

COMP170

Discrete Mathematical Tools for Computer Science

More on *“time until first success”*

Version 2.0: Last updated, May 13, 2007

Example 1

Throw a fair die until you see a 1.

Then throw it until you see a 2.

Continue until you see all of 3, 4, 5, 6, in that order.

How many times, on average, do you throw the die?

Set $X_1 = \#$ of throws until you see 1.

For $i > 1$: $X_i = \#$ of throws, starting from when you see $i - 1$ for the first time, until you see i for the first time.

$$\begin{array}{cccccc} \underbrace{2 \ 5 \ 1}_{X_1 = 3} & \underbrace{3 \ 1 \ 4 \ 2}_{X_2 = 4} & \underbrace{6 \ 5 \ 4 \ 4 \ 3}_{X_3 = 5} & \underbrace{1 \ 3 \ 4}_{X_4 = 3} & \underbrace{6 \ 4 \ 1 \ 2 \ 5}_{X_5 = 5} & \underbrace{1 \ 4 \ 3 \ 6}_{X_6 = 4} \end{array}$$

$$\begin{array}{cccccc}
 \underbrace{2 \ 5 \ 1}_{X_1=3} & \underbrace{3 \ 1 \ 4 \ 2}_{X_2=4} & \underbrace{6 \ 5 \ 4 \ 4 \ 3}_{X_3=5} & \underbrace{1 \ 3 \ 4}_{X_4=3} & \underbrace{6 \ 4 \ 1 \ 2 \ 5}_{X_5=5} & \underbrace{1 \ 4 \ 3 \ 6}_{X_6=4}
 \end{array}$$

Total number of throws is

$$X = X_1 + X_2 + \dots + X_6.$$

$$X = 19$$

X_i is a geometric random variable with $p = 1/6$,

$$\Rightarrow E(X_i) = \frac{1}{p} = 6.$$

Then, by linearity of expectation,

$$E(X) = E(X_1) + E(X_2) + \dots + E(X_6) = 6 \cdot 6 = 36.$$

Example 2

Throw a fair die until you have seen all 6 numbers.

How many times do you throw the die on average?

Let N_i be the i th new number we see.

Let $X_1 = 1$:

For $i > 1$: $X_i = \#$ of throws needed to get N_i after first time we see N_{i-1} .

Example

	2	2	5	2	3	3	2	3	6	5	2	6	5	1	2	1	5	6	1	3	4
i	1	2	3						4					5							6
N_i	2	5	3						6					1							4
X_i	1	2	2						4					5							7

Example 2 (cont'd)

$$X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6$$

$$\Rightarrow E(X) = E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) + E(X_6)$$

so we need to calculate all of the $E(X_i)$

$i = 1$: $X_1 = 1$ so $E(X_1) = 1$.

$i = 2$: Once N_1 is chosen, X_2 is the number of times we need to throw the die until we see something that is *not* N_1 . Since being “*not* N_1 ” occurs with probability $\frac{5}{6}$, X_2 is geometric with $p = \frac{5}{6}$ so $E(X_2) = \frac{1}{p} = \frac{6}{5}$.

$i = 3$: Similarly, once N_1, N_2 are chosen, X_3 is the number of times we need to throw the die until we see something that is *not* N_1, N_2 . Since being “*not* N_1, N_2 ” occurs with probability $\frac{4}{6}$, X_3 is geometric with $p = \frac{4}{6}$ so $E(X_3) = \frac{1}{p} = \frac{6}{4}$.

Example 2 (cont'd)

$$X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6$$

General i In the general case, once N_1, N_2, \dots, N_{i-1} are chosen, X_i is the number of times we need to throw the die until we see something that is *not* N_1, N_2, \dots, N_{i-1} . Since being “*not* N_1, N_2, \dots, N_{i-1} ” occurs with probability $\frac{6-(i-1)}{6}$, X_i is geometric with $p = \frac{6-(i-1)}{6}$ so $E(X_i) = \frac{1}{p} = \frac{6}{6-(i-1)}$.

$$\begin{aligned} \Rightarrow E(X) &= E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) + E(X_6) \\ &= \frac{6}{6} + \frac{6}{5} + \dots + \frac{6}{1} = 6 \sum_{i=1}^6 \frac{1}{i} = 6 \cdot \frac{49}{20} = \frac{147}{10}. \end{aligned}$$

Compare this to previous problem in which we needed $6 \cdot 6$ flips on average, to see the numbers in order.

Example 3

This is known as the coupon collectors problem.

There are n coupons.

If you collect all of them, you win a prize.

Each time you go to the store, you get a random coupon.

How long do you need (in expectation) to collect all of the coupons?

$$X = X_1 + X_2 + \dots + X_n.$$

$$X_1 = 1.$$

For $i > 1$: X_i = time needed to receive i th new coupon after having received $(i - 1)$ st new coupon.

Example 3 (cont'd)

X_i is geometric with $p = (n - (i - 1))/n$.

So,

$$E(X_i) = \frac{n}{n - (i - 1)}$$

$$E(X) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \frac{n}{n - (i - 1)} = n \sum_{i=1}^n \frac{1}{i}$$

We just showed that

$$E(X) = n \sum_{i=1}^n \frac{1}{i}$$

$H_n = \sum_{i=1}^n 1/i$ has a special name. It is called the n^{th} **harmonic number**.

It is also known that $\forall n \ |H_n - \ln n| \leq 2$.

So H_n grows like $\ln n$ and

$E(X)$ grows like $n \ln n$.