[22.-] This was not (*) but a couple of people asked me for a solution to (b)

(a) For $i = 1, 2, \ldots, 10^6$, let X_i be the i.r.v. such that $X_i = 1$ iff the ith. ballot was not misrecorded, then $X = \sum_{i=1}^{10^6} X_i$.Then if we denote $N = 10^6$, $X_i \in B(N, p)$, with $p = 0.02$. Moreover the X_i are independent. We want to bound $\mathbf{Pr}\left[\sum_{i=1}^{N}(4/100)N\right]$. As $\mu = Np = 0.02N$ and $\delta = 1$, using Chernoff:
$\mathbf{Pr}\left[\sum_{i=1}^{N} 0.04N\right] \leq \mathbf{Pr}\left[X \geq (1+\delta)\mu\right] \leq e^{-\frac{0.02N}{3}} \sim 10^{-2895.30}$.

(b) Let X be the number of votes for candidate A that are misrecorded and let Y be the number of votes for candidate B that are misrecorded. Then, candidate B wins iff $510000 - X + Y < 490000 + X - Y$, i.e. $10000 + Y < X$. As $0 \leq X \leq 510000$ and $O \leq Y \leq 490000$, for any $0 \leq l \leq 490000$, the following holds

$$\begin{aligned}
&\mathbf{Pr}\left[10000 + Y < X\right] = \mathbf{Pr}\left[(10000 + Y < X) \cap (0 \leq Y \leq 490000)\right] \\
&= \mathbf{Pr}\left[(10000 + Y < X) \cap ((0 \leq Y \leq l) \cup (l < Y \leq 490000))\right] \\
&= \mathbf{Pr}\left[((10000 + Y < X) \cap (0 \leq Y \leq l)) \cup ((10000 + Y < X) \cap (l < Y \leq 490000))\right] \\
&= \mathbf{Pr}\left[((10000 + Y < X) \cap (0 \leq Y \leq l))\right] \cup \mathbf{Pr}\left[((10000 + Y < X) \cap (l < Y \leq 490000))\right] \\
&\leq \mathbf{Pr}\left[0 \leq Y \leq l\right] + \mathbf{Pr}\left[10000 + l < X\right] \leq \mathbf{Pr}\left[Y \leq l\right] + \mathbf{Pr}\left[10000 + l \leq X\right]
\end{aligned}$$

For $i = 1, \cdots, 510000$ and $j = 1, \cdots, 490000$ let X_i and Y_j be the indicator random variables defined as $X_i = 1$ if the ith ballot for candidate A was misrecorded, and $Y_i = 1$ if the ith ballot for candidate B was misrecorded. Then, $X = \sum_{i=1}^{510000} X_i$ $Y = \sum_{j=1}^{490000} Y_j$ and the X_i and Y_j are independent. Let $\mu_X = \sum_{1}^{510000} p_i = 0.02 \times 510000 = 10200$, $\mu_Y \sum_{1}^{490000} p_j = 0.02 \times 490000 = 9800$ and $\delta = 0.48$.Using Chernoff,
$\mathbf{Pr}\left[X \geq 10000 + 5096\right] = \mathbf{Pr}\left[X \geq (1 + 0.48) \cdot 10200)\right]$
$= \mathbf{Pr}\left[X \geq (1+\delta)\mu_X\right] \leq \exp(\mu_X \delta^2/3) \sim 10^{-340.2089}$
$\mathbf{Pr}\left[Y \leq 5096\right] = \mathbf{Pr}\left[Y \leq (1 - 0.48) \cdot 9800)\right] = \mathbf{Pr}\left[Y \leq (1-\delta)\mu_Y\right]$
$\leq \exp(-\mu_y \delta^2/2) \sim 10^{-490.3011}$
If we take $l = 5096$, we can give an UB to the probability that B wins the election owing to misrecorded ballots:
$\mathbf{Pr}\left[10000 + Y < X\right] \leq \mathbf{Pr}\left[Y \leq 5096\right] + \mathbf{Pr}\left[10000 + 5096 \leq X\right] \sim 10^{-340.2089} + 10^{-490.3011}$.

[23.-]

1. First, we process ϕ so that every variable appears at most once in each clause (eliminate repeated occurrences of a literal, and delete a clause if both a literal and its negation occur). Let n denote the number of variables, and c_i the number of variables in clause C_i.

(a) size(S_i): return 2^{n-c_i} . The variables in clause i must be fixed to values that satisfy the clause, and the remaining variables may be assigned any value, ex..: if we have 5 variables and $(\bar{x}_1 \wedge x_2 \wedge \bar{x}_3)$ then $c_i = 3$ and we must have fixed $A(x_1) = 0 = A(x_3)$ and $A(x_2) = 1$, the other 2 varibles can take all combination of 0,1, so we have $2^2 = 4$ values.

(b) select(S_i): fix the variables in clause C_i to values that satisfy the clause; choose the values of the remaining variables independently and u.a.r.

(c) lowest(x): for $i = 1, 2, \dots$ test if x satisfies C_i (this test is easy); return the index of the first clause that x satisfies (undefined if it does not satisfies no clauses).

2. The problem is that S may occupy only a tiny fraction of all possible assignments in U. Thus the number of samples t would need to be huge in order to get a good estimate of q. A concrete example to make this precise. Consider $\phi = x_1 \wedge x_2 \wedge \cdots \wedge x_n$. Then $|S| = 1$ (the only satisfying assignment is when all n variables are 1). The given algorithm will output zero unless it happens to choose this assignment in one of its t samples, i.e., it outputs zero with probability $(1/2^n)^t \to 0$ for any t that is only polynomial in n. Thus the relative error of the algorithm will be arbitrarily large with probability arbitrarily close to 1.

3. Note that the first two lines of the algorithm select each pair $(x, S_i), x \in S_i$ with probability $\frac{|S_i|}{\sum_{j=1}^m |S_j|} \cdot \frac{1}{|S_i|} = \frac{1}{\sum_{j=1}^m |S_j|}$. In other words, the first 2 lines pick an element u.a.r. from the disjoint union of the sets S_i. (We really want to pick an element u.a.r. from $\cup_i S_i$). Let $\Gamma = \{(S_i, x) | \text{lowest}(x) = i\}$. (For instance in the above example, $\Gamma = \{((1011), S_2), ((0001), S_1), \dots\})$. Therefore the algorithm outputs 1 with probability $\sum_{(S_i,x)\in\Gamma} \frac{1}{\sum_{j=1}^m |S_j|} = \frac{|\Gamma|}{\sum_{j=1}^m |S_j|}$. To see that $|\Gamma| = |S|$, simply observe that every element $x \in S$ corresponds to exactly one lowest S_i, or equivalently $\Gamma = \{(x, S_{\text{lowest}(x)}) | x \in S\}$. It follows that the algorithm outputs 1 with probability $p = \frac{|S|}{\sum_{j=1}^m |S_j|}$.

4. For $i = 1, 2, \dots, m$ we have $|S_i| \leq |S|$, so that $\sum_{i=1}^m |S_j| \leq m|S|$, so that $p = \frac{|S|}{\sum_{j=1}^m |S_j|} \geq 1/m$.

5. Note that $X_1, \dots, X_t$are independent 0-1 r.v.'s with mean p, so $\mathbf{E}[X] = pt$ and by Chernoff we get $\mathbf{Pr}[|X - pt| \geq \epsilon pt] \leq 2e^{\epsilon^2 pt/3}$. The quantity on the right is bounded above by δ provided we take $t = \lceil \frac{3}{\epsilon^2 p} \ln(2/\delta) \rceil \leq \lceil \frac{3m}{\epsilon^2 p} \ln(2/\delta) \rceil$ using the fact from part (d) that $p \geq m$. Hence it suffices to take $t = O(\frac{m}{\epsilon^2} \ln \frac{1}{\delta})$.

6. Each iteration of the algorithm in (3) requires $O(1)$ operations, so the final algorithm takes $O(t) = O(\frac{m}{\epsilon^2} \ln \frac{1}{\delta})$ time. By definition we have $|S| =$

$\frac{\sum_{j=1}^{m}|S_j|}{t} \cdot tp$ and $Y = \frac{\sum_{j=1}^{m}|S_j|}{t} \cdot X$. This implies $Y \in [(1-\epsilon)|S|, (1-\epsilon)|S|)]$ iff $X \in [(1-\epsilon)tp, (1-\epsilon)tp)]$. Therefore, $\mathbf{Pr}\left[Y \in [(1-\epsilon)|S|, (1-\epsilon)|S|]\right] = \mathbf{Pr}\left[X \in [(1-\epsilon)tp, (1-\epsilon)tp]\right]$

It follows by part (5) that $\mathbf{Pr}\left[Y \in [(1-\epsilon)|S|, (1-\epsilon)|S|]\right] \geq 1 - \delta$.