Reinforcement Learning

Sebastian Jaimungal
University of Toronto

June, 2022

RL Overview

- Reinforcement Learning aims, generally, to solve stochastic control problems of the form

$$\min_{a \in \mathcal{A}} \mathbb{E}\left[\int_0^T f(s, X_s^a, a_s)\, ds\right]$$

- In financial mathematics, prototypical examples include
  - Portfolio Optimisation
  - Option Hedging
  - Statistical Arbitrage
  - Optimal switching
  - and so on...

RL Overview

- RL aims to optimise in a "model-free" (more precisely model-agnostic) manner
- With the aim of learning only from observations of
  - state
  - action
  - rewards
- The term "models" in RL refers to how optimal actions are approximated

RL Overview

- Two important aspects are:
  - Learning – exploring an unknown environment and learning from it
  - Planning – using what is known to model and then optimise

RL Overview

- There are two main flavors of optimisation
  - Value iteration – approximate the value function, and infer the policy
  - Policy iteration – approximate the policy directly, and estimate the value function

RL Overview

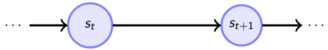
- Actions may be randomised/stochastic or deterministic
  - deterministic action: $a : \mathbb{R}^d \mapsto \mathbb{R}^m$, i.e. maps states to a unique action
  - random action: $a : \mathbb{R}^d \mapsto \mathcal{P}(\mathbb{R}^m)$, i.e. maps states to a distribution over actions

RL Overview

- **Reinforcement learning**(RL) ingredients include

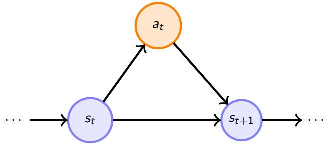
RL Overview

- **Reinforcement learning**(RL) ingredients include
  - agent observes **states** s, e.g., bitcoin prices



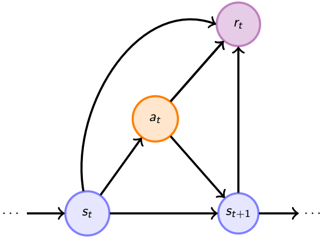
RL Overview

- **Reinforcement learning**(RL) ingredients include
  - agent observes **states** s, e.g., bitcoin prices
  - agent makes **actions** a, e.g., make a trade



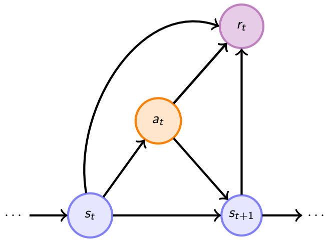
RL Overview

- **Reinforcement learning**(RL) ingredients include
  - agent observes **states** s, e.g., bitcoin prices
  - agent makes **actions** a, e.g., make a trade
  - agent receives **reward** r, e.g., more bitcoin



RL Overview

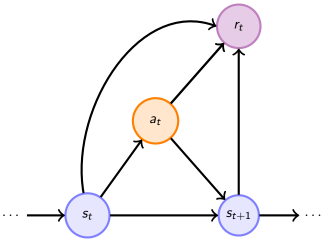
- **Reinforcement learning**(RL) ingredients include
  - agent observes **states** s, e.g., bitcoin prices
  - agent makes **actions** a, e.g., make a trade
  - agent receives **reward** r, e.g., more bitcoin
  - **environment evolves** to new state



RL Overview

- RL is unsupervised – based only on the **rewards** from **actions** & how the **system reacts**
- Can be **model-free** or **model-based**
- Prediction / Control
  - Prediction means to determine the value of a given policy $V[\pi]$
  - Control means to optimise over policies $\pi \in \mathcal{A}$
  - They are closely related as

$$\pi^* = \sup_{\pi \in \mathcal{A}} V[\pi]$$

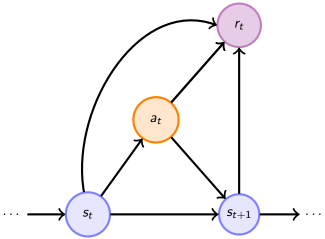


RL Overview

Goal is to **maximize performance criterion**

$$V^{\pi}(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^k \, R(a_t; S_t^{a_t}, S_{t+1}^{a_t}) \,\middle|\, S_0 = s \right]$$

- $\boldsymbol{S_t} \in \mathcal{S}$ is system **state** at time t
- $\boldsymbol{a_t} \in \mathcal{A}$ admissible set of **actions** drawn from π
- System **evolves** as $S_t \overset{a}{\mapsto} S_{t+1}^{a} \sim F(S_t; a_t)$
- $\boldsymbol{R}(a; S_t; S_{t+1})$ is **reward** when $S_t \overset{a}{\mapsto} S_{t+1}$



RL Overview

- Criterion may be risk-aware

$$V^{\pi}(s) = \rho\left(Y \middle| S_0 = s\right)$$

where

$$Y = \sum_{t=0}^{\infty} \gamma^k \, R(a_t; S_t^{a_t}, S_{t+1}^{a_t})$$

and ρ is a risk-measure

- e.g.,
  - a distortion risk-measure

$$\rho(Y) := \int_{-\infty}^{0} \Big\{ 1 - g(\mathbb{P}(U(Y) > y)) \Big\} \, dy - \int_{0}^{+\infty} (\mathbb{P}(U(Y) > y)) \, dy$$

  - standard deviation subtract mean

$$\rho(Y) := \sqrt{\mathbb{V}[Y]} - \mathbb{E}[Y]$$

Hedging Example : Black-Scholes

- Simulate a Black-Scholes model with S satisfying the SDE

$$d(\log S_t) = -\tfrac{1}{2}\sigma^2\, dt + \sqrt{\sigma}\, dW_t^X$$

- Run self-financing hedging strategy α, at hedge times $0 = \tau_0 < \tau_1 < \cdots < \tau_{N-1} < T$ (e.g., daily), with transaction costs

- Wealth process X starts with price of option minus initial hedge $X_0 = V_0 - \alpha_0\, S_0 - \kappa|\alpha_0|\, S_0$

$$X_{\tau_k} = X_{\tau_{k-1}} - \left(\alpha_{\tau_k} - \alpha_{\tau_{k-1}}\right) S_{\tau_{k-1}} - \kappa\, |\alpha_{\tau_k} - \alpha_{\tau_{k-1}}|\, S_{\tau_{k-1}}$$

and liquidate assets and pay option

$$X_{\tau_N} = X_{\tau_{N-1}} + \alpha_{\tau_{N-1}}\, S_T - \kappa\, |\alpha_{\tau_{N-1}}|\, S_T - (S_T - K)_+.$$

Hedging Example: Black-Scholes

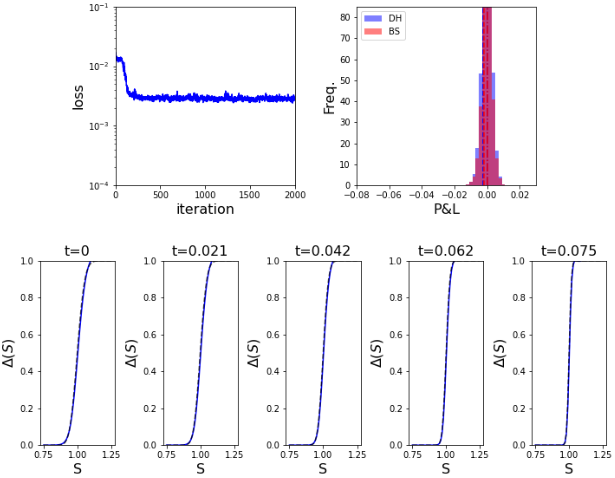


Figure: Optimised for minimising $\sqrt{\mathbb{V}[Y]} - \mathbb{E}[Y]$ no transaction costs.

Hedging Example: Black-Scholes

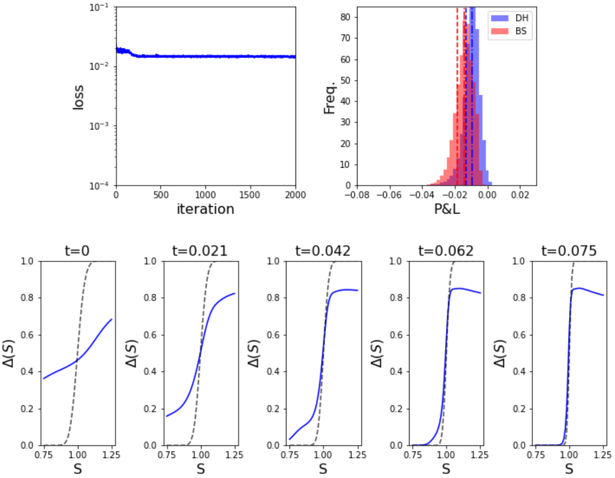


Figure: Optimised for minimising $\sqrt{\mathbb{V}[Y]} - \mathbb{E}[Y]$ with transaction costs.

Hedging Example: Black-Scholes

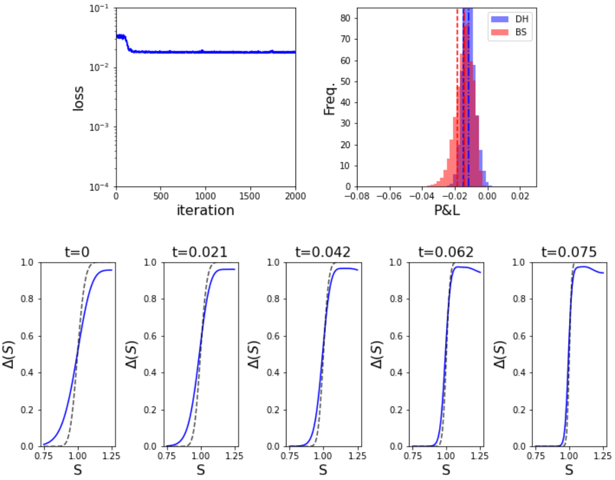


Figure: Optimised for minimise $CVaR_{10}$.

Hedging Example: Black-Scholes

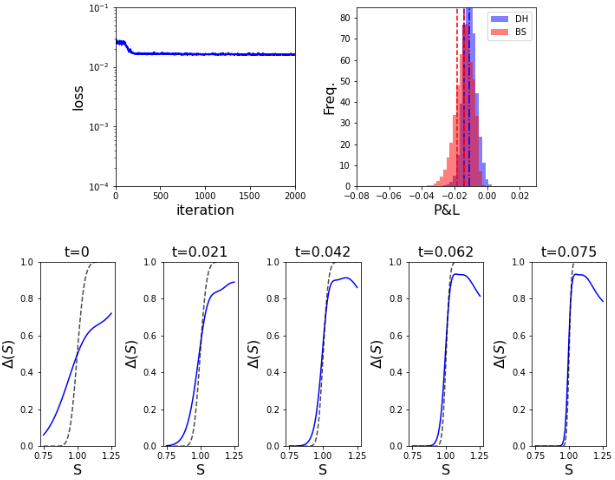


Figure: Optimised for minimise $CVaR_{20}$.

Hedging Example: Black-Scholes

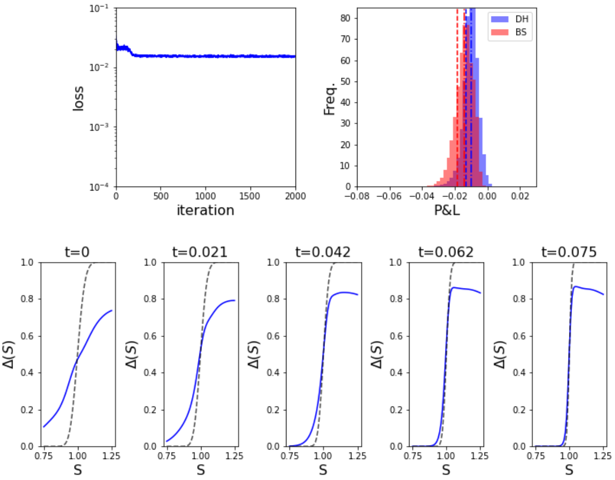


Figure: Optimised for minimise $CVaR_{30}$.

Hedging Example: Black-Scholes

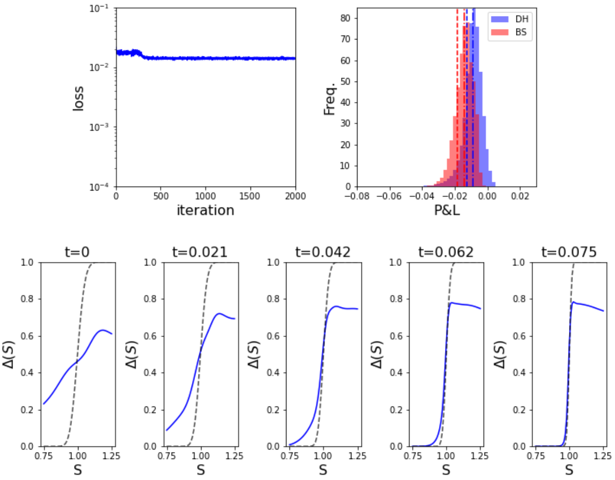


Figure: Optimised for minimise $CVaR_{40}$.

Hedging Example: Black-Scholes

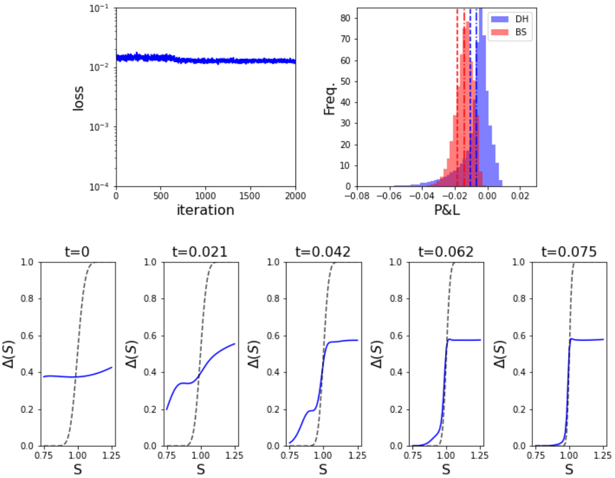


Figure: Optimised for minimise $CVaR_{50}$.

Hedging Example: Black-Scholes

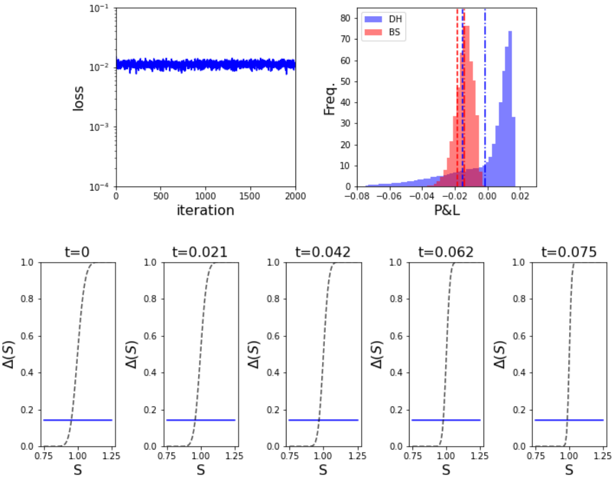


Figure: Optimised for minimise $CVaR_{60}$.

Hedging Example: Black-Scholes

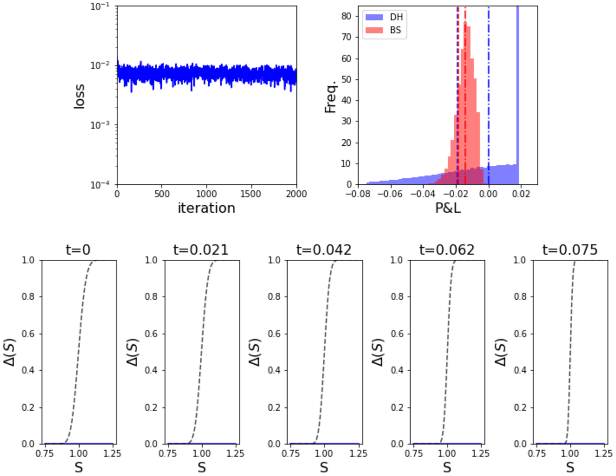


Figure: Optimised for minimise $CVaR_{70}$.

Hedging Example: Black-Scholes

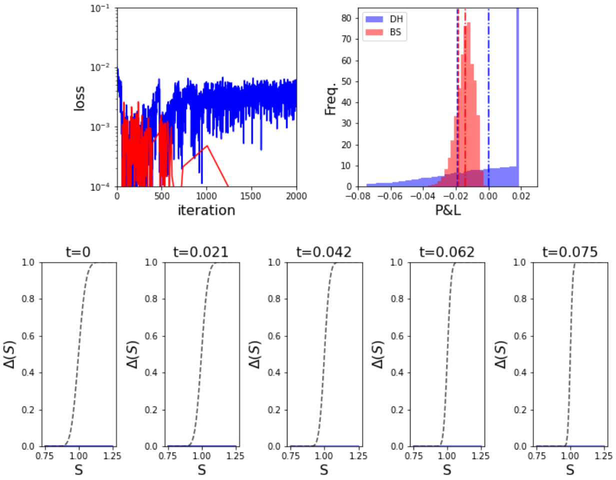


Figure: Optimised for minimise $CVaR_{80}$.

Hedging Example: Black-Scholes

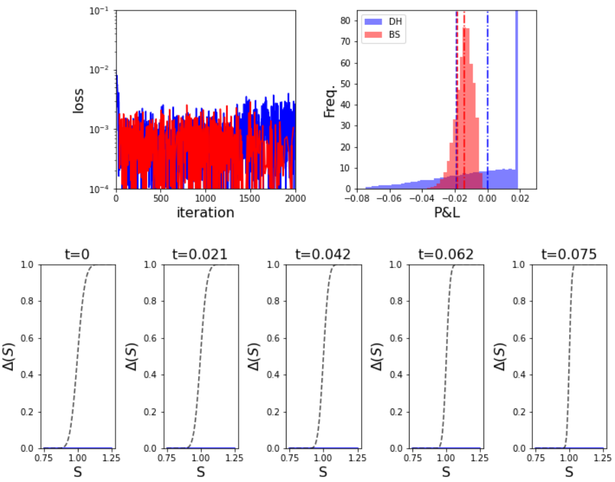


Figure: Optimised for minimise $CVaR_{90}$.

Hedging Example : Heston

- Simulate a Heston model with S satisfying the SDE

$$\begin{aligned} d(\log S_t) &= -\tfrac{1}{2} v_t \, dt + \sqrt{v_t} \, dW_t^X \\ dv_t &= \kappa(\theta - v_t) \, dt + \eta \sqrt{v_t} \, dW_t^v \end{aligned}$$

- Run self-financing hedging strategy α, at hedge times $0 = \tau_0 < \tau_1 < \cdots < \tau_{N-1} < T$ (e.g., daily), with transaction costs

- Wealth process X starts with price of option minus initial hedge $X_0 = V_0 - \alpha_0 \, S_0 - \kappa|\alpha_0| \, S_0$

$$X_{\tau_k} = X_{\tau_{k-1}} - \left(\alpha_{\tau_k} - \alpha_{\tau_{k-1}}\right) S_{\tau_{k-1}} - \kappa \left|\alpha_{\tau_k} - \alpha_{\tau_{k-1}}\right| S_{\tau_{k-1}}$$

and liquidate assets and pay option

$$X_{\tau_N} = X_{\tau_{N-1}} + \alpha_{\tau_{N-1}} \, S_T - \kappa \left|\alpha_{\tau_{N-1}}\right| S_T - (S_T - K)_+ .$$

Hedging Example: Heston

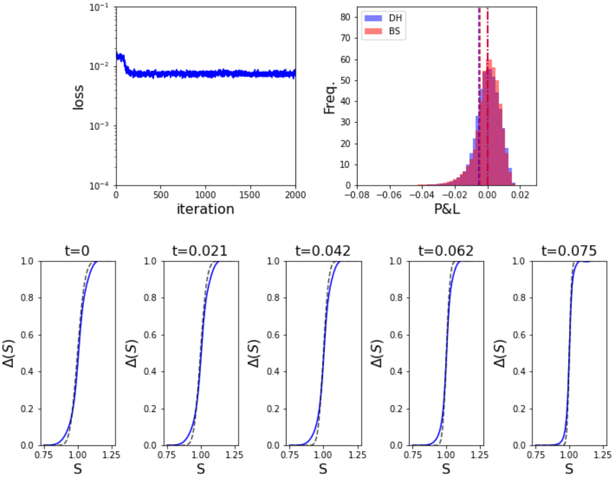


Figure: Optimised for minimising $\sqrt{\mathbb{V}[Y]} - \mathbb{E}[Y]$ no transaction costs.

Hedging Example: Heston

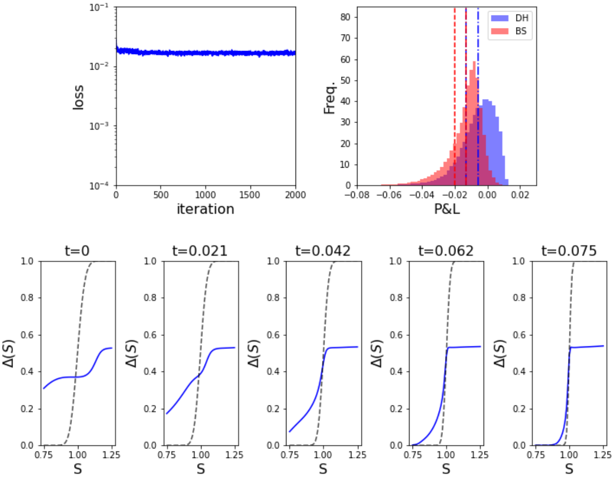


Figure: Optimised for minimising $\sqrt{\mathbb{V}[Y]} - \mathbb{E}[Y]$ with transaction costs.

Hedging Example : Heston

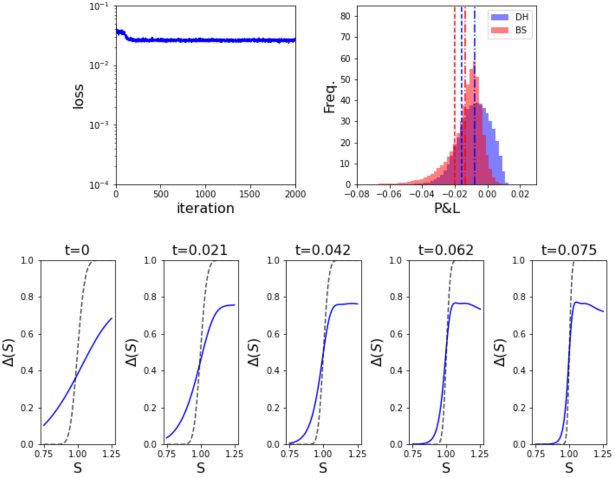


Figure: Optimised for minimise $CVaR_{10}$.

Hedging Example : Heston

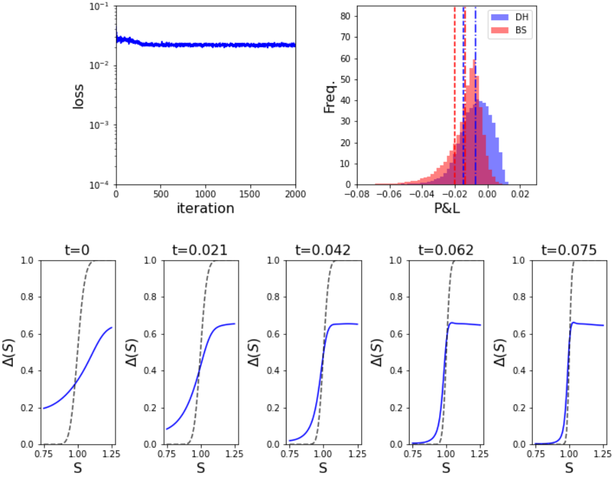


Figure: Optimised for minimise $CVaR_{20}$.

Hedging Example : Heston

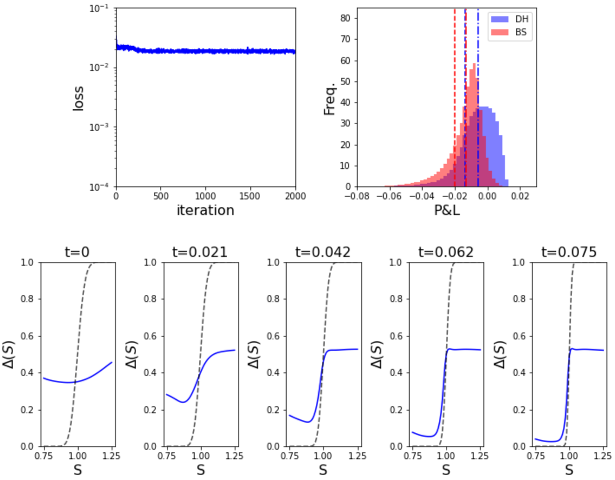


Figure: Optimised for minimise $CVaR_{30}$.

Hedging Example : Heston

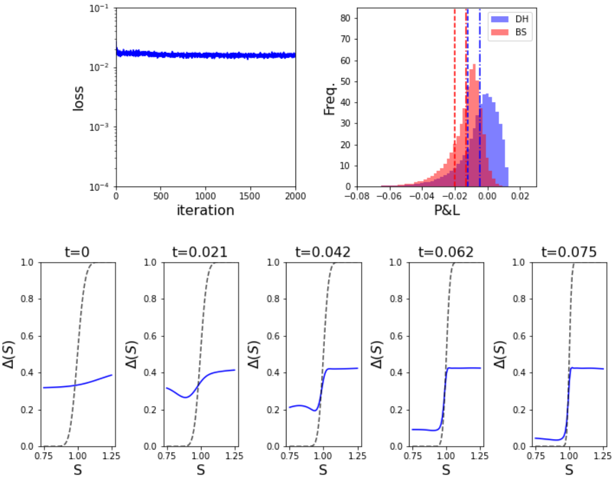


Figure: Optimised for minimise $CVaR_{40}$.

Hedging Example : Heston

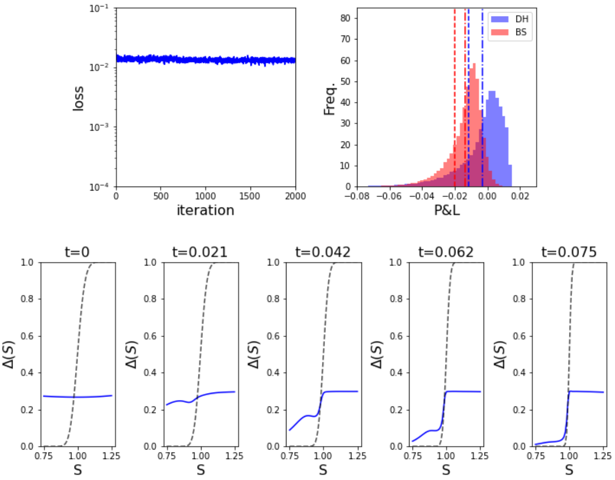


Figure: Optimised for minimise $CVaR_{50}$.

Hedging Example : Heston

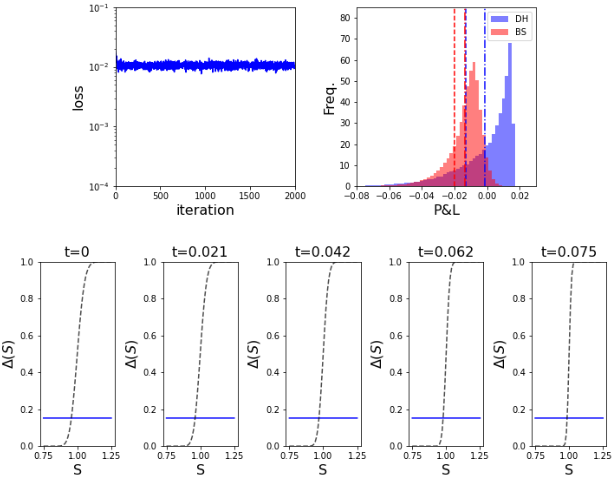


Figure: Optimised for minimise $CVaR_{60}$.

Hedging Example : Heston

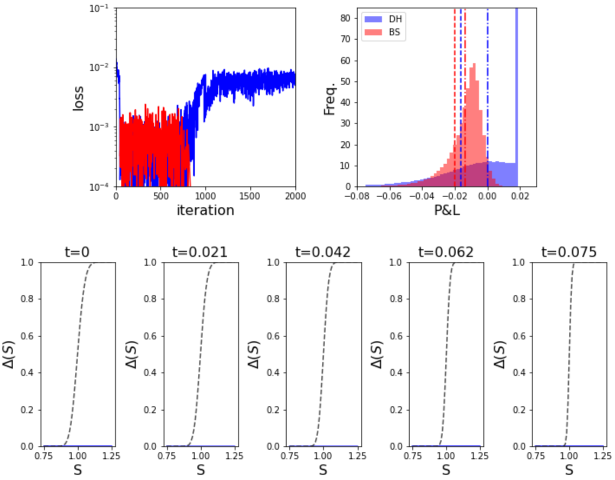


Figure: Optimised for minimise $CVaR_{70}$.

Hedging Example : Heston

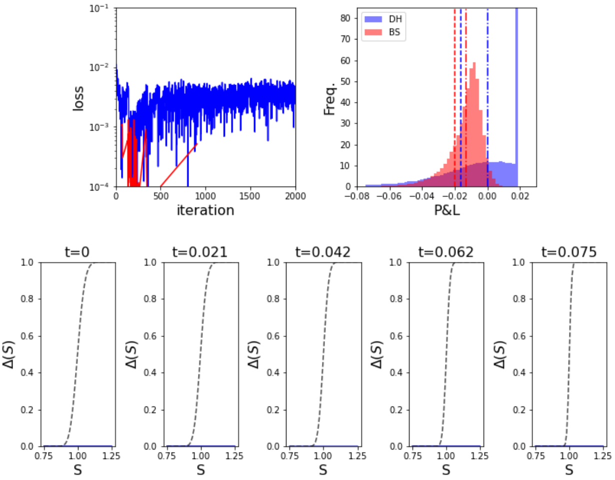


Figure: Optimised for minimise $CVaR_{80}$.

Hedging Example : Heston

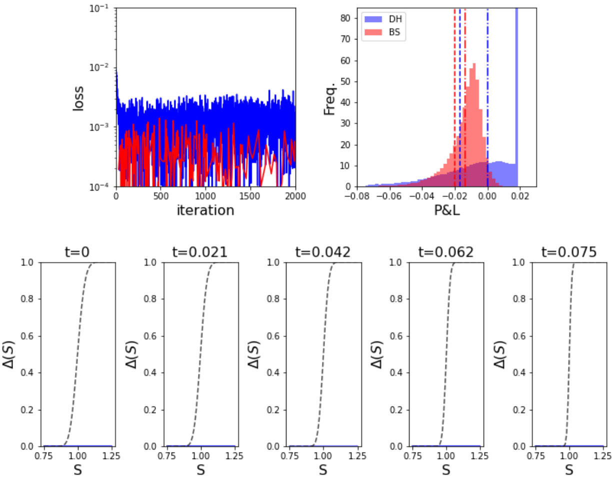


Figure: Optimised for minimise $CVaR_{90}$.