Assignment 2

Tomaž Hribernik

November 2019

1 Drop sort

First we define the indicator variable

$$X_i = \textit{event where we keep the i-th item in the array}$$

In a uniformly randomly distributed array, each sub-array i also uniformly randomly distributed. That means that the probability of the last element being largest in a sub-array of size x, is $\frac{1}{x}$, as each place in the array has an equal probability of having the largest element. This means that we keep the i-th element if it is biggest element up to this point $\Rightarrow$

$$P(X_i) = \frac{1}{i} \Rightarrow E(X_i) = \frac{1}{i}, \textit{where i is the position in the array}$$

Now we can calculate the expected value of n elements, adjusting for the first two, since we always keep them. We can also use the same way of calculation for both ascending and descending since the probabilities do not change

$$E(X) = 2 + E(X_3 + X_4 + \cdots + X_n) = 2 + E(\sum_{i=3}^{n} X_i) = 2 + \sum_{i=3}^{n} E(X_i)$$
$$E(X) = 2 + \sum_{i=1}^{n} E(X_i) - \sum_{i=1}^{2} E(X_i) = 2 + \sum_{i=1}^{n} \frac{1}{i} - \sum_{i=1}^{2} \frac{1}{i}$$
$$E(X) = 2 + H_n - \frac{1}{1} - \frac{1}{2} = \frac{1}{2} + H_n \approx 0.5 + ln(n)$$

2 Bloom filter

a) How many bits are set on 1?
Note: in both a) and b) I assume that each insert is independent
First we define the indicator variable

$$X_i = \textit{event where a specific bit is set to 0 after n inserts}$$
$$1 - \frac{1}{n} = \textit{probability of a bit being 0 after 1 insert of 1 hash function}$$
$$(1 - \frac{1}{n})^k = \textit{probability of a bit being 0 after 1 insert of k hash functions} \Rightarrow$$
$$P(X_i) = ((1 - \frac{1}{n})^k)^n = \textit{probability of a bit being 0 after n inserts of k hash functions} = ((1 - \frac{1}{n})^n)^k = e^{-k} \Rightarrow E(X_i) = e^{-k}$$

Now calculate the expected number of 0 bits in the array using the indicator variable

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$$
$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} e^{-k} = ne^{-k} = ne^{-5} = \textit{expected number of 0 bits}$$
$$E(\textit{number of 1 bits}) = \textit{total number of bits} - \textit{number of 0 bits} = n - ne^{-5}$$
$$E(\textit{number of 1 bits}) = n(1 - e^{-5}) \approx 0.993262n$$

b) What is the probability of a false positive?
We already calculated the probability of any bit being 0

$$P(\textit{a bit is 0 after n insertions}) = e^{-k} \Rightarrow$$
$$P(\textit{a bit is 1 after n insertions}) = 1 - e^{-k}$$

For a false positive to occur all k bits checked by hash functions must be 1

$$P(\textit{false positive}) = (P(\textit{a bit is 1 after n insertions}))^k = (1 - e^{-k})^k = (1 - e^{-5})^5 \approx 0.966761215$$

3 Sticker album

First we define the indicator variable

$$X_i = \textit{event where we get a sticker we don't already have} \sim G(p)$$
$$P(X_i) = \frac{n-i}{n}, \textit{ where i is the number of stickers we already own}$$

As the indicator variable X_i is distributed geometrically, it's expected value is

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

Now we can calculate the expected value for n stickers, as n repeats of X_i gives us exactly a complete set of stickers

$$E(X) = E(X_0 + X_1 + \cdots + X_{n-1}) = E(\sum_{i=0}^{n-1} X_i) = \sum_{i=0}^{n-1} E(X_i)$$
$$E(X) = \sum_{i=0}^{n-1} E(X_i) = \sum_{i=0}^{n-1} \frac{n}{n-i} = n \sum_{i=0}^{n-1} \frac{1}{n-i} = n(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + \frac{1}{1})$$
$$E(X) = n(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n-1} + \frac{1}{n}) = n * H_n \approx n * ln(n)$$