CS314H CSB Note
Geometric Distribution

According to Knuth's analysis, the expected number of **successful** probe (α = load factor) is

$$\frac{1}{2}(1+\frac{1}{1-\alpha})$$

Let's find out the intuition behind this analysis.

- What is the best case of successful probe (Min)? Found at its _____ location (1 probe).
- What is the worst case of successful probe(Max)? Found at the ____ of the cluster.
  This is the same as cluster size. (Let's call this L.) Note home is the start of the cluster.
- What is expected number of successful probe? Assuming _________ distribution (the key to find is equally likely to be placed across the cluster): $\frac{1}{2}(Min + Max)$

- What is the expected number of cluster size L?

This is the same question as: **what is the expected number of heads until a tail appears?**
Let's consider a biased coin with probability of head h is 0.3 and probability of tail is 0.7 (1-h).
Let Pr{k throws} denote the probability that exactly k dice throws are needed to get a tail.
This means the earlier k-1 throws are _______ and the last throw (the k^{th} one) is ______.

Pr{k throws} = $0.3^{k-1} \times 0.7^{1}$

The expected number of dice throw can be calculated as follows:

$E[X] = \times 0.7 + 2 \times (0.3) \times 0.7 + 3 \times (0.3)^2 \times (0.7) + 4 \times (0.3)^2 \times (0.7) \ldots$

$= \sum_1^{\infty} k \times \Pr\{k \text{ throws}\} = \sum_1^{\infty} k \times 0.3^{k-1} \times 0.7 = 0.7 \sum_1^{\infty} k \times 0.3^{k-1}$

$= (1-h) \sum_1^{\infty} k \times h^{k-1}$

In our case, what is the expected number of cluster size L?
Cluster length L+1: (k-1) consecutive slots filled followed by 1 empty slot.
(Note L = k-1 as L doesn't include the last empty slot.)
What is the probability of a slot being filled? _______.

Pr{k slots} = $\alpha^{k-1} \times (1-\alpha)$
E[X] = $\sum_1^{\infty} k \times$ Pr{k slots} = _____________________.

Z=_____________________ S= _____________________

αZ=_____________________ αS=_____________________

Z(1-α) $=1+\alpha^1+\alpha^2+\alpha^3+\ldots$ = Let this be S.
S = ____________________ (This includes the empty slot.)
S-1 is the expected number of cluster size (excluding the empty slot).

But considering another factor, Knuth adds 1 back. When you pick a "random" slot inside the cluster, you will likely pick longer clusters (not all clusters are equally chosen).

Thus, expected value for the number of longest probe is S (Max).