Mathematics and Statistics Research Competition

Question 8

Jiamu Li & Frank Tang & Edward Wang

July 30, 2022

Contents

A particle generator is emitting two types of particles (called X and Y) into a long tube. The particles will line up in order after entering the tube. Initially, the tube is empty. At each shot, either an X- or Y-particle is randomly emitted into the tube with equal probability. Different shots are assumed to be independent from each other. Suppose that n shots have been emitted.

PROBLEM 1

- What is the probability that no two X-particles are next to each other?

THEOREM 1. *The probability that no two X-particles are next to each other after n shots is given by*

$$\frac{F_{n+2}}{2^n},$$

where F_n is the nth Fibonacci number.

Proof. The probability we require can be calculated by dividing the total number of ways to arrange the contents of the tube such that there are no consecutive X-particles, by the total number of arrangements of the particles. That is to say:

$$\Pr(\text{No consecutive X-particles}) = \frac{\#\text{Arrangements w/o consecutive X-particles}}{\#\text{Total arrangements}}$$

CLAIM 1.1. *The number of arrangements with no consecutive X-particles is*

$$\sum_{k=0}^{n} \binom{n-k+1}{k}.$$

Proof. Consider a tube with n particles in it. Let the number of X-particles be equal to k, so that the number of Y-particles is $n-k$. Consider the tube without the X-particles, consisting solely of Y-particles in a line:

$$\underbrace{\text{YY}\ldots\text{YY}}_{n-k}$$

Now consider the 'gaps' between these Y-particles, indicated by a bar (|):

$$|\text{Y}|\text{Y}|\ldots|\text{Y}|\text{Y}|$$

Notice that there are exactly $n-k+1$ 'gaps'. Clearly, if we were to only place X-particles in the gaps, then there would never be any consecutive X-particles. This can be done in a total of

$$\binom{n-k+1}{k}$$

ways. However, we must consider this for any number of X-particles k, so we arrive at the sum

$$\#\text{Arrangements with no consecutive X-particles} = \sum_{k=0}^{n} \binom{n-k+1}{k}. \qquad \square$$

CLAIM 1.2.

$$\sum_{k=0}^{n} \binom{n-k+1}{k} = F_{n+2},$$

where F_n is the nth Fibonnaci number.

Proof. Figure 1 showcases a way to obtain the identity from Pascal's triangle. We will present an algebraic proof below.

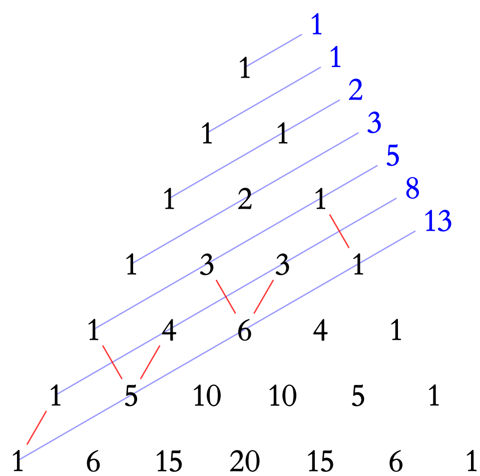

Figure 1: Each number in the row is the sum of the the numbers in the previous two rows.

Recall that the Fibonnaci numbers are defined as follows:

$$F_0 = 0,$$
$$F_1 = 1,$$
$$F_n = F_{n-1} + F_{n-2}. \qquad \text{(n > 1)}$$

We will define a function $f(n) = \sum_{k=0}^{n} \binom{n-k+1}{n}$. It is sufficient to prove that $f(1) = F_3 = 2$, $f(2) = F_4 = 3$, and that $f(n) = f(n-1) + f(n-2)$, which would then imply the result by definition of the Fibonnaci numbers.

It is obvious that $f(1) = \binom{2}{0} + \binom{1}{1} = 2$, which is equal to F_3. Next, $f(2) = \binom{3}{0} + \binom{2}{1} + \binom{1}{2} = 3$. Notice that we define $\binom{n}{k} = 0$ when $n < k$, as it is impossible to choose k things from a set with elements less than k.

We proceed to prove that $f(n) = f(n-1) + f(n-2)$, where $n > 2$. Using the fact that $\binom{n}{0} = 1$, we rewrite $f(n)$ using Pascal's identity and linearity as

$$f(n) = 1 + \sum_{k=1}^{n} \binom{n-k+1}{k} = 1 + \sum_{k=1}^{n} \binom{n-k}{k} + \sum_{k=1}^{n} \binom{n-k}{k-1}.$$

Next, we simplify, getting

$$\begin{aligned}
f(n) &= \sum_{k=0}^{n} \binom{n-k}{k} + \sum_{k=1}^{n} \binom{n-k}{k-1} \\
&= \sum_{k=0}^{n-1} \binom{n-k}{k} + \binom{n-n}{k} + \sum_{k=1}^{n} \binom{n-k}{k-1} \\
&= \sum_{k=0}^{n-1} \binom{n-k}{k} + 0 + \sum_{k=0}^{n-1} \binom{n-k-1}{k} \\
&= \sum_{k=0}^{n-1} \binom{n-k}{k} + \sum_{k=0}^{n-2} \binom{n-k-1}{k} + \binom{n-(n-1)-1}{n-1} \\
&= f(n-1) + f(n-2) + 0.
\end{aligned}$$

Hence $f(n)$ must be equivalent to F_{n+2}. □

Claim 1.3. *The number of total arrangements of a tube with n particles is*

$$2^n.$$

Proof. Each particle in the tube can be either an X-particle or a Y-particle, meaning there are 2 choices for each of the n particles. Hence, there are a total of 2^n arrangements. □

Hence by dividing the number of arrangements where there are no consecutive X-particles by the total number of arrangements, we arrive at the formula

$$\frac{F_{n+2}}{2^n}$$

which gives the desired probability. □

PROBLEM 2

Now, if two X-particles are next to each other, they will immediately collide into one single X-particle.

- Compute the average number of particles for $n = 2, 3, 4$.
- Can you find the pattern and establish an explicit formula for general n?

We found it most straightforward to proceed directly to finding a general formula. However, it should be noted that it is relatively simple to compute the averages for small n by considering every possible tube after n shots.

THEOREM 2. *The average number of particles after n shots is*

$$\frac{3}{4}n + \frac{1}{4}.$$

Proof. Let the average number of particles after n shots be T_n. Obviously, $T_1 = 1$. Next, consider the chance that after firing a shot, the number of particles *doesn't* decrease. This occurs only in the event that the last particle in the tube is an X-particle, and when the particle emitted is also an X-particle. Since both events have a probability 1/2, the probability that both occur is simply 1/4.

$$\begin{aligned} X &\to X : X \\ Y &\to X : YX \\ X &\to Y : XY \\ Y &\to Y : YY \end{aligned}$$

Hence, the probability that the number of particles *does* increase after firing a shot is $1 - 1/4 = 3/4$. Thus the expected number of particles increases by 3/4 after each shot, giving us the recursion

$$T_{n+1} = T_n + \frac{3}{4}.$$

Since $T_1 = 1$, we arrive at the formula

$$T_n = \frac{3}{4}n + \frac{1}{4}.$$ □

Using this, we can easily compute the average number of particles when $n = 2, 3, 4$:

n	T_n
1	1
2	1.75
3	2.5
4	3.25

PROBLEM 3

Suppose that at each shot, an X-particle is emitted with probability $p \in (0, 1)$.

- Under the same assumption as above, when n is very large, do you think the proportion of X-particles in the tube will eventually stabilise at a certain number? Why/why not?
- If so, can you compute this number explicitly?

THEOREM 3. *Suppose the probability of emitting an X-partical is $p \in (0, 1)$. The limiting ratio of X-particles in the tube, as the number of shots approaches infinity, is*

$$\frac{p}{1+p}.$$

Proof. We use a similar argument to that which is featured in Theorem 2 to arrive at a formula for the expected number of particles after n shots when the probability is p.

CLAIM 3.1. *The average number of particles after n shots when the probability of emitting an X-particle is $p \in (0, 1)$ is*

$$(1 - p^2)n + p^2.$$

Proof. Again, let the average number of particles after n shots be T_n. The number of particles *doesn't* increase when the last particle is an X-particle and the particle emitted is also an X-particle, which has a probabilty of p^2 of occuring. Hence, the probability that the number of particles *does* increase is $1 - p^2$, meaning that the expected number of particles increases by $1 - p^2$ after each shot, and thus we obtain

$$T_{n+1} = T_n + (1 - p^2).$$

Since $T_1 = 1$, the formula for T_n is

$$T_n = (1 - p^2)n + p^2. \qquad \square$$

We can now find the expected number of X-particles in the tube. Since the expected number of Y-particles in the tube is simply $(1 - p)n$, as the probability of emitting a Y-particle is $(1 - p)$ at each shot, the expected number of X-particles is

$$\begin{aligned} \#\text{X-particles} &= \#\text{Total particles} - \#\text{Y-particles} \\ &= T_n - (1-p)n \\ &= (1-p^2)n + p^2 - (1-p)n. \end{aligned}$$

We then divide this by the total number of particles to obtain the desired proportion and then take the limit as $n \to \infty$, to obtain

$$\lim_{n\to\infty} \frac{(1-p^2)n + p^2 - (1-p)n}{(1-p^2)n + p^2} = \lim_{n\to\infty} \frac{(1-p^2) + \frac{p^2}{n} - (1-p)}{(1-p^2) + \frac{p^2}{n}}.$$

By the algebraic limit theorem, we obtain

$$\begin{aligned}\frac{(1-p^2)+0-(1-p)}{(1-p^2)+0} &= \frac{p-p^2}{1-p^2} \\ &= \frac{p(1-p)}{(1+p)(1-p)} \\ &= \frac{p}{1+p}\end{aligned}$$

□

Generalisation

We were able to produce a generalisation of Problem 2:

- Suppose that if m X-particles are next to each other, they collapse into n X-particles.
- Assume that k particles have been emitted.
- The probability of firing an X-particle is p.
- Find the average number of particles.
- It is assumed that $m > n \geq 1$, otherwise the number of particles would simply blow up infinitely.

Theorem 4. *The average number of particles after k shots, when m X-particles collapse into n, is k when $k < m$, and otherwise is*

$$m - 1 + \sum_{b=m-1}^{k-1} \left(1 - p \left(\sum_{a=0}^{\lfloor \frac{b-m}{m-n} \rfloor} (1-p)p^{a(m-n)+m-1} + \varepsilon \right) (1 - (n - m + 1)) \right) \tag{1}$$

where $k \geq m$ and $\varepsilon = \begin{cases} \dfrac{1}{2^b}, & b - m + 1 = 0 \bmod m - n \\ 0, & k - m + 1 \neq 0 \bmod m - n. \end{cases}$

Proof. Let the average number of particles after k shots be denoted by T_k. It is obvious that $T_k = k$ when $k < m$, as it is impossible for any particles to have collapsed. Thus we must proceed to find a formula for T_k when $k \geq m$. We can achieve this by establishing a recurrence for T_k in order to find the difference between each term, and from there we may simply sum the difference to obtain the required formula.

We know that T_k is recursive since there are two events that can occur after each shot:

- Number of particles increases by 1
- Number of particles increases by $n - m + 1$

This is because if an X-particle were added to $m - 1$ consecutive X-particles, they would collapse into n X-particles. Hence, the number of particles 'increases' by $n - m + 1$, although this number is always negative or 0 (as $m > n$). Thus it remains to determine the probability that there are $m - 1$ consecutive X-particles at the end of the tube. Letting this probability be ϑ, the probability of the particles collapsing on the next shot is simply $p\vartheta$. As such, the expected number of particles increases by 1 with a probability of $1 - p\vartheta$ and increases by $n - m + 1$ with a probability of $p\vartheta$. Hence, the recurrence is:

$$\begin{aligned} T_{k+1} &= T_k + 1 - p\vartheta + p\vartheta(n - m + 1) \\ &= T_k + 1 - p\vartheta(1 - (n - m + 1)). \end{aligned} \tag{2}$$

The remaining step is to calculate ϑ.

CLAIM 4.1.

$$\vartheta = \sum_{a=0}^{\lfloor \frac{k-m}{m-n} \rfloor} (1-p)p^{a(m-n)+m-1} + \varepsilon,$$

where $\varepsilon = \begin{cases} \frac{1}{2^k}, & k-m+1 = 0 \bmod m-n \\ 0, & k-m+1 \neq 0 \bmod m-n. \end{cases}$

Proof. Recall that ϑ is the probability that there are $m-1$ consecutive X-particles at the end of the tube. In order to 'count' the number of consecutive X-particles, there must be some end to the string of X-particles. This 'end' can occur in two ways: either there is a Y-particle before last X-particle; or there are no Y-particles at all, and the tube compromises completely of X-particles.

$$\underbrace{\mathrm{X}\ldots\mathrm{X}}_{m-1}\mathrm{Y} \tag{A}$$

$$\underbrace{\mathrm{X}\ldots\mathrm{X}}_{m-1} \tag{B}$$

Let us first consider the case with a Y-particle at the end (configuration A). The probability of such a configuration occuring is simply $(1-p)p^{m-1}$. However, this probability does not account for possible previous collapses. For example, the following sequence of particles may have been emitted which would lead to the same configuration:

$$\underbrace{\mathrm{X}\ldots\mathrm{X}}_{m+(m-n)-1}\ \mathrm{Y} \quad \longrightarrow \quad \underbrace{\mathrm{X}\ldots\mathrm{X}}_{n+(m-n)-1}\ \mathrm{Y} \quad = \quad \underbrace{\mathrm{X}\ldots\mathrm{X}}_{m-1}\mathrm{Y}$$

It is clear that adding any multiple of $m-n$ X-particles to $m-1$ consecutie X-particles results in the exact same configuration. However, since only k particles have been fired, $a(m-n)+m-1+1 \leq k$ for some positive integer a. We solve this inequality for a, getting

$$\begin{aligned} a(m-n)+m &\leq k \\ a(m-n) &\leq k-m \\ a &\leq \frac{k-m}{m-n}. \end{aligned}$$

Thus the maximum value of a is $\left\lfloor \dfrac{k-m}{m-n} \right\rfloor$, so the probability of configuration A is

$$\sum_{a=0}^{\lfloor \frac{k-m}{m-n} \rfloor} (1-p)p^{a(m-n)+m-1}.$$

We now consider configuration B, which consists of X-particles only. Again, the configuration can be achieved with $a(m-n)+m-1$ consecutive X-particles, for some positive integer a. However, this time there cannot be any other particles other than these X-particles, meaning that $k = a(m-n) + m - 1$. It now becomes obvious that $k - m + 1$ must be an integer multiple of $m - n$, at which point there is a single sequence of particles which results in $m - 1$ consecutive X-particles. When $k - m + 1$ is not an integer multiple of $m - n$, there is no possible way for configuration B to exist. Thus the probability of configuration B occuring can be expressed by ε, where

$$\varepsilon = \begin{cases} \frac{1}{2^k}, & k - m + 1 = 0 \bmod m - n \\ 0, & k - m + 1 \neq 0 \bmod m - n. \end{cases}$$

We then add the probabilities of configurations A and B to get the overall probability of the tube ending in $m - 1$ consecutive X-particles, which is

$$\vartheta = \sum_{a=0}^{\lfloor \frac{k-m}{m-n} \rfloor} (1-p)p^{a(m-n)+m-1} + \varepsilon,$$

as required. □

Since we have previously determined a recursive formula for T_k in terms of ϑ and ε in Equation 2, we can simply sum the difference between each term to obtain a closed expression for the value of any term. We have

$$T_{k+1} = T_k + 1 - p\vartheta(1 - (n - m + 1)),$$

meaning the difference between terms is $1 - p\vartheta(1 - (n - m + 1))$. We then add $m - 1$ to the sum of this difference from $m - 1$ to k (since the formula is only required for $k \geq m$). This yields

$$T_k = m - 1 + \sum_{b=m-1}^{k-1} \left(1 - p \left(\sum_{a=0}^{\lfloor \frac{b-m}{m-n} \rfloor} (1-p)p^{a(m-n)+m-1} + \varepsilon \right) (1 - (n - m + 1)) \right). \qquad \square$$

It should be noted that ϑ is the sum of a geometric sequence, although for brevity, the sum is left unexpanded. The expanded sum is

$$\vartheta = \frac{(p-1)p^{m+n-1} \left(p^{(m-n)\left(\lfloor \frac{k-m}{m-n} \rfloor + 1\right)} - 1 \right)}{p^n - p^m}.$$

In addition, there also exists a closed form of ε, though it is somewhat distasteful. If we define a summation as 0 when its upper bound is less than its lower bound, then we obtain

$$\varepsilon = \sum_{a=0}^{(m-k-1)+\lfloor \frac{k-m+1}{m-n} \rfloor \cdot (m-n)} \frac{1}{2^k}.$$

The upper bound is simply the negative remainder of $k - m + 1$ when divided by $m - n$, which is 0 if and only if $k - m + 1$ is an integer multiple of $m - n$, and negative everywhere else. The lower bound, a is 0 simply to make the sum 0 whenever the upper bound is negative, and is not used elsewhere.

NICE SOLUTIONS OF GENERAL FORMULA

The general formula for the average length in Equation 1 is rather unwieldy. However, observe that many of the terms depend on $m-n$. If we just set $m-n = 1 \implies n = m-1$, then the formula can be reduced dramatically:

$$
\begin{aligned}
T_k &= m - 1 + \sum_{b=m-1}^{k-1} \left(1 - p\left(\sum_{a=0}^{\lfloor \frac{b-m}{1} \rfloor} (1-p)p^{a(1)+m-1} + \varepsilon\right)(1 - ((m-1) - m + 1))\right) \\
&= m - 1 + \sum_{b=m-1}^{k-1} \left(1 - p\left(\sum_{a=0}^{b-m} (1-p)p^{a+m-1} + \frac{1}{2^b}\right)\right) \qquad (3)
\end{aligned}
$$

Even $\varepsilon = \frac{1}{2^k}$ becomes constant, as $z = 0 \bmod 1$ for any integer z. Let us now evaluate ϑ, which is the sum of a geometric sequence:

$$
\begin{aligned}
\vartheta &= \sum_{a=0}^{k-m} (1-p)p^{a+m-1} + \frac{1}{2^k} \\
&= \sum_{a=0}^{k-m} (1-p)\left(p^{m-1}\right)p^a + \frac{1}{2^k} \\
&= (1-p)\left(p^{m-1}\right)\left(\frac{1-p^{k-m+1}}{1-p}\right) + \frac{1}{2^k} \\
&= p^{m-1}(1 - p^{k-m+1}) + \frac{1}{2^k} \\
&= p^{m-1} - p^k + \frac{1}{2^k}
\end{aligned}
$$

Now consider when $p = \frac{1}{2}$. ϑ now becomes

$$
\vartheta = \frac{1}{2^{m-1}},
$$

which is dependent only on m, meaning ϑ is constant. Hence the difference between the terms of T_k is constant when $n = m - 1$ and $p = \frac{1}{2}$, and so T_k would be linear. Let us now compute this difference more precisely. Using Equation 3, we see that the difference is

$$
1 - p\vartheta = 1 - \frac{1}{2^m}.
$$

The formula for T_k is then simply

$$
\begin{aligned}
T_k &= m - 1 + \sum_{b=m-1}^{k-1} \left(1 - \frac{1}{2^m}\right) \\
&= m - 1 + (k - m + 1)\left(1 - \frac{1}{2^m}\right). \qquad (4)
\end{aligned}
$$

Using Equation 4, we can easily derive the formula for T_k when $m = 2$ and $n = 1$, which is equivalent to Problem 2:

$$\begin{aligned} T_k &= 2 - 1 + (k - 2 + 1)\left(1 - \frac{1}{4}\right) \\ &= 1 + \frac{3}{4}(k-1) \\ &= \frac{3}{4}k + \frac{1}{4}. \end{aligned}$$

Code Implementation

```
#include <future>
#include <iostream>
#include <random>
#include <string>
#include <thread>
#include <vector>

using namespace std;

random_device rd;
mt19937 rng(rd());

const char particles[] = "XY";

unsigned int num_threads = thread::hardware_concurrency();

string gen_tube(int length, double p) {
  discrete_distribution<int> pick{p, 1 - p};
  string tube;
  for (int i = 0; i < length; i++) {
    tube += particles[pick(rng)];
  }
  return tube;
}

string annihilate(string tube) {
  string output = "";
  for (int i = 0; i < tube.length() - 1; i++) {
    if (tube[i] != tube[i + 1] && tube[i] == 'X')
      output += tube[i];
    else if (tube[i] != 'X')
      output += tube[i];
  }
  output.push_back(tube.back());
  return output;
}

bool check_consec_x(string tube) {
  for (int i = 0; i < tube.length() - 1; i++) {
    if (tube[i] == tube[i + 1] && tube[i] == 'X')
      return 0;
  }
  return 1;
}

double prob_no_two_consec_after_n(int n, int runs) {
  int count = 0;
  for (int i = 0; i < runs; i++) {
    if (check_consec_x(gen_tube(n, 0.5)))
      count++;
  }
  return double(count) / runs;
}

double average_len_after_n(int n, int runs) {
```

```cpp
  long long count = 0;
  for (int i = 0; i < runs; i++) {
    count += annihilate(gen_tube(n, 0.5)).length();
  }
  return count / double(runs);
}

double threaded_avg_len_after_n(int n, int runs) {
  double avg = 0;
  vector<future<double>> threads;
  for (int i = 0; i < num_threads; i++) {
    threads.push_back(
        async(launch::async, average_len_after_n, n, runs / num_threads));
  }
  for (auto &t : threads) {
    avg += t.get();
  }
  return avg / num_threads;
}

double proportion_x(int runs, double p) {
  string tube = "";
  vector<future<string>> threads;
  for (int i = 0; i < num_threads; i++) {
    string small_tube = gen_tube(runs / num_threads, p);
    threads.push_back(async(launch::async, annihilate, small_tube));
  }
  for (auto &t : threads) {
    tube += t.get();
  }
  int len = tube.length();
  return (len - (1 - p) * runs) / len;
}

int main() {

  unsigned long long runs = 1000000;

  cout << "q1\n";
  for (int i = 1; i <= 10; i++) {
    cout << "n = " << i << " prob = " << prob_no_two_consec_after_n(i, runs)
         << '\n';
  }
  cout << "q2\n";
  for (int i = 1; i <= 10; i++) {
    cout << "n = " << i << " avg length = " << threaded_avg_len_after_n(i, runs)
         << '\n';
  }
  cout << "q3\n";
  for (double p = 0; p <= 10; p++) {
    cout << "p = " << p / 10 << " proportion = " << proportion_x(runs, p / 10)
         << '\n';
  }

  return 0;
}
```

This code will output something similar (as it uses a random number generator) to:

```
q1
n = 1 prob = 1
n = 2 prob = 0.750097
n = 3 prob = 0.624354
n = 4 prob = 0.500569
n = 5 prob = 0.406112
n = 6 prob = 0.327949
n = 7 prob = 0.265757
n = 8 prob = 0.215079
n = 9 prob = 0.173859
n = 10 prob = 0.141454
q2
n = 1 avg length = 1
n = 2 avg length = 1.7502
n = 3 avg length = 2.49697
n = 4 avg length = 3.24526
n = 5 avg length = 3.99283
n = 6 avg length = 4.74027
n = 7 avg length = 5.49196
n = 8 avg length = 6.24677
n = 9 avg length = 6.99458
n = 10 avg length = 7.74433
q3
p = 0 proportion = 0
p = 0.1 proportion = 0.0908549
p = 0.2 proportion = 0.166695
p = 0.3 proportion = 0.231008
p = 0.4 proportion = 0.285435
p = 0.5 proportion = 0.334107
p = 0.6 proportion = 0.374551
p = 0.7 proportion = 0.412772
p = 0.8 proportion = 0.445532
p = 0.9 proportion = 0.472449
p = 1 proportion = 1
```

This output matches with the theoretical values obtained earlier.