Optimal Reporting Strategy in Insurance

July 29, 2025

Abstract

This paper presents an insurance model with N rate classes, in which the insured adopts a barrier strategy to decide whether they should report their losses in a period. In our model, the per-period loss follows a mixture distribution as seen in many real world examples such as vehicle insurance. The objective is to determine an optimal strategy that maximizes the insured's expected utility. We formulate the problem as a Markov Decision Process (MDP) and show that the known theorems of MDP hold in this setting, in particular we prove that the Bellman equation gives an optimal strategy. Further, we develop the methods of value iteration and policy iteration for examples where the Bellman equation does not suffice. Finally, we apply policy iteration to two numerical examples, then interpret and analyse the resulting strategies.

Keywords: underreporting losses, optimal insurance strategy, Markov decision process, stochastic dynamic programming

1 Introduction

In this paper we tackle a problem most of us are likely to face at some point, if I damage an insured belonging, should I report the loss or not? It is clear when damages are small we should avoid claiming because we suffer higher premiums in future periods. Similarly, it is clear that when the cost of the damages is large then we should report the losses. Many times it is unclear, the present cost may seem high but the short-term benefit of reporting the loss may not outweigh the long-term cost of entering a more expensive rate class.

We can model the insurance contract using a Markov decision process, a type of stochastic dynamic program. With this model, we aim to find an optimal strategy that tells the insured when they should report their per-period loss, this is done by maximising the value of the insurance contract, which is equivalent to minimising the burden of the costs incurred.

The theory of dynamic programming was originally developed by Richard E. Bellman in Bellman [4], this includes an introduction to Markov decision processes in chapter

eleven. Many results in the study of Markov decision processes are already well established. The main contribution of this paper is to contextualize these theorems in terms of our insurance model and show the results still hold. This will provide us with methods for finding explicit solutions, and for when this is not possible we will also develop numerical methods for finding approximate solutions. We then aim to apply these methods to some numerical examples, where we can see if the observed results match our intuition.

Research on the topic of finding an optimal reporting strategy is fairly limited. The pioneering work in this topic is Min [8], the author also uses a dynamic programming approach however each period the premium increases by an amount proportional to the size of the claim. The problem is also formulated differently in Min [8], in particular, they attempt to minimise the expected costs, whereas we minimise the expected utility.

A group of four authors have produced several papers (see, for example, Cao et al. [5, 6]; among others). In these papers they use a game theory approach, in particular the insured plays a game against themselves in each different rate class, each player tries to minimise their own incurred costs. Using this approach they compute a Nash equilibrium strategy. These papers only consider a specific number of rate classes (for example, in Cao et al. [5], they consider two and three rate classes).

The remainder of the paper is organised as follows. In section 2, we will introduce the key variables and relevant assumptions, and then delve into the formulation of the main problem. In section 3, we look at the existence of an optimal solution and some explicit results for finding the optimal solution. We then delve in to the methods of value iteration and policy iteration in section 4 to help us compute approximately optimal solutions when the explicit results of section 3 are not enough. We apply the numerical methods discussed in section 4 to two examples in section 5, here we note some key observations and see if they match our intuition of the problem. We conclude our results, discuss limitations of our model, and present future research ideas for this topic in section 6. Technical proofs which do not contribute concretely to understanding and developing the methods of this paper are placed in appendix A, and a derivation for a formula used in 5 is placed in appendix B.

2 Model

We will first introduce relevant variables necessary to study this problem, and mention any relevant assumptions about these variables. We can then use these variables to create functions which model the value of the contract in the current and future period. This allows us to formulate the main problem to be tackled in this paper.

2.1 Model variables

Consider an insurance model with N classes in discrete time such that if the insured claims in the t-th period then they move to the next class with an increased premium unless they are in the most expensive class already. Similarly, if the insured does not claim in

the t-th period then they move down a class for a reduced premium unless they are in the least expensive class already. Let the state space be $S = \{1, \ldots, N\}$, and the premiums charged to the insured in each period be $c_i > 0$ for $i \in S$ such that $0 < c_1 < c_2 < \ldots < c_n$.

Let ξ_t be independent and identically distributed random variables representing the random loss the insured incurs during the t-th period. We assume the loss is non-negative, $P(\xi_t \geq 0) = 1$, and there is a high probability the insured makes no loss, $P(\xi_t = 0) = p \approx 1$. Let F be the cumulative distribution function of ξ_t.

The insured employs a barrier strategy for claiming losses, if the loss is smaller or equal to the barrier then they choose not to claim, otherwise the loss is greater than the barrier and they will claim. Let the barrier for the i-th rate class be L_i for $i \in S$. Suppose the insured is in the i-th rate class during the t-th period, then they will choose not to report their losses (i.e. they won't claim) if $\xi_t \leq L_i$, and they will report losses if $\xi_t > L_i$. Let X_t represent the insured's rate class during the t-th period such that if $X_t = i$ for $i \in S$, then from the above,

$$X_{t+1} = \begin{cases} \max\{1, i-1\} & \text{if } \xi_t \leq L_i \text{ (insured does not claim)}, \\ \min\{N, i+1\} & \text{if } \xi_t > L_i \text{ (insured claims)}. \end{cases}$$

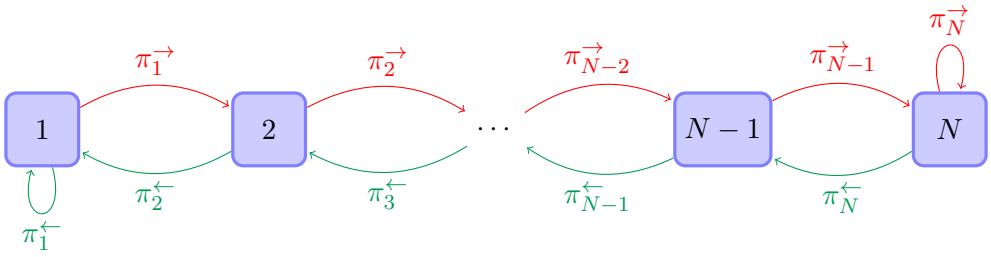
The insured follows the barrier strategy as described above, but we will model the problem in terms of the probabilities of transitioning between rate classes. To simplify notation, define $\pi_i^{\leftarrow} := P(\xi_0 \leq L_i) = F(L_i)$, the probability the insured moves to a lower rate class (does not claim). Similarly, let $\pi_i^{\rightarrow} := P(\xi_0 > L_i) = 1 - F(L_i)$, the probability the insured moves to a higher rate class (makes a claim). Clearly we have $\pi_i^{\leftarrow} + \pi_i^{\rightarrow} = 1$. The policy the insured follows is the corresponding transition matrix π. We assume the policy is stationary, i.e the transition probabilities are independent of the period. If π_{ij} is the transition probability from class i to j for $i, j \in S$, then we have

$$P(X_{t+1} = j \mid X_t = i) = \pi_{ij} = \begin{cases} \pi_i^{\leftarrow} & \text{if } j = \max\{1, i-1\}, \\ \pi_i^{\rightarrow} & \text{if } j = \min\{N, i+1\}, \\ 0 & \text{otherwise.} \end{cases}$$

So we can write the $N \times N$ policy matrix π in full,

$$\pi = \begin{bmatrix} \pi_1^{\leftarrow} & \pi_1^{\rightarrow} & 0 & \cdots & & \cdots & 0 \\ \pi_2^{\leftarrow} & 0 & \pi_2^{\rightarrow} & 0 & \cdots & \cdots & 0 \\ 0 & \pi_3^{\leftarrow} & 0 & \pi_3^{\rightarrow} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \pi_{N-2}^{\leftarrow} & 0 & \pi_{N-2}^{\rightarrow} & 0 \\ 0 & \cdots & \cdots & 0 & \pi_{N-1}^{\leftarrow} & 0 & \pi_{N-1}^{\rightarrow} \\ 0 & \cdots & & \cdots & 0 & \pi_N^{\leftarrow} & \pi_N^{\rightarrow} \end{bmatrix}.$$

It follows that $\{X_t\}_{t\geq 0}$ is a Markov process with state diagram:



2.2 Problem Formulation

Let U be the insured's utility function. As discussed in Cao et al. [5], in reference to the model presented in Arrow [3], it can be shown that the optimal indemnity a risk-neutral insured should pay is 0, i.e they shouldn't seek insurance. For this reason we will assume the insured is risk-averse. Suppose the insured is in rate class i during the t-th period: if they choose to make a claim, they are burdened with $U(-c_i)$; if they do not make a claim, then they are burdened with $U(-c_i - \xi_t)$. Let $g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow})$ be the immediate reward for being in class i, then

$$g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) := \underbrace{\pi_i^{\rightarrow} U(-c_i)}_{\text{Reward for claiming}} + \underbrace{\pi_i^{\leftarrow} E[U(-c_i - \xi_t) \mid \xi_t \leq L_i]}_{\text{Reward for not claiming}}. \tag{1}$$

We consider the value attained by following the policy π over an infinite horizon. Starting from rate class i, define

$$V^{\pi}(i) := E^{\pi}\left[\sum_{t=0}^{\infty} \delta^t g(X_t, \pi_{X_t}^{\rightarrow}, \pi_{X_t}^{\leftarrow}) \mid X_0 = i\right], \tag{2}$$

where $\delta \in (0,1)$ is a discount factor, and E^{π} denotes the expectation under the policy π.

Remark. If we know the policy π, then we can recover the barriers using the quantile function associated with ξ_t, $L_i = F^{-1}(\pi_i^{\leftarrow})$.

Problem 1. *For any class $i \in S$, we aim to determine an optimal policy π^* such that*

$$V^{\pi^*}(i) = \max_{\pi} V^{\pi}(i).$$

3 Results

This section explores the existence and uniqueness of the optimal policy. First we show that an optimal policy satisfying Problem 1 exists, then it satisfies the Bellman equation, see Bellman [4]. From this we show that the Bellman equation has a unique solution, and further that this solution is optimal.

3.1 Dynamic Programming Principles and Bellman Equation

For the discussion below, it is convenient to first define a vector for the value function in each state, and a vector for the immediate rewards, i.e., given a policy π we have

$$V^\pi = \begin{bmatrix} V^\pi(1) \\ \vdots \\ V^\pi(N) \end{bmatrix}, \quad g^\pi = \begin{bmatrix} g(1, \pi_1^{\rightarrow}, \pi_1^{\leftarrow}) \\ \vdots \\ g(N, \pi_N^{\rightarrow}, \pi_N^{\leftarrow}) \end{bmatrix}. \tag{3}$$

Theorem 1 (Dynamic Programming Principles)**.** *Given some policy π, we can compute the associated value function by solving the following system:*

$$V^\pi(i) = g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^\pi(j)\pi_{ij}, \quad \textit{for } i \in S.$$

Moreover, we can write

$$V^\pi = (I - \delta\pi)^{-1} g^\pi.$$

Proof. By definition, for any $i \in S$,

$$\begin{aligned} V^\pi(i) &= E^\pi\left[\sum_{t=0}^{\infty} \delta^t g(X_t, \pi_{X_t}^{\rightarrow}, \pi_{X_t}^{\leftarrow}) \mid X_0 = i\right] \\ &= g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta E^\pi\left[\sum_{t=1}^{\infty} \delta^{t-1} g(X_t, \pi_{X_t}^{\rightarrow}, \pi_{X_t}^{\leftarrow}) \mid X_0 = i\right] \\ &= g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} E^\pi\left[\sum_{t=1}^{\infty} \delta^{t-1} g(X_t, \pi_{X_t}^{\rightarrow}, \pi_{X_t}^{\leftarrow}) \mid X_1 = j, X_0 = i\right]\pi_{ij}, \end{aligned}$$

where the last equality holds by the law of total probability. Now, since $\{X_t\}_{t\geq 0}$ is a Markov process, we have the following equality in distributions:

$$\{X_t \mid X_1 = j, X_0 = i\} \stackrel{d}{=} \{X_t \mid X_1 = j\} \stackrel{d}{=} \{X_{t-1} \mid X_0 = j\}.$$

It follows that

$$\begin{aligned} V^\pi(i) &= g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} E^\pi\left[\sum_{t=1}^{\infty} \delta^{t-1} g(X_t, \pi_{X_t}^{\rightarrow}, \pi_{X_t}^{\leftarrow}) \mid X_1 = j\right]\pi_{ij} \\ &= g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} E^\pi\left[\sum_{t=1}^{\infty} \delta^{t-1} g(X_{t-1}, \pi_{X_{t-1}}^{\rightarrow}, \pi_{X_{t-1}}^{\leftarrow}) \mid X_0 = j\right]\pi_{ij} \\ &= g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} E^\pi\left[\sum_{t=0}^{\infty} \delta^{t} g(X_t, \pi_{X_t}^{\rightarrow}, \pi_{X_t}^{\leftarrow}) \mid X_0 = j\right]\pi_{ij} \\ &= g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^\pi(j)\pi_{ij} \qquad \text{by definition.} \end{aligned}$$

The final part of the theorem follows by expressing the above system in terms of V^π and g^π,

$$V^\pi = g^\pi + \delta\pi V^\pi,$$

and by rearranging we get

$$V^\pi = (I - \delta\pi)^{-1} g^\pi.$$

□

If the optimal policy satisfying Problem 1 exists, then we can show the optimal policy satisfies a similar expression to Theorem 1.

Theorem 2 (Bellman)**.** *The optimal policy, π^*, satisfies*

$$V^{\pi^*}(i) = \max_\pi \{g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi^*}(j)\pi_{ij}\}, \quad \textit{for } i \in S.$$

Proof. By definition of the optimal policy, π^*, we have $V^\pi(i) \leq V^{\pi^*}(i)$, for any $i \in S$ and any policy π. Then from dynamic programming principles (Theorem 1), we have

$$V^\pi(i) \leq g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi^*}(j)\pi_{ij} \leq \max_\pi \{g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi^*}(j)\pi_{ij}\}, \quad (4)$$

and notably this inequality holds for $\pi = \pi^*$.

To show the reverse inequality, we define a policy $\tilde{\pi} := (\pi', \pi^*)$, where an arbitrary policy π' is used in the first time interval, and then the optimal policy π^* is followed for any subsequent intervals. By definition of the optimal policy, we have for any $i \in S$, $V^{\pi^*}(i) \geq V^{\tilde{\pi}}(i)$. It follows that

$$V^{\pi^*}(i) \geq V^{\tilde{\pi}}(i) = g(i, \pi_i'^{\rightarrow}, \pi'^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi^*}(j)\pi'_{ij}. \quad (5)$$

Since π' is arbitrary, then we have the reverse inequality. It follows from 4 and 5 that

$$V^{\pi^*}(i) = \max_\pi \{g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi^*}(j)\pi_{ij}\}.$$

□

3.2 Existence, Uniqueness, and Optimality of Solution

We now look closer at the Bellman equation. In particular, we show it has a unique solution, and that this solution is optimal for Problem 1. This provides us with a clear method to finding the optimal policy, π^*, as we can look for a policy such that the Bellman equation is satisfied.

To show the Bellman equation has a unique solution, we first define the vector map $T : \mathbb{R}^N \to \mathbb{R}^N$, where

$$(TW)_i = \max_{\pi}\{g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} W_j \pi_{ij}\}. \tag{6}$$

We note the definition of V^π from 3. Then it follows from Theorem 2, that if an optimal policy exists, then it is a fixed point of the vector map T, i.e.

$$TV^{\pi^*} = V^{\pi^*}.$$

At this point it is not clear if the map has a unique fixed point, or a such a point at all. Moreover, if we are able to find a fixed point, is the resulting policy optimal?

Theorem 3. *Under the maximum norm, $\|W\| = \max_{1\leq i\leq n} |W_i|$, the map T is a contraction mapping, i.e. for $W, W' \in \mathbb{R}^N$, then*

$$\left\|TW - TW'\right\| \leq \delta \left\|W - W'\right\|$$

for $\delta \in (0, 1)$.

Proof. See Appendix A □

Theorem 4. *A contraction mapping $T : \mathbb{R}^N \to \mathbb{R}^N$ has a unique fixed point.*

Proof. Define the sequence $W^{(0)} \in \mathbb{R}^N$, and $W^{(n+1)} = TW^{(n)}$. We will show this sequence is a Cauchy sequence, and the limit of the sequence is a fixed point.

First, inductively we have,

$$\left\|W^{(n+1)} - W^{(n)}\right\| = \left\|TW^{(n)} - TW^{(n-1)}\right\| \leq \delta \left\|W^{(n)} - W^{(n-1)}\right\| \leq \ldots \leq \delta^n \left\|W^{(1)} - W^{(0)}\right\|.$$

Now consider $m, n \in \mathbb{N}$ such that $m > n$, it follows from the above that

$$\begin{aligned}
\left\|W^{(m)} - W^{(n)}\right\| &= \left\|(W^{(m)} - W^{(m-1)}) + (W^{(m-1)} - W^{(m-2)}) + \ldots + (W^{(n+1)} - W^{(n)})\right\| \\
&\leq \left\|W^{(m)} - W^{(m-1)}\right\| + \left\|W^{(m-1)} - W^{(m-2)}\right\| + \ldots + \left\|W^{(n+1)} - W^{(n)}\right\| \\
&\leq \delta^{m-1} \left\|W^{(1)} - W^{(0)}\right\| + \delta^{m-2} \left\|W^{(1)} - W^{(0)}\right\| + \ldots + \delta^n \left\|W^{(1)} - W^{(0)}\right\| \\
&= \delta^n \left\|W^{(1)} - W^{(0)}\right\| \sum_{i=0}^{m-n-1} \delta^i \\
&\leq \delta^n \left\|W^{(1)} - W^{(0)}\right\| \sum_{i=0}^{\infty} \delta^i \qquad \text{since } \delta > 0, \\
&= \delta^n \left\|W^{(1)} - W^{(0)}\right\| \frac{1}{1-\delta}.
\end{aligned}$$

Now, let $\epsilon > 0$, select $M \in \mathbb{N}$ large enough such that

$$\delta^M < \frac{\epsilon(1-\delta)}{\left\|W^{(1)} - W^{(0)}\right\|},$$

then, for all $m > n > M$

$$\left\|W^{(m)} - W^{(n)}\right\| < \epsilon.$$

So the sequence $(W^{(n)})_{n\in\mathbb{N}}$ is a Cauchy sequence and has a limit $W^* \in \mathbb{R}^N$. Moreover, by the continuity of contraction mappings (see Lemma 2), we have

$$W^* = \lim_{n\to\infty} W^{(n)} = \lim_{n\to\infty} TW^{(n-1)} = T \lim_{n\to\infty} W^{(n-1)} = TW^*, \tag{7}$$

so the limit point W^* is a fixed point of T.

Now we show the uniqueness of the fixed point. By way of contradiction, assume T has two fixed points W_1 and W_2 such that $W_1 \neq W_2$. Since T is a contraction mapping we have

$$\|TW_2 - TW_1\| \leq \delta \|W_2 - W_1\|.$$

But, since $\delta \in (0,1)$, and W_1 and W_2 are fixed points, we have

$$\|TW_2 - TW_1\| = \|W_2 - W_1\| > \delta \|W_2 - W_1\|,$$

a contradiction. So it follows that $W^* = \lim_{n\to\infty} W^{(n)}$ is the unique fixed point of T. □

So we have shown that the Bellman equation always has a unique fixed point, we now show that the policy corresponding to the fixed point is an optimal policy.

Theorem 5. *If V^{π^*} is a fixed point of the map T (as defined in 6), then π^* is an optimal policy, i.e.*

$$V^{\pi^*}(i) \geq V^{\pi'}(i),$$

where $i \in S$, and π' is any arbitrary policy.

Proof. First we note that the policy π^* maximises the expression

$$\max_{\pi}\{g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi^*}(j)\pi_{ij}\}, \tag{8}$$

for $i \in S$, by definition.

Now, define the policy $\tilde{\pi}^{(n)}$, where the policy π' is used for first n time intervals, and then π^* is used for the remaining intervals, i.e.

$$\tilde{\pi}^{(n)} = (\overbrace{\pi', \ldots, \pi'}^{n \text{ times}}, \pi^*, \ldots)$$

Note that $\tilde{\pi}^{(0)} = \pi^*$. Using dynamic programming principles (Theorem 1), we have

$$V^{\tilde{\pi}^{(1)}}(i) = g(i, \pi_i'^{\rightarrow}, \pi_i'^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi^*}(j)\pi'_{ij},$$

recall that π^* is the maximiser of this expression, so it follows that

$$V^{\tilde{\pi}^{(1)}}(i) \leq g(i, \pi^{*\rightarrow}, \pi^{*\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi^*}(j)\pi^*_{ij} = V^{\tilde{\pi}^{(0)}}(i).$$

Suppose $V^{\tilde{\pi}^{(n)}}(i) \leq V^{\tilde{\pi}^{(n-1)}}(i)$, we also note that following the policy $\tilde{\pi}^{(n+1)}$ is the same as following π' in the first interval then following $\tilde{\pi}^{(n)}$ for the remaining intervals. Using this fact, we have

$$V^{\tilde{\pi}^{(n+1)}}(i) = g(i, \pi_i'^{\rightarrow}, \pi_i'^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\tilde{\pi}^{(n)}}(j)\pi'_{ij} \leq g(i, \pi_i'^{\rightarrow}, \pi_i'^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\tilde{\pi}^{(n-1)}}(j)\pi'_{ij} = V^{\tilde{\pi}^{(n)}}(i).$$

So,

$$V^{\tilde{\pi}^{(n+1)}}(i) \leq V^{\tilde{\pi}^{(n)}}(i) \leq \ldots \leq V^{\tilde{\pi}^{(0)}}(i) = V^{\pi^*}(i),$$

by reverse induction. It follows that

$$\lim_{n\to\infty} V^{\tilde{\pi}^{(n)}}(i) = V^{\pi'}(i) \leq V^{\pi^*}(i)$$

for $i \in S$ and any policy π', i.e. π^* is an optimal policy. □

The above results give us a method of finding an optimal policy π^* by finding a policy that satisfies the Bellman equation in Theorem 2.

4 Numerical Methods

In this section we explore numerical methods for computing an optimal policy. The theoretical results discussed in the previous section are not always useful in practice since many times it is not possible to find the optimal policy π^* explicitly. We explore the methods of value iteration and policy iteration, each method is shown to approach the optimal policy and so can be applied to approximate the solution. Pseudocode is provided, and we will make comparisons between the two methods.

4.1 Value Iteration

The value iteration method follows from the proof of Theorem 4.

Corollary 1 (Value iteration)**.** *Start from an arbitrary value* $V^{(0)} = [V_1^{(0)}\ V_2^{(0)}\ \cdots\ V_N^{(0)}]$*, and define the sequence* $V^{(n+1)} = TV^{(n)}$*, as in Theorem 4. Then, from (7), the optimal value is the limit of the sequence,* $V^{\pi^*} = \lim_{n\to\infty} V^{(n)}$*.*

We can find an approximately optimal value numerically by performing a finite number of iterations of the value iteration method. Recall that if we have an optimal value, V^*, then by Theorem 2, we can find the optimal policy by finding the maximiser of the Bellman equation,

$$\pi^*(i) = \arg\max_{\pi}\{g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^*(j)\pi_{ij}\}.$$

In the below pseudocode, M denotes the number of iterations of value iteration will be performed.

Algorithm 4.1 Value iteration

```
procedure VALUE-ITERATION(M)
    V = arbitrary value vector of size N
    W = vector of size N
    for n = 0 to M do
        W[i] = max_π{g(i, π_i^→, π_i^←) + δ Σ_{j=1}^{N} V[j]π_ij}
        V = W
    return V
```

With the final value of V in the algorithm, we can use the bellman equation as described above to find an approximately optimal policy.

Now we look at how good of an approximation value iteration can achieve. Suppose we want to be within $\epsilon > 0$ of the optimal value in each component, i.e.

$$\|V^m - V^*\| = \max_{i \in S} |V^m(i) - V^*(i)| \leq \epsilon,$$

where m is the number of iterations performed, and V^* is the optimal value achieved. As discussed in Adams [1], such an approximation can be achieved in K iterations, where

$$K = \left\lceil \frac{\log\left(\frac{2G_{\max}}{\epsilon(1-\delta)}\right)}{\log\left(\frac{1}{\delta}\right)} \right\rceil,$$

and $G_{\max}$ is the maximal value of the immediate reward function g for any state $i \in S$ and any policy, i.e.

$$G_{\max} = \max_{\pi, i \in S} g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}).$$

4.2 Policy Iteration

Suppose we start with an arbitrary policy $\pi^{(0)} \in \mathbb{R}^{N \times N}$, and we compute the associated value,

$$V^{(0)} = V^{\pi^{(0)}} = (I - \delta\pi^{(0)})^{-1} g^{\pi^{(0)}}.$$

We apply the Bellman operator to obtain each new policy in the sequence. So, for $n \geq 1$ and $i \in S$,

$$\pi^{(n+1)}(i) = \arg\max_{\pi(i)} \{g(i, \pi_i^{\leftarrow}, \pi_i^{\rightarrow}) + \delta \sum_{y=1}^{N} V^{\pi^{(n)}}(j)\pi_{ij}\},$$

where $\pi(i)$ denotes the i-th row of a policy matrix π. Once again, we compute the associated value using the same method as before,

$$V^{(n+1)} = (I - \delta\pi^{(n+1)})^{-1} g^{\pi^{(n+1)}}.$$

We can show that with each iteration of this method the policy achieves at least as good of a value as the previous policy and it follows that the policies in this sequence approach the optimal policy.

Theorem 6 (Policy Improvement)**.** *Suppose we have an arbitrary policy π', and let*

$$\pi''(i) = \arg\max_{\pi(i)} \{g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi'}(j)\pi_{ij}\} \quad \textit{for } i \in S. \tag{9}$$

Then π'' is an improved policy over π', i.e.

$$V^{\pi''}(i) \geq V^{\pi'}(i),$$

for $i \in S$.

Proof. Suppose we have an arbitrary policy π', and the policy π'' as defined above in 9. Now define the policy $\tilde{\pi}_n$ where we use π'' for the first n time intervals, and then use the policy π' for all remaining intervals, i.e.

$$\tilde{\pi}_n = (\overbrace{\pi'', \pi'', \ldots, \pi''}^{n \text{ times}}, \pi', \ldots).$$

We first show $V^{\tilde{\pi}_1}(i) \geq V^{\tilde{\pi}_0}(i)$ for all states $i \in S$. Using dynamic programming principles (see Theorem 1), we have

$$V^{\tilde{\pi}_1}(i) = g(i, \pi''^{\rightarrow}, \pi''^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi'}(j)\pi''_{ij},$$

and by the definition of π'', given in 9, π'' is the maximiser of the above expression. It follows that

$$V^{\tilde{\pi}_1}(i) \geq g(i, \pi'^{\rightarrow}, \pi'^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\pi'}(j)\pi'_{ij} = V^{\pi'}(i),$$

where we get the last equality by dynamic programming principles. Finally we note that the policy $\tilde{\pi}_0 \equiv \pi'$, so we have $V^{\tilde{\pi}_1}(i) \geq V^{\tilde{\pi}_0}(i) = V^{\pi'}(i)$ as required.

Now, suppose that $V^{\tilde{\pi}_n}(i) \geq V^{\tilde{\pi}_{n-1}}(i)$ for $i \in S$ and consider $V^{\tilde{\pi}_{n+1}}(i)$. We note that following the policy $\tilde{\pi}_{n+1}$ is equivalent to following π'' for the first time interval, and then following $\tilde{\pi}_n$ for the remaining intervals. Using this and the inductive hypothesis, we have

$$V^{\tilde{\pi}_{n+1}}(i) = g(i, \pi''^{\rightarrow}, \pi''^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\tilde{\pi}_n}(j)\pi''_{ij} \geq g(i, \pi''^{\rightarrow}, \pi''^{\leftarrow}) + \delta \sum_{j=1}^{N} V^{\tilde{\pi}_{n-1}}(j)\pi''_{ij} = V^{\tilde{\pi}_n}(i),$$

where we obtain the last equality using dynamic programming principles.

It follows from induction that for any $n \geq 0$, $V^{\tilde{\pi}_n}(i) \geq V^{\pi'}(i)$, and moreover as $n \to \infty$ we have

$$\lim_{n \to \infty} V^{\tilde{\pi}_n}(i) = V^{\pi''}(i) \geq V^{\pi'}(i).$$

□

Since we know that each application of policy iteration we achieve at least the same value of $V^{\pi}(i)$, it follows that the sequence converges to the optimal policy.

As discussed in Adams [1] and Miller [7], the cost of each iteration of policy iteration is typically larger than value iteration. However, it is empirically observed that policy iteration usually achieves a good approximation in fewer steps than value iteration, and moreover it is typically more efficient overall than value iteration.

Similar to value iteration, in the below pseudocode, M denotes the number of iterations of policy iteration to be performed.

Algorithm 4.2 Policy Iteration

```
1: procedure POLICY-ITERATION(M)
2:     π' = arbitrary initial policy (N × N matrix)
3:     for n = 0 to M do
4:         V = (I − δπ)^{-1} g^π
5:         π' = maximiser of {g(i, π_i^→, π_i^←) + δ Σ_{j=1}^{N} V[j] π_ij} for each i ∈ S
6:     return π'
```

We note that when implementing this algorithm, the initialisation of the policy matrix π can actually be done by creating a vector of size N which holds each value of $\pi_i^{\leftarrow}$ for each $i \in S$. Then we note that $\pi_i^{\rightarrow} = 1 - \pi_i^{\leftarrow}$, and the rest of the values in the policy matrix are 0, i.e. with this vector of size N we can recover the whole policy matrix. Some care should be taken when considering $(I - \delta\pi)^{-1}$.

5 Examples

We close our analysis by looking at an informative example of how to apply the policy iteration method as seen in the previous section. We will set up a framework to analyse this example for any number of rate classes, and any parameters, then we will determine solutions to a specified version of the problem and see if this fits intuition.

5.1 Numerically solving an N Rate Class Problem

Suppose the insured has a risk aversion paramter $\gamma > 0$, and they have an exponential utility function, i.e.

$$U(x) = -e^{-\gamma x}$$

We assume the loss variable ξ_t follows a mixture of a Gamma(α, λ) distribution for positive values and a point mass at 0, i.e. $P(\xi_t = 0) = p \in (0, 1)$. Further recall that each ξ_t is independent and identically distributed. Set $\tilde{\xi}_t = \{\xi_t \mid \xi_t > 0\} \sim \text{Gamma}(\alpha, \lambda)$, then $\tilde{\xi}_t$ has density

$$f_{\tilde{\xi}}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad \text{where } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt.$$

It follows that ξ_t has cumulative distribution and quantile functions:

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ p & \text{if } x = 0, \\ p + (1-p)F_{\tilde{\xi}}(x) & \text{if } x > 0, \end{cases} \quad \text{and} \quad F^{-1}(x) = \begin{cases} 0 & \text{if } x \le p, \\ F_{\tilde{\xi}}^{-1}\left(\frac{x-p}{1-p}\right) & \text{if } x > p, \end{cases}$$

where F_ξ and F_ξ^{-1} are the cumulative distribution and quantile functions of the Gamma distribution respectively.

Assume for simplicity $\gamma \neq \lambda$. We show in Appendix B, that given a policy π, the immediate rewards can be written as

$$g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) = -e^{\gamma c_i}\left(\pi_i^{\rightarrow} + p + \frac{1-p}{\Gamma(\alpha)}\left(\frac{\lambda}{\lambda-\gamma}\right)^\alpha \tilde{\gamma}\big(\alpha, (\lambda-\gamma)F^{-1}(\pi_i^{\leftarrow})\big)\right), \tag{10}$$

where $\tilde{\gamma}$ is the lower incomplete gamma function, i.e.

$$\tilde{\gamma}(s, x) = \int_0^x t^{s-1} e^{-t}\, dt.$$

Using this, we can now apply policy iteration to compute a policy for given values of the various parameters.

In our examples we will assume the premiums follow the expected value principle, a method discussed briefly in Ali [2]. Suppose we have a vector of risk loadings Θ, where for each rate class $i \in S$ we have $\theta_i \in \Theta$ such that

$$\theta_1 < \theta_2 < \ldots < \theta_N.$$

Moreover, the premium in each rate class is given by

$$c_i = (1 + \theta_i)E[\xi_t].$$

So in the case where the loss follows a mixed gamma distribution, we have

$$c_i = \frac{\alpha}{\lambda}(1+\theta_i)(1-p). \tag{11}$$

Example 1. For our first example we use the following parameters:

Parameter	**Value**
Number of rate classes (N)	6
Risk Aversion (γ)	0.4
Shape (α)	10
Rate (λ)	0.5
Discount factor (δ)	0.8
Probability of no loss (p)	0.7
Risk loadings (Θ)	$[0, 0.2, 0.5, 0.9, 1.4, 2]$

Then, using policy iteration, we obtain the following barrier strategy:

Rate class	**Value**
L_1	3.03
L_2	4.97
L_3	5.70
L_4	5.65
L_5	4.66
L_6	1.50

So to follow this strategy suppose we are in rate class four, then we will not claim if we lose less than 5.65 in a period, and we will make a claim if our loss exceeds 5.65.

We notice a non-monotonic nature of the barrier strategy, and in particular the barrier in rate class six is significantly smaller than rate class five. This can be explained as we realise that claiming in a lower rate class (classes 1-5) means you will have to payer higher premiums in the next period but when you are already in the highest rate class, making a claim only means you suffer the same burden in the next period, hence the disincentive of reporting the losses is not as strong.

Example 2 (Variable discount factor)**.** In this example we will look at a five rate class system, the value of the parameters are:

Once again, we apply policy iteration to retrieve the barrier strategies except now with a variable discount factor δ.

We see the resulting barrier strategies in figure 1. First we note the non-monotonicity is consistent with the observations from example 1. A positive trend is seen between the discount factor and the optimal barrier strategies. Recall, the discount factor adds a weight to each term so future periods do not have as great of an effect as the current period, see 2. Intuitively, this means when the discount factor is lower, then future terms

Parameter	Value
Number of rate classes (N)	5
Risk Aversion (γ)	0.5
Shape (α)	1
Rate (λ)	1
Discount factor (δ)	Variable in $(0, 1)$
Probability of no loss (p)	0.7
Risk loadings (Θ)	$[0.5, 1.5, 3, 5, 7.5]$

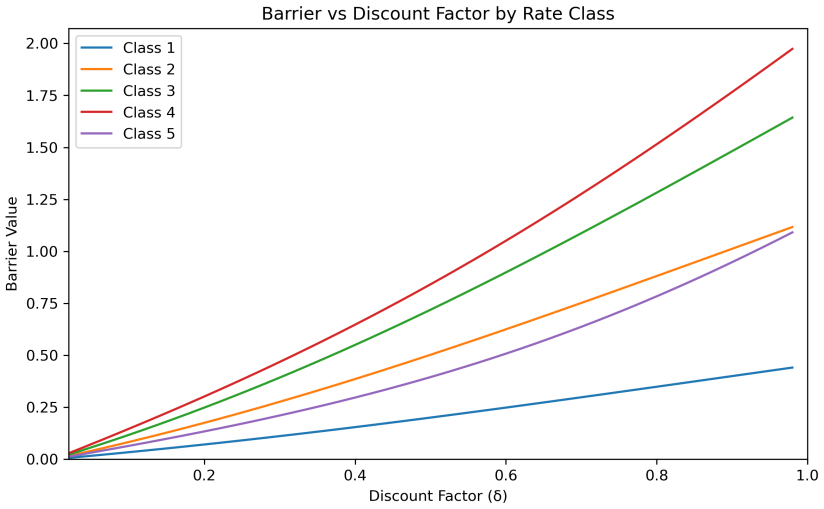


Figure 1: Impact of the discount factor δ on the optimal strategy for each rate class

have little impact on the value of V^π and thus we are more likely to claim in the present to save on the losses rather than worry about future costs.

6 Conclusion and Further Research

In the paper we studied the optimal reporting strategy for a simple N-class insurance model, where the insured moves to a more expensive rate class if they choose to claim in a period, otherwise they move to a lower rate class. The primary contribution of this paper was using the theory of Markov Decision Processes to model this insurance contract. We showed an optimal policy satisfies the Bellman equation, this provides us with a method to find an explicit solution. However, this method is typically difficult or impossible to recover an optimal solution explicitly, so we extend to two numerical methods: value iteration; and policy iteration. Applying the numerical methods, we were able to find an optimal policy for two examples where we linked the results to intuition.

The model provided in this paper makes some simplifying assumption that future research should look at more closely. There are other variables that were not included but are common in insurance contracts, in particular the excess cost (also known as a deductible) for claiming was not accounted for.

We further simplified the model by assuming claims only change the insured's rate class in single steps, but in a more general setting larger jumps should be included. Consider vehicle insurance, typically a more costly crash will result in the insured moving up multiple rate classes. Implementing this idea would involve creating a set of values that indicate how costly the damages would need to be to move up a certain number of rate classes. Once this is modeled, most of the theory discussed in this paper would still be applicable but many of the formulas would be more complicated, such as in the Bellman Equation (see Theorem 2) where there would be many terms in the summation, whereas in the scenario discussed in this paper there is only ever two terms in this sum.

A Proofs for section 3.2

Lemma 1. *Let* $W = [W_1\ W_2\ \cdots\ W_N]$ *and* $W' = [W'_1\ W'_2\ \cdots\ W'_N]$. *For* $i \in S$, *define the following functions for simplicity:*

$$A(i,\pi) = g(i,\pi_i^{\rightarrow},\pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} W_j \pi_{ij},$$
$$B(i,\pi) = g(i,\pi_i^{\rightarrow},\pi_i^{\leftarrow}) + \delta \sum_{j=1}^{N} W'_j \pi_{ij}.$$

Then we have

$$|\max_{\pi} A(i,\pi) - \max_{\pi} B(i,\pi)| \leq \max_{\pi} |A(i,\pi) - B(i,\pi)| = \delta \max_{\pi} \left| \sum_{j=1}^{N} (W_j - W'_j)\pi_{ij} \right|. \quad (12)$$

Proof. Suppose without loss of generality,

$$\max_{\pi} A(i,\pi) \geq \max_{\pi} B(i,\pi).$$

It follows that

$$\begin{aligned}
|\max_{\pi} A(i,\pi) - \max_{\pi} B(i,\pi)| &= \max_{\pi} A(i,\pi) - \max_{\pi} B(i,\pi) \\
&= \max_{\pi}(A(i,\pi) + B(i,\pi) - B(i,\pi)) - \max_{\pi} B(i,\pi) \\
&\leq \max_{\pi}(A(i,\pi) - B(i,\pi)) + \max_{\pi} B(i,\pi) - \max_{\pi} B(i,\pi) \\
&= \max_{\pi}(A(i,\pi) - B(i,\pi)) \\
&\leq \max_{\pi} |A(i,\pi) - B(i,\pi)|.
\end{aligned}$$

Trivially we have

$$A(i,\pi) - B(i,\pi) = \delta \sum_{j=1}^{N} (W_j - W'_j)\pi_{ij},$$

and so the equality in 12 follows immediately since $\delta > 0$ is constant. □

Proof of Theorem 3. Let $W = [W_1\ W_2\ \cdots\ W_N]$ and $W' = [W'_1\ W'_2\ \cdots\ W'_N]$, for $i \in S$. From Lemma 1, we have

$$\begin{aligned}
\left|(TW)_i - (TW')_i\right| &\leq \delta \max_{\pi} \left| \sum_{j=1}^{N} (W_j - W'_j)\pi_{ij} \right| \\
&\leq \delta \max_{\pi} \left\{ \sum_{j=1}^{N} \left|W_j - W'_j\right| \pi_{ij} \right\} \\
&\leq \delta \max_{\pi} \left\{ \sum_{j=1}^{N} \left\|W - W'\right\| \pi_{ij} \right\} \qquad \text{by definition of max-norm,} \\
&= \delta \left\|W - W'\right\| \max_{\pi} \left\{ \sum_{j=1}^{N} \pi_{ij} \right\} \\
&= \delta \left\|W - W'\right\|.
\end{aligned}$$

It follows immediately that

$$\left\|TW - TW'\right\| = \max_{1 \leq i \leq n} |TW_i - TW'_i| \leq \delta \left\|W - W'\right\|.$$

□

Lemma 2 (Continuity of Contraction Mappings)**.** *A contraction mapping $T : \mathbb{R}^N \to \mathbb{R}^N$ is continuous.*

Proof. For an $\epsilon > 0$, if $\|W - W'\| < \frac{\epsilon}{\delta}$, then

$$\left\|TW - TW'\right\| \leq \delta \left\|W - W'\right\| < \delta \cdot \frac{\epsilon}{\delta} < \epsilon.$$

So the mapping T is uniformly continuous, and thus continuous. □

B Computations in Examples

Immediate Rewards in Example 5.1. Recall the definition of the immediate reward function from 1:

$$g(i, \pi_i^{\rightarrow}, \pi_i^{\leftarrow}) := \pi_i^{\rightarrow} U(-c_i) + \pi_i^{\leftarrow} E[U(-c_i - \xi_t) \mid \xi_t \leq L_i],$$

where $i \in S$. For the second term, we first note the non-negativity of ξ_t, and we have

$$\pi_i^{\leftarrow} E[U(-c_i - \xi_t) \mid \xi_t \leq L_i] = -e^{c_i} E[e^{\gamma \xi_t} \cdot 1_{\{0 \leq \xi_t \leq L_i\}}].$$

And now we focus on the $E[\cdot]$ factor,

$$\begin{aligned} E[e^{\gamma\xi_t} \cdot 1_{\{0\le\xi_t\le L_i\}}] &= p + E[e^{\gamma\xi_t} \cdot 1_{\{0<\xi_t\le L_i\}}] \\ &= p + (1-p)\int_0^{L_i} e^{\gamma t}\frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}\,dt \\ &= p + \frac{(1-p)\lambda^\alpha}{\Gamma(\alpha)}\int_0^{(\lambda-\gamma)L_i} e^{-z}z^{\alpha-1}\left(\frac{1}{\lambda-\gamma}\right)^\alpha dz, \end{aligned}$$

where we obtain the last line by substituting $z = (\lambda-\gamma)t$. And so,

$$\begin{aligned} \pi_i^{\leftarrow} E[e^{\gamma\xi_t} \mid \xi_t \le L_i] &= \frac{(1-p)}{\Gamma(\alpha)}\left(\frac{\lambda}{\lambda-\gamma}\right)^\alpha \int_0^{(\lambda-\gamma)L_i} e^{-z}z^{\alpha-1}\,dz \\ &= \frac{(1-p)}{\Gamma(\alpha)}\left(\frac{\lambda}{\lambda-\gamma}\right)^\alpha \tilde{\gamma}\big(\alpha, (\lambda-\gamma)F^{-1}(L_i)\big). \end{aligned}$$

The final result follows immediately. □

References

[1] Ryan Adams. Planning via policy iteration, 2018. URL `https://www.cs.princeton.edu/courses/archive/spring19/cos324/files/policy-iteration.pdf`. Lecture notes.

[2] Mohammad Jamsher Ali. Wang's premium principle: overview and comparison with classical principles. Master's thesis, University of Tartu, Institute of Public Health, Tartu, Estonia, 2016. URL `https://core.ac.uk/download/pdf/79115112.pdf`.

[3] Kenneth J. Arrow. Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(5):941–973, 1963. URL `https://assets.aeaweb.org/asset-server/files/9442.pdf`.

[4] Richard Ernest Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1st edition, 1957. ISBN 069107951X.

[5] Jingyi Cao, Dongchen Li, Virginia R. Young, and Bin Zou. Equilibrium reporting strategy: Two rate classes and full insurance, 2023.

[6] Jingyi Cao, Dongchen Li, Virginia R. Young, and Bin Zou. Strategic underreporting and optimal deductible insurance, 2024.

[7] Tim Miller. *rl-notes*. 2023. URL `https://gibberblot.github.io/rl-notes/single-agent/policy-iteration.html`.

[8] Jae Hyung Min. Optimal reporting strategy of an insured. *Journal of the Korean Operations Research/Management Science Society*, 15(1):83–97, 1990. doi: JAKO199011919754705.