1. When a person sends a message, such as a bank transfer for \$1000 to my checking account, the message is encoded by the server as a number m. It is then compressed (or hashed) into a different number, $f(m)$, via some function f. What is known as a *birthday attack* is finding a different, possibly malicious, message m' such that $f(m) = f(m')$. For example, m' might be move the \$1000 to Matt's account.

   Suppose that the function $f(m)$ is equally likely to assume values in $[1, n]$. A hacker wants to find a *collision*—two values m and m' such that $f(m) = f(m')$. She does this by randomly generating numbers and evaluating them with f.

   What is the probability she finds a collision after k attempts?

   Solution: It looks complicated, but this is the birthday problem with 365 replaced by n. The probability k attempts fail is
   $$\frac{n \cdot (n-1) \cdots (n-k+1)}{n^k}.$$
   So the answer is 1 minus this.

2. A frugal individual collects coupons from cereal boxes. There are n coupons in total. Each box is equally likely to contain any one of them. Our goal is to discover how many boxes on average must be purchased to get all n coupons.

   (a) What is the probability the coupon collector opens $m + 1$ boxes and only has 1 distinct coupon?

   Solution: The first box will give a new coupon. For each of the remaining m boxes the probability of not getting a new coupon is $1/n$. So the probability is $(1/n)^m$.

   (b) Let X_k be the number of cereal boxes needed to go from k to $k + 1$ coupons. Explain why X_k is a geometric random variable.

   Solution: X_k is the waiting time for a success in independent trials with the same probability each attempt.

   (c) What is EX_k?

   Solution: The probability of successfully getting a new coupon is $p_k = 1 - (k/n) = \frac{n-k}{n}$. This is a geometric random variable with parameter p_k. We know that the mean is then $1/p_k = \frac{n}{n-k}$.

   (d) Let N be the total number of boxes needed to go from 0 to n coupons. Write a formula for N in terms of the X_k.

   Solution: $N = \sum_{k=0}^{k=n-1} X_k$.

   (e) What is EN?

Solution:

$$\begin{aligned} EN &= E\sum_{k=0}^{n-1} X_k \\ &= \sum_{k=0}^{n-1} EX_k \\ &= \sum_{k=0}^{n-1} \frac{n}{n-k} \\ &= n\sum_{k=0}^{n-1} \frac{1}{n-k}. \end{aligned}$$

(f) Recall from calculus that a Riemann Sum approximation gives $\sum_{k=0}^{n-1} \frac{1}{n-k} \leq \int_0^{n-1} \frac{1}{n-x}dx$. Evaluate this integral to give an upper bound on EN.

Solution: $\int_0^{n-1} \frac{1}{n-x}dx$. Let $u = n - x$ so that $-du = dx$ and the integral becomes

$$-\int_n^1 \frac{1}{u}du = \int_1^n \frac{1}{u}du = \log n - \log 1 = \log n.$$

Thus, $EN \leq n \log n$. Note that for large n this approximation becomes very good.

3. You walk into a casino with the intention of making $1 playing a game with a 30 sided dice. If you bet x dollars and the number you roll is less than or equal to 14 you win x dollars. Otherwise you lose those x dollars. Your strategy to win $1 is to first bet a dollar. If the bet wins, you walk away. If you lose the first round, then you double your bet and bet $2. So a win will cover your $1 lost from the previous round. The first time you win, you walk away. After k losses you must bet 2^k dollars to be up a dollar (check this if you want).

(a) What is the probability you lose k bets in a row?

Solution: $(16/30)^k$.

(b) What is the expected amount of money you win using this betting strategy?

Solution: $1. Because the wait time, let's call it T, is a geometric random variable with parameter $7/15$. The expected winnings are

$$\sum_{k=1}^{\infty} 1P(T = k) = 1.$$

(c) Suppose you walk in with $\sum_{k=0}^{N-1} 2^k = 2^N - 1$ dollars. So you can use the strategy at most N times. What are your expected winnings?

Solution: Like in Part (b) the expected winnings are

$$1P(T \leq N) - (2^{N+1} - 1)P(T > N)$$

Notice that $P(T > N) = (8/15)^N$. So we can write the above as

$$1 - (8/15)^N - (2^N - 1)(8/15)^N = 1 - (16/15)^N.$$

(d) Using the previous part, what are your expected earnings if you walk in with $\sum_{k=0}^{10} = 2047$ dollars?

Solution: $1 - (16/15)^{11} \approx -\$0.90$.

(e) Explain why (b) supports this as a good gambling strategy, and why (e) supports it being a bad strategy. Would you advise someone to use this gambling strategy?

Solution: Part (b) says you will eventually win \$1. However, it assumes infinitely many tries at the game, so you can't ever run out of money. In Part (e) you have a finite number of plays, so you are exposed to losing all of your money. This risk becomes worse and worse the more money you walk into the game with. Since you (presumably) will always have a finite amount of money, I would not advise using this strategy!

Answers to Homework 1

7. 3/4

10. $P(9) = 25/216 < P(1) = 27/216$

14. .73

20. (a) Yes (b) No

25. (a) .92 (b) .44

33. .4914

37. Same as two ordinary dice.

43. $-17/216$

48. 3

52. 30.4 and 1.76

56. .75