A Simple Model of Optimal Stopping

The Problem Setting

Consider a man who has to choose to each a cake on a single day between today and a fixed date in the future. Each day, his mood changes and this changes whether he would enjoy the cake or hate it and he know this fact. The cake must eventually be eated and cannot be split up. The man always knows what day it is and his mood for that day, so on that day can know exactly how much he wants to eat it. But a day earlier, he cannot say what his mood would be like on the next day–it is totally unpredictable. Each day, the man must decide whether to eat or not. Except on the last day, where he must eat since throwing the cake away is not an option. The man is completely patient, and would like to know when to eat the cake as a function of what day it is and what is mood is that day.

The Model

We represent this one-time cake eating problem as a optimal stopping problem in a Markov Decision Process (MDP):

- **State Space (S)**: $S = \{(t, \epsilon) \mid t = 1, 2, \ldots, T, \epsilon \in R\}$ represents the day at which the decision is made as well as the "mood" that the decision maker is in (on that day).
- **Action Space (A)**: $A = \{0, 1\}$, where:
  - $a_t = 1$ means consume the cake on day t.
  - $a_t = 0$ means delay consumption and move to day $t + 1$.
- **Reward Function ($R(s, a)$)**: The reward on day t is given by:

$$R(t, \epsilon_t, a_t) = \begin{cases} \epsilon_t & \text{if } a_t = 1 \\ 0 & \text{if } a_t = 0 \end{cases}$$

- **Transition Probability ($P(s'|s, a)$)**: The transitions are as follows:

$$P(t + 1, \epsilon_{t+1} \mid t, \epsilon_t, a_t) = P(t + 1 \mid t, \epsilon_t, a_t)P(\epsilon_{t+1} \mid t, \epsilon_t, a_t)$$

$$= P(t+1 \mid t, a_t) P(\epsilon_{t+1})$$

Since $\epsilon_{t+1} \perp \epsilon_t$ and only a_t affects the movement to next day. So we get,

$$P(t+1 \mid t, a_t = 0) = 1 \quad \text{and} \quad P(t+1 \mid t, a_t = 1) = 0$$

And given $\epsilon_{t+1} \sim F$

$$P(\epsilon_{t+1}) = f(\epsilon_{t+1})$$

- **Discount Factor (γ)**: $\gamma = 1$ (no discounting).

Value Function $V(t, \epsilon_t)$

We define $V(t, \epsilon_t)$ as the maximal utility at time t, given the cake has not been eaten at the start of day t but the mood has been realized. For the final day T, we have:

$$V(T, \epsilon_T) = \epsilon_T$$

Since on the last day eating must happen necessarily. For earlier days $t < T$, the value function is defined recursively:

$$V(t, \epsilon_t) = \max \{\epsilon_t, E_t V(t+1, \epsilon_{t+1})\}$$

where $E_t V(t+1, \epsilon_{t+1}) = \int_{\epsilon_{t+1}} V(t+1, \epsilon_{t+1}) f(\epsilon_{t+1}) d\epsilon_{t+1}$ is what he expects to get in the next day, not knowing the mood in the next day. It is hard to solve for this value function since it includes an expectation.

Decision Rule

Given V, the man consumes on day t as follows:

$$a_t = 1(\epsilon_t > E_t V(t+1, \epsilon_{t+1}))$$

This involves considering his mood today and comparing that with the expected maximum utility that he can extract tomorrow (after observing his mood tomorrow).

Expected Value Function $B(t)$

Let B_t be the expected maximal utility starting *after* day t is:

$$B(t) = E_t V(t+1, \epsilon_{t+1})$$

This function integrates out the mood, knowing its distribution, and is what the man expects to get from "tomorrow" on any day. Then the expected max utility yesterday about today is,

$$B(t-1) = E_{t-1} V(t, \epsilon_t)$$

$$= E_{t-1} \max\{\epsilon_t, B(t)\}$$

$$= \int_{\epsilon_t} \max\{\epsilon_t, B(t)\} f(\epsilon_t) d\epsilon_t$$

Note that, at t he will decide to eat and get ϵ_t if $\epsilon_t > B(t)$ and not eat and get $B(t)$ if $\epsilon_t < B(t)$,

$$= \int_{-\infty}^{B(t)} B(t) f(\epsilon_t) d\epsilon_t + \int_{B(t)}^{\infty} \epsilon_t f(\epsilon_t) d\epsilon_t$$

Remembering that, the expectation of a truncated variable is,

$$E[\epsilon_t|\epsilon_t > B(t)] = \frac{\int_{B(t)}^{\infty} \epsilon_t f(\epsilon)\, d\epsilon_t}{P(\epsilon_t > B(t))}$$

$$E[\epsilon_t|\epsilon_t > B(t)](1 - F(B(t)) = \int_{B(t)}^{\infty} \epsilon_t f(\epsilon_t)\, d\epsilon_t$$

So we get,

$$B(t-1) = B(t)F(B(t)) + E[\epsilon_t|\epsilon_t > B(t)](1 - F(B(t))$$

If we know F, this is a contraction map and we can use value function iteration from $t = T$ to get the $B(t)$ function.

Dynamic Optimization

The overall optimization problem is to maximize the expected total reward:

$$\max_{a_1,\dots,a_T} = E_0 \left[\sum_{t=1}^{T} r_t \right] = E_0 \left[\sum_{t=1}^{T} a_t \epsilon_t \right] \quad \text{s.t.} \quad \sum_{t=1}^{T} a_t = 1$$

where r_t depends on a_t. And the shocks are not known in advance. Trying to solve this directly would be difficult since we would need to maximize with respect to the joint distribution of the shocks.

Examples

The parameters for the simulation are set as follows:

- Time horizon: $T = 10$ days
- Number of simulations: 40 trajectories
- Distributions: Normal, Exponential, Uniform, Multimodal

The value function $B(t)$ and the stopping boundary $B(t+1)$ were computed using recursive value iteration. The stopping boundary represents the threshold at which the man decides to eat the cake based on his mood for each day.

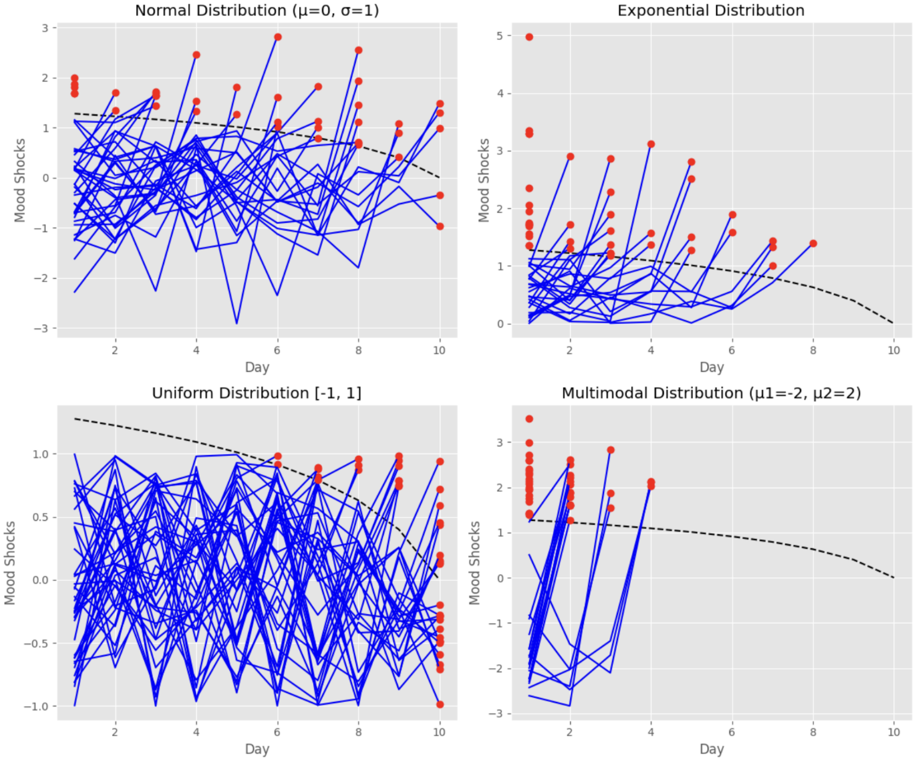

Figure 1: Caption