Foraging models and underlying processes in food-seekings behavior

One of the way classical models such as (Charnov 1976) dealt with modeling foraging in uncertain environments, was with the assumption of perfect knowledge. Animals should stay seeking for food within a patch for as long the capture rate is above the capture rate of the environment (Charnov 1976), which implicitly assumes that somehow the animal is able to compute such capture rate. While such assumptions may sound unrealistic, there is some support for this as an experienced forager may learn and integrate information about the environment to closely approximate the perefect knowledge (Marshall et al. 2016).

On the other hand, and in consequence with the priously exposed relation between foraging and uncertainty, a model presented here should account for such relation. First, the rules determining the results of the interaction between animal and environment are assumed to be unknown or only partially known due to the stochastic nature of the environment. Then, the animal may take any action a within a set of possible actions $a \in A$ for a particular state of the environment s. Any action a causes an stochastic transition from a state s to another state s'. As such the result of an interaction between animal and environment can be described by its value q which is a function of both action and current environment state $q(s, a)$. Such model of action, state and value corresponds to a markov decision process (Sutton and Barto 2018). In this model, all environment dynamics are described by the probabilities $p(s', r|s, a)$, where r is the obtained reward (interaction outcome), and such probabilities is defined for every pair of a and s. We could consider a markov decision process to include the perceptual noise which we deemed inherent to food-seeking behavior, by considering that states s are paired with an observation o made by the animal to infer state s, because state cannot be directly observed or there is some sensory noise. As such, animals consider environment states as conditional probability of any particular observation given a state $p(o|s)$, giving a belief of the current state based of perceptual information (Ma and Jazayeri 2014).

To model how an animal represents the value of a given option $q(s, a)$ in a non-stationary environment, this value is a distribution over possible values, that is updated every time an action a is executed. For the simple case were rewards are obtained or not $q(s, a)$ has a Bernoulli distribution $p(X = reward) = a$ and $(p(X = noreward) = 1 - a)$. Then, this probabilities can be modeled with the Beta distribution which takes parameters α and β. With $\alpha = 1, \beta = 1$ the Beta distributions produces a uniform distribution over $[0, 1]$ succesfully representing the uninformed prior probability for the rewards. To generate the posterior probability every time the reward process results in a reward, the parameter α increases by 1. On the other hand, if no reward is obtained the parameter β increase by 1. Finally, the mean is defined as

$$\frac{\alpha}{\alpha + \beta}$$

and its variance by

$$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

With this simple statistical properties of the Beta distribution we can represent uncertainty over the expected rewards for any given a and s. If the exploration is defined by the posterior then it can be considered a Thomposon sampling strategy (Thompson 1933). To select an action a a posterior is built for every action and updated according to the previously stated rules, then for each posterior a reward estimate $\hat{r}$ is sampled greedily so the action selected is $a = argmax_{a \in A} \hat{r}(a)$ where A is the set of possible actions within an environment (Wang and Zhou 2020). This processes must be performed for every state, limiting tractability by the number of states. In general terms a solution for this is to consider the reward vector as a weighted average over past rewards, with a step-size parameter ($(0, 1]$), the lower the value of this parameter more weight is given to recent rewards, on the other hand, if its closer to 1 then all the reward history is equally considered. More complex consideration of this problem include modeling non-stationarity as Poisson arrival process that modifies the means rewards (Ghatak 2020), bayesian approaches to modulate past observed rewards (Raj and Kalyani 2017), and explicitly modeling environment volatility in a bayasian setup (Behrens et al. 2007).

While this general model can work in non-stationary environments it doesnt consider explicitly the belief of the current state based on the perceptual information received $p(o|s)$. For this addition a probability for every $o \in O$ by state is necessary, where O is the set of all particular observations o. To model state beliefs the goal is obtain the function that finally maps observations o to action a given an underlying model that relates states with observations, a hidden markov model (HMM) represents this. HMM generates conditional probability distributions $p(o|s)$ and bayesian, among other methods for obtaining such model given only actions and observation has been proposed (Funamizu et al. 2012; Yoon, Lee, and Hovakimyan 2018; Piray and Daw 2020).

In this section we offered the elementary considerations for a model of food-seeking behavior in non-stationary environments with uncertainty over action outcomes due to perceptual limitations or noise. Thompson sampling was considered as the base for this due to its simplicity and elegance in modeling exploration/exploitation by computing uncertainty. The goal of these consideration was not to establish or to specify a complete model, but to provide a framework relating uncertainty with the exploration/exploitation dilemma and perceptual limitations shown theoretically and empirically in the previous section.

Behrens, Timothy E J, Mark W Woolrich, Mark E Walton, and Matthew F S

Rushworth. 2007. "Learning the Value of Information in an Uncertain World." *Nature Neuroscience* 10 (9): 1214–21. https://doi.org/10.1038/nn1954.

Charnov, Eric L. 1976. "Optimal Foraging, the Marginal Value Theorem." *Theoretical Population Biology* 9 (2): 129–36. https://doi.org/10.1016/0040-5809(76)90040-X.

Funamizu, Akihiro, Makoto Ito, Kenji Doya, Ryohei Kanzaki, and Hirokazu Takahashi. 2012. "Uncertainty in action-value estimation affects both action choice and learning rate of the choice behaviors of rats." *The European Journal of Neuroscience* 35 (7): 1180–89. https://doi.org/10.1111/j.1460-9568.2012.08025.x.

Ghatak, Gourab. 2020. "A Change-Detection Based Thompson Sampling Framework for Non-Stationary Bandits." *arXiv:2009.02791 [Cs, Eess]*, September. http://arxiv.org/abs/2009.02791.

Ma, Wei Ji, and Mehrdad Jazayeri. 2014. "Neural Coding of Uncertainty and Probability." *Annual Review of Neuroscience* 37 (1): 205–20. https://doi.org/10.1146/annurev-neuro-071013-014017.

Marshall, Louise, Christoph Mathys, Diane Ruge, Archy O. de Berker, Peter Dayan, Klaas E. Stephan, and Sven Bestmann. 2016. "Pharmacological Fingerprints of Contextual Uncertainty." Edited by Matthew F. S. Rushworth. *PLOS Biology* 14 (11): e1002575. https://doi.org/10.1371/journal.pbio.1002575.

Piray, Payam, and Nathaniel D. Daw. 2020. "A simple model for learning in volatile environments." *PLoS computational biology* 16 (7): e1007963. https://doi.org/10.1371/journal.pcbi.1007963.

Raj, Vishnu, and Sheetal Kalyani. 2017. "Taming Non-Stationary Bandits: A Bayesian Approach." *arXiv:1707.09727 [Cs, Stat]*, July. http://arxiv.org/abs/1707.09727.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press.

Thompson, William R. 1933. "On the Likelihood That One Unknown Probability Exceeds Another in View of the Evidence of Two Samples." *Biometrika* 25 (3/4): 285. https://doi.org/10.2307/2332286.

Wang, Zhendong, and Mingyuan Zhou. 2020. "Thompson Sampling via Local Uncertainty." *arXiv:1910.13673 [Cs, Stat]*, August. http://arxiv.org/abs/1910.13673.

Yoon, Hyung-Jin, Donghwan Lee, and Naira Hovakimyan. 2018. "Hidden Markov Model Estimation-Based Q-Learning for Partially Observable Markov Decision Process." *arXiv:1809.06401 [Cs, Stat]*, September. http://arxiv.org/abs/1809.06401.