Impact of Heavy-Tailed Rewards on Exploration Strategies in Insurance Underwriting

Joann Bartels, Karsten Bruns

Poster Presentations in context of Reinforcement Learning Lecture

LUH|AI

Leibniz Universität Hannover

1 Summary

In a heavy-tailed insurance simulation, temporally-extended **ε-greedy (EZ)** achieves the highest returns and premiums but with greater variability, **annealed ε-greedy** lowers bankruptcies and stabilizes capital, and **fixed ε-greedy** adapts poorly to regime shifts. The best strategy depends on whether profit maximization or risk control is prioritized.

2 Motivation & Problem Setting

Motivation

- Decisions in an insurance context have long-term consequences and uncertainty.
- Real-world insurance rewards follow heavy-tailed distributions with rare but extreme losses.
- Heavy tails may significantly change optimal exploration strategies.

Problem Setting

- Simulate a simplified insurance environment.
- Goal: maximize the nominal profit over an episode
- Evaluate effect of exploration strategies on efficiency

3 Approach

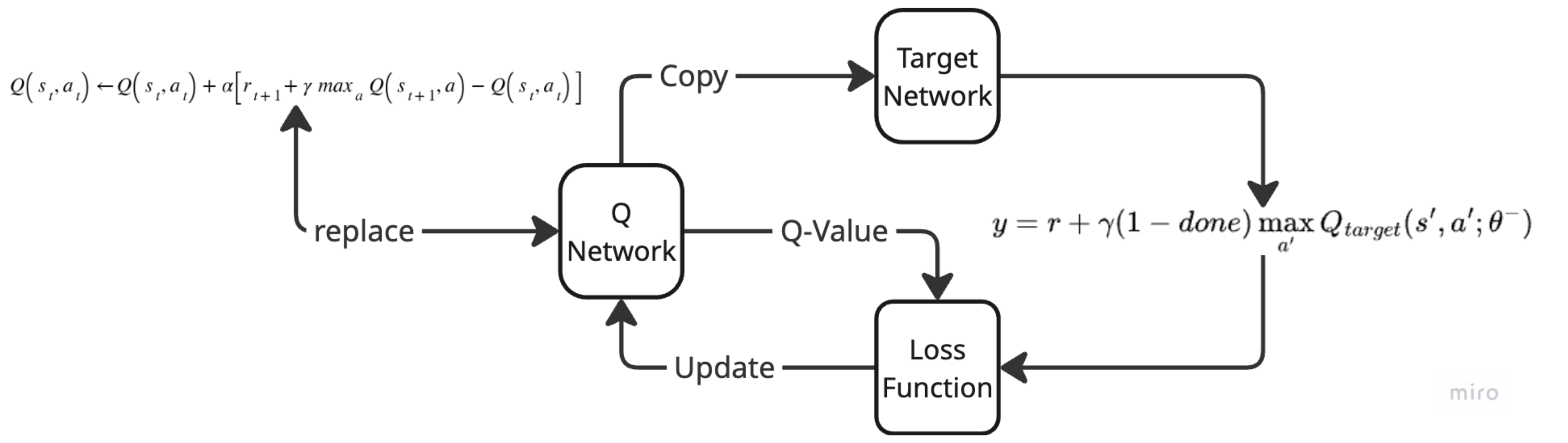
- Sample customer profiles to keep the state space manageable for tabular RL.
- **Delayed rewards**, making it harder to assign credit properly.
- **Deep Q-Network (DQN)** framework.
- Fixed **ε-greedy**, with a constant exploration probability.
- **Annealed ε-greedy,** where exploration decreases linearly over time.
- Adaptive **ez-greedy** that changes exploration based on reward variance and delay:

$$\varepsilon_{base} = \frac{\varepsilon}{k - \varepsilon \cdot (k-1)}$$

- Performance is measured by cumulative discounted returns to capture both gains and stability.
- Claim amounts follow a **heavy-tailed Pareto distribution**, introducing realistic rare but large losses:

$$p(x) = \frac{\alpha \cdot x_m^{\alpha}}{x^{\alpha+1}}$$

4 DQN: Q-Value update



5 Environment

- **Action Space:** either reject or accept with different price factors.
- **Regime Shifts**: 3 regimes with per-step switch probability, affect claim probability and loss multipliers.
- **Claim probability**: depends on risk-score, age factor, region risk of customer and regime.
- **Pricing**: base + base per age·(age−18) + region fee multiplied by price factor.
- **Delay & liabilities**: if a claim occurs, payout is queued with a random delay.
- **Capital & termination**: defined start capital, severe penalty in case of bankruptcy (capital < 0).

6 Key Insights

- **Exploration matters**: ez-greedy boosts end-capital, lowers bankruptcy while keeping returns competitive.
- **Shock windows**: synchronized drawdowns from tail events/regime shifts, exploration increases robustness but can't prevent shocks.
- **Acceptance:** premiums decline from a higher start while acceptance stays high, annealed ε raises acceptance but struggles after regime shifts.
- **Bankruptcy**: low single-digit rates in training and greedy eval; ez-greedy is typically lowest.
- **Trade-offs**: ez-greedy = upside + solvency; annealed ε = smoother but shift-sensitive; fixed ε = simple baseline.

7 Future Works

- Incorporate more complex customer features or dynamic profiles that evolve over time.
- Investigate more sophisticated risk models beyond Pareto for claims distribution.
- Add more Stabilizers (Double-DQN, prioritized replay, soft target updates).
- Explore different environment configurations (action types or losses).