Assignment 2

Problem 1

Reproduce the formatting of the following equation. You might want to use the following commands:

`\align;  \intertext;  \left; \right;  \sum; \prod \dots; \frac; \log;`

$$\begin{align}
\frac{\sum_{i=1}^{n} \log(a_i)}{n} &= \frac{\log(a_1) + \log(a_2) + \cdots + \log(a_n)}{n} \tag{1}\\
&= \frac{\log(a_1 a_2 \dots a_n)}{n} \tag{2}\\
&= \log\left[\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}}\right] \tag{3}\\
\intertext{because $\log(MN) = \log(M) + \log(N)$ and $\log(M^k) = k\log(M)$}
&= \log\left(\prod_{i=1}^{n} \sqrt[n]{a_i}\right) \tag{4}
\end{align}$$

Problem 2

Suppose a crime has been committed. Blood is found at the scene for which there is no innocent explanation. It is of a type which is present in 1% of the population.

1. The prosecutor claims: "There is a 1% chance that the defendant would have the crime blood type if he were innocent. Thus there is a 99% chance that he is guilty". This is known as the prosecutor's fallacy. What is wrong with this argument?

2. The defender claims: "The crime occurred in a city of 800,000 people. The blood type would be found in approximately 8000 people. The evidence has provided a probability of just 1 in 8000 that the defendant is guilty, and thus has absolutely **no** bearing on the investigation." This is known as the defender's fallacy. What is wrong with this argument claiming the evidence has absolutely no bearing on the investigation?

Problem 3

My neighbor has two children. Assume that the gender of a child is a coin flip. Let the genders of the children be G_1 and G_2. For both questions, write the probability symbolically (e.g. "$P(A|B)$") and give the value.

1. Suppose I happen to see one of his children run by, and it is a boy. What is the probability that the other child is a girl?

2. Suppose instead that I ask him whether he has any boys, and he says yes. What is the probability that one child is a girl?

Problem 4

Consider the Numbers game with one-sided interval hypotheses $h_{\leq x}$ for numbers 1 up to 10, where $h_{\leq x} = \{1, 2, \ldots, x\}$, $x \in \{1 \ldots 10\}$. Assume we have a uniform prior over h.

1. Show that the $h_{mle} = h_{\leq \max(S)}$ for a given set of numbers $S = \{x_1, x_2, \ldots\}$.

2. Briefly say why the MAP estimate = MLE.

3. For Parts 3 and 4, let's suppose $S = \{5, 9\}$. What is the plug-in approximation of the posterior predictive distribution for new data point x?

4. What is the full posterior predictive distribution?