Chapter 5. Bayesian statistics

5.1.

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} = \frac{\sum_k p(\mathcal{D}|\theta)p(z=k)p(\theta|z=k)}{p(\mathcal{D})}$$

Since $p(\mathcal{D}|\theta) = p(\mathcal{D}|\theta, z=k)$,

$$p(\theta|\mathcal{D}) = \sum_k \frac{p(z=k)}{p(\mathcal{D})} p(\theta|z=k)p(\mathcal{D}|\theta, z=k)$$

But meanwhile, we have

$$\begin{aligned}
&p(\theta|\mathcal{D}, z=k) = \frac{p(\mathcal{D}|\theta, z=k)p(\theta|z=k)}{p(\theta|z=k)} \\
&\Rightarrow p(\theta|\mathcal{D}) = \sum_k \frac{p(z=k)p(\mathcal{D}|z=k)}{p(\mathcal{D})} p(\theta|\mathcal{D}, z=k) \\
&= \sum_k p(z=k|\mathcal{D})p(\theta|\mathcal{D}, z=k)
\end{aligned} \tag{1}$$

5.2.
a. $\rho(a|x) = p_0 L(0,a) + p_1 L(1,a) = p_0 L(0,\hat{y}) + p_1 L(1,\hat{y})$
$= p_0 \lambda_{10}\hat{y} + p_1 \lambda_{01}(1-\hat{y})$
If $\hat{y} = 0$, $\rho(a|x) = p_1\lambda_{01}$, and if $\hat{y} = 1$, $\rho(a|x) = p_0\lambda_{10}$.
Therefore, if $p_1\lambda_{01} > p_0\lambda_{10} = (1-p_1)\lambda_{10}$, we have to take $\hat{y} = 1$. Otherwise, $\hat{y} = 0$.
The condition is equivalent to $p_1 > \theta$ where

$$\theta = \frac{\lambda_{10}}{\lambda_{01} + \lambda_{10}}.$$

b.

$$0.1 = \frac{\lambda_{10}}{\lambda_{01} + \lambda_{10}} \Rightarrow \lambda_{01} = 9\lambda_{10}.$$

5.3.
a.

$$\rho(a|x) = \sum_k p(y = k|x) L(y = k, a)$$

If we choose action $i \in \{1, 2, \cdots, C\}$,

$$\rho(a|x) = \sum_{k \neq i} p(y = k|x)\lambda_s = (1 - p(y = i|x))\lambda_s$$

If we choose action $i = C + 1$,

$$\rho(a|x) = \sum_k p(y = k|x)\lambda_r = \lambda_r$$

If we choose $Y = j$, then

$$(1 - p(y = j|x)) = \min_k (1 - p(y = k|x)) \Rightarrow p(y = j|x) = \max_k p(y = k|x)$$

$$1 - p(y = j|x) \leq \frac{\lambda_r}{\lambda_s} \Rightarrow p(y = j|x) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

otherwise rejecting option (selecting $i = C + 1$) is optimal.

b. If the relative cost of rejection increases, then the base threshold for $p(y = j|x)$ would decrease.

5.4.
a. $p(y = 0|x) = 0.8$, $p(y = 1|x) = 0.2$.
$\Rightarrow$ Selecting $\hat{y} = 0$ minimizes expected loss.

b. $p(y = 0|x) = 0.6$, $p(y = 1|x) = 0.4$.
$\Rightarrow$ Rejecting option minimizes expected loss.

c. By predicting 0, $p_1 < 1 - p_1$ and $1 - p_1 > 1 - 3/10$

$$\Rightarrow p_1 < \frac{3}{10}.$$

By predicting 1, $p_1 > 1 - p_1$ and $p_1 > 1 - 3/10$

$$\Rightarrow p_1 > \frac{7}{10}.$$

By rejection options, $p_1 < 1 - 3/10$ and $1 - p_1 < 1 - 3/10$

$$\Rightarrow \frac{3}{10} < p_1 < \frac{7}{10}.$$

5.5.

$$\begin{aligned}
&E_\pi(\theta) = (P - C)Q \int_0^\infty f(D)dD + (P - C) \int_0^Q Df(D)dD \\
&- CQ \int_0^Q f(D)dD + C \int_0^Q Df(D)dD \\
&= (P - C)Q(1 - F(Q)) + P \int_0^Q Df(D)dD - CQF(Q) \\
&= (P - C)Q - PQF(Q) + P \int_0^Q Df(D)f(D).
\end{aligned} \tag{2}$$

$$\begin{aligned}
&\frac{dE_\pi(\theta)}{d\theta} = (P - C) - PF(Q) - PQf(Q) + PQf(Q) = 0 \\
&\Leftrightarrow F(Q) = \frac{P - C}{P}.
\end{aligned} \tag{3}$$

5.6.

$$B = \frac{p(D|H_1)}{p(D|H_0)} = \frac{p(H_1|D)}{p(H_0|D)} \cdot \frac{p(H_0)}{p(H_1)} \propto \frac{p(H_1|D)}{p(H_0|D)}.$$

If $p(H_0|D)$ is constant, two values would be same, otherwise, they would be different.

5.7.

$$
\begin{aligned}
&\mathbb{E}(L(\Delta, p^m)) = \int L(\Delta, p^m) p(\Delta|D) d\Delta = \int -\log p(\Delta|m, D) p(\Delta|D) d\Delta \\
&\mathbb{E}(L(\Delta, p^{BMA})) = -\int (\log p(\Delta|D)) p(\Delta|D) d\Delta \\
&\Rightarrow \mathbb{E}(L(\Delta, p^m)) - \mathbb{E}(L(\Delta, p^{BMA})) = \int p(\Delta|D) \log \frac{p(\Delta|D)}{p(\Delta|m, D)} d\Delta \\
&= \mathbb{KL}(p(\Delta, D)|p(\Delta|m, D)) \geq 0.
\end{aligned} \tag{4}
$$

5.8.
a.

$p(x,y\|\theta)$	$y=0$	$y=1$
$x=0$	$(1-\theta_1)\theta_2$	$(1-\theta_1)(1-\theta_2)$
$x=1$	$\theta_1(1-\theta_2)$	$\theta_1\theta_2$

b.

$$\hat{\theta}_1 = \frac{4}{7}, \hat{\theta}_2 = \frac{4}{7}$$

$$p(D|\hat{\theta}, M_2) = \frac{16}{49} \cdot \frac{12}{49} \cdot \frac{12}{49} \cdot \frac{12}{49} \cdot \frac{16}{49} \cdot \frac{12}{49} \cdot \frac{9}{49} \approx 0.0000704$$

c.

$$\hat{\theta}_{0,0} = \frac{2}{7}, \hat{\theta}_{0,1} = \frac{1}{7}, \hat{\theta}_{1,0} = \frac{2}{7}, \hat{\theta}_{1,1} = \frac{2}{7}$$

$$p(D|\hat{\theta}, M_4) = \frac{2}{7} \cdot \frac{2}{7} \cdot \frac{2}{7} \cdot \frac{2}{7} \cdot \frac{2}{7} \cdot \frac{2}{7} \cdot \frac{1}{7} \approx 0.0000777$$

d. For 2-parameter models,

$$L(M_2) = \log(\frac{3}{6} \cdot \frac{3}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{2}{6} \cdot \frac{3}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{3}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{2}{6} \cdot \frac{2}{6} \cdot \frac{2}{6}) \approx -12.5423$$

For 4-parameter models,

$$L(M_4) = -\infty,$$

since $\log p(x_6, y_6|m, \hat{\theta}(D_{-6})) = \log 0$.

e. $BIC(M_2, D) \approx -9.561 - \log 7 \approx -11.507$,
$BIC(M_4, D) \approx -12.381$,
Therefore, the first model fits better.

5,9.

$$\begin{aligned}
&\rho(a|x) = \int |y-a| p(y|x)dy = \int_{y>a} (y-a)p(y|x)dy - \int_{y<a} (y-a)p(y|x)dy \\
&\frac{\partial \rho}{\partial a} = -\int_{y>a} p(y|x)dy + \int_{y<a} p(y|x)dy = 0 \\
&\Leftrightarrow P(y < a|x) = P(y \geq a|x) = 0.5
\end{aligned} \tag{5}$$

5.10.

$$\begin{aligned}
&\frac{p(y=1|x)}{p(y=0|x)} > \frac{L_{FP}}{L_{FN}} = \frac{1}{c} \\
&\Leftrightarrow \frac{c}{c+1} > p(y=0|x).
\end{aligned} \tag{6}$$