Problem Set 0

1. We need to calculate

$$\underline{p(a=1 \mid b=1)}$$

Since the data given is in inverse form (i.e., $p(b|a)$), we need to use Bayes' theorem:

$$p(a=1 \mid b=1) = \frac{p(b=1 \mid a=1) \cdot p(a=1)}{p(b=1)}$$

Now, $p(b=1 \mid a=1) = \underline{0.95}$ (given)

$p(a=1) = \underline{0.01}$ (given)

$$\begin{aligned}
p(b=1) &= p(b=1, a=1) + p(b=1, a=0) \\
&= p(b=1 \mid a=1) \cdot p(a=1) + p(b=1 \mid a=0) \cdot p(a=0) \\
&= 0.95 \times 0.01 + \underbrace{(1-0.95)}_{[p(b=0 \mid a=0),\ \text{given}]} \cdot \underbrace{(1-0.01)}_{[p(a=1)]} \\
&= 0.0095 + 0.0495 \\
&= \underline{0.059}
\end{aligned}$$

$$\Rightarrow p(a=1 \mid b=1) = \frac{0.95 \times 0.01}{0.059} = \underline{\cancel{0}0.1610}$$

OR $\underline{16.1\%}$. This may seem surprising: the test has 95% reliability, but still a positive result implies only a 16.1% chance of having the disease! This is due to the effect of the <u>small prior</u> $p(\underset{=1}{a})$; see also Example 2.3 of MacKay.

2.(a) Let the probability of getting 'i' on the die be p_i, for $i \in \{1, 2, \ldots, 6\}$. We have the following constraints:

① $\sum_{i=1}^{6} p_i = 1$

② $\sum_{i \in \{1,3,5\}} p_i = \sum_{i \in \{2,4,6\}} p_i$ (= 0.5, using ①)

These two constraints allow us to eliminate two unknowns. Suppose we are given p_1, p_2, p_3, p_4.

Then:

$$p_5 = 0.5 - p_1 - p_3$$

$$p_6 = 0.5 - p_2 - p_4$$

So we have 4 parameters here.

Let us write down the likelihood of the data (D) given these parameters:

$$\mathcal{L} = p(D \mid \{p_1, p_2, p_3, p_4\}) = \underbrace{K}_{\text{some constant}} p_1^{12} p_2^{20} p_3^{16} p_4^{24} (0.5 - p_1 - p_3)^{12} (0.5 - p_2 - p_4)^{16}$$

To find the ML estimate for any parameter, we differentiate w.r.t. that and set it to 0. E.g.:

$$\frac{\partial \mathcal{L}}{\partial p_1} = K p_2^{20} p_3^{16} p_4^{24} (0.5 - p_2 - p_4)^{16} \left[12 p_1^{11} (0.5 - p_1 - p_3)^{12} - 12 p_1^{12} (0.5 - p_1 - p_3)^{11} \right] = 0$$

$$\Rightarrow 12 p_1^{11} (0.5 - p_1 - p_3)^{11} (0.5 - p_1 - p_3 - p_1) = 0$$

Since setting either p_1 or p_5 to 0 would make $\mathcal{L} = 0$, clearly they are minima. The only possible maximum is given by:

$$0.5 - 2p_1 - p_3 = 0 \quad - \text{(I)}$$

To solve further, we also need to differentiate w.r.t. p_3 and set to 0; analogously to the above, this will give the equation

$$16 p_3^{15} (0.5 - p_1 - p_3)^{12} - 12 p_3^{16} (0.5 - p_1 - p_3)^{11} = 0$$

$$\Rightarrow 4 p_3^{15} (0.5 - p_1 - p_3)^{11} \left(4(0.5 - p_1 - p_3) - 3p_3\right) = 0$$

[For maximum] $\Rightarrow \; 2 - 4p_1 - 7p_3 = 0 \quad - \text{(II)}$

Solving (I) and (II), we get:

$$p_1 = \underline{0.15}, \quad p_3 = \underline{0.2}$$

$$\Rightarrow p_5 = \underline{0.15}$$

Note that the total probability for an odd number (0.5) has been divided up amongst 1, 3, and 5, in proportion to the number of samples of each in the data (12, 16, and 12 respectively). Similarly, the ML estimates for the other probabilities will be:

$$p_2 = \frac{1}{6} \approx \underline{0.1667}; \quad p_4 = \underline{0.2}; \quad p_6 = \frac{2}{15} \approx \underline{0.1333}$$

2. (b) Now we are discarding the assumption (2), thus our only constraint will be $\sum_{i=1}^{6} p_i = 1$, and we will need 5 parameters, $p_1, \ldots, p_5$. Given these, we know $p_6 = 1 - \sum_{i=1}^{5} p_i$.

We would like to use a prior which places equal weight/importance on odd and even numbers. The probability of an odd number is $p_1 + p_3 + p_5$. Thus, a natural choice of prior seems to be:

$$p(\underline{p}) = \underbrace{K}_{\text{const.}}(p_1+p_3+p_5)(1-p_1-p_3-p_5)$$

Here $\underline{p}$ is shorthand for $\{p_1, \ldots, p_5\}$. Note that this prior is exactly like the '$p(1-p)$' prior for coin-tossing.

Our posterior is now:

$$p(\underline{p}|D) = \frac{p(D|\underline{p}) \cdot p(\underline{p})}{p(D)}$$

$$= \underbrace{K'}_{\text{const.}} p_1^{12} p_2^{20} p_3^{16} p_4^{24} p_5^{12} \left(1-\sum_{i=1}^{5} p_i\right)^{16} (p_1+p_3+p_5)(1-p_1-p_3-p_5)$$

For MAP solution:

$$\frac{\partial p(\underline{p}|D)}{\partial p_1} = K' p_2^{20} p_3^{16} \cdot p_4^{24} p_5^{12} \left[12 p_1^{11} \left(1-\sum_{i=1}^{5} p_i\right)^{16} (p_1+p_3+p_5)(1-p_1-p_3-p_5) + \right.$$

$$-16p_1^{12}\left(1-\sum_{i=1}^{5}p_i\right)^{15}(p_1+p_3+p_5)(1-p_1-p_3-p_5)$$

$$+p_1^{12}\left(1-\sum_{i=1}^{5}p_i\right)^{16}(1-p_1-p_3-p_5)$$

$$\cancel{+}\underset{\text{(minus)}}{}\,p_1^{12}\left(1-\sum_{i=1}^{5}p_i\right)^{16}(p_1+p_3+p_5)\Big]$$

If we set this $=0$, we get

$$12\left(1-\sum_{i=1}^{5}p_i\right)(p_1+p_3+p_5)(1-p_1-p_3-p_5)-16p_1(p_1+p_3+p_5)(1-p_1-p_3-p_5)$$

$$+p_1\left(1-\sum_{i=1}^{5}p_i\right)\left[(1-p_1-p_3-p_5)-(p_1+p_3+p_5)\right]=0$$

This is complicated to solve! To simplify, let us break into two steps.

Define $p_{odd}=p_1+p_3+p_5$

$$\Rightarrow p(\underline{p}\,|\,D)=K'p_1^{12}p_2^{20}p_3^{16}p_4^{24}(p_{odd}-p_1-p_3)^{12}$$

$$(1-p_{odd}-p_2-p_4)^{16}\;p_{odd}\,(1-p_{odd})$$

By analogy with 2.(a), we can see the following:

$$\frac{\partial p(\underline{p}\,|\,D)}{\partial p_1}=0\Rightarrow p_{odd}-2p_1-p_3=0$$

$$\frac{\partial p(\underline{p}\,|\,D)}{\partial p_3}=0\Rightarrow 4p_{odd}-4p_1-7p_3=0$$

Similarly:

$$\frac{\partial p(\underline{p}\,|\,D)}{\partial p_2}=0\Rightarrow 5-5p_{odd}-9p_2-5p_4=0$$

$$\frac{\partial p(\underline{p}\,|\,D)}{\partial p_4}=0\Rightarrow 3-3p_{odd}-3p_2-5p_4=0$$

Solving these 4 equations for $p_1, \ldots, p_4$, we get:

$$p_1 = \frac{3p_{odd}}{10} \; ; \; p_3 = \frac{4p_{odd}}{10} \; ; \; p_2 = \frac{1-p_{odd}}{3} \; ; \; p_4 = \frac{4(1-p_{odd})}{10}$$

Note that the intuition behind these is exactly the same as in 2.(a). If we put $\underline{p_{odd} = 0.5}$, we recover those estimates.

Now we will rewrite the posterior purely as a function of p_{odd}, using the above identities:

$$p(p|D) = \underbrace{K''}_{\text{new const.}} p_{odd}^{12} (1-p_{odd})^{20} p_{odd}^{16} (1-p_{odd})^{24} p_{odd}^{12} (1-p_{odd})^{16} p_{odd} (1-p_{odd})$$

$$= K'' p_{odd}^{41} (1-p_{odd})^{61}$$

Now we see how this is exactly the same as the coin-tossing example!

Set $\dfrac{\partial p(p|D)}{\partial p_{odd}} = 0 \Rightarrow \underline{p_{odd} = \dfrac{41}{102}}$

So the difference with part (a) is that p_{odd} is now less than 0.5; this is because it is also being driven by the observed data, which suggests that $p_{odd} < p_{even}$. If we want to enforce $p_{odd} \simeq 0.5$ more strongly, we can use a prior of the form $p_{odd}^K (1-p_{odd})^K$; as $K \to \infty$, the MAP solution goes to the one in part (a).