Problem 1:

- **Problem details**
  - Borrow from another bank $y \geq 0$ at constant daily interest rate R
    $\therefore$ Amount owed next day = Amount Borrowed $\cdot (1 + R)$
  - Invest a portion of bank's cash in risky asset
  - Both above have no transaction cost
  - If cash amount $c < C$, pay penalty $K \cdot \cot\left(\frac{\pi c}{2C}\right)$
    - Constant $C > 0$, $K > 0$
  - **Assumptions**:
    1. $c \leq K \cot\left(\frac{\pi c}{2C}\right)$ when $c \leq C$
    2. Neglect deposits
    3. First half of the day sees withdrawals
    4. If insufficient cash for withdrawal, withdrawal attempted again next day
    5. All quantities continuous
  - T-day planning horizon
  - **Goal**: Frame the MDP
  - Define the **state space** S by:

$$S = \{(c, d, i, w) : c \in \mathbb{R}_{\geq 0}, \; d \in \mathbb{R}_{\geq 0}, \; i \in \mathbb{R}_{\geq 0}, \; w \in \mathbb{R}_{\geq 0}\}$$

where c is the amount of money the bank has in cash that day, b is the amount of debt the bank is carrying that day, i is the amount of money the bank has in risky assets that day, and w is the amount of unsatisfied withdrawal requests which are carried over to the next day.

- Define the action space by:

$$A = \{ (d', i') : d' \in [-d, \infty), \quad i' \in [-i, c] \}$$

Where d' and i' are <u>changes</u> in debt and investment amounts to be applied the following day. This means $i_{t+1} = i_t + i'_t$ and $d_{t+1} = d_t + d'$.

- The reward function accounts for the amount of interest accumulated, fees paid, and investment returns from one day to the next. Assume that daily returns from the risky asset are normally distributed with mean $\mu = f(i) \in \mathbb{R}$ and standard deviation $\sigma = f(i) \in \mathbb{R}_{\geq 0}$. Assume there is a penalty $p \cdot w$ for asking a customer to return the next day to make withdrawal w, where $p \in \mathbb{R}$. The reward as we've defined it is independent of the action. However, the action taken affects the state at the following day, which in turn affects reward. The reward function $R(s,a)$ for specific state $s = (c, d, i, w)$ in S and any action $a \in A$ is given by:

$$R(s,a) = \begin{cases} N(\mu, \sigma^2) - pw - K \cot\left(\frac{\pi c}{2C}\right) & c \leq C \\ N(\mu, \sigma^2) - pw & c > C \end{cases}$$

- The money made on the risky investment $x \in \mathbb{R}$ during the day represents one of the stochastic component of the transition probability function. Deposits and withdrawals made to the bank represent other stochastic components of the transition probability function assume that deposits for the day $a \sim N(\mu_a, \sigma_a^2)$ and withdrawals for the day are given by $b \sim N(\mu_b, \sigma_b^2)$. Also let $f(x, a, b)$ be the joint probabilities of the money made from investments x, the deposit amount a, and the withdrawal amount b for the day (i.e., $f(x, a, b) = pdf(x, N(\mu, \sigma)) \cdot pdf(a, N(\mu_{(2)}, \sigma_{(2)}^2)) \cdot pdf(b, N(\mu_{(3)}, \sigma_{(3)}^2))$). Therefore, transitions $P(s, a, s')$ from state $s_{(1)} = (c_{(1)}, d_{(1)}, i_{(1)}, w_{(1)}) \in S$ and $s_{(2)} = (c_{(2)}, d_{(2)}, i_{(2)}, w_{(2)}) \in S$ given action $a = (d', i') \in A$ is given by:

$$P(s_{(t)}, a, s_{(t+1)}) = \begin{cases} f(x, a, b) & c_{(t)} = \max\left(0,\ c_{(t)} + a - b + x - W_{(t)} - p\,u_{(t)} + d^{\wedge} - i^{\wedge}\right); \quad \text{(what's owed previously)} \\ & d_{(t)} = d_{(t)} + d^{\wedge}; \quad i_{(t)} = i_{(t)} + i^{\wedge}; \quad W_{(t)} = W_{(t)} + \\ & \min\left(0,\ c_{(t)} + a - b + x - W_{(t)} - p\,u_{(t)} + d^{\wedge} - i^{\wedge}\right) + \max\left(0,\ W_{\cdot}\right. \\ & \left. - \left(c_{(t)} + a - b + x - p\,u_{(t)} + d^{\wedge} - i^{\wedge}\right)\right); \quad c_{(t)} > C \quad \text{(what's paid off; what could not be paid off)} \\ & \text{OR} \\ & c_{(t)} = \max\left(0,\ c_{(t)} + a - b + x - W_{(t)} - p\,u_{(t)} + d^{\wedge} - i^{\wedge} - K\,\mathrm{ceil}\left[\tfrac{\pi c}{2C}\right]\right); \\ & d_{(t)} = d_{(t)} + d^{\wedge}; \quad i_{(t)} = i_{(t)} + i^{\wedge}; \quad W_{(t)} = W_{(t)} + \\ & \min\left(0,\ c_{(t)} + a - b + x - W_{(t)} - p\,u_{(t)} + d^{\wedge} - i^{\wedge} - K\,\mathrm{ceil}\left[\tfrac{\pi c}{2C}\right]\right) + \max\left(0,\ W_{\cdot}\right. \\ & \left. - \left(c_{(t)} + a - b + x - p\,u_{(t)} + d^{\wedge} - i^{\wedge} - K\,\mathrm{ceil}\left[\tfrac{\pi c}{2C}\right]\right)\right); \quad c_{(t)} \le C \\ 0 & \text{otherwise} \end{cases}$$

- **How would I solve this MDP Control Problem?**

  - Our state and action space are continuous (non-finite, intractibly large). Therefore, I would solve this MDP control problem with Approximate DP. I would represent the policy and value function using function approximation (e.g. linear, neural network), which accept continuous states as inputs. I would then solve for the optimal policy with Approximate Policy Evaluation. In approximate policy iteration, the policy evaluation step of the algorithm is performed by iteratively applying the Bellman operator on the function approximator of the value function. In an iterative process, "source" and "next" states are sampled, reward is calculated, and the value function at the source state is approximated by the avg. return + discounted VF at next state (also obtained by the function approximator). During the policy improvement step of the algorithm, changes to the policy are made by updating parameters of the policy function approximator in the direction that selects actions corresponding to the path of greatest ascent of the value function approximation.