampling

operation of each agent based on a γ-distribution.

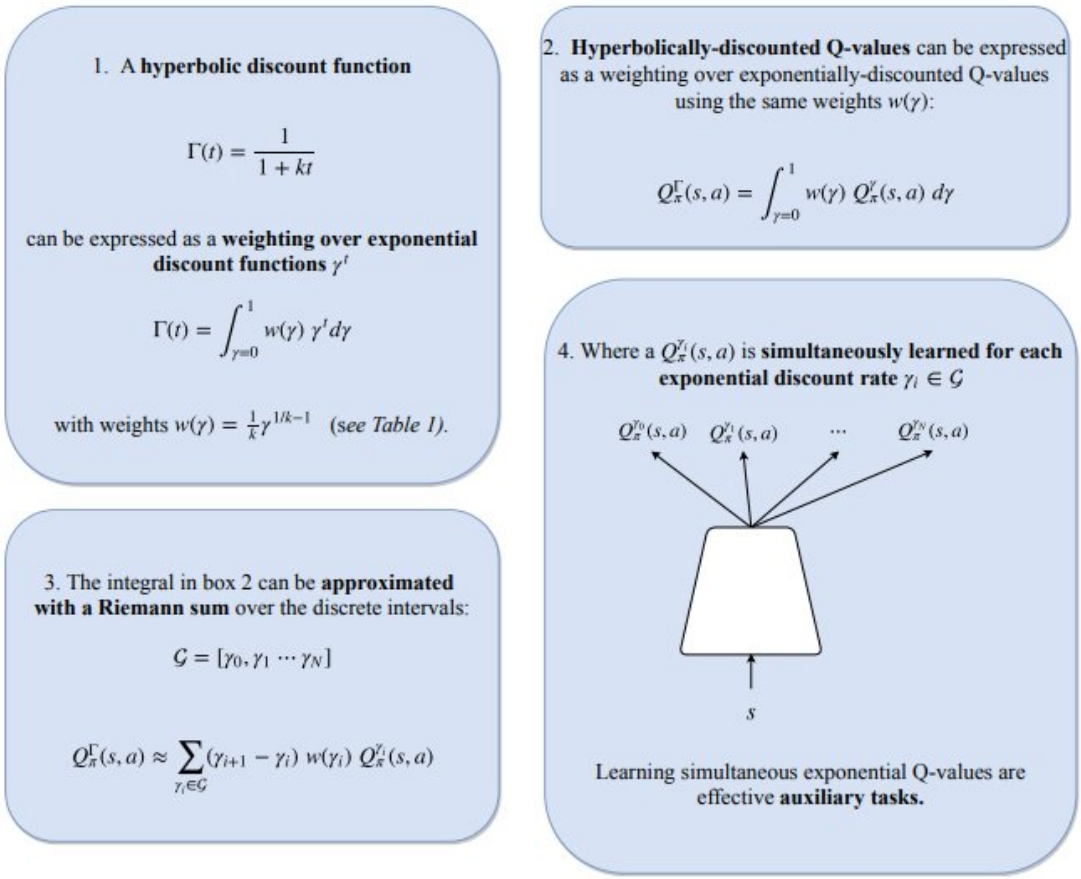

Figure 3.2: Summary of our approach to approximating hyperbolic (and other non-exponential) Qvalues via a weighted sum of exponentially-discounted Q-vaulues

Chapter 4

Experiment

Introduction

The benefits of hyperbolic discounting will be greatest under:

1. Uncertain hazard. The hazard-rate characterizing the environment is not known. For instance, an unobserved hazard-rate variable λ is drawn independently at the beginning of each episode from $H = p(\lambda)$.

2. Non-trivial intertemporal decisions. The agent faces non-trivial intertemporal decision. A non-trivial decision is one between smaller nearby rewards versus larger distant rewards.

Grid world is an environment which can be used to emulate the differences between the 2 discounting methods. Grid world consists of a square divided into chunks and a player can move in any one of the 4 directions. He can not run through the walls and if it tries to, it stays where it is. It's aim it to reach the finishing chunks where it will find it's reward and also end the episode.

Now we can tailor a scenario where the distance between the final states vary and we can compare the effect of the different discounting methods.

The experiment environment consists of uncertain hazard and it's theoretical $k = 0.05$ in the hyperbolic discounting $\frac{1}{1+kt}$.

First scenario has two rewards of immediate value 5 and 6 at a distance of 2 and 7 from the start of the environment respectively.

Second scenario has same two rewards of immediate value 5 and 6 at a distance of 6 and 11 from the start of the environment respectively. We can observe that the relative distance of these rewards has not changed.

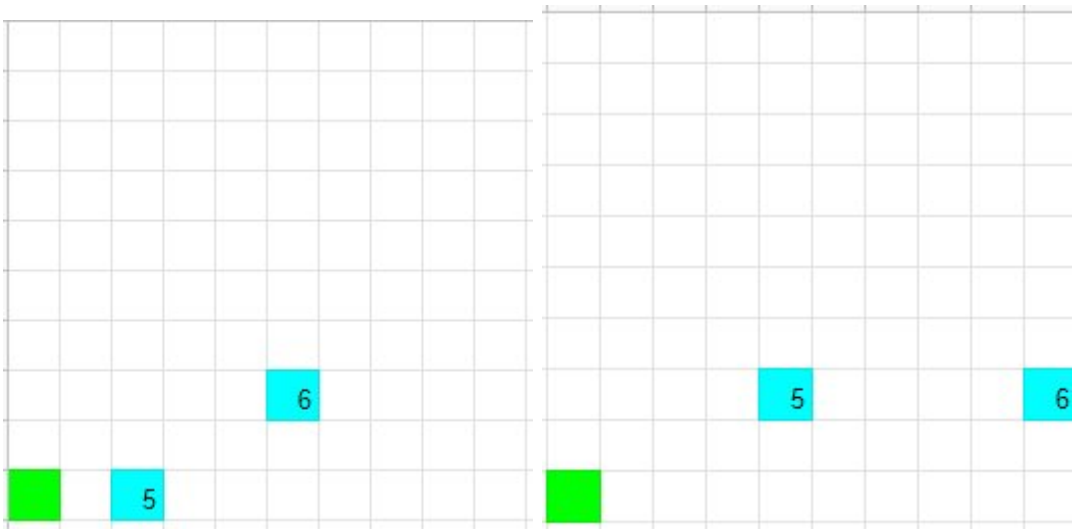

Figure 4.1: Two figures show scenario 1 on the left and scenario 2 on the right. Green is initial start point and blue boxes are end point. Number in blue boxes signifies their reward. As the relative distance of these 2 rewards never changes the exponential discounting model won't see any difference in the 2 scenarios but not hyperbolic discounting.

In both the scenarios, if exponential discounting is used with a $\gamma < 9.65$, then the agent will value the closer, smaller reward of 5 always in comparison to the reward of 6 units which is farther away. Exponential discounting does not change it's decision even with change in the scenario and keeps it's time preference.

Meanwhile the theoretical value for the reward 5 in scenario 1 at t=0 is 4.54 which is greater than the value of reward 6 in scenario 1 at t=0 which is 4.4. But when we come to scenario 2, we can see the change in decision as the Q value for reward 5 is 3.546 which is less than the value of reward 6 which is 3.87.

Hyperbolic discounting has been evolved in humans and animals over the course of their existence. This method of discounting can be very useful in some scenarios.

Chapter 5

Bibliography

William Fedus, Carles Gelada, Yoshua Bengio, Marc G. Bellemare, Hugo Larochelle: Hyperbolic discounting and learning over multiple horizons.

George Ainslie. Specious reward: a behavioral theory of impulsiveness and impulse control. Psychological bulletin.

George Ainslie. Picoeconomics: The strategic interaction of successive motivational states within the person.

William H Alexander and Joshua W Brown. Hyperbolically discounted temporal difference learning.

Marc G Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment: An evaluation platform for general agents.

Marc G Bellemare, Will Dabney, and Rémi Munos. A distributional perspective on reinforcement.

Richard Bellman. A markovian decision process. Journal of Mathematics and Mechanics.

Richard Bellman. On a routing problem. Quarterly of applied mathematics,

Dimitri P Bertsekas. Neuro-dynamic programming: an overview.

Dimitri P Bertsekas and John N Tsitsiklis. Neuro-dynamic programming.

Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. Exploration by random network distillation.

Partha Dasgupta and Eric Maskin. Uncertainty and hyperbolic discounting. American Economic

Nathaniel D Daw. Reinforcement learning models of the dopamine system and their behavioral implications.

Nathaniel D Daw and David S Touretzky. Behavioral considerations suggest an average reward td model of the dopamine system. Neurocomputing,

Peter Dayan and Geoffrey E Hinton. Feudal reinforcement learning. In Advances in neural information processing systems.

Michael Kearns and Satinder Singh. Near-optimal reinforcement learning in polynomial time.

Zeb Kurth-Nelson and A David Redish. Temporal-difference reinforcement learning with distributed representations.

Tiago V Maia. Reinforcement learning, conditioning, and the brain: Successes and challenges. Cognitive, Affective, Behavioral Neuroscience