Problem 1

Solution:
The solution to this problem uses the fact that we can model the problem with a binomial distribution. Let X denote the number of failed drives. Therefore:

1. The probability that 100 or more drives fail can be written as $1 - P(X < 100)$, where the second term is a sum of 100 different binomial probabilities. To compute this, we need to sum over the 100 expressions and the answer in this case turns out to be $\approx 0.48$. Hence, $P(X \geq 100) = 1 - 0.48668 \approx 0.51332$.

2. Similarly, the probability that 120 or more drives fail can be written as $1 - P(X < 120)$. Thus, $P(X \geq 120) = 1 - 0.97183 \approx 0.02817$.

3. By trial and error, $p \approx 0.00079$ seems to be the largest value that will lead to a 99% chance of 100 or fewer drives being returned. With $p = 0.001$ the chance of 100 or fewer drives is about 47%, so by decreasing the probability of individual failure down to $p \approx 0.00079$ we can increase the probability of 100 or fewer drives being returned from 47% to 99%.

Problem 2

Solution:
Since components fail independently of each other, it follows that the number of components remaining operative is a binomial random variable. Hence the probability that a system with four components operates correctly is

$$
\sum_{i=2}^{4} \binom{4}{i} (1-p)^{4-i}p^i = \binom{4}{2}(1-p)^2p^2 + \binom{4}{3}(1-p)^3p + \binom{4}{4}(1-p)^4
$$

$$
= 6(1-p)^2p^2 + 4(1-p)^3p + (1-p)^4
$$

Similarly, the probability that a system with three components operates correctly is
\[
\sum_{i=2}^{3} \binom{3}{i} (1 - p)^{3-i} p^i = \binom{3}{2} (1 - p)^2 p + \binom{3}{3} (1 - p)^3
\]

\[
= 3(1 - p)^2 p + (1 - p)^3
\]

When \( p = 0.1 \), the system with four components operates correctly \( \approx 99.63\% \) of the time, while the system with three components operates correctly \( \approx 97.2\% \). Therefore, the system with four components is more reliable.

**Problem 3**

Solution:

Let \( X \) denote the number of games played. Since it is impossible to get 4 wins in fewer than 4 games, and the series cannot go past 7 games without either team getting 4 wins, \( P(X = x) \) will be 0 for \( x < 4 \) and \( x > 7 \). Hence:

\[
E[X] = \sum_{i=4}^{7} xP(X = x)
\]

The probability of A winning the series in \( x \) games is a binomial with the restriction that the last game must be a win for the team that wins the series (otherwise the series will have ended earlier). Therefore:

\[
P(x, A \text{ wins}) = \binom{x-1}{3} p^4 (1-p)^{x-4}
\]

This uses \( \binom{x-1}{3} \) instead of \( \binom{x}{4} \) because of the restriction that the last win for A must be the last game in the series. Similarly,

\[
P(x, B \text{ wins}) = \binom{x-1}{3} (1-p)^4 p^{x-4}
\]

\[
P(x) = P(x, A \text{ wins}) + P(x, B \text{ wins})
\]

\[
E[X] = \sum_{i=4}^{7} x \left[ \binom{x-1}{3} p^4 (1-p)^{x-4} + \binom{x-1}{3} (1-p)^4 p^{x-4} \right]
\]

When \( p = 0.5 \), \( E[X] = \frac{93}{16} = 5.8125 \).
Problem 4
Solution:
Let $X$ denote the number of collisions and $Y$ denote the number of boxes having at least one key (each of which have exactly one key that did not cause a collision). The number of collisions will be the number of keys minus the number of keys that are put into previously empty boxes. Thus, taking expectations yields that $E[X] = r - E[Y]$.

The number of boxes that have at least one key are $k$ minus the number of empty boxes. A box will be empty if for every key, that key was not placed in that box. The probability of this will be $(1 - p_i)^r$, so:

\[
E[Y] = k - E[\text{empty boxes}] = k - \sum_{i=1}^{k} (1 - p_i)^r
\]

\[
E[X] = r - E[Y] = r - k + \sum_{i=1}^{k} (1 - p_i)^r
\]

Problem 5
Solution:
“The occurrence of B makes A more likely” is equivalent to the statement:

\[
P(A|B) > P(A)
\]

\[
P(A|B)P(B) > P(A)P(B)
\]

\[
P(A, B) > P(A)P(B)
\]

\[
P(B|A)P(A) > P(A)P(B)
\]

\[
P(B|A) > P(B)
\]

Therefore, the occurrence of A also makes B more likely.

Problem 6
Solution:
In this problem, let $X$ denote the value of the first toss, $Y$ denote the value of the second toss and $Z$ denote the sum of the two obtained values. That is, $Z = X + Y$

1. The probability that the first number is a 1 given that the total of the two numbers is 3 can be written as

\[
P(X = 1|Z = 3) = \frac{P(Z = 3|X = 1)P(X = 1)}{P(Z = 3)} = \frac{1}{2}
\]
2. The probability that the total of the two numbers is greater than 6, given that the first number is a 3 can be written as

\[ P(Z > 6 | X = 3) = 1 - P(Z \leq 6 | X = 3) = 1 - \frac{P(X = 3 | Z \leq 6)P(Z \leq 6)}{P(X = 3)} = \frac{1}{2} \]

3. The probability that the first number is a 6 given that the total of the two numbers is 7 can be written as

\[ P(X = 6 | Z = 7) = \frac{P(Z = 7 | X = 6)P(X = 6)}{P(Z = 7)} = \frac{1}{6} \]

Problem 7

Solution:

<table>
<thead>
<tr>
<th></th>
<th>non-robot</th>
<th>robot</th>
</tr>
</thead>
<tbody>
<tr>
<td>uci</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>other-edu</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>non-edu</td>
<td>0.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

1. 

\[ P(non - robot|uci) = \frac{P(non - robot \cap uci)}{P(uci)} = \frac{0.4}{0.4 + 0.0} = 1 \]

\[ P(non - robot|other - edu) = \frac{P(non - robot \cap other - edu)}{P(other - edu)} = \frac{0.2}{0.2 + 0.1} = \frac{2}{3} \]

\[ P(non - robot|non - edu) = \frac{P(non - robot \cap non - edu)}{P(non - edu)} = \frac{0.1}{0.1 + 0.2} = \frac{1}{3} \]

2. 

\[ P(uci(robot) = \frac{P(uci \cap robot)}{P(robot)} = \frac{0.0}{0.0 + 0.1 + 0.2} = 0 \]

\[ P(uci|non - robot) = \frac{P(uci \cap non - robot)}{P(non - robot)} = \frac{0.4}{0.4 + 0.2 + 0.1} = \frac{4}{7} \]

3. 

\[ P(non - edu|robot) = \frac{P(non - edu \cap robot)}{P(robot)} = \frac{0.2}{0.0 + 0.1 + 0.2} = \frac{2}{3} \]

\[ P(non - edu|non - robot) = \frac{P(non - edu \cap non - robot)}{P(non - robot)} = \frac{0.1}{0.4 + 0.2 + 0.1} = \frac{1}{7} \]
4. 
\[
P(\text{non-robot}|\text{non-uci}) = \frac{P(\text{non-robot} \cap \text{non-uci})}{P(\text{non-uci})} \\
= \frac{P(\text{non-robot} \cap \text{other-edu} \cup \text{non-edu})}{P(\text{other-edu} \cup \text{non-edu})} \\
= \frac{P((\text{non-robot} \cap \text{other-edu}) \cup (\text{non-robot} \cap \text{non-edu}))}{P(\text{other-edu} \cup \text{non-edu})} \\
= \frac{P(\text{non-robot} \cap \text{other-edu}) + P(\text{non-robot} \cap \text{non-edu})}{P(\text{other-edu}) + P(\text{non-edu})} \\
= \frac{0.2 + 0.1}{0.2 + 0.1 + 0.1 + 0.2} = \frac{1}{2}
\]
Problem 8

Solution:

ICS Web Page Request Data:

Figure 1 shows the estimated probabilities (from the session length file) of sessions of different lengths, where the x and y axes are on a log-scale. This shows a roughly linear relationship between \( \log p(L) \) (the y-axis) and \( \log L \) (the x-axis) where a linear relationship means that \( \log p(L) = a \log L + b \), or in other words

\[
p(L) = CL^{-\gamma}
\]

where \( b = \log C \) and \( a = -\gamma \). This functional form for \( p(L) \) is known as a “power-law” distribution, where \( C \) is a normalization constant (and is a function of \( \gamma \)) and \( \gamma \) is the single parameter of the model (a real-valued scalar greater than or equal to 1). Larger values of \( \gamma \) means that the distribution decays more rapidly as \( L \) increases.

\[
\text{Figure 1: Plot of estimated probabilities of session length } s \text{ for data from ICS Web page requests, on a log-log scale, with a power-law distribution fitted (dotted line).}
\]

In Figure 1 we see that the linear fit on the log-log scale is not quite perfect (there is some curvature that is non-linear and some variability in \( p(r) \) above \( 10^2 \)), but it is a pretty good fit for a probability model with only a single parameter and we can see that it broadly (“to first order”) captures the behavior of \( p(r) \) as a function of \( r \).

The slope of the best-fit line (on this log-log plot) is about \(-2\) (which can be found by either visually drawing a line on the graph or by numerically finding the \( a \) and \( b \) values that minimize the squared error between the line and the data points).

Figure 2 shows a graph where we have fitted different types of probability models to the data (in case we found the best-fit parameters for each model). We can see from this plot that although the power-law model is not perfect, it is a much better fit to the data than other models such as the geometric and Poisson model.
Figure 2: Plot of estimated probabilities of session lengths for data from ICS Web page requests, on a log-log scale, along with different specific probability models.

**Vivisimo and Excite Query Data:**

Figure 3 shows the estimated probability of the most common query, the second most common query, the third most common query, and so on, for samples of queries from the search engines *Vivisimo* and *Excite*. Again, on a log-log plot we see a near linear relationship between $p(r)$ and $r$, indicating a power-law relationship. The slopes for each graph can be estimated visually simply by drawing a straight line that approximately fits the plotted points and then calculating the slope of that line.

Figure 3: Plot of estimated probability of relative query frequency (y-axis) by rank of the query (x-axis), for Vivisimo and Excite query data.