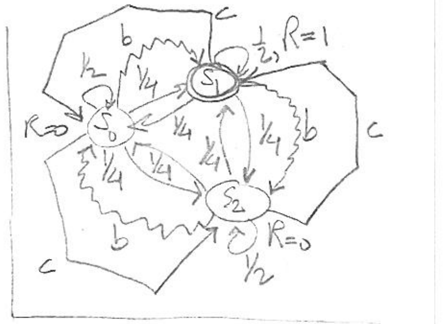
Q4 F23

$$P_a=\begin{pmatrix}1/2&1/4&1/4\\1/4&1/2&1/4\\1/4&1/4&1/2\end{pmatrix},\ P_b=\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix},\ P_c=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix},\ R=\begin{pmatrix}0\\1\\0\end{pmatrix},\ \gamma=\tfrac12$$



1) Policy $\pi_a=\begin{pmatrix}a\\a\\a\end{pmatrix}$: $V_{\pi_a}=\begin{pmatrix}0\\1\\0\end{pmatrix}+\frac12\begin{pmatrix}1/2&1/4&1/4\\1/4&1/2&1/4\\1/4&1/4&1/2\end{pmatrix}V_{\pi_a}\Rightarrow V_{\pi_a}=\begin{pmatrix}3/4&-1/8&-1/8\\-1/8&3/4&-1/8\\-1/8&-1/8&3/4\end{pmatrix}^{-1}\begin{pmatrix}0\\1\\0\end{pmatrix}$

$$\begin{vmatrix}3/4&-1/8&-1/8\\-1/8&3/4&-1/8\\-1/8&-1/8&3/4\end{vmatrix}=\frac{1}{8^3}\begin{vmatrix}6&-1&-1\\-1&6&-1\\-1&-1&6\end{vmatrix}=\frac{1}{8^3}\left(6\begin{vmatrix}6&-1\\-1&6\end{vmatrix}+\begin{vmatrix}-1&-1\\-1&6\end{vmatrix}-\begin{vmatrix}-1&6\\-1&-1\end{vmatrix}\right)=\frac{1}{8^3}\Big(\underbrace{6\times35}_{210}+\underbrace{(-7)-(7)}_{-14}\Big)=\frac{196}{8^3}$$

$$x_1=\begin{vmatrix}0&-1/8&-1/8\\1&3/4&-1/8\\0&-1/8&3/4\end{vmatrix}=\frac{1}{8^3}\begin{vmatrix}0&-1&-1\\8&6&-1\\0&-1&6\end{vmatrix}=\frac{1}{8^3}\left(-8\begin{vmatrix}-1&-1\\-1&6\end{vmatrix}\right)=-\frac{1}{8^2}(-7)=\frac{7}{8^2}$$

$$x_2=\begin{vmatrix}3/4&0&-1/8\\-1/8&1&-1/8\\-1/8&0&3/4\end{vmatrix}=\frac{1}{8^3}\begin{vmatrix}6&0&-1\\-1&8&-1\\-1&0&6\end{vmatrix}=\frac{1}{8^3}\left(8\begin{vmatrix}6&-1\\-1&6\end{vmatrix}\right)=\frac{1}{8^2}(35)$$

$$x_3=\begin{vmatrix}3/4&-1/8&0\\-1/8&3/4&1\\-1/8&-1/8&0\end{vmatrix}=\frac{1}{8^3}\begin{vmatrix}6&-1&0\\-1&6&8\\-1&-1&0\end{vmatrix}=\frac{(-1)}{8^2}\begin{vmatrix}6&-1\\-1&-1\end{vmatrix}=+\frac{7}{8^2}\Rightarrow V_{\pi_a}=\frac{1}{196}\begin{pmatrix}56\\280\\56\end{pmatrix}=\begin{pmatrix}0.286\\1.429\\0.286\end{pmatrix}$$

2) Apply Φ to V_{π_a}: $V'_{\pi_a}=\begin{pmatrix}0\\1\\0\end{pmatrix}+\frac12$

$$\left\{\begin{array}{l}
s_0:\ \max\begin{cases}a\ (\tfrac12\ \tfrac14\ \tfrac14)\times\\ b\ (0\ 1\ 0)\times\\ c\ (0\ 0\ 1)\times\end{cases}\frac{1}{196}\begin{pmatrix}56\\280\\56\end{pmatrix}=\frac{1}{196}\begin{cases}28+70+14\\280\\56\end{cases}\\[2ex]
s_1:\ \max\begin{cases}a\ (\tfrac14\ \tfrac12\ \tfrac14)\\ b\ (0\ 0\ 1)\\ c\ (1\ 0\ 0)\end{cases}\times\frac{1}{196}\begin{pmatrix}56\\280\\56\end{pmatrix}=\frac{1}{196}\begin{cases}\overbrace{14+140+14}^{168}\\56\\56\end{cases}\\[2ex]
s_2:\ \max\begin{cases}a\ (\tfrac14\ \tfrac14\ \tfrac12)\\ b\ (1\ 0\ 0)\\ c\ (0\ 1\ 0)\end{cases}\times\frac{1}{196}\begin{pmatrix}56\\280\\56\end{pmatrix}=\frac{1}{196}\begin{cases}14+70+28\\56\\280\end{cases}
\end{array}\right.$$

$$=\frac{1}{196}\begin{pmatrix}140\\196+84\\140\end{pmatrix}=\frac{1}{196}\begin{pmatrix}140\\280\\140\end{pmatrix}$$

[TRUE]

$\therefore$ π_a is not optimal

as $\underbrace{V_{\pi_a}}\le\underbrace{\Phi(V_{\pi_a})}$

$\frac{1}{196}\begin{pmatrix}56\\280\\56\end{pmatrix}$ $\frac{1}{196}\begin{pmatrix}140\\280\\140\end{pmatrix}$

3) Policy $\pi_b=\begin{pmatrix}b\\a\\c\end{pmatrix}$: $V_{\pi_b}=\begin{pmatrix}0\\1\\0\end{pmatrix}+\frac12\begin{pmatrix}0&1&0\\1/4&1/2&1/4\\0&1&0\end{pmatrix}V_{\pi_b}\Rightarrow V_{\pi_b}=\begin{pmatrix}1&-1/2&0\\-1/8&3/4&-1/8\\0&-1/2&1\end{pmatrix}^{-1}\begin{pmatrix}0\\1\\0\end{pmatrix}$

$\therefore V_{\pi_b}=\frac85\begin{pmatrix}1/2\\1\\1/2\end{pmatrix}=\begin{pmatrix}4/5\\8/5\\4/5\end{pmatrix}$

$$\begin{vmatrix}1&-1/2&0\\-1/8&3/4&-1/8\\0&-1/2&1\end{vmatrix}=\frac{1}{8^3}\begin{vmatrix}8&-4&0\\-1&6&-1\\0&-4&8\end{vmatrix}=\frac{1}{8^3}\left(8\begin{vmatrix}6&-1\\-4&8\end{vmatrix}+\begin{vmatrix}-4&0\\-4&8\end{vmatrix}\right)=\frac{1}{8^3}(8\times44+(-32))=\frac{1}{8^2}(44-4)=\frac{40}{8^2}$$

4) [TRUE] π_b beats π_a

$\begin{pmatrix}4/5\\8/5\\4/5\end{pmatrix}\ge\frac{1}{196}\begin{pmatrix}56\\280\\56\end{pmatrix}$

$$x_1=\begin{vmatrix}0&-1/2&0\\1&3/4&-1/8\\0&-1/2&1\end{vmatrix}=-\begin{vmatrix}-1/2&0\\-1/2&1\end{vmatrix}=\frac12,\ x_2=\begin{vmatrix}1&0&0\\-1/8&1&-1/8\\0&0&1\end{vmatrix}=1,\ x_3=\begin{vmatrix}1&-1/2&0\\-1/8&3/4&1\\0&-1/2&0\end{vmatrix}=-\begin{vmatrix}1&-1/2\\0&-1/2\end{vmatrix}=\frac12$$

5) Apply Φ to V_{π_b}: $V'_{\pi_b}=\begin{pmatrix}0\\1\\0\end{pmatrix}+\frac12$

$$\left\{\begin{array}{l}
s_0:\ \max\begin{cases}a\ (\tfrac12\ \tfrac14\ \tfrac14)\\ b\ (0\ 1\ 0)\\ c\ (0\ 0\ 1)\end{cases}\times\frac15\begin{pmatrix}4\\8\\4\end{pmatrix}=\frac15\begin{cases}2+2+1=5\\8\\4\end{cases}\\[2ex]
s_1:\ \max\begin{cases}a\ (\tfrac14\ \tfrac12\ \tfrac14)\\ b\ (0\ 0\ 1)\\ c\ (1\ 0\ 0)\end{cases}\times\frac15\begin{pmatrix}4\\8\\4\end{pmatrix}=\frac15\begin{cases}1+4+1=6\\4\\4\end{cases}\\[2ex]
s_2:\ \max\begin{cases}a\ (\tfrac14\ \tfrac14\ \tfrac12)\\ b\ (1\ 0\ 0)\\ c\ (0\ 1\ 0)\end{cases}\times\frac15\begin{pmatrix}4\\8\\4\end{pmatrix}=\frac15\begin{cases}1+2+2=5\\4\\8\end{cases}
\end{array}\right.$$

$$=\begin{pmatrix}0\\1\\0\end{pmatrix}+\frac15\begin{pmatrix}4\\3\\4\end{pmatrix}=\frac15\begin{pmatrix}4\\8\\4\end{pmatrix}$$

[TRUE]

$\therefore$ π_b is optimal as

$\Phi(V_{\pi_b})=V_{\pi_b}$