1-a) $p(x_1, \cdots, x_N | \mu) = p(x_1|\mu) \cdot \cdots \cdot p(x_N|\mu) = \prod_{i=1}^{N} p(x_i|\mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

Since log is monotonically increasing, let's maximize log-likelihood.

$$\log(p(x_1, \cdots, x_N|\mu)) = \sum_{i=1}^{N} \left(-\frac{(x_i-\mu)^2}{2\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)\right)$$

$$\frac{d}{d\mu}(\log(p(x_1, \cdots, x_N|\mu))) = \sum_{i=1}^{N} \frac{(x_i-\mu)}{\sigma^2}$$

Maximized when the value of the derivative is 0.

$$\therefore 0 = \sum_{i=1}^{N} \frac{(x_i-\mu)}{\sigma^2}, \quad N\mu = \sum_{i=1}^{N} x_i, \quad \mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\therefore \boxed{\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i}$$

1-b) $p(\mu | x_1, \cdots, x_N) = \frac{p(x_1, \cdots, x_N|\mu) \cdot p(\mu)}{p(x_1, \cdots, x_N)}$ (by bayes' rule)

from a), $p(x_1, \cdots, x_N|\mu) = \prod_{i=1}^{N}\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)$

from the condition, $p(\mu) = \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(\mu-\nu)^2}{2\beta^2}}$ $(\mu \sim N(\nu, \beta^2))$

$p(x_1, \cdots, x_N)$ is a constant, let's say C

$$p(\mu | x_1, \cdots, x_N) = \frac{1}{C}\cdot\left(\frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(\mu-\nu)^2}{2\beta^2}}\right) \cdot \prod_{i=1}^{N}\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)$$

$$\log(p(\mu|x_1, \cdots, x_N)) = -\log(C\sqrt{2\pi\beta^2}) - \frac{(\mu-\nu)^2}{2\beta^2} + \sum_{i=1}^{N}\left(-\frac{(x_i-\mu)^2}{2\sigma^2} - \log(\sqrt{2\pi\sigma^2})\right)$$

(Continued in next page)

$$\log(p(\mu|x_1,\cdots,x_N)) = -\log(C\sqrt{2\pi\beta^2}) - \frac{(\mu-\nu)^2}{2\beta^2} + \sum_{i=1}^{N}\left(-\frac{(x_i-\mu)^2}{2\sigma^2} - \log(\sqrt{2\pi\sigma^2})\right)$$

take partial derivatives to μ,

$$\frac{\partial}{\partial\mu}\log(p(\mu|x_1,\cdots,x_N)) = -\frac{1}{\beta^2}(\mu-\nu) + \sum_{i=1}^{N}\left(\frac{1}{\sigma^2}(x_i-\mu)\right)$$

as we've done in a), the value should be 0 to maximize

$$\therefore \frac{1}{\beta^2}(\mu-\nu) = \sum_{i=1}^{N}\left(\frac{1}{\sigma^2}(x_i-\mu)\right) \Leftrightarrow \frac{\sigma^2}{\beta^2}(\mu-\nu) = \sum_{i=1}^{N}x_i - N\mu$$

$$\Leftrightarrow \mu\left(\frac{\sigma^2}{\beta^2}+N\right) = \frac{\sigma^2}{\beta^2}\nu + \sum_{i=1}^{N}x_i$$

$$\therefore \hat{\mu} = \frac{1}{\left(\frac{\sigma^2}{\beta^2}+N\right)}\left(\frac{\sigma^2}{\beta^2}\nu + \sum_{i=1}^{N}x_i\right) = \frac{\sigma^2\nu}{\sigma^2+N\beta^2} + \frac{\beta^2\sum_{i=1}^{N}x_i}{\sigma^2+N\beta^2}$$

$$= \boxed{\frac{\sigma^2\nu + \beta^2\sum_{i=1}^{N}x_i}{\sigma^2+N\beta^2}}$$

1-c) from a), b)

$$\hat{\mu}_{MLE} = \frac{1}{N}\sum_{i=1}^{N}x_i$$

$$\hat{\mu}_{MAP} = \frac{\sigma^2\nu + \beta^2\sum_{i=1}^{N}x_i}{\sigma^2+N\beta^2} = \frac{\frac{\sigma^2\nu}{N} + \beta^2\frac{1}{N}\sum_{i=1}^{N}x_i}{\frac{\sigma^2}{N}+\beta^2}$$

as $N\to\infty$, $\hat{\mu}_{MAP} \to \frac{1}{N}\sum_{i=1}^{N}x_i$, which is $\hat{\mu}_{MLE}$.

$\therefore$ as $N\to\infty$, $\hat{\mu}_{MAP}$ and $\hat{\mu}_{MLE}$ become same value.

2 minimizing $p(x \neq f(x+w))$ is maximizing $p(x = f(x+w))$, obviously.

So, I will find $f: \mathbb{R} \to \{0,1\}$ that maximizes $p(x = f(x+w))$

since x can have only two values, 0, 1, we can write as following ;

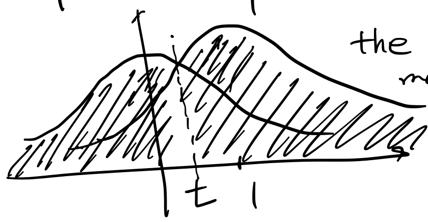
$$p(x = f(x+w)) = p(x=0) \cdot p(f(w)=0) + p(x=1) \cdot p(f(w+1)=1)$$

$$= q \cdot p(f(w)=0) + (1-q) \cdot p(f(w+1)=1)$$

$$= q \cdot p(w \in \{w \mid f(w)=0\}) + (1-q) \cdot p(w \in \{w \mid f(w+1)=1\})$$

It is maximized when

$$f(w) = \begin{cases} 0 & w < t \\ 1 & w \geq t \end{cases}$$

(∵  the area is maximized like this)

for t s.t $q \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} = (1-q) \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-1)^2}{2\sigma^2}}$

$$\therefore \frac{1-q}{q} = e^{\frac{1}{2\sigma^2}((t-1)^2 - t^2)} = e^{\frac{1}{2\sigma^2}(1-2t)}$$

$$\ln \frac{1-q}{q} = \frac{1}{2\sigma^2}(1-2t), \quad 2t = 1 - 2\sigma^2 \ln \frac{1-q}{q}$$

$$t = \frac{1}{2} - \sigma^2 \ln \frac{1-q}{q}$$

$$\therefore \boxed{f(w) = \begin{cases} 0 & w < t \\ 1 & w \geq t \end{cases} \quad \text{for } t = \frac{1}{2} - \sigma^2 \ln \frac{1-q}{q}}$$

3 $p(b=x|y=a) = \dfrac{p(y=a|b=x)\cdot p(b=x)}{p(y=a)}$ for $x=\{0,1\}$

Let $f(x|\mu,\sigma^2)$ be a pdf of $N(\mu,\sigma^2)$

when $x=0$, $p(b=0|y=a) = \dfrac{p(y=a|b=0)\cdot p(b=0)}{p(y=a)}$

$$= \frac{f(a|0.1,0.1)\cdot 0.2}{p(y=a)}$$

$$\left(\begin{aligned} *\,p(y=a) &= p(y=a|b=0)\cdot p(b=0) + p(y=a|b=1)\cdot p(b=1) \\ &= f(a|0.1,0.1)\cdot 0.2 + f(a|1.1,1.1)\cdot 0.8 \\ &= \frac{1}{5}\big(f(a|0.1,0.1) + 4f(a|1.1,1.1)\big) \end{aligned} \right)$$

$\therefore\ p(b=0|y=a) = \dfrac{1}{5p(y=a)} f(a|0.1,0.1)$

when $x=1$, $p(b=1|y=a) = \dfrac{p(y=a|b=1)\cdot p(b=1)}{p(y=a)}$

$$= \frac{f(a|1.1,1.1)\cdot 0.8}{p(y=a)}$$

$$= \frac{4}{5p(y=a)} f(a|1.1,1.1)$$

Then, we can say

$$\hat{b}(a) = \begin{cases} 1 & p(b=1|y=a) \geq p(b=0|y=a) \\ 0 & p(b=1|y=a) < p(b=0|y=a) \end{cases}$$

let's calculate the condition

continued in next page

$$P(y=a) = \frac{1}{5}\left(f(a|0.1,0.1) + 4f(a|1.1,1.1)\right)$$

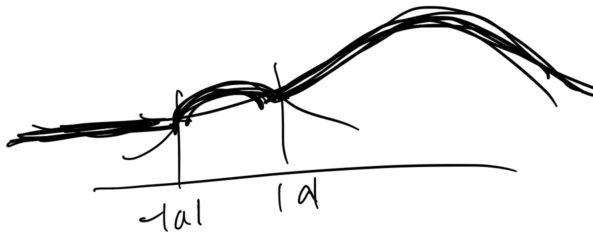
let $$g(a) = p(b=1|y=a) - p(b=0|y=a)$$
$$= \frac{1}{5p(y=a)}\left(4f(a|1.1,1.1) - f(a|0.1,0.1)\right)$$
$$= \frac{4f(a|1.1,1.1) - f(a|0.1,0.1)}{4f(a|1.1,1.1) + f(a|0.1,0.1)}$$
$$= 1 - \frac{2f(a|0.1,0.1)}{4f(a|1.1,1.1)+f(a|0.1,0.1)}.$$

$$g(a)=0 \iff 2f(a|0.1,0.1) = 4f(a|1.1,1.1) + f(a|0.1,0.1)$$
$$\iff f(a|0.1,0.1) = 4f(a|1.1,1.1)$$

$$\frac{1}{\sqrt{\cancel{2\pi}\cdot 0.1}} e^{-\frac{(a-0.1)^2}{0.2}} = \frac{4}{\sqrt{\cancel{2\pi}\cdot 1.1}} e^{-\frac{(a-1.1)^2}{2.2}}$$

$$\cancel{\sqrt{10}}\, e^{-5(a-0.1)^2} = 4\cancel{\sqrt{\frac{10}{11}}}\, e^{-\frac{5}{11}(a-1.1)^2}$$

$$\frac{\sqrt{11}}{4} = e^{\frac{5}{11}\left(11(a-0.1)^2 - (a-1.1)^2\right)}$$
$$= e^{\frac{5}{11}\left(11a^2 - \cancel{0.2\cdot 11a} + 11\cdot 0.01 - a^2 - 1.21 + \cancel{2.2a}\right)}$$
$$= e^{\frac{5}{11}(10a^2 - 1.1)}$$



$$\frac{5}{11}(10a^2 - 1.1) = \ln\sqrt{\frac{11}{16}}$$

$$\frac{5}{11}10a^2 = \frac{5}{11}\times\frac{11}{10} + \frac{1}{2}\ln\frac{11}{16} = \frac{1}{2}\left(1+\ln\frac{11}{16}\right)$$

$$\frac{100}{11}a^2 = \left(1+\ln\frac{11}{16}\right),\quad a = \pm\frac{1}{10}\sqrt{11+11\ln\frac{11}{16}}$$

$$\therefore\ \boxed{\hat{b}(a) = \begin{cases} 1 & |a| \geq \frac{1}{10}\sqrt{11+11\cdot\ln\frac{11}{16}} \\ 0 & |a| < \frac{1}{10}\sqrt{11+11\cdot\ln\frac{11}{16}} \end{cases}}$$

4-a) to win a prize by changing first choice,
I should choose wrong door at first and choose right door at second chance between n-2 doors

$\therefore$ the probability is $\frac{n-1}{n} \cdot \frac{1}{n-2} = \boxed{\frac{n-1}{n(n-2)}}$

4-b) to win a prize by not changing first choice,
I should choose right answer at first.

$\therefore$ the probability is $\boxed{\frac{1}{n}}$

5-a) Let S be a situation where an unknown panda gives birth to twins.
Let Twins be a situation where that unknown panda gives birth to twins again.

Then, $p(S) = p(S, A) + p(S, B)$

$$= p(S|A)\cdot p(A) + p(S|B)\cdot p(B)$$

$$= 0.1 \times \frac{1}{2} + 0.2 \times \frac{1}{2} = \frac{1}{20} + \frac{2}{20} = \frac{3}{20}$$

$\therefore$ $p(\text{Twins}|S) = p(\text{Twins}, A|S) + p(\text{twins}, B|S)$

$$= p(\text{Twins}|A, S)\cdot p(A|S) + p(\text{Twins}|B, S)\cdot p(B|S)$$

($\because$ S and Twins are independent)

$$= p(\text{Twins}|A)\cdot \frac{p(S|A)\cdot p(A)}{p(S)} + p(\text{Twins}|B)\cdot \frac{p(S|B)\cdot p(B)}{p(S)}$$

$$= 0.1 \times \frac{0.1 \times \frac{1}{2}}{\frac{3}{20}} + 0.2 \times \frac{0.2 \times \frac{1}{2}}{\frac{3}{20}}$$

$$= \frac{20}{3}\cdot\left(\frac{0.01}{2} + \frac{0.04}{2}\right) = \frac{20}{3}\cdot\frac{0.05}{2} = \frac{10}{3}\cdot\frac{5}{100} = \boxed{\frac{1}{6}}$$

5-b) Let S' be a situation where an unknown panda give birth to twins and then, give birth to a single infant

$$p(S') = p(S', A) + p(S', B)$$

$$= p(S'|A)\cdot p(A) + p(S'|B)\cdot p(B)$$

$$= 0.1 \times 0.9 \times \frac{1}{2} + 0.2 \times 0.8 \times \frac{1}{2} = \frac{25}{200} = \frac{5}{40} = \frac{1}{8}$$

$$p(A|S') = \frac{p(S'|A)\cdot p(A)}{p(S')} = \frac{0.1 \times 0.9 \times \frac{1}{2}}{\frac{1}{8}} = 4 \times \frac{1}{10} \times \frac{9}{10} = \frac{36}{100}$$

$$= \boxed{\frac{9}{25}}$$

6 From the definition of the conjugate prior,

if prior $p(\theta)$ is in same family with posterior $p(\theta|x)$, $p(\theta)$ is called the conjugate prior for likelihood $p(x|\theta)$

likelihood of binomial: $p(x|\theta) = \binom{n}{x} \theta^{n} (1-\theta)^{n-x}$

prior of beta : $p(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$

$$p(\theta|x) \propto p(x|\theta) \cdot p(\theta|\alpha,\beta)$$

$$\propto \binom{n}{x}\theta^{n}(1-\theta)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\propto \theta^{n+\alpha-1} (1-\theta)^{n+\beta-x-1}$$

$$= \text{beta}(n+\alpha,\ n+\beta-x)$$

$\therefore$ beta is the conjugate prior of binomial distribution.

The posterior beta's effect is that if it becomes a prior succeedingly, it still is a beta and so on, which makes it easier to handle the distribution.

2 (MC)

let $\hat{I}$ be an unbiased estimator for $\int_a^b f(x)dx$

$\int_a^b f(x)dx = (b-a)E(f(X))$ from the hint.

Since 'unbiased', $E(f(X)) \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)$ with x_i is sample of $X \sim U(a,b)$

$$\therefore \boxed{\hat{I} = \frac{(b-a)}{N}\sum_{i=1}^{N} f(x_i)}$$