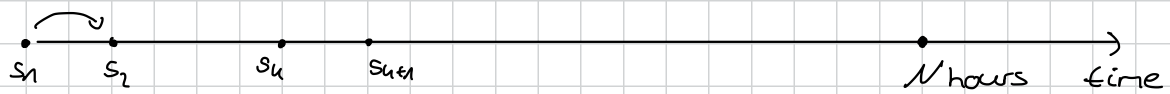
1.1 Proofreading

A manuscript must be submitted in N hours and has been typed with a known number of mistakes M. Mistakes may be found and corrected through a review. Each review takes one hour to complete and costs an amount $c_1 > 0$. On the k^{th} review, each undetected mistake is found independently with probability p_k. Each undetected mistake left in the manuscript when it is sent to the printer costs an amount $c_2 > 0$. The problem is to decide when to stop reviewing and send the manuscript to the printer.

Q1. *Model the problem as an MDP. Give a precise description of the MDP. Do not try to solve the MDP.*

N hours M number of mistakes   1 hour review, cost c_1

k reviews   probability p_k



The actions A would be either to do another review or to submit the manuscript

a_1: Do a new review

a_2: Stop and submit the manuscript

The states S can be defined as $S(k,m)$, where

- k: number of reviews done $[0,N]$
- m: number of mistakes left $[0,M]$

The probability to transition into the new state after each review is a Bernoulli distribution

$(m-m') = x$ (number of found mistakes)

$$P(t+1, m' \mid t, m) = \binom{m}{m-m'} p_k^{m-m'} (1-p_k)^{m'}$$

If the manuscript does not be submitted, the system transitions from (k,m) to $(k+1, m')$, with m' being the new number mistakes left, otherwise the process terminates

Reward:

$$R(s,a,s') = \begin{cases} -c_1 \\ -c_2 \cdot m \end{cases}$$

1.2 Selling your house

You wish to sell your house in Los Angeles. You try to sell it every spring, and start year 1. Due to climate changes, the risk of your house to burn is increasing summer after summer. In year t, the probability that your house disappears due to wildfires is b_t. Each spring you receive offers whose maximum is i.i.d. (across years) and with distribution described by $f(w)$, the probability that the best offer is w, for $w \in \{1, \ldots, W\}$. After selling your house, you place the money and enjoy an interest rate of $r\%$. Your objective is to maximize the average amount of money at the end of year $T > 1$.

Q2. *Model this problem as an MDP (describe the MDP in full detail).*

The actions are

$$A = (S, N) = \begin{cases} S & \text{sell the house} \\ W & \text{not sell the house (wait)} \end{cases}$$

The State S are defined as

$S(t, a)$

Year t (current year) $t \in [1, T]$

a: if the house is available for sale ($a = 1 \hat{=}$ available)
($a = 0$ not. avail.)

Transition probability

If actions S: money grows at interest rate r until year T

If action W: probability b_t that the house burns down -> State $(t+1, 0)$

probability $(1-b_t)$ house remains -> State $(t+1, 1)$

$$\Rightarrow P((t+1, 1) | (t, 1), W) = b_t$$

$$P((t+1), 0) | (t, 1), W) = (1 - b_t)$$

$$P((T, 0) | (t, 1), S) = 1$$

Rewards

If action S: money will accumulate over the years

$$\Rightarrow R((t, 1), S, (T, 0)) = w(1+r)^{T-t}$$

If action W: there won't be a direct reward, but a higher selling price might be achieved

$$R((t, 1), W, (t+1, 1)) = 0$$

Q3. *Establish that the optimal policy is threshold-based, i.e., you decide to accept the best offer made in year t if this offer exceeds a threshold.*

Q4. *Provide a general recursive formula satisfied by the thresholds.*

Q5. *Now assume that the best offer distribution is uniform over* $[0,1]$ *and that* $b_t = b$ *for all* t. *Answer the question c) again in this setting. When* T *is very large, what are the optimal decisions in the first years?*

Q3 For each year a threshold should be determined.
A good one is, when the value of selling and receiving $w(1+r)^{T-t}$ is equal to waiting for another year / facing risk b_t

$$V_{sell} = w \cdot (1+r)^{T-t}$$

$$V_{wait} = (1-b_t) \cdot \mathbb{E}\left[V_{t+1}(\text{available})\right]$$

To determine the threshold we can set $V_{sell} = V_{wait}$ with $w = \beta_t$

$$\Rightarrow \quad \beta_t (1+r)^{T-t} = (1-b_t)\, \mathbb{E}\left[V_{t+1}(\text{available})\right]$$

$$\Leftrightarrow \quad \beta_t = \frac{(1-b_t)\, \mathbb{E}\left[V_{t+1}(\text{available})\right]}{(1+r)^{T-t}}$$

Q4 The expeted value depends on the distribution, but be calculated:

$$(1) \quad \mathbb{E}\left[V_{t+1}(\text{available})\right] = \underbrace{\int_0^{\beta_{t+1}} V_{t+2}(\text{available})\, f(w)\, dw}_{\text{expexted value when } w < \beta_{t+1}} + \underbrace{\int_{\beta_{t+1}}^{W} w \cdot (1+r)^{T-t+1} f(w)\, dw}_{\text{expexted value, when } w > \beta_{t+1} \text{ is}}$$

Q5 Since the integral over $\int f(w) = 1$ we can simplify (1)

$$\Rightarrow \mathbb{E}\left[V_{t+1}(\text{available})\right] = \mathbb{E}\left[V_{t+2}(\text{availabe})\right] \frac{\beta_{t+1}}{W} + (1+r)^{T-(t+1)} \cdot \frac{1}{W}\left(\frac{W^2}{2} - \frac{\beta_{t+1}^2}{2}\right)$$

When T is very large, the optimal policy is to wait for higher offers than accepting lower ones $\rightarrow$ higher thresholds in the beginning (waiting for good offers)