Question 1

With a deterministic policy, we know that:

$$\pi(s) = \begin{cases} 1, & \text{if } a = \pi_D(s) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Therefore, the MDP Bellman equations become:

$$V^{\pi_D}(s) = R^{\pi_D}\left(s, \pi_D(s)\right) + \gamma \cdot \sum_{s' \in \mathcal{N}} P^{\pi_D}\left(s, \pi_D(s), s'\right) \cdot V^{\pi_D}(s') \tag{2}$$

$$V^{\pi_D}(s) = Q^{\pi_D}\left(s, \pi_D(s)\right) \tag{3}$$

$$Q^{\pi_D}\left(s, \pi_D(s)\right) = R^{\pi_D}\left(s, \pi_D(s)\right) + \gamma \cdot \sum_{s' \in \mathcal{N}} P^{\pi_D}\left(s, \pi_D(s), s'\right) \cdot V^{\pi_D}(s') \tag{4}$$

$$Q^{\pi_D}\left(s, \pi_D(s)\right) = R^{\pi_D}\left(s, \pi_D(s)\right) + \gamma \cdot \sum_{s' \in \mathcal{N}} P^{\pi_D}\left(s, \pi_D(s), s'\right) \cdot Q^{\pi_D}\left(s', \pi_D(s')\right) \tag{5}$$

Question 2

We see that:

$$
\begin{aligned}
V(s) &= \mathbb{E}\left(\sum_{t=1}^{\infty} G_t \,\middle|\, S_t = s \right) \\
\therefore V(s) &= \mathbb{E}\left(\sum_{t=1}^{\infty} \left(\frac{1}{2}\right)^{t-1} R_i \,\middle|\, S_t = s \right) \\
\therefore V(s) &= \sum_{t=1}^{\infty} \left(\frac{1}{2}\right)^{t-1} \mathbb{E}\left(R_i \,|\, S_t = s \right) \\
\therefore V(s) &= \sum_{t=1}^{\infty} \left(\frac{1}{2}\right)^{t-1} \cdot \left[(1-a)\cdot a + (1+a)\cdot(1-a)\right] \\
\therefore V(s) &= (1-a)(1+2a) \cdot \sum_{t=0}^{\infty} \left(\frac{1}{2}\right)^{i} \\
\therefore V(s) &= 2(1-a)(1+2a)
\end{aligned}
$$

The optimal value function can be found as shown below:

$$
\begin{aligned}
V^*(s) &= \max_{a\in[0,1]} 2(1-a)(1+2a) = 2\left(1-\frac{1}{4}\right)\left(1+2\cdot\frac{1}{4}\right) = \frac{9}{4} \\
\pi_D(s) &= \operatorname*{argmax}_{a\in[0,1]} 2(1-a)(1+2a) = \frac{1}{4}
\end{aligned}
$$

Question 4

Since $s' \sim N\left(s, \sigma^2\right)$, we know that $e^{as'} \sim \text{lognormal}\left(a\mu, a^2\sigma^2\right)$ and therefore, we need to minimize the following:

$$\min_{a \in \mathbb{R}} \mathbb{E}\left(e^{as'}\right) = \exp\left[a\mu + \frac{a^2\sigma^2}{2}\right]$$

By letting $f(a) = \exp\left[a\mu + \frac{a^2\sigma^2}{2}\right]$, we see that:

$$\begin{aligned} f'(a) &= \exp\left[a\mu + \frac{a^2\sigma^2}{2}\right] \cdot \left(\mu + a\sigma^2\right) \\ f''(a) &= \exp\left[a\mu + \frac{a^2\sigma^2}{2}\right] \cdot \left[\left(\mu + a\sigma^2\right)^2 + \sigma^2\right] \geq 0 \,\forall\, a \in \mathbb{R} \end{aligned}$$

Therefore, we know that $f(a)$ does have a global minimum, which is achieved at the following optimal policy:

$$a^* = -\frac{\mu}{\sigma^2}$$

The corresponding optimal value function is:

$$f\left(a^*\right) = \exp\left[-\frac{\mu}{2\sigma^2}\right]$$