APPENDIX

In this appendix, we give detailed proofs of the theorems.

Theorem 3.5. $\mathbb{E}[gain(\hat{X})] \geq \mathbb{E}[gain(X)]$ *for any* $X \subseteq X_T$.

Proof. To prove this theorem, we first prove two lemmas. We denote the sorted sequence of elements after step (1) as $\langle \hat{s}_1, \hat{s}_2, ..., \hat{s}_m \rangle$, where $\hat{s}_i$ denotes index of the ith elements in the original subsequence X_T. We further use $S(k)$ to denote the subsequence obtained after removing k elements in step (2), i.e., removing elements at indexes $\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_k$ from X_T.

Lemma 1. $\forall X \subseteq X_T, \mathbb{E}[gain(S(|ex(X)|))] \geq \mathbb{E}[gain(X)]$.

Proof. Recall that $\mathbb{E}[gain(X)] = |ex(X)|\, \Pi_i (1-p_i)^{1-x_i}$. To compare the two expected gains, we compare the two components of an expected gain, $|ex(X)|$ and $\Pi_i(1-p_i)^{1-x_i}$, separately.

- First, it is easy to see $|ex(S(|ex(X)|))| = |ex(X)|$.
- Second, we denote the indexes of the elements in $ex(X)$ in ascending order of probabilities as $s_1, s_2, ..., s_{|ex(X)|}$. Then We have $p_{\hat{s}_1} \leq p_{s_1}, p_{\hat{s}_2} \leq p_{s_2}, ..., p_{\hat{s}_{|ex(X)|}} \leq p_{s_{|ex(X)|}}$, which implies $\prod_i (1-p_{\hat{s}_i}) \geq \prod_i (1-p_{s_i})$

Therefore, the lemma holds. □

Lemma 2. *For any* $1 \leq i < j \leq m-1$*, we have that*

$$\mathbb{E}[gain(S(j))] \cdot \mathbb{E}[gain(S(i+1))] \geq \mathbb{E}[gain(S(i))] \cdot \mathbb{E}[gain(S(j+1)]$$

.

Proof.

$$\begin{aligned}
& i < j \\
\Rightarrow & 1 - p_{\hat{s}_{i+1}} \geq 1 - p_{\hat{s}_{j+1}} \\
\Rightarrow & (ij+j)(1-p_{\hat{s}_{i+1}}) \geq (ij+i)(1-p_{\hat{s}_{j+1}}) \\
\Rightarrow & j(i+1)(1-p_{\hat{s}_{i+1}}) \geq i(j+1)(1-p_{\hat{s}_{j+1}}) \\
\Rightarrow & j\Pi_{k \leq j}(1-p_{\hat{s}_k}) \cdot (i+1)\Pi_{k \leq i+1}(1-p_{\hat{s}_k}) \\
& \geq i\Pi_{k \leq i}(1-p_{\hat{s}_k}) \cdot (j+1)\Pi_{k \leq j+1}(1-p_{\hat{s}_k}) \\
\Rightarrow & \mathbb{E}[gain(S(j))] \cdot \mathbb{E}[gain(S(i+1))] \\
& \geq \mathbb{E}[gain(S(i))] \cdot \mathbb{E}[gain(S(j+1)]
\end{aligned}$$

□

Based on the two lemmas, we prove our theorem. Let X be any subsequence such that $X \subseteq X_T$. From lemma 1 we know that $\mathbb{E}[gain(S(|ex(X)|))] \geq \mathbb{E}[gain(X)]$. Below we will show $\mathbb{E}[gain(\hat{X})] \geq \mathbb{E}[gain(S(|ex(X)|))]$.

- If $|X| \geq |\hat{X}|$, we know the above formula holds directly from step (2) of the procedure.
- Now let us consider the case where $|X| < |\hat{X}|$. From step (2), we know that

  $$E[gain(\hat{X})] = E[gain(S(|ex(\hat{X})|))] > \mathbb{E}[gain(S(|ex(\hat{X})|+1))].$$

  From lemma 2, for $\forall k > |ex(\hat{X})|$, we have

  $$\begin{aligned}
  & \mathbb{E}[gain(S(k))] \cdot \mathbb{E}[gain(S(|ex(\hat{X})|+1))] \\
  \geq & \mathbb{E}[gain(S(|ex(\hat{X})|))] \cdot \mathbb{E}[gain(S(k+1)]
  \end{aligned}$$

  Therefore,

  $$\mathbb{E}[gain(S(k))] \geq \mathbb{E}[gain(S(k+1))]$$

  for $\forall k > |ex(\hat{X})|$. As a result,

  $$\begin{aligned}
  & \mathbb{E}[gain(S(|ex(\hat{X})|))] \\
  > & \mathbb{E}[gain(S(|ex(\hat{X})|+1))] \\
  \geq & \ldots \\
  \geq & \mathbb{E}[gain(S(|ex(X)|))]
  \end{aligned}$$

Putting the above together, we know that the original theorem holds. □

Theorem 4.1. *Given input with size n, the asymptotic number of tests performed by ProbDD is bounded by O(n) in the worst case.*

Proof. Recall the updating rules in Section 3.2, in each iteration, we perform one test,

- If test passes, at least one element's probability will be set to 0.
- If test fails, at least one element's probability will be increased by $\frac{p_i}{1-\Pi_j(1-p_j)^{1-x_j}} - p_i = \frac{p_i * \Pi_j (1-p_j)^{1-x_j}}{1-\Pi_j(1-p_j)^{1-x_j}}$

In order to show the total number of tests is bounded by O(n), we consider the above two cases separately.

First, the operation to set the probability to 0 can be performed no more than n times as there are n elements in total.

Second, we will show the number of possible probability increases is also bounded by O(n).

Using the notations from the proof of Theorem 3.5, below we prove three lemmas first.

Lemma 3. $|ex(\hat{X})| \leq \frac{1}{\sigma}$*, where* σ *is the initial probability,* $0 < \sigma < 1$.

Proof. First, if $|ex(\hat{X})| = 1$, then $1 < \frac{1}{\sigma}$ is satisfied.

Second, if $|ex(\hat{X})| \geq 2$,

$$\begin{aligned}
& \mathbb{E}[gain(S(|ex(\hat{X})|))] \geq \mathbb{E}[gain(S(|ex(\hat{X})|-1))] \\
\Rightarrow & (1 - p_{|ex(\hat{X})|}) * |ex(\hat{X})| \geq |ex(\hat{X})| - 1 \\
\Rightarrow & |ex(\hat{X})| \leq \frac{1}{p_{|ex(\hat{X})|}}
\end{aligned}$$

So $|ex(\hat{X})| \leq \frac{1}{\sigma}$ as $p_{|ex(\hat{X})|} \geq \sigma$. □

Lemma 4. *The number of probability increases is bounded by O(n) before the probabilities of all elements become zero or larger than 0.5.*

Proof. We will first show that the probability increase has a lower bound. Let us assume that there exist elements whose probabilities are larger than 0 and smaller than or equal to 0.5. Let the smallest probability among them be $0.5 - \epsilon$, where $\epsilon \geq 0$.

Then from Lemma 3,

$$\begin{aligned}
& \mathbb{E}[gain(\hat{X})] \geq E[gain(S(1))] = (1 - (0.5 - \epsilon)) * 1 \geq 0.5 \\
\Rightarrow & \Pi_i (1-p_i)^{1-x_i} \geq \frac{0.5}{|ex(\hat{X})|} \geq 0.5\sigma
\end{aligned}$$

As a result, the probability increment

$$\begin{aligned} & \frac{p_i * \Pi_j (1-p_j)^{1-x_j}}{1-\Pi_j (1-p_j)^{1-x_j}} \\ \geq & \frac{\sigma}{\frac{1}{\Pi_j (1-p_j)^{1-x_j}} - 1} \\ \geq & \frac{\sigma}{\frac{1}{0.5\sigma} - 1} \end{aligned}$$

So if at least one element has a positive probability less than or equal to 0.5, the minimal probability increment is $\Delta = \frac{0.5\sigma^2}{1-0.5\sigma}$, which is a constant. As a result, the maximal number of probability increases before the probabilities of all elements become zero or large than 0.5 is at most $\frac{0.5n}{\Delta} = O(n)$. □

Lemma 5. $\mathbb{E}[S(1)] > \mathbb{E}[S(2)]$ *when* $p_i > 0.5 \vee p_i = 0$ *for any* i.

Proof. Let p_a be the smallest positive probability and p_b be the second smallest positive probability. Then we have $\mathbb{E}[S(1)] = (1-p_a) > 2(1-p_b) * (1-p_a) = \mathbb{E}[S(2)]$ □

As a result, from Lemma 4, it takes O(n) probability increases before the probabilities of all remaining elements are larger than 0.5. From Lemma 5, we know the algorithm removes one element for tests at a time after the probabilities of all remaining elements are larger than 0.5. Since a failed test with one removed element will set the probability of this element to 1, we know there is at most $O(n)$ probability increases after the probabilities of all remaining elements are larger than 0.5. Therefore, the number of probability increases is bounded by O(n).

Putting the above together, we know that the original theorem holds. □