CS 630 – Fall 2024 – Lab 6
Oct 23, 2024

1. Pooled Testing

During COVID, the resources for detecting positive virus cases were limited. Therefore, people started using pooled testing: each test sample is mixed in a batch of size n. We will only retest each individual's sample if the batch tests positive. Suppose you are an employee at BU's health center and need to test $20,000$ students and faculty members, each with a positive rate of 1%.

1. What is the expected number of tests needed if we do not use pooled testing?

   Solution:
   Since we need to test everyone, the total amount of tests needed is 20,000.

2. Suppose you have a batch size of 10 for pooled sampling; what is the expected number of tests needed?

   Solution:
   The expected number can be calculate by adding the base number of test for all the batches with the expected number of additional tests. For each batch with size n, we have $(1 - 0.99^n)$ chance of getting a positive batch, meaning we need to have n more tests. In total we have $\frac{20,000}{n}$ batches. These gives us the following equation:

   $2000 + 2000 \cdot (1 - 0.99^{10}) \cdot 10 \approx 3912.36$

3. What is the optimal batch size to reduce the expected number of tests? (You don't have to provide the exact number, just the equation to solve it)

   Solution:
   To solve for the optimal batch size of the tests, we first create the equation to calculate the expected number of tests for a batch size of n:

   $$\frac{20000}{n} + \frac{20000}{n}(1 - 0.99^n)n = \frac{20000}{n} + 20000(1 - 0.99^n)$$

   Then we take the derivative of this equation:

   $$\frac{d}{dn}(\frac{20000}{n} + 20000(1 - 0.99^n)) = -\frac{20000}{x^2} - 0.99^x \ln 0.99 \cdot 20000$$

   You can find the local minimum using a calculator by setting the derivative equal to zero. Then answer is approximate to 10.51624.

2. Balls and Bins

We throw m balls into n bins uniformly at random. Answer the following questions.

1. What is the probability a particular bin will be empty?

Solution:
Let X_i the indicator random variable that takes the value of one when *ball i falls into that bin*. We see that $\Pr(X_i = 1) = 1/n$. We want to compute $\Pr(\cap_{i=1}^{m}\{X_i = 0\})$.

$$\Pr(\cap_{i=1}^{m}\{X_i = 0\}) \underset{=}{indep} \prod_{i=1}^{m} \Pr(X_i = 0)$$
$$= (1 - 1/n)^m$$

2. Use the fundamental inequality $x + 1 \leq \exp(x)$ to bound the previous probability.

Solution:
We replace x with $-1/n$ into $x + 1 \leq \exp(x)$.

$$1 - 1/n \leq \exp(-1/n)$$
$$(1 - 1/n)^m \leq exp(-m/n)$$

3. What is the probability that a particular bin has exactly k balls?

Solution:
This is equivalent to counting all possible sequences of type $\{0, 1\}^m$ that have exactly k aces. These are $\binom{m}{k}$. Let us now compute the probability for each one to happen. This is equal to $(1/n)^k(1 - 1/n)^{m-k}$ due to the independence. To sum up the probability is equal to $\binom{m}{k}(1/n)^k(1 - 1/n)^{m-k}$.

3. Shuffling Problem

n children are sitting in a row on n chairs. You can ask any two children at any time to swap places. If you do anything else with the kids it will end in chaos which you want to avoid at all cost. You want to reorder the children on their seats so that they sit in a uniform random order. Find an algorithm to do this with at most n swaps if you have access to a random number generator $rand(k)$ that takes as input an integer k and returns a random integer in the range $[1, k]$. Describe your algorithm and prove that it results in a uniform random order.

Solution: The algorithm we describe is called the Fisher-Yates Shuffle. It starts at the end of the array and iterates down to the first index. Each time it swaps the contents of the current index with a lower index at random.

1. Let n be the number of children.
2. For $i = n - 1$ down to 1, we get a random number j from $[0, i]$, then we swap the i-th child and the j-th child.

There are $n!$ possible permutations for a sequence of n elements. To prove that this algorithm yields a uniform random order of children, we have to show that the probability of any specific ordering is $\frac{1}{n!}$. The first `rand(i)` has a choice of n elements, hence the probability that any given child ends up in the last seat is $\frac{1}{n}$. The second call to `rand()` is over $n - 1$ candidates, hence the probability is $\frac{1}{n-1}$, etc., for index i the probability is $\frac{1}{i}$. The overall probability is the product of these:

$$\prod_{i=1}^{n} \frac{1}{i} = \frac{1}{n!}.$$

Probability Exercises

These are a non-exhaustive list of exercises to review probability concepts that are generally needed to follow the course content. You are expected to be comfortable with these and other concepts too which would be covered in a probability course.

1. Let A and B be two events such that $P(A) = 0.4$, $P(B) = 0.7$, $P(A \cup B) = 0.9$. And A^c denotes the complement event of A.

   (a) Find $P(A \cap B)$

   Solution:

   $$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.2$$

   (b) Find $P(A^c)$

   Solution:

   $$P(A^c) = 1 - P(A) = 0.6$$

   (c) Find $P(A^c \cap B)$

   Solution:

   $$P(A^c \cap B) = P(B) - P(A \cap B) = 0.5$$

   (d) Find $P(A \setminus B)$

   Solution:

   $$P(A \setminus B) = P(A) - P(A \cap B) = 0.2$$

   (e) Find $P(A|B)$

   Solution:

   $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2}{7}$$

2. Suppose we are rolling two six-sided dice, and let X_1 and X_2 be the random variable corresponding to the number that appears on the first and second die respectively. And let X be the random variable corresponding to the sum of the number that appears in both dice.

   (a) Find $E[X_1]$

   Solution:

   $$E[X_1] = \sum_{x=1}^{6} xP(X_i = x) = 3.5$$

   (b) Find $E[X]$

   Solution:

   $$E[X] = E[X_1 + X_2] = E[X_1] + E[X_2] = 7$$

   (c) Find $E[X|X_1 \text{ is even}]$

   Solution:

   $$\begin{aligned} E[X|X_1 \text{ is even}] &= E[X_1 + X_2|X_1 \text{ is even}] \\ &= E[X_1|X_1 \text{ is even}] + E[X_2|X_1 \text{ is even}] \\ &= E[X_1|X_1 \text{ is even}] + E[X_2] \\ &= \frac{1}{3}(2+4+6) + 3.5 \\ &= 4 + 3.5 \\ &= 7.5 \end{aligned}$$

(d) Find $E[X_1|X=9]$

Solution:
Notice that it $X = 9$ can only be achieved with these specific combination

$$\{(3,6),(4,5),(5,4),(6,3)\}$$

and since each number is equally likely to appear, we have the

$$E[X_1|X=9] = \frac{1}{4}(3+4+5+6) = 4.5$$

(e) Find $E[X|X_1 = X_2]$

Solution:

Since each number is equally likely to appear we have

$$E[X|X_1 = X_2] = \frac{1}{6}(2+4+6+8+10+12) = 7$$