Problem Set 11 Solutions

Due: Tuesday, November 29, 7:30pm

Problem 1. [15 points] In lecture we discussed the Birthday Paradox. Namely, we found that in a group of m people with N possible birthdays, if $m \ll N$, then:

$$\Pr\{\text{all } m \text{ birthdays are different}\} \sim e^{-\frac{m(m-1)}{2N}}$$

To find the number of people, m, necessary for a half chance of a match, we set the probability to 1/2 to get:

$$m \sim \sqrt{(2\ln 2)N} \approx 1.18\sqrt{N}$$

For $N = 365$ days we found m to be 23.

We could also run a different experiment. As we put on the board the birthdays of the people surveyed, we could ask the class if anyone has the same birthday. In this case, before we reached a match amongst the surveyed people, we would already have found other people in the rest of the class who have the same birthday as someone already surveyed. Let's investigate why this is.

(a) [5 pts] Consider a group of m people with N possible birthdays amongst a larger class of k people, such that $m \leq k$. Define $\Pr\{A\}$ to be the probability that m people all have different birthdays *and* none of the other $k - m$ people have the same birthday as one of the m.

Show that, if $m \ll N$, then $\Pr\{A\} \sim e^{\frac{m(m-2k)}{2N}}$. (Notice that the probability of no match is $e^{-\frac{m^2}{2N}}$ when k is m, and it gets smaller as k gets larger.)

Hints: For $m \ll N$: $\frac{N!}{(N-m)!N^m} \sim e^{-\frac{m^2}{2N}}$, and $(1 - \frac{m}{N}) \sim e^{-\frac{m}{N}}$.

Solution. We know:

$$\Pr\{A\} = \frac{N(N-1)\ldots(N-m+1)\cdot(N-m)^{k-m}}{N^k}$$

since there are N choices for the first birthday, $N - 1$ choices for the second birthday, etc., for the first m birthdays, and $N - m$ choices for each of the remaining $k - m$ birthdays. There are total N^k possible combinations of birthdays within the class.

$$
\begin{aligned}
\Pr\{A\} &= \frac{N(N-1)\dots(N-m+1)\cdot(N-m)^{k-m}}{N^k} \\
&= \frac{N!}{(N-m)!}\left(\frac{(N-m)^{k-m}}{N^k}\right) \\
&= \frac{N!}{(N-m)!N^m}\left(\frac{N-m}{N}\right)^{k-m} \\
&= \frac{N!}{(N-m)!N^m}\left(1-\frac{m}{N}\right)^{k-m} \\
&\sim e^{-\frac{m^2}{2N}}\cdot e^{-\frac{m}{N}(k-m)} && \text{(by the Hint)} \\
&= e^{\frac{m(m-2k)}{2N}}
\end{aligned}
$$

■

(b) [10 pts] Find the approximate number of people in the group, m, necessary for a half chance of a match (your answer will be in the form of a quadratic). Then simplify your answer to show that, as k gets large (such that $\sqrt{N} \ll k$), then $m \sim \frac{N \ln 2}{k}$.

Hint: For $x \ll 1$: $\sqrt{1-x} \sim (1-\frac{x}{2})$.

Solution. Setting $\Pr\{A\} = 1/2$, we get a solution for m:

$$
\begin{aligned}
1/2 &= e^{\frac{m(m-2k)}{2N}} \\
-2N\ln 2 &= m^2 - 2km \\
0 &= m^2 - 2km + (2N\ln 2) \\
m &= \frac{2k \pm \sqrt{(2k)^2 - 4(2N\ln 2)}}{2}
\end{aligned}
$$

Simplifying the solution under the assumption of large k, we find:

$$
\begin{aligned}
m &= \frac{2k - \sqrt{4k^2 - 8N\ln 2}}{2} && \text{(taking the lower positive root)} \\
&= k - k\sqrt{1 - \frac{2N\ln 2}{k^2}} \\
&\sim k - k\left(1 - \frac{2N\ln 2}{2k^2}\right) && \text{(by the Hint)} \\
&= \frac{N\ln 2}{k}
\end{aligned}
$$

■

Problem 2. [20 points]

(a) [7 pts] Suppose you repeatedly flip a fair coin until you see the sequence `HHT` or the sequence `TTH`. What is the probability you will see `HHT` first?

Hint: Use a bijection argument.

Solution. In this case the answer is 1/2. The proof is by a bijection argument on the sample space. Let A denote the event that you see `HHT` before `TTH`, and B denote the event that you see `TTH` before `HHT`.

We will define a bijection, g, between A and B so that the probability of $g(w)$ is equal to the probability of w. The bijection is quite simple. Given a sample point $w \in A$, define $g(w) = \bar{w}$, where $\bar{w}$ is the outcome where every `H` is replaced by a `T` and vice versa. For example $g(\texttt{HHT}) = \overline{\texttt{HHT}} = \texttt{TTH}$.

To show that g is a bijection, we first observe that $g : A \to B$. This follows from the fact that `HHT` precedes `TTH` in w iff $\overline{\texttt{HHT}} = \texttt{TTH}$ precedes $\overline{\texttt{TTH}} = \texttt{HHT}$ in $\bar{w}$. And g is onto by the same reasoning. Since g is clearly an injection, we can conclude that it is a bijection.

Then we observe that $\Pr(w) = \Pr(g(w))$ for any w. This is because $\Pr(H) = \Pr(T)$ and $g(w)$ has the same length as w. Hence,

$$\Pr(A) = \sum_{w \in A} \Pr(w) = \sum_{w \in A} \Pr(g(w)) = \sum_{w' \in B} \Pr(w') = \Pr(B).$$

The second equality is valid because g preserves the probability, and the third by the bijection property with $w' = g(w)$. Note that the fact that `H` and `T` are equally likely is critical in these calculations; this analysis would fail for a biased coin.

Finally we have to show that $\Pr(A \cup B) = 1$. This follows from the fact that the only way never to throw either pattern is to throw all `H`'s or all `T`'s after the first toss, and we know that the probability of there being an unbounded number of tosses of only H or only T is zero. That is, $\Pr\left(\overline{A \cup B}\right) = 0$ and so $\Pr(A \cup B) = 1$. Since A and B are disjoint, this means that $\Pr(A) + \Pr(B) = 1$ and hence

$$\Pr(A) = \frac{1}{2}.$$

■

(b) [7 pts] What is the probability you see the sequence `HTT` before you see the sequence `HHT`?

Hint: Try to find the probability that `HHT` comes before `HTT` conditioning on whether you first toss an `H` or a `T`. Somewhat surprisingly, the answer is not 1/2.

Solution. Let A be the event that `HTT` appears before `HHT`, and let $p ::= \Pr(A)$.

Suppose our first toss is `T`. Since neither of our patterns starts with `T`, the probability that A will occur from this point on is still p. That is, $\Pr(A \mid T) = p$.

Suppose our first toss is `H`. To find the probability that A will now occur, that is, to find $q ::= \Pr(A \mid H)$, we consider different cases based on the subsequent throws.

Suppose the next toss is H, that is, the first two tosses are HH. Then neither pattern appears if we continue flipping H, and when we eventually toss a T, the pattern HHT will then have appeared first. So in this case, event A will never occur. That is $\Pr(A \mid \texttt{HH}) = 0$.

Suppose the first two tosses are HT. If we toss a T again, then we have tossed HTT, so event A has occurred. If we next toss an H, then we have tossed HTH. But this puts us in the same situation we were in after rolling an H on the first toss. That is, $\Pr(A \mid \texttt{HTH}) = q$.

Summarizing this we have:

$$\begin{aligned}
\Pr\{A\} &= \Pr(A \mid T)\Pr\{T\} + \Pr(A \mid H)\Pr\{H\} && \text{(Law of Total Probability)} \\
p &= p\frac{1}{2} + q\frac{1}{2} && \text{so} \\
p &= q.
\end{aligned}$$

Continuing, we have

$$\begin{aligned}
\Pr(A \mid H) &= \Pr(A \mid \texttt{HT})\Pr\{T\} + \Pr(A \mid \texttt{HH})\Pr\{H\} && \text{(Law of Total Probability)} \\
q &= \Pr(A \mid \texttt{HT})\frac{1}{2} + 0 \cdot \frac{1}{2} && (1) \\
\Pr(A \mid \texttt{HT}) &= \Pr(A \mid \texttt{HTT})\Pr\{T\} + \Pr(A \mid \texttt{HTH})\Pr\{H\} && \text{(Law of Total Probability)} \\
\Pr(A \mid \texttt{HT}) &= 1 \cdot \frac{1}{2} + q\frac{1}{2} && (2) \\
q &= (\frac{1}{2} + \frac{q}{2})\frac{1}{2} && \text{by (??) \& (??)} \\
q &= \frac{1}{3}.
\end{aligned}$$

So HTT comes before HHT with probability

$$p = q = \frac{1}{3}.$$

These kind of events are have an amazing *intransitivity* property: if you pick *any* pattern of three tosses such as HTT, then I can pick a pattern of three tosses such as HHT. If we then bet on which pattern will appear first in a series of tosses, the odds will be in my favor. In particular, even if you instead picked the "better" pattern HHT, there is another pattern I can pick that has a more than even chance of appearing before HHT. Watch out for this intransitivity phenomenon if somebody proposes that you bet real money on coin flips. ■

(c) [6 pts] Suppose you flip three fair, mutually independent coins. Define the following events:

- Let A be the event that *the first* coin is heads.
- Let B be the event that *the second* coin is heads.

- Let C be the event that *the third* coin is heads.
- Let D be the event that *an even number of* coins are heads.

Use the four step method to determine the probability of each of A, B, C, D.

Solution. The tree is a binary tree with depth 3 and 8 leaves. The successive levels branch to show whether or not the successive events A, B, C occur. By the definitions of the characteristics *fair* and *independent*, each branch from a vertex is equally likely to be followed. So the probability space has, as outcomes, eight length-3 strings of H's and T's, each of which has probability $(1/2)^3 = 1/8$.

Each of the events A, B, C, D are true in four of the outcomes and hence has probability $1/2$. ■

Problem 3. [20 points] Suppose you have seven standard dice with faces numbered 1 to 6. Each die has a label corresponding to a letter of the alphabet (A through G). A *roll* is a sequence specifying a value for each die in alphabet order. For example, one roll is $(6, 1, 4, 1, 3, 5, 2)$ indicating that die A showed a 6, die B showed 1, die C showed 4, ...

(a) [5 pts] What is the probability of a roll where *exactly* two dice have the value 3 and the remaining five dice all have different values?

Example: $(3, 2, 3, 1, 6, 4, 5)$ is a roll of this type, but $(1, 1, 2, 6, 3, 4, 5)$ and $(3, 3, 1, 2, 4, 6, 4)$ are not.

Solution. As in the example, map a roll into an element of $B ::= R_2 \times P_5$ where P_5 is the set of permutations of $\{1, \ldots, 5\}$. A roll maps to the pair whose first element is the set of colors of the two dice with value 6, and whose second element is the sequence of values of the remaining dice (in rainbow order). So $(3, 2, 3, 1, 6, 4, 5)$ above maps to $(\{\text{A,C}\}, (2, 1, 6, 4, 5))$. By the Product rule,

$$|B| = \binom{7}{2} \cdot 5!.$$

The probability is

$$\frac{\binom{7}{2} \cdot 5!}{6^6}$$

■

(b) [5 pts] What is the probability of a roll where two dice have an even value and the remaining five dice all have different values?

Example: $(4, 2, 4, 1, 3, 6, 5)$ is a roll of this type, but $(1, 1, 2, 6, 1, 4, 5)$ and $(6, 6, 1, 2, 4, 3, 4)$ are not.

Solution. Map a roll into a triple whose first element is in S, indicating the value of the pair of matching dice, whose second element is the set of colors of the two matching dice, and whose third element is the sequence of the remaining five dice values (in rainbow order).

So $(4, 2, 4, 1, 3, 6, 5)$ above maps to $(4, \{\text{A,C}\}, (2, 1, 3, 6, 5))$. Notice that the number of choices for the third element of a triple is the number of permutations of the remaining five values, namely 5!. This mapping is a bijection, so the number of such rolls equals the number of such triples. By the Generalized Product rule, the number of such triples is

$$3 \cdot \binom{7}{2} \cdot 5!.$$

The probability is

$$\frac{3 \cdot \binom{7}{2} \cdot 5!}{6^6}$$

Alternatively, we can define a map from rolls in this part to the rolls in part (a), by replacing the value of the duplicated values with 6's and replacing any 6 in the remaining values by the value of the duplicated pair. So the roll $(4, 2, 4, 1, 3, 6, 5)$ would map to the roll $(6, 2, 6, 1, 3, 4, 5)$. Now a type a roll, r, is mapped to by exactly the rolls obtainable from r by exchanging occurrences of 6's and i's, for $i = 1, \ldots, 6$. So this map is 6-to-1, and by the Division rule, the number of rolls here is 6 times the number of rolls in part (a).

■

(c) [10 pts] What is the probability of a roll where two dice have one value, two different dice have a second value, and the remaining three dice a third value?

Example: $(6, 1, 2, 1, 2, 6, 6)$ is a roll of this type, but $(4, 4, 4, 4, 1, 3, 5)$ and $(5, 5, 5, 6, 6, 1, 2)$ are not.

Solution. Map a roll of this kind into a 4-tuple whose first element is the set of two numbers of the two pairs of matching dice, whose second element is the set of two colors of the pair of matching dice with the smaller number, whose third element is the set of two letters of the larger of the matching pairs, and whose fourth element is the value of the remaining three dice. For example, the roll $(6, 1, 2, 1, 2, 6, 6)$ maps to the triple

$$(\{1, 2\}, \{\text{orange,green}\}, \{\text{yellow,blue}\}, 6).$$

There are $\binom{6}{2}$ possible first elements of a triple, $\binom{7}{2}$ second elements, $\binom{5}{2}$ third elements since the second set of two colors must be different from the first two, and 4 ways to choose the value of the three dice since their value must differ from the values of the two pairs. So by the Generalized Product rule, there are

$$\binom{6}{2} \cdot \binom{7}{2} \cdot \binom{5}{2} \cdot 4$$

possible rolls of this kind.

The probability is

$$\frac{\binom{6}{2} \cdot \binom{7}{2} \cdot \binom{5}{2} \cdot 4}{6^6}$$

■