Q1

a) Compute the new PDF after having observed n consecutive **ones** (that is, $x_{1:n}$ is a sequence where $\forall i.x_i = 1$), for an arbitrary prior pdf $p(\theta)$. Simplify your answer as much as possible.

We want to know $p(\theta \mid X = x_1 \cdots x_n)$ where $x_1 \cdots x_n = 111 \cdots 1_n$

$\because p(x \mid y) = \frac{p(y \mid x)\, p(x)}{p(y)}$ (from lecture, Bayes' rule)

$p(x) = \int_y p(x, y)\, dy$ (from lecture, sum rule)

$p(x, y) = p(x \mid y) p(y)$ (from lecture, product rule)

$\therefore p(\theta \mid X = x_1 \cdots x_n) = \frac{p(X = x_1 \cdots x_n \mid \theta)\, p(\theta)}{p(X = x_1 \cdots x_n)}$

$\because p(X = x_1 \cdots x_n \mid \theta) = p(X = x_1 \mid \theta)\, p(X = x_2 \mid \theta) \cdots p(X = x_n \mid \theta)$ (since i.i.d)

$\therefore p(\theta \mid X = x_1 \cdots x_n) = \frac{\theta^n p(\theta)}{\int_0^1 p(X = x_1 \cdots x_n, \theta)\, d\theta} = \frac{\theta^n p(\theta)}{\int_0^1 p(X = x_1 \cdots x_n \mid \theta)\, p(\theta)\, d\theta} = \frac{\theta^n p(\theta)}{\int_0^1 \theta^n p(\theta)\, d\theta}$

b) Compute the new PDF after having observed n consecutive **zeros**, (that is, $x_{1:n}$ is a sequence where $\forall i.x_i = 0$) for an arbitrary prior pdf $p(\theta)$. Simplify your answer as much as possible.

$p(\theta \mid X = x_1 \cdots x_n)$ where $x_1 \cdots x_n = 000 \cdots 0_n$

$\because p(\theta \mid X = x_1 \cdots x_n) = \frac{p(X = x_1 \cdots x_n \mid \theta)\, p(\theta)}{p(X = x_1 \cdots x_n)}$

$= \frac{(1-\theta)^n p(\theta)}{\int_0^1 p(\theta)(1-\theta)^n}$

c) Evaluate $p(\theta \mid x_{1:n} = 1^n)$ for the uniform prior $p(\theta) = 1$.

from (a), we know $p(\theta \mid x_1 \cdots x_n = 1^n) = \frac{\theta^n p(\theta)}{\int_0^1 \theta^n p(\theta)\, d\theta}$

$\therefore p(\theta \mid x_1 \cdots x_n = 1^n) = \frac{\theta^n}{\int_0^1 \theta^n d\theta} = (n+1)\theta^n$

d) Compute the expected value μ_n of X after observing n consecutive ones, with a uniform prio $p(\theta) = 1$. Provide intuition explaining the behaviour of μ_n as $n \to \infty$.

$X = x_1 \cdots x_n = 1^n$

$\mu_n = \int_0^1 \theta p(\theta \mid X = x_1 \cdots x_n)\, d\theta = \int_0^1 \theta (n+1)\theta^n\, d\theta = \int_0^1 (n+1)\theta^{n+1}\, d\theta = \left[\frac{(n+1)\theta^{n+2}}{n+2}\right]_0^1 = \frac{n+1}{n+2} = \frac{n+2-1}{n+2} = 1 - \frac{1}{n+2}$

So when $n \to \infty$, which means we have lots of "1", and the limit is $\lim_{n\to\infty} 1 - \frac{1}{n+2} = 1$, the mean will be very close to 1

e) Compute the variance σ_n^2 of the distribution on X after observing n consecutive ones, with a uniform prior $p(\theta) = 1$. Provide intuition explaining the behaviour of σ_n^2 as $n \to \infty$.

$\sigma_n^2 = E[(\theta - \mu_n)^2] = E[\theta^2 - 2\mu_n\theta + \mu_n^2] = E[\theta^2] - 2\mu_n E[\theta] + (E[\theta])^2 = E[\theta^2] - (E[\theta])^2$

$= \int_0^1 \theta^2 p(\theta \mid X = x_1 \cdots x_n) - \left(\frac{n+1}{n+2}\right)^2 = \int_0^1 \theta^{n+2}(n+1)\, d\theta - \left(\frac{n+1}{n+2}\right)^2 = \left[\frac{n+1}{n+3}\theta^{n+3}\right]_0^1 - \left(\frac{n+1}{n+2}\right)^2$

$= \frac{n+1}{n+3} - \left(\frac{n+1}{n+2}\right)^2 = \frac{n+3-2}{n+3} - \left(\frac{n+2-1}{n+2}\right)^2 = 1 - \frac{2}{n+3} - \left(1 - \frac{1}{n+2}\right)^2$

So when $n \to \infty$, which means we have lots of "1". so the value would not change, the variance should be 0 as $\lim_{n\to\infty} 1 - \frac{2}{n+3} - (1 - \frac{1}{n+2})^2 = 0$

f) Compute the *maximum a posteriori* estimation θ_{MAP_n} of the distribution on X after observing n consecutive ones, with a uniform prior $p(\theta) = 1$. Provide intuition explaining how θ_{MAP_n} varies with n.

$p(\theta | X = x_1 \cdots x_n)$ where $x_1 \cdots x_n = 1^n$

$\because p(\theta | X = x_1 \cdots x_n) = \frac{P(X = x_1 \cdots x_n | \theta) P(\theta)}{P(X = x_1 \cdots x_n)} = (n+1)\theta^n$

$\because \frac{d((n+1)\theta^n)}{d\theta} = (n+1) n \theta^{n-1} \geq 0$

so when $\theta = 1$, $(n+1)\theta^n$ is maximum, and as n growing, the θ_{MAP_n} is still 1

g) Given we have observed n coin flips in a row, what do you think would be a better choice for the best guess of the true value of θ? μ_n or θ_{MAP}? Justify your answer.

$p(\theta | X = x_1 \cdots x_n) = (n+1)\theta^n$

$\mu_n = \frac{n+1}{n+2}$

μ_n is the best guess, because the θ_{MAP} will always be 1 even if the $n=1$ or $n=2$, however, μ_n will give a value very close to 1 such as $\frac{2}{3}$ $(n=1)$ or $\frac{3}{4}$ $(n=2)$, and when $n \to \infty$ the $\lim_{n \to \infty} \frac{n+1}{n+2}$ will become 1.

h) Plot the probability distributions $p(\theta | x_{1:n} = 1)$ over the interval $0 \leq \theta \leq 1$ for $n \in \{0, 1, 2, 3, 4\}$ to compare them.

$p(\theta | x_1 \cdots x_n) = (n+1)\theta^n$

$n=0 \quad p(\theta | x_1 \cdots x_n) = 1$

$n=1 \quad p(\theta | x_1 \cdots x_n) = 2\theta$

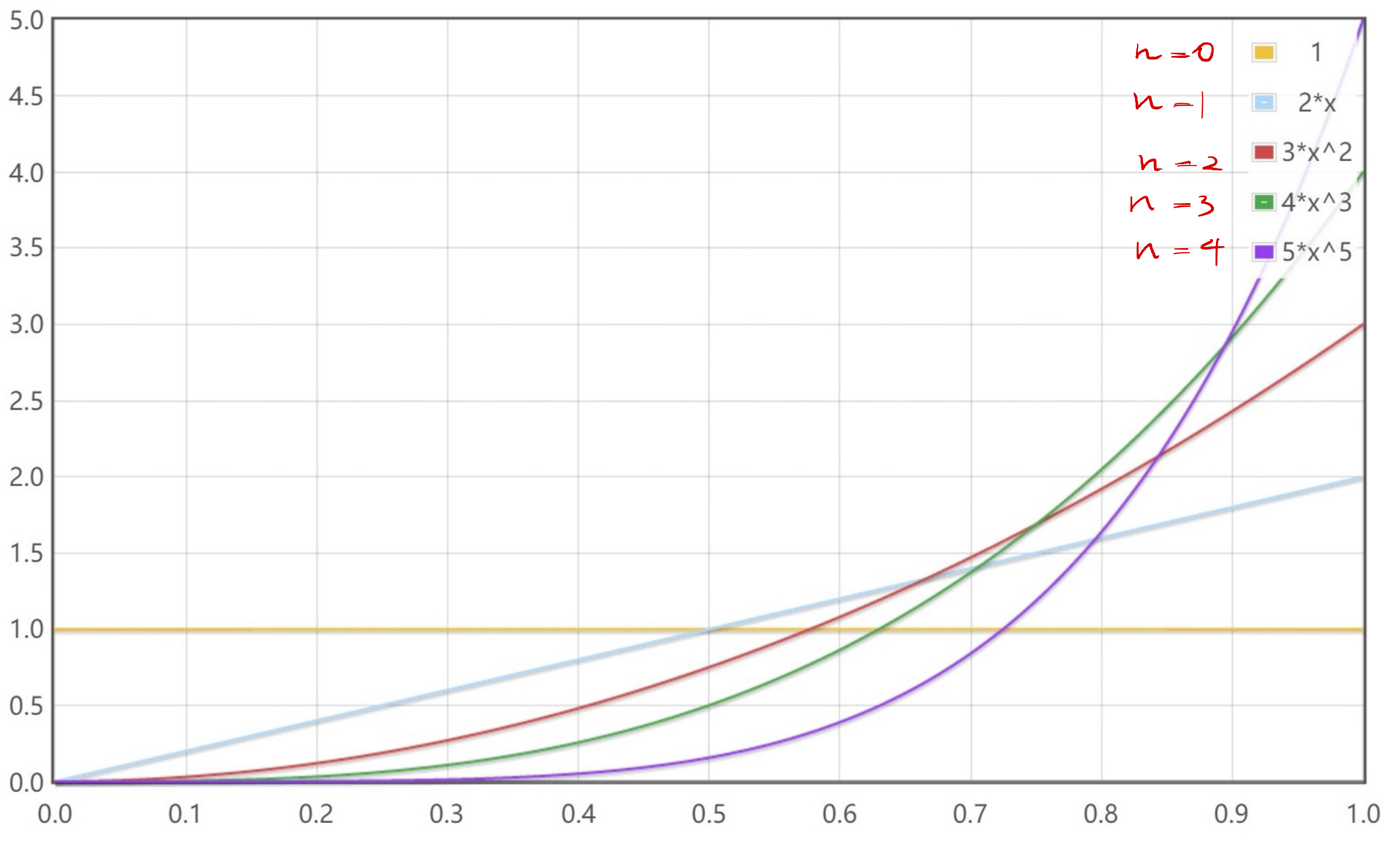
$n=2 \quad p(\theta | x_1 \cdots x_n) = 3\theta^2$

$n=3 \quad p(\theta | x_1 \cdots x_n) = 4\theta^3$

$n=4 \quad p(\theta | x_1 \cdots x_n) = 5\theta^4$

So we can see that, as we observe more "ones", we have more confidence that the biased coin is skewed to one, which means $P(\theta > 0.5)$ will increase.

We can also see in the graph that, as n increases the area of $\theta > 0.5$ will become the majority of the whole area.



Q2

a) (5 credits) Briefly comment about how the camera behaves for $\alpha = \beta = 1$, for $\alpha = \beta = 1/2$, and for $\alpha = \beta = 0$. For each of these cases, how would you expect this would change how the agent updates it's prior to a posterior on θ, given an observation of $\widehat{X}$? (No equations required.)

α, β is the probability that the camera makes no mistake

1) $\alpha = \beta = 1$:

which mean the camera can 100% observe the round time, so the ent will normally update its prior based on the fact, and produce the same result as using X

2) $\alpha = \beta = \frac{1}{2}$:

Which means the camera has 50% to correctly report the true case, and 50% to go wrong, but the agent will still correctly update the prior, and produce the same result as 1), except for the case where $\theta = 0$ or 1, which will update the majority of posterior pdf to $\theta = 0.5$

3) $\alpha = \beta = 0$

The camera never report the true fact of X. It read every 1 as 0, read every 0 as 1. The posterior will move to opposite direction of prior. and prior and posterior will be symmetric about the line $X = 0.5$.

b) (10 credits) Compute $p(\widehat{X} = x|\theta)$ for all $x \in \{0, 1\}$.

1) $p(\hat{X}=0|\theta) = P(X=0|\theta)\cdot P(\hat{X}=0|X=0,\theta) + P(X=1|\theta)\cdot P(\hat{X}=0|X=1,\theta) = (1-\theta)\alpha + \theta(1-\beta)$

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

c) (15 credits) The coin is flipped, and the camera reports seeing a one. (i.e. that $\hat{X} = 1$.) Given an arbitrary prior $p(\theta)$, compute the posterior $p(\theta|\hat{X} = 1)$. What does $p(\theta|\hat{X} = 1)$ simplify to when $\alpha = \beta = 1$? When $\alpha = \beta = 1/2$? When $\alpha = \beta = 0$? Explain your observations.

$$P(\theta|\hat{X}=1) = \frac{P(\hat{X}=1|\theta)P(\theta)}{P(\hat{X}=1)} = \frac{P(\hat{X}=1|\theta)P(\theta)}{\int_0^1 P(\hat{X}=1,\theta)\,d\theta} = \frac{P(\hat{X}=1|\theta)P(\theta)}{\int_0^1 P(\hat{X}=1|\theta)\cdot P(\theta)\,d\theta}$$

$$= \frac{P(\theta)[\theta\beta + (1-\theta)(1-\alpha)]}{\int_0^1 P(\theta)[\theta\beta + (1-\theta)(1-\alpha)]\,d\theta}$$

1) When $\alpha = \beta = 1$:

$P(\theta|\hat{X}=1) = \frac{\theta P(\theta)}{\int_0^1 \theta P(\theta)\,d\theta} = P(\theta|X=1)$, which means $X = \hat{X}$ and the prior will not change not change based on the obsomution $\hat{X}$

2) When $\alpha = \beta = \frac{1}{2}$

$P(\theta|\hat{X}=1) = \frac{P(\theta)(\frac{1}{2}\theta + \frac{1}{2}(1-\theta))}{\int_0^1 P(\theta)(\frac{1}{2}\theta + \frac{1}{2}(1-\theta))} = \frac{P(\theta)}{\int_0^1 P(\theta)\,d\theta} = P(\theta)$. Which means the posterior will not change.

3) When $\alpha=\beta=0$: $\quad 1-\theta=t \quad P$

$$p(\theta|\hat{x}=1)=\frac{p(\theta)(1-\theta)}{\int_0^1 p(\theta)(1-\theta)\,d\theta}$$, which means the $p(\theta|\hat{x}=1)$ and $p(\theta|x=1)$ are symmetric about the line $\theta=\frac{1}{2}$, $\theta\in[0,1]$, for example, $p(\theta=\frac{1}{2}|\hat{x}=1)=p(\theta=\frac{1}{2}|\hat{x}=1)$

$$=\frac{p(\theta=\frac{1}{2})}{\int_0^1 p(\theta=\frac{1}{2})}$$

and $p(\theta=0|\hat{x}=1)=p(\theta=1|x=1)$

d) Evaluate $p(\theta|\hat{X}=1)$ for the uniform prior $p(\theta)=1$. Simplify it under the assumption that $\beta:=\alpha$.

$$p(\theta|\hat{x}=1)=\frac{\alpha\theta+(1-\theta)(1-\alpha)}{2\int_0^1 \alpha\theta\,d\theta+\int_0^1 1-\alpha-\theta\,d\theta}=\frac{2\alpha\theta-\alpha-\theta+1}{\alpha+1-\alpha-\frac{1}{2}}=4\alpha\theta-2\alpha-2\theta+2$$

$$=2\theta(2\alpha-1)-2\alpha+2$$

e) (10 credits) Let $\beta=\alpha$. Plot $p(\theta|\hat{X}=1)$ as a function of θ, for all $\alpha\in\{0,\frac{1}{4},\frac{2}{4},\frac{3}{4},1\}$ on the same graph to compare them. Comment on how the shape of the distribution changes with α. Explain your observations.

$p(\theta|\hat{x}=1)=2\theta(2\alpha-1)-2\alpha+2$

$\alpha=0$: $p(\theta|\hat{x}=1)=-2\theta+2$

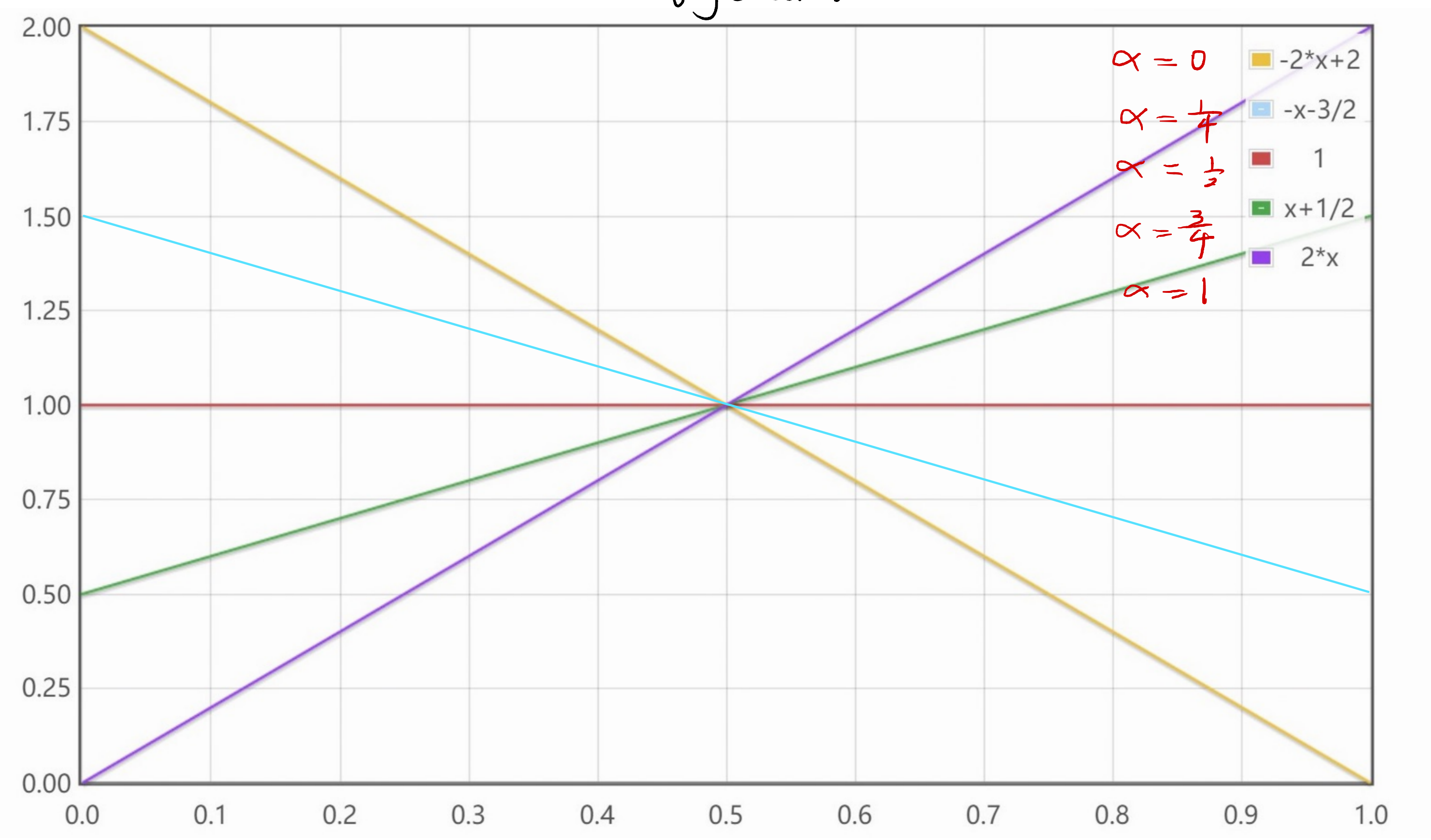
$\alpha=\frac{1}{4}$: $p(\theta|\hat{x}=1)=-\theta+\frac{3}{2}$

$\alpha=\frac{1}{2}$: $p(\theta|\hat{x}=1)=1$

$\alpha=\frac{3}{4}$: $p(\theta|\hat{x}=1)=\theta+\frac{1}{2}$

$\alpha=1$: $p(\theta|\hat{x}=1)=2\theta$

as α growing from 0 to 1, the area of $\theta<0.5$ will become smaller than the area of $\theta>0.5$. and when $\alpha=\frac{1}{2}$, the two areas are equivalent. The graph tells us as α grows from 0 to 1, the probability of θ greater than 0.5 will increase.



Q3

a) Show that the probability density function p_Y for Y is given by

$$p_Y(y) = \frac{1}{y^2} p_X(\frac{1}{y})$$

$\because P_X(X \le x) = F(x) = \int_0^x P_X(x)dx$

$\because P_Y(Y \le y) = F(y) = \int_1^y P_Y(y)dy$ ①

$\because Y = \frac{1}{X},\ X = \frac{1}{Y} \Rightarrow x = \frac{1}{y},\ y = \frac{1}{x}$

$\therefore P_Y(Y \le y) = P_Y(\frac{1}{X} \le y) = P_X(\frac{1}{y} \le X) = \int_{\frac{1}{y}}^{1} P_X(x)dx$

$\because \frac{dx}{dy} = -\frac{1}{y^2} \Rightarrow dx = -\frac{1}{y^2}dy$

$\therefore \int_{\frac{1}{\frac{1}{y}}}^{\frac{1}{1}} P_X(\frac{1}{y}) \cdot -\frac{1}{y^2}dy = \int_y^1 -P_X(\frac{1}{y})\frac{1}{y^2}dy = \int_1^y P_X(\frac{1}{y})\frac{1}{y^2}dy$ ②

from ①.② we get $P_Y(y) = P_X(\frac{1}{y})\frac{1}{y^2}$

b) Hence, or otherwise, compute the expected profit for the player under this game. Your answer will be in terms of m and p_X, and should be as simplified as possible.

We only know m is a number, but don't know whether m is an integer or real number

① m is real number:

let $f(y) = \begin{cases} m-1 & y > m \\ -1 & y \le m \end{cases}$ since c is sampled from Y

so we want to know $E_Y[f(y)]$

$\therefore E_Y[f(y)] = \int_1^\infty f(y)P_Y(y)dy = \int_1^m -1 \cdot P_Y(y)dy + \int_m^\infty (m-1)P_Y(y)dy$

$= -\int_1^m P_Y(y)dy + m\int_m^\infty P_Y(y)dy - \int_m^\infty P_Y(y)dy$

$\because \int_1^\infty P_Y(y)dy = 1$

$\therefore E_Y[f(y)] = m\int_m^\infty P_Y(y)dy - 1 = m\int_m^\infty P_X(\frac{1}{y})\frac{1}{y^2}dy - 1$

② m is an integer

$f(m) = \begin{cases} m-1 & m < c \\ -1 & m \ge c \end{cases}$ $m \in \{1, 2, 3, \cdots \infty\}$ $m \in$ int $c \in Y$

c) Suppose the casino chooses a uniform distribution over $(0, 1]$ for X, that is,

$P_Y(y) = \frac{1}{y^2}$

$$p_X(x) = \begin{cases} 1 & 0 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$m \cdot \frac{1}{m} - 1 = 0$

$0 - (-\frac{1}{m})$ [

What strategy should the player use to maximise their expected profit?

$E_Y[f(y)] = m\int_m^\infty \frac{1}{y^2}dy - 1 = m\left[-\frac{1}{y}\right]_m^\infty - 1 = 0$