Assignment 4 Student ID: u7205329

Question 1.

Part a,

Verify that $p(\theta) = 6\theta(1-\theta)$ is a valid pdf on $[0,1]$.

$$\int_0^1 p(\theta)\, d\theta = \int_0^1 6\theta(1-\theta)\, d\theta = \int_0^1 (6\theta - 6\theta^2)\, d\theta$$

$$= \left[6.\frac{\theta^2}{2} - 6.\frac{\theta^3}{3} \right]_0^1 = 1$$

We also have $p(\theta)$ is non-negative for $\theta \in [0,1]$

Indeed, let $\theta \in (0,1)$

$$\Rightarrow p(\theta) = \underbrace{6\theta}_{\text{positive}} \underbrace{(1-\theta)}_{\text{positive}} \quad \text{and so } p(\theta) > 0$$

When $\theta = 0 \Rightarrow p(0) = 6.0(1-0) = 0$, and

$\theta = 1 \Rightarrow p(1) = 6.1(1-1) = 0$.

Part b to e,

The following parts are the process of updating beliefs on the distribution of θ based on each coin flip.
Note that μ, σ^2 & θ_{MAP} are calculated after each new pdf is derived.

- Before flipping: $p(\theta) = 6\theta(1-\theta)$

$\Rightarrow$ Expected value:

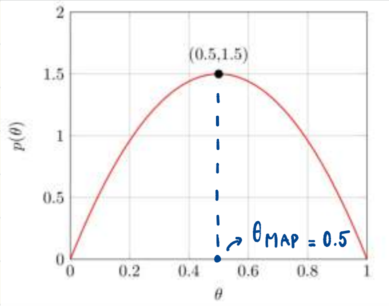
$$E[\theta] = \int_0^1 \theta \, p(\theta) \, d\theta = \int_0^1 \theta . 6\theta(1-\theta) \, d\theta = \frac{1}{2}$$

Variance:

$$Var[\theta] = E[X^2] - (E[X])^2 = \left(\int_0^1 \theta^2 p(\theta) \, d\theta \right) - \left(\frac{1}{2}\right)^2$$

$$= \left(\int_0^1 \theta^2 \, 6\theta(1-\theta) \, d\theta \right) - \left(\frac{1}{2}\right)^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

From the plot given, the θ_{MAP} which maximises $p(\theta)$ is 0.5.



- After the first flip, we have $x_1 = 0$. Using Bayes' theorem, we have

$$p(\theta \mid x_1 = 0) = \frac{p(x_1 = 0 \mid \theta) \cdot p(\theta)}{p(x_1 = 0)}$$

We have the function of observing $x_1 = 0$ given a particular value of θ
$p(x_1 = 0 \mid \theta) = 1 - \theta$ and the prior $p(\theta) = 6\theta(1-\theta)$.

To find $p(x_1 = 0)$, we need to sum the distribution function of observing $x_1 = 0$ when $\theta = \theta_i$ for all values of $\theta \in [0, 1]$.

$$\Rightarrow p(x_1 = 0) = \int_0^1 p(x_1 = 0 \text{ \& } \theta = \theta_i)\, d\theta$$

$$= \int_0^1 p(x_1 = 0 \mid \theta = \theta_i) \cdot p(\theta = \theta_i)\, d\theta$$

$$= \int_0^1 (1-\theta)\, 6\theta(1-\theta)\, d\theta = \frac{1}{2}$$

$$\Rightarrow p(\theta \mid x_1 = 0) = \frac{p(x_1 = 0 \mid \theta) \cdot p(\theta)}{p(x_1 = 0)} = \frac{(1-\theta) \cdot 6\theta(1-\theta)}{1/2} = 12\theta(1-\theta)^2$$

This density function is valid since

$$\int_0^1 p(\theta \mid x_1 = 0)\, d\theta = \int_0^1 12\theta(1-\theta)^2\, d\theta = 1$$ and $p(\theta \mid x_1 = 0) \geq 0$ for $\theta \in [0,1]$.

$\Rightarrow$ Expected value :

$$E[\theta \mid x_1 = 0] = \int_0^1 \theta\, p(\theta \mid x_1 = 0)\, d\theta = \int_0^1 12\theta^2(1-\theta)^2\, d\theta = \frac{2}{5} = 0.4.$$

Variance:

$$Var[\theta \mid x_1 = 0] = E[\theta^2 \mid x_1 = 0] - (E[\theta \mid x_1 = 0])^2$$

$$= \left(\int_0^1 \theta^2 . \, p(\theta \mid x_1 = 0) \, d\theta \right) - \left(\frac{2}{5} \right)^2$$

$$= \frac{1}{5} - \left(\frac{2}{5} \right)^2 = \frac{1}{25} = 0.04 .$$

θ_{MAP} :

We have $\frac{dp(\theta \mid x_1 = 0)}{d\theta} = \frac{d}{d\theta} \left[12\theta (1-\theta)^2 \right]$

$$= 12 \frac{d}{d\theta} \left[\theta^3 - 2\theta^2 + \theta \right]$$

$$= 12 . \left(3\theta^2 - 4\theta + \theta \right)$$

Setting $\frac{dp(\theta \mid x_1 = 0)}{d\theta} = 0 \iff 12 \left(3\theta^2 - 4\theta + \theta \right) = 0$

$$\iff (3\theta - 1)(\theta - 1) = 0$$

$$\iff \left[\begin{array}{l} \theta = 1 \\ \theta = \frac{1}{3} \end{array} \right.$$

Since $\theta = \frac{1}{3}$ maximises $p(\theta \mid x_1 = 0) \Rightarrow \theta_{MAP} = \frac{1}{3}$.

- After second flip : $x_2 = 0$. Again, using Bayes' theorem

$$p(\theta \mid x_{1:2} = 00) = \frac{p(x_{1:2} = 00 \mid \theta) . \, p(\theta)}{p(x_{1:2} = 00)}$$

As we are updating the probability distribution based on the previous flip, we can interpret $p(x_{1:2} = 00 \mid \theta)$ as the probability of observing $x_2 = 0$, given a particular value of θ and that the first flip was 0. As the coin flips are i.i.d, $p(x_{1:2} = 00 \mid \theta) = 1 - \theta$.

We also have $p(\theta) = p(\theta \mid x_1 = 0) = 12\theta (1-\theta)^2$.

To find $p(x_{1:2} = 00)$ is the same as finding the probability that $x_2 = 0$ with $\theta = \theta_i$ given $x_1 = 0$ for all values of $\theta \in [0,1]$.

$$p(x_{1:2} = 00) = \int_0^1 p(x_2 = 0 \;\&\; (\theta = \theta_i \mid x_1 = 0))\, d\theta$$

$$= \int_0^1 p(x_2 = 0 \mid (\theta = \theta_i \mid x_1 = 0)) \,.\, p(\theta = \theta_i \mid x_1 = 0)\, d\theta$$

$$= \int_0^1 p(x_{1:2} = 00 \mid \theta) \,.\, p(\theta \mid x_1 = 0)\, d\theta$$

$$= \int_0^1 (1-\theta) \,.\, 12\theta(1-\theta)^2\, d\theta = \frac{3}{5}.$$

Finally, $p(\theta \mid x_{1:2} = 00) = \dfrac{p(x_{1:2} = 00 \mid \theta) \,.\, p(\theta)}{p(x_{1:2} = 00)}$

$$= \frac{(1-\theta) \,.\, 12\theta(1-\theta)^2}{3/5}$$

$$= 20\,\theta(1-\theta)^3.$$

Again, this density function is valid as $\int_0^1 20\,\theta(1-\theta)^3 = 1$. and $p(\theta \mid x_{1:2} = 00) \geqslant 0$ for $\theta \in [0,1]$

Expected value:

$$E[X \mid \theta_{1:2} = 00] = \int_0^1 \theta \,.\, 20\,\theta(1-\theta)^3\, d\theta = \frac{1}{3}.$$

Variance:

$$Var[X \mid \theta_{1:2} = 00] = E[X^2 \mid \theta_{1:2} = 00] - (E[X \mid \theta_{1:2} = 00])^2$$

$$= \int_0^1 \theta^2 \,.\, 20\,\theta(1-\theta)^3\, d\theta - \left(\frac{1}{3}\right)^2$$

$$= \frac{1}{7} - \left(\frac{1}{3}\right)^2 = \frac{2}{63}$$

θ_{MAP}

We have $\dfrac{dp(\theta \mid x_{1:2} = 00)}{d\theta} = \dfrac{d}{d\theta}\left[20\theta(1-\theta)^3 \right]$

$$= \frac{d}{d\theta}\left[20\theta - 60\theta^2 + 60\theta^3 - 20\theta^4 \right]$$

$$= 20 - 120\theta + 180\theta^2 - 80\theta^3$$

Setting $\dfrac{dp(\theta \mid x_{1:2} = 00)}{d\theta} = 0 \Leftrightarrow 20 - 120\theta + 180\theta^2 - 80\theta^3 = 0$

$$\Leftrightarrow 1 - 6\theta + 9\theta^2 - 4\theta^3 = 0$$

$$\Leftrightarrow -(a-1)^2(4a-1) = 0$$

$$\Leftrightarrow \left[\begin{array}{l} a - 1 = 0 \\ 4a - 1 = 0 \end{array}\right. \Leftrightarrow \left[\begin{array}{l} a = 1 \\ a = 1/4 \end{array}\right.$$

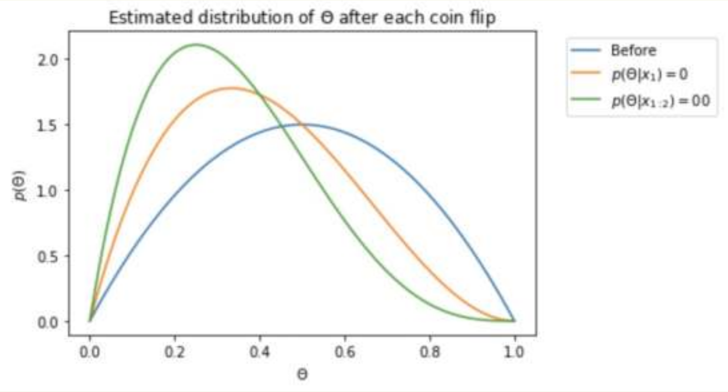
Since $\theta = \dfrac{1}{4}$ maximises $p(\theta \mid x_{1:2} = 00)$ and thus $\theta_{MAP} = \dfrac{1}{4}$.

We can present our table as follows :

Posterior	PDF	μ	σ^2	θ_{MAP}
$p(\theta)$	$6\theta(1-\theta)$	$1/2$	$1/20$	$1/2$
$p(\theta \mid x_1 = 0)$	$12\theta(1-\theta)^2$	$2/5$	$1/25$	$1/3$
$p(\theta \mid x_{1:2} = 00)$	$20\theta(1-\theta)^3$	$1/3$	$2/63$	$1/4$

Part f.

Plot each of the probability distribution



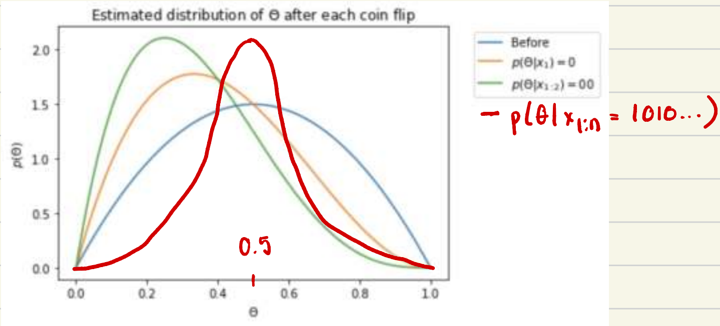
We update the probability distribution of θ after each coin flips.

Every time a new coin flip turns out to be zero, our updated beliefs of the probability distribution of θ becomes more right-skewed, with θ_{MAP} shifts closer to zero (to the left) every time the coin flip is zero.

As both of the observed coin flips are zero, our final belief so far (the green graph) is that the coin is biased towards zero.
This belief is subjected to change with any new coin flips.

Part g.

If we update the posterior distribution $p(\theta \mid x_{1:n})$ for an alternating sequence, we expect that the distribution will slowly converge to a normal distribution with the expected value $E[\theta]$ and θ_{MAP} converge gradually to 0.5. We also expect the variance σ^2 to become smaller and smaller and converge to 0 as $n \to \infty$.



Question 2.

Part a.

For $\theta = 0$: we expect that the camera always reads "1" incorrectly, i.e.

- $x = 1 \rightarrow \hat{x} = 0$
- $x = 0 \rightarrow \hat{x} = 0$

The evidence from this camera is unreliable as it always returns 0, that is, we do not know the true value of the coin flips from this camera. Therefore, the agent does not have any new information and thus cannot update the estimated distribution of θ. The agent's belief given any $\hat{x}$ shall remain the same as the prior distribution.

For $\theta = 0.5$: we expect the camera to read "1" correctly 50% of the time on average.

- if $x = 0 \rightarrow \hat{x} = 0$
- if $x = 1 \rightarrow \left[\begin{array}{ll} \hat{x} = 0 & \text{50\% of the time} \\ \hat{x} = 1 & \text{50\% of the time} \end{array}\right.$

The evidence from this camera is only accurate when $\hat{x} = 1$. Thus, the agent can only update the estimated distribution of θ when $\hat{x} = 1$. When $\hat{x} = 0$, we cannot possibly know if $x = 0$ or $x = 1$.

For $\theta = 1$: we expect that the camera always reads "1" correctly.

- if $x = 1 \rightarrow \hat{x} = 1$
- if $x = 0 \rightarrow \hat{x} = 0$.

The evidence from this camera is reliable and the agent can update the estimated distribution given an observation $\hat{x}$ in the same manner as given the true value of the coin flip x.

Part b. Compute $p(\hat{X} = x \mid \theta, \phi)$ for all $x \in \{0, 1\}$

. For $\hat{x} = 0$, we have

$$
\begin{aligned}
P(\hat{X} = 0 \mid \theta, \phi) &= P(\hat{X} = 0 \,\&\, X = 0 \mid \theta, \phi) + P(\hat{X} = 0 \,\&\, X = 1 \mid \theta, \phi) \\
&= P(\hat{X} = 0 \mid X = 0, \phi) \cdot P(X = 0 \mid \theta) + P(\hat{X} = 0 \mid X = 1, \phi) \cdot P(X = 1 \mid \theta) \\
&= 1 \cdot (1 - \theta) + (1 - \phi) \cdot \theta \\
&= (1 - \theta) + (1 - \phi)\theta \\
&= 1 - \theta + \theta - \phi\theta = 1 - \phi\theta .
\end{aligned}
$$

. For $\hat{x} = 1$,

$$
\begin{aligned}
P(\hat{X} = 1 \mid \theta, \phi) &= P(\hat{X} = 1 \,\&\, X = 0 \mid \theta, \phi) + P(\hat{X} = 1 \,\&\, X = 1 \mid \theta, \phi) \\
&= P(\hat{X} = 1 \mid X = 0, \phi) \cdot P(X = 0 \mid \theta) + P(\hat{X} = 1 \mid X = 1, \phi) \cdot P(X = 1 \mid \theta) \\
&= 0 \cdot (1 - \theta) + \phi \cdot \theta \\
&= \phi \cdot \theta .
\end{aligned}
$$

Thus, $p(\hat{X} = x \mid \theta, \phi) = \begin{cases} \phi\theta & \text{for } x = 1 \\ 1 - \phi\theta & \text{for } x = 0 \end{cases}$.

Part c. Using Bayes' Theorem, we have

$$
p(\theta \mid \hat{x}_1 = 0) = \frac{p(\hat{x}_1 = 0 \mid \theta) \cdot p(\theta)}{p(\hat{x}_1 = 0)} = \frac{(1 - \phi\theta) \cdot 6\theta(1 - \theta)}{p(\hat{x}_1 = 0)}
$$

We have $p(\hat{x}_1 = 0)$ can be calculated by integrate $p(\hat{x}_1 = 0 \mid \theta, \phi)$ for all values of $\theta \in [0, 1]$.

$\Rightarrow p(\hat{x}_1 = 0)$

$$= \int_0^1 p(\hat{x}_1 = 0 \times \theta = \theta_i, \phi) \, d\theta$$

$$= \int_0^1 p(\hat{x}_1 = 0 \mid \theta, \phi) \cdot p(\theta) \, d\theta$$

$$= \int_0^1 (1 - \phi\theta) \cdot 6\theta(1-\theta) \, d\theta = 1 - \frac{1}{2}\phi$$

Thus, $p(\theta \mid \hat{x}_1 = 0) = \dfrac{p(\hat{x}_1 = 0 \mid \theta) \cdot p(\theta)}{p(\hat{x}_1 = 0)} = \boxed{\dfrac{(1 - \phi\theta) \cdot 6\theta(1-\theta)}{1 - \frac{1}{2}\phi}}$

- When $\phi = 1$, recall that this means that the camera always reads "1" correctly. i.e. $\begin{cases} x = 1 \Rightarrow \hat{x} = 1 \\ x = 0 \Rightarrow \hat{x} = 0 \end{cases}$.

$$\Rightarrow p(\theta \mid \hat{x}_1 = 0) = \frac{(1 - 1 \cdot \theta) \cdot 6\theta(1-\theta)}{1 - \frac{1}{2} \cdot 1} = \frac{6\theta(1-\theta)^2}{1/2}$$

$$= 12\theta(1-\theta)^2$$

$$= p(\theta \mid x_1 = 0)$$

This result is expected since the camera always reads correctly, i.e. if we know what the camera returns, we also know the true value of the coin flip $\Rightarrow p(\theta \mid \hat{x} = 0)$ describes the distribution of θ in the same manner as $p(\theta \mid x = 0)$.

- When $\phi = 0$, recall that this means the camera always reads "1" incorrectly.

$$\Rightarrow p(\theta \mid \hat{x}_1 = 0) = \frac{(1 - 0.\theta).\ 6\theta(1-\theta)}{1 - \frac{1}{2}.0} = 6\theta(1-\theta) = p(\theta).$$

This result is also expected since the camera always reads "1" incorrectly, i.e. camera always returns 0 and thus we cannot derive the true outcome x from $\hat{x} = 0$.

$\Rightarrow$ Results from the camera does not give any information on the true distribution of θ, thus $p(\theta \mid \hat{x} = 0)$ should be the same as if we have not flipped any coins at all.

Part d. Now the camera reports seeing a one, $\hat{x}_1 = 1$.

$$p(\theta \mid \hat{x}_1 = 1, \phi) = \frac{p(\hat{x}_1 = 1 \mid \theta, \phi) . p(\theta \mid \phi)}{p(\hat{x}_1 = 1)}$$

We have $p(\hat{x}_1 = 1)$

$$= \int_0^1 p(\hat{x}_1 = 1 \times \theta = \theta_i, \phi) \, d\theta$$

$$= \int_0^1 p(\hat{x}_1 = 1 \mid \theta, \phi) . p(\theta) \, d\theta$$

$$= \int_0^1 \phi\theta . 6\theta(1-\theta) \, d\theta = \frac{1}{2}\phi$$

$$\Rightarrow p(\theta \mid \hat{x}_1 = 1, \phi) = \frac{p(\hat{x}_1 = 1 \mid \theta, \phi) . p(\theta \mid \phi)}{p(\hat{x}_1 = 1)} = \frac{\phi\theta . 6\theta(1-\theta)}{\phi . \frac{1}{2}}$$

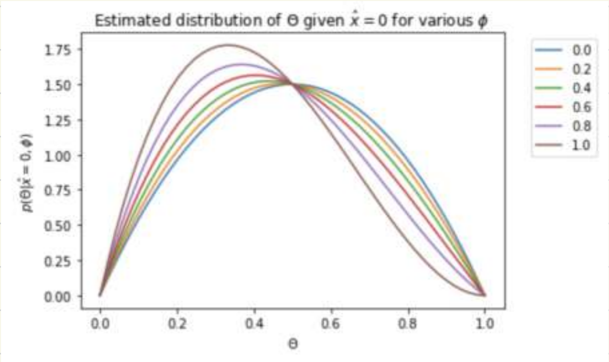
$$= 12\theta^2(1-\theta)$$

$\Rightarrow$ The result does not depend on ϕ anymore. This makes sense because:

If $\hat{x} = 1$, we know surely that $x = 1$ since under no circumstances where the camera would return $\hat{x} = 1$ when $x = 0$.

Thus $p(\theta \mid \hat{x} = 1)$ is the same as $p(\theta \mid x = 1)$ and that is why the bias of camera (ϕ) is no longer relevant here.

Part e. Plot $p(\theta \mid \hat{x}_1 = 0, \phi)$ as a function of θ



From the plot above, we see that with for lower values of ϕ, the distributions $p(\theta \mid \hat{x}_1 = 0, \phi)$ are similar to the prior distribution $6\theta(1-\theta)$, reflects the uncertainty.

For higher values, the distribution becomes more right-skewed with θ_{MAP} moves more and more to the left, closer to our updated belief of θ from the true outcome.

Question 3. We first obtain the CDF of X

$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{0}^{x} (2-2t)\, dt = \left[2t - 2 \cdot \frac{t^2}{2} \right]_0^x$$

$$= 2x - x^2$$

We have the CDF of Y

$$\begin{aligned} G_Y(y) &= P(Y < y) \\ &= P(X^2 + 1 < y) \\ &= P(X^2 < y - 1) \\ &= P(X < \sqrt{y-1}) \\ &= F_X(\sqrt{y-1}) \\ &= 2(\sqrt{y-1}) - (\sqrt{y-1})^2 \quad = 2\sqrt{y-1} - y + 1. \end{aligned}$$

Thus the pdf of Y can be found by differentiation

$$g_Y(y) = \frac{d}{dy} G_Y(y) = \frac{d}{dy}\left[2\sqrt{y-1} - y + 1 \right] = \frac{1}{\sqrt{y-1}} - 1$$

This pdf is valid as $\int_1^2 \left(\frac{1}{\sqrt{y-1}} - 1 \right) dy = 1.$