2. a) Stages = N

$$P(\underset{\text{at store } i}{\text{food we like}}) = P \quad , \quad i \in 1, \dots, N$$

if we like the food in store i:

Actions : 1. buy the food

2. do not buy the food

if we don't like the food in store i:

Action : do not buy the food

$$J_N(x_N) = \frac{1}{1-P}$$

b) $$J_k(x_k) = P\left[\min\left(N-k, J_{k+1}(x_{k+1})\right)\right] + (1-P)\left[\min\left(J_{k+1}(x_{k+1})\right)\right]$$

$$= P \cdot \min\left[n-k, J_{k+1}(x_{k+1})\right] + (1-P) \cdot J_{k+1}(x_{k+1})$$

for $k = N-1$:

$$J_{N-1}(x_{N-1}) = P\left[\min\left(1, \frac{1}{1-P}\right)\right] + \cancel{(1-P)} \cdot \frac{1}{\cancel{1-P}}$$

we know $\frac{1}{1-P} \geq 1$ as $0 \leq P \leq 1$

$\therefore J_{N-1}(x_{N-1}) = 1 + P$, Policy = buy the food for N-2 :

$$J_{N-2}(x_{N-2}) = P\left[\min(2, 1+P)\right] + (1-P)(1+P)$$

we know $1+P < 2$ as $0 \le P \le 1$

$$\therefore J_{N-2}(x_{N-2}) = P(1+P) + 1-P^2$$

$$= 1+P$$

policy : do not buy food

for N-3 :

$$J_{N-3}(x_k) = P\left[\min(3, 1+P)\right] + (1-P)(1+P)$$

$$= 1+P$$

Policy : do not buy food

we can see for all $k = 0, \dots, N-1$

$$N-k > 1+P$$

$\therefore$ In all stages except the N^{th} ~~st~~ and terminal stage we make the policy : not to buy food.

At the N^{th} stage, we can ~~either~~ ~~choose to~~ buy if the shop contains the food we like, otherwise we can ignore the shop and buy food at the terminal stage.

3) a) States : 1. Running , 2. Broken

State 1 for 1 week : profit = ₹ 1000

State 2 for 1 week : profit = ₹ 0

State 1 → preventive maintenance : Cost = ₹ 200
action : a_1

State 1 → No preventive maintenance : Cost = ₹ 0
action : a_2

action a_1 on state 1 : p(machine fail) = 0.4

action a_2 on stat 2 : p(machine fail) = 0.7

State 2 → repair : Cost = ₹ 400
action : a_3

State 2 → replaced : Cost : ₹ 1500
action : a_4

action 3 on state 2 : p(machine fail) = 0.4

action 4 on state 2 : p(machine fail) = 0

b) ₹ week 1 , week 2 , week 3 , week 4 , termi

J_0 J_1 J_2 J_3 J_4

Terminal ~~to~~ Reward :

$$J_4 = \begin{cases} 1000 & state = 1 \\ 0 & state = 2 \end{cases}$$

$$J_3(1) = \max\Big[(-200+1000)+(0.6\times1000+0.4\times0),$$
$$(-0+1000)+(0.3\times1000+0.7\cdot0)\Big]$$

$= 1400$

Policy = action 1

$$J_3(2) = \max\Big[(-400+0)+(0.6\times1000+0.4\times0),$$
$$(-1500+0)+(1\times1000+0\cdot0)\Big]$$

$= 200$

Policy = action 3

$$J_2(1) = \max\Big[(-200+1000)+(0.6\times1400+0.4\times200),$$
$$(-0+1000)+(0.3\times1400+0.7\cdot200)\Big]$$

$= 1720$

Policy = action 1

$$J_2(2) = \max\Big[(-400+0)+(0.6\times1400+0.4\times200),$$
$$(-1500+0)+(1\times1400+0\cdot200)\Big]$$

$= 520$

Policy = action 3

$$J_1(1) = \max\left[\cancel{10}(-200+1000) + (0.6\times 1720 + 0.4\cdot 520),\right.$$
$$\left.(-0+1000) + (0.3\cdot 1720 + 0.7\cdot 520)\right]$$

$$= 2040$$

Policy action = ~~$\pounds$~~ action 1

There is no possibility of $J_1(2)$ as we got a freshly replaced machine on the first week which is guaranteed to stay in state 1 throughout the week.

	State	
	Running	Broken
J_3	action 1	action 3
J_2	action 1	action 3
J_1	action 1	–

2) c) ~~At~~ ~~at~~ At each store we are not sure whether it contains food we like or not and the only cost is food carrying cost. So in order to minimize cost, we can buy at the last store available if it has the food we like or otherwise

Scanned with CamScanner

buy the food at the terminal stage.

1. a)

we know $J_\pi(x_0) = E\left[\exp\left(g_N(x_N) + \sum_{k=0}^{N-1} g_k(\mu_k, x_k, x_{k+1})\right)\right]$

for any admissible policy $\pi = \{\mu_0, \ldots \mu_{N-1}\}$

Let $\pi^k = \{\mu_k, \ldots \mu_{N-1}\}$

$J_k^*(x_k)$ be the optimal cost of tail-subproblem

$$J_k^*(x_k) = \min_{\pi^k = \mu_k, \ldots \mu_{N-1}} E\left[\exp(g_N(x_N) + \sum_{i=k}^{N-1} g_i\right]$$

Induction hypothesis:

$$J_N^*(x_N) = \cancel{g_N(x_N)}\ E\left[\exp(g_N(x_N))\right]$$

$$= \exp(g_N(x_N))$$

$$J_k^*(x_k) = \min_{\mu_k, \pi^{k+1}} E\left[\exp\left(g_N(x_N) + g_k + \sum_{i=k+1}^{N-1} g_i(x_i, \mu_i, x_{i+1})\right)\right]$$

$$= \min_{\mu_k} \underset{x_{k+1}}{E} \left[\exp\right] g_k + \min_{\pi^{k+1}} \underset{x_{k+2} \ldots x_N}{E} \left[g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i, \mu_i(x_i), x_{i+1})\right]\right]$$

$$= \min_{\mu_k} \underset{x_{k+1}}{E} \left[\underbrace{\exp(g_k)}_{\text{term 1}} \cdot \left.\exp\left(\min_{\pi^{k+1}} E\left[g_N(x_N) + \sum_{i=k+1}^{N-1} g_i\right)\right\}\right.\right. \text{term 2}$$

Scanned with CamScanner

term 2 $J^*_{k+1}(x_{k+1}) = J_{k+1}(x_{k+1})$

by induction hypothesis

$$\therefore J_k(x_k) = \min_{a_k \in A(x_k)} \mathbb{E}_{x_{k+1}} \left(\exp(g_k(x_k, a_k, x_{k+1})) J_{k+1}(x_{k+1}) \right)$$

b) given : $V_k(x_k) = \log(J_k(x_k))$

$$\therefore V_N(x_N) = \log(J_N(x_N))$$

$$= \cancel{\log}(\cancel{\exp}(g_N(x_N))$$

$$= g_N(x_N)$$

$$V_k(x_k) = \log\left(\min_{a_k \in A(x_k)} \mathbb{E}_{x_{k+1}} [\exp(g_k(x_k, a_k)) \cdot J_{k+1}(x_{k+1})]\right)$$

we can take the log inside

$$= \min_{a \in A(x_k)} g_k(x_k, a_k) + \log \mathbb{E}_{x_{k+1}} (J_{k+1}(x_{k+1}))]$$

we know $\exp(J_{k+1}(x_{k+1})) = \cancel{\exp}\ \cancel{E_x}\ V_{k+1}(x_{k+1})$

$$\therefore V_k(x_k) = \min_{a_k \in A(x_k)} g_k(x_k, a_k) + \log E_{x_{k+1}} \exp(V_{k+1}(x_{k+1})))$$

1. c) $J_{a_0, a_1} = E\left[\exp\left(\theta\left(a_0^2 + a_1^2 + (x_2 - T)^2\right)\right)\right]$

$$x_1 = (1-\alpha)x_0 + \alpha a_0 + \omega_0$$

$$x_2 = (1-\alpha)x_1 + \alpha a_1 + \omega_1$$

w.k.T: $$\int_{-\infty}^{\infty} e^{(-ax^2 - bx - c)} \cdot dx = \sqrt{\frac{\pi}{a}}\, e^{(b^2 - 4ac)/4a} \longrightarrow (1)$$

$$J_2(x_2) = e^{\theta(x_2 - T)^2} \longrightarrow (2)$$

$$J_1(x_1) = \min_{a_1} e^{\theta a_1^2} + E(J_2) \longrightarrow (3)$$

$$J_0(x_0) = \min_{a_0} e^{\theta a_0^2} + E(J_1) \longrightarrow (4)$$

$$E(J_2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\omega^2/2\sigma^2} e^{\theta((1-\alpha)x_1 + \alpha a_1 + (\omega - T))^2} \cdot d\omega$$

using (1)

$$E(J_2) = \frac{1}{\sqrt{1 - 2\sigma^2\theta}} e^{\left(\frac{a^2\theta}{1 - 2\sigma^2\theta}\right)}$$ here $a = (1-\alpha)x_1 + \alpha a_1 - T \longrightarrow (5)$

$$J_1(x_1) = \min_{a_1} e^{\theta a_1^2} E(J_2)$$

$$J_1(x_1) = \min_{a} \frac{e^{\theta a_1^2} \cdot e^{\frac{\theta a^2}{1 - 2\sigma^2\theta}}}{\sqrt{1 - 2\sigma^2\theta}}$$

$$\frac{d}{da_1} J_1(x_1) = 0 \Rightarrow$$

$$a_1^* = \frac{\theta\alpha\left(T-(1-\alpha)x_1\right)}{1+\theta\alpha^2-2\sigma^2\theta}$$

$$J_1^* x = \frac{e^{\theta(T-(1-\alpha)x_1)^2}}{\sqrt{1-2\sigma^2\theta}}$$

$$J_0(x_0) = \min_{a_0} e^{\theta a_0} \cdot E\left(J_1^*(x_1)\right)$$

$$\cancel{J_0(a_0)x_0}$$

$$J_1^*(x_1) = \frac{e^{\theta(T-(1-\alpha)x_1)^2}}{\sqrt{1-2\sigma^2\theta} \cdot \cancel{\quad}}$$

$$= e^{\theta(T-(1-\alpha)^2 x_0 - \alpha(1-\alpha)a_0 - (1-\alpha)\omega_0)}$$

$$E\left(J_1^*(x_1)\right) = \frac{1}{\sqrt{1-2\sigma^2\theta}\sqrt{1-2\sigma^2(1-\alpha)^2\theta}} e^{\frac{\theta b}{1-2\sigma^2(1-\alpha)^2\theta}}$$

here $b = T-(1-\alpha)^2 x_0 + \alpha(1-\alpha)a_0$

$$J_0(x_0) = \min e^{\theta a_0^2} \cdot E\left(J_1^*(x_0)\right)$$

$$J_0(x_0) = \min e^{\theta a_0^2} \frac{e^{\theta b}}{\sqrt{1-2\sigma^2\theta}\sqrt{1-2\sigma^2(1-\alpha^2)\theta}}$$

$$\frac{dJ_0(x_0)}{da_0} = 0$$

Scanned with CamScanner

$$a_0^* = \frac{\alpha(1-\alpha)\left(T-(1-\alpha^2)x_0\right)}{1+\alpha^2(1-\alpha^2)-2\sigma^2(1-\alpha^2)\theta}$$

$$J_0^*(x_0) = \frac{e^{\theta\left(T-(1-\alpha)^2 x_0\right)^2}}{\sqrt{(1-2\sigma^2\theta)(1-2\sigma^2(1-\alpha)^2\theta}}$$

$$J^*_{a_0,a_1}(x_0) = J_0^*(x_0)$$

$$\text{optimal expected cost} = \frac{e^{\theta\left(T-(1-\alpha)^2 x_0\right)^2}}{\sqrt{1-2\alpha^2\theta}\,\sqrt{1-2\sigma^2(1-\alpha)^2\theta}}$$

optimal policy :

Parameters Value

$$a_0^* = \frac{\alpha(1-\alpha)\left(T-(1-\alpha)^2 x_0\right)}{1+\alpha^2(1-\alpha)^2-2\sigma^2(1-\alpha)^2\theta}$$

$$a_1^* = \frac{\theta\alpha\left(T-(1-\alpha)x_0\right)}{1+\theta\alpha^2-2\sigma^2\theta}$$