ML Assn 1 20160463 [name in Korean]

1. (a)

x	1	2	3	4	5
$p(x)$	0.16	0.17	0.11	0.22	0.34

y	1	2	3
$p(y)$	0.26	0.47	0.27

$$P(X=1) = P(X=1, Y=1) + P(X=1, Y=2) + P(X=1, Y=3) = 0.01+0.05+0.1 = 0.16$$
$$P(X=2) = P(X=2, Y=1) + P(X=2, Y=2) + P(X=2, Y=3) = 0.02+0.1+0.05 = 0.17$$
$$P(X=3) = P(X=3, Y=1) + P(X=3, Y=2) + P(X=3, Y=3) = 0.03+0.05+0.03 = 0.11$$
$$P(X=4) = 0.1+0.07+0.05 = 0.22$$
$$P(X=5) = 0.1+0.2+0.04 = 0.34$$
$$P(Y=1) = P(Y=1, X=1) + P(Y=1, X=2) + P(Y=1, X=3) + P(Y=1, X=4) + P(Y=1, X=5)$$
$$= 0.01+0.02+0.03+0.1+0.1 = 0.26$$
$$P(Y=2) = 0.05+0.1+0.05+0.07+0.2 = 0.47$$
$$P(Y=3) = 0.1+0.05+0.03+0.05+0.04 = 0.27$$

(b) $E[X] = 1\cdot P(X=1) + 2\cdot P(X=2) + 3\cdot P(X=3) + 4\cdot P(X=4) + 5\cdot P(X=5)$
$= 0.16 + 2\times0.17 + 3\times0.11 + 4\times0.22 + 5\times0.34$
$= 3.41$

$E[Y] = 1\cdot P(Y=1) + 2\cdot P(Y=2) + 3\cdot P(Y=3)$
$= 1\times0.26 + 2\times0.47 + 3\times0.27$
$= 2.01$

(c)

x	1	2	3	4	5
$p(x \mid Y=1)$	$\frac{1}{26}$	$\frac{2}{26}$	$\frac{3}{26}$	$\frac{10}{26}$	$\frac{10}{26}$

y	1	2	3
$p(y \mid X=3)$	$\frac{3}{11}$	$\frac{5}{11}$	$\frac{3}{11}$

$$P(X=1 \mid Y=1) = \frac{P(X=1, Y=1)}{P(Y=1)} = \frac{0.01}{0.01+0.02+0.03+0.1+0.1} = \frac{0.01}{0.26} = \frac{1}{26}$$
$$P(X=2 \mid Y=1) = \frac{P(X=2, Y=1)}{P(Y=1)} = \frac{0.02}{0.26} = \frac{2}{26}$$
$$P(X=3 \mid Y=1) = \frac{0.03}{0.26} = \frac{3}{26} \quad P(X=4 \mid Y=1) = \frac{0.1}{0.26} = \frac{10}{26} \quad P(X=5 \mid Y=1) = \frac{0.1}{0.26} = \frac{10}{26}$$
$$P(Y=1 \mid X=3) = \frac{P(Y=1, X=3)}{P(X=3)} = \frac{0.03}{0.03+0.05+0.03} = \frac{0.03}{0.11} = \frac{3}{11}$$
$$P(Y=2 \mid X=3) = \frac{P(Y=2, X=3)}{P(X=3)} = \frac{0.05}{0.11} = \frac{5}{11} \quad P(Y=3 \mid X=3) = \frac{P(Y=3, X=3)}{P(X=3)} = \frac{0.03}{0.11} = \frac{3}{11}$$

(d) $E[X \mid Y=1] = 1\cdot P(X=1 \mid Y=1) + 2\cdot P(X=2 \mid Y=1) + 3\cdot P(X=3 \mid Y=1) + 4\cdot P(X=4 \mid Y=1) + 5\cdot P(X=5 \mid Y=1)$
$= \frac{1\cdot1+2\cdot2+3\cdot3+4\cdot10+5\cdot10}{26} = \frac{104}{26}$

$E[Y \mid X=3] = 1\cdot P(Y=1 \mid X=3) + 2\cdot P(Y=2 \mid X=3) + 3\cdot P(Y=3 \mid X=3) = \frac{1\cdot3+2\cdot5+3\cdot3}{11} = \frac{22}{11} = 2$

2. event

T : test is positive $\quad T^c$: test is negative

D : you have the disease $\quad D^c$: you don't have the disease

$$P(T|D) = 0.99 \qquad P(T^c|D^c) = 0.99 \qquad P(D) = 0.0001$$

$$P(T^c|D^c) = \frac{P(T^c, D^c)}{P(D^c)} = \frac{P(D^c) - P(T, D^c)}{P(D^c)} \quad (\because T, T^c \text{ is a partition of sample space})$$

$$= 1 - \frac{P(T, D^c)}{P(D^c)} = 1 - \frac{P(T) - P(T, D)}{P(D^c)} \quad (\because D, D^c \text{ is a partition of sample space})$$

$$= 1 - \frac{P(T) - P(T|D) \cdot P(D)}{1 - P(D)} = 1 - \frac{P(T) - 0.99 \times 0.0001}{0.9999} = 0.99$$

$$\therefore P(T) = 0.01 \times 0.9999 + 0.99 \times 0.0001 = 0.010098$$

$$\therefore P(D|T) = \frac{P(D, T)}{P(T)} = \frac{P(T|D) P(D)}{P(T)} = \frac{0.99 \times 0.0001}{0.010098} = 0.0098039\ldots\ldots$$

chances that you actually have disease : about 0.0098......

3. (a) since it has finite supports, X, Y is a discrete random variable

$$E[E[X|Y]] = E[E[X|Y=y]] = \sum_y E[X|Y=y] P(Y=y)$$

$$= \sum_y \sum_x x P(X=x|Y=y) P(Y=y) = \sum_x x \sum_y P(X=x|Y=y) P(Y=y)$$

$$= \sum_x x P(X=x) \quad (\because \text{law of total probability, partition})$$

$$= E[X]$$

(b) $$\mathrm{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = \sum_x \sum_y (x - E[X])(y - E[Y]) f(x, y)$$

$$= \sum_x \sum_y (xy - yE[X] - xE[Y] + E[X]E[Y]) f(x, y)$$

$$= \sum_x \sum_y xy f(x, y) - E[X] \sum_x \sum_y y f(x, y) - E[Y] \sum_x \sum_y x f(x, y) + E[X]E[Y] \sum_x \sum_y f(x, y)$$

$$= E[XY] - E[X]E[Y] - \cancel{E[Y]E[X]} + \cancel{E[X]E[Y]}$$

$$= E[XY] - E[X]E[Y]$$

4\. (a) $\hat{\theta}_{MLE} = \arg\max_\theta \log(p(D|\theta)) = \arg\max_\theta \sum_{i=1}^{N} \log(p(x_i|\theta))$

$= \arg\max_\theta \sum_{i=1}^{N} \log(\theta^{x_i}(1-\theta)^{1-x_i}) = \arg\max_\theta \sum_{i=1}^{N}(x_i \log\theta + (1-x_i)\log(1-\theta))$

$= \arg\max_\theta (\log\theta * N\bar{x} + \log(1-\theta) * (N - N\bar{x}))$

$$\frac{d}{d\theta}(\log\theta * N\bar{x} + \log(1-\theta) * (N-N\bar{x})) = \frac{N\bar{x}}{\theta} - \frac{N-N\bar{x}}{1-\theta} = 0$$

$N\bar{x}(1-\theta) - (N-N\bar{x})\theta = 0$

$N\bar{x} - \cancel{N\bar{x}\theta} - N\theta + \cancel{N\bar{x}\theta} = 0$

$\therefore \theta = \bar{x} = \hat{\theta}_{MLE}$

(b) (i) bias $= E(\hat{\theta}_{MLE}) - \theta = E(\frac{1}{N}\sum_{i=1}^{N} x_i) - \theta = \frac{1}{N}\sum_{i=1}^{N}\overset{\theta}{\boxed{E(x_i)}} - \theta = \frac{1}{N}N\theta - \theta = 0$

(ii) variance $= Var(\hat{\theta}_{MLE}) = Var(\frac{1}{N}\sum_{i=1}^{N} x_i) = \frac{1}{N^2}\sum_{i=1}^{N}\boxed{Var(x_i)} = \frac{1}{N^2}\sum_{i=1}^{N}\theta(1-\theta) = \frac{1}{N^2}N\theta(1-\theta) = \frac{1}{N}\theta(1-\theta)$

(iii) MSE $= E[(\hat{\theta}_{MLE} - \theta)^2] = \text{bias}^2 + \text{variance} = 0 + \frac{\theta(1-\theta)}{N} = \frac{\theta(1-\theta)}{N}$

(c) $p(\theta|D) \propto p(D|\theta)p(\theta) \propto \theta^{N\bar{x}}(1-\theta)^{N-N\bar{x}} \times \theta^{v-1}(1-\theta)^{v-1} = \theta^{v-1+N\bar{x}}(1-\theta)^{v-1+N-N\bar{x}}$

(removing the constants regarding θ)

$\therefore$ posterior $p(\theta|D) \propto \theta^{v-1+N\bar{x}}(1-\theta)^{v-1+N-N\bar{x}}$, which is $Beta(v+N\bar{x}, v+N-N\bar{x})$ distribution

$\therefore \hat{\theta}_{BE} = E(\theta|D) =$ mean of Beta distribution $= \frac{v+N\bar{x}}{v+\cancel{N\bar{x}}+v+N-\cancel{N\bar{x}}} = \frac{v+N\bar{x}}{2v+N}$

(d)(i) bias $= E(\hat{\theta}_{BE}) - \theta = E(\frac{v+N\bar{x}}{2v+N}) - \theta = \frac{1}{2v+N}(v + \sum_{i=1}^{N}E(x_i)) - \theta = \frac{v+N\theta}{2v+N} - \theta$

(ii) variance $= Var(\hat{\theta}_{BE}) = Var(\frac{v+N\bar{x}}{2v+N}) = \frac{1}{(2v+N)^2}(\underbrace{Var(v)}_{=0} + Var(N\bar{x})) = \frac{1}{(2v+N)^2}\sum_{i=1}^{N}Var(x_i) = \frac{N\theta(1-\theta)}{(2v+N)^2}$

(iii) MSE $= E[(\hat{\theta}_{BE} - \theta)^2] = \text{bias}^2 + \text{variance} = (\frac{v+N\theta}{2v+N} - \theta)^2 + \frac{N\theta(1-\theta)}{(2v+N)^2}$

(e) (i) $\hat{\theta}_{BE} = \hat{\theta}_{MLE}$, $\frac{v+N\bar{x}}{2v+N} = \bar{x}$, $v + \cancel{N\bar{x}} = 2v\bar{x} + \cancel{N\bar{x}}$, $v(1-2\bar{x}) = 0$

$\hat{\theta}_{MLE}$ is a special case of $\hat{\theta}_{BE}$ when $v = 0$

(ii) v injects a prior belief of what θ's distribution is, by $Beta(v,v) \sim \theta$

as v gets higher we have a higher belief that θ is around 0.5