COMP3670 2021 Theory Assignment 3

Yuxuan Lin – u6828533@anu.edu.au

October 10, 2021

Introduction to Machine Learning

By turning in this assignment, I agree by the ANU honor code and declare that all of this is my own work.

Exercise 1

a) ***Proof:*** We will prove by induction that, for all $n \in \mathbb{Z}_+$,

$$p(\theta|x_{1:n} = 1^n) = \frac{\theta^n p(\theta)}{\displaystyle\int_{\boldsymbol{\theta}} \theta^n p(\theta) d\theta} \tag{1}$$

Base case: When $n = 1$,

$$\begin{aligned}
p(\theta|x_1 = 1) &= \frac{p(x_1 = 1|\theta)p(\theta)}{p(x_1 = 1)} \text{...............Bayes' theorem} \\
&= \frac{p(x_1 = 1|\theta)p(\theta)}{\displaystyle\int_{\boldsymbol{\theta}} p(x_1 = 1, \theta) d\theta} \text{.............sum rule of probability} \\
&= \frac{p(x_1 = 1|\theta)p(\theta)}{\displaystyle\int_{\boldsymbol{\theta}} p(x_1 = 1|\theta)p(\theta) d\theta} \text{.........product rule of probability} \\
&= \frac{\theta p(\theta)}{\displaystyle\int_{\boldsymbol{\theta}} \theta p(\theta) d\theta} \text{...................conditional probability}
\end{aligned}$$

Induction step: To start with, I will prove **Bayes' theorem with 3 events**

$$p(x|y, z) = \frac{p(z|x, y)p(x|y)}{p(z|y)}$$

$$
\begin{aligned}
p(x|y,z) &= \frac{p(x,y,z)}{p(z,y)} \\
&= \frac{p(z|x,y)p(x,y)}{p(z,y)} \\
&= \frac{p(z|x,y)p(x|y)p(y)}{p(z|y)p(y)} \\
&= \frac{p(z|x,y)p(x|y)}{p(z|y)}.
\end{aligned}
$$

Let $k \in \mathbb{Z}_+$, $k \geq 2$ be given and suppose (1) is true for $n = k$. Then

$$
\begin{aligned}
p(\theta|x_{1:k+1} = 1^{k+1}) &= p(\theta|x_{1:k} = 1^k, x_{k+1} = 1) \\
&= \frac{p(x_{k+1} = 1|\theta, x_{1:k} = 1^k)p(\theta|x_{1:k} = 1^k)}{p(x_{k+1} = 1|x_{1:k} = 1^k)} \text{ .... Bayes' theorem with 3 events} \\
&= \frac{p(x_{k+1} = 1|\theta, x_{1:k} = 1^k)p(\theta|x_{1:k} = 1^k)}{\displaystyle\int_{\boldsymbol{\theta}} p(x_{k+1} = 1, \theta|x_{1:k} = 1^k)d\theta} \text{ .... sum rule of probability} \\
&= \frac{p(x_{k+1} = 1|\theta, x_{1:k} = 1^k)p(\theta|x_{1:k} = 1^k)}{\displaystyle\int_{\boldsymbol{\theta}} p(x_{k+1} = 1|\theta, x_{1:k} = 1^k)p(\theta|x_{1:k} = 1^k)d\theta} \text{ ...... product rule of probability} \\
&= \frac{p(x_{k+1} = 1|\theta)p(\theta|x_{1:k} = 1^k)}{\displaystyle\int_{\boldsymbol{\theta}} p(x_{k+1} = 1|\theta)p(\theta|x_{1:k} = 1^k)d\theta} \text{ ...... independent and identically distributed} \\
&= \frac{\theta p(\theta|x_{1:k} = 1^k)}{\displaystyle\int_{\boldsymbol{\theta}} \theta p(\theta|x_{1:k} = 1^k)d\theta} \\
&= \frac{\theta \frac{\theta^k p(\theta)}{\int_{\boldsymbol{\theta}} \theta^k p(\theta)d\theta}}{\displaystyle\int_{\boldsymbol{\theta}} \theta \frac{\theta^k p(\theta)}{\displaystyle\int_{\boldsymbol{\theta}} \theta^k p(\theta)d\theta} d\theta} \text{ .... induction proof assumption} \\
&= \frac{\theta^{k+1}p(\theta)}{\displaystyle\int_{\boldsymbol{\theta}} \theta^{k+1}p(\theta)d\theta} \text{ .... expected value is a constant}
\end{aligned}
$$

Thus, (1) holds for $n = k + 1$, and the proof of the induction step is complete.

Conclusion: By the principle of induction, (1) is true for all $n \in \mathbb{Z}_+$.

Since $\mathbb{E}[\theta^n] = \int_{\boldsymbol{\theta}} \theta^n p(\theta) d\theta$, we can simplify the answer as

$$p(\theta | x_{1:n} = 1^n) = \frac{\theta^n p(\theta)}{\mathbb{E}[\theta^n]}.$$

b) ***Proof:*** We will prove by induction that, for all $n \in \mathbb{Z}_+$,

$$p(\theta | x_{1:n} = 0^n) = \frac{(1-\theta)^n p(\theta)}{\int_{\boldsymbol{\theta}} (1-\theta)^n p(\theta) d\theta} \tag{2}$$

Base case: When $n = 1$,

$$\begin{aligned}
p(\theta | x_1 = 0) &= \frac{p(x_1 = 0 | \theta) p(\theta)}{p(x_1 = 0)} \text{.............. Bayes' theorem} \\
&= \frac{p(x_1 = 0 | \theta) p(\theta)}{\int_{\boldsymbol{\theta}} p(x_1 = 0, \theta) d\theta} \text{............. sum rule of probability} \\
&= \frac{p(x_1 = 0 | \theta) p(\theta)}{\int_{\boldsymbol{\theta}} p(x_1 = 0 | \theta) p(\theta) d\theta} \text{......... product rule of probability} \\
&= \frac{(1-\theta) p(\theta)}{\int_{\boldsymbol{\theta}} (1-\theta) p(\theta) d\theta} \text{.............. conditional probability}
\end{aligned}$$

Induction step: Let $k \in \mathbb{Z}_+$, $k \geq 2$ be given and suppose (2) is true for $n = k$. Then

$$
\begin{aligned}
p(\theta|x_{1:k+1} = 0^{k+1}) &= p(\theta|x_{1:k} = 0^k, x_{k+1} = 0) \\
&= \frac{p(x_{k+1} = 0|\theta, x_{1:k} = 0^k)p(\theta|x_{1:k} = 0^k)}{p(x_{k+1} = 0|x_{1:k} = 0^k)} \quad \text{.........Bayes' theorem with 3 events} \\
&= \frac{p(x_{k+1} = 0|\theta, x_{1:k} = 0^k)p(\theta|x_{1:k} = 0^k)}{\displaystyle\int_{\boldsymbol{\theta}} p(x_{k+1} = 0, \theta|x_{1:k} = 0^k)d\theta} \quad \text{............sum rule of probability} \\
&= \frac{p(x_{k+1} = 0|\theta, x_{1:k} = 0^k)p(\theta|x_{1:k} = 0^k)}{\displaystyle\int_{\boldsymbol{\theta}} p(x_{k+1} = 0|\theta, x_{1:k} = 0^k)p(\theta|x_{1:k} = 0^k)d\theta} \quad \text{......product rule of probability} \\
&= \frac{p(x_{k+1} = 0|\theta)p(\theta|x_{1:k} = 0^k)}{\displaystyle\int_{\boldsymbol{\theta}} p(x_{k+1} = 0|\theta)p(\theta|x_{1:k} = 0^k)d\theta} \quad \text{...independent and identically distributed} \\
&= \frac{(1-\theta)p(\theta|x_{1:k} = 0^k)}{\displaystyle\int_{\boldsymbol{\theta}} (1-\theta)p(\theta|x_{1:k} = 0^k)d\theta} \\
&= \frac{(1-\theta)\frac{(1-\theta)^k p(\theta)}{\int_{\boldsymbol{\theta}} (1-\theta)^k p(\theta)d\theta}}{\displaystyle\int_{\boldsymbol{\theta}} (1-\theta)\frac{(1-\theta)^k p(\theta)}{\displaystyle\int_{\boldsymbol{\theta}} (1-\theta)^k p(\theta)d\theta}d\theta} \quad \text{..................induction proof assumption} \\
&= \frac{(1-\theta)^{k+1}p(\theta)}{\displaystyle\int_{\boldsymbol{\theta}} (1-\theta)^{k+1}p(\theta)d\theta} \quad \text{...........................expected value is a constant}
\end{aligned}
$$

Thus, (2) holds for $n = k + 1$, and the proof of the induction step is complete.

Conclusion: By the principle of induction, (2) is true for all $n \in \mathbb{Z}_+$.

Since $\mathbb{E}[(1-\theta)^n] = \displaystyle\int_{\boldsymbol{\theta}} (1-\theta)^n p(\theta)d\theta$, we can simplify the answer as

$$p(\theta|x_{1:n} = 1^n) = \frac{(1-\theta)^n p(\theta)}{\mathbb{E}[(1-\theta)^n]}.$$

c) Since $p(\theta) = 1$, $\boldsymbol{\theta} = \{\theta \in \mathbb{R} : 0 \leq \theta \leq 1\}$, from conclusion of a) we have

$$
\begin{aligned}
p(\theta | x_{1:n} = 1^n) &= \frac{\theta^n}{\displaystyle\int_0^1 \theta^n d\theta} \\
&= \frac{\theta^n}{\frac{\theta^{n+1}}{n+1}\Big|_0^1} \\
&= \frac{\theta^n}{\frac{1}{n+1}} \\
&= (n+1)\theta^n
\end{aligned}
$$

Hence, the answer is $p(\theta | x_{1:n} = 1^n) = (n+1)\theta^n$

d) Since expected value $\mu_n = \displaystyle\int_{\boldsymbol{\theta}} \theta p(\theta | x_{1:n} = 1^n) d\theta$, $\boldsymbol{\theta} = \{\theta \in \mathbb{R} : 0 \leq \theta \leq 1\}$, use conclusion of c) we have

$$
\begin{aligned}
\mu_n &= \int_0^1 (n+1)\theta^{n+1} d\theta \\
&= (n+1)\frac{\theta^{n+2}}{n+2}\Big|_0^1 \\
&= (n+1)\frac{1}{n+2} \\
&= \frac{n+1}{n+2}
\end{aligned}
$$

$$
\begin{aligned}
\lim_{n\to\infty} \mu_n &= \lim_{n\to\infty} \frac{n+1}{n+2} \\
&= \lim_{n\to\infty} (1 - \frac{1}{n+2}) \\
&= 1 - \frac{1}{\infty} \\
&= 1 - 0 \\
&= 1
\end{aligned}
$$

Hence, the answers are $\mu_n = \frac{n+1}{n+2}$ and $\lim\limits_{n\to\infty} \mu_n = 1$.

e) Since variance $\sigma_n^2 = \int_{\boldsymbol{\theta}} (\theta - \mu_n)^2 p(\theta|x_{1:n} = 1^n) d\theta$, $\boldsymbol{\theta} = \{\theta \in \mathbb{R} : 0 \leq \theta \leq 1\}$, use conclusion of c) and d) we have

$$
\begin{aligned}
\sigma_n^2 &= \int_0^1 (\theta - \frac{n+1}{n+2})^2 (n+1)\theta^n d\theta \\
&= \int_0^1 \Big[\frac{(n+1)^3}{(n+2)^2}\theta^n - \frac{2(n+1)^2}{n+2}\theta^{n+1} + (n+1)\theta^{n+2}\Big] d\theta \\
&= \Big[\frac{(n+1)^2}{(n+2)^2}\theta^{n+1} - \frac{2(n+1)^2}{(n+2)^2}\theta^{n+2} + \frac{n+1}{n+3}\theta^{n+3}\Big]\Big|_0^1 \\
&= \frac{(n+1)^2}{(n+2)^2} - \frac{2(n+1)^2}{(n+2)^2} + \frac{n+1}{n+3} \ldots \text{uniform prior is 1} \\
&= \frac{1+n}{3+n} - \frac{(1+n)^2}{(2+n)^2}
\end{aligned}
$$

$$
\lim_{n\to\infty} \sigma_n^2 = \lim_{n\to\infty} (\frac{1+n}{3+n} - \frac{(1+n)^2}{(2+n)^2}) = 0 \ldots \text{calculated by Wolfram Alpha}
$$

Hence, the answers are $\sigma_n^2 = \frac{1+n}{3+n} - \frac{(1+n)^2}{(2+n)^2}$ and $\lim_{n\to\infty} \sigma_n^2 = 0$.

f) If we take the derivative of the conclusion of c), we derive

$$
p'(\theta|x_{1:n} = 1^n) = n(n+1)\theta^{n-1} \tag{3}
$$

For $\theta \in \mathbb{R} : 0 \leq \theta \leq 1$, we always have (3) ≥ 0, which means $p(\theta|x_{1:n} = 1^n)$ monotonically increases in the range of $\boldsymbol{\theta}$, then the maximum a posteriori will be always be reached when $\theta = 1$, as known as $\theta_{\mathrm{MAP}n} = 1$.

$\theta_{\mathrm{MAP}n}$ is a constant of 1 and it will not vary with n. Because the uniform prior $p(\theta) = 1$, the value selection of θ will not be biased in the range of $\boldsymbol{\theta}$. And what we observed are n consecutive ones. Consequently, we always have $\theta_{\mathrm{MAP}n} = 1$ regardless of n.

g) In my opinion, μ_n is a better choice for the best guess of the true value of θ. The reason I don't choose $\theta_{\mathrm{MAP}n}$ is: Regardless of n, $\theta_{\mathrm{MAP}n}$ is always 1. But when n is small, for instance, if we only flip the coin for once, and we will conclude that the true value of θ is 1 from $\theta_{\mathrm{MAP}n}$. I think it is illogical. As for μ_n, it will get closer to 1 when n is large enough, which is a much more logical way for the best guess of the true value of θ.

h) From the conclusion of c), we have

$$p(\theta|x_{1:n} = 1^n) = \begin{cases} 1 & n = 0 \\ 2\theta & n = 1 \\ 3\theta^2 & n = 2 \\ 4\theta^3 & n = 3 \\ 5\theta^4 & n = 4 \end{cases}$$

Generated by *Desmos*, the plot is shown as below.

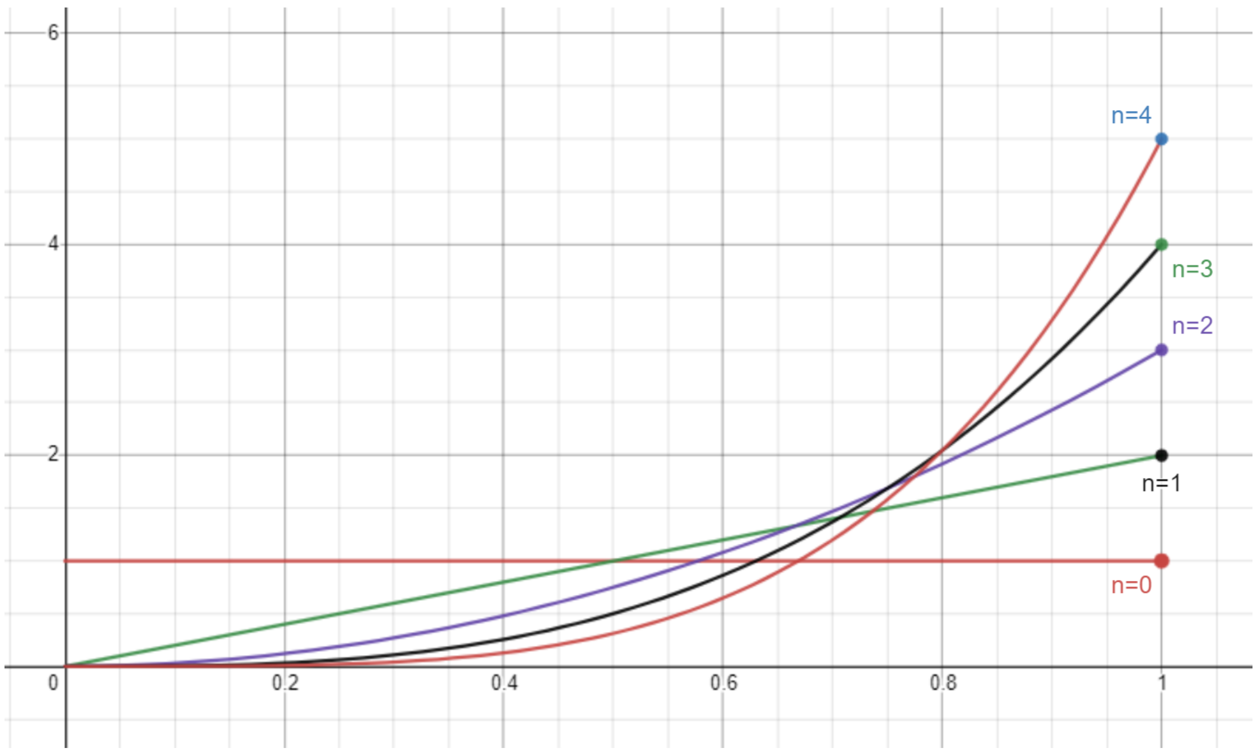

Figure 1: Probability Distributions

We observe from Figure 1 that, with the increase of n, the graphs of probability distributions $p(\theta|x_{1:n} = 1^n)$ are skewed towards $\theta = 1$.

Exercise 2

a) For $\alpha = \beta = 1$, the camera is noisy-free, and it can perfectly report back the result value. Consequently, the agent updates it's prior to a posterior on θ based on general Bayesian inference.

For $\alpha = \beta = \frac{1}{2}$, the camera has half-n-half probabilities of reporting back a 0 or a 1, regardless of the true result of a coin flip. Consequently, the way that the agent updates it's prior to a posterior on θ will be: A posteriori on θ after updating will always be the same as the prior.

For $\alpha = \beta = 0$, the camera reports back an exactly wrong result value of the true result of a coin flip. Consequently, the way that the agent updates it's prior to a posterior on *theta* will be as the following:

- When the observation value is 0, as known as the true result of a coin flip is 1, then a posteriori will skew towards 1.

- When the observation value is 1, as known as the true result of a coin flip is 0, then a posteriori will skew towards 0.

b) $\forall x \in \{0, 1\}$, we have

$$
\begin{aligned}
p(\hat{X} = x|\theta) &= \sum_{y=0}^{1} p(\hat{X} = x, X = y|\theta) \\
&= \sum_{y=0}^{1} p(\hat{X} = x|X = y, \theta)p(X = y|\theta) \\
&= p(\hat{X} = x|X = 1, \theta)p(X = 1|\theta) + p(\hat{X} = x|X = 0, \theta)p(X = 0|\theta) \\
&= \theta p(\hat{X} = x|X = 1) + (1 - \theta)p(\hat{X} = x|X = 0)
\end{aligned}
$$

For $x = 0$:

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

For $x = 1$:

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

c) Use the conclusion of b), and knowing that $\hat{X} = 1$, we have

$$
\begin{aligned}
p(\theta|\hat{X}=1) &= \frac{p(\hat{X}=1|\theta)p(\theta)}{p(\hat{X}=1)} \\
&= \frac{p(\hat{X}=1|\theta)p(\theta)}{\displaystyle\int_0^1 p(\hat{X}=1,\theta)d\theta} \\
&= \frac{[\theta\beta+(1-\theta)(1-\alpha)]p(\theta)}{\displaystyle\int_0^1 [\theta\beta+(1-\theta)(1-\alpha)]p(\theta)d\theta}
\end{aligned}
$$

For $\alpha=\beta=1$, we calculate

$$
p(\theta|\hat{X}=1) = \frac{\theta p(\theta)}{\displaystyle\int_0^1 \theta p(\theta)d\theta} = \frac{\theta p(\theta)}{\mathbb{E}[\theta]}
$$

The camera is noisy-free, and it can perfectly report back the result value. Consequently, the agent updates it's prior to a posterior on θ based on general Bayesian inference.

For $\alpha=\beta=\frac{1}{2}$, we calculate

$$
p(\theta|\hat{X}=1) == \frac{\frac{1}{2}p(\theta)}{\displaystyle\int_0^1 \frac{1}{2}p(\theta)d\theta} = \frac{\frac{1}{2}p(\theta)}{\frac{1}{2}\displaystyle\int_0^1 p(\theta)d\theta} = p(\theta)\ldots\text{integral of a PDF is 1}
$$

The camera has half-n-half probabilities of reporting back a 0 or a 1, regardless of the true result of a coin flip. Consequently, the way that the agent updates it's prior to a posterior on θ will be: A posteriori on θ after updating will always be the same as the prior.

For $\alpha=\beta=0$, we calculate

$$
p(\theta|\hat{X}=1) = \frac{(1-\theta)p(\theta)}{\displaystyle\int_0^1 (1-\theta)p(\theta)d\theta} = \frac{(1-\theta)p(\theta)}{\mathbb{E}[1-\theta]}
$$

The camera reports back an exactly wrong result value of the true result of a coin flip. Consequently, the way that the agent updates it's prior to a posterior on *theta* will be as the following:

- When the observation value is 0, as known as the true result of a coin flip is 1, then a posteriori will skew towards 1.

- When the observation value is 1, as known as the true result of a coin flip is 0, then a posteriori will skew towards 0.

d) Take $p(\theta) = 1$ into an equation we derive in c), then

$$\begin{aligned}
p(\theta|\hat{X} = 1) &= \frac{[\theta\beta + (1-\theta)(1-\alpha)]p(\theta)}{\displaystyle\int_0^1 [\theta\beta + (1-\theta)(1-\alpha)]p(\theta)d\theta} \\
&= \frac{\theta\beta + (1-\theta)(1-\alpha)}{\displaystyle\int_0^1 [\theta\beta + (1-\theta)(1-\alpha)]d\theta} \\
&= \frac{\theta\beta + (1-\theta)(1-\alpha)}{\frac{1}{2}(1-\alpha+\beta)} \ldots \text{calculated by Wolfram Alpha} \\
&= 2\theta\alpha + 2(1-\theta)(1-\alpha) \ldots \beta = \alpha \\
&= 4\theta\alpha - 2\theta - 2\alpha + 2
\end{aligned}$$

Hence, the answer is $p(\theta|\hat{X} = 1) = 4\theta\alpha - 2\theta - 2\alpha + 2$.

e) From the conclusion of d), we have

$$p(\theta|\hat{X} = 1) = \begin{cases} -2\theta + 2 & \alpha = 0 \\ -\theta + \frac{3}{2} & \alpha = 1/4 \\ 1 & \alpha = 2/4 \\ \theta + \frac{1}{2} & \alpha = 3/4 \\ 2\theta & \alpha = 1 \end{cases}$$

Generated by *Desmos*, the plot is shown as below.

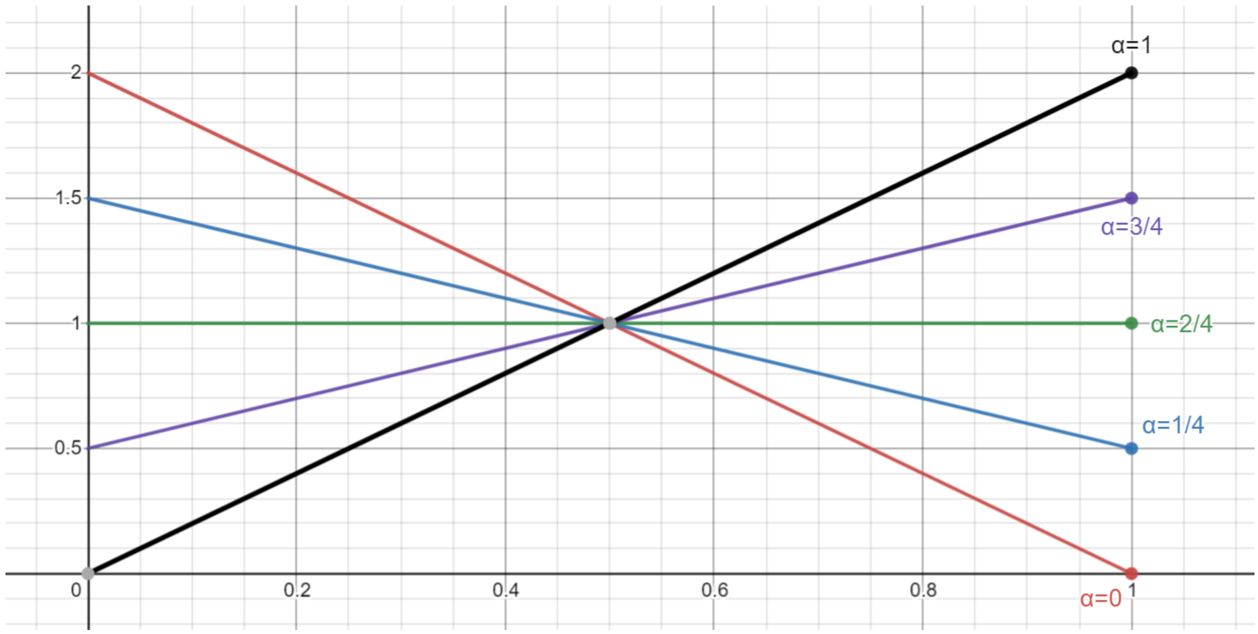

Figure 2: Posterior Distributions

We observe from Figure 2 that, with the increase of α, the graphs of posterior distributions $p(\theta|\hat{X}=1)$ are skewed from 0 to 1.

- For $\alpha > \frac{1}{2}$, the camera reports back relatively true outcome of a coin flip. And the result value is perfectly correct when $\alpha = 1$.
- For $\alpha < \frac{1}{2}$, the camera reports back relatively false outcome of a coin flip. And the result value is exactly wrong when $\alpha = 0$.
- For $\alpha = \frac{1}{2}$, the camera cannot report back any useful information of a coin flip.

Question 3

a) The cdf for X can be derived as following:

$$F_X(x) = P(X \leq x) \quad \dots \text{ by definition}$$
$$= \int_{-\infty}^{x} P_X(z)\,dz \quad \dots \text{ integral of pdf is cdf}$$
$$= \int_{0}^{x} P_X(z)\,dz \quad \dots\ 0 < x \leq 1$$

The cdf for Y can be derived as following:

$$F_Y(y) = P(Y \leq y) \quad \dots \text{ by definition}$$
$$= P\left(\frac{1}{X} \leq y\right) \quad \dots\ Y = \frac{1}{X}$$
$$= P\left(\frac{1}{y} \leq X\right) \quad \dots\ X, y > 0$$
$$= 1 - P\left(X < \frac{1}{y}\right) \quad \dots\ y \geq 1$$
$$= 1 - P\left(X \leq \frac{1}{y}\right) \quad \dots$$

X, y are continuous random variables, boundary value doesn't matter if being included or not

$$= 1 - \int_{0}^{\frac{1}{y}} P_X(z)\,dz$$

To get the pdf fo Y, we derivate the cdf for Y:

$$P_Y(y) = \frac{d}{dy}\left(1 - \int_{0}^{\frac{1}{y}} P_X(z)\,dz\right)$$
$$= -\frac{d}{dy}\left(\int_{0}^{\frac{1}{y}} P_X(z)\,dz\right)$$

Apply the chain rule, we have $\frac{d}{dx}\int_{a}^{u(x)} f(t)\,dt = u'(x) f(u(x))$, then

$$P_Y(y) = -\left(\frac{1}{y}\right)' P_X\left(\frac{1}{y}\right)$$

$$= -\left(y^{-1}\right)' P_X\left(\frac{1}{y}\right)$$

$$= y^{-2} P_X\left(\frac{1}{y}\right)$$

$$= \frac{1}{y^2} P_X\left(\frac{1}{y}\right)$$

Hence, we proved that $P_Y(y) = \frac{1}{y^2} PX\left(\frac{1}{y}\right)$.

b) If player's guess m ~~was~~ is lower than c: ($m<c$, $m \in [1,\infty)$)

player wins $m-1$, win probabilty $\int_m^{\infty} P_Y(y)dy$

If player's guess m is higher than c: ($m>c$, $m \in [1,\infty)$)

player losses 1, loss probability $\int_1^m P_Y(y)dy$

We compute the expected profit as

$$\mathbb{E} = (m-1)\int_m^{\infty} P_Y(y)dy - \int_1^m P_Y(y)dy$$

$$= m\int_m^{\infty} P_Y(y)dy - \left(\int_1^m P_Y(y)dy + \int_m^{\infty} P_Y(y)dy\right)$$

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

$$= m\int_m^{\infty} P_Y(y)dy - 1 \quad \because \text{pdf's integral is } 1$$

Hence, the answer is $m\int_m^{\infty} P_Y(y)dy - 1$,

since we proved $P_Y(y) = \frac{1}{y^2} PX\left(\frac{1}{y}\right)$ in a),

it can be expressed as $m\int_m^{\infty} \frac{1}{y^2} P_X\left(\frac{1}{y}\right) dy - 1$.

c) Take the equation of $P_X(x)$ into the conclusion of b), we derive

expected profit $\mathbb{E} = m\int_m^\infty \frac{1}{y^2}\,dy - 1$

$= m\cdot\frac{1}{m} - 1$ ... $\int_m^\infty \frac{1}{y^2}dy = \frac{1}{m}$, calculated by WolframAlpha

$= 0$

Hence, no matter what strategy the player uses, the expected profit is always 0.

d) Answer: $P_X(x) = -\ln(10)\cdot\log(x)$, $x\in(0,1]$.

my idea of finding $P_X(x)$:

① $\int_0^1 P_X(x)\,dx = 1$ and $P_X(x)\geq 0$ (Basic)

② Since X is extremely biased towards small values, and $P_X(x)\geq 0$, I think $P_X(x)$ should be monotonically decreasing from ∞ to small values when $0<X\leq 1$.

$\Rightarrow$ I first thought of $-\log(x)$, but its integral over $(0,1]$ is not 1, so I multiply it by $\ln(10)$.

$\Rightarrow$ Finally, I derive $P_X(x) = -\ln(10)\cdot\log(x)$.

Proof of validity: $\int_0^1 P_X(x)\,dx = \int_0^1 [-\ln(10)\cdot\log_{10}(x)]\,dx = 1$ (WolframAlpha)

$P_X(x)\,dx = -\ln(10)\log(x)\geq 0$ for $0<x\leq 1$.

$\mathbb{E} = m\int_m^\infty \frac{1}{y^2}P_X\left(\frac{1}{y}\right)dy - 1 = m\int_m^\infty \frac{1}{y^2}\left(-\ln(10)\log\left(\frac{1}{y}\right)dy\right) - 1 = \ln(m)$ (WolframAlpha)

For any $B>0$, $\exists\, m\geq e^B$, then $\mathbb{E}\geq \ln(e^B) = B$, expected profit is at least B.

(eg: $m = 2\cdot e^B$) $\Rightarrow E = \ln(2e^B) = \ln(2)+B \geq B$