1 Question 1

Definition 1 MLRP: \( \frac{f(x)}{g(x)} \) is increasing in \( x \).

Definition 2 FOSD: \( F(x) \leq G(x) \forall x \)

Show that MLRP \( \Rightarrow \) FOSD

Proof: MLRP implies that there exists \( x_0 \) such that \( \frac{f(x_0)}{g(x_0)} = 1 \). If not then \( f(x) > g(x) \forall x \) or \( f(x) < g(x) \forall x \). But since \( \int f(x)dx = \int g(x)dx = 1 \), neither of these cases are possible.

We have two cases:

C1: \( x < x_0 : f(x) \leq g(x) \rightarrow \int_{-\infty}^{x} f(x) \leq \int_{-\infty}^{x} g(x) \rightarrow F(x) \leq G(x) \)

C2: \( x > x_0 : f(x) > g(x) \rightarrow \int_{x}^{\infty} f(x) \geq \int_{x}^{\infty} g(x) \rightarrow 1 - F(x) \leq 1 - G(x) \rightarrow F(x) \leq G(x) \)

2 Question 2

You want to design an experiment, that is a random variable \( Y \) (that takes value in \([0;1]\), for simplicity, and is characterized by the joint distribution \( p(\theta, y) \)) such that the distribution of posteriors generated by this experiment is given by \( f(p) \). In the “experiment” the posterior will be given by \( Pr(\theta|y) \) and the probability that this posterior arises is simply \( Pr(y) \). In the statement, the probability that posterior \( p \) arises would be given by \( f(p) \). Therefore, we’d like to take \( Pr(\theta|y)=p \) and \( Pr(y)=f(p) \) (for \( y=p \)) which directly defines \( p(\theta_1,y)=pf(p) ; Pr(\theta_2|y)=1-p ; p(\theta_2,y)=(1-p)f(p) \). Does such a random variable exists?

We have \( p(\theta,y) \geq 0 \) for all \( y,\theta \) and

\[
\int p(\theta,y)dyd\theta = \int p(\theta_1,y)dy + \int p(\theta_2,y)dy = \int pf(p)dp + \int (1-p)f(p)dp \quad \text{(by construction, since } p(\theta_1,y)=pf(p) \text{ for } y=p \})
\]

\[
= p_0 + (1-p_0) = 1 \quad \text{(by hypothesis)}
\]

Therefore such an experiment exists (we can construct a random variable \( Y \) such that the joint distribution of \((Y, \theta)\) is \( p(\theta,y) \) since \( p \) is non negative and sums up to 1) and the prior is given by \( Pr(\theta_1)=\int p(\theta_1,y)dy = p_0 \) as wanted.

We can then define the likelihood functions using Bayes rule and we have \( Pr(y|\theta_1)=Pr(\theta_1/y)Pr(y)/Pr(\theta_1)=pf(p)/p_0 \) and similarly for \( Pr(y|\theta_2) \). You can then directly verify that the experiment defined by the outcome \( y \), these likelihood functions and the prior \( p_0 \) generate posteriors distributed according to \( f \).
3 Question 3

1. We are given \( u(a, s) \) therefore we can derive \( v(a, p) \):

\[
\begin{align*}
v(a_1, p) &= pu(a_1, s_1) + (1 - p) u(a_1, s_2) = 7 + 3p \\
v(a_2, p) &= pu(a_2, s_1) + (1 - p) u(a_2, s_2) = 11 - 6p
\end{align*}
\]

The upper envelope of \( v(a, p) \) will be the \( V(p) \) (it indicates the maximum expected utility that the agent can reach if faced with probability of \( s_1 \) equal to \( p \)):

\[
V(p) = \max_a v(a, p)
\]
At $p = 0.4$ the optimal decision is $a_2$ since:

\[
v(a_1, p) = 10(0.4) + 7(0.6) = 8.2 \\
v(a_2, p) = 5(0.4) + 11(0.6) = 8.6
\]

2. We are given the likelihood matrix:

\[
L = \begin{bmatrix}
\Pr(y_1|s_1) & \Pr(y_2|s_1) \\
\Pr(y_1|s_2) & \Pr(y_2|s_2)
\end{bmatrix} = \begin{bmatrix}
\lambda_1 (1 - \lambda_1) \\
\lambda_2 (1 - \lambda_2)
\end{bmatrix}
\]

The probability of observing the two signals is:

\[
\Pr(y_1) = \Pr(y_1|s_1)\Pr(s_1) + \Pr(y_1|s_2)\Pr(s_2) = 0.4\lambda_1 + 0.6\lambda_2 \\
\Pr(y_2) = \Pr(y_2|s_1)\Pr(s_1) + \Pr(y_2|s_2)\Pr(s_2) = 1 - 0.4\lambda_1 - 0.6\lambda_2
\]

Let’s calculate the posterior probabilities:

\[
\Pr(s_1|y_1) = \frac{\Pr(y_1|s_1)\Pr(s_1)}{\Pr(y_1)} = \frac{0.4\lambda_1}{0.4\lambda_1 + 0.6\lambda_2} \\
\Pr(s_1|y_2) = \frac{\Pr(y_2|s_1)\Pr(s_1)}{\Pr(y_2)} = \frac{0.4 (1 - \lambda_1)}{1 - 0.4\lambda_1 - 0.6\lambda_2}
\]

- If $\lambda_1 = \lambda_2 = \frac{1}{2}$ the information system has no value because the same signal $y_1$ is as likely to appear in the two states of nature. Whenever $\lambda_1 = \lambda_2$ the information system has no value.

- If $\lambda_1 = \frac{1}{2}$ and $\lambda_2 = 0$ the the likelihood matrix is the following:

\[
L = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} \\
0 & 1
\end{bmatrix}
\]

Posterior in this case are:

\[
\Pr(s_1|y_1) = 1 \\
\Pr(s_1|y_2) = 0.25
\]
And probabilities of the two signals are:

\[
\begin{align*}
\Pr(y_1) &= 0.2 \\
\Pr(y_2) &= 0.8
\end{align*}
\]

— When observe \( y_1 \):

\[
\begin{align*}
v(a_1, 1) &= 10 \\
v(a_2, 1) &= 5
\end{align*}
\]

therefore the optimal choice as a function of the signal is:

\[
a(y_1) = \arg\max_a v(a, p = 1) = a_1
\]

Hence:

\[
V(1) = 10
\]

— When observe \( y_2 \):

\[
\begin{align*}
v(a_1, 0.25) &= 7.75 \\
v(a_2, 0.25) &= 9.5
\end{align*}
\]

therefore the optimal choice as a function of the signal is:

\[
a(y_2) = \arg\max_a v(a, p = 0.25) = a_2
\]

hence:

\[
V(0.25) = 9.5
\]

We can know calculate \( V_Y \) as:

\[
V_Y = (9.5) 0.8 + (10) 0.2 = 9.6
\]
We can now calculate the value of information as:

\[ Z = V_Y - V(0.4) = 9.6 - 8.6 = 1 \]

3. We know that the Blackwell theorem gives general conditions under which one information system is preferred to another. So we just have to prove that the information system \((\lambda_1 = \frac{1}{2}\alpha + \frac{1}{2}\beta, \lambda_2 = \beta)\) is a garbling of the information system \((\lambda_1 = \frac{1}{2}, \lambda_2 = 0)\).

We have to find a Markov matrix \(M\):

\[
M = \begin{bmatrix}
m_{11} & m_{12} \\
m_{21} & m_{22}
\end{bmatrix}
\]

\[
m_{11} + m_{12} = 1
\]

\[
m_{21} + m_{22} = 1
\]

\[
m_{ij} \geq 0
\]

such that the following relationship between the likelihood matrices holds:

\[
\begin{bmatrix}
\frac{1}{2} & \frac{1}{2} \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
m_{11} & m_{12} \\
m_{21} & m_{22}
\end{bmatrix} =
\begin{bmatrix}
\frac{\alpha + \beta}{2} & 1 - \frac{\alpha + \beta}{2} \\
\beta & 1 - \beta
\end{bmatrix}
\]

You can verify that such matrix \(M\) exists and is equal to the following:

\[
M = \begin{bmatrix}
\alpha & 1 - \alpha \\
\beta & 1 - \beta
\end{bmatrix}
\]
4 Question 4

4.1 a)

The (Pareto) problem is:

$$\max_{e,\omega} f_e u (s_2) + (1 - f_e) u (s_1) - e$$

subject to:

$$IR) f_e (x_2 - s_2) + (1 - f_e) (x_1 - s_1) \geq 0$$

The (individual rationality) constraint (for the principal will bind) (otherwise, can just pay the agent more). Since we only have the Pareto problem here, the agent will get a constant wage (in effect, the principal is like a competitive insurance company).

$$s_1 = s_2 = f_e x_2 + (1 - f_e) x_1$$

The (effort) level should be chosen to solve:

$$\max_{e,\omega} \{ u (f_e x_2 + (1 - f_e) x_1) - e \}$$

Effort level (e_H) is (optimal) iff:

$$u (f_H x_2 + (1 - f_H) x_1) - u (f_L x_2 + (1 - f_L) x_1) \geq \epsilon H - \epsilon L$$

The (second-best) problem is the same, but also adds an incentive constraint (for implementing e_H)

$$IC) (f_H e - f_L e) (u (s_2) - u (s_1)) \geq \epsilon H - \epsilon L$$

Then, it is clear that a constant wage cannot implement e_H (since it will set the LHS to zero). Hence, the IC constraint will bind if e_H is optimal to implement. To implement e_H, we will have:

$$f_H e - f_L e = \epsilon H - \epsilon L \quad (2)$$

To implement e_L, we will have a constant wage:

$$s = f_L x_2 + (1 - f_L) x_1$$

It is efficient to implement e_H iff:

$$f_H u (s_2) + (1 - f_H) u (s_1) - e H \geq u (s) - e L$$

and (using (2),

$$\Rightarrow u (s_1) + \frac{f_L}{f_H} (e H - e L) \geq u (s)$$

and (using (2),

$$\Rightarrow u (s_1) + \frac{f_L}{f_H} (e H - e L) \geq u (s)$$
4.2 b) 
Suppose there is now a (third) effort level \(e_{Mo} > \frac{e_H}{2}(e_{Ma} + e_L)\) with \(f_{Ma} = \frac{1}{2}(f_{Ha} + f_{La})\). Then, to implement \(e_{Mo}\) we need the following incentive constraints:

\[
\text{ICH}: (f_{Ha} - f_{Ma})(u(s_2) - u(s_1)) \leq e_{Ha} - e_M \\
\text{ICL}: (f_{Ma} - f_{La})(u(s_2) - u(s_1)) \geq e_{Ma} - e_L
\]

Substituting in for \(f_{Ma}\) and \(e_{Ma}\), we get:

\[
\text{ICH}(\frac{1}{2}(f_{Ha} - f_{La})(u(s_2) - u(s_1)) < \alpha \frac{1}{2}(e_{Ha} - e_L) \\
\text{ICL}(\frac{1}{2}(f_{Ma} - f_{La})(u(s_2) - u(s_1)) > \alpha \frac{1}{2}(e_{Ma} - e_L)
\]

which cannot both hold simultaneously. (Therefore, \(e_{Mo}\) cannot be implemented.)

4.3 c) 
Now suppose \(e_{Mo} < \frac{e_H}{2}(e_{Ma} + e_L)\) and \(e_{Ha}\) was optimal to implement (in Part a). Hence, \(s_1^* \geq f_L x_2 + (1 - f_L) x_1\)

where:

\[
u(s_2) - u(s_1^*) = \frac{e_H - e_L}{f_H} - \frac{e_L}{f_L}
\]

and \((s_1^*, s_2^*)\) denote the optimal contract (in Part a). (Our constraints are now:

\[
\text{IR}(f_{Ha} x_2 (s_2) + (1 - f_{Ha})(x_1(s_1)) \geq 0 \\
\text{IC}(f_{Ha} - f_{La})(u(s_2) - u(s_1)) \geq e_{Ha} - e_L \\
\text{ICM}(\frac{1}{2}(f_{Ha} - f_{Ma})(u(s_2) - u(s_1)) \geq e_{Ha} - e_M
\]

Since we have that

\[
\frac{1}{2}(f_{Ha} - f_{La})(u(s_2) - u(s_1)) = \frac{1}{2}(e_{Ha} - e_L) < e_{Ha} - e_M
\]

constraint (ICM) fails the contract (we derived in Part a) no longer implements \(e_{Ha}\) once \(e_{Ma}\) is available. We can still implement \(e_{Ha}\) under certain conditions. (Constraint (ICM) will bind, and therefore, constraint IC) will not bind. (The sharing rule will satisfy

\[
f_{Ha}(x_2(s_2) + (1 - f_{Ha})(x_1(s_1)) = 0 \\
\frac{1}{2}(f_{Ha} - f_{La})(u(s_2) - u(s_1)) = e_{Ha} - e_M
\]