HW1 CS4501-SLGM Rishi Shah 9/7/22

① a) $H(X) = E\left[\log_2\left(\frac{1}{p(x)}\right)\right] = \sum_{x \in X} p(x) \log_2\left(\frac{1}{p(x)}\right)$

$= \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(4) + \frac{1}{8}\log_2(8) + \frac{1}{16}\log_2(16) + \frac{1}{16}\log_2(64)$

$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{4} + \frac{3}{8} = \boxed{2}$

b) $D_{KL}(q||p) = \sum_{x \in X} q(x) \log \frac{q(x)}{p(x)}$

$= \frac{1}{3}\log_2\left(\frac{2}{3}\right) + \frac{2}{3}\log_2\left(\frac{8}{3}\right) + 0\log_2(0)$

$= \frac{1}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{64}{9}\right) = \frac{1}{3}\log_2\left(\frac{128}{27}\right)$

$= \frac{1}{3}\log_2(128) - \frac{1}{3}\log_2(27) = \frac{1}{3}\log_2(128) - \log_2(3)$

$= \frac{7}{3} - \log_2(3) \approx \underline{.7484}$

$D_{KL}(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$

$= \frac{1}{2}\log\left(\frac{3}{2}\right) + \frac{1}{4}\log\left(\frac{3}{8}\right) + \frac{1}{4}\log\left(\frac{1}{0}\right) = \infty$

$\therefore D_{KL}(q||p) \neq D_{KL}(p||q)$

② a) ii)

$p(y\|x)$	$x=0$	$x=1$
$y=0$	3/5	1/5
$y=1$	1/5	2/5
$y=2$	1/5	2/5

$f^{**}(0) = E[Y|X=0]$

$= 0 \cdot \frac{3}{5} + 1 \cdot \frac{1}{5} + 2 \cdot \frac{1}{5} = \frac{3}{5}$

$f^{**}(1) = E[Y|X=1]$

$= 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} + 2 \cdot \frac{2}{5} = \frac{6}{5}$

$\boxed{f^{**}(0) = 3/5 \quad f^{**}(1) = 6/5}$

For $L(y, f(x)) = (y - f(x))^2$,

$f^{**}(x) = \underset{f}{\arg\min}\, E[L(Y, f(x))] = E[Y|X=x]$ given $p_{X,Y}(x,y)$

ii) $E[L(Y,f(x))] = p(0,0)L(0,f(0)) + p(0,1)L(1,f(0)) + p(0,2)\cdot L(2,f(0)) + p(1,0)L(0,f(1)) + p(1,1)L(1,f(1)) + p(1,2)L(2,f(1))$

$= \frac{3}{10}(0-\frac{3}{5})^2 + \frac{1}{10}(1-\frac{3}{5})^2 + \frac{1}{10}(2-\frac{3}{5})^2 + \frac{1}{10}(0-\frac{6}{5})^2 + \frac{2}{10}(1-\frac{6}{5})^2 + \frac{2}{10}(2-\frac{6}{5})^2$

$= \frac{1}{10}\left(\frac{27}{25} + \frac{4}{25} + \frac{49}{25} + \frac{36}{25} + \frac{2}{25} + \frac{32}{25}\right)$

$= \frac{1}{250}(27+4+49+36+2+32) = \frac{150}{250} = \frac{3}{5} = \boxed{.6}$

b) i) For 0-1 Loss, $f^{**}(x) = \underset{f}{\arg\min}\, E[L(Y,f(x))] = \underset{y}{\arg\max}\, p(Y=y \mid X=x)$

$f^{**}(0) = \underset{y}{\arg\max}\, p(Y=y \mid X=0)$

$\max(\frac{3}{5}, \frac{1}{5}, \frac{1}{5}) = \frac{3}{5} = p(Y=0 \mid X=0) \Rightarrow f^{**}(0) = 0$

$f^{**}(1) = \underset{y}{\arg\max}\, p(Y=y \mid X=1)$

$\max(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}) = \frac{2}{5} = p(Y=1 \mid X=0) = p(Y=2 \mid X=0)$

We will arbitrarily choose: $Y=1 \Rightarrow f^{**}(1) = 1$

$\boxed{f^{**}(0)=0, \quad f^{**}(1)=1}$

ii) $E[L(Y,f(x))] = p(0,0)L(0,f(0)) + p(0,1)L(1,f(0)) + p(0,2)L(2,f(0)) + p(1,0)L(0,f(1)) + p(1,1)L(1,f(1)) + p(1,2)L(2,f(1))$

$= \frac{3}{10}(0) + \frac{1}{10}(1) + \frac{1}{10}(1) + \frac{1}{10}(1) + \frac{2}{10}(0) + \frac{2}{10}(0) = \frac{1}{2} = \boxed{.5}$

(3) Y: $\frac{1}{6}$ chance of rolling 1,2,3,4,5,6

$$P_{X,Y}(x,y) = P_{X|Y}(x|y)\cdot P_Y(y)$$

$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)} = \frac{P_{X|Y}(x|y)\cdot P_Y(y)}{P_X(x)} = \frac{P_{X|Y}(x|y)\cdot P_Y(y)}{\sum_{y_i=1}^{6} P_{X|Y}(x|y_i)\,P_Y(y_i)}$$

$$= \frac{P_{X|Y}(x|y)}{\sum_{y_i=1}^{6} P_{X|Y}(x|y_i)}$$ since $P_Y(y) = P_Y(y_i)\ \forall y, y_i \in [1,6]$

Binomial Dist!

$$\sum_{y_i=1}^{6} P_{X|Y}(x|y_i) = \begin{cases} \frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\frac{1}{64}, & x=0 \\ \frac{1}{2}+\frac{1}{2}+\frac{3}{8}+\frac{1}{4}+\frac{5}{32}+\frac{3}{32}, & x=1 \\ 0+\frac{1}{4}+\frac{3}{8}+\frac{3}{8}+\frac{5}{16}+\frac{15}{64}, & x=2 \\ 0+0+\frac{1}{8}+\frac{1}{4}+\frac{5}{16}+\frac{5}{16}, & x=3 \\ 0+0+0+\frac{1}{16}+\frac{5}{32}+\frac{15}{64}, & x=4 \\ 0+0+0+0+\frac{1}{32}+\frac{3}{32}, & x=5 \\ 0+0+0+0+0+\frac{1}{64}, & x=6 \end{cases}$$

$$\begin{cases} 63/64, & x=0 \\ 15/8, & x=1 \\ 99/64, & x=2 \\ 1, & x=3 \\ 29/64, & x=4 \\ 1/8, & x=5 \\ 1/64, & x=6 \end{cases}$$

$P_{X\|Y}(x\|y)$	x=0	x=1	x=2	x=3	x=4	x=5	x=6
y=1	1/2	1/2	0	0	0	0	0
y=2	1/4	1/2	1/4	0	0	0	0
y=3	1/8	3/8	3/8	1/8	0	0	0
y=4	1/16	1/4	3/8	1/4	1/16	0	0
y=5	1/32	5/32	5/16	5/16	5/32	1/32	0
y=6	1/64	3/32	15/64	5/16	15/64	3/32	1/64

$P_{Y\|X}(y\|x)$	x=0	x=1	x=2	x=3	x=4	x=5	x=6
y=1	32/63	4/15	0	0	0	0	0
y=2	16/63	4/15	16/99	0	0	0	0
y=3	8/63	1/5	8/33	1/8	0	0	0
y=4	4/63	2/15	8/33	1/4	4/29	0	0
y=5	2/63	1/12	20/99	5/16	10/29	1/4	0
y=6	1/63	1/20	5/33	5/16	15/29	3/4	1

$$\boxed{f^{**}(x) = \begin{cases} 1, & x=0 \\ 2, & x=1 \\ 3, & x=2 \\ 5, & x=3 \\ 6, & x=4 \\ 6, & x=5 \\ 6, & x=6 \end{cases}}$$

For 0-1 Loss, $f^{**}(x) = \arg\max_y p(Y=y|X=x)$

④ $f^{**}(x) \doteq E[Y \mid X=x]$ for squared error loss given $p(Y=y, X=x)$

$$f^{**}(0) = 1\left(\frac{32}{63}\right) + 2\left(\frac{16}{63}\right) + 3\left(\frac{8}{63}\right) + 4\left(\frac{4}{63}\right) + 5\left(\frac{2}{63}\right) + 6\left(\frac{1}{63}\right)$$

$$= \frac{1}{63}\left(32+32+24+16+10+6\right) = \frac{120}{63} = \frac{40}{21} \approx 1.905$$

$$f^{**}(1) = 1\left(\frac{4}{15}\right) + 2\left(\frac{4}{15}\right) + 3\left(\frac{1}{5}\right) + 4\left(\frac{2}{15}\right) + 5\left(\frac{1}{12}\right) + 6\left(\frac{1}{20}\right)$$

$$= \frac{4+8+9+8}{15} + \frac{5}{12} + \frac{3}{10} = \frac{29}{15} + \frac{5}{12} + \frac{3}{10}$$

$$= \frac{1}{180}\left(348+75+54\right) = \frac{477}{180} = \frac{159}{60} = \frac{53}{20} = 2.65$$

$$f^{**}(2) = 2\left(\frac{16}{99}\right) + 3\left(\frac{8}{33}\right) + 4\left(\frac{8}{33}\right) + 5\left(\frac{20}{99}\right) + 6\left(\frac{5}{33}\right)$$

$$= \frac{32+100}{99} + \frac{24+32+30}{33} = \frac{132}{99} + \frac{86}{33} = \frac{44}{33} + \frac{86}{33} = \frac{130}{33} \approx 3.939$$

$$f^{**}(3) = 3\left(\frac{1}{8}\right) + 4\left(\frac{1}{4}\right) + 5\left(\frac{5}{16}\right) + 6\left(\frac{5}{16}\right) = \frac{3}{8} + 1 + \frac{25}{16} + \frac{15}{8} = \frac{26}{8} + \frac{25}{16}$$

$$= \frac{52+25}{16} = \frac{77}{16} = 4.8125 \approx 4.813$$

$$f^{**}(4) = 4\left(\frac{4}{29}\right) + 5\left(\frac{10}{29}\right) + 6\left(\frac{15}{29}\right) = \frac{1}{29}\left(16+50+90\right) = \frac{156}{29} \approx 5.379$$

$$f^{**}(5) = 5\left(\frac{1}{4}\right) + 6\left(\frac{3}{4}\right) = \frac{1}{4}(5+18) = \frac{23}{4} = 5.75$$

$$f^{**}(6) = 6(1) = 6$$

$$\boxed{f^{**}(x) = \begin{cases} 1.905, & x=0 \\ 2.65, & x=1 \\ 3.939, & x=2 \\ 4.813, & x=3 \\ 5.379, & x=4 \\ 5.75, & x=5 \\ 6, & x=6 \end{cases}}$$

The predictor for the squared loss seems more intuitive as the prediction increases as x increases. This is not the case for the 0-1 loss predictor as $f^{**}(4) = f^{**}(5) = f^{**}(6) = 6$. However, the 0-1 loss ~~[illegible]~~ is computationally much simpler. Both are reasonable, but I think the extra computations for the squared loss are worth it.