Probability Answers

Brief answers to the probability questions on Self Assessment #2.

> Your hash function assigns each object to a number between 1:10, each with equal probability. With 10 objects, what is the probability of a hash collision? What is the expected number of hash collisions? What is the expected number of hashes that are unused?

First, calculate the probability of not having a hash collision. There are 10^{10} hash assignments, occurring with equal probability, and there are 10! ways to permute $1, 2, \ldots, 10$ corresponding to 10! no-collision outcomes. Then the probability of having a hash collision is $1 - \frac{10!}{10^{10}} = 0.9996371$.

Let's consider a more general problem first, which will be illuminating. Suppose that we have a set S of $|S|$ elements, and that we draw n elements from S into a collection of elements C. Let I_k be 1 if the kth object in S is in C at least once and 0 otherwise. Then the expected value of I_k, $\mathbb{E}(I_k)$, is equal to p_k, the probability that the kth object in S is repesented in C. This probability is $1 - \left(\frac{|S|-1}{|S|}\right)^n$, which is *independent of k*.

Now, we would like to find the expected number of distinct elements of S represented in C. Since all the elements are identical, this amounts to finding $\mathbb{E}(|S| \times I_k)$. Then we have

$$\mathbb{E}(|S| \times I_k) = |S| \times \mathbb{E}(I_k) = |S| \times p_k = |S| \times \left(1 - \frac{|S|-1}{|S|}\right)^n$$

.

Now, specifying to the case of our hash maps, we have $|S| = n = 10$, so then the corresponding expectation is 6.513216. This is the expected number of unique hashes, so the expected number of collisisions is $10 - 6.513216 = 3.486784$.

Every unused hash corresponds to one more hash collision, and zero unused hashes means zero hash collisions, so the two expected values are identical.

> Given a fair, 6-sided dice, what's the *expected number of rolls* you have to make before each number (1, 2, …, 6) shows up at least once?

In the following, time = number of rolls.

The expected time it takes to get the 1st unique number is 1. Conditional on having gotten one unique number, the probability of getting a second unique number is $\frac{5}{6}$, so the corresponding expected time is $\frac{6}{5}$. Similarly, the expected time of getting a third unique number, conditional on having gotten two unique numbers, is $\frac{6}{4}$. Continuing the pattern, the total expected time for 6 unique numbers is $\sum_{i=1}^{6} \frac{6}{i} = 14.7$.

> Bobo the amoeba can divide into 0, 1, 2, or 3 amoebas with equal probability. (Dividing into 0 means that Bobo dies.) Each of Bobo's descendants have the same probabilities. What's the probability that Bobo's lineage eventually dies out?

Let p be the probability of Bobo's lineage dying; then $p = \frac{1}{4} + \frac{1}{4}p + \frac{1}{4}p^2 + \frac{1}{4}p^3$. A solution is $p = \sqrt{2} - 1$. (Another solution is $p = 1$ but we can rule that out computationally.)