Optimal exploitation strategies for an animal population in a Markovian environment

Stéphanie Larocque and Philip Paquette

Université de Montréal

April 24, 2018

Introduction and Motivation

- 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.*
- Models the size of the Mallard ducks population at each year (N_{t+1}) knowing number of ponds (P_t), and number of adult birds in the previous year (N_t)
- Other variables includes amount of rain fall (R_t) and number of young produced (Y_t)
- Action = number of ducks to harvest (D_t)
- Goal = provide optimal exploitation strategy without damaging population
- Good models + optimal exploitation rate = key elements required to reduce animal overexploitation

Figure 1: A Mallard duck

State variable
$X_t = (N_t, P_t)$

Action $k_t \in (0.9D_t, D_t, 1.1D_t)$
with proba (0.2, 0.6, 0.2)

Variables and equations

In year t, variables includes the size of population in May (N_t), the number of ponds (P_t), the number of young produced (Y_t) and the amount of precipitations (R_t), in inches.

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

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

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

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

where ϕ_t^a and ϕ_t^y are survival rates for adults (a) and young (y) ducks.

Let H_t be the harvest rate (nb ducks harvested/size of the population).
With additive mortality, the survival rates were estimated to be:

$$\phi_t^a = 1 - 0.37 \exp(2.78 H_t) \tag{5}$$

$$\phi_t^y = 1 - 0.49 \exp(0.90 H_t) \tag{6}$$

Expected Return and Results

State variable
$X_t = (N_t, P_t)$

Action $k_t \in (0.9D_t, D_t, 1.1D_t)$
with proba (0.2, 0.6, 0.2)

Expected return in year t :

$$\overline{r_t} = \sum_{k_t} P(k_t) r_t(X_t, D_t, k_t), \qquad (7)$$

where r_t = nb of ducks harvested.

- Shows the optimal number of ducks to harvest in year t, given N_t and P_t
- Number of ducks to harvest increase with both N_t and P_t
- Continuous obervation space of size (2,)
- Discretized action space of size (100,) : percentage of the population the agent would like to harvest for a given year
- Reward function = actual number of ducks harvested
- Objective = provide optimal exploitation strategy to maximize the harvested ducks over the years

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

Q-learning and variations

- For Q-Learning : mapped the observation space to a discrete state-action space. Optimistic initial conditions. 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.
- Different exploitation rates than in the paper (ex : top left corner, probably due to boundary conditions, immediate reward)
- Still consistent with the paper. The number of animals to harvest increases both N_t and P_t. The exploitation ratio of the Q-Learning agent is more conservative than the original paper estimate.
- Approximate DQN :more difficulty converging (stochasticy and high correlation between the trajectories, no experience replay)

Table 2: 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

Conclusion and Future Work

Policy Gradient

- Policy Gradient variations tested :
  1. Vanilla actor-critic method
  2. n-step actor-critic
  3. GAE actor-critic
  4. GAE-PPO agent
- Difficulty training (same exploitation rate for all states, updating too many states at once, couldn't converge). Probably because size of observation space was small compared to action space and that trajectories were hightly correlated.

Conclusion

- Implementation of a Markovian model of the population growth of Mallard ducks in the OpenAI gym environment
- Implementation of baselines using Q-Learning and Policy Gradient
- Different exploitation rates with the paper, but confident in our implementation
- Important model to prevent animal over-exploitation