Mallard Population

Stéphanie Larocque and Philip Paquette

April 27, 2018

1 Introduction and Motivation

We decided to implement the environment defined in the following paper : *Anderson, David R. "Optimal exploitation strategies for an animal population in a Markovian environment: a theory and an example." Ecology 56.6 (1975): 1281-1297.* It models the size of the Mallard ducks population at some time t, knowing a few important characteristics : the number of ponds (which are important for breeding young ducks), the amount of rain fall (sampled from a Gaussian random variable and influences the number of ponds) and the number of adult birds in the previous year. These three variables influence the number of young Mallards produced in that year, and thus the total size of the population. Another important component of the model is the number of ducks harvested in the previous year, because it influences the size of the population.

The goal was to provide an optimal exploitation strategy, where a good strategy would harvest the maximal number of ducks without damaging the size of the population for the next years. Having good models of population rate and optimal exploitation rates are key elements that are required to reduce animal overexploitation.

2 Model description

2.1 Variables

To model the size of the Mallard population, the following random variables were used:

- N_t : size of adult breeding population in year t (May 15th)
- P_t : number of ponds present on central breeding grounds in year t
- Y_t : number of young produced in year t (birds capable of flight on September 15th)
- R_t : precipitations (inches) during interval $t \rightarrow t+1$

The size of the adult breeding population in year $t+1$ depends on both the size of adult population in year t and on Y_t, the number of young produced in year t, together with the survival rates of young and adult ducks. Let the following be the survival rates for some birds during a specified period:

- ϕ^s : survival rate of adult ducks (between May 15th - September 15th)
- ϕ^a_t : survival rate of adults (a), in year t (between September 15th - May 15th)
- ϕ^y_t : survival rate of young (y), in year t (between September 15th - May 15th)

2.2 Equations

The size of adult breeding population in year $t+1$ is defined as

$$N_{t+1} = N_t \phi_t^a + Y_t \phi_t^y, \tag{1}$$

whereas the number of young Y_t produced in year t depends both N_t and P_t (since the presence of ponds is crucial for breeding new birds) in the following way:

$$Y_t = \left(\frac{1}{12.48 P_t^{0.851}} + \frac{0.519}{N_t} \right)^{-1} \tag{2}$$

The parameters were estimated by the authors. The number of ponds P_t in year t depends on both P_{t-1} and on the precipitations during period $t-1 \to t$.

$$P_t = -2.76 + 0.391 P_{t-1} + 0.233 R_{t-1}, \tag{3}$$

and where R_t is the amount of precipitations received during year t and can be approximated by an i.i.d. Gaussian random variable as follows:

$$R_t \sim \mathcal{N}(\mu = 16.46, \sigma^2 = 4.41) \tag{4}$$

2.3 Harvest rate

Finally, the survival rates in year t depend on the harvest rate during the same period. The harvest rate is defined as the ratio of ducks harvested to the size of the fall population as follows:

$$F_t = \phi^s N_t + Y_t \tag{5}$$

$$H_t = \frac{D_t}{F_t}, \tag{6}$$

where

- F_t : size of fall population on September 15th, year t
- D_t : number of ducks harvested in year t
- H_t : exploitation (harvest) rate in year t

2.4 Survival rates

Survival rates depends on the exploitation rate, but also on the model for mortality. Two hypothesis were tested : whether the mortality from harvesting was additive to usual mortality in the Mallard population or if it was compensative for some of the losses that would have been encountered anyways. In the additive mortality setting, the parameters were estimated to be:

$$\phi^s = 0.92 \tag{7}$$

$$\phi_t^a = 1 - 0.37 \exp(2.78 H_t) \tag{8}$$

$$\phi_t^y = 1 - 0.49 \exp(0.90 H_t) \tag{9}$$

While, in the compensative setting, the survival rates for adults (a) and young (y) in year t differed and were estimated to be :

$$\phi_t^a = \begin{cases} 0.57 - 1.2\exp(H_t - 0.25) & H_t > 0.25 \\ 0.57 & H_t \leq 0.25 \end{cases} \tag{10}$$

$$\phi_t^y = \begin{cases} 0.50 - 1.0\exp(H_t - 0.25) & H_t > 0.25 \\ 0.50 & H_t \leq 0.25 \end{cases} \tag{11}$$

2.5 Objective

The state variable X_t is defined as $X_t = (N_t, P_t)$. Our objective is to provide an optimal exploitation strategy in order to maximize the harvested ducks over the years. The more we harvest, the more reward we get, but it might affect the population at the next time step. Our goal is to obtain the greater return over the years.

The action taken will be the number of ducks to harvest in year t. If we have N_t ducks at the beginning, then the action will fall between 2M birds and (N_t)M birds. We discretized this interval in our model. Furthermore, even if our decision D_t is fixed, we have an additional stochasticity arising from the fact that the real number of ducks harvested might be different of what we planned. To model this, we used another random variable K_t, the real number of ducks harvested. We defined it as:

$$K_t(D_t) = \begin{cases} 0.9 \cdot D_t & \text{with probability } 0.2 \\ D_t & \text{with probability } 0.6 \\ 1.1 \cdot D_t & \text{with probability } 0.2 \end{cases} \tag{12}$$

The expected return in year t would thus be

$$\overline{r_t} = \sum_{k_t} P(k_t) r_t(X_t, D_t, k_t), \tag{13}$$

where r_t is simply the number of ducks harvested in year t.

3 Results

3.1 From the paper

The optimal decision matrices for the additive and compensatory mortality settings from the paper show the optimal number of ducks to harvest in year t, given the size of the population N_t and the number of ponds P_t. For example, in the additive mortality setting (Table 1), if we observe 9.0M birds with 2.0M ponds, then we would choose to harvest 2.6M of ducks. Similarly, the optimal exploitation strategy found in the paper for compensatory mortality is showed in Table 2.

3.2 Q-learning and variations

We implemented tabular and approximate DQN. We used a continuous observation space of size (2,) that contained the number of adults ducks (in millions) and the number of ponds (in

Table 1: Optimal decision matrix for additive mortality
(All figures in millions)

	Ponds (P_t)						
Breeding Population (N_t)	0.5	1.0	1.5	2.0	2.5	3.0	3.5
6	2.0	2.0	2.0	2.0	2.0	2.0	2.0
7	2.0	2.0	2.0	2.0	2.1	2.2	2.3
8	2.0	2.1	2.2	2.3	2.4	2.6	2.8
9	2.0	2.4	2.6	2.8	2.9	3.2	3.4
10	2.0	2.7	3.1	3.3	3.5	3.8	4.1
11	2.2	3.1	3.6	3.9	4.1	4.5	4.8
12	2.4	3.4	4.1	4.4	4.7	5.1	5.5
13	2.6	3.9	4.7	5.1	5.4	5.6	6.2
14	2.9	4.2	5.3	5.8	6.1	6.6	7.0
15	3.2	4.8	5.9	6.5	7.0	7.5	8.0
16	3.6	5.3	6.6	7.3	7.8	8.5	9.0
17	4.0	5.7	7.3	8.0	8.8	9.7	10.4
18	4.5	6.5	8.1	9.1	10.0	10.8	11.9

Table 2: Optimal decision matrix for compensatory mortality
(All figures in millions)

	Ponds (P_t)						
Breeding Population (N_t)	0.5	1.0	1.5	2.0	2.5	3.0	3.5
6	3.0	3.4	3.6	3.7	3.8	3.9	4.0
7	3.2	3.5	3.8	3.9	4.0	4.1	4.2
8	3.4	3.8	4.1	4.3	4.5	4.7	4.8
9	3.7	4.0	4.2	4.5	4.8	5.1	5.3
10	3.9	4.4	4.8	5.0	5.2	5.4	5.6
11	4.2	4.7	5.1	5.4	5.6	5.9	6.1
12	4.4	4.9	5.4	5.7	6.0	6.3	6.5
13	4.6	5.3	5.9	6.3	6.4	6.7	7.0
14	4.8	5.5	6.1	6.4	6.7	7.1	7.4
15	5.1	5.8	6.4	6.8	7.1	7.5	7.8
16	5.3	6.1	6.8	7.2	7.5	8.0	8.3
17	5.6	6.4	7.2	7.5	7.8	8.3	8.7
18	5.8	6.7	7.5	7.9	8.2	8.7	9.2

millions). We used a discrete action space of size (100,) representing the percentage of the population the agent wanted to kill for a given year. The reward function was the actual number of ducks killed during a given year. The episode terminates if the duck population goes below 5 million, if the agent killed less than 2 million ducks in a year (minimum annual legislative requirement), or if 25 time steps have passed.

For tabular Q-Learning, we mapped the continuous observation space to a discrete action space (i.e. number of adults from 6 to 18M, and number of adults from 0.5 to 3.5M). The total number of state-action was 9,100. To force exploration, we used optimistic initial conditions. Moreover, we used a large negative reward if the number of birds fell below 5M and less than 2M birds were killed in a year to avoid these behaviors.

The model converged to exploitation rates that were different than the paper. In the top left corner (Table 3), the model wanted to kill close to 100% of birds. This is probably due to the fact that the model was not able to avoid the minimal number of birds threshold and therefore maximized the immediate reward.

Table 3: Decision matrix for tabular Q-Learning, additive mortality
(All figures in millions)

	Ponds (P_t)						
Breeding Population (N_t)	0.5	1.0	1.5	2.0	2.5	3.0	3.5
6	5,7	5,3	2,2	2,3	2,4	2,5	2,8
7	6,4	2,2	2,4	2,2	2,2	2,2	2,3
8	2,2	2,4	2,6	2,2	2,6	2,2	2,2
9	2,3	2,2	2,4	2,3	2,9	2,5	2,4
10	2,1	2,4	2,5	2,7	2,4	2,7	2,1
11	2,1	2,2	2,2	2,3	2,9	2,9	4,3
12	2,3	2,2	2,3	2,3	2,3	3,7	3,2
13	2,1	2,2	2,7	3,0	3,1	2,2	6,0
14	2,1	2,7	2,2	3,2	3,9	4,8	6,2
15	2,1	2,7	3,5	5,0	5,1	2,4	5,6
16	2,1	3,0	3,7	5,6	2,7	6,1	7,0
17	2,0	2,7	3,9	2,9	6,0	5,1	5,1
18	2,3	2,9	4,0	5,8	7,4	7,9	8,1

The results are consistent with the paper. The number of animals to kill increases both with the current population and the number of ponds. The exploitation ratio of the Q-Learning agent is more conservative than the original paper estimate, as you can see in Table 3.

We also implemented an approximate Q-Learning method. The approximate method had more difficulty converging, probably because of the high stochasticity in the environment, and the high correlation between the trajectories (we didn't use experience replay). The approximate Q-Learning used a single hidden layer (100 units) with a relu activation.

3.3 Policy Gradient and variations

We also tried implementing variations of the policy gradient methods. More specifically, 1) a vanilla actor-critic method, 2) a n-step actor-critic, 3) a GAE actor-critic, and 4) a GAE-PPO agent.

We had difficulty training the agents. Some of them converted to the same exploitation rate for all population sizes, while others were updating too many states at once and couldn't converge. We think that the fact that the size of the observation space (2,) was small compared to the action space (100,) and that trajectories were highly correlated was some reasons that caused this problem.

We modeled the policy network using 2 hidden layers (100 units each) with a relu activation, followed by a linear layer to compute the logits. Moreover, we modeled the critic network using a single hidden layer (100 units with relu activation) followed by a linear layer to compute the state value.

4 Conclusion

We implemented a Markovian model of the population growth of Mallard ducks in the OpenAI gym environment, and also implemented examples using Q-Learning and Policy Gradient. We obtained different exploitation rates than the paper, but we are confident in our implementation. We think models of these kind are important to reduce animal over-exploitation.

Contribution of each member

- Stéphanie Larocque (50%) (Reading paper, modeling the env, results analysis, writing report, presentation)
- Philip Paquette (50%) (Reading paper, modeling, writing code, hparams search, analysis, writing report)