COMP 579 - Assignment 3

Mostafa Dadkhah K. - 261097757

Andreas Enzenhoefer - 260701043

Submission Deadline - April 29, 2022

The simulations of this assignment are carried out with the code in this link.

1 Transforming distributions

1.1 Shift and scale

Probability distribution of $\Pr(R + \gamma Z)$ where $\gamma = \frac{2}{3}$

1.1.1 $Z \sim \mathcal{N}(1, 2)$ and $R = 0.5$

This means we are scaling the distribution of Z by γ and then shifting it by R. In general, scaling by any constant c of any normally distributed random variable Z, i.e. $X = cZ$, will scale the mean $\mu_X = c\mu_Z$ and the standard deviation $\sigma_X = c\sigma_Z$. Shifting by a constant k, i.e. $Y = Z + k$, will only affect the mean $\mu_Y = \mu_Z + k$ but not the standard deviation $\sigma_Y = \sigma_Y$.

$Z \sim \mathcal{N}(1, 2)$ and $\gamma = \frac{2}{3}$ leads to $\gamma Z \sim \mathcal{N}(\gamma, 2\gamma^2) = \mathcal{N}(\frac{2}{3}, \frac{8}{9})$. Shifting it by $R = 0.5$ is then $R + \gamma Z \sim \mathcal{N}(R + \frac{2}{3}, \frac{8}{9}) = \mathcal{N}(\frac{7}{6}, \frac{8}{9})$. The probability distribution is given by $\Pr(Y) = \Pr(R + \gamma Z) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp(-\frac{(Y-\mu_y)^2}{2\sigma_y^2}) = \frac{1}{\sqrt{2\pi\frac{8}{9}}} \exp(-\frac{(Y-\frac{7}{6})^2}{2\frac{8}{9}})$

1.1.2 $Z \sim \mathcal{N}(1, 2)$ and $R \sim \mathcal{B}(\frac{3}{4})$

First, we scale the distribution of Z by γ such that $X = \gamma Z \sim \mathcal{N}(\gamma, 2\gamma^2) = \mathcal{N}(\frac{2}{3}, \frac{8}{9})$. Also, the probability can be conditioned over R because it is either zero or one.

$$
\begin{aligned}
\Pr(R + \gamma Z = Y) &= \Pr(1 + \gamma Z = Y).\Pr(R = 1) + \Pr(\gamma Z = Y).\Pr(R = 0) \\
&= \Pr(\gamma Z = Y - 1)\frac{3}{4} + \Pr(\gamma Z = Y)\frac{1}{4} \\
&= \Pr(y - 1 | y \sim \mathbb{N}(\gamma, \gamma^2))\frac{3}{4} + \Pr(y | y\ \mathbb{N} \sim (\gamma, \gamma^2))\frac{1}{4} \\
&= \Pr(y | y \sim \mathbb{N}(\gamma + 1, \gamma^2))\frac{3}{4} + \Pr(y | y\ \mathbb{N} \sim (\gamma, \gamma^2))\frac{1}{4} \\
&= \Pr(y | y \sim \mathbb{N}(\frac{5}{3}, \frac{4}{9}))\frac{3}{4} + \Pr(y | y\ \mathbb{N} \sim (\frac{2}{3}, \frac{4}{9}))\frac{1}{4}
\end{aligned}
$$

1.1.3 $Z \sim \mathcal{N}(1, 2)$ and $R \sim \mathcal{N}(0, 1)$

First, we scale the distribution of Z by γ such that $X = \gamma Z \sim \mathcal{N}(\gamma, 2\gamma^2) = \mathcal{N}(\frac{2}{3}, \frac{8}{9})$. Then, the distribution for the sum of two independent normally distributed random variables can be computed by adding their mean and variance $R + \gamma Z = \mathcal{N}(\mu_R + \mu_X, \sigma_R^2 + \sigma_X^2) = \mathcal{N}(\frac{2}{3}, \frac{17}{9})$. The probability distribution is given by $\Pr(Y) = \Pr(R + \gamma Z) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp(-\frac{(Y-\mu_y)^2}{2\sigma_y^2}) = \frac{1}{\sqrt{2\pi\frac{17}{9}}} \exp(-\frac{(Y-\frac{2}{3})^2}{2\frac{17}{9}})$

1.2 Distribution of G

$G = \sum_{t=0}^{\infty} \gamma^t R_t = \sum_{t=0}^{\infty} R_0 + \gamma R_1 + \gamma^2 R_2 + \ldots$.
According to the central limit theorem, the sum of many random variables has a normal distribution in the limit. The discount factor can be considered as a soft-form of the theorem, so we will assume that the return has a normal distribution, and then it will be verified with a one-state simulation.

1.2.1 $R \sim \mathcal{N}(0,1)$ for all t

$$G = \sum_{t=0}^{\infty} \gamma^t . R_t$$

With considering the discounting factor as a variance and using it in the characteristic function of normal distribution:

$$\begin{aligned}
\varphi_X(s) &= exp(is\mu - \sigma^2 s^2/2) = \mathbb{E}[\exp{(-\gamma^{2t} s^2/2)}] \\
\varphi_G(s) &= \lim_{n\to\infty} \mathbb{E}[\prod_{t=0}^{n} \varphi_X(s)] \\
&= E[\exp{(\sum_{t=0}^{\infty} \gamma^{2t} . \frac{-s^2}{2})}] \\
&= E[\exp{(\frac{-s^2}{2} \sum_{t=0}^{\infty} \gamma^{2t})}] \\
&= E[\exp{(\frac{1}{1-\gamma^2} . \frac{-s^2}{2}}]
\end{aligned}$$

In general, sum of discounted rewards with distribution of N$(\mu, \sigma) hasadistributionasbellow$:

$$\begin{aligned}
\varphi_G(s) &= exp(is \sum \mu_i \gamma^t - s^2/2 \sum \gamma^{2t} \sigma^2) \\
\sigma_G^2 &= \sum_{t=0}^{\infty} (\gamma)^{2t} \sigma^2 = \frac{\sigma^2}{1-\gamma^2} \\
\mu_G &= \sum_{t=0}^{\infty} \mu \gamma^t = \frac{\mu}{1-\gamma}
\end{aligned}$$

Simulation: The result shows that as the γ increases, the distribution remains Normal but wider. Also, the fitted line to the variances confirms the characteristic formulation evaluated before:

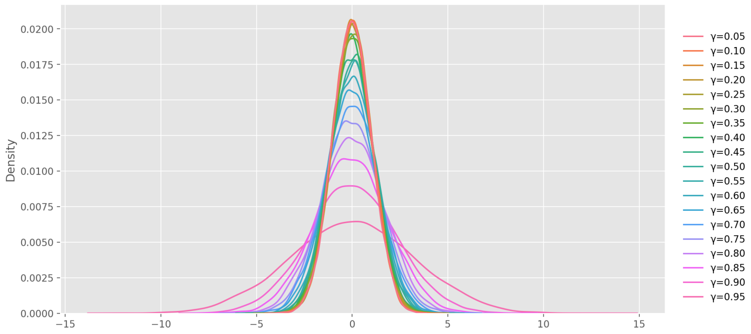
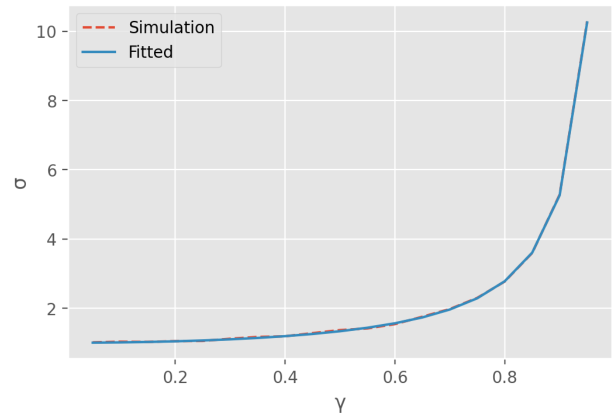
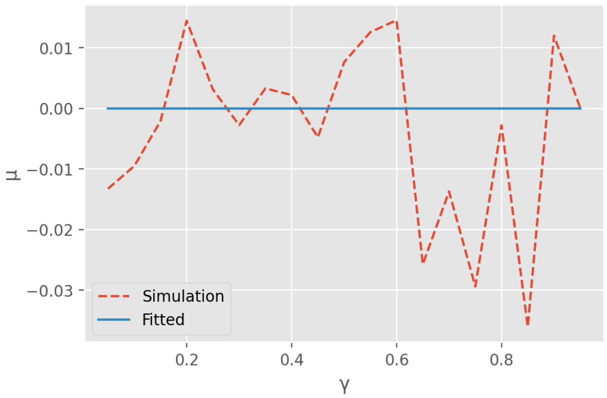


Figure 1: Distribution of returns



(a) Variance of returns



(b) Mean of returns

1.2.2 $R \sim \mathcal{U}([1,2])$

With defining a' and b' as $\frac{b-a}{2} - \frac{b-a}{2}.\gamma^t$ and $\frac{b-a}{2} + \frac{b-a}{2}.\gamma^t$ respectively and putting them in the moment generating fucntion of $U(a', b')$ we have:

$$
\begin{aligned}
M_1 &= \frac{a' + b'}{2} \\
&= \frac{\frac{b-a}{2} - \frac{b-a}{2}.\gamma^t + \frac{b-a}{2} + \frac{b-a}{2}.\gamma^t}{2} \\
&= \frac{a+b}{2} \\
M_2 &= \frac{a'^2 + b'^2 + a'b'}{3} \\
&= \frac{(\frac{b-a}{2} - \frac{b-a}{2}.\gamma^t)^2 + (\frac{b-a}{2} + \frac{b-a}{2}.\gamma^t)^2 + (\frac{b-a}{2} - \frac{b-a}{2}.\gamma^t)(\frac{b-a}{2} + \frac{b-a}{2}.\gamma^t)}{3} \\
&= \frac{3\mu^2 + \frac{(b-a)^2\gamma^{2t}}{4}}{3} \\
&= \mu^2 + \frac{(b-a)^2\gamma^{2t}}{12} \\
\sigma^2 &= M_2 - M_1^2 = \frac{(b-a)^2\gamma^{2t}}{12}
\end{aligned}
$$

Because the variables are independent, we could sum over them:

$$
\begin{aligned}
\sigma_G^2 &= \sum_{t=0}^{\infty} \frac{(b-a)^2\gamma^{2t}}{12} \\
&= \frac{1}{12}\sum_{t=0}^{\infty} \gamma^{2t} = \frac{1}{12(1-\gamma^2)}
\end{aligned}
$$

Also, for the average we can use:

$$
\begin{aligned}
\mu_G &= \sum_{t=0}^{\infty} \frac{(b+a)\gamma^t}{t} \\
&= \frac{b+a}{2}\sum_{t=0}^{\infty} \gamma^t = \frac{3}{2(1-\gamma)}
\end{aligned}
$$

Simulation: The result shows that as the γ increases, the distribution moves from uniform to a Normal distribution but shifts to higher and wider returns. Also, the fitted line to the variances is:

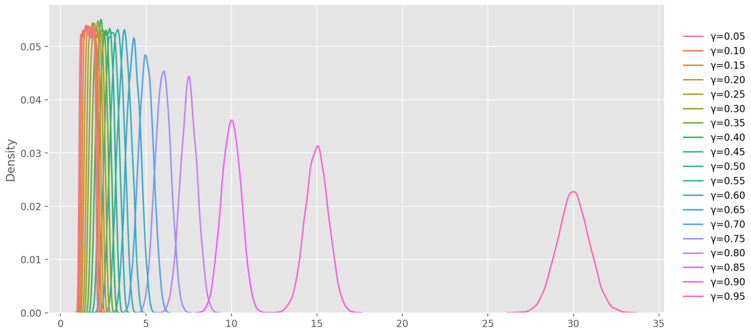
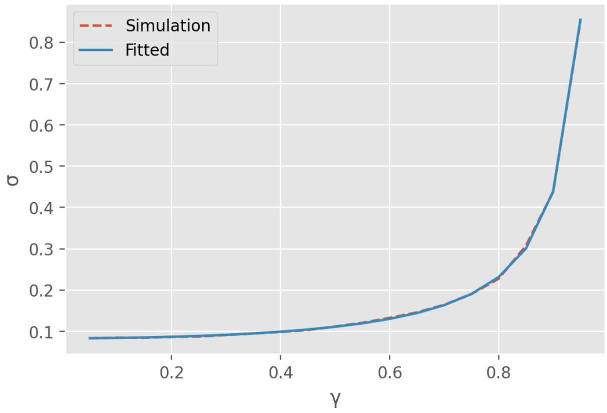
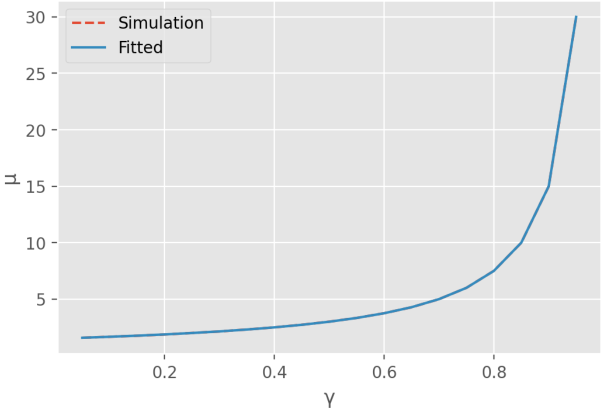


Figure 3: Distribution of returns



(a) Variance of returns



(b) Mean of returns

1.2.3 $R \sim \frac{1}{3}\delta_{-1} + \frac{1}{3}\delta_0 + \frac{1}{3}\delta_1$

With using the characteristic function of dirac distribution we have:

$$\begin{aligned}
\varphi_{\delta x}(s) &= exp(ip(s-x)) \\
\varphi_R(s) &= exp(i\frac{1}{3}(s-1)).exp(i\frac{1}{3}(s)).exp(i\frac{1}{3}(s+1)) \\
\varphi_{\gamma^t R}(s) &= exp(i\frac{\gamma^t}{3}(s-1)).exp(i\frac{\gamma^t}{3}(s)).exp(i\frac{\gamma^t}{3}(s+1)) \\
&= exp(i\gamma^t(s)) \\
\varphi_G(s) &= \prod_{t=0}^{\infty} exp(i\gamma^t(s)) \\
&= exp(\sum_{t=0}^{\infty} \gamma^t is) \\
&= exp(si\sum_{t=0}^{\infty} \gamma^t) = exp(si\frac{1}{1-\gamma})
\end{aligned}$$

Therefore, the distribution is $\frac{1}{1-\gamma}\delta_0$.
Simulation: The result shows that as the γ increases, the distribution moves from multinomial to a Normal

distribution but wider. Also, the fitted line to the variances and means are:

$$\sigma_G^2 = \frac{2}{3(1-\gamma^2)}, \mu = 0$$

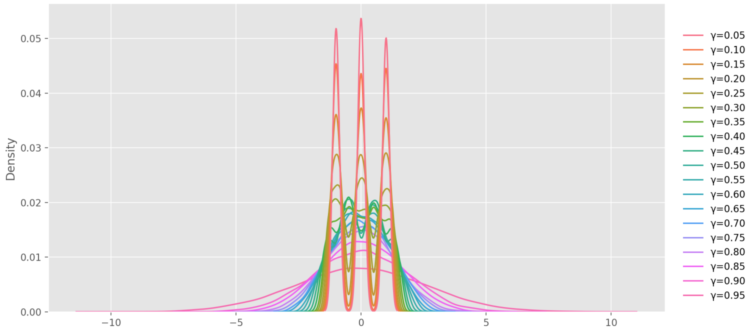
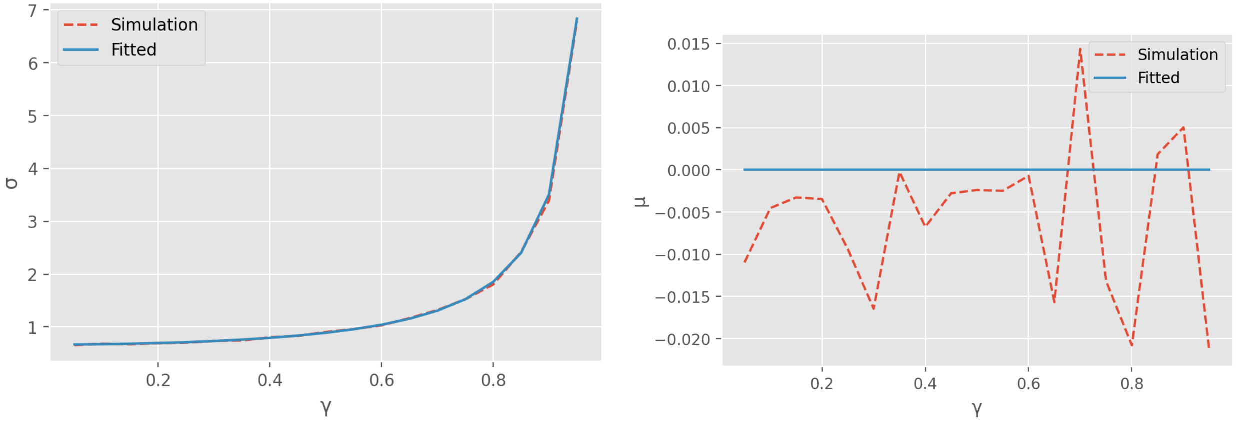


Figure 5: Distribution of returns



(a) Variance of returns

(b) Mean of returns

1.2.4 $R \sim \mathcal{N}(1, t)$

With constructing the characteristic function of Normal Distribution, the variances and mean are:

$$\begin{aligned}
\varphi_X(s) &= exp(is\mu - \sigma^2 s^2/2) \\
&= exp(is\gamma^t - \gamma^{2t} t s^2/2) \\
\varphi_G(s) &= \prod_{t=0}^{\infty} \varphi_{x_t}(s) \\
&= exp(is\sum_{t=0}^{\infty}\gamma^t + s^2/2\sum_{t=0}^{\infty}\gamma^{2t}t) \\
\gamma^{2t}t &= \frac{1}{2}\gamma(\frac{d}{d\gamma}\gamma^{2t}) \\
\sigma_G^2 = \sum_{t=0}^{\infty}\gamma^{2t}t &= \frac{1}{2}\gamma(\frac{d}{d\gamma}\sum_{t=0}^{\infty}\gamma^{2t}) \\
&= \frac{1}{2}\gamma(\frac{d}{d\gamma}\frac{1}{1-\gamma^2}) \\
&= \frac{\gamma^2}{(1-\gamma^2)^2} \\
\mu_G = \sum_{t=0}^{\infty}\gamma^t &= \frac{1}{1-\gamma}
\end{aligned}$$

Simulation: The result shows that as the γ increases, the distribution moves from uniform to a Normal distribution but shifts to higher and wider returns.

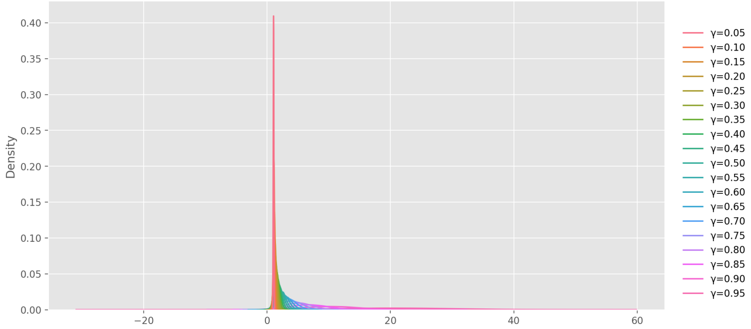
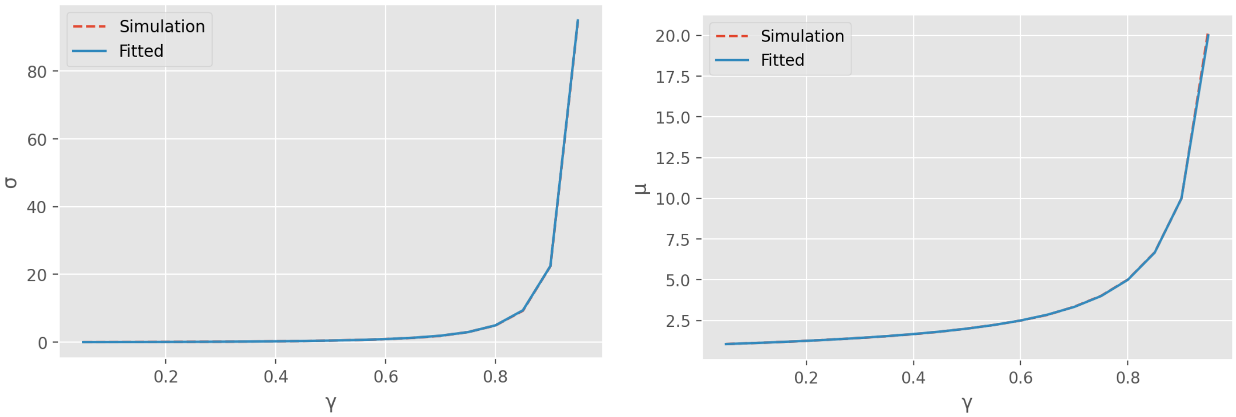


Figure 7: Distribution of returns



(a) Variance of returns

(b) Mean of returns

1.3 Return with infinite variance

Assuming $G = \sum_{t=0}^{\infty} \gamma^t R_t$

In the first part of Q(1,3), the parameters of return distribution are introduced. With the rewards of $N(0,1)$, the variance of return goes to infinity if the discount factor moves to 1^- and $\mu = 0$ or we can use a distribution such that $\sigma_R^2 \sim +\infty$.

$$\sigma_G^2 = \lim_{\gamma \to 1^-} \sum_{t=0}^{\infty} (\gamma)^{2t} \sigma^2 = \frac{\sigma^2}{1-\gamma^2} = +\infty$$

$$\mu_G = \lim_{\gamma \to 1^-} \sum_{t=0}^{\infty} \mu \gamma^t = \frac{0}{1-\gamma} = 0$$

2 Categorical dynamic programming

First, the support locations will be created. In the first step, we suggest the locations between a safe bound. For example, with a bound of $+/-3$ variance, 99 % of the distribution will be covered. Therefore, the locations in the standard normal distribution can be as below:

$$z_i = -3 + \frac{+3-(-3)}{m-1}(i+1) = 3(1 - \frac{2(i+1)}{m-1})$$

The influenced area for the support z_i is a triangular area in the vicinity of the center and decreases linearly farther from the center. Hence, the projection of the scaling the reward to standard normal distribution would be:

$$PC(\delta_y) = \begin{cases} \delta_{z_1} & \text{if: } y \leq z_1, \\ \delta_{z_i}[1 - z_i + y] & \text{if: } z_i - 1 < y < z_i, \forall i \in (2, m), \\ \delta_{z_i}[1 + z_i - y] & \text{if: } z_i < y < z_i + 1, \forall i \in (1, m-1), \\ \delta_{z_m} & \text{if: } y > z_m. \end{cases}$$

The p_i for the non-extreme supports can be calculated by the integrals over left and right of the support's location:

$$p_i = \mathbb{E}[h_i(Y)] = \int_{z_i - 1}^{z_i} (1 - z_i + y).P(y)\, dy \ + \int_{z_i}^{z_i+1} (1 + z_i - y).P(y)\, dy$$

in the above formulation, $\int y.p(y)dy$ would be:

$$\int y.P(y) = \frac{1}{\sqrt{2\pi}} \int y.e^{-\frac{y^2}{2}} = \frac{-1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = -P(y)$$

Therefore the parameter p_i is equal to:

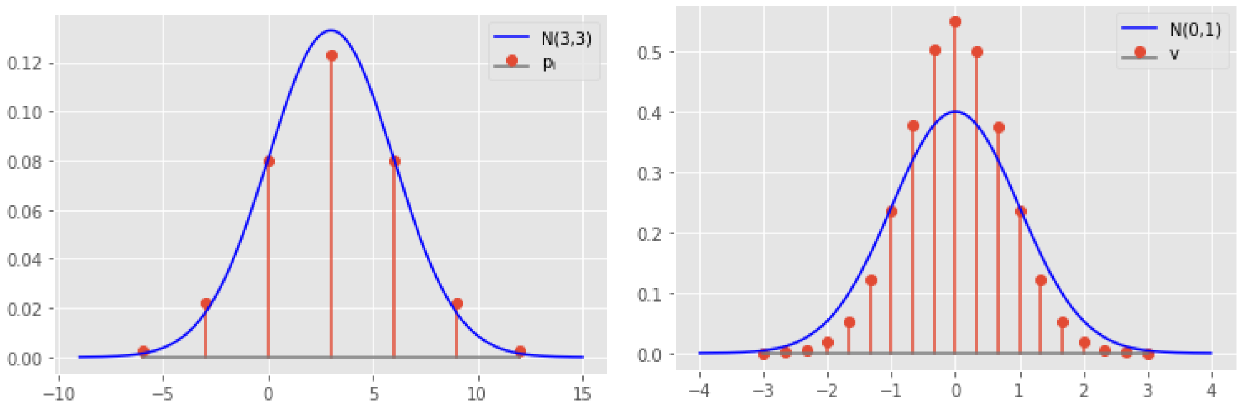
$$\begin{aligned} \mathrm{P}_i =& (1 - z_i) \int_{z_i - 1}^{z_i} \mathrm{Pr}(y)\mathrm{d}y + \int_{z_i - 1}^{z_i} y.\mathrm{Pr}(y).\mathrm{d}y + \\ & (1 + z_i) \int_{z_i}^{z_i+1} \mathrm{Pr}(y).\mathrm{d}y - \int_{z_i}^{z_i+1} y.\mathrm{Pr}(y).\mathrm{d}y \\ =& (1 - z_i)[F(z_i) - F(z_i - 1))] + [P(z_i - 1) - P(z_i)] + \\ & (1 + z_i)[F(z_i + 1) - F(z_i)] - [P(z_i) - P(z_i + 1)] \\ =& [F(z_i + 1) - F(z_i - 1)] + \\ & [F(z_i + 1) + F(z_i - 1) - 2F(z_i)]z_i + \\ & [P(z_i + 1) + P(z_i - 1) - 2P(z_i)] \end{aligned}$$

for the extreme supports we have:

$$
\begin{aligned}
p_1 = \mathbb{E}[h_1(Y)] = &\int_{-\infty}^{z_1} 1.P(y).dy + \int_{z_1}^{z_1+1} (1+z_1-y).P(y).dy \\
=&F(z_1) + (1+z_1)[F(z_1+1) - F(z_1)] - [P(z_1) - P(z_1+1)] \\
=&(1+z_1)F(z_1+1) - z_1.F(z_1) - [P(z_1) - P(z_1+1)]
\end{aligned}
$$

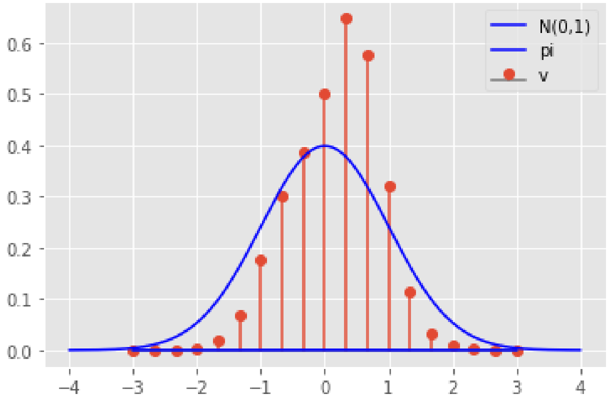
$$
\begin{aligned}
p_m = \mathbb{E}[h_m(Y)] = &\int_{z_m-1}^{z_m} (1-z_m+y).P(y).dy + \int_{z_m}^{+\infty} 1.P(y).dy \\
=&1 - F(z_m) + (1-z_m)[F(z_m) - F(z_m-1)] + [P(z_m-1) - P(z_m)] \\
=&-(1-z_m)F(z_m-1) + 1 - z_m F(z_m) + [P(z_m-1) + P(z_m)]
\end{aligned}
$$

A one-step simulation is carried out in which, in each step, the environment reveals one v for the current state. In the algorithm, first, we calculate the support positions between -3 and 3 in the standard normal distribution to cover at least 99% of the space. Then, with the projection and shifting of the revealed v to the standard normal distribution, we calculate the $PC(\delta_y)$ and multiply it to p_i. We can use the expected of p_i as calculated before which the results are in the first picture. Also, one can not use these parameters and calculate the empirical mean in each iteration which the results are in the following image.



(a) p$_i$ is expected

(b) Simulation



(a) p$_i$ is calculated

2.1 [Bonus part]

With the definition of W, the value of each support would be the probability of that point in the given Normal distribution. Therefore, the criteria could minimize the highest difference between supports of v and W.