Problem 1. (40 pts.) Maximum likelihood estimates

(a) likelihood $= f(\text{data} | \alpha) = \frac{\alpha}{2^{\alpha+1}} \cdot \frac{\alpha}{3^{\alpha+1}} \cdot \frac{\alpha}{2.5^{\alpha+1}} \cdot \frac{\alpha}{6^{\alpha+1}} = \frac{\alpha^5}{450^{\alpha+1}}$. Therefore, log likelihood $= \ln(f(\text{data} | \alpha)) = \ln(\alpha) - (\alpha + 1) \ln(450)$. We find the maximum likelihood by setting the derivative equal to 0:

$$\frac{d}{d\alpha} \ln(f(\text{data} | \alpha)) = \frac{5}{\alpha} - \ln(450) = 0.$$ 

Solving we get $\hat{\alpha} = \frac{5}{\ln(450)} \approx 0.8184314$.

(b) The hypotheses are that the urn used is urn 1, urn 2 or urn 3. The data is the the chosen balls were red, then green, then red. Call the data $RGR$. So,

$$P(RGR|\text{urn 1}) = \frac{6}{18} \cdot \frac{7}{17} \cdot \frac{5}{16} \approx 0.0429$$
$$P(RGR|\text{urn 2}) = \frac{4}{16} \cdot \frac{9}{15} \cdot \frac{3}{14} \approx 0.0321$$
$$P(RGR|\text{urn 3}) = \frac{5}{21} \cdot \frac{10}{20} \cdot \frac{4}{19} \approx 0.0251$$

So, the maximum likelihood estimate is that the urn is urn 1.

(c) The pdf of a uniform($0,b$) distribution takes two values

$$f(x|b) = \begin{cases} 1/b & \text{if } x \text{ is in } [0,b] \\ 0 & \text{otherwise} \end{cases}$$

Since the likelihood is a product of the likelihoods of each data point the likelihood function is

$$f(2.5, 19.75, 12.0, 7.0|b) = \begin{cases} (1/b)^4 & \text{if all 4 data points are in the interval } [0,b] \\ 0 & \text{otherwise} \end{cases}$$

This is maximized when $b$ is as small as possible while making sure all the data points are in $[0,b]$. This means $b$ is the the maximum of the data, i.e. $b = 19.75$.

(d) (i) Since $y_i \sim \text{N}(ax_i + b, \sigma^2)$ the likelihood with data $(x_1, y_1)$ is

$$f(x_1, y_1 | a, b, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} e^{-(y_1 - ax_1 - b)^2/(2\sigma^2)}.$$ 

The log likelihood is

$$\ln(f(x_1, y_1 | a, b, \sigma)) = -\ln(\sqrt{2\pi} \sigma) - \frac{(y_1 - ax_1 - b)^2}{2\sigma^2}.$$
(ii) The likelihood for all the data is the product of the individual likelihoods. So,

\[ f((1,1), (3,3), (1.5, 4) \mid a, b, \sigma) = \left( \frac{1}{\sqrt{2\pi \sigma}} \right)^3 e^{-((1-a-b)^2+(3-3a-b)^2+(4-1.5a-b)^2)}/(2\sigma^2) \]

Taking the log (and replacing the list of data by the word ‘data’) we get

\[ \ln(f(data \mid a, b, \sigma)) = -3 \ln(\sqrt{2\pi \sigma}) - (1-a-b)^2 + (3-3a-b)^2 + (4-1.5a-b)^2)/2\sigma^2 \]

Since we want to find \( a \) and \( b \) that maximize the likelihood we take the partial derivatives and set them to 0.

\[
\frac{\partial \ln(f(data \mid a, b, \sigma))}{\partial a} = \frac{2}{2\sigma^2}((1-a-b) + 3(3-3a-b) + 1.5(4-1.5a-b)) = 0
\]

\[
\frac{\partial \ln(f(data \mid a, b, \sigma))}{\partial b} = \frac{2}{2\sigma^2}((1-a-b) + (3-3a-b) + (4-1.5a-b)) = 0
\]

These are two equations in the unknowns \( a \) and \( b \). We simplify and solve:

\[
12.25a + 5.5b = 16
\]

\[
5.5a + 3b = 8
\]

which gives \( a \approx 0.615; \ b \approx 1.538 \).

The linear regression fit of a line to the data is \( y = ax + b \approx 0.615x + 1.538 \).

(iii) Here is the code for this plot:

```r
x = c(1,3,1.5)
y = c(1,3,4)
a = 0.6153846
b = 1.5384615
plot(x,y,pch=19,col='blue',cex=1,xlim=c(0,5),ylim=c(0,5))
#Perversely, in abline a is the intercept and b is the slope.
abline(a=b,b=a, col='cyan', lwd=2)
```

(iv) It is a happy circumstance that the solutions for \( a \) and \( b \) don’t depend on \( \sigma \)! This means that the procedure for finding the “best” line is the same no matter how large or small we think the measurement errors in \( y \) might be.
Note: we do have to assume that $\sigma$ is the same for all values of $x$. Later we will learn to describe this with the great word homoscedasticity.

**Problem 2.** (15 pts.) In all three parts to this problem we have 3 hypotheses:

$H_A = \text{the car is behind door } A$

$H_B = \text{the car is behind door } B$

$H_C = \text{the car is behind door } C$.

In all three parts the data is $D = \text{Monty opens door } B \text{ and reveals a goat}$.

**(a)** The key to our Bayesian update table is the likelihoods: Since Monty is sober he always reveals a goat.

$P(D|H_A)$: $H_A$ says the car is behind $A$. So Monty is equally likely to pick $B$ or $C$ and reveal a goat. Thus $P(D|H_A) = 1/2$.

$P(D|H_B)$: Since $H_B$ says the car is behind $B$, sober Monty will never choose $B$ (and if he did it would not reveal a goat). So $P(D|H_B) = 0$.

$P(D|H_C)$: $H_C$ says the car is behind $C$. Since sober Monty doesn’t make mistakes he will open door $B$ and reveal a goat. So $P(D|H_C) = 1$.

Here is the table for this situation.

| $H$ | $P(H)$ | $P(D|H)$ | Bayes numer. | Posterior |
|-----|--------|----------|--------------|-----------|
| $H_A$ | 1/3    | 1/2      | 1/6          | 1/3       |
| $H_B$ | 1/3    | 0        | 0            | 0         |
| $H_C$ | 1/3    | 1        | 1/3          | 2/3       |

Therefore, Elan should switch, as his chance of winning the car after switching is double that had he stayed with his initial choice.

**(b)** Some of the likelihoods change in this setting.

$P(D|H_A)$: $H_A$ says the car is behind $A$. So Monty is equally likely to show $B$ or $C$ and reveal a goat. Thus $P(D|H_A) = 1/2$.

$P(D|H_B)$: Since $H_B$ says the car is behind $B$, drunk Monty might show $B$, but if he does we won’t reveal a goat. (He will ruin the game.) So $P(D|H_B) = 0$.

$P(D|H_C)$: $H_C$ says the car is behind $C$. Drunk Monty is equally likely to show $B$ or $C$. If he chooses $B$ he’ll reveal a goat. So $P(D|H_C) = 1/2$.

Our table is now:

| $H$ | $P(H)$ | $P(D|H)$ | Bayes numer. | Posterior |
|-----|--------|----------|--------------|-----------|
| $H_A$ | 1/3    | 1/2      | 1/6          | 1/2       |
| $H_B$ | 1/3    | 0        | 0            | 0         |
| $H_C$ | 1/3    | 1/2      | 1/6          | 1/2       |

So in this case switching is just as good (or as bad) as staying with the original choice.
(c) We have to recompute the likelihoods.

$P(D|H_A)$: If the car is behind $A$ then sober or drunk Monty is equally likely to choose door $B$ and reveal a goat. Thus $P(D|H_A) = 1/2$.

$P(D|H_B)$: If the car is behind door $B$ then whether he chooses it or not Monty can’t reveal a goat behind it. So $P(D|H_B) = 0$.

$P(D|H_C)$: Let $S$ be the event that Monty is sober and $S^c$ the event he is drunk. From the table in (a), we see that $P(D|H_C, S) = 1$ and from the table in (b), we see that $P(D|H_C, S^c) = 1/2$. Thus, by the law of total probability

$$P(D|H_C) = P(D|H_C, S)P(S) + P(D|H_C, S^c)P(S^c) = 0.7 + \frac{1}{2}(0.3) = .85 = \frac{17}{20}.$$
Bayes numerator

\[ P(R_2|R_1) = P(R_2|H_4)P(H_4|R_1) + P(R_2|H_6)P(H_6|R_1) + P(R_2|H_8)P(H_8|R_1) \\
+ P(R_2|H_{12})P(H_{12}|R_1) + P(R_2|H_{20})P(H_{20}|R_1) \\
\approx 0 \cdot 0 + \frac{1}{6} \cdot 0.2913 + \frac{1}{8} \cdot 0.2913 + \frac{1}{12} \cdot 0.2427 + \frac{1}{20} \cdot 0.1748 \\
\approx 0.1139159. \]

(c) Let \( H_4, H_6, H_8, H_{12}, \) and \( H_{20} \) are the hypotheses that we have selected the 4, 6, 8, 12, or 20 sided die respectively.

The data is ‘rolled \( n \) sixes’. We compute

| Hyp. \( H \) | Prior \( P(H) \) | Likelihood \( P(\text{data}|H) \) | Bayes numerator | Posterior \( P(H|\text{data}) \) |
|---|---|---|---|---|
| \( H_4 \) | 2/20 | 0 | 0 | 0 |
| \( H_6 \) | 3/20 | \((1/6)^n\) | \(\frac{3}{20} \cdot (1/6)^n\) | \(\frac{3}{20} T (1/6)^n\) |
| \( H_8 \) | 4/20 | \((1/8)^n\) | \(\frac{4}{20} \cdot (1/8)^n\) | \(\frac{4}{20} T (1/8)^n\) |
| \( H_{12} \) | 5/20 | \((1/12)^n\) | \(\frac{5}{20} \cdot (1/12)^n\) | \(\frac{5}{20} T (1/12)^n\) |
| \( H_{20} \) | 6/20 | \((1/20)^n\) | \(\frac{6}{20} \cdot (1/20)^n\) | \(\frac{6}{20} T (1/20)^n\) |
| Total: | 1 | – | \( T = \frac{3}{20} \cdot 6^n + \frac{4}{20} \cdot 8^n + \frac{5}{20} \cdot 12^n + \frac{6}{20} \cdot 20^n \) | 1 |

The posterior probabilities are given in the table. We’ll now show that as \( n \) goes to infinity the posterior for \( H_6 \) goes to 1 and the others go to 0.

The quick way to see this is to note that the biggest term in \( T \) is \( \frac{3}{20} \cdot (1/6)^n \). So as \( n \) gets large we can ignore all the other terms in \( T \). To do the formal algebra we rewrite the posterior probabilities by multiplying numerator and denominator by \( 6^n \):

\[
\begin{align*}
P(H_6|\text{data}) &= \frac{\frac{3}{20}}{\frac{3}{20} + \frac{4}{20} \cdot \left(\frac{6}{5}\right)^n + \frac{5}{20} \cdot \left(\frac{6}{12}\right)^n + \frac{6}{20} \cdot \left(\frac{6}{20}\right)^n} \\
P(H_8|\text{data}) &= \frac{\frac{4}{20} \cdot \left(\frac{6}{8}\right)^n}{\frac{3}{20} + \frac{4}{20} \cdot \left(\frac{6}{5}\right)^n + \frac{5}{20} \cdot \left(\frac{6}{12}\right)^n + \frac{6}{20} \cdot \left(\frac{6}{20}\right)^n} \\
P(H_{12}|\text{data}) &= \frac{\frac{5}{20} \cdot \left(\frac{6}{12}\right)^n}{\frac{3}{20} + \frac{4}{20} \cdot \left(\frac{6}{5}\right)^n + \frac{5}{20} \cdot \left(\frac{6}{12}\right)^n + \frac{6}{20} \cdot \left(\frac{6}{20}\right)^n} \\
P(H_{20}|\text{data}) &= \frac{\frac{6}{20} \cdot \left(\frac{6}{20}\right)^n}{\frac{3}{20} + \frac{4}{20} \cdot \left(\frac{6}{5}\right)^n + \frac{5}{20} \cdot \left(\frac{6}{12}\right)^n + \frac{6}{20} \cdot \left(\frac{6}{20}\right)^n}
\end{align*}
\]

In every one of these expressions all the \( n \)th powers go to 0 as \( n \) goes to infinity (this is because the number raised to the \( n \) is less than 1). This means \( P(H_6|\text{data}) \) goes to \( \frac{3/20}{3/20} = 1 \) and the other posterior probabilities go to 0.
In the R code for ps5 we computed the posteriors for \( n = 1, 4, 8, 16 \). The table generated is below. It agrees with our claim that the posterior for \( H_6 \) goes to 1 and the others go to 0.

| Hyp. \( H \) | \( P(H|\text{data, } n = 1) \) | \( P(H|\text{data, } n = 4) \) | \( P(H|\text{data, } n = 8) \) | \( P(H|\text{data, } n = 16) \) |
|-------------|------------------|------------------|------------------|------------------|
| \( H_4 \)   | 0.0              | 0.0              | 0.0              | 0.0              |
| \( H_6 \)   | 0.291            | 0.648            | 0.877            | 0.987            |
| \( H_8 \)   | 0.291            | 0.274            | 0.117            | 0.013            |
| \( H_{12} \) | 0.243            | 0.068            | 0.006            | \( 2.51 \times 10^{-5} \) |
| \( H_{20} \) | 0.175            | 0.011            | \( 1.15 \times 10^{-4} \) | \( 8.5 \times 10^{-9} \) |

This makes sense because, as unlikely as it is, a string of sixes is much more likely with the six-sided die than with any of the dice with more sides.

\textbf{(d)} This is just using the law of total probability and the posterior probabilities from part (c) with \( n = 10 \). This is best computed in R. The code is posted in the usual place in ps5-sol.R. Here is the table of predictive probabilities. We simplify the table by combining columns that are identical. So the probability the next roll is a 1 is the same as the probability it’s a 2 or 3 or 4. Let \( x_{11} \) be the result of the eleventh roll. The data is 10 straight rolls of 6.

<table>
<thead>
<tr>
<th>( x_{11} )</th>
<th>1-4</th>
<th>5-6</th>
<th>7-8</th>
<th>9-12</th>
<th>13-20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(x_{11}</td>
<td>\text{data}) )</td>
<td>0.1636</td>
<td>0.1636</td>
<td>0.0088</td>
<td>( 1.26 \times 10^{-4} )</td>
</tr>
</tbody>
</table>

\textbf{Problem 4. (15 pts.) What are the odds?}

For this problem we use our fairly standard notation: \( D^+ \) is the event that the randomly chosen person has the disease; \( D^- \) is the event that they are healthy; \( T_1^+ \), \( T_2^+ \) are the events that the first and second tests are positive; \( T_1^- \), \( T_2^- \) are the events that they are negative.

\textbf{(a)} Prior odds are \( \frac{P(D^+)}{P(D^-)} = \frac{0.001}{0.999} \approx 0.001. \)

\textbf{(b)} The Bayes factor is \( \frac{P(T_1^+|D^+)}{P(T_1^+|D^-)} = \frac{0.98}{0.10} = 9.8. \) So the posterior odds are

\[ \text{Bayes factor } \times \text{ prior odds } = 9.8 \times \frac{1}{999} \approx 0.0098. \]

\textbf{(c)} The posterior predictive odds we want are \( \frac{P(T_2^+|T_1^+)}{P(T_2^-|T_1^+)} \). There is no slick formula based on the odds computed in parts (a) and (b). Instead we have to compute the posterior probabilities \( P(D^+|T_1^+) \) and \( P(D^-|T_1^+) \). Then we use the law of total probability to compute \( P(T_2^+|T_1^+) \) and \( P(T_2^-|T_1^+) \). The Bayesian update table is...
Now we can use the law of total probability to compute the predictive probabilities $P(T_2^+|T_1^+)$ and $P(T_2^-|T_1^+)$. (The law of total probability works fine if every probability is conditioned on $T_1^+$, since that just amounts to restricting the sample space to $T_1^+$.)

\[
P(T_2^+|T_1^+) = P(T_2^+|D^+, T_1^+)P(D^+|T_1^+) + P(T_2^+|D^-, T_1^+)P(D^-|T_1^+) = 0.99 \cdot 0.0097 + 0.01 \cdot 0.9903 \approx 0.0195
\]

\[
P(T_2^-|T_1^+) = 1 - P(T_2^+|T_1^+) \approx 0.9805
\]

So, the posterior predictive odds are

\[
\frac{P(T_2^+|T_1^+)}{P(T_2^-|T_1^+)} \approx \frac{0.0195}{0.9805} \approx 0.0199
\]

**Problem 5. (10 pts.) Legal Trickery**

(a) The lawyer may correctly state that $P(M|B) = 1/1000$, but the lawyer then conflates this with the probability of guilt given all the relevant data, which is really $P(M|B, K)$. The short counterargument is that while only one in a thousand abused wives are murdered, the vast majority of those that are murdered are killed by their abusers.

Working formally using Bayes’ theorem, conditioning on $B$ throughout, we have:

\[
P(M|K, B) = P(M \text{ murdered Mrs S} | \text{ he beat her & she was killed})
= \frac{P(K|M, B)P(M|B)}{P(K|B)}
= \frac{P(M|B)}{P(K|B)}.
\]

(Here we have used $P(K|M, B) = 1$, i.e. given she was murdered by her husband the probability she was killed is 1.) Now let $N$ be the event that Mrs S was murdered by someone other than her husband. Again by Bayes’ theorem

\[
P(N|K, B) = P(\text{another murdered Mrs S} | \text{ Mr S beat her & she was killed})
= \frac{P(K|N, B)P(N|B)}{P(K|B)}
= \frac{P(N|B)}{P(K|B)}.
\]

So the odds of Mr S’s guilt are

\[
O(M|K, B) = \frac{P(M|B, K)}{P(N|B, K)} = \frac{P(M|B)/P(K|B)}{P(N|B)/P(K|B)} = \frac{P(M|B)}{P(N|B)}.
\]
Now certainly $P(M|B)$ is much greater than $P(N|B)$ which tells us the odds overwhelmingly favor the hypothesis that Mr S is the murderer.

In fact, let’s make one more assumption: $P(N|B) = P(N)$, i.e. that Mrs S being murdered by someone else is independent of the fact that her husband beat her. (We should acknowledge that this assumption might not be warranted without further study.) Now our formula for the odds that Mr S is the murderer is

$$O(M|K,B) = \frac{P(M|B)}{P(N)}$$

Let’s accept the lawyer’s statistic that $P(M|B) = 1/1000$. A quick Wikipedia search shows that the murder rate in the US is about 4/100000.* A further google search shows that the murder rate of women in the US is about 2/100000** If Mr and Mrs S lived in the US that would put the odds at greater than 50 to 1 that he is the murderer. I would say the lawyer’s argument is not credible.

*The reliability of murder statistics varies from country to country. Other countries have much lower murder rates. The worldwide average is about 1/13000. The highest murder rates by country are about 1/1000.


(b) Here are four errors in the argument

1. The prosecutor arrived at “1 in 73 million” as follows: The probability that 1 child from an affluent non-smoking family dies of SIDS is 1/8543, so the probability that 2 children die is $(1/8543)^2$. However, this assumes that the SIDS death among siblings are independent. Due to genetic or environmental factors, we suspect that this assumption is invalid.

2. The use of the figure “700,000 live births in Britain each year.” The prosecutor had restricted attention only to affluent non-smoking families when erroneously computing the probability of two SIDS deaths. However, he does not similarly restrict his attention when considering the number of births.

3. The rate “once every hundred years” is not valid: The prosecutor arrived at this by multiplying the number of live births by the probability that two children die from SIDS. The result is a non-sensical rate.

4. While double SIDS is very unlikely, double infanticide may be even more unlikely. It is the odds of one explanation relative to the other given the deaths that matters, and not just how unlikely one possibility is.

The Sally Clark case is an example of the “Prosecutor’s Fallacy.” You can read about it at

There is also a video that discusses legal fallacies at
http://www.ted.com/talks/peter_donnelly_shows_how_stats_fool_juries