1. (2 marks) Power of two choices
   A. Reduces the worst case performance of chain hashing
   B. Is the central working principle in bloom filters
   C. allows for us to make multiple sample in a reservoir sample
   D. reduces the burden of choosing the best hashing function

2. (2 marks) In a Bloom filter with $k$ hash functions, why would you want $k$ to be large? (Assume $p$ is the fraction of entries in the Bloom filter that are set to 1.)
   A. The probability of false positive is $p^k$.
   B. The probability of false negative is $p^k$.
   C. The probability that a particular cell is unset after $m$ items hashed $\propto p^{-k}$.
   D. The probability that a particular cell is set after $m$ items hashed $\propto p^{-k}$.

3. (2 marks) You are a hacker and want to crack a hash function $f$ used in a website. The goal of the attack is to find two different inputs $x_1, x_2$ such that $f(x_1) = f(x_2)$. Such a pair $x_1, x_2$ is called a collision. Your methodology to find a collision is simply to evaluate the function $f$ for different input values that may be chosen randomly until the same result is found more than once. The information known about the function is that $f(x)$ yields any of $H$ different outputs with equal probability. How many times should you evaluate $f$ to find a collision with a probability of at least $1/2$?
   A. $2\exp^{\sqrt{H}}$.
   B. $H + 1$.
   C. $2\sqrt{H}$.
   D. $\log H$.

   **Solution:** Consider Birthday paradox

4. (2 marks) A fair coin is flipped $n$ times. Let $X_{ij}$, $1 \leq i < j \leq n$, be 1 if the $i_{th}$ and $j_{th}$ flip landed on the same side and 0 otherwise. What can you say about the different random variables $X_{ij}$?
   A. $X_{ij}$ are pairwise independent but not independent.
   B. $X_{ij}$ are independent.
   C. none of these.

5. (2 marks) In chain hashing what is worst case query time, if we have added so far D elements, and the number of buckets is $n$?
6. You have a 2 Bloom filters with the same size and same set of hash functions.
   (a) (2 marks) Is the bit wise OR of the two bloom filters, same as the bloom filters formed out of union of the 2 sets? Why?
   A. No, a bloom filter changes completely as we change the set of input.
   B. No, a bloom filter depends on the order of arrival of points.
   C. Yes. In fact, the two bloom filters can be combined using any binary logic operations.
   D. Yes. The OR operation works because it does not lose any “1” bit set in either of the original Bloom filters.
   (b) (2 marks) If we do bit wise AND of the two bloom filters and use it as a bloom filter for intersection of the two sets. What can you comment about false negatives?
      A. There will be no false negatives since for each member the bit values in corresponding columns in all rows will be 1 for both bloom filters and the AND operation will preserve it.
      B. There will be no false negatives since bloom filters can be combined using any binary logic operations.
      C. The false negative rate will increase since there will always exist a position in bloom filter which will be mapped to 1 in one bloom filter and 0 in other.
      D. Nothing can be said with certainty.
   (c) (2 marks) If we do bit wise AND of the two Bloom filters and use it as a Bloom filter for intersection of the two sets. What can you comment about false positives?
      1. The false positive probability in the resulting Bloom filter is at most the false-positive probability in one of the constituent Bloom filters
      2. The false positive probability may be larger than the false positive probability in the Bloom filter created from scratch using the intersection of the two sets.
      3. The false positive probability will be at most that of the Bloom filter created from scratch using the intersection of the two sets.
         A. 1 & 3
         B. 1 & 2
         C. 2
         D. 3

7. Consider the problem of Balls and Bins with $m$ balls and $n$ bins.
   (a) (2 marks) What is the probability of any two balls $i$ and $j$ colliding?
      A. $1/n$
      B. $m/n$
      C. $2/n$
      D. $1/n^2$

   **Solution:** $Pr(Collision_{ij}) = \sum_{k=1}^{n} Pr(bin_i = k)Pr(bin_j = k) = n * 1/n^2 = 1/n$
   (b) (2 marks) What is the expected number of pairwise collisions. Note $\binom{a}{b}$ means $a$ choose $b$
      A. $n - m$
      B. $m^2$
      C. $\binom{n}{2}$
      D. $\frac{1}{\binom{m}{2}}$
Solution: Let us define the random variable $X$ = The total number of collisions, and, the indicator random variable $X_{ij}$ for collision of ball $i$ and $j$. Hence, $X = \sum_{i>j} X_{ij}$.

So, $E(X) = \sum_{i>j} E[X_{ij}] = \frac{1}{n} \binom{m}{2}$