Analysis of stochastic choice in foraging

Mark D. Humphries

Version 1.0
Started: 28 July 2023

Address: School of Psychology, University of Nottingham

Corresponding author: mark.humphries@nottingham.ac.uk

Abstract

Optimal models of patch foraging assume a deterministic choice rule. We analyse here a stochastic choice model of patch foraging, based on exploit-explore action selection models from reinforcement learning.

1 Introduction

We introduce here a model of patch foraging that uses stochastic choice for its stay-or-leave decision. In particular, we propose a model that has parametric control over the exploration-exploitation tradeoff. We analyse its expected leaving times (or, equivalently, its dwell time in each patch), and the variance in leaving time as a function of these parameters.

We assume here discrete time-steps (or "trials") throughout, as this allows a direct comparison with the simulated model(s) used to fit participant data. It also makes the maths easier. The translation to continuous time is considered in Section 5.1.

Notation for the models
n: the integer time-step (or trial number)
$t(n)$: the clock time of time-step n
$r(n)$: the amount of reward in the patch at time-step n
r_0: the initial amount of reward in the patch before harvesting
α: the decay rate of $r(n)$
R: the average or "background" reward rate in the environment
β: inverse temperature parameter of the softmax function (higher values mean more exploitation – lower probability of leaving) ϵ: fixed probability of leaving per time-step

2 The stochastic choice model

We define models for the probability of leaving a patch on the nth time-step given the current reward rate $r(n)$. These models are derived from their reinforcement-learning equivalents.

The ϵ-greedy model gives a fixed probability of leaving per time-step. In the first form, ϵ is the probability of leaving:

$$p(\text{leave}|n) = \epsilon. \tag{1}$$

We will use this model throughout.

In the second form, exactly mimicking RL models, we consider staying and leaving as two possible actions to choose between, and that the default action is to stay. With some probability ϵ the agent chooses uniformly at random between the two available actions, giving a probability of 1/2 that leaving will be chosen. Hence the probability of leaving on time-step n is

$$p(\text{leave}|n) = \frac{1}{2}\epsilon. \tag{2}$$

Hence all results for model 2 can be obtained from model 1 by simply halving $p(\text{leave}|n)$.

The softmax choice model is:

$$p(\text{leave}|n) = \frac{1}{1+\exp(\beta r(n))}. \tag{3}$$

We also can consider a softmax model with a bias term

$$p(\text{leave}|n) = \frac{1}{1+\exp(c+\beta r(n))}, \tag{4}$$

which allows us to partition exploration into a fixed bias c and a reward dependent term β. The notes below are currently for model 3.

3 Reward functions

We consider here two standard decaying reward functions.

Exponential decay is given by

$$r(n) = r_0 \exp(-\alpha n). \tag{5}$$

Linear decay is given by

$$r(n) = \begin{cases} r_0 - \alpha n & \text{r}_0 \geq \alpha\text{n} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

4 MVT predictions for leaving time

The key metric for foraging is the leaving time for a patch i.e. how long an agent will stay in the patch before leaving. Optimal foraging theory (MVT) says that agents should leave as soon as the reward rate in the patch $r(n)$ is equal to the long-run average rate R for the environment. MVT predicts independent patch and environment effects on leaving time. It predicts leaving times will be longer for richer patches (the patch effect). Conversely, it predicts leaving times for a given patch will be longer in a poorer environment, one with a lower average reward rate (the environment effect). For a given patch we can calculate the predicted MVT leaving time assuming R is known.

[THESE NEED DOUBLE-CHECKING]

For the exponentially-decaying reward function we set $R = r_0 \exp(-\alpha n)$ and solve for n

$$n = -\ln\left(\frac{R}{r_0}\right)\frac{1}{\alpha}. \tag{7}$$

For the linearly-decaying reward function we set $R = r_0 - \alpha n$ and solve for n

$$n = \frac{r_0 - r}{\alpha}, \tag{8}$$

where we can ignore the second case in Eq. 6 as we can guarantee that $R \geq 0$.

Whatever the reward function, if n is a discretisation of continuous time, then the expected time is $t(n)$.

5 How to compute expectation and variance for leaving times

We want our models to replicate both patch and environment effects. We also would like them to explain overharvesting: that agents stay longer in patches than predicted by MVT. Consequently, we want to calculate the expected leaving time for each model. But given the choice is stochastic we also want the predicted variation around that expected time. We do this as follows.

The models give us $p(\text{leave}|n)$, the probability of leaving on a given time-step n. We want to know the probability $p(\text{leave} = n)$ that time-step n is the first trial at which leaving occurs. Each time-step n has only two possible outcomes (stay or leave) and is thus a Bernoulli trial; the sequence of time-steps n is a Bernoulli process. The probability $p(\text{leave} = n)$ is thus

$$p(\text{leave} = n) = p(\text{leave}|n) \prod_{i=1}^{n-1} 1 - p(\text{leave}|i), \tag{9}$$

where the first term is the probability of leaving on this time-step and the second term is the cumulative probability of not having left on all the previous time-steps.

The tasks we model can have discrete trials or states in which a stay-or-leave decision is made after every harvest in a patch. We can thus compute exact values of expectation and variance by plugging Eq. 9 into the standard equations for expectation ($\sum xp(x)$) and variance ($\sum(x-E)^2p(x)$) of a random variable (x). Here the random variable is n, the number of elapsed trials.

The expected leaving time is therefore

$$\text{E}(\text{leave}) = \sum_{n=1}^{\infty} np(\text{leave} = n). \tag{10}$$

Similarly, the variance in leaving time is therefore

$$\text{VAR}(\text{leave}) = \sum_{n=1}^{\infty} (n - \text{E}(\text{leave})))^2 \, p(\text{leave} = n). \tag{11}$$

In special cases these equations have exact solutions. Otherwise, to evaluate these numerically, we replace ∞ with some large number $n_{\max}$ as the upper limit of time-steps over which to evaluate the expectation. We need to choose that limit such that $n_{\max} \gg \text{E}(\text{leave})$.

5.1 Continuous time

For tasks with continuous time in a patch we can treat n as discrete time-steps approximating the continuous time, to match how such tasks are handled in the simulated models. Each time-step is $\delta = t(n) - t(n-1)$; equivalently: $t(n) = n\delta$.

The discrete probabilities defined above are for each trial, which have unit time (i.e. $t(n) = n$). So if we change the time-step δ from unit time then we also need to scale the probability of leaving $p(\text{leave}|n)$ to get the same probability per unit time. Namely:

$$p(\text{leave}|n)_\delta = p(\text{leave}|n)\delta. \tag{12}$$

We plug $p(\text{leave}|n)_\delta$ into equation 9 as before, obtaining $p(\text{leave} = n)_\delta$. We then compute the expectation and variance using the clock time $t(n)$ of each time-step

$$\text{E}(\text{leave}) = \sum_{n=1}^{\infty} t(n)p(\text{leave} = n)_\delta. \tag{13}$$

$$\mathrm{VAR}(\text{leave}) = \sum_{n=1}^{\infty} \left(t(n) - \mathrm{E}(\text{leave})\right)^2 p(\text{leave} = n)_\delta. \tag{14}$$

Numerical calculation of these again uses some upper limit $n_{\max}$ in place of ∞. We need to choose that limit such that $t(n_{\max}) \gg \mathrm{E}(\text{leave})$.

Numerical simulations confirm that E(leave) and VAR(leave) converge on stable values as $\delta \to 0$. Nonetheless it would be good to handle continuous time directly, rather than as the approximations above. The limiting case of the taking $\delta t \to 0$ for a Bernoulli process is the Poisson process, which we will come back to at a later date.

6 Choice behaviour for the ϵ-greedy model

This model has a fixed probability of leaving per time-step. The probability of leaving (Eq. 9) is thus independent of the reward function $r(n)$

$$\begin{aligned} p(\text{leave} = n) &= \epsilon \prod_{i=1}^{n-1} 1 - \epsilon, \\ &= \epsilon(1-\epsilon)^{n-1}. \end{aligned} \tag{15}$$

Consequently, the expectation and variation in leaving time are also independent of the reward function $r(n)$. They are found by plugging equation (15) into equations (10) and (11) for discrete trials or equations (13) and (14) for continuous time.

If we have discrete trials n then equation (15) is a geometric distribution with exact solutions for its expectation and variance (see e.g. Wikipedia)

$$\begin{aligned} \mathrm{E}(\text{leave}) &= \frac{1}{\epsilon} \\ \mathrm{VAR}(\text{leave}) &= \frac{1-\epsilon}{\epsilon^2}. \end{aligned}$$

It is thus obvious that:

- There is a monotonic relationship between ϵ and leaving time: increasing ϵ shortens the leaving time (as expected)
- There is a monotonic relationship between ϵ and the variation in leaving time: increasing ϵ reduces the variance.
- There is no patch effect as changing r_0 will not affect E(leave)
- The environment effect can be captured by changing ϵ

7 Choice behaviour for the softmax model

7.1 Exponentially decaying rewards

Putting the exponential reward decay (Eq. 5) into the softmax model gives:

$$p(\text{leave}|n) = \frac{1}{1 + \exp(\beta r_0 \exp(-\alpha n))}. \tag{16}$$

It is immediately clear that β and r_0 together form single parameter $B = \beta r_0$ that sets the explore-exploit tradeoff in foraging. As β and r_0 can be changed independently, this could allow the environment and patch effect. The environment effect between patches of the same type could thus be achieved by changing β with r_0 held constant. The patch effect in a given environment is thus a change in r_0 bewteen patches with β held constant.