Problem 1. Explain why Bloom filters do not support removing elements.

Because multiple elements may be mapped to the same bit, if removing one element, leading to clearing the corresponding bit, then there is higher probability to get a false positive since there may be other elements hashing to the same bit.

Problem 2. There are m candies on the table. Alice and Bob independently pick k random candies each. Suppose that A is the set of candies chosen by Alice, and B is the set of candies chosen by Bob. What is the expected size of the union $A \cup B$? Please, keep in mind that A and B are not necessarily disjoint.

Since $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, then, $E[|A \cup B| = E[|A|] + E[|B|] - E[|A \cap B|]$ $= 2k - E[|A \cap B|]$. Let X be a Bernoulli random variable such that for i from 1 to m, $X_i = 1$ if i is in the set of $A \cap B$ and 0 otherwise. Then, since X_i are independent, $E[|A \cap B|] = E[\sum_{i=1}^{m} X_i] = \sum_{i=1}^{m} E[X_i] = mE[X_1] = m(\frac{k}{m})^2 = \frac{k^2}{m}$ following a hypergeomtric distribution. Thus, $E[|A \cup B|] = 2k - \frac{k^2}{m}$.

Problem 3. Let X be a random integer number from 0 to $2^d - 1$. Denote by $M(X)$ the length of the maximum consecutive sequence of 1's in the binary representation of X. Find $E[M(X)]$ up to a constant multiplicative factor.

(1) $Pr\{M(x) \leq 2f(d)\} \leq \frac{1}{2}$

Let $y = C''f(d) = C''log_2(d)$ for some constant C''. Let $M'(s)$ denotes the length of the maximum consecutive sequence of 1's of a binary string s of length n. Let A_i denotes an event that there exists a subsequence s' at position $i \in [1, d+1-(y+1)]$ within a fixed binary string such that there are at least $y+1$ consecutive 1's in the subsequence. Then, $Pr\{\text{at most y consecutive 1's in the subsequence}\} = 1 - Pr\{A_i\}$. By union bound, $Pr\{M'(s') \leq y\} = 1 - Pr\{M'(s) > y\} \leq 1 - Pr\{\bigcup_{i=1}^{d+1-(y+1)} A_i\} \leq 1 - \sum_{i=1}^{d-y} P(A_i) = 1 - (d-y)P(A_i) \leq 1 - (d-y)(\frac{1}{2})^{y+1}$

Let B_n denotes the number of binary strings of length n such that $n \in [0, d]$ and $M'(n) \leq y$. $Pr\{M'(s) \leq y\} \leq \binom{d}{y}(\frac{1}{2})^d(1 - (d-y)(\frac{1}{2})^{y+1})$

$\leq (1 - (d - 2log_2 d)(\frac{1}{2})^{2log_2 d+1})(\frac{1}{2})^d(\frac{de}{2log_2 d})^{2log_2 d}$. when d = 2, this function is less than 1/2. Since it's monotonic decreasing for $d \geq 2$, $Pr\{M'(s) \leq y\} \leq \frac{1}{2}$. Since X is a random integer between 0 to $2^d - 1$, $Pr\{M(x) \leq 2f(d)\} \leq \frac{1}{2}$.

(2) $Pr\{M(x) \geq 2f(d)\} \geq \frac{1}{2}$

(3) $Pr\{M(x) \geq C'f(d)\} \leq \frac{1}{d}$

Let A_i denotes an event that there exists a subsequence s' at position $i \in [1, d+1-y]$ within a fixed binary string such that there are at least y consecutive 1's in the subsequence. Let

$y = df(d)$. Then,
$Pr\{M'(s') \leq y\} \leq Pr\{\bigcup_{i=1}^{d+1-y} A_i\} \leq \sum_{i=1}^{d+1-y} P(A_i) = (d+1-y)P(A_i) \leq (d+1-y)(\frac{1}{2})^y$
For any binary string s of length $n \in [0, d]$,
$Pr\{M'(s) \leq y\} \leq \binom{d}{y}(\frac{1}{2})^y(d+1-y)(\frac{1}{2})^y \leq (\frac{de}{dlog_2 d})^{dlog_2 d}(d-2log_2 d)(d+1-dlog_2 d)(\frac{1}{2})^{2dlog_2 d}$.

(4) conclude that $f(d) \leq E[M(x)] \leq Cf(d)$

By Markov Inequalities, $\frac{1}{2} \leq Pr\{X \geq 2f(d)\} \leq \frac{E[M(X)]}{2f(d)} \rightarrow f(d) \leq E[M(X)]$.
$E[M(x)] = \sum M(X)Pr\{M(X)\} \leq \frac{1}{2} * 2f(d) + 1 * 2f(d) + \frac{1}{d} * df(d) \leq Cf(d)$ for $C \geq 3$