Answers to PMR tutorial questions (Sheet 5)

Kian Ming Adam Chai and Chris Williams

November 6, 2008

1.

$$
\begin{aligned}
P(n_h, n_t|\theta) &= \theta^{n_h}(1-\theta)^{n_t} && \text{(Binomial likelihood)}\\
\hat{\theta}_{\mathrm{ML}} &= \frac{n_h}{n_h+n_t} && \text{(Maximum likelihood solution, see last week's tutorial)}\\
p(\theta) &\propto \theta^{\alpha_h-1}(1-\theta)^{\alpha_t-1} && \text{(Beta prior)}\\
\mathrm{E}[\theta] &= \frac{\alpha_h}{\alpha_h+\alpha_t} && \text{(Prior mean, property of beta distribution)}\\
p(\theta|n_h,n_t) &\propto P(n_h,n_t|\theta)p(\theta) \propto \theta^{\alpha_h+n_h-1}(1-\theta)^{\alpha_t+n_t-1} && \text{(Posterior is also a beta distribution)}\\
\mathrm{E}[\theta|\mathcal{D}] &= \frac{\alpha_h+n_h}{\alpha_h+\alpha_t+n_h+n_t} && \text{(Posterior mean, property of beta distribution)}\\
&= \frac{\alpha_h}{\alpha+n} + \frac{n_h}{\alpha+n} && (\alpha = \alpha_h+\alpha_t,\ n = n_h+n_t)\\
&= \frac{\alpha}{\alpha+n}\frac{\alpha_h}{\alpha} + \frac{n}{\alpha+n}\frac{n_h}{n}\\
&= \lambda\mathrm{E}[\theta] + (1-\lambda)\hat{\theta}_{\mathrm{ML}} && (\lambda = \alpha/(\alpha+n))
\end{aligned}
$$

2. Expected loss (or risk) of go ahead is $L(\text{go ahead}, \text{fair}) \times P(\text{fair}) + L(\text{go ahead}, \text{rain}) \times P(\text{rain}) = -1\times 0.4 + 2\times 0.6 = 0.8$. Expected loss of cancel is $3 \times 0.4 + 0 \times 0.6 = 1.2$. To minimise loss, should go ahead.

3. Choose C_1 if (note only the case where $1.0 \le x \le 1.1$ makes sense, if we assume the classes are complete)

$$
P(C_1|x) > P(C_2|x) \implies p(x|C_1)P(C_1) > p(x|C_2)P(C_2) \implies 10 \times 0.6 > 200(x-1) \times 0.4 \implies x < 1.075
$$

It is a good idea to sketch $P(C_i, x)$ in the domain $1.0 \le x \le 1.1$ to be clear what is going on.

Since $x = 1.075$ is the decision boundary

$$
\begin{aligned}
P(\text{error}) = \int_{1.0}^{1.1} p(\text{error}, x)\,\mathrm{d}x &= \int_{1.0}^{1.075} p(C_2, x)\,\mathrm{d}x + \int_{1.075}^{1.1} p(C_1, x)\,\mathrm{d}x\\
&= \int_{1.0}^{1.075} p(x|C_2)P(C_2)\,\mathrm{d}x + \int_{1.075}^{1.1} p(x|C_1)P(C_1)\,\mathrm{d}x\\
&= \int_{1.0}^{1.075} 200(x-1)\times 0.4\,\mathrm{d}x + \int_{1.075}^{1.1} 10\times 0.6\,\mathrm{d}x\\
&= [40x^2 - 80x]_{1.0}^{1.075} + [6x]_{1.075}^{1.1} = 0.375
\end{aligned}
$$

4.

$$
\begin{aligned}
P(\text{error}|x, \text{chosen } C_i) &= P(C_i \text{ is } \textit{not} \text{ the class}|x, \text{chosen } C_i)\\
&= P(C_i \text{ is } \textit{not} \text{ the class}|x) && \text{(whether or not } C_i \text{ is the class is independent of our choice)}\\
&= 1 - P(C_i \text{ is the class}|x)\\
&= 1 - P(C_i|x)
\end{aligned}
$$

$$
P(\text{error}|x) = \sum_i P(\text{error}|x, \text{chosen } C_i)Q(C_i|x) = \sum_i [1 - P(C_i|x)]Q(C_i|x) = 1 - \sum_i P(C_i|x)Q(C_i|x)
$$

$$
P(\text{error}) = \int P(\text{error}|x)p(x)\,\mathrm{d}x = \int \left(1 - \sum_i P(C_i|x)Q(C_i|x)\right)p(x)\,\mathrm{d}x = 1 - \int \sum_i P(C_i|x)Q(C_i|x)p(x)\,\mathrm{d}x
$$

To minimise $P(\text{error})$, we minimise $P(\text{error}|x)$ for all x, i.e. maximise $\sum_i P(C_i|x)Q(C_i|x)$. To do this, observe that $\sum_i P(C_i|x)Q(C_i|x)$ is a convex combination of $\{P(C_i|x)\}_i$, since $Q(C_i|x)$ is a probability measure. $P(\text{error}|x)$ is thus minimised by placing weight of 1 at $\max_i P(C_i|x)$, and weight of 0 at all other points. Hence, optimum $Q(C_i|x)$ is $\delta(i, \arg\max_j P(C_j|x))$.

5.

$$
\begin{aligned}
\frac{d\log p}{d\theta} &= \frac{\alpha-1}{\theta} - \frac{\beta-1}{1-\theta} \\
\frac{\theta^*}{\alpha-1} &= \frac{1-\theta^*}{\beta-1} \qquad \text{at maximum} \\
\theta^* &= \frac{\alpha-1}{\alpha+\beta-2} \qquad \text{is the value of } \theta \text{ at maximum}
\end{aligned}
$$

For same positions of mean and mode

$$
\begin{aligned}
\frac{\alpha-1}{\alpha+\beta-2} &= \frac{\alpha}{\alpha+\beta} \\
\implies \alpha &= \beta
\end{aligned}
$$