STAT 201B Homework 5

Hanze Yao
3033093286

1. (a) The parameter space Θ is the possible demand of the pairs of skies. $\therefore$ $\Theta = \{30, 40, 50, 60\}$
The possible actions A includes buying the following pairs of skies (right at the boundary of each price change).
$A = \{25, 50, 75\}$

(b) The prior distribution is $f(\theta)$
$P_f(\theta = 30) = 0.2$ $P_f(\theta = 40) = 0.4$ $P_f(\theta = 50) = 0.2$
$P_f(\theta = 60) = 0.2$

(c)

$L(\theta_i, a_j)$	$a_1 = 25$	$a_2 = 50$	$a_3 = 75$
$\theta_1 = 30$	1275	1375	1500
$\theta_2 = 40$	1325	875	1000
$\theta_3 = 50$	1375	375	500
$\theta_4 = 60$	1425	425	0

The scenario with the smallest loss is when $\theta_4 = 60$ $a_3 = 75$. Consider the loss now is 0. Then the losses of other scenarios are shown above.

(d) Bayes risk

a_1 : $r(f, a_1) = 1275 \times 0.2 + 1325 \times 0.4 + 1375 \times 0.2 + 1425 \times 0.2$
$= 1345$

a_2 : $r(f, a_2) = 1375 \times 0.2 + 875 \times 0.4 + 375 \times 0.2 + 425 \times 0.2 = 785$

a_3 : $r(f, a_3) = 1500 \times 0.2 + 1000 \times 0.4 + 500 \times 0.2 = 800$

$\because$ a_2 has the smallest Bayes risk.
$\therefore$ $a_2 = 50$ is the Bayes rule.

2. (a) $f(p \mid x) \propto f(x \mid p)\, f(p) \propto p^x (1-p)^{n-x}\, p^{\alpha-1} (1-p)^{\beta-1}$
$= p^{\alpha+x-1} (1-p)^{\beta+n-x-1}$

$\therefore$ $f(p \mid x) \sim \text{Beta}(\alpha+x, \beta+n-x)$

$r(\hat{p} \mid x) = E_{p|x}(L(p, \hat{p}(x))) = E_{p|x}((p - \hat{p}(x))^2)$
$= E_{p|x}(p^2 - 2p\hat{p}(x) + \hat{p}(x)^2) = E_{p|x}(p^2 - 2p\hat{p} + \hat{p}^2)$
$= E_{p|x}(p^2) - 2\hat{p}\, E_{p|x}(p) + \hat{p}^2$
$= \dfrac{(\alpha+x)(\alpha+x+1)}{(\alpha+\beta+n)(\alpha+\beta+n+1)} - \dfrac{2(\alpha+x)}{(\alpha+\beta+n)}\hat{p} + \hat{p}^2$

$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \dfrac{\alpha^2}{(\alpha+\beta)^2}$
$= \dfrac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$

(b) Take the first derivative of $r(\hat{p}/x)$ to $\hat{p}$.

$$\frac{d}{d\hat{p}} r(\hat{p}|x) = \frac{-2(\alpha+x)}{(\alpha+\beta+n)} + 2\hat{p} = 0$$

$$\hat{p} = \frac{\alpha+x}{\alpha+\beta+n}$$

To prove it is the global minimum value, take the second derivative.

$$\frac{d^2}{d\hat{p}^2} r(\hat{p}|x) = 2 > 0$$

$\therefore$ $\hat{p} = \frac{\alpha+x}{\alpha+\beta+n}$ minimizes $r(\hat{p}|x)$

$\because$ $\hat{p} = \frac{\alpha+x}{\alpha+\beta+n}$ is the posterior mean

$\therefore$ The Bayes estimator is also $\frac{\alpha+x}{\alpha+\beta+n}$ because of the property of $L(p, \hat{p}(x)) = (p - \hat{p}(x))^2$

(c) $\because$ $\hat{p} = \frac{\alpha+x}{\alpha+\beta+n}$

$$\therefore r(\hat{p}|x) = \frac{(\alpha+x)(\alpha+x+1)}{(\alpha+\beta+n)(\alpha+\beta+n+1)} - \frac{2(\alpha+x)}{(\alpha+\beta+n)} \times \frac{\alpha+x}{\alpha+\beta+n} + \frac{(\alpha+x)^2}{(\alpha+\beta+n)^2}$$

$$= \frac{(\alpha+x)(\alpha+x+1)(\alpha+\beta+n) - 2(\alpha+x)^2(\alpha+\beta+n+1) + (\alpha+x)^2(\alpha+\beta+n+1)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$$

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

$$= \frac{(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} \Rightarrow$$ Which is the variance of the posterior distribution.

3. Proof: The zero-one loss function is:

$$L(\theta, \hat{\theta}) = \begin{cases} 1 & \theta \neq \hat{\theta} \\ 0 & \theta = \hat{\theta} \end{cases}$$

posterior risk is: $r(\hat{\theta}|x) = \sum_{i=1}^{n} L(\theta_i, \hat{\theta}) f(\theta_i|x)$

$$= \sum_{\theta_i \neq \hat{\theta}} f(\theta_i|x) = 1 - f(\hat{\theta}|x)$$

$\because$ Bayes estimator $\hat{\theta}$ minimizes posterior risk. So $f(\hat{\theta}|x)$ is the maximum value. So $\hat{\theta}$ is the posterior mode.

4.(a) $R(\theta_1, a_1) = E_{X|\theta_1}[L(\theta_1, a_1)]$

$= P(X=0|\theta_1)\, L(\theta_1, a_1(0)) + P(X=1|\theta_1)\, L(\theta_1, a_1(1))$

$= 0.2 \times 1 + 0.8 \times 2 = 1.8$

$R(\theta_1, a_2) = E_{X|\theta_1}[L(\theta_1, a_2)]$

$= P(X=0|\theta_1)\, L(\theta_1, a_2(0)) + P(X=1|\theta_1)\, L(\theta_1, a_2(1))$

$= 0.2 \times 4 + 0.8 \times 0 = 0.8$

$R(\theta_2, a_1) = E_{X|\theta_2}[L(\theta_2, a_1)]$

$= P(X=0|\theta_2)\, L(\theta_2, a_1(0)) + P(X=1|\theta_2)\, L(\theta_2, a_1(1))$

$= 0.4 \times 3 + 0.6 \times 1 = 1.8$

$R(\theta_2, a_2) = E_{X|\theta_2}[L(\theta_2, a_2)]$

$= P(X=0|\theta_2)\, L(\theta_2, a_2(0)) + P(X=1|\theta_2)\, L(\theta_2, a_2(1))$

$= 0.4 \times 1 + 0.6 \times 4 = 2.8$

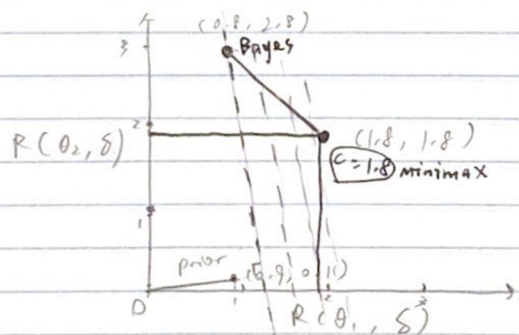
$\because \lambda \in [0, 1]$

$\therefore$ when $\lambda = 0$ $(r_1, r_2) = (R(\theta_1, a_2), R(\theta_2, a_2)) = (0.8, 2.8)$

when $\lambda = 1$ $(r_1, r_2) = (R(\theta_1, a_1), R(\theta_2, a_1)) = (1.8, 1.8)$

$\because S$ is clearly a linear function

$\therefore$ The plot of S is



$r(f, a_1) = 0.9 \times 1.8 + 0.1 \times 1.8 = 1.8$

$r(f, a_2) = 0.9 \times 0.8 + 0.1 \times 2.8 = \underbrace{1.0}_{}$ smallest

$\therefore a_2$ is the Bayes rule.

(b) The Bayes risk equals to the dot product of the prior vector and the point in the risk set. The y-intersection of the dotted lines in the graph above is proportional to the Bayes risk. So the line with the smallest Bayes risk is the line through point $(0.8, 2.8)$. So the Bayes rule is $(0.8, 2.8) \Rightarrow (R(\theta_1, a_2), R(\theta_2, a_2)) \rightarrow$ action a_2.

Those lines are perpendicular to prior vector.

(c) As for the minimax rule.

$a_1 : \max_\theta R(\theta, a_1) = \underbrace{1.8}_{}$ (smallest) $\quad a_2 : \max_\theta R(\theta, a_2) = 2.8$

Then minimax rule will be a_1. The first touch of square $Q(c)$ to Risk Set.