How Far Ahead Should We Look When Noise Persists?

Due: Wednesday, December 9 at 11:59pm

Instructions: Please submit your *handwritten* solutions as a single PDF on Gradescope.

An agent ultimately cares about the far-sighted value

$$V_{\text{target}} = V_{\gamma^\star} = \frac{\mu}{1-\gamma^\star}, \qquad 0 < \gamma^\star < 1, \qquad \mu \neq 0.$$

Across $M \geq 1$ independent episodes, a persistent environmental condition shifts every reward within episode i by the same random amount. Write $r_{i,t}$ for its reward at time t:

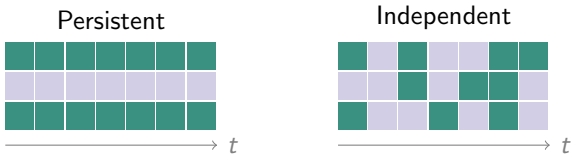
$$r_{i,t} = \mu + Z_i, \qquad \mathbb{E}[Z_i] = 0, \qquad \text{Var}(Z_i) = \tau^2.$$

The episode shocks are independent of one another. To reduce estimation noise, the agent may average the shorter returns

$$\overline{G}_{\gamma,M} = \frac{1}{M} \sum_{i=1}^{M} \sum_{t=0}^{\infty} \gamma^t r_{i,t}$$

for any $0 \leq \gamma \leq \gamma^\star$, accepting some bias relative to V_{target}.

Each row below is one possible episode. The two colors represent opposite signs of a two-valued shock.



1. Within each episode the same shock appears in every reward. What is the mean-squared error of $\overline{G}_{\gamma,M}$ as an estimator of V_{target}?

2. A shorter horizon can reduce variance at the cost of bias. Which $\gamma \in [0, \gamma^\star]$ minimizes the error from part 1?

3. Now suppose the shocks $Z_{i,t}$ are independent across time and episodes, with the same mean and positive variance. How does the variance of $\overline{G}_{\gamma,M}$ compare with its persistent-shock variance as γ approaches one?

Hint: You may use the infinite geometric-series identities and the variance rules for independent sums. A change of variable such as $t = (1-\gamma)^{-1}$ is permitted but not required.

Takeaway: *Noise that persists through an episode accumulates much faster in long returns than independent noise does. A shorter horizon can give a more accurate estimate overall, even though it introduces bias. The best horizon depends on both the noise and the number of episodes.*