I want to create a microeconomic agent based model to investigate the effect of services such as too good to go, which allow food vendors to give at a discount leftovers at the end of the day. The outcome I am interested in is in particular total production and total waste, as parameters change.

The broad setting I want to create is the following. Each day, a baker can choose to produce how many units of each type of good to produce. Consumers have specific preferences (which do not change) and they decide whether to go to the store or not based on an indepdent need probability (this introduces variability in the demand). If a consumer decides to go to the store, she demands the first kind of good of their personal preferences: if it is present, she purchases it. If not, there is a certain probability of walk-out. If she stays, she asks for the second good, and again if it is not present.

The baker has a utility function increasing in profit and decreasing in variance of profit, scaled by a risk adversion factor.

The baker has the possibility to reserve a certain share of production as too good to go bags BEFORE he is able to observe the demand. These bags are sold for certain but at a lower price and most importantly they are not defined in the kind of goods, only on quantity. Once the baker has an amount of units equal reserved bags must close shop and can't sell any more for the day.

I want to fit the parameters of the model the baker's behaviour by using evolutionary algorithms.

I want the agent to have 2 parameters to be trained:

- $\mathbf{q}_t \in \mathbb{N}$: vector of quantity produced for each good at epoch t
- $b_t \in [0, 1]$: share of production reserved as too good

NOTE THAT EACH GOOD HAS THE SAME PRODUCTION COST AND SALE PRICE

The agent has only one primitive, unchangeable parameter:

- $\gamma \in [-1, 2]$: risk aversion factor

The environment has the following parameters: (i want to both be able to set them, along with the possibility to sample them from a distribution to create variability in the environment)

- $N \in \mathbb{R}$: number of consumers
- $F \in \mathbb{R}^{nxL}$: consumer preference matrix, where each row is a consumer and each column is a good, ordered by preference
- $r \in [0, 1]$: a customer will go away with this probability for each good not found in order of preference, until either finding a good or leaving the store
- $\chi \in \mathbb{R}^+$: production cost per unit of any good
- $\rho \in \mathbb{R}^+ > \chi$: sale price per each unit of any good

- $\tau \in \mathbb{R}^+ < \chi$: sale price per each unit of too good to go bag
- $L \in \mathbb{N}$: number of different goods
- $\alpha \in [0, 1]$: independent probability of each consumer to go to the store each day

Furthermore, the evolution hyperparameters are:

- $\mu \in [0, 1]$: mutation probability per parameter
- $\sigma_q \in \mathbb{R}^+$: standard deviation of the gaussian noise added to quantities when a mutation happens
- $\sigma_b \in \mathbb{R}^+$: standard deviation of the gaussian noise added to the share of production reserved as too good to go when a mutation happens
- $D \in \mathbb{N}$: amount of days in an epoch (note that each day has the same environment parameters, but different realisations of demand due to the probabilistic nature of consumer behaviour)
- $E \in \mathbb{N}$: amount of generations to run

The evolutiion logic is the following:

- at the end of each epoch, the agent's fitness is evaluated using the following formula as reward:
$$R_\gamma = \pi - \frac{\gamma}{2} \cdot \sigma^2$$
  where
  - $\pi = \rho \sum_{j=1}^{D} (s_j + D||\mathbf{q}||_1(\tau b - \chi))$ with s_j being the amount of products sold at day j
  - $\pi^* = E[\pi]$
  - $\sigma^2 = Var(\pi)$
- after all agents have been evaluated, the top 10% are selected as parents
- new agents are created by cloning parents and applying mutation with probability μ to
  - $\mathbf{q}_{t+1,i} = \mathbf{q}_{t,i} + \epsilon_q$ where $\epsilon_q \sim \mathcal{N}(0, \sigma_q)$
  - $b_{t+1} = b_t + \epsilon_b$ where $\epsilon_b \sim \mathcal{N}(0, \sigma_b)$