4 Model selection

Because the probabilities of each model are very small we will use the exponential form to present it. It's also convenient to use *scipy.special.betaln* function in Python to calculate the results.

(a) For the first model the evidence is simply:

$$P(\mathcal{D}|\mathcal{M}_a) = \left(\frac{1}{2}\right)^{ND} = \left(\frac{1}{2}\right)^{6400} \approx \exp(-4436) \tag{4.1}$$

$$P(\mathcal{M}_a|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M}_a)P(\mathcal{M}_a) \tag{4.2}$$

(b) For the second model we only have to integrate over the single parameter with uniform prior:

$$P(\mathcal{D}|\mathcal{M}_b) = \int_0^1 P(\mathcal{D}|p_d, \mathcal{M}_b)P(p_d|\mathcal{M}_b)dp_d \tag{4.3}$$

$$= \int_0^1 \prod_{n=1}^{N}\prod_{i=1}^{D} p_d^{x_i^{(n)}}(1-p_d)^{1-x_i^{(n)}} 1dp_d \tag{4.4}$$

$$= \int_0^1 p_d^{\sum_{n=1}^N \sum_{i=1}^D x_i^{(n)}}(1-p_d)^{ND-\sum_{n=1}^N \sum_{i=1}^D x_i^{(n)}} dp_d \tag{4.5}$$

$$= \mathrm{B}(\sum_{n=1}^{N}\sum_{i=1}^{D} x_i^{(n)} + 1, ND - \sum_{n=1}^{N}\sum_{i=1}^{D} x_i^{(n)} + 1) \tag{4.6}$$

$$= \exp\left(\log \mathrm{B}(\sum_{n=1}^{N}\sum_{i=1}^{D} x_i^{(n)} + 1, ND - \sum_{n=1}^{N}\sum_{i=1}^{D} x_i^{(n)} + 1)\right) \tag{4.7}$$

We can calculate the value of the exponent for our dataset with the following code:

```
from scipy.special import betaln

log_P = betaln(Y.sum() + 1, N*D - Y.sum() + 1)
```

This gives us:

$$P(\mathcal{D}|\mathcal{M}_b) \approx \exp(-4284) \tag{4.8}$$

$$P(\mathcal{M}_b|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M}_b)P(\mathcal{M}_b) \tag{4.9}$$

(c) Finally we can calculate the evidence for the last model:

$$
\begin{aligned}
P(\mathcal{D}|\mathcal{M}_c) &= \int_{[0,1]^D} P(\mathcal{D}|\boldsymbol{p}, \mathcal{M}_c) P(\boldsymbol{p}|\mathcal{M}_c) d\boldsymbol{p} && (4.10)\\
&= \int_{[0,1]^D} \prod_{n=1}^{N} \prod_{d=1}^{D} p_d^{x_d^{(n)}} (1-p_d)^{1-x_d^{(n)}} 1 d\boldsymbol{p} && (4.11)\\
&= \prod_{d=1}^{D} \int_0^1 \prod_{n=1}^{N} p_d^{x_d^{(n)}} (1-p_d)^{1-x_d^{(n)}} dp_d && (4.12)\\
&= \prod_{d=1}^{D} \mathrm{B}\left(\sum_{n=1}^{N} x_d^{(n)} + 1, N - \sum_{n=1}^{N} x_d^{(n)} + 1 \right) && (4.13)\\
&= \exp\left(\sum_{d=1}^{D} \log \mathrm{B}\left(\sum_{n=1}^{N} x_d^{(n)} + 1, N - \sum_{n=1}^{N} x_d^{(n)} + 1 \right) \right) && (4.14)
\end{aligned}
$$

Code:

```python
from scipy.special import betaln
```

```python
log_P = sum(betaln(Y[:, d].sum() + 1, N - Y[:, d].sum() + 1) for d in range(D))
```

This gives us:

$$
\begin{aligned}
P(\mathcal{D}|\mathcal{M}_c) &\approx \exp(-3851) && (4.15)\\
P(\mathcal{M}_c|\mathcal{D}) &\propto P(\mathcal{D}|\mathcal{M}_c) P(\mathcal{M}_c) && (4.16)
\end{aligned}
$$

We observe that since we have a uniform prior on model probability it holds that $P(\mathcal{M}_i|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M}_i)$. We can calculate relative posterior probability of each model:

$$
\begin{aligned}
P(\mathcal{M}_a|\mathcal{D}) &\propto \exp(-4436) && (4.17)\\
P(\mathcal{M}_b|\mathcal{D}) &\propto \exp(-4284) && (4.18)\\
P(\mathcal{M}_c|\mathcal{D}) &\propto \exp(-3851) && (4.19)
\end{aligned}
$$

From this we can calculate (assuming we don't consider other models) that since $P(\mathcal{M}_i|\mathcal{D}) = \frac{P(\mathcal{M}_i|\mathcal{D})}{\sum_k P(\mathcal{M}_k|\mathcal{D})}$:

$$
\begin{aligned}
P(\mathcal{M}_a|\mathcal{D}) &\approx 0 && (4.20)\\
P(\mathcal{M}_b|\mathcal{D}) &\approx 0 && (4.21)\\
P(\mathcal{M}_c|\mathcal{D}) &\approx 1 && (4.22)
\end{aligned}
$$