COMP 2804 — Solutions Assignment 4

Question 1:

- Write your name and student number.

Solution: Diego Maradona[1], 10

Question 2: Consider six fair dice $D_1, \ldots, D_6$, each one having six faces. For each i with $1 \leq i \leq 6$, the die D_i has one face labeled i, whereas its other five faces are labeled zero.

You roll each of these dice once. Consider the random variable X, whose value is the sum of the results of these six rolls.

Determine the expected value $\mathbb{E}(X)$ of X.

Solution: We introduce random variables $X_1, \ldots, X_6$, where X_i is the result of rolling die D_i. Note that X_i is either i or 0. Since $X_i = i$ with probability $1/6$, and $X_i = 0$ with probability $5/6$, we have

$$\mathbb{E}(X_i) = i \cdot \Pr(X_i = i) + 0 \cdot \Pr(X_i = 0) = i/6.$$

Since $X = \sum_{i=1}^{6} X_i$, we have, using the Linearity of Expectation,

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{6} X_i\right) \\
&= \sum_{i=1}^{6} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{6} i/6 \\
&= \frac{1+2+3+4+5+6}{6} \\
&= 7/2.
\end{aligned}$$

Question 3: Consider a standard red die and a standard blue die; both of them are fair. You roll each die once. Consider the random variables

$$\begin{aligned}
X &= \text{the result of the red die plus the result of the blue die,} \\
Y &= \text{the result of the red die minus the result of the blue die.}
\end{aligned}$$

1. Prove that $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

[1] El Pibe de Oro

2. Are X and Y independent random variables? As always, justify your answer.

Solution: Let R be the result of the red die and let B be the result of the blue die. Then $X = R + B$ and $Y = R - B$.

By symmetry, $\mathbb{E}(R) = \mathbb{E}(B)$. Thus,

$$\mathbb{E}(Y) = \mathbb{E}(R - B) = \mathbb{E}(R) - \mathbb{E}(B) = 0$$

and, thus,

$$\mathbb{E}(X) \cdot \mathbb{E}(Y) = 0.$$

We next note that

$$X \cdot Y = (R + B)(R - B) = R^2 - B^2.$$

Thus,

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(R^2 - B^2) = \mathbb{E}(R^2) - \mathbb{E}(B^2).$$

By symmetry, $\mathbb{E}(R^2) = \mathbb{E}(B^2)$. Therefore, $\mathbb{E}(X \cdot Y) = 0$. This proves that

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Note: $\mathbb{E}(R^2) \neq (\mathbb{E}(R))^2$.

Are X and Y independent random variables? If $Y = 0$, then X must be even. Based on this,

$$\Pr(X = 11 \wedge Y = 0) \neq \Pr(X = 11) \cdot \Pr(Y = 0),$$

because the left-hand side is zero, whereas the right-hand side is $2/36$ times $1/6$, which is non-zero.

Question 4: When FX and his girlfriend XF have a child, this child is a boy with probability $1/2$ and a girl with probability $1/2$, independently of the sex of their other children. FX and XF stop having children as soon as they have a girl or four children.

Consider the random variables

$$\begin{aligned} C &= \text{the number of children that FX and XF have,} \\ B &= \text{the number of boys that FX and XF have.} \end{aligned}$$

Determine the expected values $\mathbb{E}(C)$ and $\mathbb{E}(B)$.

Solution: The sample space is

$$S = \{g, bg, bbg, bbbg, bbbb\}$$

and $\Pr(g) = 1/2$, $\Pr(bg) = 1/4$, $\Pr(bbg) = 1/8$, $\Pr(bbbg) = 1/16$, and $\Pr(bbbb) = 1/16$. (Sanity check: These probabilities add up to one.)

The possible values for C are $1, 2, 3, 4$. Since $\Pr(C = 1) = 1/2$, $\Pr(C = 2) = 1/4$, $\Pr(C = 3) = 1/8$, and $\Pr(C = 4) = 1/16 + 1/16 = 1/8$, we get

$$\begin{aligned}\mathbb{E}(C) &= \sum_{k=1}^{4} k \cdot \Pr(C = k) \\ &= 1 \cdot 1/2 + 2 \cdot 1/4 + 3 \cdot 1/8 + 4 \cdot 1/8 \\ &= 15/8.\end{aligned}$$

The possible values for B are $0, 1, 2, 3, 4$. Since $\Pr(B = 0) = 1/2$, $\Pr(B = 1) = 1/4$, $\Pr(B = 2) = 1/8$, $\Pr(B = 3) = 1/16$, and $\Pr(B = 4) = 1/16$, we get

$$\begin{aligned}\mathbb{E}(B) &= \sum_{k=0}^{4} k \cdot \Pr(C = k) \\ &= 0 \cdot 1/2 + 1 \cdot 1/4 + 2 \cdot 1/8 + 3 \cdot 1/16 + 4 \cdot 1/16 \\ &= 15/16.\end{aligned}$$

Remark: If G denotes the number of girls, then

$$\mathbb{E}(G) = \mathbb{E}(C - B) = \mathbb{E}(C) - \mathbb{E}(B) = 15/8 - 15/16 = 15/16 = \mathbb{E}(B).$$

Question 5: Consider the following algorithm, which takes as input an integer $n \geq 1$:

Algorithm MYSTERY(n)**:**

// all random choices made are mutually independent
$X = 0$;
for $i = 1$ **to** n
do a = random real number between -1 and 1;
    b = random real number between -1 and 1;
    if $a^2 + b^2 \leq 1$
    then $X = X + 1$
    endif
endfor;
$Y = \frac{4}{n} \cdot X$;
return Y

The output Y of this algorithm is a random variable. Determine the expected value $\mathbb{E}(Y)$ of this random variable.

Hint: If you implement this algorithm and run it several times for large values of n, then you may recognize the output. For the derivation of your value of $\mathbb{E}(Y)$, use indicator random variables.

Solution: We first determine $\mathbb{E}(X)$. We introduce indicator random variables $X_1, \ldots, X_n$, where

$$X_i = \begin{cases} 1 & \text{if } X \text{ is increased in iteration } i, \\ 0 & \text{otherwise.} \end{cases}$$

What is $\mathbb{E}(X_i)$? In iteration i, the algorithm chooses a uniformly random point (a, b) in a square with sides of length 2. This square has area 4. The value of X is increased if and only if this point is inside the circle with center at the origin and whose radius is 1. This circle has area π. Thus,

$$\mathbb{E}(X_i) = \Pr(X_i = 1) = \frac{\text{area circle}}{\text{area square}} = \pi/4.$$

Since $X = \sum_{i=1}^{n} X_i$, we have

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \\ &= \sum_{i=1}^{n} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{n} \pi/4 \\ &= \pi n/4. \end{aligned}$$

This gives

$$\mathbb{E}(Y) = \mathbb{E}\left(\frac{4}{n} \cdot X\right) = \frac{4}{n} \cdot \mathbb{E}(X) = \frac{4}{n} \cdot \pi n/4 = \pi.$$

Question 6: In the Lotto 6/49 lottery, you pick a 6-element subset X from the set $N = \{1, ..., 49\}$ and then a machine picks, uniformly at random, a 6-element subset Y from N.

1. The number $|X \cap Y|$ of numbers you picked correctly is a random variable. What is the expected value of this random variable? Give an exact answer, some Python code is provided below that can help you with the calculation.

2. [Warning: The following is an oversimplification, don't use it to make life choices.] The payout in Lotto 6/49 is relative to the Jackpot, which we will call x. The payout is defined as follows:

   - 6 correct numbers: x
   - 5 correct numbers $x/95$
   - 4 correct numbers $x/4365$
   - 3 correct numbers 10
   - 2 correct numbers 3

   If you are given one Lotto 6/49 ticket, what is your expected payout? Give an exact answer (which will include the variable x).

3. A Lotto 6/49 ticket costs \$3. What is the minimum jackpot value x that gives a payout of at least \$3?

Useful Python Code

```
morin@lauteschwein:~$ python3
Python 3.8.5 (default, Jul 28 2020, 12:59:40)
[GCC 9.3.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> from fractions import Fraction
>>> from math import factorial
>>> binom=lambda n,k: Fraction(factorial(n),factorial(k)*factorial(n-k))
>>> n = 5
>>> print(binom(n,3)/2**n)
5/16
>>> sum([binom(n,k) for k in range(n+1)])
Fraction(32, 1)
>>> sum([binom(n,k)/2**n for k in range(n+1)])
Fraction(1, 1)
```

Solution:

1. In the following, X is the 6-element subset from $\{1, ..., 49\}$ that you have picked before the actual lottery takes place; this subset is not random. During the lottery, the 6-element subset Y is chosen uniformly at random.

   The possible values for $|X \cap Y|$ are $0, 1, 2, 3, 4, 5, 6$. For each i with $0 \leq i \leq 6$, it follows from the Product Rule that

$$\Pr(|X \cap Y| = i) = \frac{\binom{6}{i}\binom{43}{6-i}}{\binom{49}{6}}. \tag{1}$$

   This gives

$$\mathbb{E}(|X \cap Y|) = \sum_{i=0}^{6} i \cdot \Pr(|X \cap Y| = i) = 36/49.$$

```
>>> pr = [ binom(6,i)*binom(43,6-i)/binom(49,6) for i in range(7)]
>>> print(", ".join([str(x) for x in pr]))
435461/998844, 68757/166474, 44075/332948, 8815/499422, 645/665896,
43/2330636, 1/13983816
>>> print(sum([i*pr[i] for i in range(7)]))
36/49
```

2. Let $p_i = \Pr(|X \cap Y| = i)$. We have seen an expression for p_i in (1). The expected payout is

$$x \cdot p_6 + \frac{x}{95} \cdot p_5 + \frac{x}{4365} \cdot p_4 + 10 \cdot p_3 + 3 \cdot p_2,$$

which is

$$\frac{683x}{1400661570} + \frac{572975}{998844}.$$

```
>>> payout_x = pr[6] + pr[5]/95 + pr[4]/4365
>>> print(payout_x)
683/1400661570
>>> payout_flat = pr[3]*10 + pr[2]*3
>>> print(payout_flat)
572975/998844
```

3. Solving

$$\frac{683x}{1400661570} + \frac{572975}{998844} \geq 3$$

gives

$$x \geq \frac{156331544285}{31418} \approx 4975859.19807117.$$

So the Jackpot needs to be nearly \$5,000,000 before you should even consider playing.

```
>>> print((3-payout_flat)/payout_x)
156331544285/31418
```

Note: People typically say that the 6/49 Jackpot needs to be at least \$14,000,000 before it's worth playing but when they do that they're only counting the probability of winning the entire jackpot $\Pr(|X \cap Y| = 6) = 1/\binom{49}{6} = 1/13983816$ (about one in 14 million).[2]

Question 7: Consider the following algorithm, which takes as input an integer $n \geq 1$:

[2] A complicating factor is that, if two or more people win the jackpot then they must split it evenly, the calculations here don't account for that at all.

```
Algorithm EULER(n):

    // all random choices made are mutually independent
    total = 0;
    i = 0;
    while total ≤ n
    do i = i + 1;
        x_i = uniformly random element in {1, 2, ..., n};
        total = total + x_i
    endwhile;
    return i
```

The output i of this algorithm is a random variable, which we denote by X.

1. What are the possible values that X can take? As always, justify your answer.
2. Let k be an integer with $0 \leq k \leq n$. Prove that

$$\Pr(X \geq k+1) = \binom{n}{k} \cdot (1/n)^k.$$

   Hint: What is the number of solutions of the inequality $x_1 + x_2 + \cdots + x_k \leq n$, where $x_1, x_2, \ldots, x_k$ are strictly positive integers?

3. Determine $\mathbb{E}(X)$.

   Hint: You may use the fact that $\mathbb{E}(X) = \sum_{k=0}^{n} \Pr(X \geq k+1)$.

4. Determine $\lim_{n \to \infty} \mathbb{E}(X)$.

Solution: We start with **part 1**:

- After the first iteration, *total* $\leq n$. Therefore, there are at least two iterations. Thus, $X \geq 2$.
- If $x_1 = n$, then $X = 2$.
- Assume that $X \geq n+2$. Then at the end of iteration $n+1$, *total* $\leq n$. However, at the end of iteration $n+1$, *total* $\geq n+1 > n$, which is a contradiction. Thus, $X \leq n+1$.
- If $x_1 = x_2 = \cdots = x_n = 1$, then $X = n+1$.
- We conclude that $X \in \{2, 3, \ldots, n+1\}$.

Next we do **part 2**:

Semi-formal proof: If $k = 0$, then, since X is always at least 2,

$$\Pr(X \geq k+1) = \Pr(X \geq 1) = 1 = \binom{n}{0} \cdot (1/n)^0.$$

Assume that $1 \leq k \leq n$. Observe that $X \geq k+1$ if and only if $x_1 + \cdots + x_k \leq n$. Since all x_i are at least 1, we rewrite this as

$$(x_1 - 1) + (x_2 - 1) + \cdots + (x_k - 1) \leq n - k.$$

Using $y_i = x_i - 1$, this becomes

$$y_1 + y_2 + \cdots + y_k \leq n - k,$$

where $y_1, \ldots, y_k \geq 0$ are integers. By Theorem 3.9.2 in the textbook, the number of solutions of this inequality is $\binom{n}{k}$. Since each y_i is a uniformly random element of $\{0, 1, \ldots, n-1\}$, the total number of possible sequences $(y_1, \ldots, y_k)$ is equal to n^k. Therefore,

$$\Pr(X \geq k+1) = \frac{\binom{n}{k}}{n^k} = \binom{n}{k} \cdot (1/n)^k.$$

A bit more formal: In case you don't like this argument, let's do this a bit more carefully. Let $t_1, \ldots, t_k$ be strictly positive integers such that $t_1 + \cdots + t_k \leq n$. (We know that there are $\binom{n}{k}$ many such sequences.) Since the random choices for the x_i are mutually independent,

$$\Pr(x_1 = t_1 \wedge x_2 = t_2 \wedge \cdots \wedge x_k = t_k)$$

is equal to

$$\Pr(x_1 = t_1) \cdot \Pr(x_2 = t_2) \cdots \Pr(x_k = t_k).$$

Since x_i is chosen uniformly at random from $\{1, 2, \ldots, n\}$, we have $\Pr(x_i = t_i) = 1/n$. It follows that

$$\Pr(x_1 = t_1 \wedge x_2 = t_2 \wedge \cdots \wedge x_k = t_k) = 1/n^k.$$

Since this is true for each of the $\binom{n}{k}$ solutions $(t_1, t_2, \ldots, t_k)$, it follows that

$$\Pr(X \geq k+1) = \frac{\binom{n}{k}}{n^k} = \binom{n}{k} \cdot (1/n)^k.$$

Rigorous proof: This is still not rigorous: Whether or not the i-th iteration[3] takes place depends on the values that are chosen for $x_1, \ldots, x_{i-1}$.

Here is a rigorous argument. We have already seen that the claim is true for $k = 0$. If $k = 1$, then, since X is always at least 2,

$$\Pr(X \geq k+1) = \Pr(X \geq 2) = 1 = \binom{n}{1} \cdot (1/n)^1.$$

[3] the iterations are numbered as $1, 2, 3, \ldots$

Assume that $2 \le k \le n$. Let $t_1, \ldots, t_k$ be strictly positive integers such that $t_1 + \cdots + t_k \le n$.

For each $i \ge 1$, let I_i be the event "iteration i takes place". Also, if the event I_i occurs, let x_i be the uniformly random element from $\{1, 2, \ldots, n\}$ that is chosen in iteration i. Finally, let A_i be the event

$$A_i = (I_i \wedge x_i = t_i).$$

The event "iteration k takes place and $x_1 = t_1, \ldots, x_k = t_k$" is equal to

$$A_1 \wedge A_2 \wedge \cdots \wedge A_k,$$

which is the same as

$$A_k \wedge A_{k-1} \wedge \cdots \wedge A_1.$$

We observe that

$$\Pr(A_k \wedge A_{k-1} \wedge \cdots \wedge A_1) = \left(\prod_{i=2}^{k} \Pr(A_i \mid A_1 \wedge \cdots \wedge A_{i-1}) \right) \cdot \Pr(A_1).$$

Where does this come from? If $k = 3$, then this becomes

$$\Pr(A_3 \wedge A_2 \wedge A_1) = \Pr(A_3 \mid A_1 \wedge A_2) \cdot \Pr(A_2 \mid A_1) \cdot \Pr(A_1),$$

which can be verified by applying the definition of conditional probability to the right-hand side and then simplifying the resulting expression.

We know that $A_1 = (I_1 \wedge x_1 = t_1)$. Since the first iteration is guaranteed to take place, $\Pr(A_1)$ is equal to the probability that x_1 is equal to t_1, which is $1/n$.

Let i be such that $2 \le i \le k$. We are going to determine

$$\Pr(A_i \mid A_1 \wedge \cdots \wedge A_{i-1}).$$

We are given that the event $A_1 \wedge \cdots \wedge A_{i-1}$ occurs. This means that all iterations $1, 2, \ldots, i-1$ take place and $x_1 = t_1, \ldots, x_{i-1} = t_{i-1}$. Therefore,

$$x_1 + \cdots + x_{i-1} = t_1 + \cdots + t_{i-1} \le t_1 + \cdots + t_k \le n$$

and, thus, iteration i takes place. It follows that $\Pr(A_i \mid A_1 \wedge \cdots \wedge A_{i-1})$ is just the probability that x_i is equal to t_i, which is $1/n$.

We conclude that

$$\Pr(A_k \wedge A_{k-1} \wedge \cdots \wedge A_1) = \left(\prod_{i=2}^{k} \frac{1}{n} \right) \cdot \frac{1}{n} = 1/n^k,$$

which is the same answer as we got before.

For **part 3**, using the hint and Newton's Binomial Theorem, we have

$$\mathbb{E}(X) = \sum_{k=0}^{n} \Pr(X \ge k+1) = \sum_{k=0}^{n} \binom{n}{k} \cdot (1/n)^k = (1 + 1/n)^n.$$

Finally, for **part 4**, using a limit that you have seen in calculus,

$$\lim_{n\to\infty} \mathbb{E}(X) = \lim_{n\to\infty} (1 + 1/n)^n = e,$$

which is Euler's number.

Question 8: Let G be an undirected graph with n vertices and m edges. This graph is not necessarily connected. Recall that the degree of a vertex u, denoted $deg(u)$, is the number of edges that are incident on u. We assume that every vertex has degree at least one. If u and v are two vertices that are connected by an edge, then we say that v is a neighbor of u.

Consider the following experiment:

- Let x be a uniformly random vertex.
- Let y be a uniformly random neighbor of x.
- Let $X = deg(x)$ and $Y = deg(y)$.

1. Let $a > 0$ and $b > 0$ be real numbers. Prove that

$$\frac{a}{b} + \frac{b}{a} \geq 2,$$

with equality if and only if $a = b$.

Hint: Rewrite this inequality until you get an equivalent inequality which obviously holds.

Solution:

$$\frac{a}{b} + \frac{b}{a} \geq 2$$

is equivalent to

$$\frac{a^2 + b^2}{ab} \geq 2,$$

which (since $ab > 0$) is equivalent to

$$a^2 + b^2 \geq 2ab,$$

which is equivalent to

$$(a - b)^2 \geq 0,$$

which obviously holds. Equality holds if and only if $a - b = 0$.

2. Prove that the expected value $\mathbb{E}(X)$ of the random variable X satisfies

$$\mathbb{E}(X) = \frac{2m}{n}.$$

Solution: To analyze X, we only need to consider the uniformly random vertex x that is chosen. The sample space S is the vertex set of G. The random variable $X : S \to \mathbb{R}$ is given by $X(u) = deg(u)$. Using the expression for $\mathbb{E}(X)$ that uses the domain of X, we get

$$\mathbb{E}(X) = \sum_{u \in S} X(u) \cdot \Pr(u) = \sum_{u \in S} deg(u) \cdot \frac{1}{n} = \frac{1}{n} \sum_{u \in S} deg(u).$$

You have learned in COMP 1805 that in the latter summation, each edge of G is counted twice. Therefore,

$$\mathbb{E}(X) = \frac{2m}{n}.$$

3. Prove that the expected value $\mathbb{E}(Y)$ of the random variable Y satisfies

$$\mathbb{E}(Y) = \frac{1}{n} \cdot \sum_{u:\text{ vertex in } G} \left(\sum_{v:\text{ neighbor of } u} \frac{deg(v)}{deg(u)} \right).$$

Solution: Since the value of Y depends on both x and y, we use the sample space

$$S' = \{(u,v) : \ u \text{ is a vertex and } v \text{ is a neighbor of } u\},$$

which is the same as

$$S' = \{(u,v) : \ \{u,v\} \text{ is an edge}\}.$$

Note that elements of S' are ordered pairs, whereas edges of the graph are unordered pairs. Each edge $\{u,v\}$ of G gives two elements of S', namely (u,v) and (v,u).

The random variable $Y : S' \to \mathbb{R}$ is given by $Y(u,v) = deg(v)$. We will use the expression for $\mathbb{E}(Y)$ that uses the domain of Y:

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{(u,v) \in S'} Y(u,v) \cdot \Pr(u,v) \\
&= \sum_{u:\text{ vertex in } G} \left(\sum_{v:\text{ neighbor of } u} Y(u,v) \cdot \Pr(u,v) \right) \\
&= \sum_{u:\text{ vertex in } G} \left(\sum_{v:\text{ neighbor of } u} deg(v) \cdot \Pr(x = u \wedge y = v) \right).
\end{aligned}$$

If v is a neighbor of u, then

$$\Pr(x = u \wedge y = v) = \Pr(y = v \mid x = u) \cdot \Pr(x = u) = \frac{1}{deg(u)} \cdot \frac{1}{n}.$$

Plugging this in gives the expression that we are asked to prove.

4. Prove that

$$\sum_{u\text{: vertex in } G} \left(\sum_{v\text{: neighbor of } u} \frac{deg(v)}{deg(u)} \right) = \sum_{\{u,v\}\text{: edge in } G} \left(\frac{deg(v)}{deg(u)} + \frac{deg(u)}{deg(v)} \right).$$

Solution: Consider an arbitrary edge $\{p, q\}$ in the graph. This edge contributes $deg(q)/deg(p) + deg(p)/deg(q)$ to the right-hand side. What does it contribute to the left-hand side:

- If $u = p$ and $v = q$, then we get the term $deg(q)/deg(p)$.
- If $u = q$ and $v = p$, then we get the term $deg(p)/deg(q)$.
- Thus, the total contribution of the edge $\{p, q\}$ to the left-hand side is $deg(q)/deg(p) + deg(p)/deg(q)$.

This shows that the left-hand side is equal to the right-hand side.

5. Prove that

$$\mathbb{E}(Y) \geq \mathbb{E}(X),$$

with equality if and only if each connected component of G is regular (i.e., all vertices in the same connected component have the same degree).

Solution: Using parts 3 and 4, we get

$$\begin{aligned} \mathbb{E}(Y) &= \frac{1}{n} \cdot \sum_{u\text{: vertex in } G} \left(\sum_{v\text{: neighbor of } u} \frac{deg(v)}{deg(u)} \right) \\ &= \frac{1}{n} \cdot \sum_{\{u,v\}\text{: edge in } G} \left(\frac{deg(v)}{deg(u)} + \frac{deg(u)}{deg(v)} \right). \end{aligned}$$

From part 1, $deg(v)/deg(u) + deg(u)/deg(v) \geq 2$, with equality if and only if $deg(u) = deg(v)$. This gives

$$\mathbb{E}(Y) \geq \frac{1}{n} \cdot \sum_{\{u,v\}\text{: edge in } G} 2 = \frac{1}{n} \cdot 2m = \mathbb{E}(X),$$

with equality if and only if for every edge $\{u, v\}$, $deg(u) = deg(v)$, which is the same as saying that each connected component is regular.

Remark: Imagine that G represents a "friendship graph": Each vertex is a person, and an edge $\{u, v\}$ indicates that u and v are friends. Informally,

- the value of $\mathbb{E}(X)$ is equal to the number of friends that Average Joe has,
- the value of $\mathbb{E}(Y)$ is equal to the number of friends that an average friend of Average Joe has.

- The latter is at least the former.
- This is known as the *friendship paradox*. Wikipedia has an article on this.