Yifan Luo u7351505

Question 1 Bayesian Inference

a) Non-negativity: We know that $p(\theta) = 30\theta^2 \cdot (1-\theta)^2$, where $\theta \in [0,1]$

It can be easily proved that $\theta^2 > 0$, let $B = 30\theta^2$, then $30\theta^2 = B \geq 0$

Let $A = 1-\theta$, $A \in [0,1]$, then $(1-\theta)^2 = A^2 \geq 0$.

Thus $p(\theta) = A \cdot B \geq 0$. i.e. $p(\theta) = 30\theta^2(1-\theta)^2 \geq 0$

Normalization: $p(\theta) = 30\theta^2(1-\theta)^2 = 30\theta^2(1-2\theta+\theta^2) = 30\theta^2 - 60\theta^3 + 30\theta^4$

Let $F'(\theta) = p(\theta)$, then $F(\theta) = 10\theta^3 - 15\theta^4 + 6\theta^5$

Then $\int_0^1 p(\theta)d\theta = F(1) - F(0) = 10 - 15 + 6 - 0 = 1$. i.e. $P(\theta)$ is normalized.

b) $p(X_{1:2}|\theta) = p(X_1=0, X_2=1|\theta) = p(X_1=0|\theta) \cdot p(X_2=1|\theta)$ (i.i.d)

$= (1-\theta)\cdot\theta$

Then, $p(X_{1:2}, \theta) = p(\theta) \cdot p(X_{1:2}|\theta)$

$= 30\theta^2(1-\theta)^2 \cdot (1-\theta)\cdot\theta$

$= 30\theta^3(1-\theta)^3$

Then, $p(X_{1:2}) = \int_\theta p(X_{1:2},\theta)d\theta = \int_0^1 p(X_{1:2},\theta)d\theta = \int_0^1 30\theta^3(1-\theta)^3 d\theta$

Let $F'(\theta) = 30\theta^3(1-\theta)^3 = 30\theta^3 - 90\theta^4 + 90\theta^5 - 30\theta^6$

Then $F(\theta) = \frac{15}{2}\theta^4 - 18\theta^5 + 15\theta^6 - \frac{30}{7}\theta^7$

Thus RHS $= F(1) - F(0) = \frac{15}{2} - 18 + 15 - \frac{30}{7} - 0 = \frac{3}{14}$

$P(X_{1:4}) = \int_\theta P(X_{1:4},\theta)d\theta = \int_0^1 P(X_{1:4},\theta)d\theta = \int_0^1 30\theta^4(1-\theta)^4 d\theta$

Let $F'(\theta) = 30\theta^4(1-\theta)^4 = 30\theta^4 - 120\theta^5 + 180\theta^6 - 120\theta^7 + 30\theta^8$

Then $F(\theta) = 6\theta^5 - 20\theta^6 + \frac{180}{7}\theta^7 - 15\theta^8 + \frac{10}{3}\theta^9$

Thus RHS $= F(1) - F(0) = 6 - 20 + \frac{180}{7} - 15 + \frac{10}{3} = \frac{1}{21}$

Then, $P(\theta|X_{1:2}) = \frac{P(\theta, X_{1:2})}{P(X_{1:2})} = \frac{30\theta^3(1-\theta)^3}{\frac{3}{14}} = 140\theta^3(1-\theta)^3$

$P(\theta|X_{1:4}) = \frac{P(\theta, X_{1:4})}{P(X_{1:4})} = \frac{30\theta^4(1-\theta)^4}{\frac{1}{21}} = 630\theta^4(1-\theta)^4$

c) Before any evidence: $\mu = E(\theta) = \int_0^1 \theta p(\theta)d\theta = \int_0^1 \theta\cdot 30\theta^2(1-\theta)^2 d\theta$

Let $F'(\theta) = \theta\cdot 30\theta^2(1-\theta)^2 = 30\theta^3 - 60\theta^4 + 30\theta^5$

Then $F(\theta) = \frac{15}{2}\theta^4 - 12\theta^5 + 5\theta^6$

Thus RHS $= F(1) - F(0) = \frac{15}{2} - 12 + 5 - 0 = \frac{1}{2}$

i.e. $E_\theta(\theta) = \frac{1}{2}$

After observing $X_{1:2}$: $E_{p(\theta|X_{1:2})}(\theta) = \int_0^1 \theta\, p(\theta|X_{1:2})\, d\theta = \int_0^1 \theta\, 140\theta^3(1-\theta)^3\, d\theta$

$RHS = \int_0^1 140\theta^4 - 420\theta^5 + 420\theta^6 - 140\theta^7\, d\theta$

Let $F'(\theta) = 140\theta^4 - 420\theta^5 + 420\theta^6 - 140\theta^7$

then $F(\theta) = 28\theta^5 - 70\theta^6 + 60\theta^7 - \frac{35}{2}\theta^8$

Thus $RHS = F(1) - F(0) = 28 - 70 + 60 - \frac{35}{2} - 0 = \frac{1}{2}$

i.e. $E_{p(\theta|X_{1:2})}(\theta) = \frac{1}{2}$

After observing $X_{1:4}$: $E_{p(\theta|X_{1:4})}(\theta) = \int_0^1 \theta\, p(\theta|X_{1:4})\, d\theta = \int_0^1 \theta\, 630\theta^4(1-\theta)^4 d\theta$

$RHS = \int_0^1 630\theta^5 - 630\times4\,\theta^6 + 630\times6\,\theta^7 - 630\times4\,\theta^8 + 630\theta^9$

Let $F'(\theta) = 630\theta^5 - 630\times4\,\theta^6 + 630\times6\,\theta^7 - 630\times4\,\theta^8 + 630\theta^9$

then $F(\theta) = 105\theta^6 - 360\theta^7 + \frac{315\times3}{2}\theta^8 - 280\theta^9 + 63\theta^{10}$

Thus $RHS = F(1) - F(0) = 105 - 360 + \frac{945}{2} - 280 + 63 - 0 = \frac{1}{2}$

i.e. $E_{p(\theta|X_{1:4})}(\theta) = \frac{1}{2}$

d) Before any evidence: $Var_\theta(\theta) = E_\theta(\theta^2) - (E_\theta(\theta))^2$

From above, we know that $E_\theta(\theta) = \frac{1}{2}$, then $(E_\theta(\theta))^2 = \frac{1}{4}$

$E_\theta(\theta^2) = \int_0^1 p(\theta)\theta^2 d\theta = \int_0^1 \theta^2\, 30\theta^2(1-\theta)^2 d\theta$

$RHS = \int_0^1 30\theta^4 - 60\theta^5 + 30\theta^6\, d\theta$

Let $F'(\theta) = 30\theta^4 - 60\theta^5 + 30\theta^6$

then $F(\theta) = 6\theta^5 - 10\theta^6 + \frac{30}{7}\theta^7$

Thus $RHS = F(1) - F(0) = 6 - 10 + \frac{30}{7} - 0 = \frac{2}{7}$, i.e. $E_\theta(\theta^2) = \frac{2}{7}$

Then, we have $Var_\theta(\theta) = \frac{2}{7} - \frac{1}{4} = \frac{1}{28}$

After observing $X_{1:2}$: $Var_{p(\theta|X_{1:2})}(\theta) = E_{p(\theta|X_{1:2})}(\theta^2) - (E_{p(\theta|X_{1:2})}(\theta))^2$

From above, we know that $E_{p(\theta|X_{1:2})}(\theta) = \frac{1}{2}$, then $(E_{p(\theta|X_{1:2})}(\theta))^2 = \frac{1}{4}$

$E_{p(\theta|X_{1:2})}(\theta^2) = \int_0^1 p(\theta|X_{1:2})\,\theta^2 d\theta = \int_0^1 140\theta^3(1-\theta)^3\theta^2 d\theta$

$RHS = \int_0^1 140\theta^5 - 420\theta^6 + 420\theta^7 - 140\theta^8\, d\theta$

Let $F'(\theta) = 140\theta^5 - 420\theta^6 + 420\theta^7 - 140\theta^8$

then $F(\theta) = \frac{70}{3}\theta^6 - 60\theta^7 + \frac{105}{2}\theta^8 - \frac{140}{9}\theta^9$

Thus, $RHS = F(1) - F(0) = \frac{70}{3} - 60 + \frac{105}{2} - \frac{140}{9} - 0 = \frac{5}{18}$, i.e. $E_{p(\theta|X_{1:2})}(\theta^2) = \frac{5}{18}$

Then, we have $Var_{p(\theta|X_{1:2})}(\theta) = \frac{5}{18} - \frac{1}{4} = \frac{1}{36}$

After observing $X_{1:4}$: $Var\, p(\theta|X_{1:4}) = E_{p(\theta|X_{1:4})}(\theta^2) - (E_{p(\theta|X_{1:4})}(\theta))^2$

From above, we know that $E_{p(\theta|X_{1:4})}(\theta) = \frac{1}{2}$, then $(E_{p(\theta|X_{1:4})}(\theta))^2 = \frac{1}{4}$

$E_{p(\theta|X_{1:4})}(\theta^2) = \int_0^1 p(\theta|X_{1:4})\,\theta^2 d\theta = \int_0^1 630\theta^4(1-\theta)^4\theta^2 d\theta = \frac{3}{11}$

Then, we have $Var\, p(\theta|X_{1:4}) = \frac{3}{11} - \frac{1}{4} = \frac{1}{44}$

When $P'(\theta)=0$, $\theta\in\{0,\frac{1}{2},1\}$ $=\frac{1}{2}$

e) $P(\theta)=30\theta^2(1-\theta)^2$, then $P'(\theta)=60\theta(2\theta^2-3\theta+1)$, $\theta_{MAP}=\underset{\theta\in\{0,\frac{1}{2},1\}}{\arg\max}\,P(\theta)=\frac{1}{2}$ $\theta_{MAP}=\underset{\theta\in\{0,\frac{1}{2},1\}}{\arg\max}\,p(\theta)$

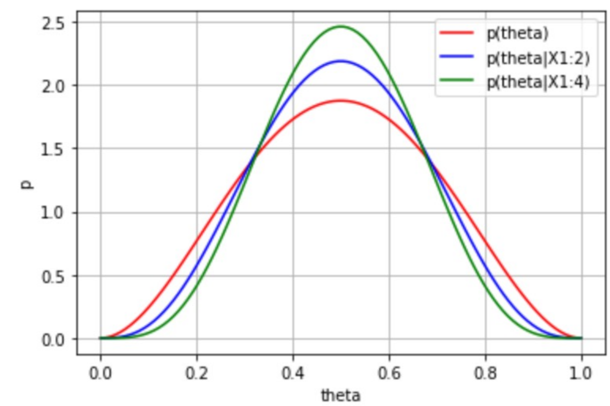
$P(\theta|X1:2)=140\theta^3(1-\theta)^3$, then $P'(\theta|X1:2)=-420(\theta-1)^2\theta^2(2\theta-1)$, when $P'(\theta)=0$, $\theta\in\{0,\frac{1}{2},1\}$

$P(\theta|X1:4)=630\theta^4(1-\theta)^4$, then $P'(\theta|X1:4)=2520(\theta-1)^3\theta^3(2\theta-1)$, when $P'(\theta|X1:4)=0$, $\theta\in\{0,\frac{1}{2},1\}$

$\theta_{MAP}=\underset{\theta\in\{0,\frac{1}{2},1\}}{\arg\max}\,P(\theta)=\frac{1}{2}$

Posterior	PDF	μ	σ^2	θ_{MAP}
$P(\theta)$	$30\theta^2(1-\theta)^2$	$\frac{1}{2}$	$\frac{1}{28}$	$\frac{1}{2}$
$P(\theta|X1:2=01)$	$140\theta^3(1-\theta)^3$	$\frac{1}{2}$	$\frac{1}{36}$	$\frac{1}{2}$
$P(\theta|X1:4=0101)$	$630\theta^4(1-\theta)^4$	$\frac{1}{2}$	$\frac{1}{44}$	$\frac{1}{2}$

f)

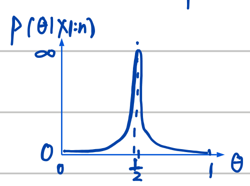


Conclusion:

The figures of the 3 distributions are all symmetric, the axis of symmetry is $\theta=\frac{1}{2}$.
The σ^2 of $p(\theta|X1:4)$ has a smaller value than $p(\theta|X1:2)$, then its shape is taller and narrower. The σ^2 of $p(\theta|X1:2)$ also has a smaller value than $P(\theta)$, and the shape is also taller and narrower.

For their max values: $P(\theta|X1:4)>P(\theta|X1:2)>P(\theta)$

g) When $n\to\infty$, $P(\theta|X1:n)=\varphi\theta^n(1-\theta)^n$, where $\varphi\propto n$ is a larger positive number.
the plot of $P(\theta|X1:n)$ has $\mu=\frac{1}{2}$, σ^2 will be smaller ($\sigma^2\to0$)
and $\max(p(\theta|X1:n))$ will be bigger. ($\max(p(\theta|X1:n))\to\infty$). $\theta_{MAP}=\frac{1}{2}$.



The shape of the distribution will have only one "spike" when $\theta=\frac{1}{2}$.
other places ($0\leq\theta\leq1$) will be flatten and the height will close to 0.

Question 2. Bayesian Inference on Imperfect Information

a) When $\phi=0$, the camera's observation will be the opposite of coins' flipping results.
$\phi=0.5$, the camera's observation is fully random, the accuracy will be 0.5.
$\phi=1$, the camera's observation will be the same as coins, i.e. the observation is correct.
when $\phi=0$ or $\phi=1$, the agent will have more confidence to update the posterior on θ.
Since the camera's observation provide more information about the coin than $\phi=0.5$.

b) From the plot, we learned that

$$P(\hat{X}=0|\theta) = P(\hat{X}=0, X=0|\theta) + P(\hat{X}=0, X=1|\theta)$$
$$= P(X=0|\theta)\cdot P(\hat{X}=0|X=0,\theta) + P(X=1|\theta)\cdot P(\hat{X}=0|X=1,\theta)$$
$$= (1-\theta)\cdot\phi + \theta\cdot(1-\phi)$$
$$= \phi + \theta - 2\theta\phi$$

$$P(\hat{X}=1|\theta) = P(\hat{X}=1, X=0|\theta) + P(\hat{X}=1, X=1|\theta)$$
$$= P(X=0|\theta)\cdot P(\hat{X}=1|X=0,\theta) + P(X=1|\theta)\cdot P(\hat{X}=1|X=1,\theta)$$
$$= (1-\theta)\cdot(1-\phi) + \theta\cdot\phi$$
$$= -\phi - \theta + 2\theta\phi + 1$$

c) $$P(\theta|\hat{X}=0) = \frac{P(\theta)P(\hat{X}=0|\theta)}{P(\hat{X}=0)} = \frac{P(\theta)P(\hat{X}=0|\theta)}{\int_\theta P(\hat{X}=0,\theta)} = \frac{P(\theta)P(\hat{X}=0|\theta)}{\int_\theta P(\theta)P(\hat{X}=0|\theta)\,d\theta}$$

$$RHS = \frac{P(\theta)(\phi+\theta-2\theta\phi)}{\int_\theta P(\theta)(\phi+\theta-2\theta\phi)\,d\theta}$$

When $\phi=1$, $RHS = \frac{P(\theta)(1-\theta)}{\int_\theta P(\theta)(1-\theta)\,d\theta}$; $\phi=0$, $RHS = \frac{P(\theta)\cdot\theta}{\int_\theta P(\theta)\cdot\theta\,d\theta}$; $\phi=\frac{1}{2}$, $RHS = \frac{P(\theta)}{\int_\theta P(\theta)\,d\theta} = P(\theta)$

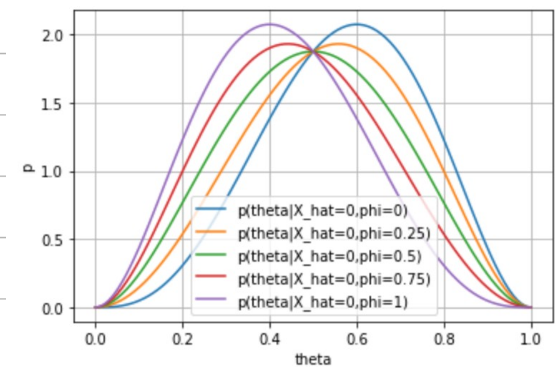
When $\phi=\frac{1}{2}$, $P(\theta|\hat{X}=0)=P(\theta)$, the camera's observation tells no information of the coins' flipping. The observation is not helpful for our inference.
From Bayes' theorem, we know that $P(\theta|\hat{X}=0) = P(\theta)\frac{P(\hat{X}=0|\theta)}{P(\hat{X}=0)}$

when $\phi=0$, $P(\theta|\hat{X}=0) = P(\theta)\cdot\frac{\theta}{\int_\theta P(\theta)\cdot\theta\,d\theta} = P(\theta)\cdot\frac{\theta}{E(\theta)}$. In this case, camera's observation will be opposite of the coin. then $\hat{X}=\neg X$, we have $P(\hat{X}=0|\theta)=P(X=1|\theta)=\theta$, and $P(\hat{X}=0) =$
$= \int_X P(\hat{X}=0,X)\,dx = \int_X P(X=1)\,dx = P(X=1) = E(\theta)$
Same as above, when $\phi=1$, then $\hat{X}=X$, $P(\hat{X}=0|\theta)=P(X=0|\theta)=1-\theta$.
and $P(\hat{X}=0) = 1-P(\hat{X}=1) = 1-P(X=1) = 1-E(\theta)$
Both $\phi=0$ and 1 provide the most information for the inference of θ.

d) $$P(\theta|\hat{X}=0) = \frac{P(\theta)(\phi+\theta-2\theta\phi)}{\int_\theta P(\theta)(\phi+\theta-2\theta\phi)\,d\theta} = \frac{30\theta^2(1-\theta)^2(\phi+\theta-2\theta\phi)}{\int_\theta 30\theta^2(1-\theta)^2(\phi+\theta-2\theta\phi)\,d\theta} = \frac{30\theta^2(1-\theta)(\phi+\theta-2\theta\phi)}{\frac{1}{2}}$$

$RHS = 60\theta^2(1-\theta)^2(\phi+\theta-2\theta\phi)$:

e)



From $\phi=0$ increases to $\phi=1$, the shape of its distribution changes from left-tailed to symmetric then to right-tailed.
The max value decreases until $\phi=0.5$, the increases later.

Question 3 Relating Random Variable (the proof method is learned from Wikipedia)

Let $g(x) := x^2+2$, then $g^{-1}(y) = \sqrt{y-2}$, where $x \in [0,1]$, $y \in [2,3]$

Monotonicity of $g(x)$: $g'(x) = 2x$, For $\forall x \in [0,1]$, $g'(x) \geq 0$.

then $g: \mathbb{R} \to \mathbb{R}$ is a monotonically non-decreasing function.

We know that the probability contained in a differential area must be invariant under change of variables, i.e. $|f_X(x)dx| = |f_Y(y)dy|$, where $f_X(x)$, $f_Y(y)$ represents PDFs of X and Y.

It implies that $f_Y(y) = f_X(g^{-1}(y)) \left|\frac{d}{dy}(g^{-1}(y))\right|$,

where $f_X := x^2+\frac{2}{3}x+\frac{1}{3}$, $g^{-1}(y) = \sqrt{y-2}$, $y \in [2,3]$

Then $\left|\frac{d}{dy}(g^{-1}(y))\right| = \left|\frac{d}{dy}(y-2)^{\frac{1}{2}}\right| = \left|\frac{1}{2}(y-2)^{-\frac{1}{2}}\right| = \frac{1}{2} \cdot \frac{1}{\sqrt{y-2}}$

$f_X(g^{-1}(y)) = (\sqrt{y-2})^2 + \frac{2}{3} \cdot \sqrt{y-2} + \frac{1}{3}$

Then $RHS = (y-2+\frac{2}{3}\sqrt{y-2}+\frac{1}{3}) \cdot \frac{1}{2\sqrt{y-2}}$

Let $z=\sqrt{y-2}$, then $RHS = (z^2+\frac{2}{3}z+\frac{1}{3}) \cdot \frac{1}{2z} = \frac{z}{2}+\frac{1}{3}+\frac{1}{6z}$

Thus $f_Y(y) = \frac{\sqrt{y-2}}{2}+\frac{1}{3}+\frac{1}{6\sqrt{y-2}}$, $y \in [2,3]$